Dirichlet–Dirichlet Domain Decomposition Methods for Elliptic Problems

h and hp Finite Element Discretizations

Vadim Glebovich Korneev

Ulrich Langer

World Scientific

Dirichlet–Dirichlet Domain Decomposition Methods for Elliptic Problems

h and hp Finite Element Discretizations

Domain decomposition (DD) methods provide powerful tools for constructing parallel numerical solution algorithms for large scale systems of algebraic equations arising from the discretization of partial differential equations. These methods are well-established and belong to a fast developing area. In this volume, the reader will find a brief historical overview, the basic results of the general theory of domain and space decomposition methods as well as the description and analysis of practical DD algorithms for parallel computing. It is typical to find in this volume that most of the presented DD solvers belong to the family of fast algorithms, where each component is efficient with respect to the arithmetical work. Readers will discover new analysis results for both the well-known basic DD solvers and some DD methods recently devised by the authors, e.g., for elliptic problems with varying chaotically piecewise constant orthotropism without restrictions on the finite aspect ratios.

The *hp* finite element discretizations of elliptic equations, in particular, by spectral elements, are given significant attention in current research and applications. This volume is the first to feature all components of Dirichlet–Dirichlet-type DD solvers for *hp* discretizations devised as numerical procedures which result in DD solvers that are almost optimal with respect to the computational work. The most important DD solvers are presented in the matrix/vector form algorithms that are convenient for practical use.

World Scientific
www.worldscientific.com
8382 hc

ISBN 978-981-4578-45-5
9 789814 578455

Dirichlet–Dirichlet Domain Decomposition Methods for Elliptic Problems

h and hp Finite Element Discretizations

Vadim Glebovich Korneev
St. Petersburg State University, Russia
St. Petersburg State Polytechnical University, Russia

Ulrich Langer
Johannes Kepler University Linz, Austria
Johann Radon Institute for Computational and Applied Mathematics,
Austrian Academy of Sciences, Linz, Austria

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

Library of Congress Cataloging-in-Publication Data

Korneev, V. G. (Vadim Glebovich), 1937–

Dirichlet-Dirichlet domain decomposition methods for elliptic problems h and hp finite element discretizations / Vadim Glebovich Korneev, St. Petersburg State University, Russia, St. Petersburg State Polytechnical University, Russia, Ulrich Langer, Johannes Kepler University Linz, Austria.

pages cm

Includes bibliographical references and index.

ISBN 978-9814578455 (hardcover : alk. paper)

1. Decomposition (Mathematics) 2. Differential equations, Partial 3. Finite element method I. Langer, Ulrich, 1952– II. Title.

QA402.2.K67 2015

515'.3533--dc23

2014040541

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

Printed in Singapore by B & Jo Enterprise Pte Ltd

Dedicated to our parents
Ekaterina and Gleb Korneev,
Gertraude and Gottfried Langer

Preface

Domain decomposition (DD) methods provide powerful tools for constructing parallel numerical solution algorithms for large scale systems of algebraic equations arising from the discretization of partial differential equations. The Alternating Schwarz Method, proposed by H. Schwarz (1869) in order to prove the existence of harmonic functions with prescribed Dirichlet data on the boundary of complicated domains, and substructuring techniques, developed by engineers in the 60s of the preceding century in order to create faster procedures for the analysis of complex structures, are commonly accepted as the origins of modern DD methods. Thus, DD methods have been developed for a long time, but most extensively since the first international DD conference that was held at Paris in 1987. This concerns both the theory and the practical use of DD techniques for creating efficient application software for massive parallel computers. The advances in DD are well documented in the proceedings of the international DD conferences[1] since 1987 and numerous papers. The first textbook on DD methods was published by B.F. Smith, P.E. Bjørstad, and W. Gropp in 1996. In 1999, A. Quarteroni and A. Vali published a monograph on "Domain Decomposition Methods for Partial Differential Equations" that provides a deep analysis of various domain decomposition methods. The authors contributed chapter 22 on "Domain Decomposition Methods and Preconditioning" to the Encyclopedia of Computational Mechanics (2004) in which they summarize the state of the art until 2003. Since then many papers on DD techniques and their use in Scientific Computing have been published. In particular, the textbook by A. Toselli and O. Widlund (2005) is now the standard reference. Another textbook, covering many interesting applications of DD methods, was published by T.P.A. Mathew in 2008. C.

[1] `http://www.ddm.org/conferences.html`

Pechstein has recently published a monograph on Finite Element Tearing and Interconnecting (FETI) and Boundary Element Tearing and Interconnecting (BETI) methods for a special class of multiscale problems. Our book is different from the textbooks mentioned above and very different from Pechstein's monograph. We mainly discuss one special class of primal substructuring methods also called Dirichlet–Dirichlet DD methods, and we emphasize the peculiarities of their application to *hp* finite element equations. In particular, we discuss and analyze the inexact versions of these DD methods which lead to optimal or, at least, almost optimal complexity and high efficiency in practical applications. The optimization of several important components of DD algorithms for *hp* discretization was achieved only quite recently. These topics are not discussed or at least not sufficiently discussed in the books mentioned above. The contributions of the authors to this field mainly appeared in journal publications including joint papers. The reader will become familiar with inexact Dirichlet–Dirichlet DD methods enjoying optimal complexity and the techniques for the numerical analysis of these methods. Thus, we hope that readers interested in both practice and theory can benefit from our book. In particular, the book will open the possibility for the reader to use such DD methods in new fields of applications like multiphysics applications in Computational Sciences.

In this book, the reader can find a brief historical overview, the basic results of the general theory of domain and space decompositions as well as the description and analysis of practical DD algorithms. Elliptic problems with strongly jumping coefficients are daily met in the engineering practice. Numerical techniques should not deteriorate in efficiency when solving such problems. Partly for this reason, we concentrate on *h* and *hp* finite element discretizations of elliptic boundary value problems and their solution by Dirichlet–Dirichlet-type DD methods. Considerable attention is paid to DD methods or preconditioners for *hp* discretizations, significant advances in the development of which having been made in the last decades. The book will add new features to the understanding of the beauty and powerfulness of DD methods, and supply the reader with a variety of modern DD algorithms. More precisely, this book contains 9 chapters where we discuss the following topics.

In the introductory **Chapter 1**, we present a brief historical retrospective of the development of modern DD methods, primarily of the Dirichlet–Dirichlet types, and discuss its two origins, which are not simply artifacts, but retained their significance in our time. One of these origins is the

iterative procedure introduced by H. Schwarz in 1869. It is called now Alternating Schwarz Method, and has led to various modern additive and multiplicative overlapping domain decomposition methods. We trace advances to two other classes of additive domain decomposition methods and methods with nonoverlapping subdomains of decomposition and, eventually, to the period of fast development of variety of DD techniques and their theory, started in the middle of the 80s. In the 60s, engineers developed Gaussian elimination algorithms, called substructuring algorithms, which is commonly considered as another origin of DD approaches. The "block scheme" of these algorithms is closely reflected, and received much more powerful filling in the subclass of Dirichlet–Dirichlet DD methods with nonoverlapping subdomains.

Chapter 2 is devoted to the *Fundamentals of Schwarz' methods.* The look at the DD method as a method based on the problem space decomposition played indispensable role in the creation of its theoretical foundations. Another feature of the modern view of DD method is that it amounts to the use of special DD preconditioners in frames of known iterative methods like the Preconditioned Conjugate Gradient (PCG) method. Apart from good relative condition numbers, such preconditioners should suggest cheap solvers for the systems of algebraic equations with these preconditioner as system matrices. For this reason, they are often termed as preconditioners-solvers. These basic facts and some of their consequences are illuminated with the use of two model problems, from which one is the scalar elliptic equation with the diagonal matrix of coefficients and the other is the system of linear elasticity equations. Convergence estimates for basic versions of DD method are formulated in terms of conditions imposed on the energy space decomposition and subspace preconditioning bilinear forms. In the modern literature, these conditions imposed on the decomposition are called stability conditions.

Chapter 3 deals with *Overlapping Domain Decomposition Methods.* A good illustration of usefulness of the FE space decomposition technique for the analysis of efficiency of DD preconditioners are overlapping DD methods. The sizes of overlap have two opposite effects: growth of the overlap improves the relative condition number of the DD preconditioner, but increases the computational overhead. We present estimates, sharp in some aspects, for the relative spectrum bounds explicitly depending on the size of overlap. They show that, for sufficiently regular domain decompositions, the relative condition number is bounded by a generic constant provided that the overlap is of the order of the diameters of the subdomains.

Chapter 4 considers *Nonoverlapping DD Methods for h FE Discretizations in 2d.* The structure of nonoverlapping DD methods of the Dirichlet–Dirichlet type, especially in 3d case, is more complex in comparison with the overlapping DD methods. In this chapter, this is illustrated for h discretizations of 2d elliptic problems and includes the discussion of the three main components composing such DD preconditioner. They are the local preconditioner-solvers for Dirichlet problems on subdomains of decomposition, preconditioner-solver for the interface subproblem, governed by the interface Schur complement matrix, and prolongation operators from inter-subdomain boundaries inside the subdomains. At present, all these components are rather well developed, and there is a number of good options for each. A few of them, which yield DD solvers of linear or almost linear computational complexity are presented in this chapter.

Chapter 5 is devoted to *BPS-type DD Preconditioners for 3d Elliptic Problems.* In fact, we consider the DD preconditioner of Bramble, Pasciak and Schatz, which is now usually referred to as BPS preconditioner in the literature. In this chapter, it is investigated under more general conditions, imposed on the domain decompositions, their subdomains and their FE discretizations, than in the renowned series of papers by J.H. Bramble, J.E. Pasciak and A.H. Schatz. In the analysis of convergence, we tried to follow these papers, where it was possible, although in several instances, it was necessary to change their proofs, in order to adapt them to a more general situation. The use of BPS-type preconditioner in PCG iterative procedure results in very efficient solvers, convenient for parallelization and incorporating fast solvers of other types in its main components. For sufficiently regular problems and triangulations the losses in the relative condition number are not significant, and there is a number of such choices for component solvers that in a whole DD solver will possess almost linear numerical complexity.

In **Chapter 6**, we consider *DD Algorithms for Discretizations with Chaotically Piecewise Variable Orthotropism.* One class of problems among the mentioned above, which is studied in this chapter, often met in practice and cause significant difficulties at numerical solution. A representative 2d model problem is described by the equation $-\nabla \cdot \boldsymbol{\rho} \nabla u = f$ with a positive diagonal coefficient matrix $\boldsymbol{\rho}$. Computational difficulties of implementation of DD method are caused by three factors. The coefficients on the diagonal of the matrix $\boldsymbol{\rho}$ are arbitrary positive numbers, different for each subdomain. The decomposition mesh is a nonuniform rectangular mesh with changing arbitrarily mesh sizes. The finite element mesh is a nonuniform rectangular mesh satisfying only one condition that it is uniform on

each subdomain. For a rectangular domain, we present a DD solver, which under the assumption of a fixed number of subdomains is optimal with respect to the arithmetical work. More precisely, it requires $\mathcal{O}(N)$ arithmetic operations, where N denotes the number of unknowns. The solver remains optimal, if the number of subdomains grows with N, but not too fast. The key problem in constructing efficient DD algorithm for the described discretization appears in preconditioning of the Schur complement, arising after elimination of the internal degrees of freedom for the subdomains of the decomposition. It interferes with the problem of deriving boundary norms for harmonic functions on slim domains and related problem of boundary norms for discrete harmonic functions corresponding to orthotropic differential operators on slim domains, discretized by means of orthotropic meshes. These problems are studied for one subdomain in Section 7.1. It is worth noting, that there are many suggestions in the literature on the efficient preconditioning of the boundary Schur complement for the tensor product orthotropic discretizations on slim rectangles. However, their use for assembling the inter-subdomain boundary Schur complement preconditioner is problematic for the reason of their incompatibility of definite sort for the described domain decompositions. Therefore, special way of constructing such a Schur complement preconditioner must be implemented, which is presented in Section 6.4.

Chapter 7 is devoted to *Nonoverlapping DD Methods for hp Discretizations of 2d Elliptic Equations*. Two-dimensional elliptic problems are quite good examples for the introduction of basic techniques employed at the derivation of nonoverlapping DD methods for more general *hp* discretization. Two typical structures of DD preconditioner-solvers, resembling structures of the DD preconditioner-solvers for h discretizations in Chapter 4 are discussed. In one of them, the vertex unknowns are split from the others in the DD preconditioner. In the other one, which can be more efficient, an inexact iterative solver is used for approximation of the Schur complement related to the whole (global) interface problem. We derive the bounds for the relative condition numbers and discuss efficient components for the both versions.

Chapter 8 deals with *Fast Dirichlet Solvers for 2d Reference Elements*. Fast solvers for p reference elements stiffness and mass matrices are of crucial importance for the performance of DD solvers for *hp* finite element discretizations. To the best of the authors' knowledge, all such solvers were obtained on the basis of finite-difference preconditioners (equivalently, finite element preconditioners induced by first order finite elements) for the

stiffness and mass matrices. For both the hierarchical-type and the spectral-type square reference elements, known solvers of linear and almost linear complexity can be divided into three groups: DD type methods, algebraic multigrid methods, and multilevel wavelet methods, *i.e.*, based on multilevel wavelet decompositions of the first order FE space. We consider representatives of all of them. Note that the operator extrapolation technique allows us to obtain fast multilevel wavelet solvers for local problems on faces, arising, *e.g.*, in Dirichlet–Dirichlet DD methods for 3d elliptic equations.

Chapter 9 is devoted to *Nonoverlapping Dirichlet–Dirichlet DD Methods for hp Discretizations of 3d Elliptic Equations.* Nonoverlapping Dirichlet–Dirichlet methods for *hp* discretizations of 3d elliptic problems are much more complex than their 2d counterparts. The chapter starts from the discussion of the general structure of DD solver, common features of main components and their interplay. Results on the relative condition number of DD preconditioner and its numerical complexity are formulated first in terms of general properties of precondioners-solvers and prolongations incorporated in the components. After that, a few specific solvers and prolongations for each component are formulated and studied. We also touch on some special cases, for instance, when discretizations possess features that are typical for the *hp* adaptivity and when incomplete finite elements are used. The presented solvers and prolongations lead to Dirichlet–Dirichlet DD preconditioner-solvers of almost linear complexity of the preconditioning operations for general second order elliptic equations in arbitrary sufficiently smooth domains with subdomainwise smooth coefficients. It is necessary to stress that the numerical complexity of preconditioning means the arithmetical cost of the PCG with this preconditioner, except for the cost required for the matrix-vector multiplications by the matrix of the finite element system of equations to be solved. The cost of such matrix-vector multiplications as well as the cost of computation of the finite element system stiffness matrix for the *hp*-version can be considerable and deserve a special study not attended in this book.

The authors would like to thank many friends and colleagues, whose interest, support and encouragement made this book possible. Special thanks go to M.M. Karchevskii for reading the manuscript at an early stage and for numerous useful suggestions for improving the text. Almost all numerical experiments, the results of which are included in the book, were performed by I.E. Anufriev whom we owe our sincere gratitude. We had many fruitful discussions with S. Beuchler, whose personal contributions

to fast solvers for *hp*-version found significant reflection in this book. The idea to write this book was born during our first Research-in-Pairs (RiP) stay at Mathematical Research Institute of Oberwolfach (MFO) in 2002 when we were working on our contribution *"Domain Decomposition Methods and Preconditioning"* to the Encyclopedia of Computational Mechanics that was published by John Wiley & Sons in 2004. Our second RiP stay in August 2013 enabled us to finish the work on this book. We are very grateful to the MFO for supporting us through the RiP programme and for providing us with a perfect research environment. We would also like to thank the Johann Radon Institute for Computational and Applied Mathematics (RICAM) of Austrian Academy of Sciences that supports the visits of the first author at Linz. A great part of the research contributed to the book by the first author was completed under support of the Russian Fund of Basic Research, grants 08-01-00676-a and 11-01-00667-a. The second author was partly supported by the National Research Network *"Geometry and Simulation"* of the Austrian Science Fund FWF under the grant NFN S11703.

St. Petersburg and Linz, February 2014 *V.G. Korneev and U. Langer*

Contents

Chapter 1

Introduction

1.1 Dirichlet–Dirichlet Domain Decomposition Methods in Retrospect

For quite some time, engineers subdivide structures, which they want to analysis or design, into substructures of simpler geometries and mechanical properties, then perform a separate analysis, and finally predict properties and behavior of the whole structure by some process of synthesis. This general procedure has found a good reflection in the approach to the analysis of the Boundary Value Problems (BVP) for Partial Differential Equations (PDE) and their numerical solution that is called now *Domain Decomposition method*, or, for brevity, DD method. The first mathematical formulation of a DD approach in the form of an *overlapping domain decomposition method* was done by [Schwarz (1869)]. It appeared in his famous paper on the existence of harmonic functions in domains of special type. Namely, he assumed that the domain could be represented by a union of two generously overlapping simpler subdomains in the sense that the existence of harmonic functions was known for each of them beforehand. The proof was based on the maximum principle and an iterative procedure, which is now called *Alternating Schwarz Method* (AltSM). The variational formulation of the AltSM was introduced by [Sobolev (1936)] almost one hundred years later. The variational setting allowed him to prove convergence in the energy space and for a wider class of domains and problems. Further pioneering contributions by [Mikhlin (1951)] and [Babuška (1958)] widened the range of problems, to which DD method could be efficiently applied and for which convergence in Sobolev spaces or pointwise can be ensured. An interpretation of DD methods as projection methods was given by [Morgenstern (1956)].

A further important step was the introduction of *nonoverlapping domain decomposition* algorithms, first studies of which can be traced back to [Katsnelson and Menshikov (1973)] and [Tsvik (1975)]. The role of the Steklov-Poincare operators in the analysis of nonoverlapping DD algorithms was shown in the book of [Lebedev and Agoshkov (1983)].

To the best of the authors' knowledge, early proofs of convergence of Schwarz algorithm, based on the discrete maximum principle, appear in [Miller (1965)] and [Starius (1977)] for finite-difference discretizations of elliptic equations. They were followed by studies of nonoverlapping algorithms for finite element approximations to elliptic boundary value problems by [Matsokin (1980)], [Smelov (1980)], [Matsokin and Nepomnyaschikh (1981)], and [Osmolovskii and Rivkind (1981)]. The alternating Schwarz method as well as many later variants, including those with nonoverlapping subdomains, fall in the category of multiplicative iterative procedures. In the 80s, new versions, which are more suitable for engineering applications and parallelization, appeared, see [Matsokin and Nepomnyaschikh (1985)], for which [Dryja and Widlund (1987)] apparently first used the notion *additive domain decomposition method* (ASM). A great influence on the promotion of DD methods must be attributed to [Lions (1988, 1989, 1990)]. He paid much attention to this method and to the creation of its general abstract theory based on stability properties of the energy space decomposition corresponding to underlying domain decomposition. In the pioneering works of [Matsokin and Nepomnyaschikh (1985)], basic general necessary and sufficient conditions for the convergence of DD methods together with rate estimates were formulated in a Hilbert space decomposition setting.

Although DD methods were recognized as a very efficient tool for solving large-scale finite-difference or finite element equations arising from the discretization of partial differential equations much earlier, the most intensive development of algorithms and their theory started in the middle of the 80s, when powerful computers became more accessible for engineering and scientific computations. The series of four papers by [Bramble *et al.* (1986, 1987, 1988, 1989)] can be considered as a landmark between the long period of formation of basic ideas of DD method and the period of an intensive development of a variety of efficient DD algorithms for a wide range of problems and the perfect tool kit of their analysis. In particular, in these papers, the main components of nonoverlapping DD algorithms for 2d and 3d elliptic equations were introduced and the influence of their choice on the relative condition number of the global DD preconditioner

was estimated as $\mathcal{O}(1 + \log^2 H/h)$. Here H/h is the maximal relation of the characteristic size H of a subdomain of decomposition to the mesh parameter h of its discretization. The DD preconditioners developed by these authors are commonly referred as BPS preconditioner.

In the BPS domain decomposition preconditioner-solver for 3d, the main components are the solvers (i) for the Dirichlet problems on subdomains of decomposition, (ii) for the internal problems in the faces, and (iii) for the problem on the wire-basket of the decomposition, and, besides, two prolongation operators, (iv) one from the subdomains boundaries to the subdomains, and (v) one from the wire-basket to subdomains boundaries. Further studies of the Dirichlet–Dirichlet DD preconditioner-solvers were aimed at perfection of their components, expansion to wider and special classes of boundary value problems, and better understanding of relations with other efficient preconditioners. Significant contributions to the development and thorough analysis of such DD preconditioners for the discretizations of the second-order elliptic equations in 3d by tetrahedral linear finite elements are related to [Dryja *et al.* (1994)] and [Toselli and Widlund (2005)]. They assumed that unions of finite numbers of tetrahedra of the coarse imbedded shape regular tetrahedral mesh served for subdomains of decomposition under some natural conditions on intersections of the subdomain faces and edges. The triangulation of each subdomain was required to be shape and size regular. The same as the given above estimate was proven for the relative condition number of the inter-subdomain Schur complement preconditioner.

The set of structural units, different from preceding studies, was used in [Korneev (2012)] for generating subdomains of decomposition for the BPS type preconditioner. By assumption, it contained a finite number of the reference convex polyhedrons, which had plain faces, in definite sense were shape regular, and had diameters equal to unity. Each subdomain of decomposition was supposed to be an image of one of the reference polygons by a sufficiently smooth and properly scaled mapping, which provided compatibility of subdomains. The proof of the bound $\mathcal{O}(1 + \log^2 H/h)$ for the relative condition number of the BPS preconditioner followed the path of [Bramble *et al.* (1989)], but required a new version of some technical tools.

In order to have a fast DD preconditioner-solver, it is not sufficient to provide an adequate structure and good component preconditioners, which result in a good relative condition number. There must exist fast solution procedures for the systems of algebraic equations with component

preconditioners for the system matrices. Designing such preconditioner-solvers for the interface (inter-subdomain boundary) Schur complement has generally proved to be most difficult, but not so for second-order elliptic problems in 2d. For the latter problems first suggestion of this sort came from [Dryja (1982, 1984)]. In particular, for any single edge block in the Schur complement interface preconditioner, he approved the use of the capacitance matrix. In the simplest case, up to the multiplier, it is the square root of the matrix of the finite element approximation of the second derivative operator on the uniform grid. Clearly, fast discrete Fourier transform is a fast solution procedure for such systems of equations. Dryja's preconditioner was later improved by [Golub and Mayers (1984)] and [Bjørstad and Widlund (1986)], and [Chan (1987)].

The transformation of the nodal basis to a multilevel basis was very successful in the creation of preconditioners for the finite element stiffness matrices and was expanded to the Schur complement preconditioning. The same transformations, as, *e.g.*, for the hierarchical Yserentant and the BPX preconditioners, but restricted to the interface nodes result in Schur complement preconditioners possessing at least the same quality as the corresponding preconditioners for the stiffness matrix. We refer here to [Smith and Widlund (1990)] and [Haase *et al.* (1991)] who proposed the hierarchical Schur complement preconditioners that behave like the BPS preconditioner in the 2d case. An asymptotically optimal, but sensitive to coefficient jumps BPX Schur complement preconditioner was introduced by [Tong *et al.* (1991)].

Some other advances in designing interface Schur complement preconditioner-solvers should be mentioned. The ways of splitting the face and the wire basket components in such preconditioner-solvers were studied in [Bramble *et al.* (1989)] and [Dryja *et al.* (1994)], where some solution procedures for the wire basket component were also considered alongside with simple and efficient algorithms of prolongations from the wire basket. Techniques borrowed from the boundary element method (see, *e.g.*, [Carstensen *et al.* (1998)], [Haase *et al.* (1997)] and [Steinbach (2003)]) represent an additional instrument for constructing interface Schur complement preconditioner-solvers.

A remarkable property of BPS preconditioners is their robustness. The same bound of the relative condition number holds, if the coefficients of the elliptic equation have strong jumps on the inter-subdomain surfaces. For example, if the coefficient matrix contains only main terms and, on each subdomain, is close to $c_j\mathbf{I}$, where $\mathbf{I}$ is the unity matrix and c_j is

the constant relative to a subdomain Ω_j. Thus, the jumps only affect the computational properties via the wire basket component, which has the smallest dimension.

Robustness was especially difficult to provide for the DD solver for discrete problems studied by [Khoromskij and Wittum (1999, 2004)], [Kwak *et al.* (2004)], [Korneev *et al.* (2007)] and [Korneev (2013a)]. They considered the elliptic equation in the 2d domain composed of a number of nests of the variable arbitrarily orthogonal mesh, at the same time playing the role of the decomposition mesh. The energy integral, induced by the differential operator, contains only squares of first derivatives with the coefficients, which are arbitrary positive numbers on each subdomain. The subdomain-wise variable orthogonal mesh of the bilinear finite elements satisfies only one condition to be uniform on each subdomain. Hence, the discrete problem is characterized by the orthotropism, varying chaotically subdomain wise with no restrictions on the aspect ratios. [Korneev (2013a)] succeeded in obtaining the DD Dirichlet–Dirichlet solver of linear complexity. The main trick was preconditioning of the inter-subdomain Schur complement implicitly by means of two preconditioners: preconditioner-multiplicator and the preconditioner-solver. The latter comes back to [Khoromskij and Wittum (1999)], the former basically was designed in [Korneev *et al.* (2007)] with the use of the new boundary norms derived there for harmonic and discrete harmonic functions in slim rectangles.

Nowdays, there is a variety of efficient exact and iterative solvers which can be used for the internal Dirichlet problems on subdomains of decomposition. Among the latter, it is sufficient to mention here the geometric and algebraic multigrid methods and the conjugate gradient method with the BPX or MDS preconditioners, which are very general, for many problems have linear complexity and are widely realized in the software, see, *e.g.*, [Bramble (1993)], [Bramble and Zhang (2000)], [Hackbusch (1994, 2004)], [Axellson (1994)] and [Shaidurov (1995)]. The important advance was to understand that they can be efficient when used as inexact solvers. Numerical experiments of [Haase *et al.* (1990, 1991)] and [Smith (1993)] demonstrated that even one multigrid V-cycle was sufficient for a good rate of convergence of solutions by DD solver. [Dyakonov (1989)] and [Nepomnyaschikh (1991a)] contributed theoretical justification of the efficiency of inexact solvers for preconditioning. However, it was proved in [Börgers (1989)] and [Haase *et al.* (1991)] that one needs at least $O(\ln(H/h))$ V-cycles, if multigrid method is applied for low energy prolongations. In a whole, components of the DD solvers for h discretizations of the 2d elliptic

second-order equations are much more thoroughly elaborated, than for *hp* discretizations. There is a wider variety of efficient iterative solvers for the internal Dirichlet problems on subdomains, only slightly less efficient at the use for prolongations inside finite elements from their boundaries and in opposite restriction operations. Besides, a number of prolongation operators, which do not assume solution of any systems of algebraic equations, have been developed, see, *e.g.*, [Matsokin and Nepomnyaschikh (1985)], [Nepomnyaschikh (1991b)] and [Haase *et al.* (1994)].

At present, DD methods are viewed as techniques for constructing special DD preconditioners for some iterative procedures such as the Preconditioned Richardson Method, the Preconditioned Conjugate Gradient Method (PCGM), or other preconditioned variational three-level or two-level iterative method. The quality of a DD preconditioner depends on several factors among which two obviously contradicting factors are most important: the relative condition number provided by the DD preconditioner, and the arithmetic cost of solving systems of algebraic equations with the preconditioner for the matrix. Suppose the relative condition number is good, in particular, bounded by a constant uniformly in the mesh parameter h and the order p of finite elements. Suppose, for the system of algebraic equations with the preconditioner as system matrix, there is a solver which is almost optimal in the computational cost. Then for shortness it is called *fast DD preconditioner-solver*. Numerous papers have been published on perfection of general computational properties of DD methods and deriving fast DD preconditioner-solvers for discretizations of various specific boundary value problems difficult for numerical solution. Contemporary literature on DD methods is extremely vast. We refer to the survey papers by [Xu (1992)], [Chan and Mathew (1994)], [Farhat and Roux (1994)], [Tallec (1994)], [Xu and Zou (1998)], [Korneev and Langer (2004)] and [Langer and Steinbach (2007)] as well as to the books of [Smith *et al.* (1996)], [Quarteroni and Vali (1999)], [Toselli and Widlund (2005)], [Mathew (2008)] and [Pechstein (2013)] for extensive references and good documentation of advances in this field. The proceedings of the 22 International DD conferences are also a reliable source for concentrated information on research in this area. We refer the reader to the DD website[1] where one can find all DD proceedings and other publications on DD methods.

The development of fast DD preconditioner-solvers for *hp*-version of the finite element method for elliptic boundary value problems is among

[1] `http://www.ddm.org/conferences.html`

directions with significant successes in the last two decades. Two popular in the numerical practice types of such *hp* discretizations are obtained by means of finite elements associated with the *spectral reference elements* and the reference elements, which we refer to as *hierarchical*, both specified on the reference cube $\tau_0 = (-1, 1)^3$. We use for the reference elements of these two types the notations $\mathcal{E}_{Sp}$ and $\mathcal{E}_H$, respectively, and $\mathcal{E}_{Sp} = \mathcal{E}_{GLL}, \mathcal{E}_{GLC}$ for the spectral reference elements with the nodes being the tensor products of the Gauss-Lobatto-Legendre and Gauss-Lobatto-Chebyshev quadrature nodes.

Polynomial spectral methods were introduced by [Gottlieb and Orszag (1977)], the description of hierarchical elements, sometimes also termed *hp* elements in the literature, may be attributed to [Peano (1975, 1976, 1979)], see also [Szabo and Babuška (1987)]. As is well known now, in many situations, *hp* discretrizations are able to provide high, *e.g.,* exponential, rates of convergence of finite element solutions with respect to the number of d.o.f. Studies and discussion of this advantageous property may be found in [Schwab (1998)], [Karniadakis and Sherwin (1999)] and [Babuška and Strouboulis (2001)]. However, gains resulting from it can be compromised by a solver for the corresponding systems of algebraic equations, which computational cost is usually higher that for the *h*-version. Indeed, as a rule, general solvers for these systems are costly. These were the reasons that the problem of creating low cost solvers for the systems of algebraic equations generated by *hp* discretizations was paid much attention in the literature. Typical structures of the Dirichlet–Dirichlet DD preconditioners have much in common for *hp* and *h* versions and depend primarily on the structure of the inter-subdomain Schur complement preconditioner. At that, domains of finite *p*-elements are reasonable to accept for subdomains of domain decomposition.

For *hp* discretization of 2d elliptic equations, [Babuška *et al.* (1991)] studied the DD Schur complement preconditioner in which the vertex d.o.f. were split from the rest d.o.f. and, besides, d.o.f. living on one edge are split from d.o.f. of the rest edges. They proved the bound $\mathcal{O}(1 + \log^2 p)$ for the relative condition number. The Dirichlet–Dirichlet DD preconditioners for 3d elliptic equations assume the representation of the finite element space by the direct sum of the subspaces induced by d.o.f. internal for each subdomain of decomposition, by d.o.f. internal for each face and the rest d.o.f. which live on the set termed the *wire basket of the domain decomposition*. In the reflecting this structure DD inter-subdomain Schur

complement preconditioner, the wire basket d.o.f. are split from the rest d.o.f., and additionally d.o.f. of each face are split from other faces. The relative condition number of such preconditioner was one of the subjects studied by [Pavarino and Widlund (1996)] and [Casarin (1997)], who proved the same bound $\mathcal{O}(1+\log^2 p)$ for this more complicated case. The ways of handling the wire basket subproblem, incorporated in the inter-subdomain Schur complement preconditioner, were also discussed in these works.

As in the case of the h-version, good relative condition number alone does not guarantee a success at the use of the preconditioner at iterative solution. It happens only if a good relative condition number is accompanied by a fast solution procedure for the system of algebraic equations with the preconditioner for the matrix. As a rule, the biggest contribution to the computational work is produced by the solver for the local discrete Dirichlet problems on the subdomains of the decomposition. At known arrangements of DD solvers, these preconditioners influence also other components of DD solvers, namely interface Schur complement preconditioning and prolongations. Under condition of the shape regularity of finite p-elements, preconditioning the local Dirichlet problems is reduced to the preconditioning the Dirichlet problems governed by the internal stiffness matrices of the reference p-elements. The search for efficient preconditioners for stiffness matrices of the reference p-elements started long ago, but resulted in a few fast preconditioner-solvers only in this century. All such fast solvers, known to the authors, are based on the finite-difference or low order finite element preconditioners.

Low order finite element preconditioners for stiffness matrices of spectral reference elements appeared in [Orszag (1980)] and [Deville and Mund (1985)]. Multiplication by some diagonal matrix transforms them into the finite-difference matrices, which will be called finite-difference preconditioners. Therefore, a solution procedure applicable to a finite-difference preconditioner is as well applicable to the respective finite element preconditioner and vice versa. For the stiffness and mass matrices of the hierarchical reference elements $\mathcal{E}_H$ finite-difference preconditioners were introduced in [Ivanov and Korneev (1996)], see additionally [Korneev and Jensen (1997, 1999)]. The two types of preconditioners mentioned for the two respective types of the reference elements $\mathcal{E}_{Sp}$ and $\mathcal{E}_H$ are substantially distinct. In the case of spectral elements $\mathcal{E}_{Sp}$, they are true finite-difference approximations of the Laplace operator on the nonuniform orthogonal grid of the spectral element nodes. A direct search for simple efficient preconditioners of stiffness matrices of the hierarchical elements resulted in the preconditioners

of another type, which look exactly as finite-difference approximations, but of a different partial differential operator with variable deteriorating coefficients. Proof of the almost uniform in p spectral equivalence to the stiffness matrices of the elements $\mathcal{E}_H$ is found in the same papers, cited above, where the preconditioners were proposed.

For the spectral reference elements $\mathcal{E}_{Sp}$, similar results on the spectral equivalence lagged a number of years behind the introduction of the preconditioners. In this respect, the results of [Bernardi and Maday (1992a,b)] on the Lagrange interpolations over the sets of Gauss-Lobatto-Legendre and Gauss-Lobatto-Chebyshev quadrature nodes were decisive. In particular, they proved the equivalence of the $L_2(-1,1)$ and $H^1(-1,1)$ norms for polynomials and their piecewise linear interpolations over the set of the Gauss-Lobatto-Legendre quadrature nodes. From here it was a shorter step to the justification of the preconditioners for the corresponding multidimensional reference elements $\mathcal{E}_{Sp}$, attested in [Canuto (1994)], [Parter and Rothman (1995)], [Bernardi and Maday (1997)] and [Casarin (1997)]. Another type of efficient preconditioners in the factorized form for the spectral elements $\mathcal{E}_{GLL}$ appeared in [Korneev and Rytov (2005a,b, 2008)]. They followed by the diagonal transformation of the low order finite element preconditioners as quadratic forms and then neglecting minor terms in the matrix obtained after the transformation. These operations yielded the factored preconditioners with three factors, two of which are diagonal matrices. It is important that the single nondiagonal factor has some essential properties in common with the finite-difference preconditioners for the hierarchical reference elements.

The finite-difference and the like preconditioners simplify inexact iterative solvers for the internal Dirichlet problems on the finite p-elements, associated by the shape regular mappings with the reference elements $\mathcal{E}_H, \mathcal{E}_{GLL}$. However, the creation of fast solvers for these preconditioners, which provide lowest contribution to the computational cost of the DD solver, was not easy, because the preconditioners represent highly irregular discrete problems. Fast solvers of different types for local Dirichlet problems, constructed on the basis of the described preconditioners, appeared in the present century. Nonoverlapping DD type almost optimal solver for Dirichlet problems on $\mathcal{E}_H$ elements was suggested in [Korneev (2002b,a)], an optimal version of which was approved in [Korneev (2013b)]. The optimal in the arithmetical cost multigrid solver and the solver, based on the multilevel decomposition of the finite element space in special wavelet subspaces, were developed by [Beuchler (2002)] and [Beuchler *et al.* (2004)], respectively. Using the

same as in [Korneev (2002b,a)] domain decomposition grid with numbers of deteriorating subdomains growing with p, the authors of [Beuchler and Nepomnyaschikh (2007)] designed an overlapping DD optimal solver. The factored preconditioners of [Korneev and Rytov (2007, 2008)] allowed them to generalize all named optimal solvers to the Dirichlet problems on the spectral reference elements $\mathcal{E}_{GLL}$.

Advances in the development of fast solvers for the discrete Dirichlet problems on the spectral reference elements paved the way to fast algorithms for some other components of DD solvers. Almost optimal in computational cost low energy prolongations inside finite elements from their boundaries can be performed by fast Dirichlet solvers for finite elements. Besides, under some conditions, solvers for the faces F can be obtained with the help of the K-intepolation technique between the spaces $L_2(F)$ and $\mathring{H}^1(F)$. Here $\mathring{H}^1(F)$ is the space of functions from $H^1(F)$ on face F with the zero traces on the boundary ∂F. Combined results on the BPS type Schur complement preconditioning, fast solvers for the Dirichlet problems on finite p-elements, and fast algorithms for other components allowed to construct DD preconditioner-solvers almost optimal in the computional work. Such preconditioner-solvers for the discretizations by hierarchical p-elements, including the discretizations suitable for the p-adaptive computations, were presented in [Korneev *et al.* (2002c,d)], and for the discretizations by spectral elements in [Korneev and Rytov (2007, 2008)].

Evaluation of the finite element stiffness matrix and load vector are unavoidable operations in the implementation of finite element methods, which sometimes is called setup. If an iterative procedure is used for solving a finite element system, then another unavoidable operation is the matrix-vector multiplication by the finite element stiffness matrix. In the h-version both operations do not cause trouble, since usually their costs are of the order of the number of degrees of freedom. However, in general, for hp discretizations the situation is different and the computational cost of these operations can be high, although they are always in the range of algebraic complexity. Until recently, these operations were not paid much attention. However, alongside with appearing fast DD preconditioner-solvers for hp discretizations, some advances in reducing the costs of the setup and matrix-vector multiplications have been done. We do not touch results in this direction and refer to [Karniadakis and Sherwin (1999)], [Melenk *et al.* (2001)] and [Nübel *et al.* (2001)] for more information.

In relation to the fact that the setup and matrix-vector multiplication procedures pose an independent problem at applications of the hp-version,

we make a few remarks on the terminology. Suppose, $\mathbf{Kv} = \mathbf{f}$ is a system of linear algebraic equations with a $n \times n$ s.p.d. (symmetric positive definite) matrix, and n is a parameter. In this book, an algorithm for solving the system, *i.e.,* solver, is termed *fast* or *(asymptotically) almost optimal* in arithmetical/computational work, if it requires $\mathcal{O}((1 + \log n)^{\nu} n)$ arithmetic operations with some fixed and not big ν. If $\nu = 0$, then it is termed *optimal*. Often for solving the system, PCGM (Preconditioned Conjugate Gradient Method) or other iterative procedure with some preconditioner $\boldsymbol{\mathcal{K}}$ is used. Our main interest is in obtaining good preconditioners, and we would like to single out their contribution to the numerical complexity of solvers. For this reason, by the *arithmetical/computational cost or complexity* of a preconditioner-solver, we understand the arithmetical cost for solving the system, except for the matrix-vector multiplications by the matrix $\mathbf{K}$ at each iteration. These multiplications are unavoidable at iterative solution. If solving procedure is not pointed out, it is always assumed PCGM. We say *preconditioner-solver* $\boldsymbol{\mathcal{K}}$ when an efficient solver for systems with the matrix $\boldsymbol{\mathcal{K}}$ is known. Accordingly, the preconditioner is called *fast, almost optimal or optimal*, if it is such in the arithmetical complexity. Therefore, a preconditioner-solver is, *e.g.,* fast, if it is at least *almost optimal in the condition*, which implies $\operatorname{cond}[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \mathcal{O}(\log^{\nu_1} n)$, and the system $\boldsymbol{\mathcal{K}}\mathbf{v} = \mathbf{f}$ with any $\mathbf{f}$ can be solved for $\mathcal{O}(n \log^{\nu_2} n)$ arithmetic operations with not too big ν_1, ν_2. If n depends on several parameters, similar terms are used with respect to any of these parameters. However, the solver is called fast, optimal etc., if it is such with all operations, including matrix-vector multiplications, taken into account.

1.2 Two Origins of Domain Decomposition Methods

Let us suppose that, for two sufficiently simple and sufficiently smooth domains Ω_1 and Ω_2, the Poisson equation

$$-\Delta u(x) = f(x)\,, \quad x \in \Omega\,, \tag{1.1}$$

$$u(x) = g(x)\,, \quad x \in \partial\Omega\,, \tag{1.2}$$

has a unique solution in $\Omega = \Omega_1$ as well as Ω_2 for any given sufficiently smooth right-hand side f and Dirichlet data g. For the special case of the Laplace equation, *i.e.* $f \equiv 0$, [Schwarz (1869)] studied the same problem in the more complicated domain $\Omega = \Omega_1 \cup \Omega_2$, which is a union of overlapping domains Ω_1 and Ω_2. Thus, he was interested in the existence of a harmonic

function with prescribed Dirichlet data g on the boundary $\partial\Omega$ of the domain Ω provided that such a result was known for the simpler domains Ω_1 and Ω_2 from which Ω is generated. Figure 1.1 provides a sketch of Schwarz' original drawing with Ω_1 and Ω_2 being a rectangle and a circle. Schwarz

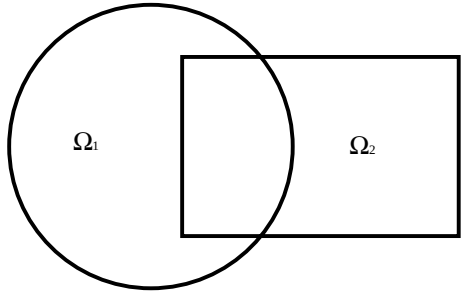

Fig. 1.1 Schwarz' overlapping domain decomposition sample.

proved that there is a unique harmonic function that solves (1.1)–(1.2) in the more complicated domain $\Omega = \Omega_1 \cup \Omega_2$. The proof was based on the maximum principle and on a new alternating iteratation procedure in which similar problems were solved in Ω_i alternately with the boundary function g defined on $\partial\Omega_i \cup \Omega_{3-i}$ from the preceding solution for the domain Ω_{3-i} with $i = 1, 2$. We also refer to [Nevanlinna (1939)] for genaralizations of the proof on the basis of the maximum principle. In a brief form, Algorithm 1.2 presents this Alternating Schwarz Method (AltSM) for solving the more general Dirichlet problem for the Poisson(1.1)–(1.2). It is obvious that the recursive application of the AltSM allows us to establish the existence and uniqueness of the solution to the Boundary Value Problem (BVP) (1.1)–(1.2) for an huge class of domains.

The variational setting of the AltSM as an alternating minimization procedure in Ω_1 and Ω_2 was given by [Sobolev (1936)], who used the linear elasticity equations as a model problem. This setting allowed him to prove convergence to the weak solution in the energy space. Returning to our model BVP (1.1)–(1.2), we introduce the Ritz functional

$$F(v) = \frac{1}{2}\, a(v, v) - \langle f, v\rangle\,, \tag{1.3}$$

where the bilinear form $a(\cdot,\cdot)$ and the linear from $\langle f, \cdot\rangle$ are defined by the relations

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx \quad \text{and} \quad \langle f, v\rangle = \int_\Omega f v \, dx\,, \tag{1.4}$$

Algorithm 1.1 AltSM for the Dirichlet problem (1.1)–(1.2).

Given initial guess u^0: $u^0 \in C^2(\Omega) \cap C(\overline{\Omega})$, $\quad u^0 = g$ on $\partial\Omega$ {initialization}

for $n = 0$ **step** 1 **until convergence** {begin iteration loop}

First step: {update in Ω_1}
find $\widetilde{u}^{n+1/2} \in C^2(\Omega_1) \cap C(\overline{\Omega}_1)$: $-\Delta\widetilde{u}^{n+1/2} = f$ in Ω_1,
$\widetilde{u}^{n+1/2} = u^n$ on $\partial\Omega_1$;
set $u^{n+1/2}(x) := \widetilde{u}^{n+1/2}(x), \quad \forall x \in \overline{\Omega}_1$;
set $u^{n+1/2}(x) := u^n(x), \quad \forall x \in \overline{\Omega}_2 \setminus \overline{\Omega}_1$.

Second step: {update in Ω_2}
find $\widetilde{u}^{n+1} \in C^2(\Omega_2) \cap C(\overline{\Omega}_2)$: $-\Delta\widetilde{u}^{n+1} = f$ in Ω_2,
$\widetilde{u}^{n+1} = u^{n+1/2}$ on $\partial\Omega_2$;
set $u^{n+1}(x) := \widetilde{u}^{n+1}(x) \quad \forall x \in \overline{\Omega}_2$;
set $u^{n+1}(x) := u^{n+1/2}(x), \quad \forall x \in \overline{\Omega}_1 \setminus \overline{\Omega}_2$.

end for {end iteration loop}

respectively. Now we can associate the minimization problem: find a function u from the Sobolev space $H^1(\Omega)$ such that

$$F(u) = \min_{v \in H^1(\Omega):\ v = g \text{ on } \partial\Omega} F(v)\,, \tag{1.5}$$

with our BVP (1.1)–(1.2), where the conditions imposed on the smoothness of the boundary $\partial\Omega$ and the given function f can be weakened in comparison with the classical formulation, see, *e.g.*, [Braess (2007)] or [Steinbach (2008)]. Algorithm 1.2 provides Sobolev's version of the AltSM, where we use the notations $\mathbb{V}_1$ and $\mathbb{V}_2$ for the Sobolev spaces $\mathring{H}^1(\Omega_1) = \{v \in H^1(\Omega_1) : v = 0 \text{ on } \partial\Omega_1\}$ and $\mathring{H}^1(\Omega_2) = \{v \in H^1(\Omega_2) : v = 0 \text{ on } \partial\Omega_2\}$, respectively, see, *e.g.*, [Adams (1975)]. It is clear that the variational setting widens the opportunities for its numerical solution. Employing Sobolev's variational setting, [Mikhlin (1951)] additionally proved uniform convergence in every closed subdomain of Ω. After these works, further studies of the AltSM were continued as a method of decomposition of Sobolev and other functional spaces (*e.g.*, corresponding to domain decomposition), in which the boundary value problem is formulated, see, *e.g.,* [Babuška (1958)].

If the AltSM is derived from the variational (weak, respectively, generalized) formulations of boundary value problems, then it can be interpreted and analyzed as projection method. An early interpretation as projection

Algorithm 1.2 AltSM as Alternating Minimization Algorithm.

Given initial guess u^0 : $u^0 \in H^1(\Omega)$, $u^0 = g$ on $\partial\Omega$ {initialization}

for $n = 0$ **step** 1 **until** convergence {begin iteration loop}

First step : {update in Ω_1}
find $u^{n+1/2}$: $F(u^{n+1/2}) = \min_{w\in\mathbb{V}_1} F(u^n + w)$;

Second step : {update in Ω_2}
find u^{n+1} : $F(u^{n+1}) = \min_{w\in\mathbb{V}_2} F(u^{n+1/2} + w)$;

end for {end iteration loop}

method appeared in the paper by [Morgenstern (1956)]. Without loss of generality, we assume homogeneous Dirichlet boundary conditions $g \equiv 0$. The variational (weak, respectively, generalized) reformulation of our BVP (1.1)–(1.2), which is equivalent to the minimization problem (1.5), reads as follows: find $u \in \mathbb{V} = \mathring{H}^1(\Omega)$ satisfying the integral identity

$$a(u, v) = \langle f, v\rangle\,, \quad \forall v \in \mathbb{V}\,, \tag{1.6}$$

whereas the corresponding form of the AltSM is given in Algorithm 1.3.

If we introduce the orthogonal projections, called also Ritz projections, $P_i : \mathbb{V} \to \mathbb{V}_i$ by the identities

$$a(P_i u, v) = a(u, v)\,, \quad \forall u \in \mathbb{V}\,, \quad \forall v \in \mathbb{V}_i\,, \tag{1.9}$$

then from (1.7) and (1.8) we observe that $w^{n+1/2} = P_1(u - u^n)$ and $w^{n+1} = P_2(u - u^{n+1/2})$, respectively. Thus, the iteration error $z^n := u - u^n$ satisfies the recurrence

$$z^{n+1} = (I - P_2)(I - P_1)z^n = (I - P_1 - P_2 + P_2P_1)z^n\,, \tag{1.10}$$

which is an alternating orthoprojection of the iteration error to the spaces $\mathbb{V}_1^\perp$ and $\mathbb{V}_2^\perp$. In either of the described forms, the AltSM can be generalized to a decomposition of Ω or, equivalently, the space $\mathbb{V}$ into many subdomains and subspaces, respectively, or to similar decompositions of finite element meshes and finite element spaces.

Due to the multiplicative nature of the error transition operator in (1.10), this kind of alternating procedures is now often called *multiplicative Schwarz method* (MSM). It has an obvious computational drawback due to the sequential character of the procedure that prevents an easy parallelization. In order to overcome this drawback or to considerably reduce its

Algorithm 1.3 Variational AltSM for (1.1)–(1.2).

$u^0 \in \mathbb{V} = \mathbb{V}_1 + \mathbb{V}_2$ given initial guess {initialization}

for $n = 0$ **step** 1 **until** Convergence {begin iteration loop}

First step: {update in $\mathbb{V}_1$}
$u^{n+1/2} = u^n + w^{n+1/2}$, with $w^{n+1/2} \in \mathbb{V}_1$:

$$\int_{\Omega_1} \nabla w^{n+1/2} \cdot \nabla v \, dx = \int_{\Omega_1} f(x) v(x) \, dx - \int_{\Omega_1} \nabla u^n \cdot \nabla v \, dx \,, \quad \forall v \in \mathbb{V}_1. \tag{1.7}$$

Second step: {update in $\mathbb{V}_2$}
$u^{n+1} = u^{n+1/2} + w^{n+1}$, with $w^{n+1} \in \mathbb{V}_2$:

$$\int_{\Omega_2} \nabla w^{n+1} \cdot \nabla v \, dx = \int_{\Omega_2} f(x) v(x) \, dx - \int_{\Omega_2} \nabla u^{n+1/2} \cdot \nabla v \, dx \,, \quad \forall v \in \mathbb{V}_2. \tag{1.8}$$

end for {end iteration loop}

influence, some remedies are used. One of them consists in a special subdomains ordering techniques (coloring) for overlapping decompositions. It is based on the usually observed property of the decompositions with many subdomains that each subdomain intersects with only a few other subdomains. This property is especially peculiar to the Schwarz methods for finite element or finite-difference discretizations of boundary value problems for PDEs. With the more radical purpose to completely avoid sequential projections, overlapping and nonoverlapping additive and other versions of Schwarz' algorithms have been proposed. These algorithm are among the most suitable methods for many engineering computations performed on parallel computers with many processors, see, *e.g.*, [Douglas *et al.* (2003)].

A forerunner of some modern versions of nonoverlapping DD methods, especially such as DD methods of the Dirichlet–Dirichlet type, were the direct substructuring techniques for solving large-scale systems of finite element equations, which were paid much attention to by engineers in the 60s. For finite dimensional problems of frame mechanics, DD type and superelement type algorithms existed for a rather long time. For some understandable reason, initially they were mostly used in the form of the

method of forces, based on the Castigliano principle and often called the method of statically indeterminate principal system. Instead of solving the problem with a high degree of static indeterminacy, one cuts the frame in sub-frames (substructures) with considerably less degrees of indeterminacy, and resolve each separately for its influence (deflection) matrix and the vector of deflections, caused by the external forces. This allows us to reduce the initial problem to a problem of a much smaller dimension, which is formulated in respect to the internal forces at the set of points of the interface between the sub-frames. Having the interface problem solved, we return to the sub-frames and get the solution for each of them separately with the boundary conditions found in the preceding step. In the absence of powerful computers, there were no use even to formulate the system of algebraic equations for the whole structure. It was only after their appearance when this method was recognized as an equivalent one to a block Gaussian elimination procedure for the system of algebraic equations for the whole structure, based on a special ordering of unknowns, *i.e.*, substructuring. After the finite element method itself was reinvented by engineers on the basis of the analogy with methods of structural mechanics, the remedy for solving large finite element systems, now primarily in the form of displacement method, was borrowed from the same source. It was efficiently applied to large systems of finite element algebraic equations and became known as *substructuring* or *superelement* and later *Schur complement* technique. Substucturing techniques are well exposed in [Przemieniecki (1963)], see also [Argyris and Kelsey (1960)].

The FE discretization of a linear elliptic boundary value problem usually results in the system of linear algebraic equations of the form

$$\mathbf{K}\mathbf{u} = \mathbf{f}\,. \tag{1.11}$$

In modern supercomputing applications, according to the complexity of high technology problems and requirements of accuracy, it may have the dimension from several tens of millions to tens of billions and more. In order to obtain accurate numerical results in an appropriate time span, the solving procedure, accounting for special properties of the FE discretizations, is of utmost importance. Let us decompose the computational domain Ω into J non-overlapping subdomains (substructures) $\Omega_1, \Omega_2, \ldots, \Omega_J$ such that

$$\overline{\Omega} = \bigcup_{j=1}^{J} \overline{\Omega}_j\,, \qquad \text{and} \qquad \Omega_i \cap \Omega_j = \varnothing \quad \text{for} \quad i \neq j\,, \tag{1.12}$$

with the boundaries $\partial\Omega_j$ not crossing finite elements. This decomposition yields a corresponding splitting of the unknowns, as well as of the related

space, into *internal* and *interface* unknowns and subspaces. They are related to the interiors of subdomains, and to the *interface boundary*, in the literature termed also *inter-subdomain boundary* and *coupling boundary*. In what follows, lower indices I and B are used for the objects (degrees of freedom, subspaces etc.), related to the interiors of the subdomains, *i.e.*, to $\Omega_I = \cup_{j=1}^{J} \Omega_j$, and to the inter-subdomain boundary

$$\Gamma_B = \bigcup_{j=1}^{J} \partial\Omega_j \setminus \overline{\Gamma}_D \,,$$

respectively. Γ_D and Γ_N denote the parts of the boundary $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$ with Dirichlet and Natural boundary conditions, respectively. Parts of the boundary with natural – Neumann and Robin – boundary conditions will be treated as a part of the inter-subdomain boundary.

We may order the unknowns in the following way: first we pick up all the *internal unknowns*, *i.e.*, the unknowns related to Ω_I, and then the *interface unknowns* which are related to Γ_B. In the subset of internal unknowns, first we enumerate the unknowns for the subdomain Ω_1, then Ω_2 and so forth, ending with Ω_J. The FE Galerkin basis Φ can be ordered accordingly, *i.e.*,

$$\Phi = (\Phi_I, \Phi_B) = \left(\phi_1, \ldots, \phi_{N_{I_1}}, \phi_{N_{I_1}+1}, \ldots, \phi_{N_I}, \phi_{N_I+1}, \ldots, \phi_{N=N_I+N_B}\right) \,, \tag{1.13}$$

where the first N_{I_1} basis functions belong to Ω_1, the next N_{I_2} functions to Ω_2 and so on until one comes to the N_{I_J} functions belonging to Ω_J and to the N_B functions belonging to Γ_B. The total number of the internal basis functions is clearly equal to $N_I = N_{I_1} + N_{I_2} + \ldots + N_{I_J}$. Due to the basic property of FE discretizations, provided that the supports of internal FE Galerkin coordinate functions for different subdomains Ω_j do not intersect each other, the FE system (1.11) obtains the block form

$$\begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{IB} \\ \mathbf{K}_{BI} & \mathbf{K}_B \end{pmatrix} \begin{pmatrix} \mathbf{u}_I \\ \mathbf{u}_B \end{pmatrix} = \begin{pmatrix} \mathbf{f}_I \\ \mathbf{f}_B \end{pmatrix} \,, \qquad \mathbf{K}_I = \text{blockdiag}\left[\mathbf{K}_{I_j}\right]_{j=1}^{J} \,. \tag{1.14}$$

Each block $\mathbf{K}_{I_j}$ on the diagonal of $\mathbf{K}_I$ is induced by the FE approximation to the restriction of the source PDE to Ω_j under homogeneous Dirichlet boundary conditions on $\partial\Omega_j$. Therefore, internal unknowns may be eliminated by Gaussian elimination for all subdomains in parallel. Instead of Gaussian elimination, more efficient solving procedures adapted to the discretizations of different subdomains may be used. In this way, (1.14) is reduced to the solution of the interface problem, which is the system with the *Schur complement* $\mathbf{S}_B = \mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB}$ as system matrix.

Having solved this system for the interface unknowns, we are able to determine the internal unknowns for each subdomain separately in parallel. This direct substructuring procedure is outlined in Algorithm 1.4. Direct

Algorithm 1.4 Direct Substructuring Algorithm.

$\widetilde{\mathbf{u}}_I = \mathbf{K}_I^{-1}\mathbf{f}_I$, *i.e.*, solve $\mathbf{K}_I\widetilde{\mathbf{u}}_I = \mathbf{f}_I$;	{elimination of internal unknowns}
$\mathbf{g}_B = \mathbf{f}_B - \mathbf{K}_{BI}\widetilde{\mathbf{u}}_I$;	{forming of the right-hand side}
$\mathbf{S}_B = \mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB}$;	{forming of Schur complement}
$\mathbf{u}_B = \mathbf{S}_B^{-1}\widetilde{\mathbf{f}}_B$;	{solving the Schur complement problem}
$\mathbf{u}_I = \widetilde{\mathbf{u}}_I - \mathbf{K}_I^{-1}\mathbf{K}_{IB}\mathbf{u}_B$;	{determination of the internal unknowns}

substructuring is well suited for parallel implementation since it reduces the solution of the FE system (1.11) to solving a number of subsystems of smaller dimensions. Still this procedure may be very costly in terms of the arithmetic cost especially forming the Schur complement $\mathbf{S}_B$, which is often time consuming in practice. However, if an iterative solver is used for the Schur complement problem, the forming of $\mathbf{S}_B$ can be completely avoided, because $\mathbf{S}_B$ becomes only involved in the matrix-by-vector multiplications $\mathbf{S}_B\mathbf{v}_B$. This is the step to the *iterative substructuring methods*, which already possess many basic features of modern non-overlapping DD methods. For the engineering applications, nonoverlapping DD methods are more attractive, because the subdomains of decomposition may exactly correspond to substructures of a real structure. This is important for many structures and problems, *e.g.*, for structures with substructures of different dimension, or quite different properties, or for structures with very complex geometry, composed from substructures of much simpler forms.

Chapter 2

Fundamentals of the Schwarz Methods

The most important properties of the DD preconditioner-solvers can be derived from general properties of finite element discretizations of the elliptic boundary value problems and the way of the decomposition in subdomains. We will illustrate this for the discretizations of some typical model problems, from the descriptions of which we start this section.

2.1 Elliptic Model Problems and their Discretizations

Let $\mathbb{V}$ be a Hilbert space, equipped with the scalar product $(\cdot,\cdot)$ and the corresponding norm $\|\cdot\| = (\cdot,\cdot)^{1/2}$, and let $a(\cdot,\cdot) : \mathbb{V}\times\mathbb{V} \to R$ be a symmetric, $\mathbb{V}$- elliptic ($\mathbb{V}$- coercive) and bounded (continuous) bilinear form. These three properties are expressed by the respective lines

$$a(u,v) = a(v,u)\,, \qquad \forall\, u,v \in \mathbb{V}\,, \tag{2.1}$$

$$\mu_1\|v\|^2 \le a(v,v)\,, \qquad \forall v \in \mathbb{V}\,, \tag{2.2}$$

$$a(u,v) \le \mu_2\,\|u\|\,\|v\|\,, \qquad \forall\, u,v \in \mathbb{V}\,, \tag{2.3}$$

where μ_k are positive constants. By the definition, the *duality product* $\langle f, v\rangle$ gives the value of the bounded (continuous) linear functional f from the dual space $\mathbb{V}^*$ at some $v \in \mathbb{V}$. The abstract variational setting of an elliptic bounded problem reads as follows: given $f \in \mathbb{V}^*$, find $u \in \mathbb{V}$ such that the variational equation

$$a(u,v) = \langle f, v\rangle\,, \quad \forall\, v \in \mathbb{V}\,, \tag{2.4}$$

holds. Under conditions (2.1)-(2.3), existence and uniqueness of the solution to the variational problem (2.4) is ensured by the Lax-Milgram lemma, see, *e.g.*, [Braess (2007); Ciarlet (1978); and Steinbach (2008)].

This abstract setting covers a multiplicity of practically important problems, including the primal for this book class of problems, described by self-adjoint, elliptic equations. Apart from the Dirichlet boundary value problem for the Poisson equation (1.1)–(1.2), some other model problems of this class are used in this book to illustrate DD techniques and results of their analysis.

Example 2.1 (Stationary Heat Conduction Problem). *Given some heat source intensity function $f \in L_2(\Omega)$, find the temperature field $u \in \mathbb{V} = \mathring{H}^1(\Omega)$ such that the variational equation (2.4) holds with the bilinear and linear forms*

$$a(u,v) = \int_\Omega \varrho(x)\,\nabla u(x)\cdot\nabla v(x)\,dx \quad \textit{and} \quad \langle f, v\rangle = \int_\Omega f(x)v(x)\,dx, \quad (2.5)$$

respectively. If it is not defined otherwise, the domain $\Omega \in R^d$ $(d = 1, 2, 3)$ is assumed to be bounded and sufficiently smooth, e.g.*, with a Lipschitz boundary. The symmetry of the bilinear form is obvious. If the given heat conduction coefficient ϱ is uniformly positive and bounded,* i.e.*, $0 < \underline{\mu} \le \varrho \le \overline{\mu}$, with $\underline{\mu}, \overline{\mu} =$ const, the $\mathbb{V}$-ellipticity (2.2) and the $\mathbb{V}$-boundness (2.3) directly follow by the Friedrichs and Cauchy inequalities, respectively, see,* e.g.*, [Braess (2007); Ciarlet (1978); and Steinbach (2008)]. The space $\mathring{H}^1(\Omega)$ assumes homogeneous Dirichlet boundary conditions,* i.e.*, vanishing temperature u on the boundary $\partial\Omega$. We mention that the heat conduction equation formally describes many other stationary processes, like diffusion or filtration in the porous media with variable permeability.*

Many practical problems are described by the elliptic equations with coefficients having jumps. A model problem of such type is presented in the Example 2.2.

Example 2.2 (Piecewise Constant Heat Conductivities). *The bilinear and linear forms of the variational problem are the same as in (2.5), but the domain Ω is a union of nonoverlapping subdomains Ω_j like in (1.12), and, in each subdomain,*

$$\varrho(x) = \varrho_j = const > 0\,, \quad x \in \Omega_j, \quad (2.6)$$

where ϱ_j are arbitrary positive constants.

As a rule, other boundary conditions like Neumann, Robin and mixed boundary conditions do not require special DD algorithms and can be treated in a similar way as the Dirichlet problem. The part of the boundary

with such boundary conditions is simply treated as a part of the intersubdomain boundary. It is worth emphasizing that in practical applications, coefficients of elliptic equations may change significantly in the domain and have jumps or deteriorate. Thus, the ratio μ_2/μ_1 is very large or even tends to infinity when approaching some points in the domain. However, this is the case when DD methods may be and, at their proper design, are very often robust, *i.e.*, allow to avoid losses in efficiency in comparison with the case of smooth coefficients. One of the examples of such algorithms is considered in Section 6.

Example 2.3 (Static Linear Elasticity Problem). *We are given a sufficiently smooth domain $\Omega \subset R^d$, occupied by a linearly elastic body clamped at some part $\Gamma_D \subset \Gamma = \partial\Omega$ of nonzero measure, volume forces $\mathbf{f} = (f_1, \ldots, f_d)^T$, surface tractions $\mathbf{t} = (t_1, \ldots, t_d)^T$ on the part $\Gamma_N = \Gamma \setminus \Gamma_D$ of the boundary Γ, and a symmetric and uniformly positive definite and bounded matrix*

$$\mathbf{D}(x) = \{D_{ijkl}(x)\}_{i,j,k,l=1}^{d}$$

of elastic coefficients such that

$$\underline{\mu}\,\mathbf{I} \le \mathbf{D}(x) \le \overline{\mu}\,\mathbf{I} \quad \textit{for almost all } x \in \overline{\Omega}\,,$$

with positive constants $\underline{\mu}$ and $\overline{\mu}$. Introducing the bilinear and linear forms

$$a(\mathbf{u}, \mathbf{v}) = \int_\Omega \sum_{i,j,k,l=1}^{d} \varepsilon_{ij}(\mathbf{u})\, D_{ijkl}(x) \varepsilon_{kl}(\mathbf{v})\, dx \quad \textit{and} \tag{2.7}$$

$$\langle \mathbf{f}, \mathbf{v} \rangle = \int_\Omega \sum_{i=1}^{d} f_i(x)\, v_i(x)\, dx + \int_{\Gamma_N} \sum_{i=1}^{d} t_i(x)\, v_i(x)\, ds, \tag{2.8}$$

respectively, where $\varepsilon_{ij}(\mathbf{u}) = 0.5(\partial u_i/\partial x_j + \partial u_j/\partial x_i)$ denotes the linearized strains, we can formulate the linear elasticity problem in variational form: find the displacements vector $\mathbf{u} = (u_1, \ldots, u_d)^T$ of the elastic body under consideration, belonging to the space $\mathbb{V} = \{\mathbf{v} = (v_1, \ldots, v_d)^T \in \mathbf{H}^1(\Omega) := (H^1(\Omega))^d : \ \mathbf{v} = 0 \text{ on } \Gamma_D\}$ such that the variational equation

$$a(\mathbf{u}, \mathbf{v}) = \langle \mathbf{f}, \mathbf{v} \rangle\,, \quad \forall\, \mathbf{v} \in \mathbb{V}\,, \tag{2.9}$$

holds. The symmetry of the bilinear form is a consequence of the symmetry of matrix $\mathbf{D}$ of elastic coefficients in the Hooke's law. The proof of existence and uniqueness of the solution is much more intricate than the corresponding proof for the heat conduction problem from Example 2.1, and

requires additionally Korn's inequality. The stated properties of the matrix $\mathbf{D}$ *together with Korn's, Friedrichs', and Cauchy's inequalities ensure the* $\mathbb{V}$*-ellipticity (2.2) and* $\mathbb{V}$*-boundness (2.3) of the bilinear form. If the volume forces and surface tractions are such that the corresponding linear functional (2.8) is continuous on* $\mathbb{V}$*, then again Lax-Milgram lemma yields existence and uniqueness of the solution of static linear elasticity problem, see,* e.g.*, [Braess (2007); Ciarlet (1978); and Steinbach (2008)].*

In the finite element method with conforming finite elements, which is a specific Galerkin method, the variational equation (2.4) is approximated by the same variational equation, but on a finite dimensional subspace of $\mathbb{V}$. Let $\mathcal{V}$ be some finite dimensional FE subspace of $\mathbb{V}$ and $\Phi = (\phi_1, \phi_2, \cdots, \phi_N)$ be the FE Galerkin basis. Therefore, $\mathcal{V} = \text{span}\Phi \subset \mathbb{V}$ and any function of $\mathcal{V}$ is represented by the sum

$$u = \Phi\mathbf{u} := \sum_{i=1}^{N} u_i \phi_i \,, \tag{2.10}$$

which maps the vector of FE parameters $\mathbf{u} = (u_i)_{i=1}^N \in R^N$ into the corresponding FE function $u \in \mathcal{V}$. The coefficients u_i, if they have the sense of the values of u and some of its derivatives at the nodes of the FE mesh, are called *nodal parameters* as well as *degrees of freedom* or *unknowns* of the system of algebraic equations to be solved. In general, N depends on the dimension of the problem, the number of finite elements of the discretization and on degrees of coordinate functions, which most often are polynomials on the reference configurations for the finite elements. To remind, when necessary, the isomorphism between $\mathbf{u}$ and $u = \Phi\mathbf{u}$, we write $\mathbf{u} \leftrightarrow u$.

The FE Galerkin solution of the variational equation (2.4) is sought as the function $\widetilde{u} \in \mathcal{V}$ satisfying the identity

$$a(\widetilde{u}, v) = \langle f, v \rangle \,, \qquad \forall\, v \in \mathcal{V} \,. \tag{2.11}$$

Once the basis Φ is chosen, the FE variational equation (2.11) is equivalent to the system of algebraic equations

$$\mathbf{K}\mathbf{u} = \mathbf{f} \,. \tag{2.12}$$

The stiffness matrix $\mathbf{K}$ and the load vector $\mathbf{f}$ are generated according to the identities

$$(\mathbf{K}\mathbf{u}, \mathbf{v}) = a(\Phi\mathbf{u}, \Phi\mathbf{v}) = a(u, v) \,, \quad \forall\, \mathbf{u}, \mathbf{v} \leftrightarrow u, v \in \mathcal{V} \,, \tag{2.13}$$

and

$$(\mathbf{f}, \mathbf{v}) = \langle f, \Phi\mathbf{v} \rangle \,, = \langle f, v \rangle \,, \quad \forall\, \mathbf{v} \leftrightarrow v \in \mathcal{V} \,, \tag{2.14}$$

respectively. Here $(\mathbf{w}, \mathbf{v}) = (\mathbf{w}, \mathbf{v})_{R^N} = \mathbf{w}^T\mathbf{v}$ denote the Euclidian scalar product of the vectors $\mathbf{w}$ and $\mathbf{v}$ in R^N. We omit the space R^N in the subscript of the Euclidian scalar product, if this is clear from the context.

2.2 Domain Decomposition Methods as Preconditioning

In a modern interpretation, the domain decomposition or the space decomposition methods are more often viewed as an efficient way of preconditioning of partial differential operators or FE stiffness matrices in iteration processes for solving boundary value problems or FE systems of algebraic equations. First of all, the construction of a DD preconditioner requires preconditioning the restrictions of the bilinear form of the problem to the subspaces of the corresponding decomposition.

2.2.1 *Simplification of Local Bilinear Forms*

In general setting, DD algorithms correspond to some decomposition of the space $\mathcal{V}$ into a sum of subspaces $\mathcal{V}_s$ of dimensions N_s, written for convenience in one of the forms

$$\mathcal{V} = \sum_{s=1}^{J} \mathcal{V}_s \,, \quad \text{or} \quad \mathcal{V} = \sum_{s=0}^{J} \mathcal{V}_s \,, \tag{2.15}$$

where $\sum N_s \geq N$. These subspaces are spanned over some bases $\Psi_s = (\psi_{s,1}, \psi_{s,2}, \ldots, \psi_{s,N_s})$, defined by means of the $N \times N_s$ transformation matrices $\mathbf{T}_s$ with ranks N_s from the originally chosen basis Φ, so that

$$\mathcal{V}_s = \operatorname{span} \Psi_s = \operatorname{span} \Phi \mathbf{T}_s \,, \quad s = 1, 2, \ldots, J \,. \tag{2.16}$$

All these subspaces except one or a few are usually defined on subdomains of decomposition like, *e.g.*, in Algorithm 1.3. By their appearance, the two decompositions in (2.15) differ only by the numbers of subspaces. The additional space $\mathcal{V}_0$ usually serves as special subspace which overlaps with all other subspaces and improves the stability of the decomposition. If the opposite is not stated, we assume $s = 1, 2, ..., J$.

Let $a_s(\cdot, \cdot)$ be the restriction of the bilinear form $a(\cdot, \cdot)$ to the subspace $\mathcal{V}_s$. DD methods reduce solving the elliptic problem to solving similar problems in subspaces and, in particular, to solving elliptic problems on subdomains of the decomposition. In general, these problems do not correspond to the bilinear forms $a_s(\cdot, \cdot)$, but to some subsidiary forms, which will be denoted $b_s(\cdot, \cdot) : \mathcal{V}_s \times \mathcal{V}_s \to R$. These subsidiary bilinear forms should satisfy two contradictive requirements. On the one hand, in order to make the numerical solution easier, they should be simpler than $a_s(\cdot, \cdot)$. Otherwise, the replacement of $a_s(\cdot, \cdot)$ by $b_s(\cdot, \cdot)$ in DD preconditioner is useless. On the other hand, they must be sufficiently close to $a_s(\cdot, \cdot)$. In particular, if forms $a_s(\cdot, \cdot)$ are $\mathcal{V}_s$-elliptic, then it is reasonable to take $b_s(\cdot, \cdot)$ also

$\mathcal{V}_s$-elliptic. Moreover, the choice of $b_s(\cdot,\cdot)$ close in the spectrum or spectrally equivalent to $a_s(\cdot,\cdot)$ will improve the relative condition number of the DD preconditioner and, therefore, reduce the number of DD iterations.

The closeness is expressed by the inequalities

$$\underline{\gamma}_s b_s(v,v) \le a_s(v,v) \le \overline{\gamma}_s b_s(v,v)\,, \quad \forall\, v \in \mathcal{V}_s\,, \tag{2.17}$$

with sufficiently good, *i.e.*, in a sense of closeness to unity, positive numbers $\underline{\gamma}_s$ and $\overline{\gamma}_s$. The subsidiary bilinear forms $b_s(\cdot,\cdot)$ induce the $N_s \times N_s$ matrices $\mathbf{B}_s$ according to the identities

$$(\mathbf{B}_s\mathbf{u}_s, \mathbf{v}_s) = b_s(u_s, v_s)\,, \quad \forall\, \mathbf{u}_s, \mathbf{v}_s \overset{\Psi_s}{\longleftrightarrow} u_s, v_s \in \mathcal{V}_s\,, \tag{2.18}$$

in which we used $\overset{\Psi_s}{\longleftrightarrow}$ instead of $\leftrightarrow$ in order to underline that the isomorphism corresponds to the basis Ψ_s.

Properties of the decomposition in subspaces and properties of the bilinear forms $b_s(\cdot,\cdot)$ are reflected in the properties of the projection and the projection-like operators which are introduced below. The use of them can significantly simplify the analysis of DD algorithms. The orthoprojection $P_s : \mathcal{V} \to \mathcal{V}_s$ of the space $\mathcal{V}$ onto its subspace $\mathcal{V}_s$ with respect to the energy inner product $a(\cdot,\cdot)$ is uniquely defined by the identity

$$a(P_s u, v_s) = a(u, v_s)\,, \quad u \in \mathcal{V}\,, \quad \forall v_s \in \mathcal{V}_s\,. \tag{2.19}$$

It is also called *Ritz*, or *energy projection*. As an orthoprojection, this operator is self-adjoint with respect to the energy inner product, *i.e.*, $P_s = P_s^*$ in the sense of the identity

$$a(P_s u, v) = a(u, P_s^* v)\,, \quad \forall u, v \in \mathcal{V}\,. \tag{2.20}$$

Of course, it also satisfies the projection relation $P_s^2 = P_s$. From (2.19), (2.16) and (2.13), it follows that, for any $u \leftrightarrow \mathbf{u}$, the orthoprojection $P_s u$ may be computed as follows

$$P_s u = \Phi \mathbf{T}_s \mathbf{u}_s\,, \quad \mathbf{u}_s = (\mathbf{T}_s^T \mathbf{K} \mathbf{T}_s)^{-1} \mathbf{T}_s^T \mathbf{K} \mathbf{u}\,, \tag{2.21}$$

which requires the solution of a smaller system with the $N_s \times N_s$ matrix $\mathbf{T}_s^T \mathbf{K} \mathbf{T}_s$ and the right-hand side $\mathbf{T}_s^T \mathbf{K} \mathbf{u}$. Having replaced $a(\cdot,\cdot)$ by $b_s(\cdot,\cdot)$ in the left part of (2.19), we come to the projection-like operator $\widetilde{P}_s : \mathcal{V} \to \mathcal{V}_s$ which is defined by the identity

$$b_s(\widetilde{P}_s u, v_s) = a(u, v_s)\,, \quad u \in \mathcal{V}\,, \quad \forall v_s \in \mathcal{V}_s\,. \tag{2.22}$$

The projection-like operator $\widetilde{P}_s$ is self-adjoint, but in general $\widetilde{P}_s^2 \ne \widetilde{P}_s$. Therefore, $\widetilde{P}_s u$, which may be calculated by the formula

$$\widetilde{P}_s u = \Phi \mathbf{T}_s \mathbf{u}_s\,, \quad \mathbf{u}_s = \mathbf{B}_s^{-1} \mathbf{T}_s^T \mathbf{K} \mathbf{u}\,, \tag{2.23}$$

is not an orthoprojection. Usually, bilinear forms $b_s(\cdot,\cdot)$ are chosen in such a way that computation of $\widetilde{P}_s u$ is much cheaper than the computation of $P_s u$, and, at the same time, the values of $\underline{\gamma}_s$ and $\overline{\gamma}_s$ are sufficiently good, *e.g.*, close to unity. In practice, for obtaining a good DD preconditioner-solver, the design of each subdomain prconditioner $\mathbf{B}_s$ can be rather involved.

2.2.2 *Additive Domain Decomposition Algorithms*

The *Additive Schwarz Method* (ASM) may be written in two forms: as an iteration process in the FE space $\mathcal{V}$, which is the function version, and as an iteration process in R^N, which is the vector/matrix version. Both forms are given in Algorithm 2.1. In a compact form, the behavior of the error

Algorithm 2.1 Inexact ASM: function $\leftrightarrow$ vector/matrix versions.

$u^0 = \Phi\mathbf{u}^0 \in \mathcal{V}$ given initial guess and iteration parameter τ {initialization}

for $n = 0$ **step** 1 **until** Convergence {begin iteration loop}

for all $s \in \{1,\dots,J\}$ **do in parallel** {subspace corrections}

$$w_s^n = \Phi\mathbf{T}_s\mathbf{w}_s^n \in \mathcal{V}_s : \; b_s(w_s^n, v_s) = \langle f, v_s\rangle - a(u^n, v_s) = a(u - u^n, v_s)\,, \quad \forall v_s \in \mathbb{V}_s$$

$$\updownarrow$$

$$\mathbf{w}_s^n \in R^{N_s} : \qquad \mathbf{B}_s\mathbf{w}_s^n = \mathbf{T}_s^T(\mathbf{f} - \mathbf{K}\mathbf{u}^\mathbf{n}) \equiv \mathbf{T}_s^T\mathbf{d}^\mathbf{n}$$

end for

$u^{n+1} = u^n + \tau\sum_{s=1}^J w_s^n$ {updating the old iterate}

$$\updownarrow$$

$$\mathbf{u}^{n+1} = \mathbf{u}^n + \tau\sum_{s=1}^J \mathbf{T}_s\mathbf{w}_s^n$$

end for {end iteration loop}

is described in terms of projection and projection-like operators introduced above. Updating the old iterate in Algorithm 2.1 is obviously equivalent to the formula

$$u^{n+1} = u^n + \tau\sum_{s=1}^J \widetilde{P}_s(u - u^n)\,,$$

from which it is clear that the error $z^n = u - u^n \in \mathcal{V}$ satisfies the iteration scheme

$$z^{n+1} = (I - \tau\widetilde{P})z^n \equiv (I - \tau\sum_{s=1}^{J}\widetilde{P}_s)z^n, \tag{2.24}$$

with the inexact ASM operator $\widetilde{P} = \sum_{s=1}^{J}\widetilde{P}_s$. If inequalities (2.17) hold and the decomposition satisfies certain properties, which we discuss in Section 2.3, the operator $\widetilde{P} = \widetilde{P}^* > 0$ is self-adjoint and positive definite with respect to the energy inner product $a(\cdot,\cdot)$. Taking the energy norm $\|\cdot\|_a = \sqrt{a(\cdot,\cdot)}$ of (2.24), we arrive at the error estimate

$$\|z^{n+1}\|_a \le \|I - \tau\widetilde{P}\|_a \|z^n\|_a. \tag{2.25}$$

We see that the convergence rate of the inexact ASM is defined by the energy norm $\|I - \tau\widetilde{P}\|_a$ of the ASM iteration operator $I - \tau\widetilde{P}$.

According to the matrix version of Algorithm 2.1, the inexact ASM is nothing but the simple iteration or Richardson iteration

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \tau\boldsymbol{\mathcal{K}}^{-1}(\mathbf{K}\mathbf{u}^n - \mathbf{f}) \tag{2.26}$$

with the preconditioner

$$\boldsymbol{\mathcal{K}}^{-1} = \sum_{s=1}^{J}\mathbf{T}_s\mathbf{B}_s^{-1}\mathbf{T}_s^T\,. \tag{2.27}$$

Hence the vector counterpart of the iteration error estimate (2.25) reads as follows:

$$\|\mathbf{z}^{n+1}\|_{\mathbf{K}} \le \|I - \tau\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}\|_{\mathbf{K}}\,\|\mathbf{z}^n\|_{\mathbf{K}}\,, \tag{2.28}$$

and

$$\|\mathbf{z}^n\|_{\mathbf{K}} \le \|I - \tau\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}\|_{\mathbf{K}}^n\,\|\mathbf{z}^0\|_{\mathbf{K}}\,, \tag{2.29}$$

where $\mathbf{z}^n = \mathbf{u} - \mathbf{u}^n$ denotes the iteration error as a vector in R^N and $\|\mathbf{v}\|_{\mathbf{A}} := (\mathbf{A}\mathbf{v},\mathbf{v})_{R^N}^{1/2}$.

Let us suppose that we have some good bounds $\underline{\gamma}$ and $\overline{\gamma}$ for the spectrum of $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$, that is equivalent to the spectrum of the operator $\widetilde{P}$, *i.e.*

$$\lambda_{\min} = \lambda_{\min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) = \lambda_{\min}(\widetilde{P}) \ge \underline{\gamma} > 0 \tag{2.30}$$

and

$$\lambda_{\max} = \lambda_{\max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) = \lambda_{\max}(\widetilde{P}) \le \overline{\gamma}\,. \tag{2.31}$$

In turn, they are equivalent to the bounds in the spectral equivalence inequalities

$$\underline{\gamma}(\boldsymbol{\mathcal{K}}\mathbf{v},\mathbf{v}) \le (\mathbf{K}\mathbf{v},\mathbf{v}) \le \overline{\gamma}(\boldsymbol{\mathcal{K}}\mathbf{v},\mathbf{v})\,, \quad \forall \mathbf{v}\in R^N\,, \tag{2.32}$$

$$\underline{\gamma}a(v,v) \le a(\widetilde{P}v,v) \le \overline{\gamma}a(v,v)\,, \quad \forall v\in\mathcal{V}\,. \tag{2.33}$$

The analysis of the norm of the iteration matrix $I-\tau\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$, arising in the preconditioned Richardson iteration can be found, *e.g.*, in [Axellson (1994); Hackbusch (1994); Meurant (1999); and Samarskii and Nikolayev (1989)] with the conclusion that, for $\tau\in(0,2/\lambda_{max})$, we have the following estimate of the convergence rate (factor):

$$\begin{aligned} q(\tau) &:= \|I-\tau\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}\|_{\mathbf{K}} = \|I-\tau\widetilde{P}\|_a = \max\{|1-\tau\lambda_{\min}|,|1-\tau\lambda_{\max}|\} \\ &\le \widetilde{q}(\tau) := \max\{|1-\tau\underline{\gamma}|,|1-\tau\overline{\gamma}|\} < 1\,. \end{aligned} \tag{2.34}$$

According to (2.34) and (2.32)–(2.33), the optimal iteration parameter $\tau_{\rm opt}$ and optimal convergence factor $q_{\rm opt}$ for the iterative process (2.26) are given by the formulas

$$\tau_{\rm opt} = \frac{2}{\lambda_{\max}+\lambda_{\min}} \approx \widetilde{\tau}_{\rm opt} := \frac{2}{\overline{\gamma}+\underline{\gamma}} \tag{2.35}$$

and

$$q_{\rm opt} = q(\tau_{\rm opt}) = \frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}} \le \widetilde{q}_{\rm opt} := \frac{\overline{\gamma}-\underline{\gamma}}{\underline{\gamma}+\overline{\gamma}} < 1\,, \tag{2.36}$$

respectively. The values of $\underline{\gamma}$ and $\overline{\gamma}$ depend on the values of $\underline{\gamma}_s$ and $\overline{\gamma}_s$ as well as on the quality of the space decomposition, which may sometimes be expressed through the angles between subspaces $\mathcal{V}_s$, see Section 2.3, but in general it requires a more involved analysis. If, for example, these spaces are orthogonal, then $\underline{\gamma}\le\min_s\underline{\gamma}_s$ and $\overline{\gamma}\ge\max_s\overline{\gamma}_s$. Approximate values $\widetilde{\tau}_{\rm opt}$ and $\widetilde{q}_{\rm opt}$ of the optimal iteration parameter and the optimal convergence rate may be considered as such only if $\underline{\gamma}$ and $\overline{\gamma}$ are sufficiently close to $\lambda_{\min}$ and $\lambda_{\max}$.

Sometimes, the error estimate

$$\|\mathbf{z}^{n+1}\|_{\mathbb{K}} = \sqrt{(\mathbf{w}^{n+1},\mathbf{d}^{n+1})_{R^N}} \le q(\tau)\|\mathbf{z}^n\|_{\mathbb{K}}\,, \quad \mathbb{K} := \mathbf{K}\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}\,, \tag{2.37}$$

with respect to $\mathbf{K}\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$-energy norm with the same $q(\tau)$ as above can be very useful. Here $\mathbf{d}^n=\mathbf{K}\mathbf{z}^n=\mathbf{f}-\mathbf{K}\mathbf{u}^n$ and $\mathbf{w}^n=\boldsymbol{\mathcal{K}}^{-1}\mathbf{d}^n$ denote the *defect* and the *preconditioned defect (correction)*, respectively. In contrast to the estimates (2.28)–(2.29) and (2.34), estimate (2.37) is computable and can be used in convergence tests of Algorithm 2.1.

Algorithm 2.2 Preconditioned Conjugate Gradient Iteration.

$\mathbf{u}^{(1)} := 0\,;\ \mathbf{r}^{(1)} := \mathbf{f}\,;\ \mathbf{q}^{(0)} := 0\,;\ \mathbf{r}^{(0)} := 0\,;\ \beta_1 := 0\,;$
for $k := 1, 2, \ldots$
$\quad \mathbf{z}^{(k)} := \boldsymbol{\mathcal{K}}^{-1}\mathbf{r}^{(k)}\,;$
$\quad$ **if** $k > 1$
$\qquad \beta_k := \dfrac{(\mathbf{z}^{(k)})^\top(\mathbf{r}^{(k)})}{(\mathbf{z}^{(k-1)})^\top\mathbf{r}^{(k-1)}}\,;$
$\quad$ **endif**
$\quad \mathbf{q}^{(k)} := \mathbf{z}^{(k)} + \beta_k\mathbf{q}^{(k-1)}\,;$
$\quad \alpha_k := ((\mathbf{z}^{(k)})^\top\mathbf{r}^{(k)}) \,/\, ((\mathbf{q}^{(k)})^\top\mathbf{K}\mathbf{q}^{(k)})\,;$
$\quad \mathbf{r}^{(k+1)} := \mathbf{r}^{(k)} - \alpha_k\mathbf{K}\mathbf{q}^{(k)}\,;$
$\quad \mathbf{u}^{(k+1)} := \mathbf{u}^{(k)} + \alpha_k\mathbf{q}^{(k)}\,;$
endfor

In practice, instead of the preconditioned Richardson iterations (2.26), the Preconditioned Conjugate Gradient (PCG) iterations are primarily used, which provide a faster convergence and which is given in Algorithm 2.2. Instead of (2.29) with (2.36), for PCGM, we have the convergence estimate

$$\|\mathbf{z}^n\|_{\mathbf{K}} \le q_{n,\mathrm{PCG}}\|\mathbf{z}^0\|_{\mathbf{K}}\,, \tag{2.38}$$

where

$$q_{n,\mathrm{PCG}} = \frac{2q^n}{1+q^{2n}} \le 2q^n \asymp q^n\,, \quad q = \frac{1-\sqrt{\underline{\gamma}/\overline{\gamma}}}{1+\sqrt{\underline{\gamma}/\overline{\gamma}}}\,, \tag{2.39}$$

see, *e.g.*, Chapter VIII, Section 3.2 in [Samarskii and Nikolayev (1989)].

If no special comments are made, the multiplication $\boldsymbol{\mathcal{K}}^{-1}\mathbf{v} =: \mathbf{w}$ by the inverse to the preconditioner is understood throughout this book as some procedure of solving the system $\boldsymbol{\mathcal{K}}\mathbf{w} = \mathbf{v}$, *e.g.*, by a direct method. In many cases, solving this system by an *inexact iterative solver* can be much more efficient. This implies that one makes only some fixed and small number of iterations. Therefore, the exact version $\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}$ is replaced by the inexact one $\boldsymbol{\mathcal{K}}_{\mathrm{it}}^{-1}\mathbf{v}$ where $\boldsymbol{\mathcal{K}}_{\mathrm{it}}$ is implicitly defined by some iterations. The influence of secondary inexact solvers on the efficiency of PCG iterations has been studied by several authors. [Golub and Ye (1999)] suggested some variant of PCG method that appears to more stable with respect to the errors

occuring at inexact solving the systems with the preconditioners. They calculated β_k according to the formula

$$\beta_k = \frac{(\mathbf{z}^{(k)})^\top(\mathbf{r}^{(k)} - \mathbf{r}^{(k-1)})}{(\mathbf{z}^{(k-1)})^\top \mathbf{r}^{(k-1)}} \,. \tag{2.40}$$

Our numerical experience support their conclusion, see [Anufriev *et al.* (2003); Anufriev and Korneev (2005)]. For general information on PCG methods, we refer to [Axellson (1994); Hackbusch (1994); Meurant (1999); and Samarskii and Nikolayev (1989)].

Summarizing, we conclude that we can look at the ASM as a technique for constructing additive and, consequently, highly parallelizable preconditioners of the form (2.27), based on the space splitting and subspace preconditioning. These preconditioners may be used in other iterative procedures as well.

In what follows, for the reason of simplicity, we often use the iterative process (2.26) in situations when the PCG method could be used with much higher efficiency.

2.2.3 *Inexact Subspaces Preconditioning*

Although the decomposition of the space or the domain is made in order to come to subproblems in the subspaces (local subproblems) that are easier than the original problem, these subproblems may still be too complicated for an efficient solution by Gaussian elimination or another sparse direct method [George and Liu (1981)]. As we will see later, this is especially true for *hp* discretizations. But even for simple local mesh problems, iterative methods can be advantageous. A remarkable feature of DD algorithms with iterative solvers for the subproblems in subspaces is that these subproblems may be solved rather rudely with a moderate loss in overall efficiency of DD solver. Moreover, a proper use of iterative solvers as inexact solvers can noticeably reduce the computational work. The origin of the approach to preconditioning based on the inexact iterative solvers and formulated in Lemma 2.1 below goes back to [Dyakonov (1989)] and [Nepomnyaschikh (1991a)].

Lemma 2.1. *Let us suppose that, for solving the system* $\mathbf{Bv} = \mathbf{f}$ *with a s.p.d. matrix* $\mathbf{B}$*, some linear iterative method is used with the zero initial guess* $\mathbf{v}^0 = \mathbf{0}$*. Furthermore, we assume that the norm* $\|\cdot\|_{\mathbf{B}}$ *of the error is reduced by some factor* $\epsilon \in (0,1)$ *after* n_ϵ *iterations,* i.e.

$$\|\mathbf{v}^{n_\epsilon} - \mathbf{v}\|_{\mathbf{B}} \le \epsilon \, \|\mathbf{v}\|_{\mathbf{B}} \,. \tag{2.41}$$

Let $\mathbf{B}_{\rm it}$ *be the s.p.d. matrix in the relationship* $\mathbf{B}_{\rm it}\mathbf{v}^{n_\epsilon} = \mathbf{f}$*, which is implicitly defined by the iterative method. Then the spectral equivalence inequalities*

$$\underline{\gamma}_{\rm it}\,\mathbf{B}_{\rm it} \le \mathbf{B} \le \overline{\gamma}_{\rm it}\,\mathbf{B}_{\rm it} \tag{2.42}$$

are valid with the spectral equivalence constants $\underline{\gamma}_{\rm it} = 1-\epsilon$ *and* $\overline{\gamma}_{\rm it} = 1+\epsilon$*.*

Proof. Using (2.41) together with the triangle inequality, we immediately get

$$(1-\epsilon)\,\|\mathbf{v}\|_{\mathbf{B}} \le \|\mathbf{v}^{n_\epsilon}\|_{\mathbf{B}} \le (1+\epsilon)\,\|\mathbf{v}\|_{\mathbf{B}}\,. \tag{2.43}$$

Setting $\mathbf{y} = \mathbf{B}_{\rm it}^{-1/2}\mathbf{f}$ and $\mathbf{C} = \mathbf{B}_{\rm it}^{-1/2}\mathbf{B}\,\mathbf{B}_{\rm it}^{-1/2}$, we obtain

$$\|\mathbf{v}\|_{\mathbf{B}}^2 = (\mathbf{f}, \mathbf{B}^{-1}\mathbf{f}) = (\mathbf{B}_{\rm it}^{-1/2}\mathbf{f}, (\mathbf{B}_{\rm it}^{1/2}\mathbf{B}^{-1}\mathbf{B}_{\rm it}^{1/2})\mathbf{B}_{\rm it}^{-1/2}\mathbf{f}) = (\mathbf{y}, \mathbf{C}^{-1}\mathbf{y})$$

and

$$\|\mathbf{v}^{n_\epsilon}\|_{\mathbf{B}}^2 = (\mathbf{B}_{\rm it}^{-1}\mathbf{f}, \mathbf{B}\,\mathbf{B}_{\rm it}^{-1}\mathbf{f}) = (\mathbf{y}, \mathbf{C}\mathbf{y})\,.$$

These relations and inequalities (2.43) imply

$$(1-\epsilon)^2 \le \frac{\|\mathbf{v}^{n_\epsilon}\|_{\mathbf{B}}^2}{\|\mathbf{v}\|_{\mathbf{B}}^2} = \frac{(\mathbf{y}, \mathbf{C}\mathbf{y})}{(\mathbf{y}, \mathbf{C}^{-1}\mathbf{y})} \le (1+\epsilon)^2\,, \quad \forall\,\mathbf{y} = \mathbf{B}_{\rm it}^{-1/2}\mathbf{f} \ne \mathbf{0}\,,$$

or in another form

$$(1-\epsilon)^2 \le \lambda(\mathbf{C}^2) \le (1+\epsilon)^2$$

that finally gives

$$(1-\epsilon) \le \lambda(\mathbf{C}) = \lambda(\mathbf{B}_{\rm it}^{-1/2}\mathbf{B}\,\mathbf{B}_{\rm it}^{-1/2}) \le (1+\epsilon)\,.$$

The latter bounds are equivalent to (2.42). □

For instance, if we set ϵ to 0.5, we get an iterative preconditioner $\mathbf{B}_{\rm it}$ which is spectrally equivalent to $\mathbf{B}$ with rather good constants.

Corollary 2.1. *Let* $\epsilon = 0.5$ *and, instead of (2.27), we set*

$$\boldsymbol{\mathcal{K}}^{-1} = \sum_{s=1}^{J} \mathbf{T}_s \mathbf{B}_{s,\rm it}^{-1} \mathbf{T}_s^T\,. \tag{2.44}$$

Then, instead of (2.32) and (2.33), we have

$$0.5\underline{\gamma}\,(\boldsymbol{\mathcal{K}}\mathbf{v}, \mathbf{v}) \le (\mathbf{K}\mathbf{v}, \mathbf{v}) \le 1.5\overline{\gamma}\,(\boldsymbol{\mathcal{K}}\mathbf{v}, \mathbf{v})\,, \quad \forall\,\mathbf{v} \in R^N\,, \tag{2.45}$$

$$0.5\underline{\gamma}\,a(v,v) \le a(\widetilde{P}v, v) \le 1.5\overline{\gamma}\,a(v,v)\,, \quad \forall\,v \in \mathcal{V}\,, \tag{2.46}$$

with the same $\underline{\gamma}$ *and* $\overline{\gamma}$*.*

The matrices $\mathbf{B}_{\rm it}$ are not given explicitly. Suppose that, for solving $\mathbf{B}_s\mathbf{v}_s = \mathbf{f}_s$, we use the Chebyshev iterations

$$\mathbf{v}_s^k = \mathbf{v}_s^{k-1} - \sigma_{k,s}\boldsymbol{\mathcal{B}}_s^{-1}(\mathbf{B}_s\mathbf{v}_s^{k-1} - \mathbf{f}_s)\,,\ k = 1, 2, \ldots, \nu_\epsilon\,, \quad \mathbf{v}_s^0 = \mathbf{0}\,, \quad (2.47)$$

with a new preconditioner $\boldsymbol{\mathcal{B}}_s^{-1}$ and variable Chebyshev iteration parameters. Then the inverse to the iterative preconditioner can formally be represented in the form

$$\mathbf{B}_{s,\rm it}^{-1} = [\mathbf{I} - \prod_{k=1}^{\nu_\epsilon}(\mathbf{I} - \sigma_{k,s}\boldsymbol{\mathcal{B}}_s^{-1}\mathbf{B}_s)]\,\mathbf{B}_s^{-1}\,. \quad (2.48)$$

Evidently, in (2.47) and (2.48), the matrix $\mathbf{B}_s$ may be replaced by $\mathbf{K}_s$. However, even in the case when $\mathbf{K}_s$ and $\mathbf{B}_s$ are spectrally equivalent, this replacement may be not good, because the multiplications by $\mathbf{B}_s$ may be much cheaper.

Since in what follows inexact iterative solvers like in (2.47) and (2.48) will be often used in different components of DD algorithms, we introduce special notations for matrices which are implicitly defined by iterative processes.

Definition 2.1. Let $\mathbf{A}$ be a s.p.d. matrix, $\boldsymbol{\mathcal{B}}$ be its s.p.d.preconditioner, and let the system $\mathbf{Ax} = \mathbf{y}$ be solved inexactly by the Chebyshev iterative process

$$\mathbf{x}^k = \mathbf{x}^{k-1} - \sigma_k\boldsymbol{\mathcal{B}}^{-1}(\mathbf{A}\mathbf{x}^{k-1} - \mathbf{y})\,, \quad \mathbf{x}^0 = \mathbf{0}\,, \quad (2.49)$$

with Chebyshev iteration parameters σ_k. Then the notation $\boldsymbol{\mathcal{I}}\,[\mathbf{A}, \boldsymbol{\mathcal{B}}, \nu]$ stands for the matrix implicitly defined by ν Chebyshev iterations of the form (2.49) such that $\boldsymbol{\mathcal{I}}\,[\mathbf{A}, \boldsymbol{\mathcal{B}}, \nu]\,\mathbf{x}^\nu = \mathbf{y}$ and

$$(\boldsymbol{\mathcal{I}}\,[\mathbf{A}, \boldsymbol{\mathcal{B}}, \nu])^{-1} = [\mathbf{I} - \prod_{k=1}^{\nu}(\mathbf{I} - \sigma_k\boldsymbol{\mathcal{B}}^{-1}\mathbf{A})]\,\mathbf{A}^{-1}\,.$$

Besides, we set $\boldsymbol{\mathcal{I}}_\circ[\mathbf{A}, \boldsymbol{\mathcal{B}}] := \boldsymbol{\mathcal{I}}\,[\mathbf{A}, \boldsymbol{\mathcal{B}}, \nu_{1/2}]$, where ν_ϵ is the least number of iterations providing the bound

$$\|\mathbf{x} - \mathbf{x}^{\nu_\epsilon}\|_{\mathbf{A}} \le \epsilon\,\|\mathbf{x}\|_{\mathbf{A}}\,.$$

Suppose that the spectral equivalence inequalities

$$\underline{\gamma}\boldsymbol{\mathcal{B}} \le \mathbf{A} \le \overline{\gamma}\boldsymbol{\mathcal{B}}$$

hold with positive spectral equivalence constants $\underline{\gamma}$ and $\overline{\gamma}$. Let us choose the optimal Richardson iteration parameter for all σ_k, *i.e.* $\sigma_k = 2/(\underline{\gamma} + \overline{\gamma})$. Then we have

$$\epsilon = q^{\nu_\epsilon}\,, \quad q = \frac{1-\xi}{1+\xi}\,, \quad \xi = \underline{\gamma}/\overline{\gamma}\,, \quad (2.50)$$

and, therefore, we get

$$\nu_\epsilon = \mathrm{int}\lfloor \frac{\ln \epsilon^{-1}}{\ln q^{-1}} \rfloor_+ , \tag{2.51}$$

where $\mathrm{int}\lfloor\, a \,\rfloor_+$ denotes the closest to a integer not less than a. Since the derivative

$$\frac{d}{d\xi} \log \frac{1}{q} = \frac{2}{1-\xi^2} , \quad \xi \in (0,1) ,$$

is monotonically growing and

$$q\big|_{\xi=0} = 1 \quad \log \frac{1}{q}\Big|_{\xi=0} = 0 , \quad \frac{d}{d\xi} \log \frac{1}{q}\Big|_{\xi=0} = 2 ,$$

we arrive at the simpler expression

$$\nu_\epsilon \le \mathrm{int}\lfloor \frac{\ln \epsilon^{-1}}{2\xi} \rfloor_+ , \tag{2.52}$$

that can be taken as ν_ϵ. If we choose Chebyshev's iteration parameters σ_k, then we get the better bound $\epsilon = q_{\nu,\mathrm{PCG}}$ from (2.39), see also [Samarskii and Nikolayev (1989)].

In general, if $\mathbf{M}_s$ is the error transition matrix for an iterative solver for the subdomain Ω_s, then we have

$$\mathbf{B}_{s,\mathrm{it}}^{-1} = [\mathbf{I} - \mathbf{M}_s^{n_\epsilon}]\, \mathbf{B}_s^{-1} , \tag{2.53}$$

and a more careful analysis can result in better estimates than those given in (2.42). Indeed, let $\mathbf{K}_s$ corresponds to some H^1-elliptic Dirichlet problem in Ω_s. [Jung and Langer (1991)] studied multigrid preconditioners and, under rather general assumptions, they derived the estimates

$$\underline{\gamma}_{s,\mathrm{it}}\, \mathbf{B}_{s,\mathrm{it}} \le \mathbf{B}_s \le \overline{\gamma}_{s,\mathrm{it}}\, \mathbf{B}_{s,\mathrm{it}} ,$$

with $\underline{\gamma}_{s,\mathrm{it}} = 1 - \eta^{n_\epsilon}$ and $\overline{\gamma}_{s,\mathrm{it}} = 1$, where $\eta = \|\mathbf{M}_s\|_{\mathbf{B}_s}$. For instance, these estimates can be guaranteed by one ($n_\epsilon = 1$) symmetric V-cycle with appropriately chosen multigrid components such that $\mathbf{B}_{\mathrm{it},s}$ is s.p.d. The estimate for the upper bound is due to the fact that the multigrid iteration operators are non-negative with respect to the energy inner product under appropriate assumptions. The operation count for the local preconditioning operation gives $\mathrm{ops}\,[\mathbf{B}_{s,\mathrm{it}}^{-1}\mathbf{d}_s] = \mathcal{O}(N_s) = \mathcal{O}((H_s/h)^d)$.

In what follows, many examples of DD algorithms are considered, which assume the use of inexact solvers in all their components. At the same time, it is worth noting that a straight forward replacement in some components of secondary (with respect to the outer DD-PCG iterations) exact solvers

by inexact ones may produce a negative result. One example is the prolongation used in DD Dirichlet–Dirichlet algorithms. In these prolongations, systems with the same matrices, as for the local Dirichlet problems, may be solved, however more iterations are required, see, *e.g.* Lemma 4.1 below.

For h discretizations of elliptic problems with not significantly changing coefficients, the tool kit of optimal (linear complexity) and almost optimal preconditioner-solvers is rather well developed. The state of art in this area is reflected, *e.g.*, in the books of [Bramble and Zhang (2000); Toselli and Widlund (2005); Vassilevski (2008)] and many others and numerous journal publications. Therefore, apart from the what has been discussed above, many other inexact solvers may be used very efficiently for the problems in the subspaces. Most of them belong to the vast realm of multilevel solvers. At the same time, the situation is different for the *hp*-version, where the computational complexity of a preconditioner-solver depends on two discretization parameters. Fast solvers for local *hp* discrete problems as well as fast DD solvers for *hp* discretizations of elliptic equations started to appear quite recently. We add that the term *preconditioner-solver* becomes now rather usable and can reflect two different situations. It is applied to preconditioners $\mathbf{B}$, such that there exists good solvers for the systems $\mathbf{Bv} = \mathbf{f}$, and to preconditioners like $\mathbf{B}_{\rm it}$, which themselves are implicitly defined by means of some efficient (inexact) iterative solvers.

2.2.4 *Multiplicative Algorithms*

The Multiplicative Schwarz method (MSM) is a direct legatee of the Alternating Schwarz method as discussed in Section 1.2. The inexact version of the MSM corresponds to the space splitting (2.15) and to the subspace bilinear forms $b_s(\cdot,\cdot)$. It can be written in the form of an iteration process in the FE space $\mathcal{V}$ (function version) and in R^N (vector/matrix version), both of which are presented in Algorithm 2.3.

Introducing the MSM iteration operator

$$E = (I - \widetilde{P}_J)(I - \widetilde{P}_{J-1}) \cdots (I - \widetilde{P}_1)\,, \tag{2.54}$$

we conclude that the iteration error $z^n = u - u^n \in \mathcal{V}$ satisfies the iteration scheme

$$z^{n+1} = Ez^n\,. \tag{2.55}$$

Therefore, the error estimate has the form

$$\|z^{n+1}\|_a \le \|E\|_a \, \|z^n\|_a\,, \tag{2.56}$$

Algorithm 2.3 Inexact MSM: function ↔ vector/matrix versions.

$u^0 = \Phi \mathbf{u^0} \in \mathcal{V}$ given initial guess {initialization}

for $n = 0$ **step** 1 **until** Convergence {begin iteration loop}

for $s = 1$ **step** 1 **until** J {subspace corrections}

$$w_s^{n+s/J} = \Phi \mathbf{G}_s \mathbf{w}_s^{n+s/J} \in \mathcal{V}_s : \; b_s(w_s^{n+s/J}, v_s) = \langle f, v_s \rangle - a(u^{n+\frac{s-1}{p}}, v_s) \quad \forall v_s \in \mathcal{V}_s$$

$$\updownarrow$$

$$\mathbf{w}_s^{n+s/J} \in R^{N_s} : \qquad \mathbf{B}_s \mathbf{w}_s^{n+s/J} = \mathbf{T}_s^T (\mathbf{f} - \mathbf{K}\mathbf{u}^{n+\frac{s-1}{J}}) \equiv \mathbf{T}_s^T \mathbf{d}^{n+\frac{s-1}{J}}$$

$$u^{n+s/J} = u^{n+\frac{s-1}{J}} + w_s^{n+s/J} = u^{n+\frac{s-1}{J}} + \widetilde{P}_s (u - u^{n+\frac{s-1}{J}}) \quad \{\text{immediate}\}$$

$$\updownarrow \qquad \{\text{updating of the iterate}\}$$

$$\mathbf{u}^{n+s/J} = \mathbf{u}^{n+\frac{s-1}{J}} + \mathbf{T}_s \mathbf{w}_s^{n+s/J}$$

end for

end for {end iteration loop}

and convergence takes place, if

$$\|E\|_a = \|\mathbf{E}\|_{\mathbf{K}} \leq q_{\text{MSM}} < 1 \,, \tag{2.57}$$

where $\mathbf{E} = (\mathbf{I} - \mathbf{T}_J \mathbf{B}_J^{-1} \mathbf{T}_J^T \mathbf{K}) \cdots (\mathbf{I} - \mathbf{T}_1 \mathbf{B}_1^{-1} \mathbf{T}_1^T \mathbf{K})$ is the matrix representation of the operator E.

Obviously, the MSM preconditioner

$$\boldsymbol{\mathcal{K}} = \mathbf{K}(\mathbf{I} - \mathbf{E})^{-1} \tag{2.58}$$

is not necessarily symmetric, and for this reason is not suitable for the use as a preconditioner in the PCG method. However, repeating the subspace corrections in Algorithm 2.3 in the reverse direction, we arrive at the *symmetric multiplicative Schwarz method* . It is characterized by the iteration operator

$$E = (I - \widetilde{P}_1) \cdots (I - \widetilde{P}_{J-1})(I - \widetilde{P}_J)(I - \widetilde{P}_J)(I - \widetilde{P}_{J-1}) \cdots (I - \widetilde{P}_1) \tag{2.59}$$

with the matrix form

$$\begin{aligned} \mathbf{E} = (\mathbf{I} - \mathbf{T}_1 \mathbf{B}_1^{-1} \mathbf{T}_1^T \mathbf{K}) \cdots (\mathbf{I} - \mathbf{T}_J \mathbf{B}_J^{-1} \mathbf{T}_J^T \mathbf{K}) \times \\ (\mathbf{I} - \mathbf{T}_J \mathbf{B}_J^{-1} \mathbf{T}_J^T \mathbf{K}) \cdots (\mathbf{I} - \mathbf{T}_1 \mathbf{B}_1^{-1} \mathbf{T}_1^T \mathbf{K}). \end{aligned} \tag{2.60}$$

Indeed, the MSM preconditioner (2.58) is now symmetric.

Evidently, if $b_s(\cdot,\cdot) = a_s(\cdot,\cdot)$, *i.e.,* $\mathbf{B}_s = \mathbf{K}_s := \mathbf{T}_s^T \mathbf{K} \mathbf{T}_s$ and consequently $\widetilde{P}_s := P_s$, we have the exact MSM. In the exact version of symmetric MSM, according to (2.59), the subspace correction in $\mathcal{V}_J$ has to be carried out only once, because $(I - P_J)(I - P_J) = (I - P_J)$.

2.2.5 *Hybrid Algorithms*

The ASM is suitable for parallelization and simpler for the computer realization, whereas MSM usually provides faster convergence. In many cases, a combination of these two methods may be advantageous. This idea leads to *Hybrid Schwarz Methods (HSM)*, where the subspace corrections are completed in part in additive and in part in multiplicative modes. There are a multitude of useful hybrid algorithms corresponding to different possibilities of the arrangement of the subspace corrections in additive and multiplicative manners. For instance, Algorithm 2.4, which presents only the function version, is a symmetric HSM with the iteration operator

$$E = (I - \widetilde{P}_1)(I - \tau(\widetilde{P}_2 + \cdots + \widetilde{P}_p))(I - \widetilde{P}_1)\,. \tag{2.61}$$

This iteration operator corresponds to a symmetric preconditioner (2.58), whereas the iteration operator, as well as the corresponding iteration matrix, are self-adjoint in the respective energy inner products. Some versions of the hybrid Schwarz algorithms were studied by [Mandel and Lett (1991)] and [Cai (1993)].

Algorithm 2.4 A Hybrid Schwarz Method which is equivalent to (2.61).

given initial guess $u^0 = \Phi\mathbf{u}^0 \in \mathcal{V}$ and iteration parameter τ initialization

for $n = 0$ **step** 1 **until** Convergence **do** {begin iteration loop}

$w_1^{n,1} \in \mathcal{V}_1 :\ b_1(w_1^{n,1}, v_1) = \langle f, v_1\rangle - a(u^n, v_1), \quad \forall v_1 \in \mathcal{V}_1,$
$u^{n,1} = u^n + w_1^{n,1} = u^n + \widetilde{P}_1(u - u^n),$ {1-st multiplicative correction}

for all $s = \overline{2, J}$ **in parallel do** {additive subspace corrections}

$w_s^{n,1} \in \mathcal{V}_s :\ b_s(w_s^{n,1}, v_s) = \langle f, v_s\rangle - a(u^{n,1}, v_s)\,, \quad \forall v_s \in \mathcal{V}_s\,,$
end for

$u^{n,2} = u^{n,1} + \tau \sum_{s=2}^{J} w_s^{n,1} = u^{n,1} + \tau \sum_{s=2}^{J} \widetilde{P}_s(u - u^{n,1})\,,$ {updating}
$w_1^{n,2} \in \mathcal{V}_1 :\ b_1(w_1^{n,2}, v_1) = \langle f, v_1\rangle - a(u^{n,2}, v_1), \quad \forall v_1 \in \mathcal{V}_1,$
$u^{n+1} = u^{n,2} + w_1^{n,2} = u^{n,2} + \widetilde{P}_1(u - u^{n,2}),$ {2-nd multipl. correction}
end for {end iteration loop}

2.2.6 *Jacobi and Gauss-Seidel Iterations*

Jacobi and Gauss-Seidel iterations may be interpreted as simplest samples of exact additive and multiplicative Schwarz methods, in which the subspaces of decomposition are one dimensional and each is spanned by one element ϕ_s of the FE basis Φ, *i.e.*,

$$\mathcal{V} = \sum_{s=1}^{N} \mathcal{V}_s\,, \quad \mathcal{V}_s = \operatorname{span}[\phi_s]\,, \quad N_s = \dim \mathcal{V}_s = 1\,. \tag{2.62}$$

In this case $\mathbf{T}_s = \mathbf{e}_s$ is the s-th unit vector $\mathbf{e}_s = (0,\ldots,0,1,0,\ldots,0)^T$ and

$$\mathbf{T}_s^T \mathbf{K} \mathbf{T}_s = \mathbf{e}_s^T \mathbf{K} \mathbf{e}_s = k_{s,s} \tag{2.63}$$

is the diagonal element of the stiffness matrix $\mathbf{K}$. Therefore,

$$\boldsymbol{\mathcal{K}}^{-1} = \sum_{s=1}^{J} \mathbf{e}_s k_{s,s}^{-1} \mathbf{e}_s^T = \mathbf{D}_{\mathbf{K}}^{-1}\,, \quad \mathbf{D}_{\mathbf{K}} := \operatorname{diag}(\mathbf{K})\,, \tag{2.64}$$

and we observe that the exact ASM (2.26), corresponding to the above decomposition, coincides with the well known (damped) Jacobi method. Furthermore, the exact MSM, corresponding to the basis splitting (2.62), gives us the Gauss-Seidel method. Indeed, from the exact version of the MSM Algorithm 2.3 and from (2.63), we see that the s-th subspace correction step in the n-th iteration step

$$\mathbf{u}^{n+s/N} = \mathbf{u}^{n+\frac{s-1}{N}} + \mathbf{T}_s \mathbf{w}_s^{n+s/N} = \mathbf{u}^{n+\frac{s-1}{N}} + \mathbf{e}_s (\mathbf{e}_s^T \mathbf{K} \mathbf{e}_s)^{-1} \mathbf{e}_s^T (\mathbf{f} - \mathbf{K}\mathbf{u}^{n+\frac{s-1}{N}}) \tag{2.65}$$

updates only the s-th component of the iterate according to the formula

$$u_s^{n+s/N} = \frac{1}{k_{s,s}} \left(f_s - \sum_{j=1}^{s-1} k_{s,j} u_j^{n+\frac{s-1}{N}} - \sum_{j=s+1}^{N} k_{s,j} u_j^{n+\frac{s-1}{N}} \right)\,, \tag{2.66}$$

where $s = 1, 2, \ldots, N$.

Clearly, we may look at ASM and MSM as the natural generalizations of the Jacobi and Gauss-Seidel methods. In a similar way, the symmetric Gauss-Seidel method falls into the category of exact symmetric MSM. Conversely, analogously with SOR and SSOR methods, one can introduce overrelaxation parameters into MSM and symmetric MSM in order to hasten the convergence. A discussion of these and other related subjects may be found in [Griebel and Oswald (1995a)].

2.3 Main Factors Influencing Convergence

In this section, we present the basic convergence results for Schwarz methods. The general conditions to be fulfilled for good DD methods (preconditioners), as they are formulated now, are the conditions of *stability and boundness of the space splitting* into subspaces. At first, for an illustrative example, we consider the case of splitting the space into a direct sum of two subspaces, which exhibits some features of the behavior of DD preconditioners common with more general cases.

2.3.1 *Convergence Analysis in the Case of Two Subspaces*

For the simplest example, let us consider the splitting of the space

$$\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2\,, \quad \text{so that} \quad \mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}\,, \tag{2.67}$$

into the direct sum of two non-trivial subspaces $\mathcal{V}_1 = \text{span}\Phi\,\mathbf{T}_1$ and $\mathcal{V}_2 = \text{span}\Phi\,\mathbf{T}_2$, and let γ be the cosine of the angle between $\mathcal{V}_1$ and $\mathcal{V}_2$:

$$\gamma = \cos\measuredangle(\mathcal{V}_1, \mathcal{V}_2) = \sup_{v_1\in\mathcal{V}_1\setminus\{0\},\, v_2\in\mathcal{V}_2\setminus\{0\}} \frac{a(v_1, v_2)}{\|v_1\|_a\,\|v_2\|_a} < 1\,. \tag{2.68}$$

Evidently, the constant γ coincides with the sharp constant in the strengthened Cauchy inequality:

$$|a(v_1, v_2)| \le \gamma\,\|v_1\|_a\,\|v_2\|_a\,, \quad \forall v_1 \in \mathcal{V}_1,\ \forall v_2 \in \mathcal{V}_2\,. \tag{2.69}$$

In the finite element method, the spaces depend on the discretization parameters such as mesh size, or, equivalently, on the number of elements, and the degree of polynomials on the associated reference elements. In particular, the numbers of degrees of freedom N, equal to dimensions of the FE spaces, depend on these parameters. The splitting (2.67) is called stable iff the constant γ stays less than 1 uniformly for $N \to \infty$, *i.e.*, uniformly then $h \to 0$, for h-version, and additionally for $p \to \infty$ in the case of the hp-version. The following simple lemma is useful for computing or estimating γ.

Lemma 2.2. *For the value of γ, the following characterizations hold:*

$$\gamma = \cos\measuredangle(\mathcal{V}_1^\perp, \mathcal{V}_2^\perp) = \sup_{v_1\in\mathcal{V}_1\setminus\{0\},\, v_2\in\mathcal{V}_2\setminus\{0\}} \frac{2a(v_1, v_2)}{\|v_1\|_a^2 + \|v_2\|_a^2}\,, \tag{2.70}$$

$$\gamma^2 = \sup_{\mathbf{v}_1\in R^{N_1}\setminus\{\mathbf{0}\}} \frac{((\mathbf{T}_1^\top\,\mathbf{K}\,\mathbf{T}_2)(\mathbf{T}_2^\top\,\mathbf{K}\,\mathbf{T}_2)^{-1}(\mathbf{T}_2^\top\,\mathbf{K}\,\mathbf{T}_1)\,\mathbf{v}_1,\,\mathbf{v}_1)}{(\mathbf{T}_1^\top\,\mathbf{K}\,\mathbf{T}_1\,\mathbf{v}_1,\,\mathbf{v}_1)}\,, \tag{2.71}$$

$$\gamma = 1 - \inf_{v_1\in\mathcal{V}_1\setminus\{0\},\, v_2\in\mathcal{V}_2\setminus\{0\}} \frac{a(v_1+v_2, v_1+v_2)}{\|v_1\|_a^2 + \|v_2\|_a^2}\,. \tag{2.72}$$

Proof. For simplifying notations, let us set

$$\mathbf{K}_l = \mathbf{T}_l^\top \mathbf{K}\, \mathbf{T}_l\,, \quad l = 1, 2,, \quad \mathbf{K}_{lm} = \mathbf{T}_l^\top \mathbf{K}\, \mathbf{T}_m\,, \quad l \neq m\,,$$

and consider

$$\Theta(\mathbf{v}_1) = \sup_{\mathbf{v}_2 \in R^{N_2} \setminus \{\mathbf{0}\}} \frac{\left(\mathbf{v}_2^\top \mathbf{K}_{21} \mathbf{v}_1\right)^2}{\mathbf{v}_2^\top \mathbf{K}_2 \mathbf{v}_2} = \sup_{\mathbf{w}_2 \in R^{N_2} \setminus \{\mathbf{0}\}} \frac{\left(\mathbf{w}_2^\top \mathbf{K}_2^{-1/2} \mathbf{K}_{21} \mathbf{v}_1\right)^2}{\mathbf{w}_2^\top \mathbf{w}_2}\,, \tag{2.73}$$

where $\mathbf{w}_2 = \mathbf{K}_2^{1/2} \mathbf{v}_2$. For $\forall\, \mathbf{z} \in R^n$, we have

$$\sup_{\mathbf{y} \in R^n \setminus \{\mathbf{0}\}} \frac{\mathbf{y}^\top \mathbf{z}\mathbf{z}^\top \mathbf{y}}{\mathbf{y}^\top \mathbf{y}} = \|\mathbf{z}\|^2\,,$$

and, therefore, picking up $n = N_2,\ \mathbf{y} = \mathbf{w}_2,\ \mathbf{z} = \mathbf{K}_2^{-1/2} \mathbf{K}_{21} \mathbf{v}_1$ and taking into account (2.73), we get

$$\Theta(\mathbf{v}_1) = \|\mathbf{K}_2^{-1/2} \mathbf{K}_{21} \mathbf{v}_1\|^2 = \mathbf{v}_1^\top \mathbf{K}_{12} \mathbf{K}_2^{-1} \mathbf{K}_{21} \mathbf{v}_1\,.$$

Now (2.71) follows from the relation above, (2.73) and (2.67).

For the proof of (2.70), we note that, from the inequality $2ab \leq a^2 + b^2,\ \forall\, a, b$, it follows that

$$\gamma \geq \sup_{v_1 \in \mathcal{V}_1 \setminus \{0\},\, v_2 \in \mathcal{V}_2 \setminus \{0\}} \frac{2a(v_1, v_2)}{\|v_1\|_a^2 + \|v_2\|_a^2}\,. \tag{2.74}$$

However, if for some v_1 and v_2 we have the equality in (2.69), then it is retained for $\overline{v}_k = v_k / \|v_k\|_a\ k = 1, 2$. Therefore, for $\overline{v}_k$ we have also the equality in (2.74), and (2.70) is proved. Clearly, in view of (2.70), we have

$$\inf_{v_1 \in \mathcal{V}_1 \setminus \{0\},\, v_2 \in \mathcal{V}_2 \setminus \{0\}} \frac{a(v_1 + v_2, v_1 + v_2)}{\|v_1\|_a^2 + \|v_2\|_a^2}$$

$$= 1 - \sup_{v_1 \in \mathcal{V}_1 \setminus \{0\},\, v_2 \in \mathcal{V}_2 \setminus \{0\}} \frac{2a(v_1, v_2)}{\|v_1\|_a^2 + \|v_2\|_a^2} = 1 - \gamma\,, \tag{2.75}$$

which is another form of (2.72). □

Original independent proofs of this lemma were given by [Axelsson and Vassilevski (1989); Bjørstad and Mandel (1991); and Haase *et al.* (1991)].

The relation (2.71) that is nothing but the Rayleigh quotient means that γ coincides with the maximal eigenvalue of the generalized eigenvalue problem

$$(\mathbf{T}_1^\top \mathbf{K} \mathbf{T}_2)(\mathbf{T}_2^\top \mathbf{K} \mathbf{T}_2)^{-1} (\mathbf{T}_2^\top \mathbf{K} \mathbf{T}_1) \mathbf{v}_1 = \lambda\, (\mathbf{T}_1^\top \mathbf{K} \mathbf{T}_1) \mathbf{v}_1. \tag{2.76}$$

The exact ASM preconditioner has the form

$$\mathcal{K} = \mathbf{T}^{-\top} \begin{pmatrix} \mathbf{T}_1^\top \mathbf{K} \mathbf{T}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{T}_2^\top \mathbf{K} \mathbf{T}_2 \end{pmatrix} \mathbf{T}^{-1}, \tag{2.77}$$

where $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)$ is a nondegenerate $N \times N$ basis transformation matrix. As it directly follows from (2.70), the exact ASM preconditioner $\mathcal{K}$ satisfies spectral equivalence inequalities (2.32 with the sharp constants

$$\underline{\gamma} = 1 - \gamma \quad \text{and} \quad \overline{\gamma} = 1 + \gamma\,, \tag{2.78}$$

see also (2.75) for $\underline{\gamma}$.

Lemma 2.3. *Assume $\gamma \in [0, 1)$. Let also $\tau = \tau_{opt} = 1$ in ASM, and u^1 is chosen such that $z^1 = (I - P_2)z^0 \in \mathcal{V}_2^\perp$ (initial orthoprojection step onto $\mathcal{V}_2$) in MSM. Then the exact Schwarz methods converge according to the estimates*

$$\|u - u^{n+1}\|_a \le \gamma^k \|u - u^n\|_a\,, \quad n = 1, 2, \dots\,, \tag{2.79}$$

with $k = 1$ for ASM and $k = 2$ for MSM.

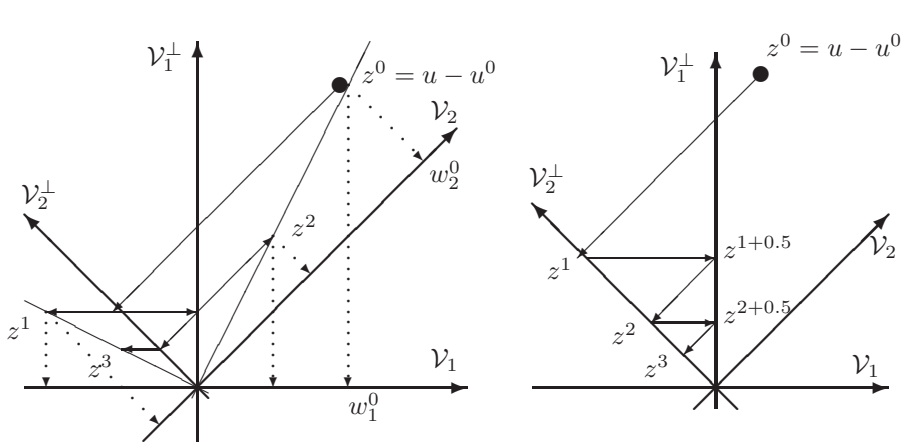

Fig. 2.1 ASM and MSM in the case of the splitting in two subspaces.

The error iteration schemes $z^{n+1} = Ez^n$, corresponding to these exact ASM and exact MSM, are illustrated at the left and right sides of Figure 2.1, respectively.

2.3.2 *Convergence of Additive Schwarz Algorithms*

In the case of the exact ASM, we denote the bounds of the spectrum of the matrix $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$ by $\underline{\beta}$ and $\overline{\beta}$, which means that, for the operator $P = P_1 + P_2 + ... + P_j$, we have

$$\underline{\beta}\, a(v,v) \leq a(Pv,v) \leq \overline{\beta}\, a(v,v)\,, \quad \forall v \in \mathcal{V}\,. \tag{2.80}$$

The values of $\underline{\beta}$, $\overline{\beta}$ characterize the *stability of the splitting*. Evidently, $\overline{\beta} \leq J$ since the energy norm of P_s is the unity. The conditions more adapted to the bounding of $\underline{\beta}$, were formulated by [Matsokin and Nepomnyaschikh (1985)].

Theorem 2.1. *Let the Hilbert space H, with the scalar product $(\cdot,\cdot)$ and the norm $\|\cdot\| = \sqrt{(\cdot,\cdot)}$, be split into the vector sum $H = H_1 + H_2 + ... + H_J$. Let p_s be the orthogonal projection on H_s and $\alpha \in [1,\infty)$. Then the two following conditions are equivalent:*

1) for any $v \in H$, there exists a representation $v = v_1 + v_2 + ... + v_J$, with $v_s \in H_s$, such that

$$\sum_{s=1}^{J} \| v_s \|^2 \leq \alpha \ \| v \|^2\,, \tag{2.81}$$

2) for $p = p_1 + p_2 + ... + p_J$ and any $v \in H$, we have

$$\alpha^{-1}\,(v,v) \leq (p\,v,v)\,. \tag{2.82}$$

Proof. Taking into account the definition of the operators p_s and (2.81), we have

$$\| v \|^2 = (v,v) = (v, v_1 + v_2 + ... + v_J) = \sum_s (p_s v, v_s) \leq$$

$$\leq (\sum_s (p_s v, v))^{1/2} (\sum_s (v_s, v_s))^{1/2} \leq \sqrt{\alpha}\, \|v\| (\sum_s (p_s v, v))^{1/2}\,, \tag{2.83}$$

proving that (2.82) is the consequence of (2.81).

In order to prove that (2.81) follows from (2.82), we note that the operator p is selfadjoint, has the upper eigenvalue bounded from above by

J and the lower eigenvalue bounded from below accordingly with (2.82). Therefore, it is a one-to-one mapping of H onto H, and, for each $v \in H$, there is some $w \in H$ such that $v = p\,w$. Now, for the representation $v = v_1 + v_2 + ... + v_J,\ v_s = p_s w$, we have

$$\sum_s (v_s, v_s) = \sum_s (p_s w, p_s w) = \sum_s (p_s w, w) = (v, w) = \\ = (v, p^{-1} v) \le \| p^{-1} \| (v, v) \le \alpha (v, v)\,, \tag{2.84}$$

which completes the proof. □

This theorem establishes that, if the condition 1) is fulfilled for $\| \cdot \| \equiv \| \cdot \|_a$, then $\lambda_{\min}(P) = \lambda_{\min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \ge \underline{\beta} = \alpha^{-1}$. Indeed, it is exactly the statement of the theorem, if H, H_s, $(\cdot, \cdot)$, $\| \cdot \|$ and p_s, p stand respectively for $\mathcal{V}$, $\mathcal{V}_s$, $a(\cdot, \cdot)$, $\| \cdot \|_a$ and P_s, P, introduced in the preceding subsection.

Let us note that if $H = H_1 + H_2 + ... + H_J$ is a direct sum of subspaces, then (2.81) obviously holds with $\alpha > 0$. That part of the above theorem, concerning the bound $\underline{\beta}$ through α, was obtained independently by [Lions (1988)]. For the case of the inexact ASM, which is considered below, the lower and upper bounds $\underline{\gamma}$ and $\overline{\gamma}$ were given in [Nepomnyaschikh (1991c)] and later by [Dryja *et al.* (1994)] under slightly different assumptions.

Now, we consider the situation when the Hilbert space H with the original scalar product $(\cdot, \cdot)$ is split in the vector sum $H = H_1 + H_2 + ... + H_J$, and we are given a linear, selfadjoint and positive definite operator $A : H \to H$ inducing the bilinear form $(v, w)_A \equiv a(v, w) \equiv (Av, w)$ and the norm $\| v \|_a = \sqrt{a}(v, v)$. The operators $P_s : H \to H_s$ denote the orthogonal projections with the respect of the scalar product $a(\cdot, \cdot)$.

Theorem 2.2. *Let there exist two positive constants α and $\overline{\beta}$ for which the following two conditions hold:*

1) for any $v \in H$ there exist the representation $v = v_1 + v_2 + ... + v_J$, with $v_s \in H_s$, such that

$$\sum_s \| v_s \|_a^2 \le \alpha \ \| v \|_a^2\,, \tag{2.85}$$

2) the operators P_s satisfy the inequality

$$((P_1 + P_2 + ... + P_J)v, v)_A \le \overline{\beta}\,(v, v)_A\,, \quad \forall\, v \in H. \tag{2.86}$$

Let also selfadjoint operators $B_s : H \to H_s$ be such that for positive constants $\underline{\gamma}_s$ and $\overline{\gamma}_s$ the spectral equivalence inequalities

$$\underline{\gamma}_s (B_s v_s, v_s) \le (v_s, v_s)_A \le \overline{\gamma}_s (B_s v_s, v_s)\,, \quad \forall\, v \in H_s\,, \tag{2.87}$$

are valid, cf. also (2.17). Then, for the constants $\underline{\gamma} \le \alpha \min_s \underline{\gamma}_s$ *and* $\overline{\gamma} \le \overline{\beta} \max_s \overline{\gamma}_s$*, we have the spectral equivalence inequalities*

$$\underline{\gamma}(A^{-1}v, v) \le (B^{-1}v, v) \le \overline{\gamma}(A^{-1}v, v)\,, \quad \forall\, v \in H\,, \tag{2.88}$$

where $B^{-1} = B_1^+ + B_2^+ + ... + B_J^+$ *and* B_s^+ *denotes the pseudo inverse operator for* B_s*.*

Proof. Let us first mention that inequalities (2.80) with $\overline{\beta}$ from (2.86) are a consequence of Theorem 2.1 and (2.86). The operator P_s can be represented in the form $P_s = (p_s A p_s)^+ A$, where $p_s : H \to H_s$ is the operator of the orthogonal projection with the respect to the scalar product $(\cdot, \cdot)$. Substituting the above expression for P_s and $v = A^{-1}w$ into (2.80) and then changing notation w for v, we get

$$\alpha(A^{-1}v, v) \le \sum_s ((p_s A p_s)^+ v, v) \le \overline{\beta}(A^{-1}v, v)\,, \quad \forall\, v \in H\,. \tag{2.89}$$

By the definition of p_s we have $(v_s, v_s)_A \equiv (Av_s, v_s) \equiv (p_s A p_s v_s, v_s)$, and, therefore, (2.87) may be rewritten in the form

$$\underline{\gamma}_s(B_s v_s, v_s) \le (p_s A p_s v_s, v_s) \le \overline{\gamma}_s(B_s v_s, v_s)\,, \quad \forall\, v \in H_s\,,$$

which is equivalent to

$$\underline{\gamma}_s((p_s A p_s)^+ v_s, v_s) \le (B_s^+ v_s, v_s) \le \overline{\gamma}_s((p_s A p_s)^+ v_s, v_s)\,, \quad \forall\, v \in H_s\,.$$

Substituting $v_s = p_s v$ we obtain

$$\underline{\gamma}_s((p_s A p_s)^+ v, v) \le (B_s^+ v, v) \le \overline{\gamma}_s((p_s A p_s)^+ v, v)\,, \quad \forall\, v \in H\,. \tag{2.90}$$

Combining (2.89) and (2.90) yields (2.88). □

The conditions leading to the left inequality (2.88), *i.e.* (2.85) and the left inequality (2.87), can be replaced by one, which is conveniently formulated by means of the bilinear forms $b_s(v, w,) \equiv (B_s v, w)$, defined for $v, w \in H_s$ and the corresponding norms $\|v\|_s = \sqrt{b_s(v, v)}$.

Lemma 2.4. *Let, for any* $v \in H$*, there exist a splitting* $v = v_1 + v_2 + ... + v_J$*,* $v_s \in H_s$ *such that*

$$\sum_s \|v_s\|_s^2 \le \alpha \|v\|^2\,, \tag{2.91}$$

and let $\widetilde{P}_s : H \to H_s$ *be the projection-like operator defined by the identity*

$$b_s(\widetilde{P}_s u, v) = a(u, v)\,, \quad \forall\, u \in H\,, \quad \forall\, v \in H_s\,. \tag{2.92}$$

Then, for $\underline{\gamma} = 1/\alpha$*, we have*

$$\underline{\gamma}\; a(v, v) \le a((\widetilde{P}_1 + \widetilde{P}_2 + ... + \widetilde{P}_J)v, v)\,, \quad \forall\, v \in H\,. \tag{2.93}$$

Proof. First of all, let us note that

$$\sum_s a(\widetilde{P}_s v, w) = \sum_s b_s(\widetilde{P}_s v, \widetilde{P}_s w)\,, \quad \forall\ v, w \in H\,.$$

Therefore, similarly with (2.83), we can write

$$\| v \|^2 = a(v,v) = a(v, v_1 + v_2 + ... + v_J) = \sum_s b_s(\widetilde{P}_s v, v_s)$$

$$\leq (\sum_s \|\widetilde{P}_s v\|_s^2)^{1/2} (\sum_s \|v_s\|_s^2)^{1/2} \leq \alpha \|v\| \sum_s \|\widetilde{P}_s v\|_s = \alpha \|v\| \sum_s a(\widetilde{P}_s v, v)\,, \tag{2.94}$$

that completes the proof of the lemma. □

The bound $\overline{\beta} \leq J$ and, as a consequence, the bound for $\overline{\gamma}$ can be improved. Let us suppose that the domains of definition of spaces H and H_s are $\operatorname{supp} H = \Omega$ and $\operatorname{supp} H_s = \Omega_s$, respectively, implying that the space decomposition corresponds to some domain decomposition. Suppose additionally that $J_O = \max_{x \in \overline{\Omega}} J_O(x)$, where $J_O(x)$ is the number of the overlapping subdomains at the point x. We can write

$$a^2(v, \sum_{s=1}^J \widetilde{P}_s v) \leq a(v,v) a(\sum_{s=1}^J \widetilde{P}_s v, \sum_{s=1}^J \widetilde{P}_s v) \leq J_O a(v,v) \sum_{s=1}^J a(\widetilde{P}_s v, \widetilde{P}_s v) \leq$$

$$\leq J_O \max_s(\overline{\gamma}_s) a(v,v) \sum_{s=1}^J b_s(\widetilde{P}_s v, \widetilde{P}_s v) = J_O \max_s(\overline{\gamma}_s) a(v,v) \sum_{s=1}^J a(v, \widetilde{P}_s v)\,, \tag{2.95}$$

concluding that

$$a(v, \sum_{s=1}^J \widetilde{P}_s v) \leq \overline{\gamma} a(v,v)\,, \quad \overline{\gamma} \leq J_O \max_s(\overline{\gamma}_s)\,. \tag{2.96}$$

Still another estimate for $\overline{\gamma}$ is used in the analysis of DD algorithms which is based on the so called subspace interaction condition, introduced by [Dryja *et al.* (1994)]. In the result formulated below, we use this condition and consider the situation most often met in DD algorithms, when the subspaces $H_1, ..., H_J$ correspond to some domain decomposition, *i.e.*, these subspaces are the Dirichlet restrictions.[1] Besides, there is an additional subspace, usually denoted by H_0, which is generated by some discretization on the coarse grid usually defined by the domain decomposition. The so-called coarse grid space H_0 serves for improving the stability.

[1] We remind that the Dirichlet restriction of the the space $\mathcal{V}(\Omega)$ to Ω_s is nothing but the subspace of functions $v \in \mathcal{V}(\Omega)$ with $\operatorname{supp} v = \Omega_s$.

Theorem 2.3. *Let $H = H_0 + H_1 + ... + H_J$, and let the numbers $\gamma_{i,j} \leq 1$ be the constants from the strengthened Cauchy inequalities*

$$(v_i, v_j)_A \leq \gamma_{i,j} (v_i, v_i)_A^{1/2} (v_j, v_j)_A^{1/2}, \quad \forall\, v_i \in H_i,\ \forall\, v_j \in H_j, \tag{2.97}$$

for $i, j = 1, \ldots, J$. Then the right inequality in (2.88) holds with $\overline{\gamma} = \max_s(\overline{\gamma}_s)[\rho(\boldsymbol{\Upsilon}) + 1]$, or, equivalently,

$$((\widetilde{P}_0 + \widetilde{P}_1 + ... + \widetilde{P}_J)v, v)_A \leq \max_s(\overline{\gamma}_s)[\rho(\boldsymbol{\Upsilon}) + 1](v, v)_A\,, \tag{2.98}$$

where $\rho(\boldsymbol{\Upsilon})$ is the spectral radius of the subspace interaction matrix $\boldsymbol{\Upsilon} = (\gamma_{i,j})_{i,j=1}^J$ that is alternatively termed subspace interaction measure*.*

Proof. First of all, we obviously have the representation

$$a(v, \widetilde{P}v) = a(v, \widetilde{P}_0 v) + a(v, \sum_{s=1}^{J} \widetilde{P}_s v)\,. \tag{2.99}$$

Combining the inequalities

$$a^2(v, \widetilde{P}_0 v) \leq a(\widetilde{P}_0 v, \widetilde{P}_0 v)\, a(v, v)$$

and

$$a(\widetilde{P}_0 v, \widetilde{P}_0 v) \leq \overline{\gamma}_0\, b_0(\widetilde{P}_0 v, \widetilde{P}_0 v) = \overline{\gamma}_0\, a(v, \widetilde{P}_0 v)\,,$$

we arrive at the following estimate of the first term of the right-hand side in (2.99):

$$a(v, \widetilde{P}_0 v) \leq \overline{\gamma}_0\, a(v, v)\,, \tag{2.100}$$

Analogously to (2.96), but with the use of the more accurate estimate

$$a(\sum_{s=1}^{J} \widetilde{P}_s v, \sum_{s=1}^{J} \widetilde{P}_s v) \leq \rho(\boldsymbol{\Upsilon}) \sum_{s=1}^{J} a(\widetilde{P}_s v, \widetilde{P}_s v)\,,$$

we get

$$a(v, \sum_{s=1}^{J} \widetilde{P}_s v) \leq \max_s(\overline{\gamma}_s)\rho(\boldsymbol{\Upsilon}) a(v, v)\,. \tag{2.101}$$

Finally, we arrive at (2.98) by combining (2.99) with (2.100) and (2.101). □

2.3.3 *Exact Representation of the Minimal and Maximal Eigenvalues of $\mathcal{K}^{-1}\mathbf{K}$*

In the part related to the rate of convergence, the quality of the inexact ASM preconditioner is defined by the bounds of the spectrum of the preconditioned matrix $\mathcal{K}^{-1}\mathbf{K}$ or, equivalently, the operator $\widetilde{P}$. A variety of approaches for the analysis of these bounds, in particular, those presented above, are based on the properties, which are reflected in the fundamental notions of stability and boundness of the space decomposition One of the condensed characterizations of these properties and, at the same time, exact representations of the minimal and maximal eigenvalues of the preconditioned matrix $\mathcal{K}^{-1}\mathbf{K}$ is obtained by means of the so called splitting norm

$$|||\ v\ |||^2 = \inf_{v=\sum_{s=1}^{J} v_s} \sum_{s=1}^{J} b_s(v_s, v_s) = \inf_{v=\sum_{s=1}^{J} v_s} \sum_{s=1}^{J} \|v_s\|_s^2 \,, \tag{2.102}$$

where the infimum has taken over all possible splittings $v = \sum_{s=1}^{J} v_s$ with $v_s \in \mathcal{V}_s$.

Theorem 2.4. *Let $\mathcal{K}$ be the ASM preconditioner. Then*

$$\lambda_{min}(\mathcal{K}^{-1}\mathbf{K}) \le \frac{a(v,v)}{|||\ v\ |||^2} \le \lambda_{max}(\mathcal{K}^{-1}\mathbf{K}) \,, \quad \forall\ v \in \mathcal{V} \,, \tag{2.103}$$

with

$$\lambda_{min}(\mathcal{K}^{-1}\mathbf{K}) = \lambda_{min}(\widetilde{P}) = \min_{v \in \mathcal{V}} \frac{a(v,v)}{|||\ v\ |||^2} \,, \tag{2.104}$$

and

$$\lambda_{max}(\mathcal{K}^{-1}\mathbf{K}) = \lambda_{max}(\widetilde{P}) = \max_{v \in \mathcal{V}} \frac{a(v,v)}{|||\ v\ |||^2} \,. \tag{2.105}$$

Proof. As soon as (2.33) has been established, the proof follows from the characterization

$$|||\ v\ |||^2 = a(\widetilde{P}^{-1}v, v) \,, \tag{2.106}$$

which we derive first. We have

$$a(\widetilde{P}^{-1}v, v) = \sum_s a(\widetilde{P}^{-1}v, v_s) = \sum_s b_s(\widetilde{P}_s\widetilde{P}^{-1}v, v_s) \le$$

$$\le (\sum_s b_s(\widetilde{P}_s\widetilde{P}^{-1}v, \widetilde{P}_s\widetilde{P}^{-1}v))^{1/2} (\sum_s b_s(v_s, v_s))^{1/2} =$$

$$= (\sum_s a(\widetilde{P}_s \widetilde{P}^{-1} v, \widetilde{P}^{-1} v))^{1/2} (\sum_s b_s(v_s, v_s))^{1/2}$$

$$= a(\widetilde{P}^{-1} v, v)^{1/2} (\sum_s b_s(v_s, v_s))^{1/2} ,$$

and therefore

$$a(\widetilde{P}^{-1} v, v) \leq \min_{\sum_s v_s = v} \sum_s b_s(v_s, v_s) .$$

Now, it is not difficult to see that the minimum is achieved for $v_s = \widetilde{P}_s \widetilde{P}^{-1} v$, proving (2.106). Substituting (2.106) in (2.104)–(2.105) after some elementary transformations, we conclude that (2.104)–(2.105) is equivalent to (2.33). □

The upper bounds given earlier for the spectrums of the operators P and $\widetilde{P}$ may be obtained as direct consequences of the representation formula (2.105). These remarkable representation formulas for the minimal and maximal eigenvalues of the preconditioned matrix $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$ appeared in [Bjøstad and Mandel (1991); Oswald (1994); Xu (1992); and Zhang (1991, 1992)].

2.3.4 *Convergence of Multiplicative Schwarz Algorithms*

The convergence analysis of inexact multiplicative or hybrid Schwarz methods is more involved, in part due to the lack of symmetry of the corresponding preconditioners. Here we only present two results concerning the inexact SMSM and the inexact MSM, introduced for the finite element discretizations in Subsection 2.2.4, without proofs. For more results and the proofs, we refer the reader to special papers on this topic, *e.g.*, [Griebel and Oswald (1995a)]

We need the so-called subspace contraction condition, stating that the inequalities

$$a(v_s, v_s) \leq b_s(\widetilde{P}_s v_s, v_s) \leq \omega\, b_s(v_s, v_s) , \ \forall\, v_s \in \mathcal{V}_s , \quad s = 1, \ldots, J , \quad (2.107)$$

hold for some constant $\omega \in (0, 2)$. These inequalities ensure the contraction of each operator $I - \widetilde{P}_s$ on the subspace $\mathcal{V}_s$ and are the strengthened right inequalities (2.87). Let us note that if (2.107) holds for some $\omega \geq 2$, than we always can change the bilinear forms $b_s(\cdot, \cdot)$ by scaling such that these inequalities will hold for $\omega \in (0, 2)$. The SMSM produces a s.p.d. preconditioner $\boldsymbol{\mathcal{K}} = \mathbf{K}(\mathbf{I} - \mathbf{E})^{-1}$ that can be used in the PCG method for solving our

linear system of FE equations. The following theorem provides bounds for the minimal and maximal eigenvalues of the preconditioned stiffness matrix $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$.

Theorem 2.5. *Let us assume that the space splitting is stable in the sense of (2.91) and that the subspace contraction condition (2.107) holds for some constant $\omega \in (1,2)$. Then the spectral bounds*

$$\lambda_{min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) = \lambda_{min}(I-E) \geq \frac{2-\omega}{\omega^2 \varkappa^2 \alpha} \tag{2.108}$$

and

$$\lambda_{max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) = \lambda_{max}(I-E) \leq 1 \tag{2.109}$$

hold with $\varkappa = \rho(\boldsymbol{\Upsilon})$ or $\varkappa = J_O$.

Proof. We refer to [Dryja and Widlund (1995); Smith *et al.* (1996)], where the estimates

$$\lambda_{min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \geq (2-\omega)/[1+2\omega^2\rho^2(\boldsymbol{\Upsilon})]\alpha\,, \quad \lambda_{max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \leq 1\,.$$

can also be found for the case $\mathcal{V} = \sum_{s=0}^{J} \mathcal{V}_s$. □

In a more general case, the analysis can be based on the exact representation of the convergence rate $\|E\|_a$ of the MSM in the energy norm presented by [Xu and Zikatanov (2002)], where E again denotes the MSM error propagation (iteration) operator defined by (2.54).

Theorem 2.6. *Let us assume that the subspace contraction condition (2.107) holds for some constant $\omega \in (0,2)$. Then the energy norm of the MSM error propagation operator E and, equivalently, of the MSM iteration matrix $\mathbf{E}$ can be represented in the form*

$$\|E\|_a = \|\mathbf{E}\|_K = q := \frac{c}{1+c} < 1\,, \tag{2.110}$$

where

$$c = \sup_{\|v\|_a=1} \inf_{v=\sum_{j=1}^{J} v_j} \sum_{i=1}^{J} a\left(\widetilde{P}_i \overline{P}_i^{\,-1} \widetilde{P}_i w_i, w_i\right) < \infty \tag{2.111}$$

with $w_i = \sum_{j=i}^{J} v_j - \widetilde{P}_i^{-1} v_i$ and $\overline{P}_i = \widetilde{P}_i^{*} + \widetilde{P}_i - \widetilde{P}_i^{*}\widetilde{P}_i = 2\widetilde{P}_i - \widetilde{P}_i^{2}$ $(\widetilde{P}_i = \widetilde{P}_i^{*})$.

Proof. See [Xu and Zikatanov (2002)] who considered also the more general case of infinite dimensional Hilbert spaces and of non-symmetric but elliptic bilinear forms. □

In the case of the exact MSM, the representation can be simplified, because $\widetilde{P}_i = P_i = P_i^* = P_i^2$ is an orthoprojection with respect to the energy inner product $a(\cdot,\cdot)$. More precisely, the expression for the constant c in Theorem 2.6 can directly be rewritten in the form

$$c = \sup_{\|v\|_a=1} \inf_{v=\sum_{j=1}^{J} v_j} \sum_{i=1}^{J-1} \|P_i \sum_{j=i+1}^{J} v_j\|_a < \infty\,. \tag{2.112}$$

Moreover, Theorem 2.6 immediately yields the convergence rate estimate

$$\|E_{SMSM}\|_a = \|E_{MSM}^* E_{MSM}\|_a \le \|E_{MSM}\|_a^2 = q^2 < 1 \tag{2.113}$$

for SMSM, where E_{MSM} and E_{SMSM} denote the error propagation operators corresponding to the MSM and SMSM, respectively. Estimate (2.113) implies the spectral equivalence inequalities

$$(1-q^2)\,\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \boldsymbol{\mathcal{K}} \tag{2.114}$$

for the SMSM preconditioner $\boldsymbol{\mathcal{K}} = \mathbf{K}(\mathbf{I} - \mathbf{E}_{SMSM})^{-1}$.

The bounds for hybrid methods can be derived in a similar way. Having in mind the analogy between the Jacobi and Gauss-Seidel methods with the additive and multiplicative Schwarz methods, one may expect that the multiplicative Schwarz methods are faster than the hybrid and additive ones. For the case of the exact versions of these kind of methods, this was proved by [Mandel (1994)].

Chapter 3

Overlapping Domain Decomposition Methods

DD methods with nonoverlapping subdomains can be superior to overlapping DD methods with respect to the efficiency, in particular, in the case of elliptic equations with coefficients having large jumps and/or anisotropies. However, the latter are much simpler in the numerical realization and theoretical analysis. Instead of five basic components of the Dirichlet–Dirichlet type DD methods, *e.g.* in 3d, they may have only two, which are related to the coarse grid problem and the solution of the Dirichlet problems on the subdomains of domain decomposition. In particular, the problems of creation and computer realization of precondioner-solvers for the interface and the prolongations, which are not easy, are absent. It is also worth to keep in mind that the multilevel versions of the overlapping domain decomposition method have clear relation to multigrid and other multilevel methods. This chapter presents basic results pertaining the overlapping domain decomposition methods.

3.1 Construction Principles

In the preceding section we discussed the conditions which must be fulfilled for DD preconditioners in order to provide good convergence rates. However, the question remains how to obtain decompositions and algorithms with good properties. The construction of such methods requires the following canonical steps:

(1) splitting of FE space into subspaces $\mathcal{V} = \sum_{j=1}^{J} \mathcal{V}_j$, or $\mathcal{V} = \sum_{j=0}^{J} \mathcal{V}_j$ when including some coarse grid space $\mathcal{V}_0$,
(2) specifying the subspace bilinear forms $b_s(\cdot, \cdot)$,

(3) the arrangement of the projection-like operations, *e.g.*, additive, multiplicative, hybrid.

These steps mark only a half of the way to an efficient DD preconditioner-solver, because they are lacking fast solvers for the problems in the subspaces. On one hand, without such solvers one can hardly expect serious gains in efficiency of the whole DD solver. On the other hand, the demand for the fast subdomain/subspace solvers may require replacement of the subspace preconditioners B_s by some others, sometimes very intelligently designed. The latter especially concerns the DD algorithms for *hp* discretizations. Therefore, the next step is necessary, which is

(4) design of fast solvers for the problems in the subspaces.

We saw that even the three first canonical steps may be completed in different ways and produce a variety of DD algorithms. Apart from the differences caused by the arrangement of the projection-like operations, *i.e.*, additive, multiplicative and hybrid, another major distinction is caused by the ways of the domain decomposition, which are directly related to the space decomposition and fall into the two classes: the overlapping and nonoverlapping domain decompositions. Although they have much in common and may be treated from a unique point of view, these two classes have their own specific construction and algorithmic features. Overlapping DD methods, from which we start, have a longer history, and the general framework of the theory of Schwarz methods, presented in the preceding section, was initially developed for their analysis.

As we saw, the design of an efficient DD algorithm for the system (1.11) is equivalent to the design of an efficient DD preconditioner $\boldsymbol{\mathcal{K}}$ and, more exactly, preconditioner-solver. The following most important properties, characterizing the quality of the preconditioner-solver, are anticipated.

i) A good condition number of the matrix $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$ as close as possible to 1.
ii) Reduction of the computational cost in comparison, *e.g.*, general purpose direct solvers. This implies a low computational cost for solving systems with the matrix $\boldsymbol{\mathcal{K}}$, and, as a prerequisite, fast solvers for the problems in the subspaces.
iii) Robustness with respect to jumps of coefficients of elliptic equations.
iv) The opportunities of the parallelization of computations and the adjustment of the algorithm to computations in an adaptive mode.

Construction of the approximate bilinear forms $b_s(\cdot,\cdot)$, which are spectrally as close as to $a_s(\cdot,\cdot) = a(\cdot,\cdot)|_{\mathcal{V}_s}$ and convenient for implementation of fast

solvers for the problems in the subspaces, is an important and common problem for the DD algorithms.

3.2 Discretizations and Generalized Quasiuniformity Conditions

In this book, we are not concerned with the problems caused by the approximation of boundaries and boundary conditions in the finite element method. For this reason and for the sake of simplicity, if the opposite is not stated, we assume that the domain $\Omega \subset R^d$, $d = 2, 3$, coincides with the computational domain $\Omega_{\rm fem}$, *i.e.*

$$\overline{\Omega} = \overline{\Omega}_{\rm fem} = \bigcup_{r=1}^{\mathcal{R}} \overline{\tau}_r \,,$$

occupied by the assemblage of geometrically conform and in general curvilinear finite elements τ_r, the number of which is denoted by $\mathcal{R}$. The domains τ_r of the finite elements are specified by nondegenerate mappings $x = \mathcal{X}^{(r)}(y) : \bar{\tau}_0 \to \bar{\tau}_r$ with positive Jacobian's , where τ_0 is the reference square/cube $\tau_0 = (-1, 1)^d$, or the reference triangle/tetrahedron/simplex $\tau_0 = \{x : 0 < x_k, \sum_{i=1}^d x_i < 1,\ 1 \le k \le d\}$. For the components of $\mathcal{X}^{(r)}(y)$, we use the notation $\mathcal{X}_k^{(r)}(y)$. Sometimes, for defining simplicial reference elements, it is more convenient to use the unilateral triangle τ_0 with the vertices $(-1, 0), (1, 0), (0, \sqrt{3})$, the tetrahedron and simplexes with the equal edges. If some finite element $\mathcal{E}_r$ is defined on τ_r by means of a nondegenerate mapping $\mathcal{X}^{(r)}(y) : \bar{\tau}_0 \to \bar{\tau}_r$ with the use of some reference element $\mathcal{E}_0$ defined on the reference simplex τ_0, we call such elements, their domains, nodes etc. *associated*.

In what follows, results of analysis of DD algorithms are often formulated without mentioning any quasiuniformity conditions. In these cases it is always assumed that the *generalized conditions of shape regularity*, which in the known cases are equivalent to the generalized conditions of angular regularity, are fulfilled. However, not rarely the finite element assemblage must satisfy the *generalized conditions of quasiuniformity*, which are stronger and assume that not only angles and shapes, but also sizes of all elements are quasiuniform in the generalized sense. The word *generalized*, in most occasions omitted for brevity, is used for the reason that the conditions are applicable to the range of finite elements which includes

curvilinear elements. One convenient way to formulate these conditions is to represent each mapping as a superposition of two non-degenerate mappings $\mathcal{X}^{(r)}(y) = \widetilde{\mathcal{X}}^{(r)}(\mathcal{Z}^{(r)}(y))$, where $x = \widetilde{\mathcal{X}}^{(r)}(z) : \bar{\tau}'_r \to \bar{\tau}_r$ is a nonlinear mapping and $z = \mathcal{Z}^{(r)}(y) : \bar{\tau}_0 \to \bar{\tau}'_r$ is an affine or polylinear mapping of a τ_0-simplex or a square/cube, respectively. Here τ'_r is a simplex, if the reference element is the reference simplex, and is a guadrangle/octagon, if the reference element is square/cube. In particular, if τ_r is a curvilinear simplex, one can take for τ'_r the simplex with plain faces, having common vertices with τ_r. If τ_r is a curvilinear guadrangle/octagon, then τ'_r may be uniquely defined by the conditions that the vertices of τ_r and τ'_r coincide and $\mathcal{Z}^{(r)}(y)$ is a polylinear mapping.

The generalized conditions of the quasiuniformity require that, uniformly in $r = 1, 2, .., \mathcal{R}$ and the mesh parameter h,

α) polyhedrons τ'_r must be quasiuniform in the usual geometric sense, and

β) for the nonlinear mappings and their inverses, the modulus of Jacobians and components of Jacobi matrices must be bounded, *i.e.,*

$$\underline{c} \le \mathbb{J}[\widetilde{\mathcal{X}}^{(r)}(z)]\,, \qquad \left|\partial \widetilde{\mathcal{X}}^{(r)}_k / \partial z_l\right| \le \overline{c}\,, \qquad k, l = 1, 2, .., d\,, \quad \forall\ z \in \overline{\tau}'_r\,, \tag{3.1}$$

where $\underline{c}, \overline{c}$ are positive constants and $\mathbb{J}[\widetilde{\mathcal{X}}^{(r)}]$ is Jacobian, or the determinant of the Jacobi matrix $\mathbf{J}[\widetilde{\mathcal{X}}^{(r)}]$ of the mapping $\widetilde{\mathcal{X}}^{(r)}$.

For definiteness, it is assumed that Jacobians of mappings $\mathcal{X}^{(r)}, \widetilde{\mathcal{X}}^{(r)}$ and $\mathcal{Z}^{(r)}$ are always positive.

Now we turn to the polyhedrons τ'_r and discuss the geometric conditions of their quasiuniformity. Obviously, the conditions of the shape and angular quasiuniformity are equivalent only for the elements associated with the simplical reference elements. For instance, in a rectangular finite element all angles are good, but at the same time the aspect ratio of edges may be arbitrary, and, therefore, the shape regularity is not a consequence of the angular regularity .

The geometric conditions of qusiuniformity for polyhedrons τ'_r assume existence of positive constants $\alpha^{(1)}$ and θ such that for any h and r we have

$$\alpha^{(1)} h \le \overline{\rho}_r \le h\,, \qquad \theta \le \underline{\rho}_r / \overline{\rho}_r\,, \tag{3.2}$$

where $\underline{\rho}_r$ and $\overline{\rho}_r$ are, respectively, the radii of the largest inscribed and smallest circumscribed circles relative to τ'_r. These conditions are necessary and sufficient only for simplexes. The generalized conditions of shape or, *i.e.,* angular quasiuniformity imply that in α) only the shape or angular

quasiuniformity is required. Consequently, for simplexes only the inequality with θ in (3.2) must hold with the maximal size of elements bounded by h.

Let $h_{i,j}^{(r)}$ be the lengths of the edges of τ_r' and $\theta_{i,j}^{(r)}$ be the angle between the edge $\overline{x^{(i)}, x^{(j)}}$ joining the vertices $x^{(i)}, x^{(j)} \in \overline{\tau}_r$ and the plane, containing the rest edges converging at the vertex $x^{(i)}$. For simplexes, equivalent to (3.2) conditions can be written in a form:

$$\alpha^{(1)} h \le h_{i,j}^{(r)} \le h\,, \qquad \theta \le \theta_{i,j}^{(r)} \le \pi - \theta \tag{3.3}$$

with some positive constants $\alpha^{(1)}$ and θ. Shape and angular quasiuniformity assume that only inequalities for the angles hold, whereas first pair of inequalities (3.3) is the consequence of second ones with h replaced by $h^{(r)}$ and $h^{(r)}$ understood as the maximal among lengths of edges of the simplex τ_r. In this case, the notation h is as a rule used for $\max_r h^{(r)}$.

It is easy to see that the equations in (3.3) are also the necessary and sufficient conditions to be quasiuniform for an assemblage of finite elements, associated with a cubic reference element by polylinear mappings.

However, second pair of inequalities (3.3) is insufficient for the shape regularity of such a FE assemblage, and both pairs of the inequalities

$$\alpha^{(1)} h^{(r)} \le h_{i,j}^{(r)} \le h^{(r)}\,, \qquad \theta \le \theta_{i,j}^{(r)} \le \pi - 2\theta \tag{3.4}$$

are independent. In order words, additionally to the restriction from below and above for the angles of all elements, it is required that $h_{\min}^{(r)}/h_{\max}^{(r)} \ge \alpha^{(1)}$ uniformly for all elements, where $h_{\min}^{(r)}$ and $h_{\max}^{(r)} = h^{(r)}$ are the minimal and maximal lengths of edges of an element.

In the proofs of the estimates of approximation and convergence of FE solutions, stronger conditions must hold, depending first of all on the order of finite elements coordinate functions and the smoothness of the approximated solution. They can be termed the *generalized conditions of quasiuniformity of the order* κ, $\kappa \ge 1$, and imply that additionally to α, β) it is required that

β^+) the nonlinear mappings $x = \widetilde{\mathcal{X}}^{(r)}(z)$ satisfy the inequalities

$$|D_z^q \widetilde{\mathcal{X}}_l^{(r)}(z)| \le c\,, \qquad 1 < |q| \le \kappa\,, \qquad z \in \overline{\tau}_r'\,, \tag{3.5}$$

with the constant independent of r and h. Such conditions were used in early studies of approximation and convergence of FE methods with the curvilinear finite elements associated with triangular and tetrahedral reference p-th order elements, see, *e.g.*, [Ciarlet and Raviart (1972); and Korneev (1970)].

The conditions of the quasiuniformity, allowing to introduce the mesh parameter h for curvilinear finite elements, may be formulated directly

for the mappings $\mathcal{X}^{(r)}(y)$. The length change is characterized by Lamé coefficients

$$H_k^{(r)} = \left[\sum_{l=1}^{d} \left(\frac{\partial \mathcal{X}_l^{(r)}}{\partial y_k} \right)^2 \right]^{1/2} ,$$

where $\mathcal{X}_l^{(r)}$ are the components of the vector-function $\mathcal{X}^{(r)} = (\mathcal{X}_l^{(r)})_{l=1}^d$. Also, let $\mathbf{i}_k(y)$ be the unit vector in the space of variables x tangent to the line $y_l = \text{const}$, $l \neq k$ and directed towards increasing y_k. We introduce $\theta_k^{(r)}(y)$ as the angle between $\mathbf{i}_k(y)$ and the plane containing $\mathbf{i}_l(y)$, $l \neq k$. An equivalent formulation of the quasiuniformity conditions of the 1-st order is

$$\alpha^{(1)} h \leq H_k^{(r)} \leq h \,, \qquad \theta \leq \theta_k^{(r)} \leq \pi - \theta \,, \qquad 0 < \alpha^{(1)}, \theta = \text{const} \,, \tag{3.6}$$

whereas conditions of the order κ for simplical finite elements assume additionally the inequalities

$$|D_y^q \mathcal{X}_k^{(r)}(y)| \leq c h^{|q|} \,, \qquad 1 < |q| \leq \kappa \,, \qquad y \in \overline{\tau}_0 \,, \tag{3.7}$$

see, *e.g.*, [Korneev (1977a)]. In the case of cubic reference elements, and, *e.g.,* tensor product complete polynomial spaces $\mathcal{Q}_p$, conditions (3.7) and (3.5) for the mixed derivatives may be relaxed, indeed omitted, if to take into account the approximation theory developed in [Bramble and Hilbert (1971); Ciarlet and Raviart (1972)]; and [Ciarlet and Lions (1991)]. Obviously, for the same finite element assemblage, the values $\alpha^{(1)}, \theta, h, c$ in (3.3), (3.5) and (3.6), (3.7) or (3.2) will be different. However, this will not cause confusion, because almost everywhere we do not specify explicitly the form of dependence of constants in the bounds related to DD preconditioning on these values.

Proposition 3.1. *Let τ_0 be a sufficiently smooth domain, $\mathcal{X}(y) : \bar{\tau}_0 \to \bar{\tau}$ be a nondegenerate mapping $\mathcal{X}(y) \in C^1(\overline{\tau}_0)$ and H_k and θ_k be Lamé coefficients and the angles, introduced as above. Then, for any $v(x) \in C^1(\overline{\tau})$, we have*

$$\begin{aligned} &\left[\textstyle\sum_{k=1}^d H_k^2 \right]^{-1} \textstyle\sum_{k=1}^d (\frac{\partial v}{\partial y_k})^2 \leq \textstyle\sum_{k=1}^d (\frac{\partial v}{\partial x_k})^2 \leq \\ &\leq \left[\textstyle\sum_{k=1}^d \frac{1}{(H_k \sin\theta_k)^2} \right] \textstyle\sum_{k=1}^d (\frac{\partial v}{\partial y_k})^2 \,, \qquad \forall\, y \in \overline{\tau}_0 \,. \end{aligned} \tag{3.8}$$

Proof. The left bound is proved by means of Cauchy inequality and the definition of Lamé coefficients:

$$\sum_{k=1}^{d}(\tfrac{\partial v}{\partial y_k})^2 = \sum_{k=1}^{d}\left(\sum_{l=1}^{d}\tfrac{\partial v}{\partial x_l}\tfrac{\partial x_l}{\partial y_k}\right)^2 \le$$

$$\le \sum_{k=1}^{d}\left(\sum_{l=1}^{d}(\tfrac{\partial v}{\partial x_l})^2\right)\left(\sum_{l=1}^{d}(\tfrac{\partial x_l}{\partial y_k})^2\right) =$$

$$= \left(\sum_{k=1}^{d}H_k^2\right)\sum_{k=1}^{d}(\tfrac{\partial v}{\partial x_k})^2.$$

Suppose that $z=(z_1,z_2,..,z_d)$ is the system of the oblique-angled coordinates with the origin at the point $x=\mathcal{X}(\mathring{y})$ and the axes directed along the unit vectors $\{\mathbf{i}_k(\mathring{y})\}_{k=1}^d$. If we introduce the linear mapping

$$x_k = \mathcal{X}_{k,y}(z) := \mathring{y}_k + \sum_{l=1}^{d} H_l(\mathring{y})\,(\mathbf{i}_l(\mathring{y})\cdot\mathbf{e}_k)\,z_k\,, \tag{3.9}$$

then $\partial\mathcal{X}_{k,y}/\partial z_l = \partial\mathcal{X}_k/\partial y_l$ at $y=\mathring{y}$ from where we get the relations $\partial z_l/\partial x_k = \partial y_l/\partial x_k$ and $y=\mathring{y}$. Let $\{\mathbf{i}^k(y)\}_{k=1}^d$ be the unit vectors dual to $\{\mathbf{i}_k(y)\}_{k=1}^d$. Inverting (3.9), we obtain

$$z_k = \frac{1}{H_k(\mathbf{i}_k(\mathring{y})\cdot\mathbf{i}^k(\mathring{y}))}\sum_{l=1}^{d}(\mathbf{i}_k(\mathring{y})\cdot\mathbf{e}_l)\,(x_l-\mathring{y}_l)\,. \tag{3.10}$$

Now we see that

$$\sum_{k=1}^{d}\left(\frac{\partial v}{\partial x_k}\right)^2 = \sum_{k=1}^{d}\left(\sum_{l=1}^{d}\frac{\partial v}{\partial z_l}\frac{\partial z_l}{\partial x_k}\right)^2 \le \sum_{k=1}^{d}\left(\frac{\partial v}{\partial z_k}\right)^2\sum_{k=1}^{d}\sum_{l=1}^{d}\left(\frac{\partial z_k}{\partial x_l}\right)^2, \tag{3.11}$$

and according to (3.10), we have

$$\sum_{l=1}^{d}\left(\frac{\partial z_k}{\partial x_l}\right)^2 = \frac{1}{[H_k(\mathring{y})(\mathbf{i}_k(\mathring{y})\cdot\mathbf{i}^k(\mathring{y}))]^2} = \frac{1}{[H_k(\mathring{y})\sin\theta_k(\mathring{y})]^2}\,.$$

The right bound (3.8) follows from (3.11), the above equality and the equalities $\partial z_l/\partial x_k = \partial y_l/\partial x_k$ for $y=\mathring{y}$. □

The bounds

$$\max_l\left(\sin^{-1}\theta_l\prod_{k=1}^{d}H_k\sin\theta_k\right) \le |\mathbb{J}[\mathcal{X}(y)]| \le \prod_{k=1}^{d}H_k \tag{3.12}$$

for the modulus of the Jacobian $\mathbb{J}[\mathcal{X}(y)]$ of the mapping $\mathcal{X}(y)$, Proposition 3.1 and the generalized conditions of quasiuniformity allow us to write

$$\frac{1}{d}\,(\alpha^{(1)}\sin\theta)^{d-1}h^{d-2}|v(\mathcal{X}(y))|^2_{1,\tau_0} \le |v(x)|^2_{1,\tau} \le \frac{d}{\alpha^{(1)}\sin\theta}h^{d-2}|v(\mathcal{X}(y))|^2_{1,\tau_0}\,. \tag{3.13}$$

In the case of an affine mapping, these bounds can be written in the form

$$\frac{1}{d\max_k(H_k{}^2)}\,|v(\mathcal{X}(y))|^2_{1,\tau_0} \le \frac{|v(x)|^2_{1,\tau}}{|\mathbb{J}[\mathcal{X}(y)]|} \le \frac{d}{\min_k(H_k^2\sin^2\theta_k)}\,|v(\mathcal{X}(y))|^2_{1,\tau_0}\,, \tag{3.14}$$

and, if additionally τ_0 is a reference triangle, then

$$\frac{1}{3}(\sin\theta)\,|v(\mathcal{X}(y))|^2_{1,\tau_0} \le |v(x)|^2_{1,\tau} \le 3(\sin\theta)^{-1}\,|v(\mathcal{X}(y))|^2_{1,\tau_0}\,. \tag{3.15}$$

If one uses (3.3) instead of (3.6), bounds (3.12), (3.13) are easily transformed. For instance, (3.13) retains, but additional constants appear in the left and right inequalities, depending on $\underline{c}, \overline{c}$ from (3.1). If also (3.5), or (3.7), hold for $\kappa = 2$, then these additional constants are close to one up to the values of the order h. For complementary information on these matters, we refer to Korneev [Korneev (1970, 1977a, 1979a,b)]

3.3 Algorithms with Generous Overlap

In the overlapping DD method, similar to that studied by Schwarz, the splitting

$$\mathcal{V} = \sum_{j=1}^{J} \mathcal{V}_j\,, \qquad \mathcal{V} = \mathcal{V}_h(\Omega) \subset \mathring{H}^1(\Omega)\,, \tag{3.16}$$

of FE space corresponds to the representation of the computational domain Ω by the union

$$\Omega = \bigcup_{j=1}^{J} \Omega_j \tag{3.17}$$

of overlapping subdomains Ω_j. This implies that for each Ω_j there is at least one subdomain Ω_i such that $j \neq i$ and $\Omega_j \cap \Omega_j \neq \varnothing$. Then $\mathcal{V}_j = \mathcal{V} \cap \mathring{H}^1(\Omega_j)$, and, therefore, subspace problems are the local Dirichlet FE problems. There are various ways of constructing overlapping domain decompositions, depending on different factors. Often overlapping decompositions are created by means of a *coarse triangulation* $\mathcal{T}_H$ of the computational domain, for the nests of which we use the notation $\tau_{H,k}\,,\ k = 1, 2, .., \mathcal{R}_H$. In applied software packages, the coarse triangulation $\mathcal{T}_H$, when possible, is created first and then is used for a cheap generation of a fine triangulation $\mathcal{T}_h$ by triangulating each subdomain $\tau_{H,k}$. These additional trangulations of the

nests $\tau_{H,k}$ of the coarse mesh are denoted $\mathcal{T}_{h,k}$ and can be constructed in such a way that the fine triangulation is the union of these triangulations,

$$\mathcal{T}_h = \bigcup_{k=1}^{\mathcal{R}_H} \mathcal{T}_{h,k}\,, \tag{3.18}$$

and is conform.

Two simple techniques may be used for producing overlapping decompositions. Let us assume for simplicity that the subdomains τ_r and $\tau_{H,k}$ of the fine and coarse triangulations, respectively, are simplexes, *i.e.,* triangles at $d = 2$ and tetrahedrons at $d = 3$, the both triangulations are quasiuniform and h and H are their mesh parameters entering (3.2) or (3.6). If $y_H^{(j)}$ denote vertices of the coarse triangulation, then in one technique each $\overline{\Omega}_j$ is defined as the star of the simplexes $\overline{\tau}_{H,k}$, having $y_H^{(j)}$ for the common vertex, *i.e.,*

$$\overline{\Omega}_j = \cup_{k\in B_{H,j}} \overline{\tau}_{H,k}\,, \quad \text{with } B_{H,j} = \{k : y_H^{(j)} \in \overline{\tau}_{H,k}\}\,.$$

Therefore, the number of subdomains coincides with the number of the vertices of the coarse triangulation. For this technique, we use the abbreviation ODD-1.

Alternatively, each $\overline{\Omega}_j$ may be defined as the union of one simplex $\overline{\tau}_{H,k_j}$ and all the simplexes $\overline{\tau}_{H,l}$, having at least one common vertex with $\overline{\tau}_{H,k_j}$. This technique of decomposition, denoted ODD-2, provides a more generous overlap, and the number of the subdomains is equal to the number of simplexes of the coarse triangulation. Several generalizations of these procedures are feasible. In particular, the overlapping decomposition ODD-2 is a special case of the following more general one. First, one builds a nonoverlapping domain decomposition with the tentative subdomains $\widetilde{\Omega}_j$, consisting of one or several coarse grid elements, and then extends the subdomains by adding some layers of fine grid elements around their boundaries.

Obviously, in a strict sense, the decomposition ODD-1 and ODD-2 can be realized only when the computational domain coincides with the domains of the coarse and fine triangulations, *i.e.,*

$$\Omega = \Omega_H = \Omega_h\,,$$

with

$$\overline{\Omega}_H = \cup_k \overline{\tau}_{H,k}\,, \quad \text{and} \quad \overline{\Omega}_h = \cup_r \overline{\tau}_r = \cup_r \overline{\tau}_{h,r}\,.$$

However, if the coarse triangulation does not provide sufficiently accurate approximation of a curved boundary, there are remedies to effectively employ the same ideas. For instance, the domain Ω may be first represented

(or approximated) by a union of curvilinear polygons which are smoothly mapped on the compatible polygons with the plain faces. These mappings should be compatible, *i.e.,* define the piecewise smooth mapping of the class $L^1_\infty(\Omega)$ of Ω on some polygonal domain Ω_P. Then subsidiary coarse and fine triangulations may be produced for Ω_P and mapped back on Ω, or alteratively one can solve the transformed elliptic problem on Ω_P. It is also worth noting that the field of computer generation of nonoverlapping and overlapping domain decompositions by clustering and other techniques is now rather well developed not only in relation to DD solvers, but a number of other numerical methods. For instance, overlapping domain decompositions are used in algebraic two– and multilevel methods, based on aggregation (agglomeration), for solving FE systems, see *e.g.,* [Korneev and Fish (2000); Lallemand *et al.* (1992); Lonsdale (1993)] and the references there.

For different types of overlapping decompositions, the measure δ of the overlap may be introduced differently. Suppose that each subdomain Ω_j is not completely covered by other subdomains and $\overline{\Omega}'_j = \overline{\Omega}_j \setminus \cup_{i\neq j} (\Omega_i \cap \Omega_j)$ contains at least one point. Then, for Ω_j, the measure δ_j of the overlap may be introduced as the minimal thickness of the overlap, *i.e.,* $\delta_j = \operatorname{dist}(\partial\Omega_j, \overline{\Omega}'_j)$. For a given domain decomposition, the measure of the overlap is $\delta = \min \delta_i$. If each subdomain is completely covered by others, we can adopt $\delta = \mathcal{O}(H)$. Evidently, that for the two described above techniques ODD-1 and ODD-2 of creating the overlapping decompositions, one has $\delta = \mathcal{O}(H)$, which is usually referred as a generous overlap

Let $\boldsymbol{\mathcal{K}}$ be the (exact) additive Schwarz preconditioner

$$\boldsymbol{\mathcal{K}}^{-1} = \sum_{j=1}^{J} \mathbf{T}_j (\mathbf{T}_j^T \mathbf{K} \mathbf{T}_j)^{-1} \mathbf{T}_j^T = \sum_{j=1}^{J} \mathbf{T}_j \mathbf{K}_j^{-1} \mathbf{T}_j^T , \tag{3.19}$$

corresponding to the space splitting

$$\mathcal{V} = \sum_{j=1}^{J} \mathcal{V}_j , \qquad \mathcal{V}_j = \mathcal{V} \cap \mathring{H}^1(\Omega_j) = \operatorname{span} \Phi \mathbf{T}_j , \tag{3.20}$$

which is based on one of the overlapping domain decompositions $\Omega = \bigcup_{j=1}^{J} \Omega_j$ with $\delta = O(H)$. The $N \times N_j$ matrix $\mathbf{T}_j$ picks exactly those basis functions from the fine grid nodal basis Φ which belong to the inner nodes of Ω_j. The s.p.d. $N_j \times N_j$ matrix $\mathbf{K}_j$ is nothing but the stiffness matrix of the local FE Dirichlet problem in Ω_j. Thanks to the generous $O(H)$ overlap, the relative spectral condition number does not depend on

the fine grid discretization parameter h, but depends on the DD parameter H. Using the general theory of the Schwarz method, presented in the preceding chapter, it can be proved that

$$\operatorname{cond}[\mathcal{K}^{-1}\mathbf{K}] = \mathcal{O}(H^{-2})\,. \tag{3.21}$$

The bad dependence on H is due to the absence of some coarse grid solver managing the global information transport that is essential for elliptic problems. For the overlapping domain decompositions ODD-1 and ODD-2, a natural way certainly consists in adding the coarse grid FE space $\mathcal{V}_0 := \mathcal{V}_H = \operatorname{span}\Phi\mathbf{T}_0 \subset \mathcal{V} \subset H_0^1(\Omega)$ to the splitting (3.20), *i.e.,*

$$\mathcal{V} = \mathcal{V}_0 + \sum_{j=1}^{J} \mathcal{V}_j\,. \tag{3.22}$$

For the case under consideration, $\mathcal{V}_0$ is the space of continuous piecewise linear functions on the triangulation $\mathcal{T}_H$. Now, the corresponding *two level* (coarse level and fine level) overlapping ASM preconditioner

$$\mathcal{K}^{-1} = \sum_{j=0}^{J} \mathbf{T}_j \mathbf{K}_j^{-1} \mathbf{T}_j^T \tag{3.23}$$

yields an optimal relative spectral condition number estimate.

Theorem 3.1. *The exact two-level ASM preconditioner (3.23) based on an overlapping domain decomposition with a uniform overlap $\delta = \mathcal{O}(H)$ is optimal in condition, so that* $\operatorname{cond}[\mathcal{K}^{-1}\mathbf{K}] \le c$ *with some positive constant c not depending on h, H, and J.*

Proof. In the next section, we prove a more general result. □

Clearly, the quality of the overlapping DD preconditioner is greatly influenced by solvers which can be used for local discrete Dirichlet problems. Taking into account that in general the coefficients of elliptic problem will change inside subdomains Ω_j and shapes of subdomains may be far from canonical, we can hardly hope that efficient direct solvers exist for all matrices $\mathbf{K}_j$. Therefore, as a rule iterative solvers will be used for systems governed by these matrices, and in this case the most efficient strategy assumes the use of iterative solvers as inexact ones. Obviously, the theorem remains true for the inexact version where the local stiffness matrices $\mathbf{K}_j$ in the DD preconditioner are replaced by suitable spectrally equivalent preconditioners $\mathcal{K}_j$. In other words,

$$\mathcal{K}^{-1} = \sum_{j=0}^{J} \mathbf{T}_j \mathcal{K}_j^{-1} \mathbf{T}_j^T\,, \tag{3.24}$$

where the local preconditioners $\boldsymbol{\mathcal{K}}_j$ satisfy the spectral equivalence inequalities

$$\underline{\gamma}_j \boldsymbol{\mathcal{K}}_j \le \mathbf{K}_j \le \overline{\gamma}_j \boldsymbol{\mathcal{K}}_j \tag{3.25}$$

with positive constants $\underline{\gamma}_j$ and $\overline{\gamma}_j$. In the simplest case we can choose $\boldsymbol{\mathcal{K}}_j = \mathbf{B}_j$ as in (2.18), where the matrices $\mathbf{B}_j$ are induced by subsidiary bilinear forms $b_j(\cdot,\cdot)$, or according to the discussion in Subsection 2.2.3 as matrices $\boldsymbol{\mathcal{K}}_j = \mathbf{B}_{j,\mathrm{it}}$, implicitly defined by some inexact iterative solvers, see, *e.g.,* (2.48). Also, as we noted in Subsection 2.2.1 and as it will be seen from what follows, there are much more involved ways of deriving good local preconditioners.

Generation of initial nonoverlapping domain decompositions, used for creating overlapping domain decompositions, can be performed by means of the graph partitioning and graph coloring techniques applied to the graph representations of the finite element systems of equations. A few such as well as other techniques for creating overlapping decompositions can be found in [Brelaz (1979); and Hendrickson and Leland (1995)]. In p and hp methods, p-finite elements are most often considered as subdomains of nonoverlapping decompositions. Overlapping DD solvers for spectral discretizations were studied by [Pothen *et al.* (1990)] and [Pavarino (1994)].

Two-level ASM preconditioners with a generous overlap have a drawback. The larger the overlap is, the more difficult parallel implementation of the DD solver becomes due to the more intensive exchange data between processors. In general, there is no easily handled data structure, which could provide simple communications between global vectors and local vectors, related to subdomains of decomposition. Robustness with respect to coefficients of the elliptic equation having jumps is also compromised. Piecewise constant coefficients of the elliptic problem create an obstacle in producing such an overlapping domain decomposition that inside each subdomain Ω_j coefficients will not have jumps. In turn, it becomes more difficult to create fast subdomain solvers, unless a secondary nonoverlapping decompositions are carried out for subdomains. The problem of the same type becomes even more unpleasant in engineering applications to structures with complicated geometries, composed of substructures of much more simple geometries, often of different dimensions and essentially different materials. In this case, each unit of an overlapping decomposition may include parts of different substructures, which makes such decompositions practically unacceptable. Besides, at a considerable overlap, the computational overhead may be not negligible.

3.4 Loss in Convergence Due to Small Overlap

If we assume $H = \mathcal{V}$, $H_s = \mathcal{V}_s$, $\boldsymbol{\mathcal{K}}_s = \mathbf{B}_s$, $(\mathbf{B}_s\mathbf{v}_s, \mathbf{v}_s) = (B_s v_s, v_s) = b(v_s, v_s)$, then, according to Subsection 2.3.4, the analysis of the efficiency of DD preconditioners with respect to the relative condition number first of all requires the verification of the stability condition (2.85) for a chosen space splitting. It also requires the calculation of the values of the local spectral equivalence constants $\underline{\gamma}_s$ and $\overline{\gamma}_s$ in the local spectral equivalence inequalities (2.87) and the subspace interaction measure $\rho(\boldsymbol{\Upsilon}_b)$, figuring in Theorem 2.3. The most involved one is the verification of the stability condition of the domain decomposition , through which the small overlap influences the bound of the lower relative eigenvalue from below. At the same time, the subspace interaction measure can usually be bounded with the use of Cauchy's inequality. Let us remind that $\delta_s = \mathrm{dist}(\partial\Omega_s, \Omega_s')$ is defined for $s \neq 0$, where Ω_s' is the internal part of Ω_s, which is not covered by any of the other subdomains. It is here assumed that Ω_s' contains at least one point.

Theorem 3.2. *The two-level ASM preconditioner (3.24) based on an overlapping domain decomposition satisfies the estimates*

$$\lambda_{\min}^{-1}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \leq c_1 \frac{J_O}{\underline{\gamma}_b}\left(1 + c_2 \max_{s>1} \frac{H_s}{\delta_s}\right) \quad \textit{and} \quad \lambda_{\max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \leq \overline{\gamma}_b J_O \tag{3.26}$$

of the minimal and maximal eigenvalues of the preconditioned matrix $\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}$. *Therefore, we have*

$$\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] := \frac{\lambda_{\max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K})}{\lambda_{\min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K})} \leq c_1\, J_O^2\, \frac{\overline{\gamma}_b}{\underline{\gamma}_b}\left(1 + c_2 \max_{s>1} \frac{H_s}{\delta_s}\right), \tag{3.27}$$

where $\underline{\gamma}_b = \min_{s\geq 0}(\underline{\gamma}_s)$, $\overline{\gamma}_b = \max_{s\geq 0}(\overline{\gamma}_s)$, *and the positive constants* c_k *do not depend on* h_s, H_s, J *and* δ_s.

Proof. The bound for $\lambda_{\max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K})$ is the bound (2.96) for the case when the coarse space $\mathcal{V}_0$ is included in the space decomposition. Looking at formula (2.104) we conclude that a good bound for $\lambda_{\min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K})$ may be obtained by means of a low energy decomposition $v = \sum_{s=0}^{J} v_s$, see (2.102) and (2.104). The way of definition of such a decomposition is based on some subsidiary results and requires new notations, from descriptions of which we now start.

Let $Q_0 u$ be the L_2-projection of $u \in \mathring{H}^1(\Omega)$ in $\mathcal{V}_0$, *i.e.,*

$$(Q_0 u, v) = (u, v)\,, \qquad \forall\ v \in \mathcal{V}_0\,. \tag{3.28}$$

As was established by [Bramble and Xu (1991)], this projection is stable simplexwise in $H^1(\Omega)$ and approximates any function $u \in H^1(\Omega)$ in $L_2(\Omega)$. These properties are expressed by the inequality

$$H^{-2}\|u - Q_0 u\|^2_{0,\tau_H} + |Q_0 u|^2_{1,\tau_H} \le c\,|u|^2_{1,\tau_H}\,, \tag{3.29}$$

that holds for each simplex τ_H of the coarse triangulation, where H is the diameter of τ_H. We will need additionally the operator of the piecewise linear quasi-interpolation $I_h(v) : H^1(\Omega) \to \mathcal{V}(\Omega)$ and functions $\theta_s \in \mathcal{V}$ of the partition of unity. We describe these objects below. The main requirements for the interpolation operator, which, for definiteness, will be introduced for the fine mesh, is stability in $H^1(\Omega)$ and approximation. In the literature, there are known several operators satisfying these requirements and, *e.g.,* those considered by [Clément (1975); Scott and Zhang (1990); Verfürth (1996)]. For the proof of the above theorem, we use the quasi-interpolation operator of [Scott and Zhang (1990)]. Properties of this operator, approved by the belonging to these authors Lemma 5.2, are formulated as follows:

a) $I_h(v) : H^1(\Omega) \to \mathcal{V}(\Omega)$, and if $v \in \mathcal{V}(\Omega)$, then $I_h(v) = v$,
b) $(v - I_h(v)) \in \overset{\circ}{H}{}^1(\Omega)$, if $v \in H^1(\Omega)$ and $v|_{\partial\Omega} \in \mathcal{V}_{\rm tr}(\partial\Omega)$, where $\mathcal{V}_{\rm tr}(\partial\Omega)$ is the space of traces on $\partial\Omega$ of FE fuctions from $\mathcal{V}(\Omega)$,
c) $\|v - I_h(v)\|_{t,\Omega} \prec h^{s-t}\|v\|_{s,\Omega}$, for $t = 0, 1,\ s = 1, 2$,
d) $|I_h(v)|_{1,\Omega} \prec |\,v\,|_{1,\Omega}$ and $\|I_h(v)\|_{1,\Omega} \prec \|\,v\,\|_{1,\Omega}$ for all $v \in H^1(\Omega)$.

Functions θ_s of the partition of unity must satisfy equality $\sum_{s=1}^{J} \theta_s \equiv 1$ for all $x \in \overline{\Omega}$. For our purpose, it suffices to chose nonnegative functions $\theta_s \in C(\overline{\Omega}) \cap H^1(\Omega)$ in such a way that each $\theta_s = 1$ in the interior part $\Omega'_s \subset \Omega_s$ of the domain Ω_s, not covered by other subdomains. Besides, we require that it satisfies the inequality

$$|\nabla\theta_s| \le c_\sqcap/\delta_s\,, \qquad \forall\ x \in \Omega_{\delta,s} := \Omega_s \setminus \overline{\Omega}'_s\,, \tag{3.30}$$

in the rest part of Ω_s. Obviously, according to this definition, $\theta_s = 0$ outside of Ω_s. Note also that due to the introduced overlapping, construction of such a partition of unity is always possible under assumptions of quasi-uniformity of the fine and coarse triangulations. Now, for each $u \in \mathcal{V}$ one can set $u_0 = Q_0 u,\ w = u - u_0,$ and $u_s = I_h(\theta_s w),\ s = 1, 2, .., J$. As a consequence of this , we have

$$u = u_0 + \sum_{s=1}^{J} u_s\,.$$

First we estimate $a(u_s, u_s)$ by $a(w, w)$. Clearly, in view of a) and the definition of θ_s,

$$a_{\Omega'_s}(u_s, u_s) = a_{\Omega'_s}(w, w)\,, \tag{3.31}$$

where $a_{\Omega'_s}(\cdot,\cdot)$ is the restriction of $a(\cdot,\cdot) = a_\Omega(\cdot,\cdot)$ to Ω'_s. If $\tau \in \Omega_{s,\delta}$ and $\overline{\theta}_{s,\tau}$ is the average of θ_s over the element τ, then using the definition of u_s, property *d)* of the interpolation operator I_h, inequalities $\theta_s \le 1$ and (3.30), we find that

$$\begin{aligned} a_\tau(u_s, u_s) &= a_\tau(I_h(\theta_s w), I_h(\theta_s w)) \le a_\tau(\theta_s w, \theta_s w) \\ &\le c_{\rm a}\left(a_\tau(w, w) + c_\sqcap\, \delta_s^{-2}\|w\|^2_{0,\tau}\right)\,, \end{aligned} \tag{3.32}$$

with an absolute constant $c_{\rm a}$.

Summation over all $\tau_r \in \Omega_{\delta,s}$ and (3.31) allow us to write

$$a_{\Omega_s}(u_s, u_s) \le c_{\rm a}\left(a_{\Omega_s}(w, w) + c_\sqcap\, \delta_s^{-2}\|w\|^2_{0,\Omega_{\delta,s}}\right)\,, \tag{3.33}$$

and $\|w\|^2_{0,\Omega_{\delta,s}}$ can be estimated from above by means of the following inequality

$$\|v\|_{0,\Omega_{\delta,s}} \le c\,\delta_s^2\left((1+\frac{H_s}{\delta_s})|v|^2_{1,\Omega_s} + \frac{1}{H_s\delta_s}\|v\|^2_{0,\Omega_{\delta,s}}\right)\,,\ v \in H^1(\Omega_s)\,, \tag{3.34}$$

proved in [Dryja and Widlund (1994)]. Now, (3.34) allows us to rewrite (3.33) in the form

$$|u_s|^2_{1,\Omega_s} \le c\left((1+\frac{H_s}{\delta_s})|w|^2_{1,\Omega_s} + \frac{1}{H_s\delta_s}\|w\|^2_{0,\Omega_{\delta,s}}\right)\,, \tag{3.35}$$

whereas application of (3.29) yields

$$|u_s|^2_{1,\Omega_s} \le c\left((1+\frac{H_s}{\delta_s})|u|^2_{1,\Omega_s} + \frac{H_s}{\delta_s}|u|^2_{1,\Omega_{\delta,s}}\right) \le c\left(1+\frac{H_s}{\delta_s}\right)|u|^2_{1,\Omega_s}\,. \tag{3.36}$$

Since each subdomain overlaps with not more than J_O other subdomains and, as follows from (3.29),

$$|u_0|^2_{1,\Omega} \le |u|^2_{1,\Omega}, \tag{3.37}$$

we have

$$\sum_{s=0}^{J}|u_s|^2_{1,\Omega_s} \le c(1+J_O)\left(1+\max_s\frac{H_s}{\delta_s}\right)|u|^2_{1,\Omega}\,. \tag{3.38}$$

Taking additionally into account (3.25), we immediately get the estimate

$$\sum_{s=0}^{J}\mathbf{u}_s^\top\boldsymbol{\mathcal{K}}_s\mathbf{u}_s \le c(1+J_O)\frac{1}{\underline{\gamma}_b}\left(1+\max_s\frac{H_s}{\delta_s}\right)\mathbf{u}^\top\mathbf{K}\mathbf{u}\,, \tag{3.39}$$

where $\mathbf{u}$ and $\mathbf{u}_s$ are the vector representations of u and u_s, respectively. This bound is another form of the first bound (3.26). □

The low energy decomposition, used above for the estimation of the minimal relative eigenvalue of the DD preconditioner, was produced by [Dryja and Widlund (1994)], where they gave the proof of the theorem, see additionally Widlund [Widlund (1992)] and Le Tallec [Tallec (1994)]. The case of more general coarse meshes was studied by [Chan *et al.* (1996)].

For the case of two subdomains, in the same paper [Dryja and Widlund (1994)] showed that the result of Theorem 3.2 is sharp with respect to the dependence of the relative condition number on H/δ. If the overlap is minimal, *i.e.,* $\delta = h$, Brenner [Brenner (2000)] approved the sharpness of the estimate (3.27) for decompositions with many subdomains. For quasiuniform fine and coarse triangulations, this implies that there exists a positive constant c independent of h, H, and J for which $\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \geq c(H/h)$. In the same paper, overlapping decompositions and ASM preconditioners for finite element solvers of fourth-order elliptic boundary value problems were studied. Under the condition of the minimal overlap, it was proved that $\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \mathcal{O}((H/h)^3)$. Therefore, on the one hand, a small overlap really damages the quality of the two-level ASM preconditioner (3.23). On the other hand, the $\mathcal{O}(H)$ overlap means that additional $\mathcal{O}((H/h)^d)$ unknowns are added to the local problems in contrast to $\mathcal{O}((H/h)^{d-1})$ unknowns in the case of an $\mathcal{O}(h)$ overlap. But there exists an additional alternative. [Bank *et al.* (2002)] proposed a two-level hierarchical overlapping ASM preconditioner that adds only $\mathcal{O}((H/h)^{d-1})$ unknowns to the local problems as in the case of the minimal overlap, but has a relative condition number that is uniformly bounded as in Theorem 3.1.

3.5 Multilevel Versions

The two-level Schwarz methods, described above, use a fine (h) and a coarse (H) meshes, which are aimed to capture the local high frequency and the global low frequency parts of the solution, respectively, and of the iteration error. This approach can be efficient, if efficient local and global preconditioner-solvers are available. For large problems, one has two alternatives: to choose relatively small H and have many simple local problems and a difficult global problem, or to choose relatively big H and have a small global problem while local problems on subdomains of decomposition become more difficult. However, there is a third opportunity which can be much more efficient: one chooses relatively small H, but solves the corresponding global problem by applying again a two-level

algorithm on some coarser grid. A recursive application of the coarsening and corresponding two-level algorithms results in a special multilevel ASM preconditioner. To be more precise, we assume that the coarse triangulation $\mathcal{T}_0 = \mathcal{T}_H$ is refined L times giving the finer and finer imbedded triangulations $\mathcal{T}_1, \ldots, \mathcal{T}_{L-1}$, and $\mathcal{T}_L = \mathcal{T}_h$. For each level $l = 1, 2, \ldots, L$, with exception of the coarsest level $l = 0$, we construct some overlapping domain decomposition $\Omega = \bigcup_{j=1}^{J_l} \Omega_{l,j}$ and relate to such multilevel overlapping domain decomposition the multilevel splitting of the FE space

$$\mathcal{V} = \mathcal{V}_0 + \sum_{l=1}^{L} \sum_{j=1}^{J_l} \mathcal{V}_{l,j}. \tag{3.40}$$

The subspaces $\mathcal{V}_{l,j} = \operatorname{span} \Phi \, \mathbf{T}_{l,j}$ can again be generated by using the $N \times N_{l,j}$ basis transformation matrices $\mathbf{T}_{l,j}$. Now the corresponding (exact) multilevel overlapping ASM preconditioner can be written in the form

$$\boldsymbol{\mathcal{K}}^{-1} = \mathbf{T}_0 \mathbf{K}_0^{-1} \mathbf{T}_0^T + \sum_{l=1}^{L} \sum_{j=1}^{J_l} \mathbf{T}_{l,j} \mathbf{K}_{l,j}^{-1} \mathbf{T}_{l,j}^T \,. \tag{3.41}$$

An extreme case is when the subdomains $\Omega_{l,j}$ are simply the supports of the nodal basis functions $\phi_{l,j}$, with each function related to the node $x_{l,j}$ of the l-level triangulation $\mathcal{T}_l$. Now, the subspaces $\boldsymbol{\mathcal{V}}_{l,j} = \operatorname{span} [\phi_{l,j}] = \operatorname{span} [\Phi \mathbf{T}_{l,j}]$ are one-dimensional, where $\Phi = \Phi_h = \Phi_L = \{\phi_{L,j}\}$ denotes the fine grid basis and the $N \times 1$ matrix $\mathbf{T}_{l,j}$ provides the representation of the basis function $\phi_{l,j}$ in the fine grid basis. The overlapping DD, just described, resembles ODD1, but now the elements around the node $x_{l,j}$ are taken from the triangulation $\mathcal{T}_l$ and not $\mathcal{T}_{l-1}$. In the case of quasiuniform triangulations, this preconditioner may be greatly simplified. The simplification is based on the fact that, for second-order elliptic problems, the 1×1 matrices $\mathbf{K}_{l,j}$ obviously behaves like h_l^{d-2}. Therefore, the multilevel overlapping ASM preconditioner (3.41) yield the preconditioner

$$\boldsymbol{\mathcal{K}}^{-1} = \mathbf{T}_0 \mathbf{K}_0^{-1} \mathbf{T}_0^T + \sum_{l=1}^{L-1} \sum_{j=1}^{J_l} h_l^{2d} \mathbf{T}_{l,j} \mathbf{T}_{l,j}^T + h_L^{2d} \mathbf{I} \,. \tag{3.42}$$

It can be shown, that the multiplication $\mathbf{z} := \boldsymbol{\mathcal{K}}^{-1} \mathbf{v}$ may be completed for $\mathcal{O}(N)$ arithmetic operations, if it is arranged as outlined in Algorithm 3.1. In this algorithm, the notation $\mathbf{T}_{l,l+1,j}$ stands for the matrix providing the representation of a l-level basis function $\phi_{l,j}$ in the basis of the level $l+1$.

This remarkable idea of designing preconditioners in the way presented above is due to [Bramble *et al.* (1990)], and the preconditioner (3.42) is commonly referred as BPX (Bramble-Pasciak-Xu) preconditioner, while (3.41)

Algorithm 3.1 Procedure $\mathbf{z} := \boldsymbol{\mathcal{K}}^{-1}\mathbf{v}$.

for $l = L-1, L-2, .., 1, 0$ **do** $\mathbf{v}_l := \mathbf{T}^T_{l,l+1,j}\mathbf{v}_{l+1}$, **end for**

$\mathbf{u}_0 := \mathbf{K}_0^{-1}\mathbf{v}_0$,

for $l = 1, 2, .., L-1$ **do** $\mathbf{u}_l := h_l^{2d}\mathbf{v}_l$, **end for**

$\mathbf{u}_L := h_L^{2d}\mathbf{v}_L = h^{2d}\mathbf{v}$,

for $l = 1, 2, .., L-1$ **do** $\mathbf{w}_{l+1} := \mathbf{T}_{l,l+1,j}(\mathbf{w}_l + \mathbf{u}_l)$, **end for**

$\mathbf{z} := \mathbf{w}_L + \mathbf{u}_L$.

is termed Multilevel Diagonal Scaling (MDS) preconditioner and was introduced by [Zhang (1992)].

Theorem 3.3. *Let imbedded triangulations of the sequence be quasiuniform and $h_k/h_{k-1} \le \sigma < 1$. Then the BPX and MDS preconditioners are optimal with respect to the relative condition numbers and arithmetical costs. In other words, for some positive constant c, not depending on h and L, we have* cond $[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \le c$, *and the arithmetical cost* ops$(\boldsymbol{\mathcal{K}}^{-1}\mathbf{d})$ *of the preconditioning operation is proportional to the number of unknowns $N = \mathcal{O}(h^{-d})$ on the finest grid.*

Proof. The original proof of Bramble, Pasciak and Xu [Bramble *et al.* (1990)] provided weaker non-optimal bounds, which depended on the number of refinement levels $L = \mathcal{O}(\log(1 + (H/h)))$, see also Zhang [Zhang (1992)] for the MDS preconditioner. The optimality of the BPX preconditioner in the relative condition was established by Oswald [Oswald (1992)], who employed Besov's space techniques based on the characterization of Besov's spaces by means of finite element spaces, see also Oswald [Oswald (1994)]. □

It is a straight forward way to multiplicative and hybrid versions of these multilevel Schwarz methods, which turn out to be closely related to traditional multigrid methods. A discussion of these relations can be found in [Bramble and Zhang (2000)]. The BPX type preconditioners turned out to be very efficient in the practical solution of engineering problems. Additional features of them, which are very attractive for practical computations, are the fact that they are easily parallelizable, robust, may be adjusted to adaptive computations and allow us to use not imbedded sequences of triangulations. Below, we give a result illustrating only one of

these features, namely, robustness with respect to jumps in the coefficients of the elliptic problem. Following [Oswald (1999b)], we relate a triangulation to the *edge type*, if each of its tetrahedrons has at least one edge on $\partial\Omega$.

Theorem 3.4. *Let $d = 2, 3$ and let the triangulation $\mathcal{T}_0$ be of the edge type for the case $d = 3$. Let also $\boldsymbol{\mathcal{K}}$ be the BPX preconditioner for the Model Problem 2.2 with $\varrho(x) = \varrho_j = \text{const}$ for $x \in \tau_{0,j}$, where $\tau_{0,j}$, $j = 1, 2, .., J_0$, are the coarse grid finite elements. Then*

$$\text{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \leq c\, L^2,$$

where the constant c is independent of L as well as ϱ_j.

Proof. For the case $d = 2$, this estimate is due to [Bramble and Xu (1991)]. The $3d$ case was studied by [Dryja *et al.* (1996)] and [Oswald (1999b)]. The estimates

$$\underline{c} \leq \lambda_{\max}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}) \leq \overline{c} \tag{3.43}$$

hold without the restriction on the coarse triangulation to be of the edge type. For functions from the space $\mathcal{V}$, let us introduce the notation Q_l for the L_2-orthogonal projection operator into $\mathcal{V}_l$, the norm

$$\|v\|^2_{(0,\varrho)} = \sum_{j=1}^{J_0} \varrho_j \|v\|^2_{0,\tau_{0,j}}\,,$$

and the norms

$$|||\ v\ |||^2_{\varrho} = \inf_{v_0\in\mathcal{V}_0,\, v_{l,j}\in\mathcal{V}_{l,j}:\,(S)} \left[a_\Omega(v_0,v_0) + \textstyle\sum_{l=1}^{L}\sum_{j}^{J_l} a_\Omega(v_{l,j},v_{l,j})\right]\,,$$

$$|||\ v\ |||^2_{0,\varrho} = \inf_{v_l\in\mathcal{V}_l:\, v=\sum_{l=0}^{L} v_l} \left[a_\Omega(v_0,v_0) + \textstyle\sum_{l=1}^{L} 2^{2l}\|v_l\|^2_{(0,\varrho)}\right]\,,$$

$$|||\ v\ |||^2_{*} = \inf_{v_l\in\mathcal{V}_l:\, v=\sum_{l=0}^{L} v_l} \textstyle\sum_{l=1}^{L} 2^{2l}\|v_l\|^2_{0,\Omega}\,,$$

$$|||\ v\ |||^2_{E} = \|v\|_{0,\Omega} + \textstyle\sum_{l=1}^{L-1} 2^{2l} E_l(v)^2_{0,\Omega}\,, \quad E_l(v)^2_{0,\Omega} = \inf_{v_l\in\mathcal{V}_l}\|v - v_j\|_{0,\Omega}\,,$$

$$|||\ v\ |||^2_{Q} = \|Q_0 v\|_{0,\Omega} + \textstyle\sum_{l=1}^{L} 2^{2l}\|Q_l v - Q_{l-1}v\|_{0,\Omega}\,,$$

(3.44)

where (S) means that the infimum is taken over all splittings of the form $v = v_0 + \sum_{l=1}^{L} \sum_{j}^{J_l} v_{l,j}$. At $\varrho \equiv 1$, for any $v \in \mathcal{V}$ and $\mathcal{V} = \mathring{\mathcal{V}}$ or $\mathcal{V} = \overline{\mathcal{V}}$, the norm $\|v\|_{1,\Omega}$, is equivalent to any of the triple bar norms (3.44), uniformly in L. This is the result of Oswald [Oswald (1994)], which is Theorem 3.5 given at the end of this subsection. Let us remind that the FE functions from the space $\overline{\mathcal{V}}$ are not required to satisfy any boundary conditions. By locally applying the pointed out result for the triple bar norm $|||\cdot|||_*$ to $(v - v_0)|_{\tau_{0,j}} \in \mathcal{V}_{\tau_{0,j}} = \mathcal{V}|_{\tau_{0,j}}$, we can write

$$
\begin{aligned}
a_\Omega(v,v) &\le c \inf_{v_0 \in \mathcal{V}_0} \left[a_\Omega(v_0,v_0) + \sum_{l=j}^{J_0} \varrho_j |v - v_0|_{1,\Omega} \right] \\
&\le c \inf_{v_0 \in \mathcal{V}_0} \Bigg[a_\Omega(v_0,v_0) + \\
&\quad + \sum_{j=1}^{J_0} \varrho_j \inf_{v_{l,\tau_{0,j}} \in \mathcal{V}_{l,\tau_{0,j}} : \, (S_j)} \left(\sum_{l=1}^{L} 2^{2l} \|v_{l,\tau_{0,j}}\|_{0,\tau_{0,j}} \right) \Bigg] \\
&\le c \inf_{v_l \in \mathcal{V}_l : \, v = v_0 + \sum_{l=1}^{L} v_l} \left[a_\Omega(v_0,v_0) + \sum_{l=1}^{L} 2^{2l} \|v_l\|^2_{(0,\varrho)} \right] \\
&= c \, ||| \, v \, |||^2_{0,\varrho} \le c \, ||| \, v \, |||^2_{\varrho} , \qquad (3.45)
\end{aligned}
$$

where (S_j) means that the infimum is taken over all splittings of the form $(v - v_0)|_{\tau_{0,j}} = \sum_{l=1}^{L} v_{l,\tau_{0,j}}$. Since

$$
||| \, \phi_{l,j} \, |||^2_{\varrho} \le a_\Omega(\phi_{l,j}, \phi_{l,j}) ,
$$

we conclude that the left estmate (3.43) is valid as well.

Let $L_{2,\varrho}(\Omega)$ be the space with the scalar product

$$
(v,w)_{0,\varrho} = \sum_{j=1}^{J_0} \varrho_j (v,w)_{\Omega_j} , \qquad (3.46)
$$

and the norm $\|\cdot\|_{(0,\varrho)}$. For estimation of $\lambda_{\min}(\boldsymbol{\mathcal{K}}^{-1}\mathbf{K})$ from below, we introduce the operator $Q_{l,\varrho} : \mathcal{V} \to \mathcal{V}_l$ of the $L_{2,\varrho}$-orthogonal projection. If conditions of Theorem 3.6 below are fulfilled, then, in distinction with (3.29), for any $v \in \mathcal{V}$ we have inequality (3.48). Applying Theorems 3.5

and 3.6, we easily get

$$\begin{aligned} |||\ v\ |||_{\varrho}^2 &\le c\left[a_\Omega(Q_{0,\varrho}v, Q_{0,\varrho}v) + \sum_{l=1}^{J_0} 2^{2l}\|Q_{l,\varrho}v - Q_{l-1,\varrho}v\|_{(0,\varrho)}^2\right] \\ &\le c\left[a_\Omega(Q_{0,\varrho}v, Q_{0,\varrho}v) + \sum_{l=1}^{J_0} 2^{2l}\|v - Q_{l,\varrho}v\|_{(0,\varrho)}^2\right] \\ &\le c\left(\sum_{l=1}^{J_0} L\right) a_\Omega(v,v) \le c\,L^2 a_\Omega(v,v)\,, \end{aligned} \tag{3.47}$$

completing the proof of Theorem 3.4. □

It is possible to obtain stronger results, if the distribution of the numbers ϱ_j over the subdomains $\tau_{0,j}$ belongs to the special class of *guasi-monotone distributions*, according to the terminology of [Dryja *et al.* (1996)]. In this case an optimal condition number $\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \le c$ may be guaranteed. Moreover, [Oswald (1999b)] showed that, on certain types of coarse triangulations, any distribution of coefficients ϱ_j will be quasi-monotone.

Theorem 3.5. *At $\varrho \equiv 1$, for any $v \in \mathcal{V}$, the norm $\|v\|_{1,\Omega}$, is equivalent to any of the triple bar norms, introduced in (3.44), uniformly in L.*

Proof. See Theorem 15 in [Oswald (1994)]. □

Theorem 3.6. *Let $d = 2, 3$ and let the triangulation $\mathcal{T}_0$ be of the edge type for the case $d = 3$. Then, for any $v \in \mathcal{V}$, we have*

$$\begin{aligned} a_\Omega(Q_{l,\varrho}v, Q_{l,\varrho}v) + 2^{2l}\|v - Q_{l,\varrho}v\|_{(0,\varrho)}^2 \le \\ c\,a_\Omega(v,v)\cdot\begin{cases} L-j+1\,, & d=2\,, \\ L\,, & d=3\,. \end{cases} \end{aligned} \tag{3.48}$$

Proof. This result can be found in [Bramble and Xu (1991)] (Theorem 4.5) and [Oswald (1999b)] (Theorem 2.3) for the case $d = 2$ and $d = 3$, respectively. □

Chapter 4

Nonoverlapping DD Methods for h FE Discretizations in 2d

Striving for more robust algorithms with respect to jumps of coefficients of elliptic equations and for minimizing computational overhead resulted in the development of nonoverlapping DD methods. Such methods are much more suitable for engineering applications, because they admit decompositions in subdomains exactly corresponding to real substructures of the structure under analysis. Therefore, simpler and more efficient solvers for the problems on subdomains may be used. The predecessors of these methods were the direct substructuring elimination techniques, briefly touched in the Introduction, see also Algorithm 1.4 in Section 1.2, and indeed there are many more efficient versions of such substructuring techniques. In really efficient nonoverlapping DD methods, elimination procedures will be as a rule replaced by iterative solvers with carefully chosen preconditioners. Nonoverlapping methods, though having much in common with overlapping methods, are more complex logically. In this class of DD methods, one finds a multiplicity of subclasses, like Dirichlet–Dirichlet, Dirichlet-Neumann, Neumann-Neumann mortar, FETI algorithms etc, see, *e.g.*, [Smith *et al.* (1996); and Toselli and Widlund (2005)]. It is also worth noting that nonoverlapping algorithms for 2d and 3d elliptic problems are significantly different in their flow charts, components and difficulties related to their design and analysis. The same can be said about nonoverlapping algorithms for *hp*-versions of the finite element method. Indeed, until now, algorithms of this type are rather well developed for h-version, whereas for hp FE discretizations some components of the nonoverlapping DD algorithms are still far from optimal and their improvement will require a considerable effort. We here concentrate on Dirichlet–Dirichlet type of DD nonoverlapping algorithms.

4.1 Schur Complement Algorithms for h Discretizations

4.1.1 *Exact Schur Complement Algorithms*

For definiteness, we consider the heat conduction problem (2.5)–(2.6) from Example 2.2 as a model problem. Typically for practice, the heat conduction coefficient ϱ has jumps due to different materials. In these situations, the computational domain is usually decomposed into non-overlapping subdomains,

$$\overline{\Omega} = \bigcup_{j=1}^{J} \overline{\Omega}_j \,, \tag{4.1}$$

in such a way that coefficient jumps occur only along the inter-subdomain boundary. For simplicity, we assume that in each subdomain Ω_j the coefficient ϱ has a constant positive value ϱ_j and that the domain decomposition in subdomains is quasiuniform. The latter can be understood in a sense that subdomains are images of some canonical reference domain, *e.g.,* unit simplex, cube or a polygon, of a sufficiently regular form and the diameter is equal to one, or a few such reference domains, by mappings satisfying some quasiuniformity conditions with the scaling parameter H. In other words, for these mappings, the generalized quasiuniformity conditions are fulfilled with the mesh parameter H, which at the same time can be viewed as a typical subdomain diameter, so that $J = \mathcal{O}(H^{-d})$. As described in Section 1.2, we provide every subdomain with a quasiuniform triangulation $\mathcal{T}_j$ such that the triangulation $\mathcal{T}_h$ of the computational domain Ω,

$$\mathcal{T}_h = \bigcup_{j=1}^{J} \mathcal{T}_j \,, \qquad \overline{\Omega} = \bigcup_{j=1}^{J} \bigcup_{\tau_r \in \mathcal{T}_j} \overline{\tau}_r \,, \tag{4.2}$$

is conform and satisfies the generalized conditions of the quasiuniformity with the mesh parameter h. Again, in this case h may be a typical or maximal element diameter and, thus, $\mathcal{R} = \mathcal{R}_h = O(h^{-d})$ and the number of the internal subdomain d.o.f., denoted by N_{I_j}, behaves like $\mathcal{O}((H/h)^d)$. The domain decomposition can be produced by some quasiuniform coarse mesh as well. Indeed, more general domain decompositions are allowed, satisfying only shape regularity type conditions, some set of which will be discussed later in relation to BPS preconditioner in Section 5.1.

The FE discretization with the ordering (1.13) of the FE basis results in the block structure (1.14) of system $\mathbf{Ku} = \mathbf{f}$ the FE equations. The stiffness matrix $\mathbf{K}$ and the load vector $\mathbf{f}$ can obviously be represented in

the forms

$$\mathbf{K} = \sum_{j=1}^{J} \mathbb{T}_j \mathbf{K}_j \mathbb{T}_j^\top \quad \text{and} \quad \mathbf{f} = \sum_{j=1}^{J} \mathbb{T}_j \mathbf{f}_j \,, \tag{4.3}$$

where $\mathbb{T}_j^\top$ is the $N_j \times N$ Boolean subdomain connectivity matrix. From any vector $\mathbf{u} \in R^N$ of all nodal values, it picks up the subvector $\mathbf{u}_j = \mathbb{T}_j^\top \mathbf{u} \in R^{N_j}$ of the nodal values of the closed subdomain $\overline{\Omega}_j$. The $N_j \times N_j$ subdomain stiffness matrices $\mathbf{K}_j$, and the subdomain load vector $\mathbf{f}_j \in R^{N_j}$ may be structured in the same way as we have structured $\mathbf{K}$ and $\mathbf{f}$ in (1.14), *i.e.*,

$$\mathbf{K}_j = \begin{pmatrix} \mathbf{K}_{I_j} & \mathbf{K}_{I_j B_j} \\ \mathbf{K}_{B_j I_j} & \mathbf{K}_{B_j} \end{pmatrix} , \qquad \mathbf{f}_j = \begin{pmatrix} \mathbf{f}_{I_j} \\ \mathbf{f}_{B_j} \end{pmatrix} , \tag{4.4}$$

and the matrices $\mathbf{K}_{I_j}$ correspond to the local homogeneous Dirichlet problems, whereas each matrix $\mathbf{K}_j$ arises from the FE discretization of the local problem in Ω_j with the homogeneous Neumann boundary condition on $\partial\Omega_j \setminus \partial\Omega$. For the model problem with the differential operator contaning only second order derivatives, the latter matrices are singular, if $\partial\Omega_j \cap \partial\Omega = \varnothing$. If the FE basis is generated by the elementwise Lagrange interpolation, then the kernel (null space) $\ker(\mathbf{K}_j)$ is spanned by a single vector $\mathbf{1}_j = (1, 1, \ldots, 1)^\top \in R^{N_j}$.

Independent for each block $\mathbf{K}_j$ Gaussian elimination of the internal subdomain unknowns $\mathbf{u}_{I_j}$, commonly termed *condensation of the internal unknowns*, reduces solution of (1.14) to solving the Schur complement problem

$$\mathbf{S}_B \mathbf{u}_B = \mathbf{g}_B \,, \tag{4.5}$$

where $\mathbf{S}_B = \mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB}$ and $\mathbf{g}_B = \mathbf{f}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{f}_I$. The procedure of reducing (1.14) to (4.5) of *condensation of the internal unknowns* is one of the two keys procedures of substructuring algorithms, an example of which is Algorithm 1.4.

System (4.5) is explicitly formed and can be directly solved as in the classical substructuring Algorithm 1.4. The Schur complement $\mathbf{S}_B$ and the right-hand side $\mathbf{g}_B$ may be assembled from the local Schur complements $\mathbf{S}_{B_j}$ and the local right-hand sides $\mathbf{g}_{B_j}$ in the same way as $\mathbf{K}$ and $\mathbf{f}$ was assembled in (4.3) from $\mathbf{K}_j$ and $\mathbf{f}_j$, respectively, *i.e.*,

$$\mathbf{S}_B = \sum_{j=1}^{J} \mathbb{T}_{B_j} \mathbf{S}_{B_j} \mathbb{T}_{B_j}^\top = \biguplus_{j=1}^{J} \mathbf{S}_{B_j} \,, \qquad \mathbf{g}_B = \sum_{j=1}^{J} \mathbb{T}_{B_j} \mathbf{g}_{B_j} = \biguplus_{j=1}^{J} \mathbf{g}_{B_j} \,, \tag{4.6}$$

with the sign $\biguplus$ standing for the assembling procedure.

Forming the Schur complement $\mathbf{S}_B$ and solving the system (4.5) by direct methods may be expensive and indeed comparable in the order to the cost of solving the FE system (1.14). One of the reasons is the fill in of $\mathbf{S}_B$, which in general is of the order $J(H/h)^{2(d-1)}$. The iterative solution of (4.5), *e.g.,* by some PCGM avoids the expensive forming of the Schur complement $\mathbf{S}_B$ and results in a more efficient algorithm, which is the simplest iterative substructuring method called Schur complement PCG method. In each iteration step of the Schur complement PCG method, we need one matrix-by-vector multiplication of the form

$$\mathbf{S}_B\mathbf{v}_B^n = \sum_{j=1}^{J} \mathbb{T}_{B_j}\mathbf{S}_{B_j}\mathbb{T}_{B_j}^{\top}\mathbf{v}_B^n = \sum_{j=1}^{J} \mathbb{T}_{B_j}(\mathbf{K}_{B_j} - \mathbf{K}_{B_jI_j}\mathbf{K}_{I_j}^{-1}\mathbf{K}_{I_jB_j})\mathbb{T}_{B_j}^{\top}\mathbf{v}_B^n \tag{4.7}$$

requiring the direct solution of J systems (local Dirichlet problems)

$$\mathbf{K}_{I_j}\mathbf{w}_{I_j}^n = \mathbf{K}_{I_jB_j}\mathbb{T}_{B_j}^{\top}\mathbf{v}_B^n\,, \qquad j = 1,\ldots,J\,, \tag{4.8}$$

which can be done completely in parallel. Moreover, the factorization of the matrices $\mathbf{K}_{I_j}$ in a preprocessing step and the use of sparse direct techniques can make this multiplication operation very efficient, see, *e.g.,* the monograph by [George and Liu (1981)], [Demmel *et al.* (1999), Gupta (2002)] and more recent publications on $\mathcal{H}$-matrices and tensor-train decompositions techniques of [Hackbusch *et al.* (2005)], [Bebendorf (2008)], [Hackbusch (2009)], [Dolgov *et al.* (2011)], [Khoromskij (2011)] and [Hackbusch (2012)]. Nevertheless, for real large scale problems, this operation is a bottleneck of the Schur complement PCG. The use of inexact iterative solvers for the local Dirichlet problems (4.8) in (4.7) helps to circumvent this bottleneck, but may easily result in wrong results or even in divergence of the PCG iterations , if used without sufficient care.

The forming of the Schur complement can be interpreted as some kind of a preconditioning operation. Indeed, under the conditions of quasiuniformity, the spectral condition number of the Schur complement

$$\mathrm{cond}\,[\mathbf{S}_B] = \lambda_{max}(\mathbf{S}_B)/\lambda_{min}(\mathbf{S}_B) = O(H^{-1}h^{-1})$$

is much better than the spectral condition number of the original stiffness matrix $\mathrm{cond}\,[\mathbf{K}] = O(h^{-2})$, since $h \ll H$, see, *e.g.,* [Brenner (1999)]. However, $\mathrm{cond}\,[\mathbf{S}_B]$ still depends on the DD parameter H, on the global discretization parameter h and on the coefficients jumps in a bad way. Therefore, we need such a s.p.d. Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$

that cond $[\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{S}_B]$ does not depend too much on these parameters and that the preconditioning operation $\mathbf{w}_B^n = \boldsymbol{\mathcal{S}}_B^{-1}\mathbf{d}_B^n$, mapping the defect $\mathbf{d}_B^n$ into preconditioned defect $\mathbf{w}_B^n$, is sufficiently cheap. Incorporating inexact solvers for the local Dirichlet problems and the interface problem in the described above algorithm leads to the class of *DD Dirichlet–Dirichlet type algorithms*, which for many practical problems can be more efficient, but at the same time have a more complicated structure. The Schur complement preconditioner-solvers are among the most important ingredients of this as well as many other classes of iterative substructuring methods. Some general facts related to the Schur complement preconditioning, are discussed in the next subsubsection.

4.1.2 *Schur Complement Preconditioning*

Speaking more precisely, we would like to have s.p.d. Schur complement preconditioners $\boldsymbol{\mathcal{S}}_B$ satisfying the following conditions:

1) Spectral equivalence or closeness in the spectrum to $\mathbf{S}_B$, which assumes that the spectral equivalence inequalities

$$\underline{\gamma}_B\,\boldsymbol{\mathcal{S}}_B \;\le\; \mathbf{S}_B \;\le\; \overline{\gamma}_B\,\boldsymbol{\mathcal{S}}_B \tag{4.9}$$

hold with positive $\underline{\gamma}_B$ and $\overline{\gamma}_B$, such that cond$[\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{S}_B] \le \overline{\gamma}_B/\underline{\gamma}_B$ does not, or only weakly depends on h, H, and the coefficient jumps. The latter property is often referred as *robustness against coefficient jumps*.

2) The number of arithmetical operations ops$[\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{d}_B^n]$ needed for the preconditioning operation should be in the best case of the order $\mathcal{O}(N_B)$, or at least, should not enlarge the order of the overall complexity of the DD algorithm too much, where N_B is the number of d.o.f. living on the inter-subdomain boundary

$$\Gamma_B = \cup_{j=1}^{J}\partial\Omega_j \setminus \partial\Omega\,.$$

3) The preconditioning operation $\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{d}_B^n$ should not worsen, at least, worsen not too much, the efficiency of parallelization of the total algorithm.

Taking into consideration 3), we however should be aware that, in many Schur complement preconditioners, some coarse grid solver, managing the global information transport, is hidden. Therefore, at least, some additional communication between processors is inevitable.

In the literature, there are several basic proposals for constructing Schur complement preconditioners. Many of them rest on the fact that under, *e.g.,* conditions of quasiuniformity of FE mesh, the Schur complement energy $(\mathbf{S}_B\mathbf{v}_B, \mathbf{v}_B)$ is equivalent to the square of the so called *broken weighted* $H^{1/2}$*-seminorm*

$$|v_B|^2_{\alpha,1/2,\Gamma_B} = \sum_{j=1}^{J} \alpha_j |v_{B_j}|^2_{1/2,\partial\Omega_j}$$
$$= \sum_{j=1}^{J} \alpha_j \int_{\partial\Omega_j}\int_{\partial\Omega_j} \frac{|v_{B_j}(y) - v_{B_j}(x)|^2}{|y-x|^d} ds_y ds_x \,, \quad (4.10)$$

assuming that α_j are properly chosen, *e.g.,* $\alpha_j = \varrho_j$. This assumes the existence of positive constants $\underline{\delta}_B$ and $\overline{\delta}_B$, which are independent of h, H and the coefficient jumps, such that

$$\underline{\delta}_B \, |v_B|^2_{\varrho,1/2,\Gamma_B} \le (\mathbf{S}_B\mathbf{v}_B, \mathbf{v}_B) \le \overline{\delta}_B \, |v_B|^2_{\varrho,1/2,\Gamma_B} \,, \quad (4.11)$$

which hold for all $v_B \leftrightarrow \mathbf{v}_B \in R^{N_B}$, where v_B is the restrictions of the functions $v \in \mathring{\mathcal{V}}$ to the interface boundary Γ_B. The constants $\underline{\delta}_B$ and $\overline{\delta}_B$ depend on the constants entering the quasiuniformity conditions, see (3.1)–(3.2). Note that, on the subspace $\mathring{\mathcal{V}}$, the seminorm $|v|_{1/2,\Gamma_B}$ is a norm. The inequalities (4.11) are a consequence of similar inequalities for subdomains

$$\underline{\delta}_{B_j} \, |v_{B_j}|^2_{1/2,\partial\Omega_j} \le \varrho_j^{-1}(\mathbf{S}_{B_j}\mathbf{v}_{B_j}, \mathbf{v}_{B_j}) \le \overline{\delta}_{B_j} \, |v_{B_j}|^2_{1/2,\partial\Omega_j} \,, \quad (4.12)$$

which hold for all $v_{B_j} \leftrightarrow \mathbf{v}_{B_j} \in R^{N_{B_j}}$ with uniformly bounded constants $\underline{\delta}_{B_j} \ge \underline{\delta}_B$ and $\overline{\delta}_{B_j} \le \overline{\delta}_B$.

The equivalences (4.11)–(4.12) reflect the following facts. According to the formula $\mathbf{S}_{B_j} = \mathbf{K}_{B_j} - \mathbf{K}_{B_jI_j}\mathbf{K}_{I_j}^{-1}\mathbf{K}_{I_jB_j}$, the subdomain Schur complement is the result of the condensation of the internal unknowns for the subdomain $\partial\Omega_j$. Hence, in the case of $\varrho \equiv 1$, the inequalities (4.12) become

$$\underline{\delta}_{B_j} \, |v_{B_j}|^2_{1/2,\partial\Omega_j} \le \inf_{\phi|_{\partial\Omega_j} = v_{B_j}} |\phi|_{1,\Omega_j} \le \overline{\delta}_{B_j} \, |v_{B_j}|^2_{1/2,\partial\Omega_j} \,, \; \forall \, v_{B_j} \in \overline{\mathcal{V}}_j \,, \quad (4.13)$$

where $\overline{\mathcal{V}}_j$ is the restriction of $\mathcal{V}$ to subdomain Ω_j. Therefore, the association of the quadratic form $\varrho_j^{-1}(\mathbf{v}_{B_j}, \mathbf{S}_{B_j}\mathbf{v}_{B_j})$ with the squared seminorm $|v_{B_j}|^2_{1/2,\partial\Omega_j} = |v_{B_j}|^2_{H^{1/2}(\partial\Omega_j)}$, where $v_{B_j} \leftrightarrow \mathbf{v}_{B_j}$, becomes transparent. In the functional analysis, for the case of v_{B_j} being the trace of $\phi \in H^1(\Omega_j)$, the equivalence (4.13) is proved by the trace (left inequality) and prolongation (right inequality) theorems, see, *e.g.*, [Babich and Slobodetski (1956); Gagliardo (1957); and Maz'ya and Poborchi (1998)]. For FE functions, the

trace theorem retains since the FE space is a subspace of $H^1(\Omega_j)$, whereas the proof of the prolongation theorem should be completed anew due to the same reason. Contemporary literature on the existence and design of the prolongation operators which possess this property, including the operators cheap in computer realization, is very vast. In relation to DD methods, these problems were studied, *e.g.,* by [Nepomnyaschikh (1991b)]. Prolongation operators, providing the right inequality (4.13) for FE functions, sometimes can serve as convenient and cheap numerical prolongation operators. For this reason, they play yet another important role in the practical nonoverlapping DD algorithms which are discussed in Section 4.2, where additional references may be found.

The inequalities (4.12) give an instrument for obtaining subdomain Schur complement preconditioners $\boldsymbol{\mathcal{S}}_{B_j}$ as matrices $\varrho_j \boldsymbol{\mathcal{S}}_{1/2,j}$, where $\boldsymbol{\mathcal{S}}_{1/2,j}$ is the matrix of the quadratic form $|v_{B_j}|^2_{1/2,\partial\Omega_j}$ or the matrix spectrally close to it. In turn the latter, which in general is the completely filled in matrix, may be defined by means of the K-interpolation technique between easily calculated matrices of the quadratic forms $|\mathbf{v}_{B_j}|^2_{0,\partial\Omega_j}$ and $|\mathbf{v}_{B_j}|^2_{1,\partial\Omega_j}$.

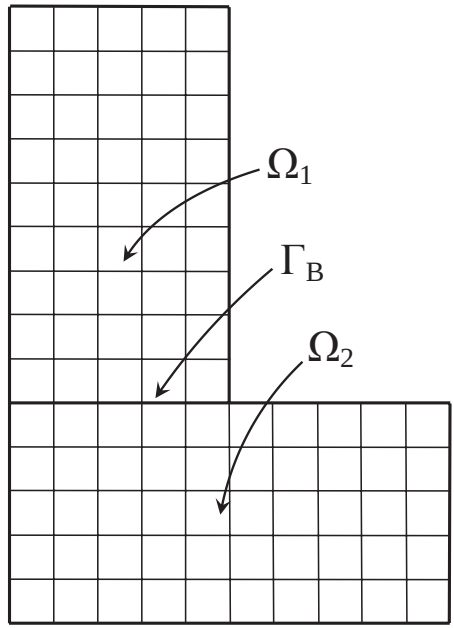

Fig. 4.1 Two subdomains and their interface.

The outlined approach is best illustrated by Dryja's Schur complement preconditioner that was approved for the interface boundary of the decomposition of the L-shaped domain in two nonoverlapping subdomains, as shown in Fig. 4.1. Here the interface Γ_B consists of only one straight piece with N_B nodal points inside Γ_B. For simplicity, assume $\varrho \equiv 1$ and consider the matrix

$$\boldsymbol{\mathcal{S}}_B = \boldsymbol{\Delta}_B^{1/2} := \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}^{1/2} , \tag{4.14}$$

as Schur complement preconditioner. It represents the multiplied by $\sqrt{h}$ square root of the discretized 1d Laplacian Δ_B along the interface, subjected to homogeneous Dirichlet boundary conditions at the end points of Γ_B. For the proof of the spectral equivalence inequalities (4.9) with h-independent constants $\underline{\gamma}_B$ and $\overline{\gamma}_B$, it is sufficient to take into account two facts. One is that, due to the homogeneous Dirichlet boundary conditions at the end points of Γ_B, (4.11) becomes

$$\underline{\delta}_{B\,00}|v_B|^2_{1/2,\Gamma_B} \le (\mathbf{S}_B\mathbf{v}_B, \mathbf{v}_B) \le \overline{\delta}_{B\,00}|v_B|^2_{1/2,\Gamma_B}, \quad \forall\, v_B \leftrightarrow \mathbf{v}_B \in R^{N_B}. \tag{4.15}$$

Another lies in the characterization of ${}_{00}|\cdot|_{1/2,\Gamma_B}$ as the result of the K-interpolation between the spaces $\mathring{H}^1(\Gamma_B)$ and $L_2(\Gamma_B)$ supplied with the norms $|\cdot|_{1,\Gamma_B}$ and $||\cdot||_{0,\Gamma_B}$, respectively. Therefore, the spectrally equivalent to $\mathbf{S}_B$ Schur complement preconditioner can be obtained by the K-interpolation between the quadratic forms induced by the 1d stiffness and mass FE matrices. Completition of the latter procedure directly leads to (4.14).

The suggestion to use the matrix (4.14) for the Schur complement preconditioner and its justification goes back to [Dryja (1982)], see additionally [Dryja (1984)], Theorem 1. Some improvements of Dryja's preconditioner were discussed later by [Golub and Mayers (1984); Bjørstad and Widlund (1986); Chan (1987)] and others. We also refer to [Andreev (1972, 1973)], where similar to (4.14) representations of the discrete norm ${}_{00}|\cdot|_{1/2,\Gamma_B}$ can be found in relation to finite-difference approximations.

In defiance of the matrix $\boldsymbol{\Delta}_B^{1/2}$ implicit definition due to its complicated form – *e.g.,* it is completely filled in – it is easily and efficiently handled numerically. The matrix $\boldsymbol{\Delta}_B$ has the eigenvalues and the corresponding

orthonormal eigenvectors

$$\lambda_k = 4\sin^2\frac{k\pi}{2(N_B+1)} \text{ and } \mathbf{v}_{B,k} = \sqrt{\frac{2}{N_B+1}}\left(\sin\frac{k\pi l}{N_B+1}\right)_{l=1}^{N_B},$$

respectively, where $k = 1, 2, \ldots, N_B$. The matrix $\boldsymbol{\Delta}_B^{1/2}$ has the same eigenvectors and the eigenvalues $\lambda_k^{1/2}$. Therefore, having defined the diagonal matrix $\boldsymbol{\Lambda}_B^{1/2} = \mathrm{diag}(\lambda_k^{1/2})$ and the Fourier matrix $\mathbf{F}_B = (\mathbf{v}_{B,1}, \ldots, \mathbf{v}_{B,N_B})$, we conclude that

$$\boldsymbol{\Delta}_B^{1/2} = \mathbf{F}_B^T\, \boldsymbol{\Lambda}_B^{1/2}\, \mathbf{F}_B\,. \tag{4.16}$$

Since the Fourier matrix $\mathbf{F}_B$ is orthogonal, *i.e.,* $\mathbf{F}_B^{-1} = \mathbf{F}_B^T$, the preconditioning equation $\boldsymbol{\mathcal{S}}_B\mathbf{w}_B = \mathbf{d}_B$ may be solved in the following three steps:

Fast discrete Fourier analysis:	$\mathbf{x}_B = \mathbf{F}\mathbf{d}_B\,,$
Diagonal scaling:	$\mathbf{y}_B = \boldsymbol{\Lambda}_B^{-1/2}\mathbf{x}_B\,,$
Fast discrete Fourier synthesis:	$\mathbf{w}_B = \mathbf{F}^T\mathbf{y}_B\,.$

The total complexity of the preconditioning operation is of the order $N_B \ln(N_B)$ with the main contributions from Fourier analysis and Fourier synthesis.

The construction of good Schur complement preconditioners in the general case of many non-overlapping subdomains is more involved. In the two-dimensional case, a cheap and highly parallelizable Schur complement preconditioner was suggested by [Bramble *et al.* (1986)]. Its design is the most simple if the domain decomposition grid is the coarse FE grid embedded in the discretization grid. For simplicity, we assume this below at the brief description of BPS preconditioner $\boldsymbol{\mathcal{S}}_B$, how it is commonly referred in the literature. The BPS preconditioner has the form

$$\boldsymbol{\mathcal{S}}_B^{-1} = \sum_{E_k} \mathbb{T}_{E_k}\boldsymbol{\mathcal{S}}_{E_k}^{-1}\mathbb{T}_{E_k}^{\top} + \mathbf{T}_V\boldsymbol{\mathcal{K}}_{V,\mathrm{coarse}}^{-1}\mathbf{T}_V^{\top} \tag{4.17}$$

in which summation is done over all edges $E_k \subset \Omega$ of the subdomains of the coarse domain decomposition grid. Here $\mathbb{T}_{E_k}^{\top}$ picks up the edge degrees of freedom, belonging to the edge E_k, from the vector of V_B of the interface finite element degrees of freedom, $\mathbf{T}_V$ transforms a coarse grid vector, corresponding to (hierarchical) vertex basis functions, into the vector of the space V_B, $\boldsymbol{\mathcal{S}}_{E_k}$ is an edge Schur complement preconditioner, and $\boldsymbol{\mathcal{K}}_{V,\mathrm{coarse}}$ denotes a coarse grid preconditioner. The latter one manages the global information exchange and can be, *e.g.,* the stiffness matrix $\mathbf{K}_{V,\mathrm{coarse}}$ induced by the coarse grid discretization. Dryja's type Schur complement preconditioners may be used for edges. More exactly, if $E_k \in \partial\Omega_j \cap \partial\Omega_i$ then

$\boldsymbol{\mathcal{S}}_{E_k} = (\varrho_j + \varrho_i)\boldsymbol{\Delta}_{E_k}^{1/2}$, where $\boldsymbol{\Delta}_{E_k}^{1/2}$ is similarly defined to $\boldsymbol{\Delta}_B^{1/2}$ in (4.14). From (4.17), we conclude that the preconditioner requires solving in parallel systems for each edge and for vertex degrees of freedom. [Bramble *et al.* (1986)] proved that, for the BPS preconditioner $\boldsymbol{\mathcal{S}}_B$, the relative condition number grows at most according to the bound

$$\kappa(\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{S}_B) \leq c_{\mathrm{BPS}}(1+\ln(H/h))^2\,, \tag{4.18}$$

and $c_{\mathrm{BPS}} = \text{const}$ independent of h and of jumps of the coefficient ϱ. Another important positive quality is that it is fast. However, the condition number of $\mathbf{K}_{V,\mathrm{coarse}}$ depends on jumps of the coefficients of the elliptic problem, and special care should be taken when solving systems with this matrix.

A simpler, but less efficient Schur complement preconditioner for the $2d$ case was introduced by [Nepomnyaschikh (1991a)]. It results from the inexact iterative solver for another Schur complement preconditioner, in which vertex degrees of freedom are not split from ones living on the edges. Let $\mathcal{V}_{\partial_j}(\partial\Omega_j)$ be the subspace of all the FE boundary functions on $\partial\Omega_j$, *i.e.,* the restriction to $\partial\Omega_j$ of the space $\overline{\mathcal{V}}(\Omega_j)$. With the purpose to obtain such a preconditioner, we first establish that, for all j, the inequalities

$$\underline{\beta}_B\,|v_{\partial_j}|^2_{1/2,\partial\Omega_j} \;\leq\; (\mathbb{S}_{1/2,j}\mathbf{v}_{\partial_j}, \mathbf{v}_{\partial_j}) \;\leq\; \overline{\beta}_B\,|v_{B_j}|^2_{1/2,\partial\Omega_j} \tag{4.19}$$

are valid for all $\mathbf{v}_{B_j} \leftrightarrow v_{B_j} \in \mathcal{V}_{B_j}(\partial\Omega_j)$, where $\underline{\beta}_B$ and $\overline{\beta}_B$ are positive constants, the matrix in the middle term is defined as follows:

$$\mathbb{S}_{1/2,j} = \boldsymbol{\Delta}_{\partial\Omega_j}^{1/2} := \begin{pmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{pmatrix}^{1/2} . \tag{4.20}$$

The matrix $\boldsymbol{\Delta}_{\partial\Omega_j}$ is a $N_{\partial\Omega_j} \times N_{\partial\Omega_j}$ cyclic tridiagonal matrix, and $N_{\partial\Omega_j}$ is the number of nodes on $\partial\Omega_j$. The seminorm $|\cdot|_{1/2,\partial\Omega_j}$ for functions from the space $H^{1/2}(\partial\Omega_j)$ is equivalent to one obtained by the method of K-interpolation between the factor spaces $H^1(\partial\Omega_j)/R$ and $L_2(\partial\Omega_j)/R$ with the norms $|\cdot|_{1,\partial\Omega_j}$ and $|\cdot|_{0,\partial\Omega_j}$,

$$|v|_{0,\partial\Omega_j} = \inf_{c\in R} ||v-c||_{0,\partial\Omega_j}\,, \qquad \forall\, v \in L_2(\partial\Omega_j)\,.$$

As a consequence, the matrix of the quadratic form $|v_{\partial_j}|^2_{1/2,\partial\Omega_j}$, induced by the FE space $\mathcal{V}_{\partial_j}(\partial\Omega_j)$, can be obtained by the K-interpolation between

the matrices of the quadratic forms $|v_{\partial_j}|^2_{1,\partial\Omega_j}$ and $|v_{\partial_j}|^2_{0,\partial\Omega_j}$, induced by the same space. Implementation of the interpolation procedure easily leads to the conclusion that $\mathbb{S}_{1/2,j}$ is a proper representation of the quadratic form $|v_{\partial_j}|^2_{1/2,\partial\Omega_j}$ and (4.19) holds. If the decomposition grid is a coarse FE grid imbedded in the discretization FE grid, then the constants $\underline{\beta}_B$ and $\overline{\beta}_B$ depend on the quasiuniformity conditions for both grids. If subdomains are images of several reference domains, then they depend on the quasiuniformity conditions for the mappings and on configurations of the reference domains.

Now, for each subdomain $\overline{\Omega}_j \subset \Omega$, we can pick up the matrix $\boldsymbol{\mathcal{S}}_{1/2,j}$ satisfying

$$(\mathbf{v}_{B_j}, \boldsymbol{\mathcal{S}}_{1/2,j}\mathbf{v}_{B_j}) = (\mathbf{v}_{\partial\Omega_j}, \mathbb{S}_{1/2,j}\mathbf{v}_{\partial\Omega_j}) \tag{4.21}$$

for all $\mathbf{v}_{B_j} \in R^{N_{B_j}}$ and the vectors $\mathbf{v}_{\partial\Omega_j} \in R^{N_{\partial\Omega_j}}$, which are obtained by continuation of each $\mathbf{v}_{B_j}$ by zero entries on all nodes belonging to the set $\partial\Omega_j \cap \partial\Omega$. For subdomains, which boundaries have nonempty intersections with $\partial\Omega$, *i.e.*, $\partial\Omega_j \cap \partial\Omega \neq \varnothing$, one can also use for $\boldsymbol{\mathcal{S}}_{1/2,j}$ the Dryja preconditioners described above. By the analogy with (4.6), we set

$$\boldsymbol{\mathcal{S}}_B = \sum_{j=1}^{J} \varrho_j \mathbb{T}_{B_j} \boldsymbol{\mathcal{S}}_{1/2,j} \mathbb{T}^{\top}_{B_j} = \biguplus_{j=1}^{J} \varrho_j \boldsymbol{\mathcal{S}}_{1/2,j}\,, \tag{4.22}$$

and from (4.11) and (4.19) and the definition of the preconditioner $\boldsymbol{\mathcal{S}}_B$ conclude that

$$(\underline{\delta}_B/\overline{\beta}_B)\ \boldsymbol{\mathcal{S}}_B \leq \mathbf{S}_B \leq\ (\overline{\delta}_B/\underline{\beta}_B)\ \boldsymbol{\mathcal{S}}_B\,. \tag{4.23}$$

This Schur complement preconditioner has a too complicated structure for solving systems $\boldsymbol{\mathcal{S}}_B\mathbf{w}_B = \mathbf{d}_B$ by direct methods. However, the matrix-vector multiplications by this matrix are cheap, since each matrix $\mathbb{S}_{1/2,j}$ is represented in the form similar to (4.16)

$$\mathbb{S}_{1/2,j} = \mathbf{F}^T_{\partial\Omega_j}\, \boldsymbol{\Lambda}^{1/2}_{\partial\Omega_j}\, \mathbf{F}_{\partial\Omega_j} \tag{4.24}$$

with the similarly defined orthonormal matrices $\mathbf{F}_{\partial\Omega_j}$ of FDFT (fast discrete Fourier transforms) and diagonal matrices $\boldsymbol{\Lambda}_{\partial\Omega_j}$. Note that each matrix $\boldsymbol{\Lambda}_{\partial\Omega_j}$ contains one zero eigenvalue on diagonal, corresponding to the constant eigenvector, but since matrix $\mathbf{F}_{\partial\Omega_j}$ is not singular, $\boldsymbol{\Lambda}_{\partial\Omega_j}$ is uniquely defined. Therefore, the multiplication $\mathbf{w}_B := \boldsymbol{\mathcal{S}}_B\mathbf{v}_B$ is produced

by completing the steps:

$$\begin{aligned}
&\text{Finding restrictions:} && \mathbf{v}_{B_j} = \mathbb{T}_{B_j}^{\top}\,\mathbf{v}_B\,,\\
&\text{In parallel for} j = 1,2,..,J:\\
&\qquad\text{Fast discrete Fourier analysis:} && \mathbf{x}_{\partial\Omega_j} = \mathbf{F}_{\partial\Omega_j}\,\mathbf{v}_{\partial\Omega_j}\,,\\
&\qquad\text{Diagonal scaling:} && \mathbf{y}_{\partial\Omega_j} = \mathbf{\Lambda}^{1/2}_{\partial\Omega_j}\,\mathbf{x}_{\partial\Omega_j}\,,\\
&\qquad\text{Fast discrete Fourier synthesis:} && \mathbf{w}_{\partial\Omega_j} = \mathbf{F}^{T}_{\partial\Omega_j}\,\mathbf{y}_{\partial\Omega_j}\,,\\
&\text{Assembling:} && \mathbf{w}_B := \biguplus_{j=1}^{J}\mathbf{w}_{B_j}\,,
\end{aligned} \tag{4.25}$$

where $\mathbf{v}_{\partial\Omega_j}$ is related to $\mathbf{v}_{B_j}$ as in (4.21) and $\mathbf{w}_{B_j}$ is the restriction of $\mathbf{w}_{\partial\Omega_j}$ to the nodes of $\partial\Omega_j \setminus \partial\Omega$. The total computational work is dominated again by FDFT, which which behaves like $\mathcal{O}(N_{B_j}\log N_{B_j})$ for every j. Thus, $\mathrm{ops}\,[\boldsymbol{\mathcal{S}}_B\mathbf{v}_B] = \mathcal{O}(N_B\log(\max_j\, N_{B_j}))$.

The new Schur complement preconditioner, which is denoted $\boldsymbol{\mathcal{S}}_{\mathrm{it},B}$, can be defined in a similar way to the one presented in Subsection 2.2.3. Namely, it can be defined as the result of inexact solution of the system $\boldsymbol{\mathcal{S}}_B\mathbf{w}_B = \mathbf{d}_B$ by means of the iterative process

$$\mathbf{w}_B^{k+1} = \mathbf{w}_B^{k} - \sigma_{k+1}\boldsymbol{\mathcal{D}}_B^{-1}(\boldsymbol{\mathcal{S}}_B\mathbf{w}_B^{k} - \mathbf{d}_B)\,, \qquad \mathbf{v}_B^{0} = \mathbf{0}\,, \tag{4.26}$$

with the diagonal matrix $\boldsymbol{\mathcal{D}}_B$, having the same main diagonal with the matrix $\boldsymbol{\mathcal{S}}_B$, and variable, *e.g.* Chebyshev, iteration parameters σ_{k+1}. Suppose that at each DD-PCGM iteration, we make a fixed number n_ϵ of iterations (4.26), which provide the reduction of the error in the norm $\|\cdot\|_{\boldsymbol{\mathcal{S}}_B}$ in half, *i.e.,* $\epsilon = 1/2$. Then we implicitly define the preconditioner, the inverse to which is

$$\boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1} := \boldsymbol{\mathcal{I}}_{\circ}^{-1}[\boldsymbol{\mathcal{S}}_B,\boldsymbol{\mathcal{D}}_B] = [\mathbf{I} - \prod_{k=1}^{n_\epsilon}(\mathbf{I} - \sigma_k\boldsymbol{\mathcal{D}}_B^{-1}\boldsymbol{\mathcal{S}}_B)]\,\boldsymbol{\mathcal{S}}_B^{-1}\,, \qquad \epsilon = 0.5\,, \tag{4.27}$$

and which satisfies the inequalities

$$\frac{1}{2}\,\boldsymbol{\mathcal{S}}_{\mathrm{it},B} \le \boldsymbol{\mathcal{S}}_B \le \frac{3}{2}\,\boldsymbol{\mathcal{S}}_{\mathrm{it},B} \text{ and } \frac{\underline{\delta}_B}{2\,\overline{\beta}_B}\,\boldsymbol{\mathcal{S}}_B \le \mathbf{S}_B \le \frac{3\,\overline{\delta}_B}{2\,\underline{\beta}_B}\,\boldsymbol{\mathcal{S}}_B\,. \tag{4.28}$$

The condition number of the preconditioner $\boldsymbol{\mathcal{D}}_B$ relative to the matrix $\boldsymbol{\mathcal{S}}_B$ is $\mathcal{O}(H/h)$. Therefore, at the choice of the Chebyshev iteration parameters according to [Axellson (1994); and Samarskii and Nikolayev (1989)], the inequalities (4.28) are guaranteed at the number of the iterations $n_\epsilon = \mathcal{O}(\sqrt{H/h})$. This is a consequence of Lemma 2.1 and estimate (2.52). However, the main contribution to the computational cost at each iteration is from the matrix-vector multiplication (4.25), and the total cost of

the inexact solution of the system $\boldsymbol{\mathcal{S}}_B \mathbf{w}_B = \mathbf{d}_B$ or, more precisely, of the operation $\boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1}\mathbf{d}_B$ is bounded according to

$$\mathrm{ops}\,[\boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1}\mathbf{d}_B] \le c_1\sqrt{H/h}\,N_B \log(\max_j N_{B_j}) \le c_2(Hh)^{-1}\sqrt{H/h}\,|\log H/h|\,,$$

where c_1 and $c_2 = \mathrm{const}$. If, for instance, we have a fast solver for the internal Dirichlet problems requiring $\mathcal{O}(N^2(\log N))$ operations, than the whole Schur complement algorithm has the same cost, since the cost of solving the system with Schur complement is subordinate. Let us emphasize that the described solver is robust against jumps of coefficients of the elliptic equation, if they occur on the inter-subdomain boundary and the coefficients do not change much inside each subdomain.

Schur complement preconditioners in many cases can be designed as complete counterparts of BPX preconditioners for the FE stiffness matrices. Such preconditioners were introduced by [Tong *et al.* (1991)]. Similarly to BPX preconditioner, it is based on the fast forth-back transformations of the FE nodal basis to the Riesz multilevel basis, called also *generating system*, and the diagonal representation of the preconditioner in this basis. Whereas the BPX preconditioner for the matrix $\mathbf{K}$, corresponding to the Poisson equation in the domain Ω, requires an equivalent multilevel representation of the $H^1(\Omega)$-norm for FE functions, the corresponding BPX preconditioner for the Schur complement requires similar representation for the $H^{1/2}(\Gamma_B)$-norm. Indeed, for the Riesz basis in $\mathcal{V}_B(\Gamma_B)$, it can be taken the restriction of the Riesz basis in the FE space $\mathcal{V}(\Omega)$ to Γ_B. [Smith and Widlund (1990)] and [Haase *et al.* (1991)] also proposed Schur complement preconditioners of the hierarchical type, which asymptotically behave like the BPS preconditioner in the 2d case.

There are various other proposals for Schur complement preconditioners. Let us here mention only the *probing technique* proposed by [Chan and Mathew (1994)], see also [Keyes and Gropp (1987)], and the techniques based on the boundary element method, see, *e.g.*, [Carstensen *et al.* (1998); Haase *et al.* (1997); and Steinbach (2003)].

4.2 Dirichlet–Dirichlet DD Algorithms

4.2.1 *Flow Chart of DD Algorithms*

In general, a fast solver for local Dirichlet problems will be an iterative one. In this case, according to theoretical results and numerical practice, the most efficient strategy for solving the FE system (2.12) will be the

PGC method with some DD preconditioner based on fast inexact solvers for both the internal Dirichlet subproblems in subdomains and for the interface subproblem. The so-called *Dirichlet–Dirichlet DD algorithms* compose one of the classes of such algorithms. We are now turning to the consideration of such kind of DD methods for 2d elliptic problems.

It is natural to distinguish *internal, edge* and *vertex* degrees of freedom, and, respectively, decompose the vector space

$$V = V_I \oplus V_E \oplus V_V \, .$$

For designing DD solvers or their parts, it is useful to introduce also the decompositions

$$V = V_I \oplus V_B \quad \text{and} \quad V = V_\Sigma \oplus V_V ,$$

where $V_B = V_E \oplus V_V$ is termed the subspace of the *(coupling) interface boundary* degrees of freedom and $V_\Sigma = V_I \oplus V_E$. For the corresponding FE subspaces, we use the notations $\mathcal{V}_I$, $\mathcal{V}_B$, $\mathcal{V}_E$, $\mathcal{V}_V$ etc. According to these subspaces, the FE stiffness matrix may be represented in either of the block forms

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,B} \\ \mathbf{K}_{B,I} & \mathbf{K}_B \end{pmatrix} = \begin{pmatrix} \mathbf{K}_\Sigma & \mathbf{K}_{\Sigma,V} \\ \mathbf{K}_{V,\Sigma} & \mathbf{K}_V \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,E} & \mathbf{K}_{I,V} \\ \mathbf{K}_{E,I} & \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,I} & \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix},$$

$$\mathbf{K}_B = \begin{pmatrix} \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix}, \quad \mathbf{K}_\Sigma = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,E} \\ \mathbf{K}_{E,I} & \mathbf{K}_E \end{pmatrix}, \tag{4.29}$$

where $\mathbf{K}_L$ are $N_L \times N_L$, $L = I, E, V$, sub-matrices corresponding to the internal, edge and vertex unknowns, respectively. We assume also that the decomposition in subdomains is induced by some imbedded coarse FE mesh with the corresponding coarse FE subspace denoted $\mathcal{V}_{V,\mathrm{coarse}}$ and spanned by the coarse grid basis functions.

A popular Dirichlet–Dirichlet DD preconditioner is associated with the BPS Schur complement preconditioner and has the following four major components:

- Preconditioner-solvers for the Dirichlet problems on the subdomains Ω_j which are blocks of the preconditioner-solver $\boldsymbol{\mathcal{K}}_I = \mathrm{diag}\,[\boldsymbol{\mathcal{K}}_{I_j}]_{j=1}^J$ for the matrix $\mathbf{K}_I$.
- Preconditioner-solver $\boldsymbol{\mathcal{S}}_E = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{E_q}]_{q=1}^Q$ for the edge Schur complement, where Q is the number of edges forming the interface boundary.
- Prolongation matrix $\boldsymbol{\mathcal{P}} : V_E \to V$ from the edges onto the whole domain.

- Preconditioner-solver $\boldsymbol{\mathcal{K}}_{V,\mathrm{coarse}}$ for the global coarse grid vertex problem.

In general, the inverse to this preconditioner may be represented by expressions

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \boldsymbol{\mathcal{P}}_B \boldsymbol{\mathcal{S}}_B^{-1} \boldsymbol{\mathcal{P}}_B^T = \boldsymbol{\mathcal{K}}_I^{+} + \boldsymbol{\mathcal{P}}_E \boldsymbol{\mathcal{S}}_E^{-1} \boldsymbol{\mathcal{P}}_E^T + \boldsymbol{\mathcal{K}}_V^{+} = \textstyle\sum_{j=1}^{J} \mathbb{T}_{I_j} \boldsymbol{\mathcal{K}}_{I_j}^{-1} \mathbb{T}_{I_j}^{\top} +$$

$$+ \textstyle\sum_{q=1}^{Q} (\mathbb{T}_{I_{j_q}} \boldsymbol{\mathcal{P}}_{I_{j_q} \leftarrow E_q} + \mathbb{T}_{I_{i_q}} \boldsymbol{\mathcal{P}}_{I_{i_q} \leftarrow E_q} + \mathbb{T}_{E_q}) \boldsymbol{\mathcal{S}}_{E_q}^{-1} (\mathbb{T}_{E_q}^{\top} + \boldsymbol{\mathcal{P}}_{I_{i_q} \leftarrow E_q}^{\top} \mathbb{T}_{I_{i_q}}^{\top} +$$

$$+ \boldsymbol{\mathcal{P}}_{I_{j_q} \leftarrow E_q}^{T} \mathbb{T}_{I_{j_q}}^{\top}) + \mathbf{T}_V \boldsymbol{\mathcal{K}}_{V,\mathrm{coarse}}^{-1} \mathbf{T}_V^T , \tag{4.30}$$

where $\mathbb{T}_{I_j}^{\top}$ picks up components of a vector from the space V, belonging to the interior of a subdomain Ω_j, $\mathbf{T}_V$ transforms a coarse grid vector, corresponding to (hierarchical) vertex basis functions, into the vector of the space V, the matrices $\boldsymbol{\mathcal{S}}_{E_q}$ and $\boldsymbol{\mathcal{K}}_{V,\mathrm{coarse}}$ are the same as in (4.17). In (4.30), i_q and j_q stand for the numbers of subdomains adjacent to the edge E_q. To be defined below matrices

$$\boldsymbol{\mathcal{P}}_B = \begin{pmatrix} \boldsymbol{\mathcal{P}}_{I \leftarrow B} \\ \mathbf{I}_B \end{pmatrix} : V_B \to V \quad \text{and} \quad \boldsymbol{\mathcal{P}}_E = \begin{pmatrix} \boldsymbol{\mathcal{P}}_{I \leftarrow E} \\ \mathbf{I}_E \\ \mathbf{0} \end{pmatrix} : V_E \to V \tag{4.31}$$

are called prolongation matrices, whereas $\boldsymbol{\mathcal{P}}_{I_{j_q} \leftarrow E_q}$ is the restriction of $\boldsymbol{\mathcal{P}}_{I \leftarrow E}$ or $\boldsymbol{\mathcal{P}}_{I \leftarrow B}$, defining the prolongation from the edge E_q inside the subdomain Ω_{j_q} to which this edge belongs. Notations of the unit matrices in (4.31) are supplied with the indices characterizing degrees of the freedom on which they are defined. Let $\overline{V}_j$ be the restriction of V to $\overline{\Omega}_j$. Usually, when designing the prolongation operator, we first define the prolongation matrices $\boldsymbol{\mathcal{P}}_{B_j} : V_{B_j} \to \overline{V}_j$ for each subdomain and then assemble the global prolongation matrix $\boldsymbol{\mathcal{P}}_B$ in such a way that the former matrices are the restrictions of the latter to subdomains $\overline{\Omega}_j$. We denote by $\mathcal{P}_B$, $\mathcal{P}_E$, $\mathcal{P}_{B_j}$ etc. the corresponding maps in the FE space, where, *e.g.*, $\mathcal{P}_{B_j}$ is the map $V_{B_j} \to \overline{V}_j$ and V_{B_j} is the restriction of V to $\partial\Omega_j \setminus \partial\Omega$.

As can be seen from (4.30), DD-PCG algorithm with this preconditioner completely avoids operations with the Schur complement as well as its implicit calculations. The procedure for solving the system $\boldsymbol{\mathcal{K}}\mathbf{v} = \mathbf{f}$ and, therefore, the whole DD algorithm are highly parallelizable. This procedure may be arranged according to Algorithm 4.1, in which $E_{j,q}$ stands for the edge "q" of the subdomain Ω_j. In distinction with the Schur complement

Algorithm 4.1 Solving system $\mathcal{K}\mathbf{v} = \mathbf{f}$ for 2d h discretization with the Dirichlet–Dirichlet DD preconditioner, vector/matrix version.

for $j = 1, 2, .., J$ **do**

1) $\mathbf{v}_{I_j} := \mathcal{K}_{I_j}^{-1}\mathbf{f}_{I_j}$, { solving local Dirichlet problems for each subdomain Ω_j in parallel}

for q **such that** $E_{j,q} \in \partial\Omega_j$ **do**

2) $\mathbf{f}_{E_{j,q}}^{(1)} := \mathcal{P}_{I_j \leftarrow E_q}^{T}\mathbf{f}_{I_j}$, { restrictions to subdomain edges in parallel}

end for

end for

for $q = 1, 2, .., Q$ **do**

for $j = j', j''$ **such that** $\partial\Omega_{j'} \cap \partial\Omega_{j''} = \overline{E}_q$ **do**

4) $\mathbf{f}_{E_q} := \mathbf{f}_{E_q} + \mathbf{f}_{E_{j,q}}^{(1)}$, { updating edge subvectors of the right part in parallel}

end for

5) $\mathbf{v}_{E_q} := \mathcal{S}_{E_q}^{-1}\mathbf{f}_{E_q}$, {in parallel edgewise solving systems with edge Schur complement
preconditioners and updated subvectors $\mathbf{f}_{E_q}$ by FDFT}

end for

for $j = 1, 2, .., J$ **do**

for q **such that** $E_q \in \partial\Omega_j$ **do**

6) $\mathbf{v}_{I_{j,q}}^{(1)} := \mathcal{P}_{I_j \leftarrow E_q}\mathbf{v}_{E_q}$, { prolongations inside subdomains from their edges in parallel}

7) $\mathbf{v}_{I_j} := \mathbf{v}_{I_j} + \mathbf{v}_{I_{j,q}}^{(1)}$, {subdomain-edgewise updating internal subvectors in parallel}

end for

end for

8) $\mathbf{v}_V := \mathcal{K}_V^{-1}\mathbf{f}_V$, {solving global vertex subproblem}

9) **Set** $\mathbf{v}_{I_r}, \mathbf{v}_{E_q}, \mathbf{v}_V$ **for components of** $\mathbf{v}$, ***i.e.***,

$\mathbf{v}^\top := (\mathbf{v}_{I_1}^\top, .., \mathbf{v}_{I_R}^\top, \mathbf{v}_{E_1}^\top, .., \mathbf{v}_{E_Q}^\top, \mathbf{v}_V^\top)$**.**

methods, we have two additional components: solvers for the local Dirichlet problems and prolongations. Indeed, both of them are implicitely present in the Schur complement algorithms. They are used in the preprocessor for reducing the FE equations to the Schur complement FE problem and in the postprocessor for calculating the internal unknowns for subdomains, respectively, and are assumed to be exact, even if they are based on iterative procedures. In general in DD-PCG solver, the both of them can be inexact with several new options for prolongations.

The DD preconditioner incorporating the Schur complement preconditioner (4.27) can be written in the form

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \boldsymbol{\mathcal{P}}_B \boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1} \boldsymbol{\mathcal{P}}_B^T , \tag{4.32}$$

which is also easy to rewrite in a subdomainwise parallelized form.

Among the components of DD algorithms, solving local discrete Dirichlet problems on subdomains of decomposition will usually contribute to the main bulk of the computational work. This is even more so for the overlapping DD algorithms. Therefore, the optimization of the local Dirichlet solvers is the most important task. If vertex d.o.f.'s are split from others, the corresponding global discrete problem is similar to the one mentioned, as well as to the global coarse grid problem in the overlapping DD algorithms. It is worth noting, however, that different mesh parameters h and H corresponding to local and global subproblems, respectively, result often in implementation of different solvers. Only in a restricted number of situations, the operation $\boldsymbol{\mathcal{K}}_{I_j}^{-1}\mathbf{v}_{I_j}$ may be performed by some simple fast solver like FDFT , or by a combination of 1d FDFT and 1d direct methods, *e.g.,* the Thomas algorithm called *progonka* in the Russian literature, or by any other direct method with the complexity close to linear. Practically, these cases are limited to the subdomains of decomposition, whose triangulations $\mathcal{T}_{h,j}$ are topologically equivalent to regular triangulations of a rectangle. Besides, the continuous mapping of $\mathcal{T}_{h,j}$ on the corresponding triangulation of the rectangle, which defines the equivalent FE discretization on the rectangle, must be sufficiently good. If $\mathcal{T}_{h,j}$ ia a quasiuniform triangulation, then the requirement to the mapping is automatically fulfilled. If the triangulation $\mathcal{T}_{h,j}$ is obtained by, *e.g.,* the diadic refinement of a quasiuniform coarse grid triangulation, then the BPX subdomain preconditioner-solver $\boldsymbol{\mathcal{K}}_{I_j}$ of the form (3.42) will be fast. Other multilevel preconditioner-solvers, such as MDS preconditioners, based on multiscale stable decompositions may be also mentioned in this contest. In general, the preconditioner-solvers, produced by means of any optimal and almost optimal inexact iterative solvers will be another good choice. The ways of implementation and

justification of such preconditioner-solvers $\boldsymbol{\mathcal{K}}_{I_j} = \mathbf{B}_{I_j,\mathrm{it}}$ for these problems (in particular multigrid ones) were discussed earlier in Subsection 2.2.3. If $\varrho = \varrho_j = \text{const}$ or is changing slightly in Ω_j, then indeed many efficient fast multilevel iterative solvers may be incorporated in DD code. In this tool box one may find classical, sometimes also called geometric, multigrid methods presented in the books by [Bramble and Zhang (2000); Hackbusch (1985); and Shaidurov (1995)] algebraic multigrid methods, see review article by [Stüben (2001)], and algebraic multilevel methods of [Axelsson and Vassilevski (1989)], see also monographs by [Kraus and Margenov (2009); and Vassilevski (2008)]. Besides, efficient multigrid algorithms have been developed for many sophisticated elliptic problems, including problems with orthotropic and deteriorating coefficients, problems with singularities etc.

We gave above only references on a few sources from the vast realm of publications on multigrid or multilevel methods, and other efficient methods for solving finite element systems of equations for elliptic problems.

4.2.2 *Relative Condition Number of DD Preconditioners*

We are now going to estimate the relative condition number of Dirichlet–Dirichlet DD preconditioners, which defines the rate of convergence of iterative processes with this kind of preconditioners.

Theorem 4.1. *Let us suppose that the matrices $\boldsymbol{\mathcal{K}}_{I_j}$ and the operators $\mathcal{P}_{B_j}$ satisfy the inequalities*

$$\underline{\gamma}_j \boldsymbol{\mathcal{K}}_{I_j} \le \mathbf{K}_{I_j} \le \overline{\gamma}_j \boldsymbol{\mathcal{K}}_{I_j} \quad \text{and} \quad |\mathcal{P}_{B_j} v|^2_{1,\Omega_j} \le c_{j,P}\, |v|^2_{1/2,\partial\Omega_j} \tag{4.33}$$

for all $v \in \mathcal{V}(\partial\Omega_j)$, and let us further assume that the bounds (4.9) hold for the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$. Then

$$\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \le 2\, c_P^2 \underline{\beta}_B^{-2} \max(\overline{\gamma}, \overline{\gamma}_B) / \min(\underline{\gamma}, \underline{\gamma}_B)\,, \tag{4.34}$$

where $\overline{\gamma} = \max \overline{\gamma}_j$, $\underline{\gamma} = \min \underline{\gamma}_j$ $c_P = \max c_{j,P}$, and $\underline{\beta}_B$ from (4.11) .

Proof. Taking the definition of the matrix $\mathbf{K}_j$, relations (4.10)–(4.12) and inequality (4.33) for $\mathcal{P}_{B_j}$ into account, we obtain the estimate $\|\boldsymbol{\mathcal{P}}_{B_j}\mathbf{v}\|_{\mathbf{K}_j} \le c_{j,P}\underline{\beta}_j^{-1}\|\mathbf{v}\|_{\mathbf{S}_j}$ for any $\mathbf{v} \in V_{B_j}$ and j. Therefore, we have

$$\|\boldsymbol{\mathcal{P}}\mathbf{v}\|_{\mathbf{K}} \le c_P \underline{\beta}^{-1} \|\mathbf{v}\|_{\mathbf{S}_B}\,, \quad \forall\ \mathbf{v} \in V_B\,. \tag{4.35}$$

Clearly, due to (4.33) for $\boldsymbol{\mathcal{K}}_j$, we have

$$\underline{\gamma}_I \boldsymbol{\mathcal{K}}_I \le \mathbf{K}_I \le \overline{\gamma}_I \boldsymbol{\mathcal{K}}_I \tag{4.36}$$

Now, the proof follows from (4.36), (4.35), (4.9), the definition of DD preconditioner in (4.30), and Theorem 4.2 and its Corollary 4.1, see below, originally proved by [Haase *et al.* (1994)]. □

An important conclusion from the above theorem is that the quality of the DD preconditioner is completely defined by the quality of the preconditioners for the local internal subproblems and Schur complement as well as the quality of the prolongation operator.

Now, we will formulate similar results in a purely algebraic form. Let $\mathbf{S} = \mathbf{A}_2 - \mathbf{A}_{21}\mathbf{A}_1^{-1}\mathbf{A}_{12}$ be the Schur complement for the $n \times n$ nonnegative symmetric matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_2 \end{pmatrix}$$

with the $n_k \times n_k$ blocks $\mathbf{A}_k$ on diagonal, $n_1 + n_2 = n$, and $\mathbf{A}_1$ be a positive matrix. We remind that if $\ker[\mathbf{A}] \neq \{\mathbf{0}\}$, then $\ker[\mathbf{A}] = \ker[\mathbf{S}]$. In the case of $\mathbf{A}$ with a nontrivial kernel, we term $(\cdot,\cdot)_{\mathbf{A}} := (\cdot, \mathbf{A}\cdot)$ and $\|\cdot\|_{\mathbf{A}} := \sqrt{(\cdot,\cdot)_{\mathbf{A}}}$ by the scalar product and the norm, respectively, with the understanding that they are such in the factor space $V/(\ker[\mathbf{A}])$, and $\|\cdot\|_{\mathbf{A}}$ is only a seminorm in V. Let also $\boldsymbol{\mathcal{P}}_2 : R^{n_2} \to R^n$ be the "prolongation" matrix

$$\boldsymbol{\mathcal{P}}_2 = \begin{pmatrix} \boldsymbol{\mathcal{P}}_{21} \\ \boldsymbol{\mathcal{P}}_{22} \end{pmatrix}, \quad \boldsymbol{\mathcal{P}}_{22} = \mathbf{I},$$

such that for some $c_S > 0$

$$\|\boldsymbol{\mathcal{P}}_2\mathbf{v}_2\|_{\mathbf{A}}^2 \leq c_S \|\mathbf{v}_2\|_{\mathbf{S}}^2, \quad \forall\ \mathbf{v}_2 \in R^{n_2}. \tag{4.37}$$

Note that, if $\mathbf{v}_2 \in \ker[\mathbf{S}]$, this inequality assumes the inclusion $\boldsymbol{\mathcal{P}}_2\mathbf{v}_2 \in \ker[\mathbf{A}]$ and otherwise it can't be fulfilled. By means of the transformation matrix $\mathbf{C}$, we introduce the matrix $\mathbf{B} = \mathbf{C}^T\mathbf{A}\mathbf{C}$, where

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & \boldsymbol{\mathcal{P}}_{21} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_2 \end{pmatrix},$$

and the block diagonal matrix $\mathbf{B}_{\mathrm{diag}} = \mathrm{diag}[\mathbf{B}_1, \mathbf{B}_2]$. Obviously $\mathbf{B}_1 = \mathbf{A}_1$. We denote by V and V_k the spaces R^n and R^{n_k}, respectively, and introduce the space $\widetilde{V}_2 := \boldsymbol{\mathcal{P}}_2 V_2$. According to the theorem below, the splitting of the space V into the direct sum of the subspaces $\widetilde{V}_1 = V_1$ and $\widetilde{V}_2$ is stable, under the condition that c_S is a constant.

Theorem 4.2. *Let* $\mathbf{A}$ *be a positive symmetric matrix, the prolongation matrix* $\boldsymbol{\mathcal{P}}_2$ *satisfy (4.37), and* $\boldsymbol{\mathcal{A}}$ *be the matrix defined through its inverse*

$$\boldsymbol{\mathcal{A}}^{-1} := \mathbf{A}_1^{+} + \boldsymbol{\mathcal{P}}_2\mathbf{S}^{-1}\boldsymbol{\mathcal{P}}_2^T. \tag{4.38}$$

Then

$$\underline{c}_B\mathbf{B}_{\rm diag} \le \mathbf{B} \le \overline{c}_B\mathbf{B}_{\rm diag} \quad \textit{and} \quad \underline{c}_B\boldsymbol{\mathcal{A}} \le \mathbf{A} \le c_S\overline{c}_B\boldsymbol{\mathcal{A}}, \tag{4.39}$$

where $\underline{c}_B = 1-\sqrt{1-c_S^{-1}} \ge (2c_S)^{-1}$ *and* $\overline{c}_B = 1+\sqrt{1-c_S^{-1}} \le 2-c_S^{-1} \le 2$.

Proof. The basic for the proof of the theorem are the inequalities

$$\begin{aligned}\gamma_B := \cos\measuredangle(V_1,\widetilde{V}_2) &= \sup_{\mathbf{v}_k\in V_k\setminus\{\mathbf{0}\}} \frac{\mathbf{v}_1^T\mathbf{B}\mathbf{v}_2}{\|\mathbf{v}_1\|_{\mathbf{B}}\|\mathbf{v}_2\|_{\mathbf{B}}}\\ &= \sup_{\mathbf{v}_k\in V_k\setminus\{\mathbf{0}\}} \frac{\mathbf{v}_1^T\mathbf{B}_{12}\mathbf{v}_2}{\|\mathbf{v}_1\|_{\mathbf{B}_1}\|\mathbf{v}_2\|_{\mathbf{B}_2}}\\ &\le \sqrt{1-c_S^{-1}}\,,\end{aligned} \tag{4.40}$$

originally obtained by [Haase *et al.* (1991)], and one of characterizations of the constant γ_B:

$$\gamma_B^2 = \sup_{\mathbf{v}_2\in V_2\setminus\{\mathbf{0}\}} \frac{\mathbf{v}_2^\top\mathbf{B}_{21}\mathbf{B}_1^{-1}\mathbf{B}_{12}\mathbf{v}_2}{||\mathbf{v}_2||_B^2}\,, \tag{4.41}$$

independent proofs of which were given by [Axelsson and Vassilevski (1989); Bjørstad and Mandel (1991); and Haase *et al.* (1994)]. For deriving (4.40), we note that the equalities

$$\mathbf{S} = \mathbf{B}_2 - \mathbf{B}_{21}\mathbf{B}_1^{-1}\mathbf{B}_{12} = \mathbf{A}_2 - \mathbf{A}_{21}\mathbf{A}_1^{-1}\mathbf{A}_{12} \quad \text{and} \quad ||\mathbf{v}_2||_B^2 = \|\boldsymbol{\mathcal{P}}_2\mathbf{v}_2\|_{\mathbf{A}}^2\,,$$

follow from the definitions of the involved matrices. By means of these inequalities, we transform the representation (4.41) and come to the squared inequality (4.40) as follows:

$$\begin{aligned}\gamma_B^2 &= \sup_{\mathbf{v}_2\in V_2\setminus\{\mathbf{0}\}} \frac{\mathbf{v}_2^\top\mathbf{B}_{21}\mathbf{B}_1^{-1}\mathbf{B}_{12}\mathbf{v}_2}{||\mathbf{v}_2||_B^2}\\ &= \sup_{\mathbf{v}_2\in V_2\setminus\{\mathbf{0}\}} \{1 - \frac{\mathbf{v}_2^\top[\mathbf{B}_2 - \mathbf{B}_{21}\mathbf{B}_1^{-1}\mathbf{B}_{12}]\mathbf{v}_2}{||\mathbf{v}_2||_B^2}\}\\ &= \sup_{\mathbf{v}_2\in V_2\setminus\{\mathbf{0}\}} \{1 - \frac{\mathbf{v}_2^\top\mathbf{S}\mathbf{v}_2}{\|\boldsymbol{\mathcal{P}}_2\mathbf{v}_2\|_{\mathbf{A}}^2}\}\\ &\le 1 - \inf_{\mathbf{v}_2\in V_2\setminus\{\mathbf{0}\}} \frac{\mathbf{v}_2^\top\mathbf{S}\mathbf{v}_2}{\|\boldsymbol{\mathcal{P}}_2\mathbf{v}_2\|_{\mathbf{A}}^2} \le 1 - c_s^{-1}\,.\end{aligned}$$

It can be seen from the above relations that, if the constant c_s in (4.37) is sharp, then the estimate (4.40) is also sharp. In this case, the sign $\le$ in the last inequality above can be replaced by the sign of equality.

The inequalities of the theorem with the matrix $\mathbf{B}$ are the consequence of (4.40) and the equality $\|\boldsymbol{\mathcal{P}}_2\mathbf{v}_2\|_{\mathbf{A}}^2 = \mathbf{v}_2^T\mathbf{B}_2\mathbf{v}_2$. Now we introduce the block

diagonal matrix $\mathbf{A}_{_S} = \mathrm{diag}\,[\mathbf{A}_1, \mathbf{S}]$ and note that an equivalent definition of $\boldsymbol{\mathcal{A}}$ is

$$\boldsymbol{\mathcal{A}} = \mathbf{C}^{-T}\mathbf{A}_{_S}\mathbf{C}^{-1}\,.$$

For this reason and in view of (4.37) and the pair of inequalities (4.39) for the matrix $\mathbf{B}$, we can write

$$\mathbf{A} = \mathbf{C}^{-T}\mathbf{B}\,\mathbf{C}^{-1} \le \overline{c}_B\mathbf{C}^{-T}\mathbf{B}_{\mathrm{diag}}\mathbf{C}^{-1} \le \overline{c}_B c_{_S}\mathbf{C}^{-T}\mathbf{A}_{_S}\mathbf{C}^{-1} = \overline{c}_B c_{_S}\boldsymbol{\mathcal{A}}\,,$$

and

$$\mathbf{A} \ge \underline{c}_B\mathbf{C}^{-T}\mathbf{B}_{\mathrm{diag}}\mathbf{C}^{-1} \ge \underline{c}_B\mathbf{C}^{-T}\mathbf{A}_{_S}\mathbf{C}^{-1} = \underline{c}_B\boldsymbol{\mathcal{A}}\,.$$

This completes the proof of the theorem. □

Corollary 4.1. *Suppose the matrices $\boldsymbol{\mathcal{A}}_1$ and $\boldsymbol{\mathcal{S}}$ satisfy the inequalities*

$$\underline{c}_A\boldsymbol{\mathcal{A}}_1 \le \mathbf{A}_1 \le \overline{c}_A\boldsymbol{\mathcal{A}}_1 \quad \textit{and} \quad \underline{c}_S\boldsymbol{\mathcal{S}} \le \mathbf{S} \le \overline{c}_S\boldsymbol{\mathcal{S}}\,, \tag{4.42}$$

and $\boldsymbol{\mathcal{A}}^{-1} = \boldsymbol{\mathcal{A}}_1^{+} + \boldsymbol{\mathcal{P}}_2\boldsymbol{\mathcal{S}}^{-1}\boldsymbol{\mathcal{P}}_2^{T}$. Then

$$\mathrm{cond}[\,\boldsymbol{\mathcal{A}}^{-1}\mathbf{A}] \le c_{_S}\,\frac{1+\sqrt{1-c_{_S}^{-1}}\,\max\{\overline{c}_A,\overline{c}_S\}}{1-\sqrt{1-c_{_S}^{-1}}\,\min\{\underline{c}_A,\underline{c}_S\}} \le 4c_{_S}^2\,\frac{\max\{\overline{c}_A,\overline{c}_S\}}{\min\{\underline{c}_A,\underline{c}_S\}}\,.$$

If $\boldsymbol{\mathcal{S}}$ is constructed by means of approximation of the matrix $\mathbf{B}_2$, and, instead of inequalities (4.42) for $\mathbf{S}$, there hold the inequalities $\underline{c}_S\boldsymbol{\mathcal{S}} \le \mathbf{B}_2 \le \overline{c}_S\boldsymbol{\mathcal{S}}$, then

$$\mathrm{cond}[\,\boldsymbol{\mathcal{A}}^{-1}\mathbf{A}] \le \frac{\overline{c}_B}{\underline{c}_B}\,\frac{\max\{\overline{c}_A,\overline{c}_S\}}{\min\{\underline{c}_A,\underline{c}_S\}} \le 4c_S\,\frac{\max\{\overline{c}_A,\overline{c}_S\}}{\min\{\underline{c}_A,\underline{c}_S\}}\,.$$

Proof. The corollary directly follows from (4.37), (4.39) and (4.42). □

At defining the interface preconditioner $\boldsymbol{\mathcal{S}}$, a direct approximation of $\mathbf{B}_2$ may give better condition estimates, especially if the prolongation operator is not close to a discrete harmonic. For a suitable set of conditions, one can adopt the bounds

$$\underline{c}_A\mathbf{v}_1^{\top}\boldsymbol{\mathcal{A}}_1\mathbf{v}_1 \le \mathbf{v}^{\top}\mathbf{B}\mathbf{v}\,, \quad \mathbf{A}_1 \le \overline{c}_A\boldsymbol{\mathcal{A}}_1\,, \quad \underline{c}_S\boldsymbol{\mathcal{S}} \le \mathbf{S}\,, \quad \mathbf{B}_2 \le \overline{c}_S\boldsymbol{\mathcal{S}}\,. \tag{4.43}$$

Corollary 4.2. *Under conditions (4.43), we have the bound*

$$\mathrm{cond}\,[\,\boldsymbol{\mathcal{A}}^{-1}\mathbf{A}] \le 2(\underline{c}_A^{-1} + \underline{c}_S^{-1})\,\max(\overline{c}_A,\overline{c}_S)\,. \tag{4.44}$$

Proof. For the preconditioner $\boldsymbol{\mathcal{B}} = \mathrm{diag}\,[\boldsymbol{\mathcal{A}}_1, \boldsymbol{\mathcal{S}}]$ of $\mathbf{B}$, the same energy equivalence estimates hold as for the preconditioner $\boldsymbol{\mathcal{A}}$ of $\mathbf{A}$, from where we immediately get (4.44). □

Remark 4.1. Let us assume that $\mathbf{A}$ is a nonnegative symmetric matrix and $\mathbf{A}_1$ is positive. Then Theorem 4.2 remains valid with $\boldsymbol{\mathcal{A}}^{+} := \mathbf{A}_1^{+} + \boldsymbol{\mathcal{P}}_2\mathbf{S}^{+}\boldsymbol{\mathcal{P}}_2^{T}$.

4.2.3 *Discrete Low Energy Prolongations*

4.2.3.1. Direct Discrete Low Energy Prolongations. The structure of the DD preconditioner closely reflects the factorization

$$\mathbf{K}^{-1} = \begin{pmatrix} \mathbf{I} & -\mathbf{K}_I^{-1}\mathbf{K}_{I,B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{K}_I^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K}_{B,I}\mathbf{K}_I^{-1} & \mathbf{I} \end{pmatrix}, \tag{4.45}$$

with the matrices $\boldsymbol{\mathcal{K}}_I$, $\boldsymbol{\mathcal{S}}_B$ and

$$\boldsymbol{\mathcal{P}}_B = \begin{pmatrix} \boldsymbol{\mathcal{P}}_{I,B} \\ \boldsymbol{\mathcal{P}}_{B,B} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mathcal{P}}_{I,B} \\ \mathbf{I}_B \end{pmatrix}$$

replacing $\mathbf{K}_I$, $\mathbf{S}_B$ and

$$\mathbf{P}_B = \begin{pmatrix} \mathbf{P}_{I,B} \\ \mathbf{P}_{B,B} \end{pmatrix} := \begin{pmatrix} -\mathbf{K}_I^{-1}\mathbf{K}_{I,B} \\ \mathbf{I}_B \end{pmatrix},$$

respectively. The prolongation $\mathbf{u} = \mathbf{P}_B\mathbf{v}_B$ specifies the FE function $u \leftrightarrow \mathbf{u}$, which is *discrete harmonic* inside each subdomain, *i.e.,* u is the FE approximation of the harmonic prolongation. For this reason, the prolongation satisfying the right inequality (4.33) is called *bounded* or *almost discrete harmonic* or *low energy prolongation*, respectively, *extension.* Alongside with *prolongation/extension* the term *lifting* is also widely used in the literature. Prolongations used in DD algorithms can be subdivided in two groups. In the first group, operations $\mathbf{K}_{I_j}^{-1}\mathbf{K}_{I_j,B_j}\mathbf{v}_{B_j}$ are performed approximately by means of inexact solvers for systems with the system matrices $\mathbf{K}_{I_j}$ and the corresponding replacement for the matrices $\mathbf{K}_{I_j,B_j}$. Thus, we have two options for the prolongation matrix:

$$\boldsymbol{\mathcal{P}}_B = \begin{pmatrix} -\boldsymbol{\mathcal{K}}_I^{-1}\boldsymbol{\mathcal{K}}_{I,B} \\ \mathbf{I}_B \end{pmatrix} \quad \text{and} \quad \boldsymbol{\mathcal{P}}_B = \begin{pmatrix} -\boldsymbol{\mathcal{B}}_{I,\mathrm{it}}^{-1}\boldsymbol{\mathcal{K}}_{I,B} \\ \mathbf{I}_B \end{pmatrix}.$$

In the other group, the solution of local Dirichlet problems for prolongations are completely avoided and replaced by the multiplications $\boldsymbol{\mathcal{P}}_{I_j}\mathbf{v}_{B_j}$ with cheap and explicitly defined matrices $\boldsymbol{\mathcal{P}}_{I_j}$. For the construction of such matrices $\boldsymbol{\mathcal{P}}_{I_j}$, different ideas may be used. Some of them, like averaging, which we only refer below, bear similarity to the techniques used for the constructive definition of the prolongations $\mathcal{P}_{I_j} : H^{1/2}(\partial\Omega_j) \to H^1(\Omega_j)$ in Sobolev spaces.

A computable extension procedure of that kind was proposed by [Matsokin and Nepomnyaschikh (1985)], see also [Nepomnyaschikh (1991b)]. It provides a uniform in h and H bound (4.33) with the constant $c_{j,P}$, depending only on the constants in the conditions of quasiuniformity, and is optimal in the arithmetic operations count. [Haase *et al.* (1994)] introduced an alternative approach termed the hierarchical extension which

is almost optimal in the arithmetic operations count. In 2d, it leads to a $\log(H/h)$ growth of $c_{j,\mathcal{P}}$ that can be compensated by $\mathcal{O}(\log\log(H/h))$ multigrid iterations. In 3d this extension is not so efficient. However, the variant of a multilevel extension that was studied by [Haase and Nepomnyaschikh (1997)] is almost optimal in 2d and 3d. When the subdomains of decomposition are introduced by a coarse grid triangulation, their fine triangulations $\mathcal{T}_{H,j}$ may be topologically equivalent to regular triangulations of some canonical reference subdomains. In this case, prolongations of the described and other types can be especially efficient and simple, and easier generalized to singular in some aspects discretizations, *e.g.,* discretizations of thin rectangular domains, orthotropic discretizations with high aspect ratios etc. The literature on low energy prolongations for boundary value problems of different sort is now very rich. It is also worth noting that the problems of designing good Schur complement preconditioners, *e.g.,* of BPX type, and low energy prolongations are closely interrelated. In this connection, we refer to [Griebel and Oswald (1995b); Khoromskij and Wittum (2004); Korneev *et al.* (2007); and Oswald (1999a)].

4.2.3.2. Prolongations by means of inexact Dirichlet solvers. Performing prolongations by means of inexact solvers for the local Dirichlet problems seems to be a reasonable approach. However, a naive replacement of an exact solution procedure, designated by $\mathbf{K}_I^{-1}$ in the left and right factors of the factorization (4.45), by an inexact solver may not produce a satisfactory prolongation. Suppose, similarly to the use of inexact preconditioner-solvers for local Dirichlet problems, discussed in Subsection 2.2.3, the operation $\mathbf{K}_I^{-1}$ is replaced by one multigrid cycle, *i.e.,* by $(\mathbf{I}_I - \mathbf{M}_I^s)\mathbf{K}_I^{-1}$ with $s = 1$. In general, one may not obtain a good prolongation and, consequently, a stable space decomposition even for a multigrid cycle for which the energy norm of the multigrid iteration operator less than unity independently of the discretization parameter. It is proven in [Börgers (1989)] and [Haase *et al.* (1991)] that one needs at least $s = \mathcal{O}(\log(H/h))$ multigrid cycles to get a satisfactory extension. In other words, in contrast to the use of the same inexact solver for the internal Dirichlet problems, see (2.53) and Lemma 2.1, the number of iterations depends on H/h. This conclusion also holds for other inexact iterative solvers with linear error transition operators. Below, we consider two examples.

We first turn to our model problem (2.4) with the the bilinear form and linear form from Example 2.1 and with $\mathbb{V} = \mathring{H}^1(\Omega)$. Suppose, we have good preconditioner-solvers $\boldsymbol{\mathcal{K}}_{I_j}$ for the Dirichlet problems on subdomains,

satisfying the inequalities

$$\underline{\gamma}_j \boldsymbol{\mathcal{K}}_{I_j} \le \mathbf{K}_{I_j} \le \overline{\gamma}_j \boldsymbol{\mathcal{K}}_{I_j}\,, \qquad \underline{\gamma}_j > 0\,. \tag{4.46}$$

Since the trace theorem is used to estimate the number of iterations, the related conditions on subdomains of decompositions are required. It is sufficient, *e.g.,* to assume that the subdomains are the nests of some coarse shape regular mesh. Alternatively, we can assume that the subdomains Ω_j are associated with the finite number of the canonical reference domains $\mathcal{D}_\varkappa$, $\varkappa = 1, 2, .., \varkappa_\circ$, by some mappings $\Upsilon_j(y) : \overline{\mathcal{D}}_\varkappa \to \overline{\Omega}_j$ with $\varkappa = \varkappa(j)$. Suppose also $\mathcal{H}_{j,k}$ and $\theta_{j,k}$ are Lamé coefficients and the angles for these mappings, introduced similarly to Lamé coefficients and the angles $H_k^{(r)}, \theta_k^{(r)}$ in Section 3.2 for the mappings $\mathcal{X}^{(r)}(y)$. Then similar to (3.6) conditions of shape regularity can be written as

$$\alpha_D^{(1)} \mathcal{H}_j \le \mathcal{H}_{j,k}(y) \le \mathcal{H}_j\,,\ 0 < \theta_D \le \theta_{j,k}(y) \le \pi - \theta_D\,,\ \forall\ y \in \mathcal{D}_{\varkappa(j)}\,, \tag{4.47}$$

with some positive constants $\alpha_D^{(1)}$ and θ_D. In this section, we use the notation $n_j = \mathcal{H}_j/h$.

Lemma 4.1, formulated below, is indeed fulfilled under the assumptions that the pre-image of the FE triangulation $\mathcal{T}_j$ on each reference subdomain $\mathcal{D}_{\varkappa(j)}$ is quasiuniform, but mappings $\Upsilon^{j,\varkappa}(y)$ are only shape regular. For simplicity, however, in this section it is assumed that the FE triangulation of the domain Ω and mappings are quasiuniform, and, therefore, it can be adopted that $\mathcal{H}_j \equiv \mathcal{H}$.

For each subdomain of decomposition, the prolongation operator is defined independently, and the definition depends on the restriction $a_{\Omega_j}(u, v)$ of the bilinear form $a(u, v) = a_\Omega(u, v)$ to Ω_j. We will distinguish the following two situations:

$$\text{A)} \quad \begin{cases} \mu_{1,j} \|v\|^2_{1,\Omega_j} \le a_{\Omega_j}(v, v)\,, & \forall v \in H^1(\Omega_j)\,, \\ a_{\Omega_j}(u, v) \le \mu_{2,j}\, \|u\|_{1,\Omega_j}\, \|v\|_{1,\Omega_j}\,, & \forall\, u, v \in H^1(\Omega_j)\,, \end{cases} \tag{4.48}$$

and

$$\text{B)} \quad \begin{cases} \mu_{1,j} |v|^2_{1,\Omega_j} \le a_{\Omega_j}(v, v)\,, & \forall v \in H^1(\Omega_j)\,, \\ a_{\Omega_j}(u, v) \le \mu_{2,j}\, |u|_{1,\Omega_j}\, |v|_{1,\Omega_j}\,, & \forall\, u, v \in H^1(\Omega_j)\,, \end{cases} \tag{4.49}$$

with some positive constants $\mu_{k,j}$.

In the case A), for any $\mathbf{v}_{B_j} \in V_{B_j}$, the prolongation $\mathbf{u}_j = \boldsymbol{\mathcal{P}}_{B_j} \mathbf{v}_{B_j}$ is produced in fact by means of the implicitly defined matrix $\boldsymbol{\mathcal{P}}_{B_j} = \boldsymbol{\mathcal{P}}_{B_j,\mathrm{it}}$ in the following way. If $\mathbf{u}_{I_j}$ and $\mathbf{u}_{B_j}$ are the subvectors of $\mathbf{u}_j$, then $\mathbf{u}_{B_j} = \mathbf{v}_{B_j}$, and $\mathbf{u}_{I_j} = \mathbf{u}_{I_j}^{l_j}$, is found from the iteration process

$$\mathbf{u}_{I_j}^{k+1} = \mathbf{u}_{I_j}^{k} - \sigma_{k+1} \boldsymbol{\mathcal{K}}_{I_j}^{-1} (\mathbf{K}_{I_j} \mathbf{u}_{I_j}^{k} + \mathbf{K}_{I_j B_j} \mathbf{v}_{B_j})\,, \qquad \mathbf{u}_{I_j}^{0} = \mathbf{0}\,, \tag{4.50}$$

with Chebyshev iteration parameters σ_k for some fixed number l_j of iterations.

Case B) is different, the right inequality (4.33), obviously, can't be fulfilled, if the prolongation operator does not reproduce the constant function from the constant function given on the boundary. There is a few remedies to provide this property. One of the simplest is to define the prolongation operator as

$$\boldsymbol{\mathcal{P}}_{B_j}\mathbf{v}_{B_j} = \overline{\mathbf{v}} + \boldsymbol{\mathcal{P}}_{B_j,\mathrm{it}}\mathbf{w}_{B_j}\,, \qquad \mathbf{w}_{B_j} := (\mathbf{v}_{B_j} - \overline{\mathbf{v}}_{B_j})\,, \tag{4.51}$$

where $\boldsymbol{\mathcal{P}}_{B_j,\mathrm{it}}$ is the prolongation operator, defined as above by an inexact iterative solver. All entries of the vectors $\overline{\mathbf{v}}_{B_j}$ and $\overline{\mathbf{v}}$ have the average value on $\partial\Omega_j$ of the FE function $v_{B_j} \leftrightarrow \mathbf{v}_{B_j}$, *i.e.*

$$\overline{v}_{B_j} = \frac{1}{\mathrm{mes}\,\partial\Omega_j}\int_{\partial\Omega_j} v_{B_j}\,ds\,.$$

We will use the notation $\mathbf{w} := \boldsymbol{\mathcal{P}}_{B_j,\mathrm{it}}\mathbf{w}_{B_j} = \mathbf{w}^{l_j}$, where $\mathbf{w}^k$ has the subvectors $\mathbf{w}^k_{I_j}$ and $\mathbf{w}_{B_j}$, and

$$\mathbf{w}^{k+1}_{I_j} = \mathbf{w}^k_{I_j} - \sigma_{k+1}\boldsymbol{\mathcal{K}}^{-1}_{I_j}(\mathbf{K}_{I_j}\mathbf{w}^k_{I_j} + \mathbf{K}_{I_jB_j}\mathbf{w}_{B_j})\,, \quad \text{with} \quad \mathbf{w}^0_{I_j} = \mathbf{0}\,. \tag{4.52}$$

Let us turn to the model problem formulated in Example 2.2, *i.e.* the heat conduction problem from Example 2.1, but with piecewise constant heat conduction coefficients, and its FE discretization by means of linear triangular elements.

Lemma 4.1. *Let us assume that, for the discretization of the problem in Example 2.2, the generalized quasiuniformity conditions, condition (4.46) and condition (4.47) are fulfilled. Let also*

$$\nu_j \geq (c + \log n_j)/(\log \rho_j^{-1})\,,$$

where $\rho_j = (1-\vartheta_j)/(1+\vartheta_j)$, $\vartheta_j = \sqrt{\underline{\gamma}_j/\overline{\gamma}_j}$, *and* c *is the constant independent of* h. *Then, for any* $l_j \geq \nu_j$, *we have*

$$\|\boldsymbol{\mathcal{P}}_{B_j}\mathbf{v}_{B_j}\|_{\mathbf{K}_j} \leq 2\|\mathbf{v}_{B_j}\|_{\mathbf{S}_{B_j}}\,. \tag{4.53}$$

Proof. The prolongations are performed for each subdomain separately. Therefore, in the proof, we omit indices j and $\varkappa$ for brevity. Let $\boldsymbol{\varphi}_I$ be the solution of the system $\mathbf{K}_I\boldsymbol{\varphi}_I = \mathbf{K}_{I,B}\mathbf{w}_B$ and $\boldsymbol{\varphi}$ has for the subvectors $\boldsymbol{\varphi}_I$ and $\mathbf{w}_B$. We have the convergence estimate

$$\|\mathbf{w}^k_I - \boldsymbol{\varphi}_I\|_{\mathbf{K}_I} \leq \rho^k\|\boldsymbol{\varphi}_I\|_{\mathbf{K}_I}\,,$$

from which it follows

$$\|\mathbf{w}^k - \boldsymbol{\varphi}\|_{\mathbf{K}} \le \rho^k (\|\boldsymbol{\varphi}\|_{\mathbf{K}} + \|\mathbf{w}_B\|_{\mathbf{K}_B}) . \tag{4.54}$$

FE function v_B is nonzero only on one layer of finite elements adjacent to the boundary. In order to estimate the norm of any $\mathbf{v}_B \in V_B$, we take additionally into account the quasiuniformity conditions for the FE assemblage see Section 3.2, and apply Markov's type inequality:

$$\|\mathbf{v}_B\|^2_{\mathbf{K}_B} = \varrho |v_B|^2_{1,\Omega} \le c_1 \varrho [h^{-1} \|v_B\|^2_{0,\partial\Omega} + h |v_B|^2_{1,\partial\Omega}] \le c_2 \varrho h^{-1} \|v_B\|^2_{0,\partial\Omega} , \tag{4.55}$$

here constants depend only on the pointed out quasiuniformity conditions. On the basis of this inequality, we obtain for $\mathbf{w}_B$:

$$\|\mathbf{w}_B\|^2_{\mathbf{K}_B} \le c_2 \varrho h^{-1} \inf_{c \in R} \|v_B - c\|^2_{0,\partial\Omega} \le c_2 \varrho h^{-1} \mathcal{H} \inf_{c \in R} \|v_B - c\|^2_{0,\partial\mathcal{D}} . \tag{4.56}$$

Furthermore, we have

$$\inf_{c \in R} \|v_B - c\|^2_{0,\partial\mathcal{D}} \le \frac{(\operatorname{mes} \partial\mathcal{D})}{2\pi} |v_B|^2_{1/2,\partial\mathcal{D}} = c_3 |v_B|^2_{1/2,\partial\mathcal{D}} \tag{4.57}$$

and

$$\begin{aligned} |v_B|^2_{1/2,\partial\mathcal{D}} &\le c_4 \inf_{v_I} |v_B - v_I|^2_{1,\mathcal{D}} \le \frac{c_5}{\alpha_D^{(1)} \sin\theta_D} \inf_{v_I} |v_B - v_I|^2_{1,\Omega} \\ &= \varrho^{-1} \frac{c_5}{\alpha_D^{(1)} \sin\theta_D} \|\mathbf{v}_B\|^2_{\mathbf{S}_B} . \end{aligned} \tag{4.58}$$

The first inequality is the consequence of inequality

$$\inf_{c \in R} \|v_B - c\|^2_{0,\partial\mathcal{D}} \le \left(\frac{\operatorname{mes} \partial\mathcal{D}}{2\pi} \right)^2 |v_B|^2_{1,\partial\mathcal{D}} , \quad \forall\, v_B \in H^1(\partial\mathcal{D}) , \tag{4.59}$$

from which it is obtained by the K-interpolation. The inequalities of the second line follow from definitions of $|v_B|_{1/2,\partial\mathcal{D}}$ and the Schur complement $\mathbf{S}_B$, the quasiuniformity conditions for the domain decomposition (4.47) and the inequality

$$|v(\mathcal{X}(y))|^2_{1,\mathcal{D}} \le \frac{2}{\alpha_D^{(1)} \sin\theta_D} |v(x)|^2_{1,\Omega} , \quad \forall\, v \in H^1(\Omega) , \tag{4.60}$$

similar to the left inequality (3.13).

Combining (4.56), (4.57) and (4.58), we come to the inequality

$$\|\mathbf{v}_B\|^2_{\mathbf{K}_B} \le c_6 \frac{\mathcal{H}}{h} \|\mathbf{v}_B\|^2_{\mathbf{S}_B} , \quad c_6 = c_2 c_5 / \alpha_D^{(1)} \sin\theta_D , \tag{4.61}$$

which together with (4.54) mean that

$$\|\mathbf{w}^k\|_{\mathbf{K}} \le [1 + \rho^k c_6 \frac{\mathcal{H}}{h}] \, \|\mathbf{w}_B\|_{\mathbf{S}_B} . \tag{4.62}$$

Now it is left to take into account the definition of the prolongation in (4.51), which allows to rewrite (4.62) for $\mathbf{u}^k = \overline{\mathbf{v}} + \mathbf{w}^k$:

$$\|\mathbf{u}^k\|_{\mathbf{K}} \leq [1 + \rho^k c_6 \frac{\mathcal{H}}{h}] \, \|\mathbf{v}_B\|_{\mathbf{S}_B} \,, \tag{4.63}$$

from where Lemma immediately follows with $c = \log c_6$. □

Turning to case A), we look at the elliptic problem (2.4) with $\mathbb{V} = \mathring{H}^1(\Omega)$. Instead of (4.55)–(4.56), we have now

$$\begin{aligned} \|\mathbf{v}_B\|^2_{\mathbf{K}_B} &\leq \mu_2 \| v_B\|^2_{1,\Omega} \leq c_1 \mu_2 [\, h^{-1} \|v_B\|^2_{0,\partial\Omega} + h |v_B|^2_{1,\partial\Omega}] \leq \\ &\leq c_2 \mu_2 h^{-1} \|v_B\|^2_{1/2,\partial\Omega} \leq c_3 \mu_2 h^{-1} \max(1, \mathcal{H}) \, \|v_B\|^2_{1/2,\partial\mathcal{D}} \leq \\ &\leq c_4 \mu_2 h^{-1} \max(1, \mathcal{H}) \inf_{v_I} \|v_B + v_I\|^2_{1,\mathcal{D}} \,, \end{aligned} \tag{4.64}$$

and the constants c_k depend only on the characteristics of the quasiuniformity for finite elements, whereas c_5 depends additionally on $\mathcal{D}$. The first line of these inequalities holds for the reason that FE function v_B is nonzero only on one layer of finite elements adjacent to the boundary. In the next line, the norm $\|v_B\|_{1/2,\partial\Omega}$ defined on the boundary $\partial\Omega$ is transformed to the norm $\|v_B\|_{1/2,\partial\mathcal{D}}$ defined on the boundary $\partial\mathcal{D}$, and then the trace theorem for domain $\mathcal{D}$ is used. By the transformation back to the variables x with the use of similar to (3.8) and (3.12) bounds and the definition of the Schur complement, we get

$$\begin{aligned} \inf_{v_I} \|v_B + v_I\|^2_{1,\mathcal{D}} &\leq \frac{c_5}{\alpha_D^{(1)} \sin\theta_D} \max\left(\frac{1}{\alpha_D^{(1)} \mathcal{H}^2}, 1\right) \inf_{v_I} \|v_B + v_I\|^2_{1,\Omega} \leq \\ &\leq \frac{c_5}{\mu_1 \alpha_D^{(1)} \sin\theta_D} \max\left(\frac{1}{\alpha_D^{(1)} \mathcal{H}^2}, 1\right) \|\mathbf{v}_B\|^2_{\mathbf{S}_B} \,. \end{aligned}$$

When combined with (4.64), this results in the estimate

$$\|\mathbf{v}_B\|^2_{\mathbf{K}_B} \leq c_6^2 \, \Psi \, \|\mathbf{v}_B\|^2_{\mathbf{S}_B} \,, \tag{4.65}$$

with

$$\Psi = \frac{\mu_2 \max(\mathcal{H}, 1)}{\mu_1 h} \, \max\left(\frac{1}{\alpha_D^{(1)} \mathcal{H}^2}, 1\right) \quad \text{and} \quad c_6^2 = \frac{c_4 c_5}{\alpha_D^{(1)} \sin\theta_D} \,.$$

Therefore, we have

$$\nu = \nu_j \geq (c_6 + \log \Psi_j)/(\log \rho_j^{-1}) \,.$$

If $\alpha_D^{(1)}\mathcal{H}_j \geq 1$, *i.e.*, the influence of the minor term in the bilinear form $a_{\Omega_j}(\cdot,\cdot)$ is not significant, this estimate and the estimate of Lemma 4.1 are basically the same.

In definite cases reproducing constant functions by the iterative prolongation operators, defined as in Lemma 4.1, is provided automatically. For instance, it is the case when subdomains of decomposition are the nests of some coarse grid, the subspace $\mathcal{V}_V$ is the coarse grid FE space that exactly represents constants on the subdomains – finite elements of the coarse grid.

4.2.4 *Numerical Complexity of DD Algorithms*

Suppose for simplicity that the domain decomposition is produced by some embedded coarse mesh and the fine and coarse meshes are quasi-uniform. Concerning our model problem from Example 2.2 with piecewise constant coefficient ϱ (constant on each subdomain Ω_j), one surely comes to the optimistic conclusion that, in many cases, optimal subdomain preconditioner-solvers $\boldsymbol{\mathcal{K}}_j$ of several types can be found. In other words, such preconditioner-solvers $\boldsymbol{\mathcal{K}}_j$, for which the first pair of inequalities (4.33) holds with some spectral equivalence constants $\underline{\gamma}_j$ and $\overline{\gamma}_j$, depending only on the constants from the quasiuniformity conditions, and $\operatorname{ops}[\boldsymbol{\mathcal{K}}_j^{-1}\mathbf{v_j}]$ are estimated by $\mathcal{O}(N_j)$. If the prolongation operators $\mathcal{P}_{B_j}$ are also optimal, *i.e.,* satisfy the inequality (4.33) for $\mathcal{P}_{B_j}$ and require each $\mathcal{O}(N_j)$ arithmetic operations, then the optimality of the DD-PCG solver may only be damaged by the Schur complement preconditioner.

Let us assume that the BPS preconditioner (4.17) for the Schur complement, satisfying (4.18), is implemented in the DD preconditioner of the form (4.30). Then $\mathcal{O}(\log H/h)$ PCG iterations are required and on each $\mathcal{O}(N)$ arithmetic operations are performed. Therefore, the total arithmetic complexity of the DD algorithm is of the order $N \log N$.

We can retain the BPS preconditioner for the main solver of the interface problem, but attain optimality of the DD algorithm. It is sufficient to use additionally the preconditioner $\boldsymbol{\mathcal{S}}_B$ that is cheap with respect to the vector-matrix multiplication. For instance, it can be the preconditioner defined by (4.20)–(4.22). Formally, we use the DD preconditioner (4.32) with $\boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1}$ defined by the formula

$$\boldsymbol{\mathcal{S}}_{\mathrm{it},B}^{-1} = [\mathbf{I} - \prod_{k=1}^{n_\epsilon}(\mathbf{I} - \sigma_k \boldsymbol{\mathcal{S}}_{\mathrm{BPS}}^{-1}\boldsymbol{\mathcal{S}}_B)]\,\boldsymbol{\mathcal{S}}_B^{-1}\,, \tag{4.66}$$

where $\boldsymbol{\mathcal{S}}_{\mathrm{BPS}}$ stands for the BPS preconditioner (4.17) and, according to (4.18), (4.23) and Lemma 2.1, n_ϵ must be a fixed integer of the order $\log H/h$ in order to fulfil the first pair of inequalities (4.28). Now, on each DD-PCGM iteration one performs $\mathcal{O}(\log H/h)$ secondary iterations (4.66) with each having the cost $\mathcal{O}((Hh)^{-1}\log H/h)$. Due to (4.28), $\boldsymbol{\mathcal{S}}_{\mathrm{it},B}$ is spectrally equivalent to $\mathbf{S}$. Therefore, the total cost of DD-PCG is $\max\left(\mathcal{O}(N), \mathcal{O}((Hh)^{-1}\log H/h)\right) = \mathcal{O}(N)$.

The DD preconditioner (4.32) with the Schur complement preconditioner (4.27) is also optimal under the above assumptions imposed on the local Dirichlet solvers and prolongation operators. The solution procedure for the interface problem costs more than that with BPS preconditioner, but less than $\mathcal{O}(N)$ and its realization is much simpler.

Chapter 5

BPS-type DD Preconditioners for 3d Elliptic Problems

In the famous series of papers on the domain decomposition methods, [Bramble *et al.* (1986, 1987, 1988, 1989)] presented an efficient and rather general DD preconditioner for h finite element discretizations of 3d elliptic partial differential equations. Since then, these preconditioners are commonly called BPS preconditioners with the abbreviation coming from the first letters in the surnames of the authors. BPS preconditioners have been the origin of the whole family of efficient Dirichlet–Dirichlet type DD preconditioner-solvers for h as well as for hp discretizations of elliptic problems. At first, it was expanded to more general h discretizations and domain decompositions with the state of art in this area is well reflected in the DD books by [Quarteroni and Vali (1999); Smith *et al.* (1996); Toselli and Widlund (2005); and Mathew (2008)], see also the review paper by [Korneev and Langer (2004)].

A vast bibliography of related papers can also be found in these publications. For the developments of fast DD preconditioner-solvers of BPS type for hp discretizations of 3d elliptic problems, apart from the cited books we refer to [Casarin (1997); Korneev *et al.* (2003b); Korneev *et al.* (2003a); Korneev and Rytov (2005a, 2008); and Pavarino and Widlund (1996, 2000a,b)].

As we noted, generalizations of BPS preconditioners were related not only to the types of finite element discretizations, but also to the domain decompositions. [Bramble *et al.* (1989)] assumed that the subdomains were images of the reference unit cube by trilinear mappings, see assumptions A.1–A.4 on p. 4. Therefore, the discretization mesh of each subdomain was assumed to be a rectangular mesh on the reference cube. [Toselli and Widlund (2005)] considered decompositions which are obtained from coarser tetrahedral meshes, see Assumption 4.3 on p. 90, and more general finite element discretization meshes. This required the development of some

new technical tools for the analysis of the interface component of DD algorithms. In this section, we consider a wider in some aspects family of domain decompositions and discretizations to which BPS-type preconditioners can efficiently be applied. We assume that there is a finite number of the *convex reference polyhedrons* $\tau^\circ_\varkappa$, $\varkappa = 1, 2, \ldots, \varkappa_\circ$, which have plain faces, are shape regular in the sense made specific in Subsection 5.1.1, and have diameters equal to unity. This family includes decompositions

$$\overline{\Omega} = \bigcup\nolimits_{j=1}^{J} \overline{\Omega}_j \tag{5.1}$$

of Ω into subdomains Ω_j, which are images of the corresponding reference polyhedrons $\tau^\circ_{\varkappa(j)}$ by the nondegenerate mappings $x = \Upsilon_j(y) : \overline{\tau}^\circ_{\varkappa(j)} \to \overline{\Omega}_j$, $\Upsilon_j(y) \in [L^1_\infty(\overline{\tau}^\circ_{\varkappa(j)})]^3$, see [Sobolev (1991)] for the definition of the L^k_p spaces. These mappings can be characterized by scaling parameters, controlling the changes of lengths and angles and satisfying compatibility conditions for the finite element discretization and the domain decomposition. Since the mappings are continuous, on each boundary $\partial\Omega_j$, we can distinguish faces, edges and vertices, which are uniquely defined as the images of faces, edges and vertices of the polyhedron $\tau^\circ_{\varkappa(j)}$. For simplicity, for these polyhedrons, we use also the notation $\tau_{0,j} = \tau^\circ_{\varkappa(j)}$. Compatibility of the decomposition assumes the property:

A.1. *For each pair $i \neq j$ the intersection $\overline{\Omega}_i \cap \overline{\Omega}_j$ is either empty or common for the pair of subdomains face, or edge or vertex.*

Under these general conditions, finite elements of the FE (finite element) assemblage can be curvilinear. Nevertheless, for the FE mesh, we use the term *triangulation* and assume that the following natural requirement is fulfilled:

A.2. *Let $\mathcal{T}_j$ and $\mathcal{V}(\Omega_j) \subset H^1(\Omega_j)$ be the triangulations of the subdomains Ω_j and the corresponding finite element spaces, and let us assume that they define the triangulation $\mathcal{T}$ of the domain Ω and the finite element space $\mathcal{V}(\Omega) \subset H^1(\Omega)$ in such a way that $\mathcal{T}_j$ and $\mathcal{V}(\Omega_j)$ are the restrictions to Ω_j of $\mathcal{T}$ and $\mathcal{V}(\Omega)$, respectively.*

In Subsection 5.1.1, we will add specific details pertaining to the requirements to the regularity of decompositions, used throughout Chapter 5.

Let us consider the boundary value problem formulated in Example 2.2, the weak formulation of which reads as follows: find $u \in \mathring{H}^1(\Omega)$ satisfying the integral identity

$$a_\Omega(u, v) = (f, v)_\Omega\,, \quad \forall\, v \in \mathring{H}^1(\Omega)\,, \tag{5.2}$$

where, for all w and $v \in H^1(\Omega)$,

$$a_\Omega(w, v) \equiv \int_\Omega \varrho \nabla w \cdot \nabla v dx \quad \text{and} \quad (w, v)_\Omega \equiv \int_\Omega wv dx\,. \tag{5.3}$$

The system of FE (finite element) algebraic equations

$$\mathbf{K}\mathbf{u} = \mathbf{f} \tag{5.4}$$

is equivalent to the integral identity for the FE solution $u_{\rm fem} \in \mathcal{V}(\Omega)$:

$$a_\Omega(u_{\rm fem}, v) = (f, v)_\Omega\,, \quad \forall\, v \in \mathcal{V}(\Omega)\,, \tag{5.5}$$

on the subspace $\mathcal{V}(\Omega) \subset \mathring{H}^1(\Omega)$ of finite element functions. For obtaining an efficient iterative solver for (5.4), it is necessary to create an efficient preconditioner-solver for the FE stiffness matrix $\mathbf{K}$. We will construct DD preconditioner-solver $\boldsymbol{\mathcal{K}}$ with the relative condition number satisfying the bound

$$\mathrm{cond}[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \le c(1 + \log^2 1/\hbar)\,, \tag{5.6}$$

where $\hbar$ is the parameter, characterizing the finite element meshes on the reference subdomains, and c is the constant independent of $\hbar$ and ϱ. This bound is in essential of the same form as in [Bramble *et al.* (1989)] and [Toselli and Widlund (2005)], and it testifies that the preconditioner $\boldsymbol{\mathcal{K}}$ is almost optimal in respect to the relative condition number.

Let us suppose that the system of algebraic equations $\boldsymbol{\mathcal{K}}\mathbf{v} = \boldsymbol{\phi}$ with any vector $\boldsymbol{\phi}$ can be solved for $\mathcal{O}(\hbar^{-3}(1 + \log 1/\hbar)^\gamma)$ arithmetic operations with some γ of the order one. Then, according to (5.6), the system (5.4) can be solved, *e.g.,* by the preconditioned conjugate gradient method, for $\mathcal{O}(\hbar^{-3}(1+\log 1/\hbar)^{\gamma+1})$ arithmetic operations. Therefore, this DD solver has almost optimal/linear numerical complexity. As we will see from the later description of the BPS type preconditioner $\boldsymbol{\mathcal{K}}$, it contains five components or modules, which completely define its arithmetical cost. Among them practically only one, commonly referred as the face component, requires additional optimization in order the above bounds of the arithmetical work were not violated. For the remaining four there are a few suggestions in the literature the use of which do not compromise these bounds. However, we do not elaborate on the complete specifications of all components and concentrate on the analysis of the relative condition number.

The proof of (5.6) is separated in two parts. Formally, the preconditioner, similarly looking to the DD preconditioner, can be defined for the general s.p.d. matrix. The bound for the relative condition number of such preconditioner is conventionally termed the *abstract bound*, since it does not reflect properties of the boundary value problem and its discretization. The derivation of such a bound, convenient for further use, composes the first part of the proof. In the second part, we establish the dependence of

the values, entering the abstract bound on the properties of the boundary value problem, its discretization and decomposition. In a general setting, abstract estimates for the boundaries of the relative spectrum of DD preconditioners are traditionally obtained by means of the space decomposition approach, *i.e.*, as a direct consequence of the space decomposition stability, see for this notion, *e.g.*, [Zhang (1991, 1992)] and the survey paper of [Korneev and Langer (2004)]. It has been developed in numerous papers with the results and bibliography well reflected in the above cited books and papers. We refer here additionally only to the pioneering papers of [Matsokin and Nepomnyaschikh (1985)] and [Lions (1988)]. In this section, we employ an approach, that looks more algebraic, and is based on Theorem 4.2 and Corollaries 4.1 and 4.2. It seems to be a fast way to find out main factors influencing the quality of the DD preconditioner. In second part of the proof, the most involved one is the analysis of consequences of splitting faces in the inter-subdomain boundary Schur complement precondtitioner – which is a module of the DD preconditioner being designed – and of some other structural features of this preconditioner. At that, the difference of our assumptions on the domain decomposition from ones used earlier has most significant implications in the proof. Among some seemingly possible ways of analysis in the second part of the proof, we choose to develop further the approach of [Bramble *et al.* (1989)], adapting it to more general domain decompositions.

The DD preconditioner-solver, we deal with, belongs to the so called Dirichlet–Dirichlet type and possesses several advantages. At each primal DD iteration, it reduces solution of the finite element system of algebraic equations to solution of local problems on subdomains of decomposition, local problems on faces, also local prolongation and restriction operations, and the wire-basket problem, which is the single global problem of the algorithm. Only computational properties of the latter problem are influenced by jumps of the coefficient ϱ across the inter-subdomain boundary, but as a rule it has a much smaller dimension, than the initial finite element problem. DD algorithm, can be treated as having three or four hierarchical stages, depending on the arrangement of the wire basket solver. At the corresponding stages, solution of local problems and local operations of prolongations/restrictions can be performed in parallel.

The chapter is arranged as follows. Section 5.1 is primarily devoted to the description of the structure and main components of DD preconditioner, although in it some components are analyzed alongside with their specification. Our main assumptions, imposed on the type of the domain

decompositions and finite element discretizations under consideration, are formulated in Subsection 5.1.1. Subsection 5.1.2 outlines the general structure of the DD preconditioner, whereas Subsections 5.1.3, 5.1.4 and 5.1.5 present

- preconditioners for solvers of local Dirichlet problems on subdomains and prolongation operators from inter-subdomain boundary,
- arrangement of the preconditioner for the subproblem arising on faces of the inter-subdomain boundary, and
- the preconditioner-solver for the subproblem arising on the wire basket of the decomposition, which is the union of edges and vertices of subdomains of decomposition,

respectively. The abstract bound for the relative condition number is derived in Subsection 5.2.1, while Theorem 5.2, representing the main result of Chapter 5, is proved in Subsection 5.2.2. Subsection 5.2.3 contains some auxiliary results, used in the proof of the main bounds.

5.1 DD Algorithms and their Main Components

5.1.1 *Decomposition and Finite Element Meshes*

We will use (5.2)–(5.3), where $\varrho_j = \varrho|_{\Omega_j}$ can be arbitrary positive numbers, as a model problem for constructing efficient DD preconditioner-solvers, in which only the component of the smallest dimension depends on the jumps of the coefficient ϱ. As usual, we will make no distinction between the physical domain of the problem and the computational domain, occupied by the assemblage of finite elements. Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$ imply that the inter-subdomain boundary, by which we mean the set

$$\Gamma_B = \left(\bigcup\nolimits_{j=1}^{J} \partial\Omega_j\right) \setminus \partial\Omega,$$

fits the inter-element boundaries, and that jumps of the coefficient ϱ are allowed to occur only across the inter-subdomain boundaries. The next assumptions, we are going to formulate, will be on the *quasiuniformity* of the FE discretization and *shape regularity* of the domain decomposition.

In this section, we call a FE discretization quasiuniform with an understanding this term in a slightly weaker than usual sense. Namely, the triangulation of each reference subdomain is assumed to be quasiuniform, with the same for all reference subdomains mesh parameter. At the same

time, the set of mappings $\overline{\tau}_{0,j} \to \overline{\Omega}_j$ is assumed to satisfy the conditions typical for the angular regularity. In the case of a fixed number of subdomains, which is only considered in what follows, the resulting FE discretization in quasiuniform in a commonly used sense. However, if the number of subdomains is sufficiently large, then, in general, the sizes of finite elements can vary considerably.

Let $\mathcal{T}_{0,j}$ and δ^r be the inverse images of the triangulation $\mathcal{T}_j$ of a subdomain Ω_j and of a FE domain τ_r, $r \in \mathbb{R}_j$, respectively, where $\mathbb{R}_j = \{r : \tau_r \subset \Omega_j\}$ is the subset of numbers of finite elements belonging to subdomain Ω_j. In the agreement with the above description, we suppose that each triangulation $\mathcal{T}_{0,j}$ is quasiuniform with the same mesh parameter $\hbar$ for all $j = 1, 2, \ldots, J$, and that the FE domains δ^r have plain faces. Basically this means that, in general, the domains τ_r are associated with the unit tetrahedra or cube τ_0 by nondegenerate linear or trilinear mappings, respectively, and that usual geometric quasiuniformity conditions are fulfilled. Let $\overline{y^{(i)}, y^{(j)}}$ be the edge joining the vertices $y^{(i)}$ and $y^{(j)} \in \overline{\delta}^r$, $\hbar_{i,j}^{(r)}$ be its length, and $\hat{\theta}_{i,j}^{(r)}$ be the angle between the edge $\overline{y^{(i)}, y^{(j)}}$ and the plane, containing the rest edges converging at the vertex $y^{(i)}$. Then for some positive constants $\hat{\alpha}^{(1)}$ and $\hat{\theta}$ the quasiuniformity conditions for the triangulation $\mathcal{T}_{0,j}$ can be written in the form

$$\hat{\alpha}^{(1)}\hbar \le \hbar_{i,j}^{(r)} \le \hbar\,, \qquad \hat{\theta} \le \hat{\theta}_{i,j}^{(r)} \le \pi - \hat{\theta}\,, \tag{5.7}$$

or in the equivalent form

$$0 < c_{\triangle} \le \underline{\rho}_r/\overline{\rho}_r\,, \quad \hat{\alpha}^{(1)}\hbar \le \overline{\rho}_r \le \hbar\,, \tag{5.8}$$

where $\underline{\rho}_r$ and $\overline{\rho}_r$ are the radii of the largest inscribed and the smallest circumscribed spheres for the tetrahedron τ_r. The mesh parameter $\hbar$ is termed also mesh size.

In this chapter, requirements to configurations of subdomains of decomposition and, therefore, conditions imposed upon reference subdomains and their mappings on subdomains of decomposition play a much more substantial role and are more detailed. For this reason, we define some new values and terms related to these matters. For each mapping $x = \Upsilon_j(y)$, let $\mathbf{i}_{j,k}(y)$ be the unit vector in the space of variables x tangent to the line $y_l = \text{const}$, $l \ne k$, and directed towards the growth of y_k. The quantity $\theta_{j,k}(y)$ denotes the angle between $\mathbf{i}_{j,k}(y)$ and the plane containing $\mathbf{i}_{j,l}(y)$, $l \ne k$. Now let $\mathcal{H}_{j,k}$ be Lame's coefficients

$$\mathcal{H}_{j,k} = \left[\sum_{l=1}^{3}\left(\frac{\partial \Upsilon_{j,l}}{\partial y_k}\right)^2\right]^{1/2}. \tag{5.9}$$

Shape regularity or equally shape quasiuniformity of a domain decomposition implies existence of such positive constants $\underline{\alpha}_D, \underline{\theta}_D$ and $H_{D,j}$, $j = 1, 2, \ldots, J$, that, for each $x \in \overline{\tau}_{0,J}$,

$$\underline{\alpha}_D \, H_{D,j} \le \mathcal{H}_{j,k} \le H_{D,j} \quad \text{and} \quad \underline{\theta}_D \le \theta_{j,k} \le \pi - \underline{\theta}_D \,, \tag{5.10}$$

with positive constants $\underline{\alpha}_D$ and $\underline{\theta}_D$. These inequalities will be referred as the shape regularity conditions for mappings $x = \Upsilon_j(y)$. It additionally implies that reference polyhedrons $\tau_{0,j}$ satisfy some conditions of shape and size regularity , also called quasiuniformity conditions of reference subdomains. Let $\underline{\wp}_j$, $\underline{\wp}_j^{(k)}$ and $\overline{\wp}_j$, $\overline{\wp}_j^{(k)}$ be the radii of the largest inscribed and the smallest circumscribed spheres for the reference polyhedron $\tau_{0,j}$ and its faces $\hat{F}_j^{(k)}$, respectively. Let $\hat{l}_j^{(k)}$ be the lengths of edges of $\tau_{0,j}$. Then a part of these regularity conditions is expressed by the inequalities

$$0 < c_\circ \le \underline{\wp}_j / \overline{\wp}_j \,, \quad \overline{\wp}_j = 1 \,, \tag{5.11}$$

$$0 < c_\circ \le \underline{\wp}_j^{(k)} / \overline{\wp}_j^{(k)} \,, \quad 0 < c_\circ \le \overline{\wp}_j^{(k)}, \hat{l}_j^{(k)} \le 1 \,. \tag{5.12}$$

Indeed, if reference subdomains are tetrahedrons, then (5.11) are sufficient to provide all needed properties of the decompositions and, in particular, quasiuniformity of faces, *i.e.,* (5.12), and angles. In a more general case, additional conditions should be imposed on the angles, namely, on the angles between adjacent faces and between adjacent edges of the faces of polyhedrons $\tau_{0,j}$. These angles should be separated from zero and from 2π by some constant $\hat{\theta}_D > 0$.

For obtaining some bounds in Subsection 5.2.1, we use one property, which apparently is a consequence of the shape regularity of the domain decomposition and quasiuniformity of the reference triangulations. According to it, each triangulation $\mathcal{T}_{0,j}$ can be expanded on c-vicinity of $\tau_{0,j}$, $c = \text{const}$, in such a way that the formulated above conditions of quasiuniformity remain fulfilled, possibly with new constants $\hat{\alpha}^{(1)}$, $\hat{\theta}$, $c_\triangle$, for which, however, we retain the same notations. We do not prove this property under conditions, imposed above, and take it for an additional assumption.

The additional requirements, introduced above for the domain decomposition and the finite element discretizations, we summarize in the form of the assumptions given below:

$\mathcal{A}.3$. *The domain decomposition is shape quasiuniform, i.e., conditions (5.10), (5.11), (5.12) and the condition on the angles are fulfilled.*

$\mathcal{A}.4$. *Triangulations $\mathcal{T}_{0,j}$ of the reference subdomains have the same mesh size $\hbar$ and satisfy quasiuniformity conditions (5.7),(5.8).*

A.5. Each triangulation $\mathcal{T}_{0,j}$ can be expanded on c-vicinity of $\tau_{0,j}$, $c = \text{const} \geq 1$ in such a way that quasiuniformity of the triangulation will be retained.

The description of the BPS preconditioner, presented in this chapter, practically does not depend on the reference element, used in the FE discretization. The proof of the estimate for the relative condition number in the frame of h-version is also almost not affected by the type of the reference element. However, for the sake of simplicity, we assume that finite elements are tetrahedrons with linear coordinate functions.

5.1.2 *Structure of DD Preconditioners*

Orderings of d.o.f. (degrees of freedom), adapted to the substructuring algorithms and to the corresponding Dirichlet–Dirichlet-type DD algorithms, are reflected in the representations of the FE stiffness matrices in the block forms

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,B} \\ \mathbf{K}_{B,I} & \mathbf{K}_B \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,F} & \mathbf{K}_{I,W} \\ \mathbf{K}_{F,I} & \mathbf{K}_F & \mathbf{K}_{F,W} \\ \mathbf{K}_{W,I} & \mathbf{K}_{W,F} & \mathbf{K}_W \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,F} & \mathbf{K}_{I,E} & \mathbf{K}_{I,V} \\ \mathbf{K}_{F,I} & \mathbf{K}_F & \mathbf{K}_{F,E} & \mathbf{K}_{F,V} \\ \mathbf{K}_{E,I} & \mathbf{K}_{E,F} & \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,I} & \mathbf{K}_{V,F} & \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix}, \quad (5.13)$$

where lower indices mark the sets of d.o.f. living on

- I – interiors of subdomains Ω_j,
- B – the inter-subdomain boundary,
- F – interiors of the faces of subdomains,
- W – the wire basket of DD mesh,
- E – interiors of the edges, and
- V – vertices of subdomains of decomposition.

The wire basket of the DD mesh is nothing else but union of edges and vertices of subdomains of the domain decomposition. In the sets of vertices, edges and faces are included only those, which are not on the part of the boundary, where the Dirichlet boundary condition is imposed. If natural boundary conditions are prescribed on a part of $\partial\Omega$, then this part is usually included in the interface boundary. At the same time, it is necessary to note that, in the analysis of DD algorithms, it is sometimes convenient to include all vertices, edges and faces on $\partial\Omega$ into consideration. In what follows, it is done without special remarks. As usual, additional indexation will be used

reflecting different ways of ordering of elements of the introduced sets. For instance, F_j and F_j^k will denote the union of faces of subdomain Ω_j and its separate faces, respectively. At the same time separate faces ordered by some global ordering will be denoted by F^l. Similar indices will be used for the sets related to edges.

According to the given above block forms of the matrix $\mathbf{K}$, the vector space of FE d.o.f., which is denoted by V, can be decomposed into the direct sums of corresponding subspaces

$$V = V_I \oplus V_B\,, \quad V = V_I \oplus V_F \oplus V_W\,, \quad V = V_I \oplus V_F \oplus V_E \oplus V_V\,,$$

where $V_B = V_F \oplus V_E \oplus V_V$ and $V_W = V_E \oplus V_V$ are the subspaces of the inter-subdomain boundary and wire basket d.o.f., respectively. In the notations of the corresponding subspaces of FE functions, we replace V by $\mathcal{V}$ so that, *e.g.*, $V_E \leftrightarrow \mathcal{V}_E(\Omega)$, whereas $\mathcal{V}_E(E)$ is the subspace of traces on E. Corresponding subspaces for a particular subdomain Ω_j are supplied with additional indices j. In agreement with this $\mathcal{V}_F(\Omega_j)$ is the subspace of face functions from $\mathcal{V}(\Omega_j)$ and $\mathcal{V}_F(F_j)$ is the space of their traces on F_j. For the FE spaces on the pre-image domain $\tau_{0,j}$, on the union $\hat{F}_j$ of its faces, and on the wire basket $\hat{W}_j$, similar notations are used with $\mathbb{V}$ standing instead of $\mathcal{V}$. Hence, $\mathbb{V}_E(\tau_{0,j})$ and $\mathbb{V}_E(E_j)$ are the subspace of edge functions from $\mathbb{V}(\tau_{0,j})$ and the space of their traces on $\hat{E}_j$.

Let us turn to the 3×3 block form (5.13) of the matrix $\mathbf{K}$ and vectors $\mathbf{v} \in V = V_I \oplus V_F \oplus V_W$. The inverse to the DD preconditioner-solver $\boldsymbol{\mathcal{K}}$ for the FE stiffness matrix $\mathbf{K}$ can be expressed by the general formulas

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \mathbf{P}_{V_B \to V} \boldsymbol{\mathcal{S}}_B^{-1} \mathbf{P}_{V_B \to V}^{\top}\,, \tag{5.14}$$

$$\boldsymbol{\mathcal{S}}_B^{-1} = \boldsymbol{\mathcal{S}}_F^{+} + \mathbf{P}_{V_W \to V_B} (\boldsymbol{\mathcal{S}}_W^F)^{-1} \mathbf{P}_{V_W \to V_B}^{\top}\,. \tag{5.15}$$

The s.p.d. matrices $\boldsymbol{\mathcal{K}}_I, \boldsymbol{\mathcal{S}}_F$ and $\boldsymbol{\mathcal{S}}_W^F$ are defined on the subspaces V_I, V_F and V_W of d.o.f. living on interior parts of subdomains of domain decomposition, their faces and on the wire basket, respectively. Indeed, these matrices can be defined implicitly, and, in what follows, we need only to define procedures realizing multiplications of vectors by $\boldsymbol{\mathcal{K}}_I^{+}$, by the pseudo-inverse $\boldsymbol{\mathcal{S}}_F^{+}$ of $\boldsymbol{\mathcal{S}}_F$, continued by zero entries up to the quadratic form on $V_B \times V_B$, and by $(\boldsymbol{\mathcal{S}}_W^F)^{-1}$. The rectangular matrices

$$\mathbf{P}_{V_B \to V} : V_B \to V \quad \text{and} \quad \mathbf{P}_{V_W \to V_B} : V_W \to V_B$$

realize the prolongation operations.

The structure of the preconditioning operation, which is the procedure designated as multiplication by $\boldsymbol{\mathcal{K}}^{-1}$, mimics the multiplication by $\mathbf{K}^{-1}$ in

the implementation of the three stage block Gauss elimination procedure according to the substructuring approach. In this procedure, the d.o.f. internal for subdomains of decomposition are eliminated at the first stage, d.o.f. living on faces – at the second stage, and d.o.f. of the wire basket – at the third stage. In order to become certain of this, let us look at the factorization

$$\mathbf{K} = \mathbf{C}_I \begin{pmatrix} \mathbf{K}_I & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B \end{pmatrix} \mathbf{C}_I^\top , \quad \text{with } \mathbf{C}_I = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}_{BI}\mathbf{K}_I^{-1} & \mathbf{I} \end{pmatrix} , \tag{5.16}$$

where $\mathbf{S}_B$ is also factored into the product of three matrices

$$\mathbf{S}_B = \begin{pmatrix} \mathbf{S}_F & \mathbf{S}_{FW} \\ \mathbf{S}_{WF} & \mathbf{S}_W \end{pmatrix} = \mathbf{C}_B \begin{pmatrix} \mathbf{S}_F & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_W^F \end{pmatrix} \mathbf{C}_B^\top, \ \ \mathbf{C}_B = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{S}_{WF}\mathbf{S}_F^{-1} & \mathbf{I} \end{pmatrix}, \tag{5.17}$$

and where, clearly, $\mathbf{S}_B$ and $\mathbf{S}_W^F$ are the Schur complements

$$\mathbf{S}_B = \mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB} \quad \text{and} \quad \mathbf{S}_W^F = \mathbf{S}_W - \mathbf{S}_{WF}\mathbf{S}_F^{-1}\mathbf{S}_{FW} . \tag{5.18}$$

The following factorization of $\mathbf{K}^{-1}$ now immediately follows from (5.16)–(5.18):

$$\mathbf{K}^{-1} = (\mathbf{C}_I^\top)^{-1} \begin{pmatrix} \mathbf{K}_I^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B^{-1} \end{pmatrix} \mathbf{C}_I^{-1}, \text{ with } \mathbf{C}_I^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{K}_{BI}\mathbf{K}_I^{-1} & \mathbf{I} \end{pmatrix}, \tag{5.19}$$

$$\mathbf{S}_B^{-1} = (\mathbf{C}_B^\top)^{-1} \begin{pmatrix} \mathbf{S}_F^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{S}_W^F)^{-1} \end{pmatrix} \mathbf{C}_B^{-1}, \text{ with } \mathbf{C}_B^{-1} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{S}_{WF}\mathbf{S}_F^{-1} & \mathbf{I} \end{pmatrix} . \tag{5.20}$$

If we introduce the prolongation matrices

$$\mathbf{P}_{B,\text{exact}} = \begin{pmatrix} -\mathbf{K}_I^{-1}\mathbf{K}_{IB} \\ \mathbf{I} \end{pmatrix} \quad \text{and} \quad \mathbf{P}_{W,\text{exact}} = \begin{pmatrix} -\mathbf{S}_F^{-1}\mathbf{S}_{FW} \\ \mathbf{I} \end{pmatrix}, \tag{5.21}$$

based on exact solution procedures, then $\mathbf{K}^{-1}$ can be defined by the equivalent to (5.19)–(5.20) expressions

$$\mathbf{K}^{-1} = \mathbf{K}_I^+ + \mathbf{P}_{B,\text{exact}}\mathbf{S}_B^{-1}\mathbf{P}_{B,\text{exact}}^\top , \tag{5.22}$$

$$\mathbf{S}_B^{-1} = \mathbf{S}_F^+ + \mathbf{P}_{W,\text{exact}}(\mathbf{S}_W^F)^{-1}\mathbf{P}_{W,\text{exact}}^\top . \tag{5.23}$$

Therefore, formally, (5.14)–(5.15) is the result of the replacement of the s.p.d. matrices $\mathbf{K}_I, \mathbf{S}_F, \mathbf{S}_W^F$ by their preconditioners and the replacement of the rectangular prolongation matrices $\mathbf{P}_{B,\text{exact}}$ and $\mathbf{P}_{W,\text{exact}}$ by their appropriate approximations $\mathbf{P}_{V_B \to V}$ and $\mathbf{P}_{V_W \to V_B}$ in (5.22)–(5.23).

The DD preconditioner-solver will be designed in such a way that the procedure of the multiplication by $\mathcal{K}^{-1}$ of any vector

$$\mathbf{v} = \begin{pmatrix} \mathbf{v}_I \\ \mathbf{v}_F \\ \mathbf{v}_W \end{pmatrix}$$

will assume the sequence of operations

$$\boldsymbol{\mathcal{K}}^{-1}\mathbf{v} = \{\boldsymbol{\mathcal{K}}_I^{+} + \mathbf{P}_{V_B\to V}\,[\,\boldsymbol{\mathcal{S}}_B^{-1} = \boldsymbol{\mathcal{S}}_F^{+} + \mathbf{P}_{V_W\to V_B}\boldsymbol{\mathcal{S}}_W^{-1}\mathbf{P}_{V_W\to V_B}^{\top}\,]\,\mathbf{P}_{V_B\to V}^{\top}\}\,\mathbf{v}\,, \tag{5.24}$$

performed consequently. Let *s.-wise* means *subdomainwise*, *f.-wise* means *facewise*, *ass.* means *vector assembling*, which assumes that, after completing local operations, the obtained vectors are assembled in the vector of a larger dimension. Then the preconditioning operation (5.24) yields the preconditioning procedure presented in Algorithm 5.1. In step 2), $\mathbf{v}_F^{(1)}$ is

Algorithm 5.1 Preconditioning operation: $\mathbf{w} = \boldsymbol{\mathcal{K}}^{-1}\mathbf{v}$.

1) $\mathbf{v}_I^{(1)} = \boldsymbol{\mathcal{K}}_I^{+}\mathbf{v}_I \;\Leftarrow$ s-wise ; f.-wise + ass. $\Rightarrow\; \mathbf{v}_B^{(1)} = \mathbf{P}_{V_B\to V}^{\top}\mathbf{v}$;
2) f.-wise $\Rightarrow\; \mathbf{v}_F^{(2)} = \boldsymbol{\mathcal{S}}_F^{-1}\mathbf{v}_F^{(1)}$; f.-wise + ass. $\Rightarrow\; \mathbf{v}_W^{(2)} = \mathbf{P}_{V_W\to V_B}^{\top}\mathbf{v}_B^{(1)}$;
3) $\mathbf{v}_W^{(3)} = \boldsymbol{\mathcal{S}}_W^{-1}\mathbf{v}_W^{(2)}$;
4) $\mathbf{v}_B^{(4)} = \mathbf{P}_{V_W\to V_B}\mathbf{v}_W^{(3)} \;\Leftarrow$ f.-wise ;
5) $\mathbf{v}_B^{(5)} = \mathbf{v}_F^{(2)} + \mathbf{v}_B^{(4)} \;\Leftarrow$ f.-wise ;
6) $\mathbf{v}^{(6)} = \mathbf{P}_{V_B\to V}\mathbf{v}_B^{(5)} \;\Leftarrow$ subdomainwise ;
7) $\mathbf{w} = \mathbf{v}_I^{(1)} + \mathbf{v}^{(6)} \;\Leftarrow$ subdomainwise .

a subvector of $\mathbf{v}_B^{(1)}$, sums of vectors in steps 5) and 7) are understood as topological sums, *i.e.*, $\mathbf{v}_F^{(2)}$ and $\mathbf{v}_I^{(1)}$ are considered as continued by zero entries on all nodes participating in the respective sum.

5.1.3 *Local Dirichlet Problems and Prolongation*

5.1.3.1 *Local Dirichlet Problems*

In view of the assumption $\mathcal{A}.1$ and the accepted ordering of d.o.f., the block $\mathbf{K}_I = \mathrm{diag}\,[\mathbf{K}_{I_j}]_{j=1}^{J}$ is the block diagonal matrix where each block $\mathbf{K}_{I_j}$ is related to the internal unknowns of the corresponding subdomain Ω_j. Naturally, the preconditioner-solver for the local Dirichlet problems has the same typical block-diagonal structure, so that

$$\boldsymbol{\mathcal{K}}_I^{+} := \begin{pmatrix} \boldsymbol{\mathcal{K}}_I^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \quad \text{with} \quad \boldsymbol{\mathcal{K}}_I = \mathrm{diag}\,[\,\boldsymbol{\mathcal{K}}_{I_1}, \boldsymbol{\mathcal{K}}_{I_2}, \ldots, \boldsymbol{\mathcal{K}}_{I_J}\,]\,, \tag{5.25}$$

and it is assumed that

$$\underline{\gamma}_{I_j}\boldsymbol{\mathcal{K}}_{I_j} \prec \mathbf{K}_{I_j} \prec \overline{\gamma}_{I_j}\boldsymbol{\mathcal{K}}_{I_j}\,, \quad 0 < \underline{\gamma}_I \le \underline{\gamma}_{I_j}\,, \quad \overline{\gamma}_{I_j} \le \overline{\gamma}_I\,, \tag{5.26}$$

for some $\underline{\gamma}_I$ and $\overline{\gamma}_I$ independent of $j \in \{1, \ldots, J\}$.

The preconditioners involved in DD algorithm should satisfy two contradictory requirements, which have the following form for $\boldsymbol{\mathcal{K}}_{I_j}$:

α) The values of $\underline{\gamma}_I$ and $\overline{\gamma}_I$ should be as close as possible to unity, uniformly in j.
β) There exist sufficiently fast solvers for systems of algebraic equations with the system matrices $\boldsymbol{\mathcal{K}}_{I_j}$.

In general, when we have no fast exact solver for the system with the matrix $\mathbf{K}_{I_j}$ and it is not reasonable to take $\boldsymbol{\mathcal{K}}_{I_j} = \mathbf{K}_{I_j}$, there indeed exists a number of options for defining subdomain preconditioners $\boldsymbol{\mathcal{K}}_{I_j}$ efficient in a sense of the requirements α) and β). The situation is one of the simplest, when reference subdomains are the unit cube, *i.e.*, $\tau_{0,j} \equiv \tau_0$, triangulated by the uniform rectangular mesh of sizes $\hbar_1 \times \hbar_2 \times \hbar_3$. The preconditioner $\boldsymbol{\mathcal{K}}_{I_j}$ can be defined as the matrix of the FD approximation of the operator $-\varrho_j \Delta v$ in τ_0 with homogeneous Dirichlet boundary conditons $v = 0$ on $\partial\tau_0$, multiplied by $\hbar_1\hbar_2\hbar_3$. In this case, $\underline{\gamma}_I$ and $\overline{\gamma}_I$ are constants depending only on the shape regularity conditions for the mappings Υ_j. Then systems with this preconditioner $\boldsymbol{\mathcal{K}}_{I_j}$ can be solved by fast direct methods with at least almost optimal complexity, see, *e.g.*, [Samarskii and Nikolayev (1989)]. Different multigrid and multilevel preconditioners for FE stiffness matrices have been developed, which can be efficient at the use as $\boldsymbol{\mathcal{K}}_{I_j}$ for rather general decompositions and triangulations. We refer here only to the early paper on BPX preconditioners of [Bramble *et al.* (1990)] and to the books on multigrid and other preconditioning methods by [Bramble and Zhang (2000); Hackbusch (1985); Kraus and Margenov (2009); and Vassilevski (2008)] as well as to the survey paper by [Hackbusch (2004)]. The preconditioner $\boldsymbol{\mathcal{K}}_{I_j}$ can be defined also implicitly by means of some inexact iterative solver for the SLAE with the matrix $\mathbf{K}_{I_j}$ according to the scheme, presented in Subsection 2.2.3, or some other scheme. For example, one multigrid iteration can produce very efficient local preconditioner, see [Jung and Langer (1991); and Jung *et al.* (1989)].

5.1.3.2 *Prolongation from the inter-subdomain boundaries*

Obviously, the matrices $\mathbf{P}_{B,\mathrm{exact}} : V_B \to V$ and $\mathbf{P}_{B_j,\mathrm{exact}} : V_{B_j} \to V_j$ of the subdomainwise discrete harmonic global and local discrete harmonic prolongations satisfy the equalities

$$\|\mathbf{P}_{B,\mathrm{exact}}\mathbf{v}_B\|_{\mathbf{K}} = \|\mathbf{v}_B\|_{\mathbf{S}_B} \quad \text{and} \quad \|\mathbf{P}_{B_j,\mathrm{exact}}\mathbf{v}_{B_j}\|_{\mathbf{K_j}} = \|\mathbf{v}_{B_j}\|_{\mathbf{S}_{B_j}}$$

for all $\mathbf{v}_B \in \mathbf{V}_B$ and $\mathbf{v}_{B_j} \in \mathbf{V}_{B_j}$, respectively. In Subsection 5.2.1, we prove that in order to obtain DD preconditioners which yield good relative condition numbers it is sufficient to implement prolongations satisfying

$$\|\mathbf{P}_{V_B \to V}\mathbf{v}_B\|_{\mathbf{K}} \le c_{P_B}\|\mathbf{v}_B\|_{\mathbf{S}_B}\,, \quad \forall\, \mathbf{v}_B \in \mathbf{V}_B\,, \tag{5.27}$$

with some good (*e.g.*, close to 1) constant c_{P_B}.

The global prolongation operator is defined on the basis of local prolongation operators $\mathbf{P}_{V_{B_j} \to V_j}$ in such a way that the restriction of $\mathbf{P}_{V_B \to V}$ to any subdomain $\overline{\Omega}_j$ is $\mathbf{P}_{V_{B_j} \to V_j}$. Therefore, for establishing (5.27), it is sufficient that

$$\|\mathbf{P}_{V_{B_j} \to V_j}\mathbf{v}_{B_j}\|_{\mathbf{K}_j} \le c_{P_B}\|\mathbf{v}_{B_j}\|_{\mathbf{S}_{B_j}}\,, \quad \forall\, \mathbf{v}_{B_j} \in \mathbf{V}_{B_j}\,, \tag{5.28}$$

uniformly for all subdomains.

Let $\mathcal{P}_{B_j} : \mathcal{V}_B(\partial\Omega_j) \to \mathcal{V}(\Omega_j)$ be the prolongation operator, whose matrix representation is $\mathbf{P}_{V_{B_j} \to V_j}$. Under the stated conditions imposed on the function ϱ, the reference domains, their triangulations, and the FE assemblage, inequality (5.28) is equivalent to

$$|\mathcal{P}_{B_j} v_{B_j}|_{1,\Omega_j} \le c_{\mathcal{P}_B}\, |v_{B_j}|_{1/2,\partial\Omega_j}\,, \quad \forall\; v_{B_j} \in \mathcal{V}_B(\partial\Omega_j), \tag{5.29}$$

with some constant $c_{\mathcal{P}_B}$ different from c_{P_B}, if

$$\varrho_j |v_{B_j}|^2_{1/2,\partial\Omega_j} \ne \|\mathbf{v}_{B_j}\|^2_{\mathbf{S}_{B_j}}\,, \quad \forall v_{B_j} \leftrightarrow \mathbf{v}_{B_j} \in \mathbf{V}_{B_j}\,.$$

Inequality (5.29) immediately follows from (5.28), the definition of the local FE matrices $\mathbf{K}_j$, and the right inequality in Lemma 5.1 given below.

Lemma 5.1. *Let assumptions $\mathcal{A}_1$ – $\mathcal{A}_4$ be fulfilled. Then, for positive constants $\underline{\gamma}_S$ and $\overline{\gamma}_S$ depending only on $\underline{\alpha}_D^{(1)}$, $\underline{\theta}_D$ and on the reference subdomains $\tau^{\circ}_{\varkappa}$, $\varkappa = 1, 2, \ldots, \varkappa_{\circ}$, the inqualities*

$$\underline{\gamma}_S \varrho_j\, |v_{B_j}|^2_{1/2,\partial\Omega_j} \le \|\mathbf{v}_{B_j}\|^2_{\mathbf{S}_{B_j}} \le \varrho_j \overline{\gamma}_S\, |v_{B_j}|^2_{1/2,\partial\Omega_j} \tag{5.30}$$

hold for all $\mathbf{v}_{B_j} \leftrightarrow v_{B_j} \in \mathcal{V}_B(\partial\Omega_j)$.

Proof. According to our assumptions imposed on the reference subdomains, for any $v_{B_j} \in H^{1/2}(\partial\tau_{0,j})$, the function $u \in H^1(\tau_{0,j})$, defined by

$$|\, u\,|^2_{1,\tau_{0,j}} = \inf_{v \in H^1(\tau_{0,j}):\, v = v_{B_j} \text{ on } \tau_{0,j}} |\, v\,|^2_{1,\tau_{0,j}}\,,$$

exists and satisfies the inequalities

$$\underline{\beta}\,|\, v_{B_j}\,|^2_{1/2,\partial\tau_{0,j}} \le |\, u\,|^2_{1,\tau_{0,j}} \le \overline{\beta}\,|\, v_{B_j}\,|^2_{1/2,\partial\tau_{0,j}} \tag{5.31}$$

for all $j = 1, 2, \ldots, J$, with some positive constants $\underline{\beta}$ and $\overline{\beta}$. From (5.31), we come to its discrete counterpart: for $\forall\, v_{B_j} \in \mathcal{V}_{B_j}(\partial\tau_{0,j})$ and v, which minimizes $|\, v\,|_{1,\tau_{0,j}}$ among all $v \in \mathcal{V}(\tau_{0,j})$ coinciding with v_{B_j} on $\partial\tau_{0,j}$, we have

$$\underline{\beta}\,|\, v_{B_j}\,|^2_{1/2,\partial\tau_{0,j}} \le \inf_{v\in\mathcal{V}(\tau_{0,j})\colon v=v_{B_j}\ \text{on}\ \partial\tau_{0,j}} |\, v\,|^2_{1,\tau_{0,j}} \le \hat{\beta}\,|\, v_{B_j}\,|^2_{1/2,\partial\tau_{0,j}}. \quad (5.32)$$

Indeed, the left inequality (5.32) is a particular case of the left inequality (5.31). In order to get the right inequality (5.32), it is sufficient to use a suitable quasi-interpolation operator. We pick up the quasi-interpolation operator $\mathcal{I}_h$, suggested by [Scott and Zhang (1990)], the properties of which are summarized in Lemma 5.2 presented after the proof. Now, for an arbitrary $v_{B_j} \in \mathcal{V}_B(\partial\tau_{0,j})$, we consider the corresponding function u satisfying (5.31), its interpolation $\mathcal{I}_h u$ and v which is equal to v_{B_j} on the boundary $\partial\tau_{0,j}$ and minimize $|\, v\,|_{1,\tau_{0,j}}$. The right inequality (5.32) with $\hat{\beta} = \overline{\beta} c_{\rm int}$ follows by taking into account properties b) and d), formulated in Lemma 5.2, and (5.31). Indeed, we have

$$\inf_{v\in\mathcal{V}(\tau_{0,j})\colon v=v_{B_j}\ \text{on}\ \partial\tau_{0,j}} |\, v\,|^2_{1,\tau_{0,j}} \le |\, \mathcal{I}_h u\,|^2_{1,\tau_{0,j}} \le$$
$$c_{\rm int}\,|\, u\,|^2_{1,\tau_{0,j}} \le \overline{\beta}\, c_{\rm int}\,|\, u\,|^2_{1/2,\partial\tau_{0,j}}$$

where $|\, u\,|_{1/2,\partial\tau_{0,j}} = |\, v_{B_j}\,|_{1/2,\partial\tau_{0,j}}$. In order to complete the proof, we transform (5.32) to (5.30) by means of the mapping $x = \Upsilon_j(y) : \overline{\tau}_{0,j} \to \overline{\Omega}_j$, and use (5.10). □

Now we are going to describe the quasi-interpolation operator of [Scott and Zhang (1990)]. Let $\Omega \subset R^n$ be the n-dimensional bounded Lipschitz domain, and let $\mathcal{T}_h$ be an arbitrary quasiuniform triangulation with the nodal points $x^{(i)}$, $i = 1, 2, \ldots, I$, and maximal edge size h. To each node $x^{(i)}$, we relate the $(n-1)$-dimensional simplex $\tau_i^{(n-1)}$, which is the face of one of the n-dimensional simplexes of the triangulation $\mathcal{T}_h$ having the vertex $x^{(i)}$. For n vertices of the simplex $\tau_i^{(n-1)}$, we use also notations $z_l^{(i)}$, $l = 1, 2, \ldots, n$, assuming for definiteness that $z_1^{(i)} = x^{(i)}$. The choice of $\tau_i^{(n-1)}$ is not unique, but, for $x^{(i)} \in \partial\Omega$, we always take $\tau_i^{(n-1)} \subset \partial\Omega$. By $\mathcal{V}_\triangle(\Omega)$ and $\mathcal{V}_{\rm tr}(\partial\Omega)$ are denoted the space of functions, which are continuous on $\overline{\Omega}$ and linear on each simplex of the triangulation, and the space of their traces on $\partial\Omega$, respectively. Let $\theta_i \in \mathcal{P}_1(\tau_i^{(n-1)})$ be the function, satisfying

$$\int_{\tau_i^{(n-1)}} \theta_i \lambda_l^{(i)}\, dx = \delta_{1,l}\,, \quad l = 1, 2, \ldots, n\,, \quad (5.33)$$

where $\lambda_l^{(i)}$ are the barycentric coordinates in $\tau_i^{(n-1)}$ related to its vertices $z_l^{(i)}$ and $\delta_{i,l}$ is Kronecker's symbol. If $\phi_i \in \mathcal{V}_\triangle(\Omega)$ are the Galerkin FE basis functions such that $\phi_i(x_j) = \delta_{i,j},\ i,j = 1,2,..,I,$ then for each $v \in H^1(\Omega)$ the quasi-interpolation $\mathcal{I}_h v \in \mathcal{V}_\triangle(\Omega)$ is defined as

$$\mathcal{I}_h v = \sum_{i=1}^{I} \left(\int_{\tau_i^{(n-1)}} \theta_i v \, dx \right) \phi_i(x) \,. \tag{5.34}$$

Lemma 5.2. *The quasi-interpolation operator $\mathcal{I}_h$ satisfies*

a) $\mathcal{I}_h v : H^1(\Omega) \mapsto \mathcal{V}_\triangle(\Omega)$, *and, if* $v \in \mathcal{V}_\triangle(\Omega)$, *then* $\mathcal{I}_h v = v$,

b) $(v - \mathcal{I}_h v) \in \overset{\circ}{H}{}^1(\Omega)$, *if* $v|_{\partial\Omega} \in \mathcal{V}_{\rm tr}(\partial\Omega)$,

c) $\|v - \mathcal{I}_h v\|_{t,\Omega} \le c_{\rm int} h^{s-t} \|v\|_{s,\Omega}$ *for* $t = 0,1,$ *and* $s = 1,2,$

d) $|\mathcal{I}_h v|_{1,\Omega} \le c_{\rm int} |\, v\,|_{1,\Omega}$ *and* $\|\mathcal{I}_h v\|_{1,\Omega} \le c_{\rm int} \|\, v\,\|_{1,\Omega}$ *for all* $v \in H^1(\Omega)$, *where* $c_{\rm int}$ *are generic positive constants, depending only on* $\hat\alpha^{(1)}, \hat\theta$ *from the quasiuniformity conditions (5.7). Clearly the operator* $\mathcal{I}_h$ *is a projection onto the space* $v \in \mathcal{V}_\triangle(\Omega)$.

Proof. The proof of this lemma was given by [Scott and Zhang (1990)]. In the above form, the lemma can be found [Xu and Zou (1998)]. □

In the literature, there are suggestions on fast prolongation operators in 3d, which provide inequalities (5.28), (5.29) with good constant c_{P_B} and have linear or almost linear computational complexity, but mostly they are related to the subdomains of a canonical geometrical form or their images.

Below, we formulate the result on the efficiency of prolongations from the inter-subdomain boundary upon the whole domain by means of inexact two-stage solvers for local Dirichlet problems on subdomains Ω_j.

They have at least two advantages against others: the same solvers can be used in two main components of DD algorithm and they are applicable for rather general subdomains of decomposition and FE meshes.

Suppose, $\boldsymbol{\mathcal{K}}_I = \mathrm{diag}\,[\,\boldsymbol{\mathcal{K}}_{I_1}, \boldsymbol{\mathcal{K}}_{I_2}, \ldots, \boldsymbol{\mathcal{K}}_{I_J}\,]$ is the preconditioner-solver, for which (5.26) hold with good $\underline{\gamma}_I$ and $\overline{\gamma}_I$, and solving SLAE with matrices $\boldsymbol{\mathcal{K}}_{I_j}$ is relatively cheap. In this case, a natural way of defining the prolongation operator $\mathbf{P}_{V_B \to V} : V_B \to V$ is by means of the inexact solver with the preconditioner $\boldsymbol{\mathcal{K}}_{I_j}$ for each local discrete Dirichlet problem governed by the matrix $\mathbf{K}_{I_j}$. In order to define such a prolongation, we need some new notations.

We introduce vectors $\mathbf{1}_{B_j} \in V_{B_j}$ and $\mathbf{1}_j \in V_j$, with all entries equal to the unity and the mass matrix $\mathbf{M}_{B_j}$ induced by $\|v_{B_j}\|^2_{0,\partial\tau_{0,j}}$ and the space

of FE traces $\mathbb{V}_{\hat{B}_i}(\partial\tau_{0,j})$. Furthermore, we introduce the matrices

$$\mathbf{\Theta}_{B_j} = \frac{1}{\mathrm{mes}_1(\partial\tau_{0,j})}\mathbf{1}_{B_j}\mathbf{1}_{B_j}^{\top}\mathbf{M}_{B_j}\,, \quad \mathbf{\Theta}_j = \frac{1}{\mathrm{mes}_1(\partial\tau_j)}\mathbf{1}_j\mathbf{1}_{B_j}^{\top}\mathbf{M}_{B_j}\,,$$

and $\mathbf{K}_{I_j,\mathrm{it}} = \mathcal{I}_\circ[\mathbf{K}_{I_j}, \mathcal{K}_{I_j}]$ such that, for a given vector $\mathbf{v}_{B_j}$, the products $\mathbf{\Theta}_{B_j}\mathbf{v}_{B_j}$ and $\mathbf{\Theta}_j\mathbf{v}_{B_j}$ yield vectors in the spaces V_{B_j} and V_j, respectively, with all entries equal to the mean value of v_{B_j} on $\partial\tau_j$. According to Definition 2.1, the matrix $\mathbf{K}_{I_j,\mathrm{it}}$ is produced by an inexact iterative process which is formally defined by Definition 2.1. Now, for the prolongation matrix, we can take

$$\mathbf{P}_{V_{B_j}\to V_j} = \mathbf{\Theta}_j + \begin{pmatrix} -\mathbf{K}_{I_j,\mathrm{it}}^{-1}\mathbf{K}_{I_j,B_j}(\mathbf{I}_{B_j} - \mathbf{\Theta}_{B_j}) \\ \mathbf{I}_{B_j} \end{pmatrix}, \tag{5.35}$$

where $\mathbf{I}_{B_j}$ is the unity matrix.

Lemma 5.3. *Let assumptions $\mathcal{A}_1$ – $\mathcal{A}_4$ for the finite element discretization of the elliptic problem (5.2)–(5.3) and the domain decomposition be fulfilled and let inequalities (5.26) hold. Then, for*

$$\nu_{j,1/2} = c(1 + \log(\hat{\alpha}^{(1)}\hbar)^{-1})/(\log\rho_j^{-1})$$

and for any number k_j of iterations with $k_j \geq \nu_{j,1/2}$, the prolongation matrix, defined in (5.35), satifies the estimate

$$\|\mathbf{P}_{V_{B_j}\to V_j}\mathbf{v}_{B_j}\|_{\mathbf{K}_j} \leq 2\|\mathbf{v}_{B_j}\|_{\mathbf{S}_{B_j}}\,, \tag{5.36}$$

where $\rho_j = (1-\vartheta_j)/(1+\vartheta_j)$, $\vartheta_j = \sqrt{\underline{\gamma}_{I_j}/\overline{\gamma}_{I_j}}$, and c is a positive constant independent of $\hbar$.

Proof. We omit the proof, which is quite similar to the proofs of the bounds for analogous prolongation operators derived in [Korneev *et al.* (2003a)]. □

Remark 5.1. Suppose the arithmetical cost of solving the system $\mathcal{K}_I\mathbf{w}_I = \mathbf{w}_B$ is less than $cJ\hbar^{-3}(1+\log\hbar^{-1})^\gamma$. Since the matrix-vector multiplication $\mathbf{K}_I\mathbf{w}_I$ requires $cJ\hbar^{-3}$ a.o., the arithmetical cost of the prolongation $\mathbf{P}_{V_B\to V}\mathbf{v}_B$ is $\mathcal{O}(J\hbar^{-3}(1+\log\hbar^{-1})^{1+\gamma})$, and computational time still can be reduced in a parallel computation with J processors.

In some situations, met, *e.g.*, in the *hp*-version, the cost of the matrix-vector multiplications by $\mathbf{K}_{I_j}$ turns out to be not cheap, but there exists a preconditioner cheap for matrix-vector multiplications. In this case, it can be advantageous to use two preconditioners for the internal stiffness

matrices $\mathbf{K}_{I_j}$. If $\boldsymbol{\mathcal{K}}_{I_j}$ and $\boldsymbol{\mathcal{B}}_{I_j}$ are preconditioners for $\mathbf{K}_{I_j}$, the prolongation matrix can be defined by the expressions

$$\mathbf{P}_{V_{B_j}\to V_j} = \boldsymbol{\Theta}_j + \begin{pmatrix} -\boldsymbol{\mathcal{K}}_{I_j,\mathrm{it}}^{-1}\boldsymbol{\mathcal{K}}_{I_j,B_j}(\mathbf{I}_{B_j} - \boldsymbol{\Theta}_{B_j}) \\ \mathbf{I}_{B_j} \end{pmatrix}, \tag{5.37}$$

where $\boldsymbol{\mathcal{K}}_{\mathrm{it}} = \boldsymbol{\mathcal{I}}_{\circ}[\boldsymbol{\mathcal{K}}_j, \boldsymbol{\mathcal{B}}_j]$. We here assume that $\boldsymbol{\mathcal{K}}_{I_j}$ is cheap for matrix-vector multiplications, but solving systems of SLAE with the matrix $\boldsymbol{\mathcal{B}}_{I_j}$ is much cheaper than solving systems with the system matrix $\boldsymbol{\mathcal{K}}_{I_j}$. If the triangulation of the subdomain is topologically equivalent to a regular triangulation, a good preconditioner $\boldsymbol{\mathcal{K}}_{I_j}$ can be a finite-difference preconditioner. Still, systems with the matrix $\boldsymbol{\mathcal{K}}_{I_j}$ can be not easy to solve. In this case, the preconditioner $\boldsymbol{\mathcal{B}}_{I_j}$ can be the matrix resulting from, *e.g.*, incomplete Cholesky decomposition for $\boldsymbol{\mathcal{K}}_{I_j}$, the matrix corresponding to a few multigrid or two-grid iterations applied to $\boldsymbol{\mathcal{K}}_{I_j}$, see [Jung and Langer (1991); Jung *et al.* (1989)] some kind of multilevel preconditioners like the one of [Bramble *et al.* (1990)] or others.

5.1.4 *Face Component*

Among the most difficult problems in designing DD algorithms for 3d elliptic problems are those related to the preconditioning of the interface and wire basket subproblems. It is necessary to derive preconditioner-solvers $\boldsymbol{\mathcal{S}}_F$ and $\boldsymbol{\mathcal{S}}_W^F$ for the Schur complements $\mathbf{S}_F$ and $\mathbf{S}_W^F$, respectively, and the prolongation matrix $\mathbf{P}_{V_W\to V_B}$, which possess two basic properties, similar to ones mentioned above in relation with the preconditioner $\mathcal{K}_I$. In particular, Schur complement preconditioners should provide good spectral equivalence bounds

$$\underline{\gamma}_B\boldsymbol{\mathcal{S}}_B \le \mathbf{S}_B \le \overline{\gamma}_B\boldsymbol{\mathcal{S}}_B\,, \tag{5.38}$$

$$\underline{\gamma}_W\boldsymbol{\mathcal{S}}_W^F \le \mathbf{S}_W^F \le \overline{\gamma}_W\boldsymbol{\mathcal{S}}_W^F, \tag{5.39}$$

and, at the same time, allow a fast solution of SLAE with these preconditioners as system matrices.

Often, the additional requirement of parallelization of computations is taken into account, and similar to (5.25) $\boldsymbol{\mathcal{S}}_F$ is defined as a block diagonal matrix

$$\boldsymbol{\mathcal{S}}_F = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{F^1}, \boldsymbol{\mathcal{S}}_{F^2}, \ldots, \boldsymbol{\mathcal{S}}_{F^Q}]\,, \tag{5.40}$$

where Q is the number of the FE faces inside the computational domain and each block $\boldsymbol{\mathcal{S}}_{F^k}$ corresponds to d.o.f. of one face F^k. In general, the

Schur complement $\mathbf{S}_F$ is not block diagonal, and, moreover, all faces of the subdomain Schur complement $\mathbf{S}_{F_j}$ can be coupled. Decoupling faces in the preconditioner $\boldsymbol{\mathcal{S}}_F$ affects the relative condition number, including the case when one takes the blocks $\mathbf{S}_{F^k}$ on the diagonal of $\mathbf{S}_F$ for the blocks $\boldsymbol{\mathcal{S}}_{F^k}$ in the preconditioner (5.40). However, these losses are not significant, and under the stated conditions we will show that they can be estimated by the multiplier $\mathcal{O}(\log^2 \hbar^{-1})$. Note that, in the 2d case, we paid the same loss in the complexity order for decoupling edges in the Schur complement preconditioner.

Let us note that one can define preconditioner $\boldsymbol{\mathcal{S}}_B$ (or $\boldsymbol{\mathcal{S}}_F$) by an inexact iterative solver for $\mathbf{S}_B$ (or $\mathbf{S}_F$) or even to reduce solving (5.4) to solving system with the Schur complement $\mathbf{S}_B$. At an iterative solution of the SLAE with the matrix $\mathbf{S}_B$, this matrix is used only for the matrix-vector multiplications $\mathbf{S}_B\mathbf{v}_B$. If there is a fast solver for the SLAE with the matrix $\mathbf{K}_I$, then it is reasonable not to store $\mathbf{S}_B$, but at each iteration to perform the sequence of operations $(\mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB})\mathbf{v}_B$ with the use of the fast solver for realization of $\mathbf{K}_I^{-1}$. Clearly, in this case the computational cost of the operation, implicitly realizing $\mathbf{S}_B\mathbf{v}_B$, is not less than the cost for the operation $\boldsymbol{\mathcal{K}}_I^{-1}\mathbf{v}_I$. Nevertheless, it can be significantly cheaper than the multiplication by the explicitly given $\mathbf{S}_B$. However, the cost is usually reduced by means of defining these preconditioner-solvers in a more elaborate way.

All preconditioners $\boldsymbol{\mathcal{K}}_I$, $\boldsymbol{\mathcal{S}}_B$, $\boldsymbol{\mathcal{S}}_F$, $\boldsymbol{\mathcal{S}}_W^F$ and the prolongation matrices $\mathbf{P}_{V_B\to V}$, $\mathbf{P}_{V_W\to V_B}$ can be assembled from the preconditioners and prolongation matrices $\boldsymbol{\mathcal{K}}_{I_j}$, $\boldsymbol{\mathcal{S}}_{B_j}$, $\boldsymbol{\mathcal{S}}_{F_j}$, $\boldsymbol{\mathcal{S}}_{W_j}^F$, $\mathbf{P}_{V_{B_j}\to V_j}$, $\mathbf{P}_{V_{W_j}\to V_{B_j}}$, similarly defined for each single subdomain Ω_j of the domain decomposition. In accordance with (5.40), the preconditioner for the face subproblem is the block diagonal matrix

$$\boldsymbol{\mathcal{S}}_{F_j} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{F_j^1}, \boldsymbol{\mathcal{S}}_{F_j^2}, \ldots, \boldsymbol{\mathcal{S}}_{F_j^{Q_j}}]\,, \tag{5.41}$$

where the notation F_j^l stands for a face of the subdomain Ω_j, Q_j is the number of free faces (implying $F_j^l \subset \Omega$) of Ω_j and l is the local number of these faces. Let i and j be the numbers of two subdomains having a common face $F^k = \overline{\Omega}_i \bigcap \overline{\Omega}_l$. Further, let s and l be numbers of the face F^k in local orderings of faces for subdomains Ω_i and Ω_j, respectively. Therefore, $F^k = F_i^s = F_j^l$ and $\boldsymbol{\mathcal{S}}_{F^k}$ is the sum of the corresponding blocks of Schur complement preconditioners $\boldsymbol{\mathcal{S}}_{F_i}$ and $\boldsymbol{\mathcal{S}}_{F_j}$:

$$\boldsymbol{\mathcal{S}}_{F^k} = \boldsymbol{\mathcal{S}}_{F_i^s} + \boldsymbol{\mathcal{S}}_{F_j^l}\,, \tag{5.42}$$

where, for simplicity, we assume that d.o.f., living on face F^k, have the same local numbers on the faces F_i^s and F_j^l.

Let $\mathbf{S}_{00,F_j^l}$ be the matrix generated by the quadratic form

$$\mathbf{v}^\top \mathbf{S}_{00,F_j^l}\,\mathbf{v} \equiv {}_{00}|v|^2_{1/2,F_j^l}\,, \qquad \mathbf{v} \leftrightarrow v \in \mathring{\mathcal{V}}(F_j^l)\,, \tag{5.43}$$

where $\mathring{\mathcal{V}}(F_j^l)$ is the subspace of the traces on F_j^l of FE functions vanishing on the wire basket W_j. Then we can set

$$\boldsymbol{\mathcal{S}}_{F^k} = \varrho_i \mathbf{S}_{00,F_i^s} + \varrho_j \mathbf{S}_{00,F_j^l} = (\varrho_i + \varrho_j)\mathbf{S}_{00,F^k}\,. \tag{5.44}$$

Matrices $\mathbf{S}_{00,F_j^p}$ can be replaced by some preconditioners close to them in the spectrum. Suppose that each pair of the FE spaces $\mathcal{V}(\Omega_j)$ and $\mathbb{V}(\tau_{0,j})$ are associated by the mapping $\Upsilon_j(y) \in [\mathbb{V}_j(\tau_{0,j})]^3$ satisfying the shape quasiuniformity conditions. Then the matrices $\mathbf{S}_{00,F_j^p}$ can be replaced by matrices $\mathbb{S}_{00,\hat{F}_j^p}$ generated by the simpler quadratic forms

$$\mathbf{v}^\top \mathbb{S}_{00,\hat{F}_j^p}\mathbf{v} \equiv {}_{00}|v|^2_{1/2,\hat{F}_j^p}\,, \qquad \mathbf{v} \leftrightarrow v \in \mathring{\mathcal{V}}(\hat{F}_j^p)\,, \tag{5.45}$$

multiplied by $H_{D,j}$, where $\mathring{\mathcal{V}}(\hat{F}_j^p)$ is the space of traces of FE functions on the face $\hat{F}_j^p$ of the reference subdomain $\tau_{0,j}$, vanishing on its wire basket. Moreover, the spectral inequalities

$$\underline{\gamma}_\circ H_{D,j}\,\mathbb{S}_{00,\hat{F}_j^p} \;\le\; \mathbf{S}_{00,F_j^p} \;\le\; \overline{\gamma}_\circ\, H_{D,j}\,\mathbb{S}_{00,\hat{F}_j^p} \tag{5.46}$$

as a rule hold with positive constants $\underline{\gamma}_\circ$ and $\overline{\gamma}_\circ$ depending only on shape quasiuniformity conditions, in other words, on $\underline{\alpha}_D, \underline{\theta}_D$ from (5.10), $c_\circ$ from (5.11) and $\theta_\circ$. In spite of all these simplifications, in general, the one-face preconditioner

$$\boldsymbol{\mathcal{S}}_{F^k} = \varrho_i\, H_{D,i}\,\mathbb{S}_{00,\hat{F}_i^s} \;+\; \varrho_j\, H_{D,j}\,\mathbb{S}_{00,\hat{F}_l^j} \tag{5.47}$$

still does not provide good procedures for solving SLAE, governed by it, as well as for its evaluation. In view of this, further simplifications are usually done which take into account particular forms of reference subdomains and their faces as well as their reference discretizations. These simplifications usually result in some preconditioners $\boldsymbol{\mathcal{S}}_{00,\hat{F}_j^p}$ satisfying the spectral equivalence inequalities

$$\underline{\gamma}_{00}\boldsymbol{\mathcal{S}}_{00,\hat{F}_j^p} \;\le\; \mathbb{S}_{00,\hat{F}_j^p} \;\le\; \overline{\gamma}_{00}\,\boldsymbol{\mathcal{S}}_{00,\hat{F}_j^p}\,, \tag{5.48}$$

with some positive spectral constants $\underline{\gamma}_{00}$ and $\overline{\gamma}_{00}$. Then, for the face preconditioners, we can take the matrices

$$\boldsymbol{\mathcal{S}}_{F^k} = \varrho_i\, H_{D,i}\,\boldsymbol{\mathcal{S}}_{00,\hat{F}_i^s} \;+\; \varrho_j\, H_{D,j}\,\boldsymbol{\mathcal{S}}_{00,\hat{F}_j^l}. \tag{5.49}$$

The analysis, completed in Section 5.2, allows us to conclude that, under assumptions $\mathcal{A}.1 - \mathcal{A}.5$ and $\underline{\gamma}_{00}, \overline{\gamma}_{00} = \text{const}$, the presented way of preconditioning of face problems guarantees the spectral inqualities

$$\underline{\gamma}_F \boldsymbol{\mathcal{S}}_F \le \mathbf{S}_F \le \overline{\gamma}_F \boldsymbol{\mathcal{S}}_F \,, \tag{5.50}$$

with the relative spectrum bounds

$$\underline{\gamma}_F \ge \frac{c}{1 + \log^2 \hbar^{-1}} \quad \text{and} \quad \overline{\gamma}_F = \text{const} > 0 \,, \tag{5.51}$$

where c is some positive constant.

Let us note that (5.50) and (5.51) is a collateral result. Indeed, for the control of the relative condition number of the DD preconditioner $\boldsymbol{\mathcal{K}}$ and inter-subdomain Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$, we do not directly use (5.50) and (5.51), but we prove the left inequality of (5.50), see, *e.g.,* Lemma 5.5.

5.1.5 *Wire Basket Component*

The wire basket preconditioner $\boldsymbol{\mathcal{S}}_W^F$ is a single global matrix, governing subsystems of algebraic equations to be solved at each iteration of the DD algorithm. Under the assumption of the subdomainwise constant ϱ, it is also a single preconditioner in the DD algorithm, the inversion of which is essentially influenced by the jumps of ϱ across the inter-subdomain boundaries. Obviously, the preconditioner $\boldsymbol{\mathcal{S}}_W^F$ can be assembled from subdomain preconditioners $\boldsymbol{\mathcal{S}}_{W_j}^F, j = 1, 2, \ldots, J$, and, in view of assumptions $\mathcal{A}.1 - \mathcal{A}.3$, these subdomain preconditioners can be defined by means of the reference subdomains. Until now, no trace theorem has been proved assuring that, for domains $\mathcal{D} \subset R^2$ with piecewise smooth boundaries, the traces on $\partial\mathcal{D}$ of all functions from the space $H^{1/2}(\mathcal{D})$ belong to the space $L_2(\partial\mathcal{D})$. This created one of the difficulties in designing efficient wire basket preconditioners. Basic results in the analysis of the BPS preconditioner in [Bramble *et al.* (1986, 1987, 1988, 1989)] have been established with the use of the "seminorm" $|\cdot|_{0,\partial\mathcal{D}}$ in the space $L_2(\partial\mathcal{D})$ as the norm in the space of traces of FE functions from $L_2^{1/2}(\mathcal{D})$. This seminorm is introduced by the equality

$$|\, v \,|_{0,\partial\mathcal{D}}^2 \stackrel{\text{def}}{=} \inf_{c \in R} \| \, v - c \, \|_{0,\partial\mathcal{D}}^2 \,, \qquad \forall\, v \in L_2(\partial\mathcal{D}) \,. \tag{5.52}$$

In Dirichlet–Dirichlet DD algorithms, the roles of $\mathcal{D}$ and $\partial\mathcal{D}$ are played by faces and their boundaries, and, respectively, by the interface boundary and the wire basket. For the reference subdomain $\tau_{0,j}$, the roles of $\mathcal{D}$ and $\partial\mathcal{D}$ are played by an edge of $\tau_{0,j}$ and its boundary, or by the boundary

$\partial\tau_{0,j}$ and the wire basket $\hat{W}_j$. Later, in Lemma 5.6, we give the bounds of seminorms $|\,v\,|_{0,\hat{W}_j}$ for the traces on $\hat{W}_j$ of FE functions by their seminorms $|\,v\,|_{1/2,\partial\tau_{0,j}}$, and the bounds in the opposite direction for the appropriately chosen prolongations. Motivated by these bounds, the wire basket preconditioner is usually obtained via preconditioning of the quadratic forms

$$\mathbf{v}_{W_j}^{\top}\mathbf{W}_j\,\mathbf{v}_{W_j} \;=\; |\,v\,|^2_{0,\hat{W}_j} \stackrel{\rm def}{=} \inf_{c_j\in R} \|\,v-c_j\,\|^2_{0,\hat{W}_j} \tag{5.53}$$

that hold for all $\mathbf{v}_{W_j} \leftrightarrow v \in \mathbb{V}_W(\hat{W}_j)$. In particular, we can set

$$\boldsymbol{\mathcal{S}}^F_{W_j} \;=\; \varrho_j\, H_{D,j}\,(1+\log\hbar^{-1})\,\boldsymbol{\mathcal{W}}_j\,, \tag{5.54}$$

where $\boldsymbol{\mathcal{W}}_j$ is some good preconditioner for $\mathbf{W}_j$. The space of traces $\mathbb{V}_W(\hat{W}_j)$ contains continuous piecewise linear functions. For this reason, if $\mathbf{M}_{\hat{W}_j}$ is the mass matrix defined by the identity

$$\mathbf{v}_{W_j}^{\top}\mathbf{M}_{\hat{W}_j}\mathbf{v}_{W_j} = ||v||^2_{0,\hat{W}_j}\,, \qquad \forall\ \mathbf{v}_{W_j} \leftrightarrow v \in \mathbb{V}_W(\hat{W}_j)\,,$$

and vector $\mathbf{z}_j$ contains unity for all entries, then simple computations give

$$\int_{\hat{W}_j} v^2\,ds - \frac{(\int_{\hat{W}_j} v\,ds)^2}{\int_{\hat{W}_j} ds} = \mathbf{v}_{W_j}^{\top}\left(\mathbf{M}_{\hat{W}_j} - \frac{(\mathbf{M}_{\hat{W}_j}\mathbf{z}_j)(\mathbf{M}_{\hat{W}_j}\mathbf{z}_j)^{\top}}{\mathbf{z}_j^{\top}\mathbf{M}_{\hat{W}_j}\mathbf{z}_j}\right)\mathbf{v}_{W_j}\,.$$

Therefore, we have

$$\mathbf{W}_j = \mathbf{M}_{\hat{W}_j} - \frac{1}{\mathrm{mes}_1(\hat{W}_j)}(\mathbf{M}_{\hat{W}_j}\mathbf{z}_j)(\mathbf{M}_{\hat{W}_j}\mathbf{z}_j)^{\top}\,, \tag{5.55}$$

where $\mathrm{mes}_1(\hat{W}_j)$ is the length of the wire basket $\hat{W}_j$ of the reference subdomain.

The solution procedure for the SLAE with the matrix $\boldsymbol{\mathcal{S}}^F_W$ simplifies, if we replace the mass matrix $\mathbf{M}_{\hat{W}_j}$ by its diagonal preconditioner, for which we use the notation $\mathbf{D}_j$. We assemble it from matrices $\mathbf{D}_{\hat{E}^k_j}$ defined for each edge $\hat{E}^k_j$ of the reference domain. It is sufficient to describe the matrix $\mathbf{D}_{\hat{E}^k_j}$ for one edge $\hat{E}^k_j$, which without loss of generality can be described as the set $\hat{E}^k_j = \{x : 0 < x_1 < \ell,\ x_2, x_3 \equiv 0\}$. Suppose, points $x_1 = x_1^{(i)},\ i = 0, 1, .., n,\ \hbar_i = x_1^{(i)} - x_1^{(i-1)} > 0,\ x^{(0)} = 0,\ x^{(n)} = \ell,$ are the nodes of discretization, and

$$\eta_0 = \hbar_1/2, \quad \eta_i = (\hbar_i + \hbar_{i+1})/2 \quad \text{for} \quad i = 1, 2.., n-1, \quad \eta_n = \hbar_n/2\,.$$

Then we set

$$\mathbf{D}_{\hat{E}^k_j} = \mathrm{diag}\,[\eta_i]_{i=0}^n\,. \tag{5.56}$$

When assembling the matrix $\mathbf{D}_j$, assembling is indeed needed only for vertices of the polygon $\tau_{0,j}$ and not for other wire-basket nodes. If vertex $x^{(i)} \in \hat{W}_j$ is common for $\kappa \geq 3$ edges, then the entry on the diagonal of $\mathbf{D}_j$, corresponding to this vertex, is the sum of κ those single entries in κ edge matrices $\mathbf{D}_{\hat{E}_j^k}$, which correspond to the same vertex.

Defining the matrix $\boldsymbol{\mathcal{W}}_j$ by the relation

$$\mathbf{v}_{W_j}^{\top} \boldsymbol{\mathcal{W}}_j \, \mathbf{v}_{W_j} \equiv \inf_{c_j \in R} \sum_{y^{(i)} \in \hat{W}_i} \eta_i (v_i - c_j)^2 \tag{5.57}$$

with summation over the nodes $y^{(i)}$ on the wire basket $\hat{W}_i$, we come to the equality

$$\boldsymbol{\mathcal{W}}_j = \mathbf{D}_j - \frac{1}{\mathrm{mes}_1(\hat{W}_j)} (\mathbf{D}_j \mathbf{z}_j)(\mathbf{D}_j \mathbf{z}_j)^{\top} \tag{5.58}$$

in the same way as above. Then we define $\boldsymbol{\mathcal{S}}_{W_j}^F$ by (5.54). As a consequence of the spectral equivalence of matrices $\mathbf{M}_{\hat{W}_j}$ and $\mathbf{D}_j$, the two matrices $\mathbf{W}_j$ and $\boldsymbol{\mathcal{W}}_j$, defined in (5.55) and in (5.58), respectively, are also spectrally equivalent. The wire basket Schur complement preconditioner is defined now by assembling the subdomain preconditioners $\boldsymbol{\mathcal{S}}_{W_j}^F$:

$$\boldsymbol{\mathcal{S}}_W^F = \biguplus_{j=1}^{J} \boldsymbol{\mathcal{S}}_{W_j}^F = \biguplus_{j=1}^{J} \varrho_j \, H_{D,j} \, (1 + \log \hbar^{-1}) \, \boldsymbol{\mathcal{W}}_j, \tag{5.59}$$

where the sign $\biguplus_{j=1}^{J}$ stands for the assembling procedure over all subdomains.

The solution procedure for the system of algebraic equations

$$\mathbf{S}_W^F \mathbf{v} = \mathbf{f}_W \tag{5.60}$$

still require consideration due to the presence of the second term in the right part of (5.58). It is equivalent to the minimization of the quadratic functional, *i.e.*, finding

$$\min_{\mathbf{v}} I(\mathbf{v}) = \min_{\mathbf{v}} \left(\frac{1}{2} \sum_{j=1}^{J} \min_{c_j} [(\mathbf{v}^{(j)} - \mathbf{z}_j c_j)^{\top} \mathbf{D}_W^{(j)} (\mathbf{v}^{(j)} - \mathbf{z}_j c_j)] - \mathbf{v}^{\top} \mathbf{f} \right), \tag{5.61}$$

where $\mathbf{D}_W^{(j)} = \varrho_j \, H_{D,j} \, (1 + \log h^{-1}) \mathbf{D}_j$ and $\mathbf{v}^{(j)}$ denotes the restriction to the wire basket of the reference domain. When it does not cause confusion, we omit index "W" in the notations of vectors, living on the wire basket. We describe first the procedure for finding the minimizer of $I(\mathbf{v})$ in the case of the Neumann boundary conditions on $\partial\Omega$, when all nodes of

the discretization are free nodes. Let $\mathbf{D}_W$ be the matrix assembled from matrices $\mathbf{D}_W^{(j)}$. The conditions of minimum of quadratic functional (5.61) with respect to components of the vector $\mathbf{v}$ and constants c_j lead to the system of algebraic equations

$$\mathbf{D}_W\mathbf{v} - \biguplus_j(\mathbf{D}_W^{(j)}\mathbf{z}_j c_j) = \mathbf{f} \quad \text{and} \quad \mathbf{z}_j^\top \mathbf{D}_W^{(j)}\mathbf{v}^{(j)} = c_j a_j \tag{5.62}$$

with $a_j = \mathbf{z}_j^\top \mathbf{D}_W^{(j)}\mathbf{z}_j = \varrho_j H_{D,j}(1+\log h^{-1})\,\mathrm{mes}_1(\hat{W}_j)$. Since $\mathbf{D}_W$ is a diagonal matrix, we can easily express $\mathbf{v}$ through the unknown constants c_j, *i.e.,*

$$\mathbf{v} = \mathbf{D}_W^{-1}\left(\biguplus_j(\mathbf{D}_W^{(j)}\mathbf{z}_j c_j) + \mathbf{f}\right), \tag{5.63}$$

and substitute in the second group of equations (5.62). This produces the system of algebraic equations

$$c_j a_j - \mathbf{z}_j^\top \mathbf{D}_W^{(j)}\left[\mathbf{D}_W^{-1}\biguplus_l(\mathbf{D}_W^{(l)}\mathbf{z}_l c_l)\right]_j = \mathbf{z}_j^\top \mathbf{D}_W^{(j)}\left[\mathbf{D}_W^{-1}\mathbf{f}\right]_j, \tag{5.64}$$

where $[\mathbf{s}]_j$ stands for the subvector $[\mathbf{s}]_j \in V_{W_j}$ of the vector $\mathbf{s} \in V_W$. The product $\mathbf{d}_j := \mathbf{D}_W^{(j)}\mathbf{z}_j$ is the vector that contains the diagonal elements of $\mathbf{D}_W^{(j)}$ as entries. On this account, systems (5.63) and (5.64) can now be rewritten as

$$\mathbf{v} = \mathbf{D}_W^{-1}\left(\biguplus_j(\mathbf{d}_j c_j) + \mathbf{f}\right) \tag{5.65}$$

and

$$c_j a_j - \mathbf{d}_j^\top\left[\mathbf{D}_W^{-1}\biguplus_l(\mathbf{d}_l c_l)\right]_j = \mathbf{d}_j^\top\left[\mathbf{D}_W^{-1}\mathbf{f}\right]_j, \tag{5.66}$$

respectively. After solving system (5.66), we substitute the solution in (5.65) and calculate vector $\mathbf{v}$.

It is necessary to make several remarks. System (5.66) resembles FE/FD systems of algebraic equations. The matrix of this system is nonnegative. The sum of its coefficients in each row is zero. In order to establish that, it is necessary to substitute all $c_l = 1$ and to note that the entries of the vector inside square brackets are first of all positive and are equal to 1. From that it follows that the vector $\mathbf{c} = \{c_j\}_{j=1}^j$ is found up to an arbitrary constant vector $c\mathbf{1}$, *i.e.*, if $\mathbf{c}$ is the solution, then $\mathbf{c} + c\mathbf{1}$ is also solution.

As a consequence, the vector, which solves (5.65), is also defined up to the same constant vector $c\mathbf{1}$. At the Dirichlet boundary condition the values of a_j, see (5.62), are not changed. This means that in expression (5.62) for a_j matrix $\mathbf{D}_W^{(j)}$ is written as no boundary conditions are imposed on $\partial\hat{\Omega}_j$ and has the dimension $N_{W_j}^+ \times N_{W_j}^+$, where $N_{W_j}^+$ is the number of all nodes on W_j. However, in relations (5.62)–(5.66), the matrices $\mathbf{D}_W$ and $\mathbf{D}_W^{(i)}$ take into account the Dirichlet boundary conditions and, *e.g.*, the latter have dimensions $N_{W_i} \times N_{W_i}$, where N_{W_i} is the number of free nodes on the wire basket W_j. Respectively, vectors $\mathbf{d}_i$ have dimension N_{W_i}. As a consequence, in all rows, corresponding to c_j with subdomain boundary $\partial\Omega_j$ having nodes on $\partial\Omega$, there is some diagonal predominance, the matrix of the system (5.66) is s.p.d., and this system has a unique solution.

Remark 5.2. If all coefficients before c_j in (5.66) are calculated, then the solution of (5.66) by the general Gauss elimination procedure will require cJ^3=const a.o. More efficient solution procedures can be used, and, in particular, sparsity of the matrix of the system (5.66) can be taken into account. If the vector in brackets in (5.65) is known, then one has to spend $cJ\hbar^{-1} = \mathcal{O}(\hbar^{-1})$ a.o. in order to find $\mathbf{v}$. It is easy to see, that the arithmetical cost of the coefficients before c_j in (5.66) and the vector in brackets in (5.65) is $cJ\hbar^{-1} = \mathcal{O}(\hbar^{-1})$ a.o. Therefore, if J is fixed, the arithmetical cost of solving the system (5.60) is $\mathcal{O}(\hbar^{-1})$.

The prolongation matrix $\mathbf{P}_{V_W \to V_B}$ can be defined face by face, *i.e.*, by means of the matrices

$$\mathbf{P}_{V_{W^q} \to V_{B^q}}\,, \qquad q = 1, 2, .., Q\,,$$

where $\mathbf{P}_{V_{W^q} \to V_{B^q}}$ is the prolongation matrix from the boundary $W^q = \partial F^q$ on the closure $\overline{F}^q$ of a face F^q. If $\mathbf{v}_{W^q}$ is a vector living on W^q, then by $\widetilde{\mathbf{P}}^{(q)}$ is denoted such prolongation matrix that $\widetilde{\mathbf{P}}^{(q)}\mathbf{v}_{W^q}$ is the prolongation by zero entries to all internal nodes of the face. We set

$$\mathbf{P}_{V_{W^q} \to V_{\overline{F}^q}} \mathbf{v}_{W^q} = \mathrm{m_{ean}}(\mathbf{v}_{W^q})\mathbf{1}_{\overline{F}^q} + \widetilde{\mathbf{P}}^{(q)}(\mathbf{v}_{W^q} - \mathrm{m_{ean}}(\mathbf{v}_{W^q})\mathbf{1}_{W^q})\,, \quad (5.67)$$

where $\mathbf{1}_{\overline{F}^q}$ and $\mathbf{1}_{W^q}$ are vectors living on $\overline{F}^q$ and W^q, respectively, and having unities for components, and $\mathrm{m_{ean}}(\mathbf{v})$ is the arithmetic mean value of components of the vector $\mathbf{v}$. Obviously, the so defined prolongation $\mathbf{P}_{V_W \to V_B}\mathbf{v}$ is cheap and requires $J\hbar^{-2}$ a.o. Later we will establish that this prolongation is also efficient in providing a good relative condition number for the DD preconditioners, see Theorem 5.2.

5.2 Condition Number and Complexity Estimates

We start with abstract estimates of the relative spectrum of the DD preconditioners under consideration. They are expressed through relative spectrum bounds of the component preconditioners and bounds for the norms of prolongation operators. Then, in Subsection 5.2.2, we give a more accurate analysis of the quantities entering these abstract bounds, and illuminate the dependence of the relative condition numbers of DD preconditioners on the parameters characterizing the elliptic boundary value problem, the FE discretizations and the decompositions of the domain into subdomains. In particular, a more detailed representation of the interface Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$, which is more adjusted to fast and parallelized computations, is taken into account. In correspondence with the previously given description, it is composed of the face preconditioner with a splitting of the d.o.f. living on different faces, the specific wire basket preconditioner, and some prolongation operator from the wire basket to the interface. The final result of this analysis is stated in Theorem 5.2. Several subsidiary results are collected in Subsection 5.2.3.

5.2.1 *Abstract Bound for the Relative Condition Number*

According to the above outline, we first derive the abstract relative spectrum bounds, that is here completed in two steps. Basically, in compliance with the representation (5.14) and (5.15) of the DD preconditioner, we apply Corollary 4.2 twice. This allows us to estimate the relative condition number of the DD preconditioner under quite general assumptions, pertaining to all five components of the DD preconditioner, namely to the three component preconditioners and the two prolongation operators.

Lemma 5.4. *Let inequalities (5.26) for the preconditioner $\boldsymbol{\mathcal{K}}_I$, inequalities (5.28) for the prolongation operators $\mathbf{P}_{V_{B_j}\to V_j}$ and inequalities (5.38) for the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$ hold. Then*

$$\underline{\gamma}\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{\gamma}\boldsymbol{\mathcal{K}} \tag{5.68}$$

with the relative spectrum bounds

$$\underline{\gamma} \ge \frac{\underline{\gamma}_I}{2(1 + c^2_{P_B} + \underline{\gamma}_B^{-1}\underline{\gamma}_I)} \quad \textit{and} \quad \overline{\gamma} \le 2\max(\overline{\gamma}_I, c^2_{P_B}\overline{\gamma}_B)\,. \tag{5.69}$$

Proof. Let us consider $\mathbf{v} \in V$ represented in two forms $\mathbf{v} = \mathbf{v}_I + \mathbf{v}_B = \overline{\mathbf{v}}_I + \overline{\mathbf{v}}_B$, where $\overline{\mathbf{v}}_B := \mathbf{P}_{V_B\to V}\mathbf{v}_B$ and $\overline{\mathbf{v}}_I := \mathbf{v} - \overline{\mathbf{v}}_B$. Using the left inequality

of (5.26), estimating the square of a sum by the sum of the squares, and applying (5.28), we arrive at the estimates

$$\begin{aligned}\overline{\mathbf{v}}_I^\top \boldsymbol{\mathcal{K}}_I \overline{\mathbf{v}}_I &\le 2\underline{\gamma}_I^{-1}(\mathbf{v}^\top \mathbf{K}\mathbf{v} + \overline{\mathbf{v}}_B^\top \mathbf{K}\overline{\mathbf{v}}_B) \le 2\underline{\gamma}_I^{-1}(\mathbf{v}^\top \mathbf{K}\mathbf{v} + c_{P_B}^2 \|\mathbf{v}_B\|_{\mathbf{S}_B}^2) \\ &\le 2\underline{\gamma}_I^{-1}(1 + c_{P_B}^2)\mathbf{v}^\top \mathbf{K}\mathbf{v}\,.\end{aligned}$$

Another bound needed is

$$\|\mathbf{P}_{V_{B_j} \to V_j} \mathbf{v}_{B_j}\|_{\mathbf{K_j}}^2 \le c_{P_B}^2 \overline{\gamma}_B \|\mathbf{v}_{B_j}\|_{\boldsymbol{\mathcal{S}}_{B_j}}\,, \qquad \forall\, \mathbf{v}_{B_j} \in \mathbf{V}_{B_j}^2\,, \tag{5.70}$$

which is a consequence of (5.28) and the right inequality (5.38). Now we see that we can apply Corollary 4.2 by setting

$$\underline{c}_A = \frac{\underline{\gamma}_I}{2(1 + c_{P_B}^2)}\,, \qquad \overline{c}_A = \overline{\gamma}_I\,, \qquad \underline{c}_S = \underline{\gamma}_B\,, \qquad \overline{c}_S = c_{P_B}^2 \overline{\gamma}_B\,,$$

which completes the proof. □

Lemma 5.5. *Suppose that inequalities (5.39) for the wire basket Schur complement preconditioner, the inequalities*

$$\widetilde{\underline{\gamma}}_F \boldsymbol{\mathcal{S}}_F \le \mathbf{S}_B\,, \qquad \mathbf{S}_F \le \overline{\gamma}_F \boldsymbol{\mathcal{S}}_F \tag{5.71}$$

for the face subproblem preconditioner, and the inequality

$$\|\mathbf{P}_{V_W \to V_B} \mathbf{v}_W\|_{\mathbf{S}_B}^2 \le c_{P_W} \mathbf{v}_W^\top \boldsymbol{\mathcal{S}}_W^F \mathbf{v}_W \tag{5.72}$$

for the prolongation from the wire basket to the interface hold. Then the spectral inequalities

$$\underline{\gamma}_B \boldsymbol{\mathcal{S}}_B \le \mathbf{S}_B \le \overline{\gamma}_B \boldsymbol{\mathcal{S}}_B \tag{5.73}$$

are valid with the relative spectrum bounds

$$\underline{\gamma}_B \ge \frac{1}{\widetilde{\underline{\gamma}}_F^{-1} + \underline{\gamma}_W^{-1}} \quad \textit{and} \quad \overline{\gamma}_B \le 2\max(\overline{\gamma}_F, c_{P_W})\,. \tag{5.74}$$

Proof. With the setting

$$\underline{c}_A = \widetilde{\underline{\gamma}}_F\,, \qquad \overline{c}_A = \overline{\gamma}_F\,, \qquad \underline{c}_S = \underline{\gamma}_W\,, \qquad \overline{c}_S = c_{P_W}\,,$$

the proof directly follows from Corollary 4.2. □

Now, we are in the position to formulate an abstract estimates of the relative spectral bounds for our DD preconditioner.

Theorem 5.1. *Under the assumptions of Lemmas 5.4 and 5.5, the spectral equivalence inequalities (5.68) hold with*

$$\underline{\gamma} \ge \frac{\underline{\gamma}_I}{2(1 + c_{P_B}^2 + \underline{\gamma}_I(\widetilde{\underline{\gamma}}_F^{-1} + \underline{\gamma}_W^{-1}))} \quad \textit{and} \quad \overline{\gamma} \le 4\max(\frac{\overline{\gamma}_I}{2}, c_{P_B}^2 \overline{\gamma}_F, c_{P_B}^2 c_{P_W})\,. \tag{5.75}$$

Proof. For the proof, it is sufficient to plug (5.74) in (5.69). □

5.2.2 *Dependence of the Relative Condition Number on Discretization and Decomposition Parameters*

In order to find out the dependence of $\underline{\gamma}$ and $\overline{\gamma}$ on the parameters of the FE discretization and domain decomposition, we have to establish the dependence of the values entering the right parts of the bounds (5.75) on these parameters. For this purpose, we start from summarizing all our specific assumptions imposed on the component preconditioners and prolongation operators for DD preconditioner (5.14)–(5.15). First of all, we fix the set of assumptions allowing to elucidate the dependence on $\hbar$. This dependence, which though can be characterized as weak, is irreducible within the structure of DD preconditioner under consideration:

- `i)` Assumptions $\mathcal{A}.1 - \mathcal{A}.5$ are fulfilled.
- `ii)` Interior subdomain preconditioners $\boldsymbol{\mathcal{K}}_{I_j}$ satisfy inequalities (5.26).
- `iii)` The Schur complement preconditioner $\boldsymbol{\mathcal{S}}_F$ is defined according to (5.40) and (5.49) and inequalities (5.48) hold.
- `iv)` The second level Schur complement preconditioner $\boldsymbol{\mathcal{S}}^F_W$ is defined according to (5.58) and (5.59).
- `v)` The prolongation $\mathbf{P}_{V_B \to V}$ from the inter-subdomain boundary to the whole computational domain satisfies (5.28).
- `vi)` The prolongation $\mathbf{P}_{V_{W_j} \to V_{B_j}}$ from the wire basket to the inter-subdomain boundary is defined by (5.67).

Most multipliers in front of matrices, norms and other values, entering the inequalities participating in conditions `i)`–`vi)`, are considered as constants independent of $\hbar$. In particular, these are the constants $\underline{\alpha}_D, \underline{\theta}_D, c_\circ, \wp_\circ, \theta_\circ$ from the shape regularity conditions for the mappings $x = \Upsilon_j(y) : \overline{\tau}_{0,j} \to \overline{\Omega}_j$ and size and shape quasiuniformity conditions for the reference subdomains, the constants $\hat{\alpha}^{(1)}, \hat{\theta}$ from the quasiuniformity conditions for FE discretizations of reference subdomains, the constants $\underline{\gamma}_I, \overline{\gamma}_I, \underline{\gamma}_{00}, \overline{\gamma}_{00}$ from the energy inequalities (5.26) and (5.48), and the constant c_{P_B} from (5.28), which bounds the energy norm of the prolongation operator from the inter-subdomain boundary.

Theorem 5.2. *Under the stated conditions* `i) - vi)`*, the inequalities*

$$\frac{\underline{c}}{(1+\log \hbar^{-1})^2}\,\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{c}\,\boldsymbol{\mathcal{K}} \tag{5.76}$$

hold with positive constants $\underline{c}$ *and* $\overline{c}$ *which are independent of* $\hbar$.

Proof. The $\hbar$-independent bound for the factor c_{P_B} in inequalities (5.27) and (5.28) directly follows from Lemma 5.3 and the definition of the

prolongation matrix $\mathbf{P}_{V_B \to V}$. Therefore, for the proof of the Theorem, it is necessary to establish the dependence of quantities $\underline{\gamma}_W, c_{P_W}, c_F, \overline{\gamma}_F$ in (5.75) on $\hbar$. It will be done below with the help of subsidiary results collected in Subsection 5.2.3.

1. Bound for $\underline{\gamma}_W$: Since the preconditioner $\boldsymbol{\mathcal{S}}^F_W$ defined in (5.54), (5.58), (5.59), and the preconditioner defined in (5.54), (5.55), (5.59) are spectrally equivalent, it is sufficient to consider only the latter. According to (5.54)–(5.55) and inequality (5.95) of Lemma 5.6 in Subsection 5.2.3, we have

$$\begin{aligned}\mathbf{v}^\top_{W_j} \boldsymbol{\mathcal{S}}^F_{W_j} \mathbf{v}_{W_j} &= \varrho_j \, H_{D,j} \, (1 + \log \hbar^{-1}) |v_{\hat{W}_j}|^2_{0,\hat{W}_j} \\ &\le c \, \varrho_j \, H_{D,j} \, (1 + \log^2 \hbar^{-1}) |v_{\hat{B}_j}|^2_{1/2,\partial\tau_{0,j}} \,, \end{aligned} \tag{5.77}$$

where $v_{\hat{W}_j}$ is the FE function on $\hat{W}_j$, corresponding to vector $\mathbf{v}_{W_j}$, and $v_{\hat{B}_j} \in \mathbb{V}_{B_j}$ is any FE function coinciding with $\mathbf{v}_{W_j}$ at the nodes on $\hat{W}_j$. By the conditions of shape regularity (5.10) for mappings $x = \Upsilon_j(y) : \tau_{0,j} \to \Omega_j$, we have

$$c_1 H_{D,j} |v_{\hat{B}_j}|^2_{1/2,\partial\tau_{0,j}} \le |u_{B_j}|^2_{1/2,\partial\Omega_j} \le c_2 H_{D,j} |v_{\hat{B}_j}|^2_{1/2,\partial\tau_{0,j}} , \tag{5.78}$$

with the constants $c_k = c_k(\underline{\alpha}_D, \underline{\theta_D})$ and $u_{B_j}(\Upsilon_j(y)) = v_{\hat{B}_j}(y)$. Therefore, for any $\mathbf{v}_{B_j}$ with the subvector $\mathbf{v}_{W_j}$, it follows by (5.77), (5.78) and (5.30) that

$$\mathbf{v}^\top_{W_j} \boldsymbol{\mathcal{S}}^F_{W_j} \mathbf{v}_{W_j} \le c\varrho_j \, (1 + \log \hbar^{-1})^2 |u_{B_j}|^2_{1/2,\partial\Omega_j} \le c(1 + \log^2 \hbar^{-1}) \mathbf{v}^\top_{B_j} \mathbf{S}_{B_j} \mathbf{v}_{B_j} \,. \tag{5.79}$$

The inequality (5.79) holds for any vector $\mathbf{v}_{B_j}$ and, consequently, for the vector $\mathbf{v}'_{B_j}$ minimizing the norm $\|\cdot\|_{\mathbf{S}_{B_j}}$ among all vectors $\mathbf{v}_{B_j}$ with the fixed subvector $\mathbf{v}_{W_j}$. This approves the left inequality (5.39) for the subdomain Ω_j with the specific value of $\underline{\gamma}_W$:

$$\underline{\gamma}_W \boldsymbol{\mathcal{S}}^F_{W_j} \le \mathbf{S}^F_{W_j} \,, \qquad \underline{\gamma}_W = \frac{c}{1 + \log^2 \hbar^{-1}} > 0 \,, \tag{5.80}$$

and the inequality (5.39) follows from (5.80) by assembling.

2. Bound for c_{P_W}: The value of c_{P_W} in (5.72) depends on the choice of the preconditioner $\boldsymbol{\mathcal{S}}^F_W$. Again it is sufficient to give the proof for one of the preconditioners introduced in Subsection 5.1.5, namely for the one defined in (5.54), (5.55), (5.59). Let $\mathbf{v}_B = \mathbf{P}_{V_W \to V_B} \mathbf{v}_W$ and $v_W \in \mathbb{V}_W$, $v_B \in \mathbb{V}_B$ be the corresponding FE functions. Applying the right inequalities (5.30) and (5.78), the inequality (5.94) of Lemma 5.6, and the definition of $\boldsymbol{\mathcal{S}}^F_{W_j}$ by (5.54), (5.55), we arrive at the following bound:

$$\begin{aligned}\|\mathbf{P}_{V_{W_j} \to V_{B_j}} \mathbf{v}_{W_j}\|^2_{\mathbf{S}_{B_j}} &\le c_2 \overline{\gamma}_S \varrho_j H_{D,j} |v_{\hat{B}_j}|^2_{1/2,\partial\tau_{0,j}} \le c\varrho_j H_{D,j} \, |v_{\hat{W}}|^2_{0,\hat{W}_j} \\ &\le \frac{c}{1 + \log \hbar^{-1}} \, \mathbf{v}^\top_{W_j} \boldsymbol{\mathcal{S}}^F_{W_j} \mathbf{v}_{W_j} \,. \end{aligned} \tag{5.81}$$

This means that

$$c_{P_W} \le c/(1+\log\hbar^{-1}). \tag{5.82}$$

Therefore, c_{P_W} is small for small $\hbar$.

3. Bound for $\underline{\widetilde{\gamma}}_F$: Alongside with (5.108), for the same FE functions, we get by a factor space argument:

$$_{00}|v_F|^2_{1/2,\mathcal{F}} \le c_{\mathcal{F}}(1+\log^2\hbar^{-1})|v|^2_{1/2,\partial\tau_0}\,. \tag{5.83}$$

In the face preconditioner $\boldsymbol{\mathcal{S}}_{F_j}$ for the subdomain Ω_j, we consider one block $\boldsymbol{\mathcal{S}}_{F_j^p}$ on its diagonal corresponding to one of the faces F_j^p. From the definition of this block, see (5.45), (5.48), (5.49), it follows

$$\mathbf{v}^\top_{F_j^p}\boldsymbol{\mathcal{S}}_{F_j^p}\mathbf{v}_{F_j^p} = \rho_j\,H_{D,j}\mathbf{v}^\top_{F_j^p}\boldsymbol{\mathcal{S}}_{00,\hat{F}_j^p}\mathbf{v}_{F_j^p} \le \underline{\gamma}_{00}^{-1}\rho_j\,H_{D,j}|v_{\hat{F}_j^p}|^2_{1/2,\hat{F}_j^p}\,. \tag{5.84}$$

Let FE functions $v_{\hat{F}_j^p} \in \mathbb{V}_F(\hat{F}_j^p)$ and $u_{F_j^p} \in \mathcal{V}_F(F_j^p)$, such that $u_{F_j^p}(\Upsilon_j(y)) = v_{\hat{F}_j}(y)$, correspond to an arbitrary vector $\mathbf{v}_{F_j^p} \in V_{F_j^p}$. By the use of (5.83), the shape regularity conditions (5.10) for mappings $x = \Upsilon_j(y) : \tau_{0,j} \to \Omega_j$, and (5.30), we get the estimates

$$\begin{aligned}
\underline{\gamma}_{00}^{-1}\rho_j\,H_{D,j}|v_{\hat{F}_j^p}|^2_{1/2,\hat{F}_j^p} &\le c_{\mathcal{F}}\underline{\gamma}_{00}^{-1}\rho_j\,H_{D,j}(1+\log^2\hbar^{-1})|v|^2_{1/2,\partial\tau_{0,j}}\\
&\le cc_{\mathcal{F}}\underline{\gamma}_{00}^{-1}\rho_j\,(1+\log^2\hbar^{-1})|u|^2_{1/2,\partial\Omega_j}\\
&\le cc_{\mathcal{F}}\underline{\gamma}_{00}^{-1}\underline{\gamma}_S^{-1}(1+\log^2\hbar^{-1})\mathbf{v}_j^\top\mathbf{S}_j\mathbf{v}_j\,.
\end{aligned} \tag{5.85}$$

From (5.84) and (5.85), the definition of the face preconditioner $\boldsymbol{\mathcal{S}}_{F_j}$, and the shape regularity of, in general curvilinear polygons Ω_j, it follows that

$$\underline{\widetilde{\gamma}}_F^{-1} \le cc_{\mathcal{F}}\underline{\gamma}_{00}^{-1}\underline{\gamma}_S^{-1}(1+\log^2\hbar^{-1})\,. \tag{5.86}$$

4. Bound for $\overline{\gamma}_F$: Let $\mathcal{F} = F^p$ be one of the faces of the reference subdomain $\tau_{0,j}$, which satisfies the shape regularity conditions. We remind that, for $v_{\mathcal{F}} \in H^{1/2}(\tau_{0,j})$, $\mathrm{supp}(v_{\mathcal{F}}) = \mathcal{F}$, we have

$$\|v_{\mathcal{F}}\|_{1/2,\tau_{0,j}} \approx {}_{00}|v_{\mathcal{F}}|_{1/2,\mathcal{F}} \tag{5.87}$$

by the definition of ${}_{00}|\cdot|_{1/2,\mathcal{F}}$. Furthermore, the number of faces of each subdomain is bounded independently of j. We now consider an arbitrary vector $\mathbf{v}_{F_j}$ and two FE functions corresponding to it, namely $v_{\hat{F}_j} \in \mathbb{V}_F(\hat{F}_j)$ and $u_{F_j} \in \mathcal{V}_F(F_j)$ such that $u_{F_j}(\Upsilon_j(y)) = v_{\hat{F}_j}(y)$. Taking (5.30), the Cauchy inequality, the shape regularity conditions for mappings $x = \Upsilon_j(y) : \tau_{0,j} \to \Omega_j$, and (5.87) into account, we can write

$$\begin{aligned}
&\|\mathbf{v}_{F_j}\|_{\mathbf{S}_{F_j}} \le \overline{\gamma}_S\rho_j|u_{F_j}|_{1/2,\partial\Omega_j} \le c\overline{\gamma}_S\rho_j H_{D,j}|v_{F_j}|_{1/2,\tau_{0,j}} \le\\
&\le c\overline{\gamma}_S\rho_j H_{D,j}\textstyle\sum_p|v_{F_j^p}|_{1/2,\tau_{0,j}} \approx c\overline{\gamma}_S\rho_j H_{D,j}\sum_p{}_{00}|v_{F_j^p}|_{1/2,F_j^p}\,.
\end{aligned} \tag{5.88}$$

The definition of the face preconditioner, (5.48) and (5.49) yield the estimate

$$\|\mathbf{v}_{F_j}\|_{\mathbf{S}_{F_j}} \leq c\,\overline{\gamma}_S\,\overline{\gamma}_{00} \sum_p \|\mathbf{v}_{\hat{F}_j^p}\|_{\boldsymbol{\mathcal{S}}_{00,\hat{F}_j}}\,. \tag{5.89}$$

Together with the relationships (5.41) and (5.47), used in the definition of the face preconditioner, this bound implies

$$\overline{\gamma}_F \leq c\,\overline{\gamma}_S\,\overline{\gamma}_{00}\,. \tag{5.90}$$

Now, we can directly return to the proof of Theorem 5.2. Indeed, the estimate $c_{P_B} \leq 2$ for the constant in the inequalities (5.27) and (5.28) follows from (5.36) and the definition of the prolongation matrix $\mathbf{P}_{V_B \to V}$. Substituting the estimates, obtained above for $\underline{\gamma}_W, c_{P_W}, c_F, \overline{\gamma}_F$, and the estimate $c_{P_B} \leq 2$ into the estimates (5.75) of Theorem 5.1 completes the proof of Theorem 5.2. □

Suppose that the preconditioners $\boldsymbol{\mathcal{K}}_I$ and $\boldsymbol{\mathcal{S}}_F$, satisfying the assumptions of Theorem 5.2, are such that solving SLAE with these matrix requires not more than $\mathcal{O}(\hbar^{-3})$ a.o. Then, due to Theorem 5.2 and the presented analysis of other components of the BPS preconditioner, the solution of the SLAE with $\boldsymbol{\mathcal{K}}$ as system matrix requires not more than $\mathcal{O}(\hbar^{-3}(1+\log^2\hbar^{-1})^2)$ a.o., see also Remarks 5.1 and 5.2 as well as the earlier discussion of the cost of the prolongation (5.67).

5.2.3 *Subsidiary Results*

The proof of Theorem 5.2, given in the preceding subsection, uses subsidiary results which are summarized below in one theorem and three lemmas. Among them, Theorem 5.3 belongs to [Xu (1989)] and is here presented without proof. It is related to discrete Sobolev type inequalities for the limiting cases, when the imbedding theorem for the Sobolev spaces fails. Another example of such type is inequality (5.95). For subdomains of simpler forms, it can be found in several sources, see, *e.g.*, [Toselli and Widlund (2005)] for references. We prove it for convex hexahedrons under the shape regularity conditions given in Subsection 5.1.1. For such subdomains, we below also prove the bound for the prolongation operator from the wire basket. Besides we prove the bound for the traces of the FE face functions, *i.e.*, having nonzero nodal values only at the nodes living on faces of the subdomains.

Theorem 5.3. *Let $\Omega \subset R^d$ be a bounded Lipschitz domain and $p > 1$. Then there exists a constant $c = c(d,p,\Omega)$ such that the inequality*

$$\|v\|_{C(\overline{\Omega})} \leq c\,\{|\log\epsilon|^{1-\frac{1}{d}}\|v\|_{W_p^{d/p}(\Omega)} + \epsilon^{\lambda}\|v\|_{C^{0,\lambda}(\overline{\Omega})}\} \tag{5.91}$$

holds for all $v \in W_p^{d/p}(\Omega) \cap C^{0,\lambda}(\overline{\Omega})$ *and for any fixed* $\epsilon > 0$ *and* $\lambda \in (0,1]$.

Proof. We refer to Theorem 2.1 in [Xu (1989)] for the proof and also to an earlier proof in [Xu and Long (1988)] for $p = d = 2$ and $\lambda = 1$. □

For $p = d = 2$ and $\lambda = 1$, inequality (5.91) can be written in the form

$$\|v\|_{L_\infty(\Omega)} \le c\,\{|\log\epsilon|^{1/2}\|v\|_{H^1(\Omega)} + \epsilon\|v\|_{W^1_\infty(\Omega)}\}\,. \tag{5.92}$$

Before formulating the next lemma, let us remind that $\mathbb{V}_j(\tau_{0,j})$ stands for the FE space on $\tau_{0,j}$, induced by the FE assemblage satisfying the quasiuniformity conditions, $\mathbb{V}_B(\partial\tau_{0,j})$ is the FE space of traces on $\partial\tau_{0,j}$, $\hat{W}_j$ is the union of edges and vertices of a polygon $\tau_{0,j}$, $\hat{F}_j = \partial\tau_{0,j} \setminus \hat{W}_j$, and $|v|_{0,\hat{W}_j}$ denotes the seminorm

$$|v|^2_{0,\hat{W}_j} = \inf_{c\in R}\|v-c\|^2_{0,\hat{W}_j}\,. \tag{5.93}$$

Lemma 5.6. *For any FE function* $v_W \in \mathbb{V}_W(\tau_{0,j})$, i.e., *vanishing at all nodes which do not belong* $\hat{W}_j$, *and for any* $u_B \in \mathbb{V}_B(\partial\tau_{0,j})$, *the inequalities*

$$|v_W|^2_{1/2,\partial\tau_{0,j}} \le c|v_W|^2_{0,\hat{W}_j}\,, \tag{5.94}$$

and

$$|u_B|^2_{0,\hat{W}_j} \le c(1+\log\hbar^{-1})|u_B|^2_{1/2,\partial\tau_{0,j}} \tag{5.95}$$

are valid with some constants $c = \text{const}$ *which are independent of* $\hbar$.

Before giving the proof, we make one remark. Clearly, (5.94) bounds the prolongation operator $\mathbf{P}_{V_{W_j}\to V_{B_j}}$ defined for one face in (5.67). Indeed, let us denote by $\hat{\mathcal{P}}_{\mathbb{V}_{W_j}\to\mathbb{V}_{B_j}}$ the prolongation operator in the FE subspace $\mathbb{V}_B(\partial\tau_{0,j})$ on the boundary of the reference subdomain and imply by ϕ_W the FE function specified on W_j and equal there to v_W. Then $v_W = \hat{\mathcal{P}}_{\mathbb{V}_{W_j}\to\mathbb{V}_{B_j}}\phi_W$, and a consequence of (5.94) is

$$\mathbf{v}_W^\top\mathbf{S}_W^F\mathbf{v}_W \le |\hat{\mathcal{P}}_{\mathbb{V}_{W_j}\to\mathbb{V}_{B_j}}v_W|^2_{1/2,\partial\tau_{0,j}} \le c|v_W|^2_{0,\hat{W}_j}\,. \tag{5.96}$$

At the same time, (5.95) characterizes traces of FE functions on $\hat{W}_j$.

Proof. In view of the quasiunifomity conditions for FE discretizaton (5.7) or (5.8), the function v_W is distinct from zero only in $\mathcal{O}(\hbar)$-vicinity of the wire basket $\hat{W}_j$. The proof of the bound

$$\|v_W\|^2_{1,\tau_{0,j}} \le c\|v_W\|^2_{0,\hat{W}_j}\,. \tag{5.97}$$

is completed in a similar way with the cases of subdomains of a simpler forms, see *e.g.,* [Bramble *et al.* (1986)]-[Bramble *et al.* (1989)] and [Toselli and Widlund (2005)]. From (5.97), by the trace theorem, it follows that

$$\|v_W\|^2_{1/2,\partial\tau_{0,j}} \le c\|v_W\|^2_{0,\hat{W}_j}\,. \tag{5.98}$$

For obtaining (5.94), it is left to apply the factor space argument.

Inequality (5.95) for subdomains of simpler types can be found in many sources, *e.g.*, [Bramble *et al.* (1989)] and [Bramble and Xu (1991)]. An easy proof of (5.95) for the case $\tau_{0,j} = \tau_0 := \{x : 0 < x_1, x_2, x_3 < 1\}$ is based on inequality (5.92). We consider $u \in H^1(\tau_0)$ and write (5.92) for the point $x = (0, 0, x_3)$, $0 < x_3 < 1$, of the square $\Pi_{x_3} = \{x : 0 < x_1, x_2 < 1,\ x_3 \equiv \text{const}\}$:

$$|u(0,0,x_3)|^2 \le\ c\,\{|\log\epsilon|\,\|u\|^2_{H^1(\Pi_{x_3})} + \epsilon^2\|u\|^2_{W^1_\infty(\Pi_{x_3})}\}\,. \tag{5.99}$$

The inequality

$$\|u\|^2_{0,\mathcal{E}} \le c\,\{|\log\epsilon|\,\|u\|^2_{H^1(\tau_{0,j})} + \epsilon^2\|u\|^2_{W^1_\infty(\tau_{0,j})}\}$$

for the edge $\mathcal{E} = \{x = (0, 0, x_3),\ 0 < x_3 < 1\}$ follows by integrating (5.99) with respect to x_3. In the case of the FE h-version, we can use some trivial inequalities. Namely, if $v \in \mathbb{V}_B(\tau_{0,j})$ and τ is the domain of any finite element of a quasiuniform reference FE assemblage on $\tau_{0,j}$, then

$$\hbar^{3/2}|v|_{L_\infty(\tau)} \le c\|v\|_{0,\tau} \quad \text{and} \quad \hbar^{3/2}|v|_{W^1_\infty(\tau)} \le c\,|v|_{H^1(\tau)} \tag{5.100}$$

for all $v \in \mathbb{V}_B(\tau_{0,j})$, see, *e.g.,* [Ciarlet (1978)] and [Korneev (1977a)]. Now, by taking $\epsilon = \hbar^2$ and applying (5.100), we arrive at the estimate

$$\|u\|^2_{0,\mathcal{E}} \le c\,(1 + |\log\hbar|)\,\|u\|^2_{1,\tau_{0,j}}\,. \tag{5.101}$$

Summation overall edges belonging to $\hat{W}_j$ allows us to conclude that

$$\|u\|^2_{0,\hat{W}_j} \le c\,(1 + |\log\hbar|)\,\|u\|^2_{1,\tau_{0,j}}\,. \tag{5.102}$$

Inequality (5.102) also holds for discrete harmonic functions. Thus, inequality

$$\|u\|^2_{0,\hat{W}_j} \le c\,(1 + |\log\hbar|)\,\|u\|^2_{1/2,\partial\tau_{0,j}} \tag{5.103}$$

directly follows from (5.30). In order to obtain (5.95), it is sufficient to use the definition of the seminorm $|\cdot|_{0,\mathcal{E}}$ and the characterizations of the seminorm in the space $H^{1/2}(\partial\tau_{0,j})$ by means of the relation $|u|^2_{1/2,\partial\tau_{0,j}} \approx \inf_{c\in R^1}\|u + c\|^2_{1/2,\partial\tau_{0,j}}$.

Now we will prove inequality (5.95) for the case of polyhedrons $\tau_{0,j}$, satisfying our assumptions of size and shape quasiuniformity. Again, we start from the consideration of one edge, and, without restriction of generality, we can assume that such an edge is $\mathcal{E} = \{x = (0, 0, x_3),\ 0 < x_3 < 1\}$. For simplicity, the length of the edge can be set to unity, because, according to the shape regularity condition for the reference subdomains, the length of each edge of $\tau_{0,j}$ is $\mathcal{O}(1)$. Let $\pi_{j,\mathcal{E}}$ be the unit cube, which has $\mathcal{E}$ for the edge and which intersection with $\tau_{0,j}$ is the largest. If $\pi_{j,\mathcal{E}} \cap \tau_{0,j} \neq \pi_{j,\mathcal{E}}$, then we consider the least c-vicinity $S_{j,c}$ of $\tau_{0,j}$, which covers $\pi_{j,\mathcal{E}}$ and denote by $\mathcal{T}_{j,\mathrm{ext}}$ an extension of the reference triangulation $\mathcal{T}_{0,j}$ on $S_{j,c}$, satisfying the quasiuniformity conditions (5.7) and (5.8). More exactly, let $\mathcal{T}_{j,\mathrm{ext}}$ be the triangulation, each nest of which has nonempty intersection with $S_{j,c}$ such that $S_{j,c} \subset \tau_{j,\mathrm{ext}}$, where $\tau_{j,\mathrm{ext}}$ is the domain of the extended triangulation. We consider the FE space $\mathbb{V}_j(\tau_{j,\mathrm{ext}})$, induced by the triangulation $\mathcal{T}_{j,\mathrm{ext}}$, and its restriction $\mathbb{V}_j(\pi_{j,\mathcal{E}})$ to the unit cube $\pi_{j,\mathcal{E}}$. Then, for any function $v \in \mathbb{V}_j(\tau_{0,j})$, there exists an expansion $v_{\mathrm{ext}} \in \mathbb{V}_j(\tau_{j,\mathrm{ext}})$, for which

$$\|v_{\mathrm{ext}}\|_{1,\tau_{j,\mathrm{ext}}} \leq c\,\|v\|_{1,\tau_{0,j}}\,. \tag{5.104}$$

We note that, under our assumptions, the constant c is independent of j.

For small $\epsilon = \mathcal{O}(\hbar^k)$, $k > 1$, which we use, inequality (5.92) can be rewritten as

$$\|v\|_{L_\infty(\Omega)} \leq c\,\{|\log\epsilon|^{1/2}\|v\|_{H^1(\Omega)} + \epsilon|v|_{W^1_\infty(\Omega)}\}\,, \tag{5.105}$$

with the term $\epsilon\|v\|_{L_\infty(\Omega)}$ shifted from the right to the left part of (5.92), that changes only the constant. Let $\sqcap_{x_3}$ be the intersection of the plane $x_3 = \mathrm{const}$ with $\pi_{j,\mathcal{E}}$. Then taking square of (5.105), written for $\sqcap_{x_3}$, and integrating over $x_3 \in (0, 1)$, we obtain

$$\begin{aligned}\|v\|^2_{2,\mathcal{E}} \leq\ & c\,\{|\log\epsilon| \int_0^1 \|v_{\mathrm{ext}}\|^2_{1,\sqcap_{x_3}}\,dx_3 + \epsilon^2 \int_0^1 |v_{\mathrm{ext}}|^2_{W^1_\infty(\sqcap_{x_3})}\,dx_3 \leq \\ \leq\ & c\,\{|\log\epsilon|\|v_{\mathrm{ext}}\|^2_{1,\pi_{j,\mathcal{E}}} + \epsilon^2|v_{\mathrm{ext}}|^2_{W^1_\infty(\pi_{j,\mathcal{E}})}\}\,.\end{aligned} \tag{5.106}$$

Inequalities (5.100) hold for any $w \in \mathbb{V}_j(\tau_{j,\mathrm{exp}})$ and $\tau \in \mathcal{T}_{j,\mathrm{ext}}$. Thus, the choice $\epsilon = \hbar^2$ and the use of (5.100) and (5.104) yield (5.101):

$$\|v\|^2_{0,\mathcal{E}} \leq c\,(1 + |\log\hbar|)\,\|v_{\mathrm{ext}}\|^2_{1,\tau_{j,\mathrm{ext}}} \leq c\,(1 + |\log\hbar|)\,\|v\|^2_{1,\tau_{0,j}}\,. \tag{5.107}$$

The transition to (5.95) is the same as for the unit cube $\tau_{0,j} = \tau_0$. □

For simplification of notations, in the next Lemma and mostly in its proof, we omit index j, *i.e.*, τ_0 stands for $\tau_{0,j}$ etc.

Lemma 5.7. *Let τ_0 be a size and shape regular convex polyhedron provided with a quasiuniform triangulation, and $\mathcal{F}$ be any face of it. Further, let us represent a function $v \in \mathbb{V}_B(\tau_0)$ as sum $v = v_F + v_W$ of two functions v_F and $v_W \in \mathbb{V}_B(\partial\tau_0)$ the nodal values of which are not zero only on $\hat{F}$ and $\hat{W}$, respectively. Then*

$$ {}_{00}\|v_F\|^2_{1/2,\mathcal{F}} \le c(1+\log^2 \hbar^{-1})\|v\|^2_{1/2,\partial\tau_0}\,. \tag{5.108}$$

Proof. The proof can be produced with the help of the representation of the norm in $H^{1/2}(\partial\tau_0)$ by the expression

$$\|v\|^2_{1/2,\partial\tau_0} = \int_{\partial\tau_0}\int_{\partial\tau_0} \frac{(v(x)-v(y))^2}{|x-y|^3}\,ds(x)ds(y) + \|v\|^2_{0,\partial\tau_0}, \tag{5.109}$$

in which the double integral for the function $v_F \in \mathbb{V}_B(\partial\tau_0)$, vanishing on $\partial\tau_0 \setminus \mathcal{F}$, can be rewritten in the form

$$\begin{aligned} |v_F|^2_{1/2,\partial\tau_0} = &\int_{\mathcal{F}}\int_{\mathcal{F}} \frac{(v_F(x)-v_F(y))^2}{|x-y|^3}\,ds(x)ds(y) \\ &+2\int_{\mathcal{F}}\int_{\partial\tau_0\setminus\mathcal{F}} \frac{v_F^2(y)}{|x-y|^3}\,ds(x)\,ds(y)\,. \end{aligned} \tag{5.110}$$

Using Lemma 5.6, we get the estimates

$$\begin{aligned} &\int_{\mathcal{F}}\int_{\mathcal{F}} \frac{(v_F(x)-v_F(y))^2}{|x-y|^3}\,ds(x)ds(y) \le |v_F|^2_{1/2,\partial\tau_0} = |v-v_E|^2_{1/2,\partial\tau_0} \\ &\le 2(|v|^2_{1/2,\partial\tau_0} + |v_E|^2_{1/2,\partial\tau_0}) \le c(1+\log h^{-1})|v|^2_{1/2,\partial\tau_0}\,. \end{aligned} \tag{5.111}$$

Thus, it is left to estimate only the second term in the right part of (5.110).

The set $\partial\tau_0\setminus\mathcal{F}$ can be separated in two parts $\mathcal{B}_1$ and $\mathcal{B}_2$ with $\mathcal{B}_1\cap\mathcal{B}_2 = \varnothing$ such that $\overline{\mathcal{B}}_1$ includes the closures of the faces, having with the face $\mathcal{F}$ common vertices or edges and vertices, and $\overline{\mathcal{B}}_2 = \partial\tau_0 \setminus (\overline{\mathcal{B}}_1 \cup \mathcal{F})$. For instance, we have $\mathcal{B}_2 = \varnothing$, if $\tau_{0,j}$ is a tetrahedra, and $\mathcal{B}_2$ contains only the edge opposite to $\mathcal{F}$, if $\tau_{0,j}$ is a cube. Additionally, we subdivide $\mathcal{B}_1$ in two nonoverlapping subsets $\mathcal{B}_1 = \mathcal{B}_{11}\cup\mathcal{B}_{12}$, where $\mathcal{B}_{12}$ contains faces which have common edges with $\mathcal{F}$. Note that $\mathcal{B}_{11} = \varnothing$ for a tetrahedra or a cube. Due to the size and shape regularity of the reference subdomains, we conclude that $\operatorname{dist}[\mathcal{F},\mathcal{B}_2] \ge c_\triangle$ with $c_\triangle = \text{const}$. Therefore,

$$\int_{\mathcal{F}}\int_{\mathcal{B}_2} \frac{u^2(y)}{|x-y|^3}\,ds(x)\,ds(y) \le \frac{c}{c_\triangle^3}\,\|u\|^2_{0,\mathcal{F}}\,. \tag{5.112}$$

Turning to faces in $\mathcal{B}_1$, we can assume without loss of generality that the face $\mathcal{F}$ is in the plane $x_3 \equiv 0$. Let $\mathcal{F}' = \mathcal{F}'_k,\ k = 1, 2, .., \mathbb{k}$, be the faces in $\mathcal{B}_1$ and $\mathcal{L} = \mathcal{L}_k$ be the straight line which is the intersection of the plane, containing the face $\mathcal{F}'$, with the plane $x_3 \equiv 0$. Let for definiteness, $k = 1, 2, .., \mathbb{k}_\circ$ correspond to $\mathcal{F}' = \mathcal{F}'_k \subset \mathcal{B}_{12}$. Taking again into account that $\tau_{0,j}$ have diameters equal to 1 and are shape regular, in a sense explained in Subsection 5.1.1, we can establish the inequality

$$\int_{\mathcal{B}_{12}} \frac{1}{|x-y|^3}\, ds(x) \le \overline{c} \sum_{k=1}^{\mathbb{k}_2} \frac{1}{\operatorname{dist}[y, \mathcal{L}_k]}\,. \tag{5.113}$$

In order to prove (5.113), we consider one face $\mathcal{F}' = \mathcal{F}'_k$, which have a common edge $\mathcal{E}$ with $\mathcal{F}$, and, without loss of generality, we assume that the line $\mathcal{L}$ coincides with the axis x_2. It is easy to see that it is sufficient to consider only the case when the angle between faces $\mathcal{F}$ and $\mathcal{F}'$ is not greater than $\pi/2$. In the opposite case, a stronger bound holds.

Let $z = \mathcal{Z}(x) : \tau_{0,j} \to \tau'_{0,j}$ be a nondegenerate mapping which satisfies the conditions:

A) $\mathcal{Z}(\mathcal{F}) = \mathcal{F}$, whereas the plane of the face $\mathcal{F}'$ is rotated around the axis x_2 at the smallest angle sufficient to make it orthogonal to the plane of the face $\mathcal{F}$,

B) the Jacobian $|\mathbb{J}(\mathcal{Z})|$ of the Jacobi matrix $\mathbb{J}(\mathcal{Z}) = (\partial z_k/\partial x_l)_{k,l=1}^3$ satisfies the inequalities

$$c \le |\mathbb{J}(\mathcal{Z})| \le C\,, \qquad \forall\, x \in \overline{\tau}_{0,j}\,, \tag{5.114}$$

where c and C are positive constants depending only on $\tau_{0,j}$.

For our purpose, it is sufficient to use affine maps. For simplicity, we assume that faces $\mathcal{F}$ and $\mathcal{F}'$ are in the positive parts of the half-spaces with respect to the planes $x_1 \equiv 0$ and $x_3 \equiv 0$, respectively. Let γ denote the angle between these faces. Then we can set $z_1 = x_1 + x_3 \tan(\frac{\pi}{2} - \gamma),\ z_k = x_k,\ k = 2, 3$. Furthermore, let $\mathcal{F}'' = \mathcal{Z}(\mathcal{F}')$ and Π be the smallest rectangle containing $\mathcal{F}''$. Without loss of generality, we assume that the origin of the coordinates x and z is at the vertex of Π, so that $\Pi = \{(x_2, x_3) \in (0, a_2) \times (0, a_3)\}$, where a_2 and a_3 are the sizes of the edges of Π. Then, under these assumptions, we have the following estimates:

$$\begin{aligned}\int_{\mathcal{F}'} \frac{1}{|x-y|^3}\, ds(x) &\le c \int_{\mathcal{F}''} \frac{1}{|z-y|^3}\, ds(z) \\ &\le \int_0^{a_2}\int_0^{a_3} \frac{1}{[y_1^2 + (z_2 - y_2)^2 + z_3^2]^{3/2}}\, dz_3\, dz_2\,,\end{aligned}$$

$$\int_0^{a_3} \frac{1}{(y_1^2+(z_2-y_2)^2+z_3^2)^{3/2}}\,dz_3$$
$$= \frac{z_3}{(y_1^2+(z_2-y_2)^2)[y_1^2+(z_2-y_2)^2+z_3^2]^{1/2}}\Big|_0^{a_3}$$
$$= \frac{a_3}{(y_1^2+(z_2-y_2)^2)[y_1^2+(z_2-y_2)^2+a_3^2]^{1/2}} \le \frac{1}{(y_1^2+(z_2-y_2)^2)},$$

and

$$\int_0^{a_2} \frac{1}{(y_1^2+(z_2-y_2)^2)}\,dz_2 = \frac{1}{y_1}\arctan\frac{z_2-y_2}{y_1}\Big|_0^{a_2} \le \frac{\pi}{y_1}\,.$$

Thus, we have shown that

$$\int_{\mathcal{F}}\int_{\mathcal{F}'} \frac{v_F^2(y)}{|x-y|^3}\,ds(x)\,ds(y) \le c\int_{\mathcal{F}} \frac{v_F^2(y)}{\operatorname{dist}[y,\pounds]}\,dy_1\,dy_2$$
$$= c\int_{\mathcal{F}} \frac{v_F^2(y)}{y_1}\,dy_1\,dy_2. \qquad (5.115)$$

It is easy to see that, in the case of convex-shaped and size-regular polyhedrons $\tau_{0,j}$, the proof practically does not depend on whether $\mathcal{F}' \subset \mathcal{B}_{12}$ or $\mathcal{F}' \subset \mathcal{B}_{11}$. Therefore, (5.115) also holds in the latter case.

Bounding the integral in (5.115) requires the consideration of the situation in the plane containing the face $\mathcal{F}$. In this connection, we often speak about finite elements assuming 2d linear elements belonging to the triangulation of the face $\mathcal{F}$ defined by the 3d triangulation of the reference subdomain $\tau_{0,j}$.

Let $\overline{\mathcal{F}}^{\hbar}$ be the union of the closures of domains of one layer of finite elements with vertices or edges on the boundary $\partial\mathcal{F}$. Quasiuniformity conditions for the reference triangulation allow us to conclude that there exists $c_\dagger = \text{const} > 0$ for which the intersection of the straight line $x_1 \equiv c_\dagger\hbar,\ x_3 \equiv 0$, with $\overline{\mathcal{F}}^{\hbar}$ and with $\overline{\mathcal{F}}$ coincide. We denote this intersection by $\mathcal{L}$. Two cases must be distinguished: $\mathcal{L} \subset \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$ and $\mathcal{L}$ not in $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$, where $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$ is the union of the closures of domains of one layer finite elements with vertices or edges on $\overline{\mathcal{E}}$. We start with the first case. We subdivide $\mathcal{F}$ in two parts $\mathcal{F} = \mathcal{F}_\alpha \cup \mathcal{F}_\beta,\ \mathcal{F}_\alpha = \{x \in \mathcal{F} : x_1 < c_\dagger\hbar\},\ \mathcal{F}_\beta = \mathcal{F} \setminus \overline{\mathcal{F}}_\alpha$. Accordingly we represent the integral in the right part of (5.115) by the sum

$$\mathcal{I} := \int_{\mathcal{F}} \frac{v_F^2}{x_1}\,dx_1dx_2 = \mathcal{I}_\alpha + \mathcal{I}_\beta\,, \qquad (5.116)$$

of the integrals

$$\mathcal{I}_\alpha = \int_{\mathcal{F}_\alpha} \frac{v_F^2(x_1,x_2,0)}{x_1}\,dx_1dx_2 \quad\text{and}\quad \mathcal{I}_\beta = \int_{\mathcal{F}_\beta} \frac{v_F^2(x_1,x_2,0)}{x_1}\,dx_1dx_2\,.$$

According to Lemma 5.8, proved at the end of this section, the first integral $\mathcal{I}_\alpha$ is bounded by the integral along the line $\mathcal{L}$, *i.e.*

$$\mathcal{I}_\alpha \le c \int_{\mathcal{L}} v_F^2(x)\, dx_2\,. \tag{5.117}$$

In turn, the last integral can be estimated in the same way as the integral in the left part of (5.107). Therefore, we have

$$\mathcal{I}_\alpha \le c(1+|\log\hbar|)\|v_F\|^2_{1,\tau_0}\,. \tag{5.118}$$

Indeed, this estimate is obtained by the use of (5.107) as follows:

$$\begin{aligned} \mathcal{I}_\alpha &\le c(1+|\log\hbar|)\|v-v_W\|^2_{1,\tau_0} \le \\ &\le 2c(1+|\log\hbar|)(\|v\|^2_{1,\tau_0}+\|v_W\|^2_{1,\tau_0}) \le (1+\log^2\hbar)\|v\|^2_{1,\tau_0}\,. \end{aligned} \tag{5.119}$$

For bounding integral $\mathcal{I}_\beta$ in (5.116), it is again convenient to use prolongations of FE functions to some vicinity of $\tau_{0,j}$ and then cut out an appropriate cube with the edges of the length $\mathcal{O}(1)$. We use the chosen above system of coordinates $x=(x_1,x_2,x_3)$ such that $\mathcal{F}$ is in the plane (x_1,x_2), $\mathcal{F}''$ is in the plane (x_2,x_3), and the line $\pounds$ is on the axis x_2. Let $\pi_j := (0,\ell)\times(a,b)\times(0,\ell)$, $b=a+\ell$, be the least cube, one face of which is in the plane $x_3\equiv 0$ and covers $\mathcal{F}$. We consider the c-vicinity $S_{c,j}$ of $\tau_{0,j}$ and the extended triangulation $\mathcal{T}_{j,\mathrm{ext}}$, occupying the domain $\tau_{j,\mathrm{ext}}$, where c is such that $\pi_j\subset\tau_{j,\mathrm{ext}}$. Besides, we introduce the space $\mathbb{V}_j(\pi_j)$ as the restriction of $\mathbb{V}_j(\tau_{j,\mathrm{ext}})$ to the cube π_j. The extensions v_{ext} of functions $v\in\mathbb{V}(\tau_{0,j})$ in the space $\mathbb{V}_j(\tau_{j,\mathrm{ext}})$ are assumed to be produced in such a way that the inequality (5.104) holds.

Let, for $x_2\in(a,b)$, each line $x_2\equiv$ const and $x_3\equiv 0$ has nonempty intersection with the face $\mathcal{F}$ with x_1 in the interval denoted $\Lambda=\Lambda(x_2)$. Then

$$\begin{aligned} \mathcal{I}_\beta &:= \textstyle\int_a^b \int_{\Lambda\setminus(0,c_\dagger\hbar)} \frac{v_F^2(x_1,x_2,0)}{x_1}\, dx_1 dx_2 \\ &\le \ln((c_\dagger\hbar)^{-1}) \textstyle\int_a^b \|v_F(x_1,x_2,0)\|^2_{L_\infty(\Lambda)}\, dx_2 \\ &\le 2\ln\left(\tfrac{1}{c_\dagger\hbar}\right)\left\{\textstyle\int_a^b \|v(x_1,x_2,0)\|^2_{L_\infty(\Lambda)}\, dx_2 + \int_a^b \|v_W(x_1,x_2,0)\|^2_{L_\infty(\Lambda)}\, dx_2\right\}, \end{aligned} \tag{5.120}$$

where we declare that $\|\cdot\|_{L_\infty(\Lambda)}=0$, if $\Lambda=\Lambda(x_2)=\varnothing$. For the reason that v_W is distinct from zero only at the nodes of the wire basket, it can be shown that

$$\int_a^b \|v_W(x_1,x_2,0)\|^2_{L_\infty(\Lambda)}\, dx_2 \le c\,\|v\|^2_{2,\hat{W}_j}\,.$$

Combining this estimate with (5.107) yields

$$\int_0^{a_2} \|v_W(x_1,x_2,0)\|^2_{L_\infty(\Lambda)}\,dx_2 \le c\,(1+|\log\hbar|)\,\|v\|^2_{1,\tau_{0,j}}\,. \tag{5.121}$$

The first integral in the right part of (5.120) can be estimated similarly to (5.106). Let $\pi_j(x_2)$ be the cross section of π_j by the plane $x_2 \equiv \text{const}$. Writing (5.105) squared for $\pi_j(x_2)$ and integrating over $x_2 \in (a,b)$ give the following estimates:

$$\begin{aligned}
&\textstyle\int_a^b \|v(x_1,x_2,0)\|^2_{L_\infty(\Lambda(x_2))}\,dx_2 \le \int_a^b \|v_{\exp}(x_1,x_2,0)\|^2_{L_\infty(\Lambda(x_2))}\,dx_2\\
&\le\ c\,\{|\log\epsilon|\textstyle\int_a^b \|v_{\exp}\|^2_{1,\pi_j(x_2)}\,dx_2 + \epsilon^2\int_a^b |v_{\exp}|^2_{W^1_\infty(\pi_j(x_2))}\,dx_2\\
&\le\ c\,\{|\log\epsilon|\|v_{\exp}\|^2_{1,\pi_j} + \epsilon^2|v_{\exp}|^2_{W^1_\infty(\pi_j)}\}\\
&\le\ c\,\{|\log\epsilon|\|v_{\exp}\|^2_{1,\tau_{j,\exp}} + \epsilon^2|v_{\exp}|^2_{W^1_\infty(\tau_{j,\exp})}\}\,.
\end{aligned} \tag{5.122}$$

Inequalities (5.100) are valid for all functions v belonging to the extended FE space, *i.e.*,

$$\hbar^{3/2}|v|_{L_\infty(\tau)} \le c\|v\|_{0,\tau} \quad\text{and}\quad \hbar^{3/2}|v|_{W^1_\infty(\tau)} \le c\,|v|_{H^1(\tau)} \tag{5.123}$$

for all $v \in \mathbb{V}(\tau_{j,\exp})$. We now set $\epsilon = \hbar^2$, apply (5.123) to the second term in the right part, and then apply (5.104), repeating last steps in deriving (5.107). This is reflected in the inequalities

$$\begin{aligned}
\int_a^b \|v(x_1,x_2,0)\|^2_{L_\infty(\Lambda(x_2))}\,dx_2 &\le c\,(1+|\log\hbar|)\,\|v_{\exp}\|^2_{1,\tau_{j,\exp}}\\
&\le c\,(1+|\log\hbar|)\,\|v\|^2_{1,\tau_{0,j}}\,.
\end{aligned} \tag{5.124}$$

Combining inequalities (5.116), (5.119)–(5.121), and (5.124) yields

$$\mathcal{I} \le c\,(1+\log^2\hbar)\,\|v\|^2_{1,\tau_{0,j}}\,. \tag{5.125}$$

From (5.113), (5.115), (5.116) and (5.125), we conclude that

$$\int_{\mathcal{F}}\int_{\mathcal{F}'} \frac{v_F^2(y)}{|x-y|^3}\,ds(x)\,ds(y) \le c\,(1+\log^2\hbar)\|v\|^2_{1,\tau_{0,j}}\,. \tag{5.126}$$

Heretofore, our considerations were restricted by the assumption $\mathcal{L} \subset \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$. Let the angles between edge $\mathcal{E}$ and the adjacent to $\mathcal{E}$ edges be separated from π by some constant $\theta_\circ$, $0 < \theta_\circ \le \pi/4$. Then, for sufficiently small $\hbar$, we can always choose such $c_\dagger = c_\dagger(\theta_\circ) = \text{const}$ that $\mathcal{L} \subset \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$. The mentioned assumption will be violated, if the angles between the adjacent to $\mathcal{E}$ edges and the edge $\mathcal{E}$, or one of these angles, approach π. In general, the

line $\mathcal{L}$ can intersect with several domains $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_k}$ neighboring the domain $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}}$. However, from further considerations, it will became clear that a sufficiently representative case is when the line $\mathcal{L}$ intersects with two domains. In order to retain some notations, we adopt that $\mathcal{E} = \mathcal{E}_1$ and that the line $\mathcal{L}$ intersects with $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_1}$ and $\overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_2}$. For definiteness, we assume that the edges of the face $\mathcal{F}$ are counted counter-clockwise.

We consider triangles with the common vertex at the point $\overline{\mathcal{E}}_k \cap \overline{\mathcal{E}}_{k+1}$ and denote by $\mathcal{S}_k$ the one of their sides in $\mathcal{F}^{\hbar}$, which is closest to the bisector of the angle between $\mathcal{E}_{k-1}$ and $\mathcal{E}_k$. It is convenient to change the definition of the line $\mathcal{L}$. Let $\overline{\mathcal{F}}^{\hbar}_{\dagger} = \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_1} \cup \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_2}$ and $\mathcal{F}^{\hbar}_k$, $k = 1, 2$, be the nonoverlapping subdomains of $\mathcal{F}^{\hbar}_{\dagger}$ separated by the line $\mathcal{S}_1$. We define $\mathcal{L}_{\dagger} \subset \overline{\mathcal{F}}^{\hbar}_{\dagger}$ as the broken line, which on each domain $\overline{\mathcal{F}}^{\hbar}_k$, $k = 1, 2$, is a straight line parallel to $\mathcal{E}_k$ and such that $\mathrm{dist}[y, \partial\mathcal{F}] \geq c_{\dagger}\hbar$, $\forall\, y \in \mathcal{L}_{\dagger}$, for some constant $c_{\dagger} > 0$. Now, the line $\mathcal{L}$ is defined as the intersection of the straight line $x_1 \equiv c_{\dagger}\hbar$ with $\overline{\mathcal{F}}^{\hbar}$. If $\mathcal{L}$ crosses the line $\mathcal{E}_2$, we proceed as follows. The line $\mathcal{L}$ defines the least simply connected mesh domain $\mathcal{F}^{\hbar,2} \subset \mathcal{F}^{\hbar}_2$, the left border of which is $\mathcal{S}_1$ and the right one is defined by the condition that each line $x_2 = \mathrm{const}$, crossing $\mathcal{L}$, crosses as well $\mathcal{L}_{\dagger}$ on $\overline{\mathcal{F}}^{\hbar,2}$. We introduce also the sets $\overline{\mathcal{F}}^{\hbar}(\mathcal{E}) := \overline{\mathcal{F}}^{\hbar}_1 \cup \overline{\mathcal{F}}^{\hbar,2}$ and $Ł_{\dagger} := \mathcal{L}_{\dagger} \cap \overline{\mathcal{F}}^{\hbar}(\mathcal{E})$. If $\mathcal{L}$ does not cross $\mathcal{E}_2$, then we simply set $\mathcal{F}^{\hbar,2} = \mathcal{F}^{\hbar}_2$ with the consequences: $\mathcal{F}^{\hbar}(\mathcal{E}) = \mathcal{F}^{\hbar}_{\dagger}$ and $\mathcal{L}_{\dagger} = Ł_{\dagger}$.

In general, the line $Ł_{\dagger}$ separates each of the domains $\mathcal{F}^{\hbar}_1$ and $\mathcal{F}^{\hbar,2}$ in two parts, the parts adjacent to $\partial\mathcal{F}$ are denoted by Π_1 and Π_2, respectively. Obviously, we have

$$\mathcal{I}_{\alpha} = \int_{\mathcal{F}_{\alpha}} \frac{v_F^2}{x_1}\, dx_1 dx_2 \leq \sum_{k=1,2} \int_{\Pi_k} \frac{v_F^2}{x_1}\, dx_1 dx_2. \tag{5.127}$$

Lemma 5.8 allows us to obtain bounds similar to (5.117) for each integral in the right part. For this reason, we get

$$\mathcal{I}_{\alpha} \leq \int_{Ł_{\dagger}} v_F^2 ds\,,$$

where s is the element of the line $Ł_{\dagger}$. Further transition from this bound to the bound (5.119) and then to the bounds (5.125)–(5.126) are quite similar to the presented above. Thus, summarizing, we obtain

$$\int_{\mathcal{F}} \int_{\mathcal{B}_{12}} \frac{v_F^2(y)}{|x-y|^3}\, ds(x)\, ds(y) \leq c\,(1 + \log^2 \hbar) \|v\|^2_{1,\tau_{0,j}}\,. \tag{5.128}$$

If the part $\mathcal{L}$ of the line $\mathcal{L}_{\infty} := \{x = (c_{\dagger}\hbar, x_2, 0)\}$ with $c_{\dagger}$ defined by the condition

$$\mathcal{L}_{\infty} \cap \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}} = \mathcal{L}_{\infty} \cap \overline{\mathcal{F}}^{\hbar}$$

crosses several domains $\overline{\mathcal{F}}_{\mathcal{E}_k}$, $k = 1, 2, \ldots, \Bbbk_+$, then first of all we redefine $\mathcal{L}$. We consider the domain

$$\overline{\mathcal{F}}^{\hbar}_{\dagger} = \cup_{k=1}^{\Bbbk_+} \overline{\mathcal{F}}^{\hbar}_{\mathcal{E}_k}$$

and subdivide it into $\Bbbk_+$ nonoverlapping subdomains $\mathcal{F}^{\hbar}_k$ by the edges $\mathcal{S}_k$, $k = 1, 2, \ldots, \Bbbk_+ - 1$. The broken line $\mathcal{L}_\dagger$ is defined by the conditions:

α) $\mathcal{L}_\dagger \cap \overline{\mathcal{F}}^{\hbar}_k$ is a straight line, $k = 1, 2, \ldots, \Bbbk_+$, parallel to the edge $\mathcal{E}_k$,

β) $\operatorname{dist}[\mathcal{L}_\dagger, \mathcal{F}^{\hbar}] \geq c_\dagger \hbar$ for some $c_\dagger = \text{const} > 0$.

The right border of the domain $\mathcal{F}^{\hbar, \Bbbk_+}$ is defined in the same way as for $\Bbbk_+ = 2$, and it is adopted

$$\overline{\mathcal{F}}^{\hbar}(\mathcal{E}) = (\cup_{k=1}^{\Bbbk_+ - 1} \overline{\mathcal{F}}^{\hbar}_k) \cup \overline{\mathcal{F}}^{\hbar, \Bbbk_+} .$$

Further considerations toward the bound (5.126) also repeat the way used for the case $\Bbbk_+ = 2$, except that we now have $\Bbbk_+$ integrals instead of two in the right part of (5.127), each of which is bounded by Lemma 5.8 for the first step.

Clearly, the case when the line $x_1 \equiv c_\dagger \hbar$ crosses the domains $\mathcal{F}^{\hbar}_{\mathcal{E}_k}$ on the both sides of $\mathcal{F}^{\hbar}_{\mathcal{E}}$ introduces minor differences in the proof.

It is left to derive the bound

$$\int_{\mathcal{F}} \int_{\mathcal{B}_{11}} \frac{v_F^2(y)}{|x-y|^3}\, ds(x)\, ds(y) \leq c\,(1 + \log^2 \hbar) \|v\|^2_{1,\tau_{0,j}} , \tag{5.129}$$

but since (5.115) holds for $\mathcal{F}' \subset \mathcal{B}_{11}$, the proof closely repeats the proof of (5.128). Indeed, technically, it is a part of the proof of (5.128).

Relations (5.110), (5.111), (5.128), (5.129) and the definition of the norm ${}_{00}\|\cdot\|_{1/2,\mathcal{F}}$ show that

$${}_{00}\|v_F\|_{1/2,\mathcal{F}} \leq c\,(1 + \log^2 \hbar) \|u\|^2_{1,\tau_{0,j}} .$$

Since this bound holds also for discrete harmonic functions, for which $\|v\|_{1,\tau_{0,j}} \leq \|v\|_{1/2,\partial\tau_{0,j}}$, we have completed the proof. □

Now we turn to Lemma 5.8 used in the proof of Lemma 5.7. In the plane of variables $x = (x_1, x_2)$, we consider the quasiuniform triangulation $\Im_\hbar$ which has $\mathcal{E} = \{x : x_1 \equiv 0,\ x_2 \in (a, b),\ -\infty < a < b < \infty\}$ for the part of the boundary and which has triangles δ, having on $\overline{\mathcal{E}}$ a vertex or an edge. The domain of the quasiuniform triangulation is denoted by $\Pi_\hbar$. Two additional conditions are supposed to be fulfilled. 1) For some constant $c_\dagger > 0$, the intersection Ł of the line $x_1 \equiv c_\dagger \hbar$ is connected and crosses all triangles $\overline{\delta}$. 2) If $\overline{\delta}$ has a single vertex on $\overline{\mathcal{E}}$ and E_δ is the edge opposite to this vertex, than $x_1 \geq c_\dagger \hbar$ for $\forall\, x \in E_\delta$. Note that, under condition 1),

condition 2) is fulfilled automatically, if the vertex is in $\mathcal{E}$. However, under condition 1), it can be violated for the triangles having vertices at the ends of $\mathcal{E}$.

Lemma 5.8. *Let $\Pi_{\hbar,\mathcal{E}} = \{x \in \Pi_\hbar : x_1 \le c_\dagger \hbar\}$. Further, let v be a function continuous on $\Pi_\hbar$, linear on each triangle $\delta \in \Im_\hbar$ and satisfying the condition $v\big|_{\mathcal{E}} = 0$. Then*

$$\int_{\Pi_{\hbar,\mathcal{E}}} \frac{v^2(x)}{x_1}\, dx_2 \le c \int_{Ł} v^2(x)\, dx_2\,, \tag{5.130}$$

with some constant c depending only on the quasiuniformity conditions.

Proof. Let $\overline{\mathcal{E}}'$ be the part of $\partial\Pi'_\hbar$, $\Pi'_\hbar = \Pi_\hbar \setminus \Pi_{\hbar,\mathcal{E}}$, containing only whole edges of triangles $\delta \in \Im_\hbar$. Then $\partial\Pi_\hbar \setminus (\mathcal{E} \cup \mathcal{E}')$ contains only two edges of two triangles which have vertices at the ends of $\mathcal{E}$. Obviously, on each group of triangles δ, having one common vertex on $\overline{\mathcal{E}}'$ and edges on $\overline{\mathcal{E}}$, the function v is linear. Therefore, every such cluster of triangles can be considered as one triangle δ'', and, instead of the triangulation $\Im_\hbar$, we can consider the rarefied triangulation $\Im'_\hbar$ consisting of the triangles δ'' and triangles $\delta' = \delta$ having edges on $\mathcal{E}$. Each triangle δ'', which does not have edges in $\partial\Pi_\hbar \setminus (\mathcal{E} \cup \mathcal{E}')$, has two adjacent triangles δ' and vice versa. The conditions of quasiuniformity for the triangulation $\Im_\hbar$ can be written in the similar to (5.7) form

$$\beta^{(1)}\hbar \le h_{i,j}^{(r)} \le \beta^{(2)}\hbar \quad \text{and} \quad \theta \le \theta_i^{(r)} \le \pi - 2\theta\,, \tag{5.131}$$

where $h_{i,j}^{(r)}$ are the lengths of edges, $\theta_i^{(r)}$ are the angles at the vertices of triangles of the triangulation $\Im_\hbar$, and $\beta^{(1)}, \beta^{(2)}, \theta = \text{const} > 0$ are positive constants. In view of (5.131), similar quasiuniformity conditions are retained for the triangulation $\Im'_\hbar$, in general, with other constants, for which, however, we will use the same notations.

Considering one triangle δ', we introduce the notations A' for the length of the line $\mathcal{A}' = Ł \cap \overline{\delta}'$, v_L and v_R for the values of v at the "left" and the "right" ends of $\mathcal{A}'$, respectively, and $\delta'_A := \delta' \cap \Pi_{\hbar,\mathcal{E}}$. For the values of v at the ends of the intersection $\mathcal{A}'_\circ(x_1)$ of the line $x_1 = \text{const}$ with $\overline{\delta}'$, we use the notations $u_L = u_L(x_1)$ and $u_R = u_R(x_1)$, so that $u_L(c_\dagger\hbar) = v_L$ and $u_R(c_\dagger\hbar) = v_R$.

Let the affine mapping $\eta = \eta(x)$, $x \in \delta'$, does not change the positions of the two vertices of the triangle δ', which are on $\overline{\mathcal{E}}'$, but shifts the third

vertex of δ' along $\overline{\mathcal{E}}$ to such nearest position that the triangle $\Delta' = \eta(\delta')$ becomes rectangular. Obviously, we have

$$\int_{\delta'_A} \frac{v^2}{x_1}\, dx = \int_{\Delta'_A} \frac{v^2}{\eta_1}\, d\eta\,,$$

where $\Delta'_A = \eta(\delta'_A)$. Hence, without loss of generality, we can adopt that δ' itself is a rectangular triangle with one cathetus crossing the left end of $\mathcal{A}'$, and that $x_2 = 0$ for this end. Therefore, we have

$$v(x) = u_L(x_1) + \frac{u_R(x_1) - u_L(x_1)}{A_\circ(x_1)}\, x_2\,, \qquad x_2 \in (0, A_\circ(x_1))\,,$$

with $A_\circ(x_1)$ being the length of $\mathcal{A}'_\circ(x_1)$, and obviously such a function v satisfies the inequalities

$$\underline{c} A_\circ(x_1)(u_L^2(x_1)+u_R^2(x_1)) \le \int_0^{A_\circ(x_1)} v^2(x) dx_2 \le \overline{c} A_\circ(x_1)(u_L^2(x_1)+u_R^2(x_1)) \tag{5.132}$$

with generic positive constants $\underline{c}$ and $\overline{c}$. Using this estimate, we get

$$\begin{aligned}
\mathcal{I}_{\delta'_A} &:= \int_{\delta'_A} \frac{v^2(x)}{x_1}\, dx \\
&= \int_0^{c_\dagger \hbar} \frac{1}{x_1} \int_0^{A_\circ(x_1)} [u_L(x_1) + \frac{u_R(x_1) - u_L(x_1)}{A_\circ(x_1)}\, x_2]^2\, dx_2 dx_1 \\
&\le \overline{c} \int_0^{c_\dagger \hbar} \frac{A_\circ(x_1)}{x_1} [u_L^2(x_1) + u_R^2(x_1)] dx_1\,.
\end{aligned} \tag{5.133}$$

Using the relation

$$(u_L(x_1), u_R(x_1), A_\circ(x_1)) = \mu(x_1)(v_L, v_R, A')\,, \tag{5.134}$$

with $\mu(x_1) = x_1/(c_\dagger \hbar)$, and estimates (5.132)–(5.133), we obtain

$$\begin{aligned}
\mathcal{I}_{\delta'_A} &\le \overline{c}\, \frac{A'}{(c_\dagger \hbar)^3} [v_L^2(x_1) + v_R^2] \int_0^{c_\dagger \hbar} x_1^2\, dx_1 \\
&\le \overline{c}\, \frac{A'}{3} [v_L^2(x_1) + v_R^2] \le \frac{\overline{c}}{3\underline{c}} \int_{\mathcal{A}'} v^2 dx_2\,.
\end{aligned} \tag{5.135}$$

Now we will consider an arbitrary triangle δ'', which has an edge on $\mathcal{E}$, the length of which will be denoted by B''. We also introduce the notations: $\delta''_A := \delta'' \cap \Pi_{\hbar,\mathcal{E}}$, A'' for the length of the line $\mathcal{A}'' = Ł \cap \overline{\delta}''$, and $A''_\circ(x_1)$ for the intersection of the line $x_1 = \text{const}$ with $\overline{\delta}''$. It is clear that, on each line $x_1 = \text{const}$ crossing $\overline{\delta}''_A$, the function v has a constant value $v = (v_A/c_\dagger \hbar) x_1$ and

$$A''_\circ(x_1) = A'' + \frac{B'' - A''}{c_\dagger \hbar} x_1\,.$$

Therefore,

$$\mathcal{I}_{\delta''_A} := \int_{\delta''_A} \frac{v^2(x)}{x_1}\,dx = (\frac{v_A}{c_\dagger\hbar})^2 \int_0^{c_\dagger\hbar} x_1[A'' + \frac{B''-A''}{c_\dagger\hbar}x_1]dx_1$$
$$= \left[\frac{A''}{2} + \frac{B''-A''}{3}\right] v_A^2 = \frac{1}{2}\int_{\mathcal{A}''} v^2 dx_2 + \frac{B''-A''}{3}v_A^2\,. \quad (5.136)$$

As was mentioned earlier, there is at least one triangle δ'' adjacent to δ', and we have $x_1 = c_\dagger\hbar$ for all $x \in \mathcal{A}'$. Furthermore, the quasiuniformity conditions (5.131) hold. In view of this, for any triangle δ', the value A' is bounded from below and from above, *i.e.*,

$$c_-\hbar \le A' \le c_+\hbar\,, \quad (5.137)$$

with positive constants c_- and c_+ depending only on the constants $\beta^{(1)}$ and θ from (5.131). We pick up a triangle δ' adjacent to δ'', take into account that $v_A = v_L$ or $v_A = v_R$, and estimate the second term in (5.136) as follows:

$$\frac{B''-A''}{3}v_A^2 \le \frac{\hbar}{3}[v_L^2 + v_R^2] \le \frac{A'}{3c_-}[v_L^2 + v_R^2] \le \frac{1}{3c_-\underline{c}}\int_{\mathcal{A}'} v^2 dx_2\,, \quad (5.138)$$

where we used (5.131), (5.137) and (5.132). Now, it only remains to combine (5.135), (6.105), (5.138) in order to arrive at the bound

$$\int_{\delta_A^+} \frac{v^2(x)}{x_1}\,dx \le \max[\frac{1}{2}, \frac{1}{3\underline{c}}(1+\frac{1}{c_-})]\int_{\mathcal{A}^+} v^2(x)\,dx_2\,, \quad (5.139)$$

where $\delta_A^+ = \delta'_A \cup \delta''_A$ and $\overline{\mathcal{A}}^+ = \overline{\mathcal{A}}' \cup \overline{\mathcal{A}}''$. If the number n of triangles δ' and δ'' in the triangulation $\Im'_\hbar$ is even, then the domain $\Pi_{\hbar,\mathcal{E}}$ is decomposed in $n/2$ nonoverlapping pairs δ_A^+, and the bound (5.130) holds with the same constant c as in (5.139). In the case of an odd n, $\Pi_{\hbar,\mathcal{E}}$ is decomposed in $(n-1)/2$ nonoverlapping pairs δ_A^+ and one overlapping pair, so that c is increased by the factor 2. □

Chapter 6

DD Algorithms for Discretizations with Chaotically Piecewise Variable Orthotropism

This chapter illustrates that Dirichlet–Dirichlet type DD methods, the key problem of which is Schur complement preconditioning, can be robust in situations, which seem difficult for providing this property. In designing the interface components, we often relayed on the fact that the problem of construction of boundary norms for discrete harmonic FE functions and the problem of Schur complement preconditioning are closely interrelated. If the equation is uniformly elliptic in Ω or at least in each subdomain of decomposition, the domain Ω is sufficiently good, the decomposition and FE meshes are quasiuniform, then the situation is well understood. In this case, for each subdomain, the energy space is equivalent to $H^1(\Omega_j)$, and a proper boundary norm for discrete harmonic FE functions is the norm in the space $H^{1/2}(\partial\Omega_j)$. This fact together with relatively simple, at least in 2d, discrete representations of the norms in $H^{1/2}(\partial\Omega_j)$ for the traces of FE functions on boundaries of 2d domains considerably simpifies designing computationally cheap interface preconditioners for DD algorithms. The situation changes significantly when some or all of the conditions mentioned above are lacking. One of such situations arises when subdomains of the domain decomposition becomes shape irregular, the elliptic equation as well as the finite element mesh are characterized by independent subdomainwise variable orthotropisms. A discrete model problem of this sort, which we aim to solve in this section, is the following. Suppose that the domain $\Omega = (0,1) \times (0,1)$ is decomposed into the subdomains

$$\Omega_j = (z_{1,j_1-1}, z_{1,j_1}) \times (z_{2,j_2-1}, z_{2,j_2}), \quad j = (j_1, j_2), \tag{6.1}$$

by the rectangular coarse *decomposition grid*

$$x_k = z_{k,j_k}, \quad j_k = 0, 1, \ldots, J_k, \quad z_{k,0} = 0, \quad z_{k,J_k} = 1, \tag{6.2}$$

which is arbitrary in each direction x_k, $k = 1, 2$. We set $H_{k,j_k} = z_{k,j_k} - z_{k,j_k-1} > 0$. The decomposition grid is imbedded in the nonuniform

rectangular finer *source grid*, termed also *fine grid*:

$$x_k = x_{k,i_k}, \quad i_k = 0, 1, \ldots, N_k, \quad x_{k,0} = 0, \quad x_{k,N_k} = 1, \tag{6.3}$$

such that $x_{k,\gamma_k} = z_{k,j_k}$ for some numbers $\gamma_k = \varkappa_k(j_k)$. In the case of rectangular finite elements, this grid serves as a *discretization grid*. For simplicity, it is assumed to be uniform on each subdomain and otherwise arbitrary. Its mesh sizes are denoted by $h_{k,j_k} = H_{k,j_k}/n_{k,j_k}$, where n_{k,j_k} is the number of the source grid intervals on the decomposition grid interval (z_{k,j_k-1}, z_{k,j_k}).

It is worth underlining that according to the above definitions there are no restrictions on the finite aspect ratios of the nests of the decomposition and discretization grids. The decomposition and discretization grids are illustrated in Figure 6.1

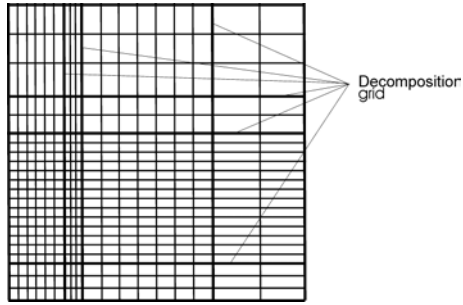

Fig. 6.1 DD and subdomainwise uniform rectangular discretization grids.

We now consider the variational problem

$$\alpha_\Omega(u, v) = \langle f, v\rangle, \quad \forall\, v \in \mathring{H}^1(\Omega), \tag{6.4}$$

where

$$\alpha_\Omega(u, v) \equiv \int_\Omega \nabla u(x) \cdot \boldsymbol{\rho}(x)\, \nabla v(x)\, dx,$$

and $\boldsymbol{\rho} = \operatorname{diag}[\rho_1, \rho_2]$ is a 2×2 diagonal matrix with positive piecewise constant functions $\rho_k(x)$ such that $\rho_k(x) = \rho_{k,j} = \text{const}$ for any $x \in \Omega_j$. The FE approximation of the variational problem (6.4) yields the large-scale system of finite element equations

$$\mathbf{Ku} = \mathbf{f}. \tag{6.5}$$

An intriguing question arises whether there exists a DD algorithm which is fast and robust with respect to the bad parameters H_{k,j_k}, $\rho_{k,j}$ and h_{k,j_k}.

The Results of this chapter allow us to give a positive answer. In Section 6.3, we present the DD preconditioner-solver $\boldsymbol{\mathcal{K}}_{\rm DD}$ that is of linear arithmetical complexity and robust. Under the assumption that number of subdomains is constant, *i.e.* $J_1 J_2 = \text{const}$, the total arithmetical work can be estimated by $\mathcal{O}(N_\Omega)$, where N_Ω is the total number of degrees of freedom. More precisely, let the matrix $\boldsymbol{\mathcal{I}}\,[\mathbf{K}, \boldsymbol{\mathcal{K}}_{\rm DD}, n_\epsilon]$ be defined according to Definition 2.1, and let ops$[\cdot]$ be the the number of arithmetic operations for the procedure in the square brackets. Then the determination of an approximate solution $\mathbf{u}_{\rm DD} = \boldsymbol{\mathcal{I}}^{-1}\mathbf{f}$ of the system (6.5) of FE equations, which has the accuracy $\|\mathbf{u} - \mathbf{u}_{\rm DD}\|_{\mathbf{K}} \le \epsilon$, by means of the preconditioned two level method with Chebyshev iteration parameters or by the PGG method requires not more than

$$\text{ops}\left[\,(\boldsymbol{\mathcal{I}}[\mathbf{K}, \boldsymbol{\mathcal{K}}_{\rm DD}, n_\epsilon])^{-1}\mathbf{f}\,\right] \le c\,\sqrt{\text{cond}\,[\boldsymbol{\mathcal{K}}_{\rm DD}^{-1}\mathbf{K}]}\cdot \text{ops}\left[\boldsymbol{\mathcal{K}}_{\rm DD}^{-1}\mathbf{f}\,\right] = \mathcal{O}(N_\Omega) \tag{6.6}$$

arithmetical operations, where c is a generic constant. The bound (6.6) holds for any rigth-hand side vector $\mathbf{f}$. Note that the right part of (6.6) does not depend on any of the bad parameters contributing to the orthotropism of the discretization on subdomains. This bound also retains in the case of J_k growing not too fast with the respect to growing N_k. We remind that we omit the multipier $\ln \varepsilon^{-1}$ in the bounds of arithmetical work at the iterative solution of systems of algebraic equations, where ε denotes the relative accuracy.

The model problem (6.4) represents a family of boundary value problems to which the results can be easily generalized. Indeed, Ω can be a union of any number of nests of the arbitrary orthogonal decomposition mesh, the discretization mesh can be a tensor product of two piecewise quasiuniform meshes, while the bilinear form can be of the form

$$\alpha_\Omega(u, v) \equiv \int_\Omega \nabla u(x)\cdot \wp(x)\,\nabla v(x)\,dx, \tag{6.7}$$

where the 2×2 matrix $\wp(x)$ satisfies the spectral inequalities

$$\mu_1\,\boldsymbol{\rho}(x) \le \wp(x) \le \mu_2\,\boldsymbol{\rho}(x) \tag{6.8}$$

with positive constants $\mu_1, \mu_2 = \text{const} > 0$, and with the piecewise constant matrix $\boldsymbol{\rho}$ described above. Clearly, instead of bilinear rectangular finite elements, we can use linear triangular finite elements with vertices at the nodes of the discretization mesh. For both types of discretizations, the preconditioners will be often defined by means of triangular elements, which provide simpler explicit representations.

A crucial role in the DD solver for the problem under consideration, satisfying the bound (6.6), is played by the inter-subdomain Schur complement preconditioner-solver. It results from the secondary inexact iterative procedure in which two other explicitely given Schur complement preconditioners are employed. We term these preconditioners primal Schur complement preconditioner, preconditioner-multiplicator and preconditioner-solver. The choice of the later two preconditioners is essential for robustness and optimality of the DD solver. Their definition and justification occupy a considerable part of this chapter. In this arrangement, the preconditioner-multiplicator is obtained by means of a new boundary norm for discrete harmonic finite element functions on slim rectangles. This is the reason why we start with the introduction of a suitable norm for the traces of discrete harmonic functions on the boundary of a single slim rectangular subdomain that can have an arbitrary finite aspect ratio of its edges.

6.1 Single Slim Domain

We start our considerations from a boundary norm for harmonic functions on the slim rectangular domain $\Omega = (0,1) \times (0,\epsilon)$, which is equivalent to the H^1-norm uniformly in $\epsilon \in (0,1]$. This norm was derived by [Korneev *et al.* (2007)] along the lines which following [Maz'ya and Poborchi (1998)] in their study of boundary norms for bad domains. This norm is strongly influenced by the values of ϵ. Therefore, it is a shape dependent norm. Then we establish that the shape dependent boundary norm, here denoted by $]|\cdot|[_{\partial\Omega}$, see (6.14), remains a proper norm for discrete harmonic finite element functions on a shape regular meshes. This is done by means of the special quasi-interpolation operator by [Scott and Zhang (1990)] defined for functions from $H^1(\Omega)$, see also Lemma 5.2 of the preceding section. After that, in a way similar to deriving discrete norms equivalent to $H^{1/2}$-norms for FE functions, the shape dependent norm $]|\cdot|[_{\partial\Omega}$ is replaced by simpler finite-difference norms. It is worth noting that the results, obtained for discrete boundary norms, reflect the fact sometimes referred to as *the absorbtion of singularities* and especially known from applications of the boundary element methods. In the case under consideration it reveals in the following property. Suppose that we have a spectrally equivalent preconditioner for the boundary Schur complement resulted from a FE discretization on a quasiuniform mesh. Then, this preconditioner retains the spectral equivalence to the Schur complement generated on any shape

regular mesh which is imbedded in some quasiuniform mesh and coincides with it on the boundary. Even more general meshes are admitted.

6.1.1 *Boundary Norms for Harmonic Functions*

Before formulating the main result of this paragraph, we recall the definition of the $H^{1/2}$-norm. Let Γ be a rectifiable curve in R^2. A function $f \in L_2(\Gamma)$ belongs to the space $H^{1/2}(\Gamma)$, if the norm

$$\|f\|_{1/2,\Gamma} = \left(\|f\|^2_{0,\Gamma} + [f]^2_\Gamma\right)^{1/2}$$

is finite, where $[f]_\Gamma = |f|_{1/2,\Gamma}$ is the seminorm defined by

$$[f]^2_\Gamma = \int_\Gamma \int_\Gamma \frac{|f(p) - f(q)|^2}{|p-q|^2} ds_p ds_q, \tag{6.9}$$

and ds_p, ds_q are the length elements of Γ.

The following "Poincaré inequality" holds for all $f \in H^{1/2}(\Gamma)$:

$$\|f - \overline{f}\|_{0,\Gamma} \le (\operatorname{mes}\Gamma)^{1/2} [f]_\Gamma, \tag{6.10}$$

where $(\operatorname{mes}\Gamma)$ is the length of Γ and

$$\overline{f} = \frac{1}{(\operatorname{mes}\Gamma)} \int_\Gamma f(p) ds_p$$

is the mean value of f on Γ. Inequality (6.10) is a simple consequence of Hölder's inequality and the above definition.

The theorem stated below is due to [Aronszajn (1955)], [Babich and Slobodetski (1956)], and [Gagliardo (1957)], see also [Maz'ya and Poborchi (1998)].

Theorem 6.1. *Let Ω be a planar, simply connected domain with Lipschitz boundary $\Gamma = \partial\Omega$. Then every function $u \in H^1(\Omega)$ has a well-defined trace $f = u|_{\partial\Omega}$ on $\partial\Omega$. The space of all these traces coincides with $H^{1/2}(\partial\Omega)$. Moreover, the norm $\|f\|_{1/2,\partial\Omega}$ is equivalent to the norm $\inf\{\|u\|_{1,\partial\Omega} : u \in H^1(\Omega),\ u|_{\partial\Omega} = f\}$, and there exists a linear continuous extension operator $E : H^{1/2}(\partial\Omega) \to H^1(\Omega)$.*

We remind that, for positive quantities a and b, $a \asymp b$ stands for two-sided inequalities that hold up to positive generic constants or constants depending on secondary parameters.

Corollary 6.1. *Under the assumptions of Theorem 6.1, the following equivalence relation holds:*

$$\inf\{\|\nabla u\|_{0,\Omega} : u \in H^1(\Omega),\ u|_{\partial\Omega} = f\} \asymp [f]_{\partial\Omega}, \tag{6.11}$$

i.e. there exist positive generic constants c_1 *and* c_2 *such that*

$$c_1[f]_\Gamma \le \inf\{\|\nabla u\|_{0,\Omega} : u \in H^1(\Omega),\ u|_\Gamma = f\} \le c_2[f]_\Gamma, \quad \forall f \in H^{1/2}(\Gamma).$$

Proof. The equivalence

$$|u|_{1,\Omega} \le \inf_{\underline{c}\in R} \|u - \underline{c}\,\|_{1,\Omega} \le c\,|u|_{1,\Omega}$$

is not obvious only in its right part that follows by the Poincaré inequality

$$\|v\|^2_{0,\Omega} \le c_0(\int_\Omega v\,dx)^2 + c_1 \int_\Omega \nabla v \cdot \nabla v\,dx, \quad \forall\ v \in H^1(\Omega)\,. \tag{6.12}$$

Now, if $u \in H^1(\Omega)$, $u|_{\partial\Omega} = f$, then by Theorem 6.1 and equivalence (6.12), we have

$$[f]_{\partial\Omega} = \inf_{\underline{c}\in R} [f - \underline{c}\,]_{\partial\Omega} \le c \inf_{\underline{c}\in R} \|u - \underline{c}\,\|_{1,\Omega} \le c\,|u|_{1,\Omega}\,.$$

In order to check the reverse inequality, we fix some $f \in H^{1/2}(\partial\Omega)$ and put $u = \overline{f} + E(f - \overline{f})$, where E is the extension operator from Theorem 6.1. Then $u \in H^1(\Omega)$, $u|_{\partial\Omega} = f$ and the estimate

$$\|\nabla u\|_{L_2(\Omega)} \le \mathrm{c}\,\|f - \overline{f}\|_{1/2,\partial\Omega}$$

is true by the same Theorem 6.1. It remains to observe that the right part of the last inequality is not greater than $c[f]_{\partial\Omega}$ in view of (6.10). □

Let $\Omega = (0,1)\times(0,\epsilon)$ and ϵ, δ satisfy $0 < \epsilon, \delta \le 1$. For the trace of some functions $v \in H^1(\Omega)$ on $\partial\Omega$, we consider two norms and two seminorms. The norm and the seminorm of one pair is defined by the minimization of the properly scaled H^1-norm and H^1-seminorm

$$|\!|\!|\, v \,|\!|\!|^2_{\partial\Omega} = \inf_{\phi|_{\partial\Omega}=v} \left(\frac{\delta^2}{\epsilon^2}\|\phi\|^2_{0,\Omega} + \|\nabla\phi\|^2_{0,\Omega}\right) \text{ and } |\!|\, v \,|\!|^2_{\partial\Omega} = \inf_{\phi|_{\partial\Omega}=v} \|\nabla\phi\|^2_{0,\Omega}, \tag{6.13}$$

respectively, among all functions $\phi \in H^1(\Omega)$ coinciding with a given function v on the boundary on $\partial\Omega$. Another pair, denoted by $]|\cdot|[_{\partial\Omega}$ and $]\cdot[^2_{\partial\Omega}$,

is defined by the expressions

$$
\begin{aligned}
]|v|[^2_{\partial\Omega} &= \delta^2\epsilon^{-1}\|v\|^2_{0,\partial\Omega} +]v[^2_{\partial\Omega},\\
]v[^2_{\partial\Omega} &= \epsilon^{-1}\int_0^1 (v(x_1,\epsilon)-v(x_1,0))^2 dx_1\\
&\quad + \int_0^1\int_{|x_1-y_1|\le\epsilon} \frac{(v(x_1,0)-v(y_1,0))^2}{(x_1-y_1)^2} dx_1dy_1\\
&\quad + \int_0^1\int_{|x_1-y_1|\le\epsilon} \frac{(v(x_1,\epsilon)-v(y_1,\epsilon))^2}{(x_1-y_1)^2} dx_1dy_1\\
&\quad + \int_{\Gamma_0}\int_{\Gamma_0} \frac{(v(s)-v(\overline{s}))^2}{(s-\overline{s})^2} ds d\overline{s}\\
&\quad + \int_{\Gamma_1}\int_{\Gamma_1} \frac{(v(s)-v(\overline{s}))^2}{(s-\overline{s})^2} ds d\overline{s}\,.
\end{aligned}
\tag{6.14}
$$

Here $\Gamma_0 = \{x \in \partial\Omega : x_1 < \epsilon\}$, $\Gamma_1 = \{x \in \partial\Omega_\epsilon : x_1 > 1-\epsilon\}$ and ds, $d\overline{s}$ are the length elements of $\partial\Omega$. The set Γ_1 is symmetric to Γ_0 with respect to the line $x_1 \equiv 1/2$, see Figure 6.2 on page 159. We will term $]|\cdot|[_{\partial\Omega}$ and $]\cdot[_{\partial\Omega}$ the shape dependent norm and seminorm for boundary functions, respectively.

Theorem 6.2. *For the traces of functions from the space $H^1(\Omega)$ on $\partial\Omega$, the seminorm and the norm (6.13) are equivalent to the seminorm and the norm (6.14), respectively, uniformly in $\epsilon, \delta \in (0,1]$.*

Proof. In the proof, sometimes it is more convenient for us to use the notations (x,y) for the space variables instead of $x = (x_1, x_2)$. The similarity transformation with coefficient ε^{-1} transfers the rectangle $\Omega = (0,1)\times(0,\epsilon)$ to a long rectangle $\Omega = (0,N)\times(0,1)$, with $N = \varepsilon^{-1}$, and the desired equivalence relations take the forms

$$
\inf\{\|\nabla u\|_{0,\Omega} : u \in H^1(\Omega),\ u|_{\partial\Omega} = f\} \asymp \langle f\rangle, \tag{6.15}
$$

and

$$
\inf\{\delta^2\|u\|^2_{0,\Omega} + \|\nabla u\|^2_{0,\Omega} : u \in H^1(\Omega),\ u|_{\partial\Omega} = f\} \asymp \delta^2\|f\|^2_{0,\partial\Omega} + \langle f\rangle^2, \tag{6.16}
$$

where

$$
\langle f\rangle^2 = \sum_{i=0}^{1} \iint\limits_{\{t,\tau\in(0,N):|t-\tau|<1\}} \frac{|f(t,i)-f(\tau,i)|^2}{(t-\tau)^2}\, dtd\tau +
$$

$$+\int_0^N |f(t,1)-f(t,0)|^2 dt + \sum_{i=0,N} \iint\limits_{\Gamma_i\times\Gamma_i} \frac{|f(p)-f(q)|^2}{|p-q|^2}\, ds_p ds_q,$$

$\Gamma_0 = \{(x,y)\in\partial\Omega : x<1\}$, $\Gamma_N = \{(x,y)\in\partial\Omega : x>N-1\}$ and ds_p, ds_q are the length elements of $\partial\Omega$ at the points p and q. We can assume without loss of generality that $N=\varepsilon^{-1}$ is an integer[1], $N>2$.

Let $u\in H^1(\Omega)$, $u|_{\partial\Omega}=f$. Now we establish the estimate

$$\langle f\rangle \prec \|\nabla u\|_{0,\Omega}, \tag{6.17}$$

which is a part of the equivalence (6.15). Since u is absolutely continuous on almost all vertical intervals, it follows that

$$f(t,1)-f(t,0)=\int_0^1 \frac{\partial u}{\partial\tau}(t,\tau)d\tau \quad \text{for a.e. } t\in(0,N).$$

By using Hölder's inequality and by integrating over $(0,N)$, we find that

$$\int_0^N |f(t,1)-f(t,0)|^2 dt \le \int_0^N dt \int_0^1 \left|\frac{\partial u}{\partial\tau}(t,\tau)\right|^2 d\tau \le \|\nabla u\|^2_{0,\Omega}\,. \tag{6.18}$$

Next, (6.11) implies the estimate

$$[f]^2_{\Gamma_0} \le c \iint\limits_{(0,1)\times(0,1)} |\nabla u(x,y)|^2 dxdy\,. \tag{6.19}$$

The estimate, which appears if the integration is made over $\Gamma_N\times\Gamma_N$ on the left of (6.19) and over $(N-1)\times(0,1)$ on the right of (6.19), is verified in the same way.

If we fix $i=0$ or $i=1$, then

$$\iint\limits_{\{t,\tau\in(0,N):|t-\tau|<1\}} |f(t,i)-f(\tau,i)|^2 \frac{dtd\tau}{(t-\tau)^2}$$

$$=\sum_{k=0}^{N-1}\int_k^{k+1} dt \int_{\{\tau\in(0,N):|t-\tau|<1\}} |f(t,i)-f(\tau,i)|^2 \frac{d\tau}{(t-\tau)^2} \le$$

$$\le \sum_{k=1}^{N-1} \iint\limits_{\Delta_k\times\Delta_k} |f(t,i)-f(\tau,i)|^2 \frac{dtd\tau}{(t-\tau)^2}\,,$$

[1]If N is not an integer and $\mathrm{int}\lfloor N\rfloor$ is its integer part, we can transform Ω on the rectangle $(0,\mathrm{int}\lfloor N\rfloor+1)\times(0,1)$ with the aid of transformation $\Omega(x,y)\mapsto(\xi,\eta)$, $\xi=(\mathrm{int}\lfloor N\rfloor+1)N^{-1}x$, $\eta=y$ with coefficient $(\mathrm{int}\lfloor N\rfloor+1)N^{-1}\in(1,1+\varepsilon)$.

where $\Delta_1 = (0,2)$, $\Delta_k = (k-1, k+2)$ for $k = 2, \ldots, N-2$ and $\Delta_{N-1} = (N-2, N)$. By Theorem 6.1 and Corollary 6.1, the general term of the last sum is dominated by $c\,\|\nabla u\|^2_{0,\Omega_k}$ with $\Omega_k = \Delta_k \times (0,1)$. The sum does not exceed $c\,\|\nabla u\|^2_{0,\Omega}$, because every point of Ω belongs to at most to three rectangles Ω_k. Combining this with (6.18), (6.19) yields (6.17).

The following estimate

$$\|f\|^2_{0,\partial\Omega} \le \sum_{k=0}^{N-1} \|u\|^2_{0,\partial Q_k},$$

is obviously true for $u \in H^1(\Omega)$ with $u|_{\partial\Omega} = f$, where $Q_k = (k, k+1)\times(0,1)$. Using Theorem 6.1 for each square Q_k, we can dominate the last sum by

$$c \sum_{k=0}^{N-1} \|u\|^2_{1,Q_k} = c\|u\|^2_{1,\Omega}\,.$$

A combination of the latter with (6.17) results in the estimate

$$\delta^2\|f\|^2_{0,\partial\Omega} + \langle f\rangle^2 \le c\,\left(\delta^2\|u\|^2_{0,\Omega} + \|\nabla u\|^2_{0,\Omega}\right)\,. \tag{6.20}$$

To conclude the proof of (6.15) and (6.16), we should construct an extension $u \in H^1(\Omega)$ of a function $f \in H^{1/2}(\partial\Omega)$ for which the reverse inequalities (6.17) and (6.20) are valid. The required extension will be defined step by step. We note that most functions specified at some interval or a domain and vanishing at their boundaries are considered, when necessary, as continued by zero outside on the needed range of space variables. By $L^1_{\infty,\omega}(-1,1) = \{v \in L^1_\infty(-1,1) : v = 0 \text{ for } x \notin (-1+\omega, 1-\omega)\}$ we denote the subspace of functions from the space $L^1_\infty(-1,1)$, vanishing outside of the interval $(-1+\omega, 1-\omega)$ for some fixed $\omega \in (0, 1/2)$. For the first step, we consider a function μ such that

$$\mu \in L^1_{\infty,\omega}(-1,1),\quad 0 \le \mu \le 1,\quad \sum_{k=-\infty}^{\infty} \mu(t-k) = 1,\quad t \in R^1.$$

For instance, μ can be a continuous piecewise linear function. Let $\mu_k(t) = \mu(2t-k)$ for $t \in R^1$, $k = 0, \pm1, \pm2, \ldots$. Then the collection $\{\mu_k\}$ is a uniform partition of unity on R^1 subordinate to the covering by intervals $\gamma_k = ((k-1)/2, (k+1)/2)$. In particular

$$\sum_{k=1}^{2N-1} \mu_k(t) = 1 \quad \text{for} \quad t \in [1/2, N-1/2]\,.$$

We also introduce a function $\lambda \in L^1_{\infty,\omega/2}(-1,1))$ with the property $\lambda\mu = \mu$. Let $\lambda_k(t) = \lambda(2t-k)$, $t \in R^1$, $k = 0, \pm1, \ldots$. It is clear that $\mu_k\lambda_k = \mu_k$ for

all integer k. It is worth emphasizing that constants in the bounds below may depend on the choice of the function μ, defining the partition of unity, and the function λ, but we do not specially mention this again.

Fix $i = 0$ or $i = 1$. Let $\gamma_{k,i} = \gamma_k \times i$, where $\gamma_k = ((k-1)/2, (k+1)/2)$, $k = 1, \dots, 2N-1$, and let $\bar{f}_{k,i}$ denote the mean value of f on $\gamma_{k,i}$. Put $g_k = \gamma_k \times (0,1)$. The function $\mu_k \cdot (f - \bar{f}_{k,i})$ is distinct from zero on $\gamma_{k,i}$. We extend this function by zero on $\partial g_k \setminus \gamma_{k,i}$. Since the support of μ_k is distant from the endpoints of γ_k, we have

$$\|\mu_k(f - \bar{f}_{k,i})\|_{1/2,\partial g_k} \le c\,\|\mu_k(f - \bar{f}_{k,i})\|_{1/2,\gamma_{k,i}}\,.$$

By Theorem 6.1, there exists a linear continuous extension operator $E_k : H^{1/2}(\partial g_k) \to H^1(g_k)$. Combining this with the above inequality, we provide an extension $\mu_k(f - \bar{f}_{k,i}) \mapsto w_{k,i}$ from the boundary segment $\gamma_{k,i}$ into the square g_k such that

$$\|w_{k,i}\|_{1,g_k} \le c\,\|\mu_k(f - \bar{f}_{k,i})\|_{1/2,\gamma_{k,i}} \text{ for } i = 0,1,\ \ k = 1, \dots, 2N-1\,. \quad (6.21)$$

Let

$$u_i = v_i + w_i, \quad (6.22)$$

with

$$v_i(x,y) = \sum_{k=1}^{2N-1} \bar{f}_{k,i}\mu_k(x) \quad \text{and} \quad w_i(x,y) = \sum_{k=1}^{2N-1} \lambda_k(x) w_{k,i}(x,y), \quad (6.23)$$

where $(x,y) \in \Omega$ and $i = 0,1$. For convenience, we accept the convention that $\lambda_k(x)w_{k,i}(x,y) = 0$ outside g_k. It easily follows from our construction that $u_i(x,i) = f(x,i)$ for $i = 0,1$ and $x \in (1/2, N-1/2)$.

Some corrections are required to satisfy boundary conditions near vertical segments $0 \times (0,1)$, $N \times (0,1)$. Let $\bar{f}_0$ be the mean value of f on the broken line Γ_0 and $\bar{f}_N$ the mean value on Γ_N. We extend the function $\mu_0(f - \bar{f}_0)$ by zero from the broken line Γ_0 to ∂Q_0, where $Q_0 = (0,1) \times (0,1)$, and apply Theorem 6.1 to extend the trace of this function on ∂Q_0 from ∂Q_0 into Q_0. This way provides an extension

$$H^{1/2}(\Gamma_0) \supset \mu_0(f - \bar{f}_0) \mapsto U \in H^1(Q_0) \quad (6.24)$$

such that

$$\|U\|_{1,Q_0} \le c\,\|\mu_0(f - \bar{f}_0)\|_{1/2,\Gamma_0}\,. \quad (6.25)$$

In the same way an extension $V \in H^1(Q_{N-1})$ of the trace $\mu_{2N}(f - \bar{f}_N)\big|_{Q_{N-1}} \in H^{1/2}(\Gamma_N)$ is defined, where $Q_{N-1} = (N-1,1) \times (0,1)$, and the following estimate holds:

$$\|V\|_{1,Q_{N-1}} \le c\,\|\mu_{2N}(f - \bar{f}_N)\|_{1/2,\Gamma_N}\,. \quad (6.26)$$

Furthermore, we define

$$W_0(x,y) = \bar{f}_0\mu_0(x) + \lambda_0(x)U(x,y), \quad (x,y) \in \Omega, \tag{6.27}$$

and

$$W_N(x,y) = \bar{f}_N\mu_{2N}(x) + \lambda_{2N}(x)V(x,y), \quad (x,y) \in \Omega, \tag{6.28}$$

where $\lambda_0 U = 0$ outside $(0,1/2)\times(0,1)$ and $\lambda_{2N}V = 0$ outside $(N-1/2,N)\times(0,1)$.

Now the function

$$u(x,y) = W_0(x,y) + W_N(x,y) + u_0(x,y) + y(u_1(x,y) - u_0(x,y))$$

can be considered as a candidate for the required extension u of f. We claim that this is really the case, *i.e.*, $u|_{\partial\Omega} = f$ and the following estimates are true:

$$\|\nabla u\|_{0,\Omega} \le c\,\langle f\rangle \tag{6.29}$$

and

$$\left(\delta^2\|u\|^2_{0,\Omega} + \|\nabla u\|^2_{0,\Omega}\right) \le c\,(\delta^2\|f\|^2_{0,\partial\Omega} + \langle f\rangle^2)\,. \tag{6.30}$$

As we noted above, the functions u_0 and u_1, introduced by (6.22)–(6.23), satisfy $u_i(x,i) = f(x,i)$ for $i = 0,1$, $x \in (1/2, N-1/2)$. Since $W_0(x,y) = W_N(x,y) = 0$ for the same x, we have $u(x,i) = f(x,i)$ for $x \in (1/2, N-1/2)$, $i = 0,1$. Let $x \in [0,1/2)$. Then, we have

$$\begin{aligned} u(x,0) = u_0(x,0) + W_0(x,0) &= \bar{f}_{1,0}\mu_1(x) + \lambda_1(x)\mu_1(x)(f(x,0) - \bar{f}_{1,0}) \\ &+\bar{f}_0\mu_0(x) + \lambda_0(x)\mu_0(x)(f(x,0) - \bar{f}_0)\,. \end{aligned}$$

Since $\lambda_k\mu_k = \mu_k$, we obtain

$$u(x,0) = \mu_1(x)f(x,0) + \mu_0(x)f(x,0) = f(x,0)\,.$$

Further, obviously $\lambda_0(0) = \mu_0(0) = 1$, from where we get

$$u(0,y) = W_0(0,y) = \mu_0(0)\bar{f}_0 + \lambda_0(0)\mu_0(0)(f(0,y) - \bar{f}_0) = f(0,y)\,.$$

In an analogous way one can check that $u(x,1) = f(x,1)$ for $x \in (0,1/2)$ and $u(x,y) = f(x,y)$ for $(x,y) \in \partial\Omega$, $x > N - 1/2$. Thus, u satisfies the boundary condition $u|_{\partial\Omega} = f$.

Turning to the proof of (6.29), we observe that the definition of u implies the estimate

$$\|\nabla u\|_{0,\Omega} \le \|\nabla W_0 + \nabla W_N + \nabla v_0\|_{0,\Omega} + \|v_1 - v_0\|_{1,\Omega} + \sum_{i=0,1}\|w_i\|_{1,\Omega}, \tag{6.31}$$

where v_i and w_i are given by (6.23). Below we bound each term on the right of (6.31).

Let

$$G_k = (k/2, (k+1)/2) \times (0,1), \quad k = 0, \ldots, 2N-1\,.$$

Since $W_N|_{G_0} = 0$, from definitions (6.23) and (6.27), we obtain the estimates

$$\|\nabla W_0 + \nabla W_N + \nabla v_0\|_{0,G_0} = \|\nabla W_0 + \nabla v_0\|_{0,G_0}$$

$$\leq \|\bar{f}_{1,0}\mu_1' + \bar{f}_0\mu_0'\|_{0,G_0} + \|\nabla(\lambda_0 U)\|_{0,G_0}$$

$$\leq \|\mu_0'(\bar{f}_0 - \bar{f}_{1,0})\|_{0,G_0} + c\,\|U\|_{1,G_0}\,.$$

At the last step, we have used that $\mu_0(x)+\mu_1(x) = 1$ for $x \in [0, 1/2]$. Next, the first term on the right can be estimated as follows:

$$\begin{aligned} c\,|\bar{f}_0 - \bar{f}_{1,0})| &\leq c\,\|\bar{f} - f(\cdot,0)\|_{0,(0,1/2)} + c\,\|f(\cdot,0) - \bar{f}_{1,0}\|_{0,(0,1/2)} \\ &\leq c\,\|f - \bar{f}_0\|_{0,\Gamma_0} + c\,\|f(\cdot,0) - \bar{f}_{1,0}\|_{0,(0,1/2)}\,. \end{aligned}$$

According to the inequality (6.10), the sum of the last two summands does not exceed $c\,[f]_{\Gamma_0}$, where $[\cdot]_{\Gamma_0}$ is the seminorm (6.9). Hence, we have established the estimate

$$\|\mu_0'(\bar{f}_0 - \bar{f}_{1,0})\|_{0,G_0} \leq c\,[f]_{\Gamma_0}.$$

Furthermore, the application of estimate (6.25) yields

$$\|U\|_{H^1(G_0)} \leq c\,\|\mu_0(f - \bar{f}_0)\|_{1/2,\Gamma_0}.$$

Note that μ_0 is a smooth function with a compact support in Γ_0. Therefore, the last norm is not greater than $c\,\|f - \bar{f}_0\|_{1/2,\Gamma_0}$ which, according to (6.11), does not exceed $c\,[f]_{\Gamma_0}$. Thus, we have verified the estimate

$$\|\nabla W_0 + \nabla W_N + \nabla v_0\|_{0,G_0} \leq c\,[f]_{\Gamma_0}. \tag{6.32}$$

The estimate

$$\|\nabla W_0 + \nabla W_N + \nabla v_0\|_{0,G_{2N-1}} \leq c\,[f]_{\Gamma_N} \tag{6.33}$$

is verified in an analogous way. The only difference is that we should now apply inequality (6.26) instead of (6.25) in order to bound the norm of the last term on the right part of (6.28).

If $x \in (1/2, N-1/2)$, then $W_0(x,y) = W_N(x,y) = 0$. Therefore,

$$\|\nabla W_0 + \nabla W_N + \nabla v_0\|^2_{0,((1/2,N-1/2)\times(0,1))} = \sum_{k=1}^{2N-2} \|\nabla v_0\|^2_{0,G_k}.$$

Clearly,

$$v_0|_{G_k} = \bar{f}_{k,0}\mu_k + \bar{f}_{k+1,0}\mu_{k+1} = \bar{f}_{k,0} + (\bar{f}_{k+1,0} - \bar{f}_{k,0})\mu_{k+1},$$

and hence

$$|\nabla v_0| \le c\,|\bar{f}_{k+1,0} - \bar{f}_{k,0}| \quad \text{on } G_k.$$

This implies that

$$\|\nabla v_0\|^2_{0,G_k} \le c \sum_{j=k}^{k+1} \|f(\cdot,0) - \bar{f}_{j,0}\|^2_{0,\gamma_j},$$

with $\gamma_j = ((j-1)/2, (j+1)/2)$. Using again the Poincaré inequality (6.10), we arrive at

$$\|\nabla v_0\|^2_{0,G_k} \le c \sum_{j=k}^{k+1} \iint\limits_{\gamma_j\times\gamma_j} \frac{|f(t,0) - f(\tau,0)|^2}{(t-\tau)^2}\, dtd\tau\ .$$

After summation over $k = 1, \dots, 2N-2$, we obtain

$$\int_{1/2}^{N-1/2} \int_0^1 |\nabla v_0(x,y)|^2 dxdy \le c \iint\limits_{\omega} \frac{|f(t,0) - f(\tau,0)|^2}{(t-\tau)^2}\, dtd\tau\,, \qquad (6.34)$$

with $\omega = \{t, \tau \in (0,N) : |t-\tau| < 1\}$. To bound $\|v_1 - v_0\|_{1,\Omega}$, we first write

$$v_1(x,y) - v_0(x,y) = \sum_{k=1}^{2N-1} (\bar{f}_{k,1} - \bar{f}_{k,0})\mu_k(x), \quad (x,y) \in \Omega,$$

which follows from (6.23). The last sum contains at most two nonzero summands for each fixed $x \in (0,1)$. Therefore, we get the estimates

$$\|v_1 - v_0\|^2_{1,\Omega} \le c \sum_{k=1}^{2N-1} |\bar{f}_{k,1} - \bar{f}_{k,0}|^2$$

$$\le c \sum_{k=1}^{2N-1} \int_{\gamma_k} \left(|f(x,0) - \bar{f}_{k,0}|^2 + |f(x,1) - f(x,0)|^2 + |f(x,1) - \bar{f}_{k,1}|^2\right) dx\,.$$

In view of (6.10), we have

$$\int_{\gamma_k} |f(x,i) - \bar{f}_{k,i}|^2 dx \le c \iint\limits_{\gamma_k\times\gamma_k} \frac{|f(t,i) - f(\tau,i)|^2}{(t-\tau)^2}\, dtd\tau$$

and, consequently,

$$\begin{aligned} \|v_1 - v_0\|^2_{1,\Omega} &\le c \textstyle\int_0^1 |f(x,1) - f(x,0)|^2 dx + \\ &+ c \textstyle\sum_{i=0,1} \iint\limits_{\{t,\tau\in(0,1):|t-\tau|<1\}} \frac{|f(t,i)-f(\tau,i)|^2}{(t-\tau)^2}\, dtd\tau\,. \end{aligned} \qquad (6.35)$$

We now bound the last term in (6.31). It follows from (6.21) and (6.23) that

$$\|w_i\|^2_{1,\Omega} \le c \sum_{k=1}^{2N-1} \|\lambda_k w_{k,i}\|^2_{1,\Omega}$$
$$\le c \sum_{k=1}^{2N-1} \|w_{k,i}\|^2_{1,g_k} \le c \sum_{k=1}^{2N-1} \|\mu_k(f - \bar{f}_{k,i})\|^2_{1/2,\gamma_{k,i}}.$$

Since μ_k has compact support in γ_k and μ_k' is uniformly bounded, the general term of the last sum does not exceed $c\,\|f - \bar{f}_{k,i}\|^2_{1/2,\gamma_{k,i}}$ which is dominated by $c\,[f]^2_{\gamma_{k,i}}$ because of inequality (6.10). Thus, we have

$$\|w_i\|^2_{1,\Omega} \le c \sum_{k=1}^{2N-1} \iint_{\gamma_k\times\gamma_k} \frac{|f(t,i) - f(\tau,i)|^2}{(t-\tau)^2}\, dtd\tau$$
$$\le \iint_{\{t,\tau\in(0,1):|t-\tau|<1\}} \frac{|f(t,i) - f(\tau,i)|^2}{(t-\tau)^2}\, dtd\tau\,.$$

A combination of the last estimate with (6.31)–(6.35) gives (6.29).

In order to conclude the proof of the lemma, it remains to check (6.30). Clearly (6.30) follows from (6.29) and the estimate

$$\|u\|_{0,\Omega} \le c\,(\|f\|_{0,\partial\Omega} + \langle f\rangle), \tag{6.36}$$

where u is the extension of f constructed above. From the definition of u, we get the estimate

$$\|u\|_{0,\Omega} \le \|v_0\|_{0,\Omega} + \|W_0\|_{0,\Omega} + \|W_N\|_{0,\Omega} + \|v_1 - v_0\|_{0,\Omega} + \sum_{i=0,1} \|w_i\|_{0,\Omega}.$$

According to what has been said above, the two last terms are dominated by the right part of (6.29). Thus, (6.36) is a consequence of the estimate

$$\|v_0\|_{0,\Omega} + \|W_0\|_{0,\Omega} + \|W_N\|_{0,\Omega} \le c\,(\|f\|_{0,\partial\Omega} + \langle f\rangle) \tag{6.37}$$

to be checked below. Indeed, it follows from (6.23), that

$$\|v_0\|^2_{0,\Omega} \le c \sum_{k=1}^{2N-1} |\bar{f}_{k,0}|^2 \le c \sum_{k=1}^{2N-1} \|f(\cdot,0)\|^2_{0,\gamma_k} \le c \int_0^1 |f(x,0|^2 dx.$$

Next, by using (6.25) and (6.27), we obtain

$$\|W_0\|_{0,\Omega} \le c\,|\bar{f}_0| + c\,\|\mu_0(f - \bar{f}_0)\|_{1/2,\Gamma_0}.$$

The first term on the right does not exceed $c\,\|f\|_{L_2(\Gamma_0)}$, while the last term is not greater than $c\,\|f - \bar{f}_0\|_{1/2,\Gamma_0}$ which, in view of inequality (6.10), is dominated by $c\,[f]_{\Gamma_0}$. Therefore, we have

$$\|W_0\|_{0,\Omega} \le \|f\|_{1/2,\Gamma_0}.$$

In the same way, from (6.26) and (6.28), we deduce the estimate

$$\|W_N\|_{0,\Omega} \le \|f\|_{1/2,\Gamma_N},$$

which concludes the proof of (6.37) and Theorem 6.2. □

6.1.2 *Boundary Norms for Discrete Harmonic Functions in Slim Rectangles*

Let us consider the slim domain $\Omega = (0,1) \times (0,\epsilon)$, and let us introduce the rectangular mesh

$$x_k \equiv x_{k,l},\ l = 1, 2, \ldots, n_k, \quad x_{k,0} = 0, \quad x_{1,n_1} = 1, \quad x_{2,n_2} = \epsilon,$$

with the step-size $h_{k,l} = x_{k,l} - x_{k,l-1},$ satisfying the quasiuniformity conditions

$$\underline{c}h \le h_{k,l} \le \overline{c}h, \tag{6.38}$$

with positive constants $\underline{c}$ and $\overline{c}$. Further, we consider the FE space $\mathcal{V}(\Omega)$ of continuous on Ω functions which are bilinear on each rectangular element of the mesh. The space of traces of functions from $\mathcal{V}(\Omega)$ on $\partial\Omega$ is denoted by $\mathcal{V}_{\rm tr}(\partial\Omega)$.

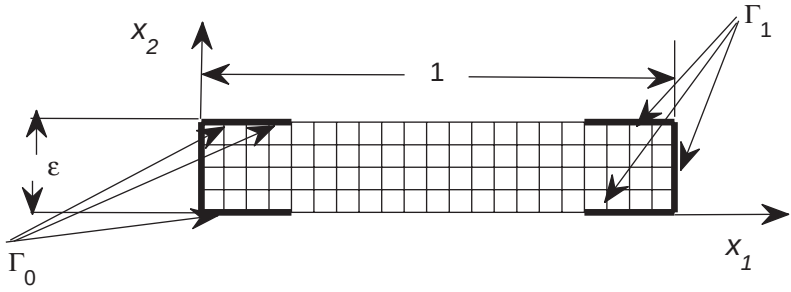

Fig. 6.2 High aspect ratio rectangular domain triangulated by a square mesh.

In this chapter, we are primarily concerned with discretizations on rectangular meshes. However, most of the results are applicable to much more general meshes and FE discretizations. For instance, the above described mesh may represent a *skeleton mesh* of some sort or *skeleton triangulation*, while calculations are performed on a finer mesh, termed the *source mesh*. The latter may

α) be finer in the interior of the domain,
β) have the same with the skeleton mesh trace on the boundary and
γ) cover the skeleton mesh, whereas

δ) the skeleton mesh itself may be a general quasiuniform, quadrangular unstructured mesh with the mesh parameter h.

Obviously, the skeleton mesh can be a triangular quasiuniform mesh as well as the source mesh. In order to simplify explicit representations of the boundary norms for FE functions and solution procedures, besides the presence of mesh nodes at the vertices of Ω, it is worth to assume that there are also mesh nodes at the ends of the sets $\Gamma_k,\ k = 0, 1$, and that the number of nodes on the opposite edges of Ω are equal. The sets $\Gamma_k,\ k = 0, 1$, were introduced in the previous subsection and are shown in Fig. 6.2.

For the traces of FE functions $v \in \mathcal{V}_{\rm tr}(\partial\Omega)$, we also define norm and seminorm

$$\begin{aligned} &\| v \|^2_{h,\partial\Omega} = \inf_{\phi\in\mathcal{V}(\Omega)\,;\,\phi|_{\partial\Omega}=v} \left(\delta^{-2}\|\phi\|^2_{0,\Omega} + \|\nabla\phi\|^2_{0,\Omega}\right) \ \text{and} \\ &| v |^2_{h,\partial\Omega} = \inf_{\phi\in\mathcal{V}(\Omega)\,;\,\phi|_{\partial\Omega}=v} \|\nabla\phi\|^2_{0,\Omega}, \end{aligned} \tag{6.39}$$

in which, in distinction with (6.13), the infimum is taken over the subspace of FE functions.

Theorem 6.3. *Let the FE space $\mathcal{V}(\Omega)$ be induced by the quasiuniform triangulation with the mesh parameter h or by its refinement satisfying $\alpha)$–$\gamma)$. Then, for any $v \in \mathcal{V}_{\rm tr}(\partial\Omega)$, the norms and seminorms defined by (6.39) and (6.14), respectively, are equivalent uniformly in $h > 0$ and $\epsilon, \delta \in (0, 1]$.*

Proof. The proof is based, firstly, on the similar result on the equivalence of the norms (6.13)–(6.14) for the traces of functions from space $H^1(\Omega)$, as formulated in Theorem 6.2, and, secondly, on the quasi-interpolation results of Lemma 5.2 and Lemma 6.1 given below after the proof of the theorem.

Since $\mathcal{V}(\Omega) \subset H^1(\Omega)$, we have the inequalities

$$]|v|[_{\,\partial\Omega} \prec |\, v \,\|_{\partial\Omega} \le \| v \|_{h,\partial\Omega}, \quad \forall\ v \in \mathcal{V}_{\rm tr}(\partial\Omega), \tag{6.40}$$

with the first one following from Theorem 6.2 and the definitions of the norms $\| \cdot |_{\,\partial\Omega}$ and $\| \cdot \|_{h,\partial\Omega}$. For the proof of the opposite bound

$$\| v \|_{h,\partial\Omega} \prec]|v|[_{\partial\Omega}, \quad \forall\ v \in \mathcal{V}_{\rm tr}(\partial\Omega), \tag{6.41}$$

it is sufficient to use additionally Lemma 6.1.

Indeed, let $\mathcal{H}(\Omega)$ be the subspace of $\mathcal{V}(\Omega)$, induced by the skeleton quasiuniform triangulation, and let $v \in \mathcal{V}_{\rm tr}(\partial\Omega)$. Suppose also that $v_{\rm inf} \in H^1(\Omega)$ and $v_{d/{\rm inf}} \in \mathcal{V}(\Omega)$ are the functions at which the infimums in the

first relationships of (6.13) and (6.39), respectively, are reached. Let also $\widetilde{v}$ be the interpolation of $v_{\inf}$ from the space $\mathcal{H}(\Omega)$ satisfying i) and ii) of Lemma 6.1. First of all, we note that, according to Lemma 6.1, we have

$$(\delta\epsilon^{-1})^2\|\widetilde{v}\|^2_{0,\Omega} + \|\nabla\widetilde{v}\|^2_{0,\Omega} \prec (\delta\epsilon^{-1})^2\|v_{\inf}\|^2_{0,\Omega} + \|\nabla v_{\inf}\|^2_{0,\Omega}\,. \tag{6.42}$$

Therefore, we can write

$$\begin{aligned} |\,v\,|^2_{h,\partial\Omega} &:= (\delta\epsilon^{-1})^2\|v_{d/\inf}\|^2_{0,\Omega} + \|\nabla v_{d/\inf}\|^2_{0,\Omega} \\ &\prec (\delta\epsilon^{-1})^2\|\widetilde{v}\|^2_{0,\Omega} + \|\nabla\widetilde{v}\|^2_{0,\Omega} \\ &\prec (\delta\epsilon^{-1})^2\|v_{\inf}\|^2_{0,\Omega} + \|\nabla v_{\inf}\|^2_{0,\Omega} \prec]|v|[_{\,\partial\Omega}, \end{aligned} \tag{6.43}$$

where the first inequality follows by the definition of $|\,v\,|^2_{h,\partial\Omega}$, the second one results from the definition of the same norm and the inclusion $\mathcal{H}(\Omega) \subset \mathcal{V}(\Omega)$, the third inequality is simply (6.42), and the last one is a consequence of Theorem 6.2. □

Now we are going to formulate the interpolation result used in the proof above. We suppose that $\mathcal{V}(\Omega)$ is the FE space induced by the first-order quadrangular finite elements. By means of Lemma 5.2, we can define the quasi-interpolation operator $\Pi_{\square} v : H^1(\Omega) \mapsto \mathcal{V}(\Omega)$. For that purpose, we define a triangulation of Ω by subdividing each quadrangular element of the mesh in two triangles by one of the diagonals of the element. Then we introduce the space $\mathcal{V}_{\triangle}(\Omega)$ of continuous piecewise linear functions induced by this triangulation. For each $v \in H^1(\Omega)$, we define $\Pi_{\square} v$ by the equality $(\Pi_{\square} v)(x^{(i)}) = (\mathcal{I}_h v)(x^{(i)})$ for all nodes $x^{(i)}$ of the triangulation, where $\mathcal{I}_h$ is the quasi-interpolation operator of [Scott and Zhang (1990)], see page 115. It is easy to see that $\Pi_{\square}$ is not a projection operator, but it retains other properties of the operator $\mathcal{I}_h$.

Lemma 6.1. *For any $v \in H^1(\Omega)$, the interpolation $\Pi_{\square} v \in \mathcal{V}(\Omega)$ is such that*

i) if $v|_{\partial\Omega} \in \mathcal{V}_{\rm tr}(\partial\Omega)$, then the traces of $\Pi_{\square} v$ and v on the boundary $\partial\Omega$ coincide,

ii) the interpolation satisfies the stability estimates

$$|\,\Pi_{\square} v\,|_{1,\Omega} \prec |\,v\,|_{1,\Omega}, \quad \|\,\Pi_{\square} v\,\|_{1,\Omega} \prec \|\,v\,\|_{1,\Omega}, \tag{6.44}$$

iii) and the approximation estimates

$$\|v - \Pi_{\square} v\|_{t,\Omega} \prec h^{s-t}\|v\|_{s,\Omega}, \quad t = 0, 1,\ s = 1, 2\,. \tag{6.45}$$

Proof. Let $v_\square \in \mathcal{V}(\Omega)$ and $v_\triangle \in \mathcal{V}_\triangle(\Omega)$ coincide at the nodes. Then, as it is known, see, *e.g.,* [Korneev (1977b)], the equivalences

$$|\, v_\square \,|_{1,\Omega} \prec |\, v_\triangle \,|_{1,\Omega} \prec |\, v_\square \,|_{1,\Omega} \text{ and } \|\, v_\square \,\|_{1,\Omega} \prec \|\, v_\triangle \,\|_{1,\Omega} \prec \|\, v_\square \,\|_{1,\Omega} \quad (6.46)$$

hold. Combining (6.46) with Lemma 5.2 gives (6.44). In order to get (6.45), for any $v \in H^s(\Omega)$, $s = 1, 2$, we set $v_\triangle = \Pi_h v$ and $v_\square = \Pi_\square v$. We consider first the case $s = 1$. Statement c) of Lemma 5.2 yields

$$\|v - v_\square\|_{t,\Omega} \le \|v_\triangle - v_\square\|_{t,\Omega} + \|v - v_\triangle\|_{t,\Omega} \prec \|v_\triangle - v_\square\|_{t,\Omega} + h^{s-t}\|v\|_{s,\Omega} \quad (6.47)$$

for $t = 0, 1$ and $s = 1, 2$. Let $\square_l$ be the domain of a quadrangular finite element. Since $v_\triangle$ is the Lagrange interpolation of $v_\square$, for estimation of the first summand in the right part of (6.47), we can use the interpolation estimate elementwise

$$\|v_\triangle - v_\square\|_{t,\square_l} \prec h^{2-t}|v_\square|_{2,\square_l}, \quad (6.48)$$

then the inverse inequality

$$|v_\square|_{2,\square_l} \prec h^{k-2}|v_\square|_{k,\square_l}, \quad k = 0, 1, \quad (6.49)$$

and come to

$$\|v_\triangle - v_\square\|_{t,\Omega} \prec h^{1-t}|v_\square|_{1,\Omega}\,.$$

Application of (6.44) to the last inequality transforms it into the next one

$$\|v_\triangle - v_\square\|_{t,\Omega} \prec h^{1-t}|v|_{1,\Omega},$$

substitution of which into (6.47) proves (6.45) for $s = 1$.

Turning to the case $s = 2$, we introduce $\widetilde{v}_\triangle$ and $\widetilde{v}_\square$ which are the Lagrange interpolations of v belonging to $\mathcal{V}_\triangle(\Omega)$ and $\mathcal{V}(\Omega)$, respectively, and start from the inequality

$$\|v - v_\square\|_{t,\Omega} \le \|v - \widetilde{v}_\square\|_{t,\Omega} + \|\delta\|_{t,\Omega}, \quad (6.50)$$

with $\delta = \widetilde{v}_\square - v_\square$. For the first term in the right part of it, we can use the interpolation error bound

$$\|v - \widetilde{v}_\square\|_{t,\Omega} \prec h^{2-t}|\, v \,|_{2,\Omega}\,. \quad (6.51)$$

Since functions $v_\triangle$ and $v_\square$ coincide at the FE nodes, as well as functions v and $\widetilde{v}_\square$, we have

$$\delta(x^{(i)}) = \widetilde{v}_\square(x^{(i)}) - v_\square(x^{(i)}) = v(x^{(i)}) - v_\triangle(x^{(i)})\,.$$

In view of this equality, the properties *a*) of the operator $\mathcal{I}_h$ defined in (5.34), and the Cauchy inequality, we can write

$$|\delta(x^{(i)})| = |\int_{F_i} \theta_i(v - \widetilde{v}_\triangle)\,dx| \le \|\theta_i\|_{0,F_i}\|v - \widetilde{v}_\triangle\|_{0,F_i}, \quad (6.52)$$

where F_i stands for $\tau_i^{(d-1)}$ introduced in the definition of $\mathcal{I}_h$. In general, the function θ_i in (5.33) has the form

$$\theta_i = \frac{1}{\operatorname{mes}_{d-1}(F_i)} \sum_1^d a_k \lambda_k^{(i)} \tag{6.53}$$

with coefficients a_k which are independent of the form of the $(d-1)$-dimensional simplex F_i. In order to prove (6.53), it is sufficient to write down the system of algebraic equations defining the coefficients b_k in the expansion

$$\theta_i = \sum_1^d b_k \lambda_k^{(i)} .$$

From (6.53), it follows that

$$\max\nolimits_{x\in\bar{F}_i} |\,\theta_i| \prec h^{1-d},$$

and by means of (6.52), the trace and approximation theorems, we have

$$|\delta(x^{(i)})| \prec \|v - \widetilde{v}_\Delta\|_{0,F_i} \prec \|v - \widetilde{v}_\Delta\|_{1,\tau_{r(i)}} \prec h|v|_{2,\tau_{r(i)}}, \tag{6.54}$$

where $\tau_{r(i)}$ is any triangle of the triangulation having the face F_i. Now, we note that

$$\|\delta\|_{t,\tau_r} \prec h^{0.5d-t} \max_{x^{(i)}\in\bar{\tau}_r} |\delta(x^{(i)})| .$$

If the maximum is reached at $x^{(j)}$, then, due to (6.54), we have

$$\|\delta\|_{t,\tau_r} \prec h^{2-t} \sum |v|_{2,\tau_{r(j)}} . \tag{6.55}$$

The number of coinciding numbers $r(j)$ is bounded by d. From this fact and from (6.55), we get

$$\|\delta\|_{t,\Omega} \prec h^{2-t} \sum |v|_{2,\Omega} . \tag{6.56}$$

Now, estimate (6.45) for $s = 2$ is the consequence of (6.50), (6.51) and (6.56). This completes the proof of the lemma. □

Obviously, the norm (6.14) defines some matrix $\mathbb{B}_{\partial\Omega}$ such that $\|\mathbf{v}\|_{\mathbb{B}_{\partial\Omega}} =]v[_{\partial\Omega}, \forall\, \mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\partial\Omega)$. Let also the FE matrix $\mathbf{A}$ be induced by the Dirichlet integral over Ω and the FE space $\mathcal{V}(\Omega)$ with the nodal basis functions. Representing it in the block form

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_I & \mathbf{A}_{I,B} \\ \mathbf{A}_{B,I} & \mathbf{A}_B \end{pmatrix}, \tag{6.57}$$

we consider the Schur complement

$$\mathbf{B} = \mathbf{A}_B - \mathbf{A}_{B,I}\mathbf{A}_I^{-1}\mathbf{A}_{I,B}, \tag{6.58}$$

where the lower indices I and B are related to the degrees of freedom at the nodes living on the interior of the domain Ω and on its boundary $\partial\Omega$, respectively.

Corollary 6.2. *The matrices* $\mathbf{B}$ *and* $\mathbb{B}_{\partial\Omega}$ *are spectrally equivalent uniformly with respect to* h, i.e., $\mathbb{B}_{\partial\Omega} \prec \mathbf{B} \prec \mathbb{B}_{\partial\Omega}$.

Thus, $\mathbb{B}_{\partial\Omega}$ can be used as a spectrally equivalent preconditioner for the Schur complement $\mathbf{B}$. However, the matrix $\mathbb{B}_{\partial\Omega}$ has an $\mathcal{O}(n_1 n_2)$ fill-in. Therefore, the direct procedure of the matrix-vector multiplication by $\mathbb{B}_{\partial\Omega}$ requires $\mathcal{O}(n_1 n_2)$ arithmetic operations, where n_1 and n_2 are the numbers of the mesh intervals on the horizontal and vertical edges of Ω, respectively. It will be seen later that such computational cost of matrix-vector multiplications is not irreducible for spectrally equivalent Schur complement preconditioners. Unfortunately, we do also not know any fast solving procedures for the systems with this system matrix. Therefore, in our next steps, we will consider simpler boundary norms and preconditioners induced by them, which can be represented in a more explicit form, save computational work for the matrix-vector multiplications or/and admit fast solving procedures for the corresponding systems of algebraic equations. In the next paragraph, we rebuild $]v[_{\partial\Omega}$ into a simpler finite-difference boundary norm which saves computational work, at least, for matrix-vector multiplications.

6.1.3 *Finite-Difference Shape Dependent Boundary Norm*

Let us turn to a simpler case of the skeleton mesh, the trace of which on $\partial\Omega$ coincides with the trace of the auxiliary orthogonal quasiuniform grid satisfying (6.38). We denote the left and the right vertical edges of Ω by γ_1 and γ_2, respectively, whereas the horizontal lower and upper edges are denoted by γ_3 and γ_4, respectively.

The auxiliary *coarse (quasiuniform) grid*, by its definition, is the coarsest rectangular imbedded quasiuniform grid. It has rectangular elements as much as possible close to the square $\epsilon \times \epsilon$ and is obtained by subdividing the domain Ω by vertical lines $x_1 = t_{1,i}$ in such a way that $t_{1,0} = 0$, $t_{1,n_\epsilon} = 1$ for some integer $n_\epsilon \geq 1$ and that sizes $\eta_{1,i} := t_{1,i} - t_{1,i-1},\ i = 1, 2, \ldots, n_\epsilon$, satisfy the inequalities

$$\underline{c}_\circ \epsilon \leq \eta_{1,i} \leq \overline{c}_\circ \epsilon, \tag{6.59}$$

with positive constants $\underline{c}_\circ$ and $\overline{c}_\circ$. with $0 < \underline{c}_\circ, \overline{c}_\circ = \text{const}$. The notation $\mathcal{V}_c(\Omega)$ will be used for the space of finite element functions which are continuous on Ω and bilinear on each element of the coarse quasiuniform mesh. This space will be termed the *coarse finite element space.*

Simultaneously, we have defined overlapping intervals

$$\tau_i = (t_{1,i-1}, t_{1,i+1}),\ i = 0, 2, \ldots, n_\epsilon, \tag{6.60}$$

and intersections $\gamma_{k,i}$ of τ_i, $i = 1, 2, \ldots, n_\epsilon - 1$, with the edges γ_k, $k = 2, 3$, where we agree that $t_{1,-1} = 0$ and $t_{1,n+1} = 1$. Furthermore, we use the notations $\gamma^0 = \Gamma_0$, $\gamma^1 = \Gamma_1$. For simplicity, we assume that $\overline{\tau}_0 \cap \partial\Omega = \overline{\Gamma}_0$ and $\overline{\tau}_{n_\epsilon} \cap \partial\Omega = \overline{\Gamma}_1$, and that the numbers of nodes on these sets are the same and equal to ν. Let

$$\boldsymbol{\Delta}_{1/2,k} = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & \mathbf{0} & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \ddots \\ & \mathbf{0} & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}^{1/2}, \qquad k = 0, 1, \tag{6.61}$$

be $\nu \times \nu$ matrices, acting on vectors of d.o.f corresponding to the nodes on $\overline{\gamma}_k$. If ν_i is the number of the nodes on $\overline{\tau}_i$, then $\boldsymbol{\Delta}_{1/2,k,i}$ is the similar to (6.61) $\nu_i \times \nu_i$ matrix related to the segment $\overline{\gamma}_{k,i}$, $k = 2, 3$. In the $2(n_1 + 1) \times 2(n_1 + 1)$ matrix

$$\boldsymbol{\nabla} = \frac{h}{\epsilon} \begin{pmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{pmatrix},$$

the unity matrices $\mathbf{I}$ on diagonal correspond to the nodes on the edges $\overline{\gamma}_2$ and $\overline{\gamma}_3$, respectively.

In the expressions for the matrices $\boldsymbol{\Delta}_{1/2,k}$, $\boldsymbol{\Delta}_{1/2,k,i}$ and $\boldsymbol{\nabla}$ local orderings of degrees of freedom are assumed. We can assume that rows of these matrices correspond to successive orderings of the nodes in agreement with their "global" numbers on $\partial\Omega$. The mesh nodes on the boundary $\partial\Omega$ can be numbered, *e.g.*, counter-clockwise, starting from the left lower vertex. In accordance with their definitions and our general agreement, these matrices are considered as defined on d.o.f. at the nodes of the sets $\overline{\Gamma}_0, \overline{\Gamma}_1$, $\overline{\gamma}_{2,i}, \overline{\gamma}_{3,i}$ and $\overline{\gamma}_2, \overline{\gamma}_3$, respectively, and continued by zeroes on the d.o.f. of the remaining nodes of the boundary, when necessary. This implies that sums like (6.62) below should be understood as *topological* sums.

Lemma 6.2. *The matrix*

$$\mathbf{C} = \boldsymbol{\nabla} + \boldsymbol{\Delta}_{1/2,0} + \boldsymbol{\Delta}_{1/2,1} + \sum_{k=2,3} \sum_{i=1}^{n_\epsilon - 1} \boldsymbol{\Delta}_{1/2,k,i} \tag{6.62}$$

is spectrally equivalent to the matrix $\mathbb{B}_{\partial\Omega}$ *uniformly in* h *and* $\epsilon \in (0,1]$. *Moreover, for any vector* $\mathbf{v}_B$, *the arithmetical cost of the matrix-vector multiplication* $\mathbf{C}\mathbf{v}_B$ *is* $\mathrm{ops}\,[\mathbf{C}\mathbf{v}_B] = \mathcal{O}((n_1+n_2)(1+\log n_2))$.

Proof. The proof is completed in two steps. In the first step, we introduce the seminorm $\rfloor\cdot\lfloor_{\partial\Omega}$ by the expression

$$\rfloor v\lfloor^2_{\partial\Omega} = \epsilon^{-1}\int_0^1 (v(x_1,\epsilon) - v(x_1,0))^2 dx_1$$

$$+ \sum_{i=1}^{n_\epsilon-1} \int_{\tau_i}\int_{\tau_i} \left[\frac{(v(x_1,0)-v(y_1,0))^2}{(x_1-y_1)^2} + \frac{(v(x_1,\epsilon)-v(y_1,\epsilon))^2}{(x_1-y_1)^2}\right] dx_1 dy_1 \tag{6.63}$$

$$+ \int_{\Gamma_0}\int_{\Gamma_0} \frac{(v(s)-v(\overline{s}))^2}{(s-\overline{s})^2} ds d\overline{s} + \int_{\Gamma_1}\int_{\Gamma_1} \frac{(v(s)-v(\overline{s}))^2}{(s-\overline{s})^2} ds d\overline{s},$$

and show that it is equivalent to the seminorm $\rceil v \lceil_{\partial\Omega}$, *i.e.*,

$$\underline{\gamma}_0 \rfloor v\lfloor_{\partial\Omega} \le \rceil v \lceil_{\partial\Omega} \le \overline{\gamma}_0 \rfloor v\lfloor_{\partial\Omega}, \quad \forall\, v \in H^1(\Omega), \tag{6.64}$$

with positive constants $\underline{\gamma}_0$ and $\overline{\gamma}_0$, whereas its matrix is equivalent to $\mathbf{C}$. Indeed, for any $f(x_1,y_1) \in L_2(\pi_1)$, $\pi_1 = (0,1)\times(0,1)$, we have

$$\int_0^1 \int_{|x_1-y_1|\le\epsilon} f^2(x_1,y_1) dy_1 dx_1 = \sum_{i=2}^{n_\epsilon-1} \int_{t_{i-1}}^{t_i} \int_{|x_1-y_1|\le\epsilon} f^2(x_1,y_1) dx_1 dy_1$$
$$\le \sum_{i=1}^{n_\epsilon-1} \int_{\tau_i}\int_{\tau_i} f^2(x_1,y_1) dy_1 dx_1,$$

and, therefore,

$$\rceil v \lceil_{\partial\Omega} \le \rfloor v\lfloor_{\partial\Omega}, \quad \forall\, v \in H^1(\Omega)\,. \tag{6.65}$$

To get the opposite inequality, we consider a function $v(x) \in H^{1/2}(0,1)$ of one variable and note that

$$\sum_{i=1}^{n_\epsilon-1} \int_{\tau_i}\int_{\tau_i} \frac{(v(x)-v(y))^2}{(x-y)^2} dxdy \le 2\int_0^1 \int_{|x-y|\le\varkappa} \frac{(v(x)-v(y))^2}{(x-y)^2} dxdy,$$

where $\varkappa \le c\epsilon$, $c = 2\overline{c}_\circ$. On the other hand, for any $\varkappa \le 1$ and $z = (x+y)/2$, we have $|z-x| = |y-x|/2$, $|y-z| = |y-x|/2$, and as a consequence

$$\int_0^1 \int_{|x-y|\le\varkappa} \frac{(v(x)-v(y))^2}{(x-y)^2} dxdy \le 2\int_0^1 \int_{|x-y|\le\varkappa} \frac{(v(x)-v(z))^2+(v(z)-v(y))^2}{(x-y)^2} dydx$$

$$= 4\int_0^1 \int_{|x-z|\le\varkappa/2} \frac{(v(x)-v(z))^2}{(x-z)^2} dzdx + 4\int_0^1 \int_{|z-y|\le\varkappa/2} \frac{(v(z)-v(y))^2}{(z-y)^2} dzdy$$

$$= 8\int_0^1 \int_{|x-y|\le\varkappa/2} \frac{(v(x)-v(y))^2}{(x-y)^2} dydx\,.$$

From the above inequality, it follows that there exists $\underline{\gamma}_0 = \text{const} > 0$, for which

$$\underline{\gamma}_0^2 \sum_{i=1}^{n_\epsilon - 1} \int_{\tau_i}\int_{\tau_i} \frac{(v(x)-v(y))^2}{(x-y)^2}\,dxdy \le \int_0^1 \int_{|x-y|\le\epsilon} \frac{(v(x)-v(y))^2}{(x-y)^2}\,dydx\,.$$

This inequality, (6.14) and (6.63) directly lead to the opposite to (6.65) bound, so that combining with (6.65) we get (6.64).

The restriction of the FE space $\mathcal{V}(\Omega)$ to the boundary is denoted by $\mathcal{V}_{\rm tr}(\partial\Omega)$. For the second step, it remains to show the inequalities

$$\underline{\gamma}_C \mathbf{v}^\top \mathbf{C}\mathbf{v} \le \lfloor v \rfloor^2_{\partial\Omega} \le \overline{\gamma}_C \mathbf{v}^\top \mathbf{C}\mathbf{v}, \quad \forall\ \mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\partial\Omega), \tag{6.66}$$

with some positive constants $\underline{\gamma}_C$ and $\overline{\gamma}_C > 0$. These bounds follow from the equivalences

$$\begin{aligned}
&\epsilon^{-1}\textstyle\int_0^1 (v(x_1,\epsilon)-v(x_1,0))^2 dx_1 \asymp \mathbf{v}^\top \boldsymbol{\nabla}\mathbf{v}, && \forall\,\mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\gamma_2\cup\gamma_3),\\
&\textstyle\int_{\tau_i}\int_{\tau_i} \frac{(v(x_1,0)-v(y_1,0))^2}{(x_1-y_1)^2}\,dx_1dy_1 \asymp \mathbf{v}^\top \boldsymbol{\Delta}_{1/2,2,i}\mathbf{v}, && \forall\ \mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\gamma_2),\\
&\textstyle\int_{\tau_i}\int_{\tau_i} \frac{(v(x_1,\epsilon)-v(y_1,\epsilon))^2}{(x_1-y_1)^2}\,dx_1dy_1 \asymp \mathbf{v}^\top \boldsymbol{\Delta}_{1/2,3,i}\mathbf{v}, && \forall\ \mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\gamma_3),\\
&\textstyle\int_{\Gamma_k}\int_{\Gamma_k} \frac{(v(s)-v(\overline{s}))^2}{(s-\overline{s})^2} ds d\overline{s} \asymp \mathbf{v}^\top \boldsymbol{\Delta}_{1/2,k}\mathbf{v}, && \forall\ \mathbf{v} \Leftrightarrow v \in \mathcal{V}_{\rm tr}(\Gamma_k),
\end{aligned} \tag{6.67}$$

with $k = 0, 1$, where $\mathcal{V}_{\rm tr}(\gamma_2\cup\gamma_3)$, $\mathcal{V}_{\rm tr}(\gamma_2)$, $\mathcal{V}_{\rm tr}(\gamma_3)$ and $\mathcal{V}_{\rm tr}(\Gamma_k)$ are the spaces of traces of the FE spaces on the corresponding subsets of the boundary. The bounds of first line follow from the spectral equivalence of the FE 1d mass matrix and its diagonal. The remaining bounds express the well-known fact proved in [Andreev (1972, 1973)]. Suppose that some interval τ is subdivided by a quasiuniform grid into ν intervals, and that $\mathcal{H}(\tau)$ is the corresponding space of continuous piecewise linear functions. Then, according to the cited works, the matrix corresponding to the quadratic form $|\,v\,|^2_{1/2,\tau}$ on the space $\mathcal{H}(\tau)$ is spectrally equivalent to the matrix $\boldsymbol{\Delta}_{1/2}$ of the form (6.61).

Now we can conclude that (6.66) holds, which together with (6.64) results in the inequalities

$$\underline{\beta}_C \mathbf{C} \le \mathbb{B}_{\partial\Omega} \le \overline{\beta}_C \mathbf{C} \tag{6.68}$$

with some positive constants $\underline{\beta}_C$ and $\overline{\beta}_C$ which depend only on the constants in the quasiuniformity conditions (6.38).

The matrix-vector multiplication $\boldsymbol{\nabla}\mathbf{v}$ requires $6n_2+1$ a.o., matrix-vector multiplications with the matrices $\boldsymbol{\Delta}_{1/2,k}$ or $\boldsymbol{\Delta}_{1/2,k,i}$ can be completed by FDFT with $\mathcal{O}(n_2 \log n_2)$ a.o. Hence, the matrix-vector multiplication by the topological sum of these matrices requires $\mathcal{O}((n_1+n_2)(1+\log n_2))$ a.o. This approves the estimate of the arithmetic work for the multiplication $\mathbf{Cv}$ given in the lemma. □

Corollary 6.3. *The matrix* $\mathbf{C}$ *is spectrally equivalent to the Schur complement* $\mathbf{B}$ *of (6.58),* i.e., $\mathbf{B} \asymp \mathbf{C}$.

Proof. The proof follows from Theorem 6.3, the definition of the matrix $\mathbb{B}_{\partial\Omega}$ and Lemma 6.2. □

6.2 Schur Complement Preconditioning by DD

In the case of a slim domain and the quasiuniform FE mesh, preconditioners for boundary Schur complements can be obtained by means of DD techniques employing both overlapping and nonoverlapping domain decompositions. The main tool is the decomposition of a shape irregular domain by a subsidiary coarse grid into an assemblage of shape regular subdomains, for which the situation is well studied. In this way, we may come to preconditioners, which are different from the preconditioners of the preceding section, but provide the same or close estimates of the relative condition number. Besides, they allow us to apply solvers for the systems of algebraic equations with these preconditioners as the system matrices, which are almost optimal in the computational complexity. There are two more specific purposes for their derivation. One is to minimize the damage of the relative condition number caused by splitting the d.o.f. belonging to the vertices of the domain from those of the remaining unknowns in the preconditioner. Such a splitting will be used in the DD algorithm for the problem (6.1)–(6.4) with piecewise variable orthotropism, see Section 6.4. Another is to create instruments for the analysis of preconditioner-solvers, obtained in a more direct way from the Schur Complement matrix. This analysis will be accomplished by comparison of different preconditioners.

6.2.1 *Preconditioning by Nonoverlapping DD Techniques*

In this paragraph, we consider a thin rectangle and derive the preconditioner for the boundary Schur complement by the implementation of the

DD procedure with nonoverlapping subdomains. It allows us to split the d.o.f. of each vertical edge and the d.o.f. of the pair of longest edges from other d.o.f. Then we additionally split d.o.f. belonging to the vertices of the rectangle from the remaining d.o.f.

The coarse grid, introduced in the preceding subsection, defines the nonoverlapping domain decomposition of the rectangle Ω into the subdomains

$$\Omega^i = (t_{1,i-1}, t_{1,i}) \times (0, \epsilon), \quad i = 1, 2, \ldots, n_\epsilon.$$

For their edges, we use the notations Γ_k^i, $k = 0, 1, 2, 3$, and order them counter-clockwise starting from the lower edge of Ω^i. The FE space is the direct sum

$$\mathcal{V}(\Omega) = \mathcal{W}_{\rm r}(\Omega) \oplus \mathcal{V}_{\rm c}(\Omega), \tag{6.69}$$

where $\mathcal{V}_{\rm c}(\Omega)$ is the coarse space containing continuous functions which are bilinear on each subdomain Ω^i. The second subspace in (6.69) is supplied by the index "r", because in the sequel the mesh, to which it corresponds, will result from the rarefication of the orthotropic mesh. Note that functions $v_r \in \mathcal{W}_{\rm r}(\Omega)$ vanish at the nodes of the coarse mesh. Notations $\mathcal{V}_{\rm c}(\Omega^i)$ and $\mathcal{W}_{\rm r}(\Omega^i)$ will stand for restrictions of the spaces $\mathcal{V}_{\rm c}(\Omega)$ and $\mathcal{W}_{\rm r}(\Omega)$ to Ω^i.

We now consider any $v = v_{\rm c} + v_W \in \mathcal{V}(\Omega^i)$ such that $v_{\rm c} \in \mathcal{V}_{\rm c}(\Omega^i)$, $v_W \in \mathcal{W}_{\rm r}(\Omega^i)$ and is discrete harmonic in Ω^i. The subdomains Ω^i are shape regular, and according to the inequalities (4.18), obtained by [Bramble *et al.* (1986)],

$$\frac{1}{(1+\log n_2)^2} \left(|\, v_{\rm c} \,|^2_{1,\Omega^i} + \sum_{k=0}^{3} {}_{00}|\, v_W \,|^2_{1/2,\Gamma_k^i} \right) \prec |\, v \,|^2_{1,\Omega^i}$$

$$\prec |\, v_{\rm c} \,|^2_{1,\Omega^i} + \sum_{k=0}^{3} {}_{00}|\, v_W \,|^2_{1/2,\Gamma_k^i}. \tag{6.70}$$

Let

$$\mathcal{S}_{\partial\Omega}(v) = |\, v_{\rm c} \,|^2_{1,\Omega} + \mathcal{B}_W(v_W)$$

with

$$\mathcal{B}_W(v_W) = {}_{00}|\, v_W \,|^2_{1/2,\Gamma_3^1} + {}_{00}|\, v_W \,|^2_{1/2,\Gamma_1^{n_\epsilon}} + \sum_{i=1}^{n_\epsilon} \sum_{k=0,2} {}_{00}|\, v_W \,|^2_{1/2,\Gamma_k^i}\,.$$

For a FE function $v \in \mathcal{V}(\Omega)$ that is discrete harmonic in Ω, we directly come from (6.70) to the inequalities

$$\frac{1}{(1+\log n_2)^2}\, \mathcal{S}_{\partial\Omega}(v) \;\prec |\, v \,|^2_{1,\Omega} \prec \mathcal{S}_{\partial\Omega}(v)\,. \tag{6.71}$$

If the matrix $\boldsymbol{\mathcal{B}}_W$ is spectrally equivalent to the matrix corresponding to the quadratic form $\mathcal{B}_W(v_W)$ on the subspace $\mathcal{W}_{\rm r}(\Omega)$, *i.e.,*

$$\mathbf{v}_W^\top \boldsymbol{\mathcal{B}}_W \mathbf{v}_W \prec \mathcal{B}_W(v_W) \prec \mathbf{v}_W^\top \boldsymbol{\mathcal{B}}_W \mathbf{v}_W, \tag{6.72}$$

then (6.71) is equivalent to the inequality

$$\frac{1}{(1+\log n_2)^2} \mathbf{v}^\top \boldsymbol{\mathcal{C}}_{\rm hi}\, \mathbf{v} \prec \mathbf{v}^\top \mathbf{B}_{\rm hi}\, \mathbf{v} \prec \mathbf{v}^\top \boldsymbol{\mathcal{C}}_{\rm hi}\, \mathbf{v}, \tag{6.73}$$

where

$$\boldsymbol{\mathcal{C}}_{\rm hi} = \begin{pmatrix} \boldsymbol{\mathcal{B}}_W & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\rm c} \end{pmatrix},$$

and $\mathbf{B}_{\rm c}$ is the block that corresponds to the subspace $\mathcal{V}_{\rm c}(\Omega)$. The notation $\mathbf{B}_{\rm hi}$ for the Schur complement reflects the fact that it is written in the two level basis corresponding to the representation (6.69) of the FE space.

Let $\mathbf{B}_{\rm Hi}$ stand for the Schur complement corresponding to the three level representation of the FE space

$$\mathcal{V}(\Omega) = \mathcal{W}_{\rm r}(\Omega) \oplus \mathcal{W}_{\rm c}(\Omega) \oplus \mathcal{V}_0(\Omega), \tag{6.74}$$

resulting from (6.69) and

$$\mathcal{V}_{\rm c}(\Omega) = \mathcal{W}_{\rm c}(\Omega) \oplus \mathcal{V}_0(\Omega), \tag{6.75}$$

where $\mathcal{V}_0(\Omega)$ is the subspace of bilinear functions and the subspace $\mathcal{W}_{\rm c}(\Omega)$ contains functions vanishing at the vertices. Let also $\mathbf{B}^{(\rm w)}$ and $\mathbf{B}^{(\rm v)}$ denote the blocks on the diagonal of $\mathbf{B}_{\rm Hi}$ corresponding to the subspaces $\mathcal{W}_{\rm c}(\Omega)$ and $\mathcal{V}_0(\Omega)$, respectively, and $\underline{n} = \min(n_1, n_2)$, $\overline{n} = \max(n_1, n_2)$. If we slightly change reasoning, then we can also get the inequalities

$$\frac{1}{1+\log \underline{n}} \min\Big(\frac{1}{n_\epsilon}, \frac{1}{1+\log \underline{n}}\Big) \mathbf{v}^\top \boldsymbol{\mathcal{C}}_{\rm Hi}\, \mathbf{v} \prec \mathbf{v}^\top \mathbf{B}_{\rm Hi} \mathbf{v} \prec \mathbf{v}^\top \boldsymbol{\mathcal{C}}_{\rm Hi}\, \mathbf{v}, \tag{6.76}$$

where

$$\boldsymbol{\mathcal{C}}_{\rm Hi} = \begin{pmatrix} \boldsymbol{\mathcal{B}}_W & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{(\rm w)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}^{(\rm v)} \end{pmatrix}. \tag{6.77}$$

We reorder sets $\Gamma_k^i \subset \partial\Omega$ consecutively, *e.g.,* counter-clockwise, starting from Γ_0^1, introduce the notations $\mathcal{T}^i$, $i = 1, 2, \ldots, n_\epsilon$, for these sets, an $\hat{\nu}_i$ denote the number of the intervals of the source mesh on $\mathcal{T}^i$. Estimates (6.72) and, therefore, (6.73) as well as (6.76) hold for the block diagonal matrix

$$\boldsymbol{\mathcal{B}}_W = \operatorname{diag}\,[\hat{\boldsymbol{\Delta}}_{1/2,i}]_{i=1}^{2(n_\epsilon+1)} \tag{6.78}$$

with the blocks $\hat{\boldsymbol{\Delta}}^{\,2}_{1/2,i} = \operatorname{tridiag}\,[-1, 2, -1]_1^{\hat{\nu}_i - 1}$. Thus, we have just proved the following lemma.

Lemma 6.3. *Let $\boldsymbol{\mathcal{B}}_W$ be defined by (6.78). Then estimates (6.76) hold for all $\epsilon \in (0, 1]$ and $h \le (0, \epsilon]$.*

Systems of algebraic equations with the matrices $\boldsymbol{\mathcal{C}}_{\rm hi}$ and $\boldsymbol{\mathcal{C}}_{\rm Hi}$ can be solved with $\mathcal{O}(n_\epsilon \underline{n} \log \underline{n}) = \mathcal{O}(\overline{n} \log \underline{n})$ a.o., where here $\overline{n} = n_1$, $\underline{n} = n_2$. However, as follows from (6.76), additional splitting of the vertices of the subdomain in the latter preconditioner seriously damages its relative condition number.

6.2.2 *Preconditioning by Overlapping DD*

The quasiuniform coarse grid satisfying (6.59) introduces a decomposition of the domain Ω into overlapping subdomains $\Omega^i,\ i = 0, 1, \ldots, n_\epsilon,$ where

$$\Omega^i = (t_{1,i-1}, t_{1,i+1}) \times (0, \epsilon), \qquad i = 1, 2, \ldots, n_\epsilon - 1,$$

$$\Omega^0 = (0, t_{1,1}) \times (0, \epsilon), \qquad \Omega^{n_\epsilon} = (t_{n_\epsilon - 1}, 1) \times (0, \epsilon).$$

Accordingly, the FE space is decomposed into the direct sum

$$\mathcal{V}(\Omega) = \mathcal{V}_{\rm c}(\Omega) \oplus \mathcal{V}^0(\Omega^0) \oplus \mathcal{V}^1(\Omega^1) \oplus .. \oplus \mathcal{V}^{n_\epsilon}(\Omega^{n_\epsilon}) \tag{6.79}$$

of $n_\epsilon + 2$ subspaces. In this decomposition, $\mathcal{V}_{\rm c}(\Omega)$ is the subspace of finite element functions which are piecewise bilinear on the coarse mesh and continuous on Ω. The remaining spaces are defined as follows:

$$\begin{aligned}
\mathcal{V}^0(\Omega^0) &= \{v \in \mathcal{V}(\Omega) : v = 0 \ \text{ for } x_1 \geq t_{1,1}\},\\
\mathcal{V}^{n_\epsilon}(\Omega^{n_\epsilon}) &= \{v \in \mathcal{V}(\Omega) : v = 0 \ \text{ for } x_1 \leq t_{1,n_\epsilon - 1}\},\\
\mathcal{V}^i(\Omega^i) &= \{v \in \mathcal{V}(\Omega) : v = 0 \ \text{ for } x_1 \leq t_{1,i-1}, \ \text{and } x_1 \geq t_{1,i+1}\},
\end{aligned}$$

where $i = 1, 2, \ldots, n_\epsilon - 1$. The spaces of traces of functions from the spaces $\mathcal{V}^i(\Omega^i)$ will be denoted by $\mathcal{V}^i_{\rm tr}(\partial\Omega^i)$. For the vector spaces, corresponding to the introduced FE spaces $\mathcal{V}, \mathcal{V}_{\rm c}$, and $\mathcal{V}^i$, we use the notations $V, V_{\rm c}$ and V^i, respectively. Analogously, the notations $V_{\rm tr}, V_{\rm c,tr}, V^i_{\rm tr}$ are related to $\mathcal{V}_{\rm tr}(\partial\Omega), \mathcal{V}_{\rm c,tr}(\partial\Omega), \mathcal{V}^i_{\rm tr}(\partial\Omega^i)$.

Let $\overline{T}^i = \partial\Omega_i \cap \partial\Omega,\ k_i = \nu - 1$ be the number of the fine mesh intervals on T^i for $i = 0, n_\epsilon$, and $k_i = \nu_i - 1$ for $i = 1, 2, \ldots, n_\epsilon - 1$. Let also

$$\mathbb{\Delta}_{1/2,i} = \mathbb{\Delta}_i^{1/2}, \quad \text{with} \quad \mathbb{\Delta}_i = \text{tridiag}\,[-1, 2, -1]_1^{k_i}\,.$$

For $i = 0, n_\epsilon$, the matrices $\mathbb{\Delta}_{1/2,i}$ are assumed to be related to the internal nodes of the corresponding part T^i of the boundary $\partial\Omega$. Additionally, for $i = 1, 2, \ldots, n_\epsilon - 1$, we introduce matrices

$$\mathbb{\Delta}_{i,=} = \text{diag}\,[\mathbb{\Delta}_{1/2,i}, \mathbb{\Delta}_{1/2,i}]$$

with two independent blocks, each of which is related the internal nodes of one of the two disjoint parts of T^i lying on the lines $x_2 \equiv 0, \epsilon$, respectively. The matrix $\boldsymbol{\mathcal{I}}_{\rm c}$ denotes the usual interpolation matrix such that, if

$v_{\rm c} \in \mathcal{V}_{\rm c}(\Omega)$ and $\mathbf{v}_{\rm c}$ its vector representation in the space $V_{\rm c}$, then $\boldsymbol{\mathcal{I}}_{\rm c}\mathbf{v}_{\rm c}$ is the vector representation of v_0 in V. Finally, we define the matrix $\mathbf{B}_{\rm c}$ as generated by the Dirichlet integral on the space $\mathcal{V}_{\rm c}(\Omega)$, and set

$$\mathbb{B}_{\rm c}^{+} = \boldsymbol{\mathcal{I}}_{\rm c}\mathbf{B}_{\rm c}^{-1}\boldsymbol{\mathcal{I}}_{\rm c}^{\top} .$$

Now, the inverse to the overlapping DD preconditioner for the Schur complement can be defined as

$$\boldsymbol{\mathcal{C}}^{-1} = \mathbb{B}_{\rm c}^{+} + \mathbb{A}_{1/2,0}^{+} + \mathbb{A}_{1/2,n_\epsilon}^{+} + \sum_{i=1}^{n_\epsilon - 1} \mathbb{A}_{i,=}^{+} . \tag{6.80}$$

Lemma 6.4. *The matrices* $\mathbf{B}$ *and* $\boldsymbol{\mathcal{C}}$ *are spectrally equivalent uniformly in* h, i.e.,

$$\boldsymbol{\mathcal{C}} \prec \mathbf{B} \prec \boldsymbol{\mathcal{C}} . \tag{6.81}$$

Proof. According to Theorems 3.1 and 3.2, the space decomposition (6.79) is stable. Therefore, Theorem 2.4 is applicable. This means that the seminorm $|||\cdot|||_1$ defined by

$$|||v|||_1^2 = \min_{v=w_{\rm c}+v_0+v_1+\ldots+v_{n_\epsilon} \,:\, w_{\rm c}\in\mathcal{V}_{\rm c}(\Omega),\ v_i\in\mathcal{V}^i(\Omega^i)} \left(|w_{\rm c}|_{1,\Omega}^2 + \sum_{i=0}^{n_\epsilon} |v_i|_{1,\Omega^i}^2 \right),$$

where minimum is taken over all specified decompositions of v, is equivalent to $|\cdot|_{1,\Omega}$ on $\mathcal{V}(\Omega)$, *i.e.*

$$|||v|||_1^2 \prec |v|_{1,\Omega}^2 \prec |||v|||_1^2, \quad \forall\ v \in \mathcal{V}(\Omega) . \tag{6.82}$$

Let $\mathcal{V}_{\rm harm}(\Omega) \subset \mathcal{V}(\Omega)$ and $\mathcal{V}_{\rm harm}^i(\Omega^i) \subset \mathcal{V}^i(\Omega^i)$ be the subspaces of the discrete harmonic functions in Ω^i, whereas $\mathcal{V}_{\rm tr}(\partial\Omega)$ and $\mathcal{V}_{\rm tr}^i(\partial\Omega^i)$ be the spaces of traces introduced above. For $u \in \mathcal{V}_{\rm harm}(\Omega)$, we can also introduce the boundary seminorm $|||\cdot|||_{\partial\Omega}$,

$$|||u|||_{\partial\Omega}^2 = \min_{u=w_{\rm c}+u_0+u_1+\ldots+u_{n_\epsilon} \,:\, w_{\rm c}\in\mathcal{V}_{\rm c}(\Omega),\ u_i\in\mathcal{V}_{\rm tr}^i(\partial\Omega^i)} \mathcal{S}_{\rm r}^2(w_{\rm c}, u_0, u_1, \ldots u_{n_\epsilon}), \tag{6.83}$$

where

$$\mathcal{S}_{\rm r}^2(w_{\rm c}, u_0, u_1, \ldots u_{n_\epsilon}) = \left(|w_{\rm c}|_{1,\Omega}^2 + \sum_{i=0}^{n_\epsilon} |u_i|_{1/2,\partial\Omega^i}^2 \right),$$

and show that

$$\begin{aligned} |||u|||_{\partial\Omega}^2 \prec |u|_{1,\Omega}^2 \prec |||u|||_{\partial\Omega}^2, \quad \forall\ u \in \mathcal{V}_{\rm harm}(\Omega), \\ |||v|||_{\partial\Omega}^2 \prec |v|_{1/2,\Omega}^2 \prec |||v|||_{\partial\Omega}^2, \quad \forall\ v \in \mathcal{V}_{\rm tr}(\partial\Omega) . \end{aligned} \tag{6.84}$$

Obviously it is sufficient to prove only one pair of these inequalities, and we turn to the inequalities of the first line in (6.84). In order to get the left inequality, we take the decomposition of $u = w_{\rm c} + \widetilde{u}_1 + ... + \widetilde{u}_{n_\epsilon}$ entering $|||u|||_1$, then replace $\widetilde{u}_i$ by $u_i \in \mathcal{V}^i_{\rm harm}(\Omega^i)$, $u_i = |_{\partial\Omega^i}\widetilde{u}_i$, and use the equivalences

$$|\phi|_{1,\Omega^i} \asymp |\phi|_{1/2,\partial\Omega^i}, \quad \forall\ \phi \in \mathcal{V}^i_{\rm harm}(\Omega^i)\,. \tag{6.85}$$

Suppose now that, for a given $u \in \mathcal{V}_{\rm harm}(\Omega)$, functions $w_{\rm c} \in \mathcal{V}_{\rm c}(\Omega)$ and $u_i \in \mathcal{V}^i_{\rm tr}(\partial\Omega^i)$ realize the minimum in (6.83). Let also $u_i \in \mathcal{V}^i_{\rm harm}(\Omega_i)$ denote the corresponding discrete harmonic functions and $\widetilde{u} = w_{\rm c}+u_0+u_1+ ... + u_{n_\epsilon}$. Then, since (6.82) holds for $v = \widetilde{u}$, the inequality $|u|_{1,\Omega} \le |\widetilde{u}|_{1,\Omega}$ also holds. Due to the fact that not more than three functions among the functions $w_{\rm c}$ and $\widetilde{u}_i$ are distinct from zero at each point of $\overline{\Omega}$ and by means of Cauchy's inequality and (6.85), we have

$$|u|^2_{1,\Omega} \le 3\left(|w_{\rm c}|^2_{1,\Omega} + \sum_{i=0}^{n_\epsilon} |u_i|^2_{1,\Omega^i}\right) \prec |w_{\rm c}|^2_{1,\Omega} + \sum_{i=0}^{n_\epsilon} |u_i|^2_{1/2,\partial\Omega^i}. \tag{6.86}$$

Therefore, (6.84) is indeed valid.

Let, for a subdomain $\Omega_i,\ 1 \le i \le n_\epsilon - 1$,

- $\mathtt{v}^i_k$, $k = 0,1,2,3$ be vertices of Ω^i counted, *e.g.*, counter-clockwise, starting from $\mathtt{v}_{0^i} = (t_{1,i-1}, 0)$,
- T^i_k be the edges situated between vertices $\mathtt{v}^i_k, \mathtt{v}^i_{k+1}$,
- $t \in (0,1)$ be the coordinate of a point $x \in T^i_l$, $l = k, k+1$, equal to the distance from the vertex $\mathtt{v}^i_k$ divided by the length of the edge T^i_l.

Taking into account that the ratio of largest/smallest edges of $\partial\Omega^i$ can be made less than 2, the seminorm $|\cdot|_{1/2,\partial\Omega^i}$ may be defined similarly to its definition for the unit square, *i.e.,*

$$|v|^2_{1/2,\partial\Omega^i} = \sum_{k=0}^{3} |v|^2_{1/2,T^i_k} + \sum_{k=0}^{3} \int_0^1 \frac{(v_k(t) - v_{k-1}(t))^2}{|t|}\, dt, \tag{6.87}$$

where $v_l(t)$ is the value of v on the edge T^i_l at the "distance" t from the vertex $\mathtt{v}^i_k$.

For $i = 0, n_\epsilon$, traces of functions from $\mathcal{V}^i_{\rm harm}(\Omega^i)$ vanish at the vertical edges on the lines $x_1 \equiv t_{1,1}, t_{1,n_\epsilon-1}$, respectively. For $i = 1, 2, \ldots, n_\epsilon - 1$, traces of functions from each subspace $\mathcal{V}^i_{\rm harm}(\Omega^i)$ vanish at the two opposite vertical edges, which are on the lines $x_1 \equiv t_{1,i-1}, t_{1,i+1}$. Due to these facts and the characterization (6.87), for any $v \in \mathcal{V}^i_{\rm harm}(\Omega^i)$ we obtain the

following representations of the boundary $H^{1/2}$-seminorms:

$$|v|^2_{1/2,\partial\Omega^0} = {}_{00}|v|^2_{1/2,T^0}, \qquad |v|^2_{1/2,\partial\Omega^{n_\epsilon}} = {}_{00}|v|^2_{1/2,T^{n_\epsilon}},$$
$$|v|^2_{1/2,\partial\Omega^i} = {}_{00}|v|^2_{1/2,T_0^i} + {}_{00}|v|^2_{1/2,T_2^i}, \quad i = 1,2,\dots,n_\epsilon - 1, \tag{6.88}$$

which, in turn, allow us to write

$$\mathcal{S}^2_{\rm r}(w_{\rm c}, u_0, u_1, \dots u_{n_\epsilon}) = |w_{\rm c}|^2_{1,\Omega} + {}_{00}|u_0|^2_{1/2,T^0} + {}_{00}|u_{n_\epsilon}|^2_{1/2,T^{n_\epsilon}} + \sum_{i=1}^{n_\epsilon-1} \sum_{k=0,2} {}_{00}|u_i|^2_{1/2,T_k^i}\,. \tag{6.89}$$

Moreover, we have the relations

$$\begin{aligned} &\mathbf{v}_{\rm c}^\top \mathbf{B}_{\rm c} \mathbf{v}_{\rm c} = |v_{\rm c}|^2_{1,\Omega}, \\ &\mathbf{v}_i^\top \mathbb{\Delta}_{i,=} \mathbf{v}_i \asymp {}_{00}|v_i|^2_{1/2,T_0^i} + {}_{00}|v_i|^2_{1/2,T_2^i}, \quad i = 1,2,\dots,n_\epsilon - 1, \\ &\mathbf{v}_0^\top \mathbb{\Delta}_0 \mathbf{v}_0 \asymp {}_{00}|v_0|^2_{1/2,T^0}, \quad \mathbf{v}_{n_\epsilon}^\top \mathbb{\Delta}_{n_\epsilon} \mathbf{v}_{n_\epsilon} \asymp {}_{00}|v_{n_\epsilon}|^2_{1/2,T^{n_\epsilon}} \end{aligned} \tag{6.90}$$

for any FE functions $v_{\rm c} \in \mathcal{V}_{\rm c}(\Omega)$ and $v_i \in \mathcal{V}^i_{\rm tr}(\Omega^i)$ and vectors of their nodal representations $\mathbf{v}_{\rm c} \Leftrightarrow v_{\rm c}$, $\mathbf{v}_i \Leftrightarrow v_i$. The first equality holds by definition, whereas the remaining equivalence relations are one and the same relation for $H^{1/2}$ norm written for different 1d sets discretized by quasiuniform meshes with different numbers of nodes. They are proved similarly to (6.67), see [Andreev (1972, 1973)].

The proof of (6.81) now follows from (6.84), (6.89), (6.90) and Theorem 2.4. □

6.2.3 *Schur Complement Preconditioners by Overlapping DD: Some Simplifications*

We consider two preconditioners resulting from simplifications in the former one. In the first one, the degrees of freedom at the vertices of the domain are decoupled from the remaining degrees of freedom, whereas in another one, additionally, the degrees of freedom at each vertical edge are decoupled from all others. These decouplings are made in the same way as similar simplifications in the preconditiones of the preceding subsection have been made.

The mentioned couplings are located in the first three summands of (6.80). Let V_0 and $V_{\rm c}$ be the spaces of vector representations for functions from the spaces $\mathcal{V}_0(\Omega)$ and $\mathcal{V}_{\rm c}(\Omega)$, respectively, in the corresponding nodal bases, and let $\boldsymbol{\mathcal{I}}_0$ be the interpolation operator such that, for any

$\mathbf{v}_0 \in V_0,\ \mathbf{v}_0 \Leftrightarrow v_0 \in \mathcal{V}_0(\Omega)$, the vector $\mathbf{v}_c = \boldsymbol{\mathcal{I}}_0 \mathbf{v}_0$ represents v_0 in the basis of the space $\mathcal{V}_c(\Omega)$. In the same way as in the case of defining $\mathbb{B}_c^+$, we introduce the coarse grid preconditioner

$$\widehat{\mathbb{B}}_c^{-1} = \boldsymbol{\mathcal{I}}_c \boldsymbol{\mathcal{B}}_c^{-1} \boldsymbol{\mathcal{I}}_c^\top, \quad \text{with } \boldsymbol{\mathcal{B}}_c^{-1} := (\mathbf{B}^{(w)})^+ + \boldsymbol{\mathcal{I}}_0 (\mathbf{B}^{(v)})^{-1} \boldsymbol{\mathcal{I}}_0^\top .$$

The matrix $\mathbb{A}_{1/2,0} : V_{tr}^0 \to V_{tr}^0$ can be represented in the block form

$$\mathbb{A}_{1/2,0} = \begin{pmatrix} \mathbb{A}_{1/2,0}^{(e)} & \mathbb{A}_{1/2,0}^{(e,v)} \\ \mathbb{A}_{1/2,0}^{(v,e)} & \mathbb{A}_{1/2,0}^{(v)} \end{pmatrix}, \tag{6.91}$$

reflecting the decomposition $V_{tr}^0 = V_{tr}^{0,e} \oplus V_{tr}^{0,v}$, where $V_{tr}^{0,e}$ is the subspace of vectors with zero entries for the vertices of Ω^0, and $V_{tr}^{0,v}$ is the 2d subspace with nonzero entries only for the two vertices common for Ω^0 and Ω. Let the space decomposition $\mathcal{V}_{tr}^0(\partial\Omega^0) = \mathcal{V}_{tr}^{0,e}(\partial\Omega^0) \oplus \mathcal{V}_{tr}^{0,v}(\partial\Omega^0)$ implies that the latter subspace contains traces of the nodal bilinear polynomials on Ω^0. Then the vertex degrees of freedom are split in the respective block diagonal preconditioner, *i.e.*,

$$\widetilde{\mathbb{A}}_{1/2,0} = \begin{pmatrix} \mathbb{A}_{1/2,0}^{(e)} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_0^{(v)} \end{pmatrix},$$

in which

$$\boldsymbol{\Delta}_0^{(v)} = \frac{1}{6} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

is the block of the stiffness matrix for the bilinear reference element induced by the Dirichlet integral over the reference square. Now we can take

$$\widehat{\mathbb{A}}_{1/2,0}^{-1} := (\mathbb{A}_{1/2,0}^{(e)})^+ + \boldsymbol{\mathcal{I}}_c^{(v)} (\boldsymbol{\Delta}_0^{(v)})^{-1} (\boldsymbol{\mathcal{I}}_c^{(v)})^\top \tag{6.92}$$

as the inverse to the preconditioner of the matrix (6.91), where $\boldsymbol{\mathcal{I}}_c^{(v)} : V_{tr}^{0,v} \to V_{tr}^0$ is the interpolation matrix.

Similar to (6.92), the Schur complement preconditioner

$$\boldsymbol{\mathcal{C}}_1^{-1} = \widehat{\mathbb{B}}_c^{-1} + \widehat{\mathbb{A}}_{1/2,0}^{+} + \widehat{\mathbb{A}}_{1/2,n_\epsilon}^{+} + \sum_{i=1}^{n_\epsilon - 1} \mathbb{A}_{i,=}^{+} . \tag{6.93}$$

is obtained in the same way as the preconditioner $\widehat{\mathbb{A}}_{1/2,n_\epsilon}^{-1}$ for $\mathbb{A}_{1/2,n_\epsilon}^{-1}$.

Additionally, the block $\mathbb{A}_{1/2,0}^{(e)}$ in (6.92) can be preconditioned by the block diagonal matrices

$$\widehat{\mathbb{A}}_{1/2,0}^{(e)} = \mathrm{diag}[\hat{\Delta}_{1/2,1}, \hat{\Delta}_{1/2,2}, \hat{\Delta}_{1/2,2n_\epsilon}] = \mathrm{diag}[\hat{\Delta}_{1/2,1}, \mathbb{A}_{0,=}] .$$

Thus, $\widehat{\mathbb{A}}^{-1}_{1/2,0}$ can be replaced by

$$\mathring{\mathbb{A}}^{-1}_{1/2,0} := (\widehat{\mathbb{A}}^{(\mathrm{e})}_{1/2,0})^{+} + \boldsymbol{\mathcal{I}}^{(\mathrm{v})}_{\mathrm{c}} (\boldsymbol{\Delta}^{(\mathrm{v})}_{0})^{-1} (\boldsymbol{\mathcal{I}}^{(\mathrm{v})}_{\mathrm{c}})^{\top} .$$

Therefore, we arrive at another preconditioner that looks like

$$\boldsymbol{\mathcal{C}}^{-1}_{2} = \widehat{\mathbb{B}}^{-1}_{\mathrm{c}} + \mathring{\mathbb{A}}^{+}_{1/2,0} + \mathring{\mathbb{A}}^{+}_{1/2,n_\epsilon} + \sum_{i=1}^{n_\epsilon - 1} \mathbb{A}^{+}_{i,=}, \tag{6.94}$$

where $\mathring{\mathbb{A}}^{+}_{1/2,n_\epsilon}$ is the matrix that is obtained in the same way as $\mathring{\mathbb{A}}^{+}_{1/2,0}$ on the basis of symmetry considerations.

Lemma 6.5. *For $\epsilon \in (0,1]$, $h \in (0,\epsilon]$, and $\underline{n} = \min(n_1, n_2)$, the spectral inequalities*

$$\min\left(\frac{1}{n_\epsilon}, \frac{1}{1+\log \underline{n}}\right) \boldsymbol{\mathcal{C}}_1 \prec \mathbf{B} \prec \boldsymbol{\mathcal{C}}_1, \quad \min\left(\frac{1}{n_\epsilon}, \frac{1}{1+\log^2 \underline{n}}\right) \boldsymbol{\mathcal{C}}_2 \prec \mathbf{B} \prec \boldsymbol{\mathcal{C}}_2$$

are valid.

Proof. We omit the proof, because, on the basis of Lemma 6.4, the further steps of the proof are based on the same facts as the proof of Lemma 6.3. □

6.3 Orthotropic Discretizations with Arbitrary Aspect Ratios on Thin Rectangles

A more complicated situation arises when we need to solve a heat conduction problem with different heat conduction coefficients in different directions in a slim rectangle which is discretized by a uniform rectangular mesh. Let us assume that the aspect ratios of the conductivity coefficients and sizes of the mesh are finite but arbitrary. Therefore, we consider the following model problem in a slim domain $\Omega = (0,1)\times(0,\epsilon)$: given $f \in \mathbb{V}(\Omega)^*$, find $u \in \mathbb{V}(\Omega) \subseteq H^1(\Omega)$ such that

$$\alpha_\Omega(u,v) = \langle f, v\rangle, \quad \forall\, v \in \mathbb{V}(\Omega), \tag{6.95}$$

where the bilinear form is given by the expression

$$\alpha_\Omega(u,v) = \int_\Omega \nabla u(x) \cdot \boldsymbol{\rho}(x)\, \nabla v(x)\, dx,$$

and $\boldsymbol{\rho} = \mathrm{diag}\,[\rho_1, \rho_2]$ is a 2×2 diagonal matrix with the arbitrary positive constants ρ_1 and ρ_2. For simplicity, we restrict ourselves to a uniform

rectangular mesh of arbitrary sizes $h_1, h_2 > 0$, although our consideration are easily adjustable to more general situations, *e.g.,* of a quasiuniform tensor product mesh or of mesh that is topologically equivalent to such a mesh.

For the ease of future references, in relation to a new problem, we use different from previously used notations for the FE stiffness matrix

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_I & \mathbf{Q}_{I,B} \\ \mathbf{Q}_{B,I} & \mathbf{Q}_B \end{pmatrix} \tag{6.96}$$

and its boundary Schur complement

$$\mathbf{Y} = \mathbf{Q}_B - \mathbf{Q}_{B,I}\mathbf{Q}_I^{-1}\mathbf{Q}_{I,B}\,. \tag{6.97}$$

The derivation of a good preconditioner for the Schur complement will be completed in three steps. At step 1), we change variables and reduce the problem (6.95) to a transformed isotropic problem on some domain Ω_ξ. At step 2), we introduce the rarefied transformed mesh, which is the finest and, if possible, quasiuniform mesh imbedded in the transformed source mesh. It is obtained by rarefication of the transformed source mesh in one direction, corresponding to the smallest size of the latter mesh. Then the block diagonal preconditioner for the FE matrix $\mathbf{Q}$ is introduced, containing two independent blocks on diagonal, one of which is the FE stiffness matrix, induced by the rarefied transformed mesh. In turn, the preconditioner for the FE matrix $\mathbf{Q}$ allows us to obtain the block diagonal preconditioner for the Schur complement $\mathbf{Y}$ with two independent blocks. At step 3), further decoupling is accomplished. The block of the Schur complement preconditioner, corresponding to the transformed rarefied mesh, obviously, can be handled as in the preceding section. Another block does not require additional treatment, because it itself is a block diagonal matrix with simple explicitly written down blocks, specified on the unknowns, subjected to rarefication.

6.3.1 *Reducing to Isotropic Discretization*

The change of variables $\xi_1 = x_1$ and $\xi_2 = \sqrt{\rho_1/\rho_2}\, x_2$ transforms the bilinear form $\alpha_\Omega(\cdot,\cdot)$ into

$$\alpha_\Omega(u,v) = \sqrt{\rho_1\rho_2}\;\widetilde{\alpha}(u,v), \quad \text{with} \quad \widetilde{\alpha}(u,v) = \int_{\Omega_\xi} \nabla_\xi u \cdot \nabla_\xi v \, d\xi, \tag{6.98}$$

where ∇_ξ denotes the gradient with respect to the new variables ξ, $\Omega_\xi = (0,1)\times(0,\widetilde{\epsilon})$ is the new domain and $\widetilde{\epsilon} = \epsilon\sqrt{\rho_1/\rho_2}$. The FE space $\mathcal{V}(\Omega)$ is

transformed into the space $\mathbb{V}(\Omega_\xi)$ of the piecewise bilinear functions on the rectangular transformed source mesh of the sizes $\hbar_1 = h_1,\ \hbar_2 = \sqrt{\rho_1/\rho_2}\, h_2$ with the mesh lines $\xi_k \equiv \xi_{k,l} = l\hbar_k$. We have $\mathbf{Q} = \sqrt{\rho_1\rho_2}\,\mathbb{Q},\ \mathbf{Y} = \sqrt{\rho_1\rho_2}\,\mathbb{Y}$, and

$$\mathbb{Y} = \mathbb{Q}_B - \mathbb{Q}_{B,I}\mathbb{Q}_I^{-1}\mathbb{Q}_{I,B}, \tag{6.99}$$

where $\mathbb{Q}_I, \mathbb{Q}_{I,B}, \mathbb{Q}_B$ are blocks of the stiffness matrix $\mathbb{Q}$ generated by the bilinear form $\widetilde{\alpha}(u,v)$ on the space $\mathbb{V}(\Omega_\xi)$. Therefore, the preconditioning of $\mathbf{Y}$ is reduced to the preconditioning of the Schur complement $\mathbb{Y}$.

We can restrict ourselves to the consideration of the case $\widetilde{\epsilon} < 1$, since the case $\widetilde{\epsilon} > 1$ is reduced to the former one by the interchange of variables. Under the condition $\widetilde{\epsilon} \le 1$, the following three cases can be distinguished:

$$i)\ \ \hbar_2 \le \hbar_1 \le \widetilde{\epsilon}, \quad ii)\ \ \hbar_2 \le \widetilde{\epsilon} \le \hbar_1, \quad iii)\ \ \hbar_1 \le \hbar_2 \le \widetilde{\epsilon}. \tag{6.100}$$

Let us start from i). Under the stated conditions, the embedded rarefied quasiuniform rectangular grid $\xi_k \equiv \widetilde{\xi}_{k,i}$ for $k = 1, 2$, with the mesh-size $\mathfrak{h}_{k,i} = \widetilde{\xi}_{k,i} - \widetilde{\xi}_{k,i-1}$, is introduced by coarsening only in one direction ξ_2. In other words, it is the same uniform grid in the direction ξ_1 with $\mathfrak{h}_{1,i} \equiv \hbar_1 \equiv h_1$ and nonuniform in the direction ξ_2 with the sizes $\mathfrak{h}_{2,j}$ as much close as possible to $\hbar_1$. The mesh lines $\xi_2 \equiv \widetilde{\xi}_{2,j}$ can be defined as follows. We find $m_2 = \text{integer}\,[\widetilde{\epsilon}/h_1]$, then define the uniform mesh $\zeta_{2,j} = j\,\widetilde{\epsilon}\,/m_2,\ j = 0, 1, \ldots, m_2,$ and then shift the lines of this uniform nonembedded coarse mesh $\xi_2 = \zeta_{2,j}$ to the nearest lines $\xi_2 \equiv \xi_{2,l} = l\hbar_2$ of the transformed source mesh of the size $\hbar_2$ in the direction ξ_2. We retain the notation m_2 for the number of the rarefied mesh intervals in the direction ξ_2 whereas the number of the rarefied mesh intervals in the direction ξ_1 is $m_1 = n_1$. Obviously, the sizes of this mesh satisfy inequalities

$$\underline{c}h_1 \le \mathfrak{h}_{k,i} \le \overline{c}h_1, \quad k = 1, 2, \tag{6.101}$$

with positive constants $\underline{c}$ and $\overline{c}$, for which we retain the notations from (6.38). The situation is similar as shown in Figure 6.3.

The space $\mathbb{V}(\Omega_\xi)$ may be represented by the direct sum

$$\mathbb{V}(\Omega_\xi) = \mathbb{V}_{\rm r}(\Omega_\xi) \oplus \mathbb{W}(\Omega_\xi), \tag{6.102}$$

where $\mathbb{V}_{\rm r}(\Omega_\xi)$ is the space of FE functions which are continuous on $\overline{\Omega}_\xi$ and bilinear on each element of the rarefied grid. Obviously, the space $\mathbb{W}(\Omega_\xi)$ contains FE functions which vanish on the lines $\xi_2 \equiv \widetilde{\xi}_{2,j}$ of the rarefied grid. Let $\mathbb{V}_{\rm tr}(\partial\Omega_\xi)$ be the space of traces of functions from $\mathbb{V}(\Omega_\xi)$ on $\partial\Omega_\xi$. We now introduce the norms

$$|u|_{1,\Omega_\xi} = |\!|\!|u|\!|\!|_{\Omega_\xi} = (\widetilde{\alpha}(u,u))^{1/2} \quad \text{and} \quad |v|\!|\!|_{h,\partial\Omega_\xi} = \inf_{\phi\in\mathbb{V}(\Omega_\xi)\,;\,\phi|_{\partial\Omega}=v} |\!|\!|\phi|\!|\!|_{\Omega_\xi}\,. \tag{6.103}$$

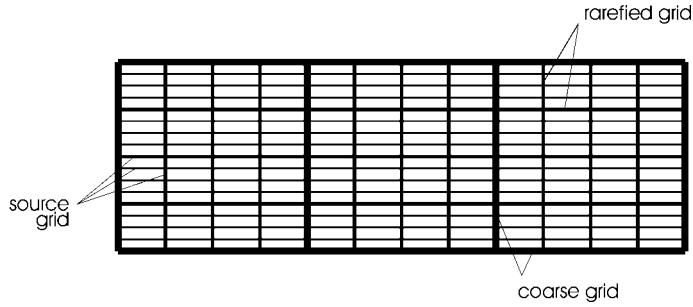

Fig. 6.3 Transformed rectangular domain and the source, rarefied and coarse grids.

for $u \in \mathbb{V}(\Omega_\xi)$ and $v \in \mathbb{V}_{\rm tr}(\partial\Omega_\xi)$, respectively.

The matrix $\mathbb{Q}$ corresponding to the quadratic form $|u|^2_{\Omega_\xi}$, $u \in \mathbb{V}(\Omega_\xi)$, can be represented in the block form

$$\mathbb{Q} = \begin{pmatrix} \mathbb{Q}_{\rm s} & \mathbb{Q}_{\rm sr} \\ \mathbb{Q}_{\rm rs} & \mathbb{Q}_{\rm r} \end{pmatrix}, \tag{6.104}$$

with blocks $\mathbb{Q}_{\rm s}$ and $\mathbb{Q}_{\rm r}$ corresponding to subspaces $\mathbb{W}(\Omega_\xi)$ and $\mathbb{V}_{\rm r}(\Omega_\xi)$, respectively. It is assumed that the basis in the subspace $\mathbb{V}_{\rm r}(\Omega_\xi)$ is the nodal basis corresponding to the rarefied grid. We also remind that in (6.57) and in other block representations, like the one above, different orderings of the degrees of freedom, convenient for the specific purposes, are used.

Let $\nu_{2,j}$ be the number of the fine mesh intervals on the rarefied mesh interval $(\widetilde{\xi}_{2,j-1}, \widetilde{\xi}_{2,j})$ and

$$\boldsymbol{\Delta}_{2,j} = \mathrm{tridiag}\,[-1, 2, -1]_1^{\nu_{2,j}-1}. \tag{6.105}$$

An intermediate preconditioner $\boldsymbol{\mathcal{Q}}_{\rm d}$ for $\mathbb{Q}$ may be defined in the following block form:

$$\boldsymbol{\mathcal{Q}}_{\rm d} = \mathrm{diag}\,[\boldsymbol{\mathcal{Q}}_{\rm s}, \mathbb{Q}_{\rm r}], \quad \boldsymbol{\mathcal{Q}}_{\rm s} = \frac{\hbar_1}{\hbar_2}\,\mathrm{diag}\,[\underbrace{\boldsymbol{\Delta}_{2,j}, \boldsymbol{\Delta}_{2,j}, \ldots, \boldsymbol{\Delta}_{2,j}}_{(n_1+1)\ \text{times}}]_{j=1}^{m_2}. \tag{6.106}$$

Note that $\ker \boldsymbol{\mathcal{Q}}^* = \ker \mathbb{Q}_{\rm r}$. For a given j, the i-th block $\boldsymbol{\Delta}_{2,j}$ in the square brackets is related to the nodes on the interval $(\widetilde{\xi}_{2,j-1}, \widetilde{\xi}_{2,j})$ of the mesh line $\xi_1 \equiv \widetilde{\xi}_{1,i}$.

In case *ii*), the rarefied quasiuniform mesh of the characteristic size $\widetilde{\epsilon}$ does not exist, and we introduce the *transformed rarefied uniform rectangular mesh* of the sizes h_1 and $\widetilde{\epsilon}$. Therefore, we have only one layer of n_1 cells $h_1 \times \widetilde{\epsilon}$, meaning $m_2 = 1$, and, similarly to (6.106), we can set

$$\boldsymbol{\mathcal{Q}}_{\rm d} = \mathrm{diag}\,[\boldsymbol{\mathcal{Q}}_{\rm s}, \mathbb{Q}_{\rm r}], \quad \boldsymbol{\mathcal{Q}}_{\rm s} = \frac{\hbar_1}{\hbar_2}\,\mathrm{diag}\,[\underbrace{\boldsymbol{\Delta}_2, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_2}_{(n_1+1)\ \text{times}}], \tag{6.107}$$

with the $(n_2-1)\times(n_2-1)$ blocks $\boldsymbol{\Delta}_2$, in the notation of which second lower index 1 is omitted. The matrix $\mathbb{Q}_{\rm r}$ is defined on the uniform rectangular coarse transformed grid, which coincides with the rarefied transformed grid and has all nodes on $\partial\Omega_\xi$. This matrix is block-tridiagonal with 2×2 blocks, and does not require any preconditioning.

In case *iii*), the mesh parameter for the quasiuniform rectangular rarefied grid is $\hbar_2$, and satisfies the relations

$$\underline{c}\hbar_2 \le \mathfrak{h}_{k,i} \le \overline{c}\hbar_2, \quad k=1,2, \tag{6.108}$$

with positive constants $\underline{c}$ and $\underline{c}$. At the proper ordering of unknowns, we have again $\boldsymbol{\mathcal{Q}}_{\rm d} = \mathrm{diag}\,[\boldsymbol{\mathcal{Q}}_{\rm s}, \mathbb{Q}_{\rm r}]$, but now with

$$\boldsymbol{\mathcal{Q}}_{\rm s} = \frac{\hbar_2}{\hbar_1}\,\mathrm{diag}\,[\underbrace{\boldsymbol{\Delta}_{1,i}, \boldsymbol{\Delta}_{1,i}, \ldots, \boldsymbol{\Delta}_{1,i}}_{(n_2+1)\ \text{times}}\,]_{j=1}^{m_1}, \tag{6.109}$$

where $\boldsymbol{\Delta}_{1,i} = \mathrm{tridiag}\,[-1,2,-1]_1^{\nu_{1,i}-1}$ and $\nu_{1,i}$ is the number of the fine mesh intervals on the coarse mesh interval $(\widetilde{\xi}_{1,i-1}, \widetilde{\xi}_{1,i})$. For a given i, the j-th block $\boldsymbol{\Delta}_{1,i}$ in the square brackets is related to the nodes on the interval $(\widetilde{\xi}_{1,i-1}, \widetilde{\xi}_{1,i})$ of the mesh line $\xi_2 \equiv \widetilde{\xi}_{2,j}$.

Lemma 6.6. *For any positive h_1, h_2, ρ_1, ρ_2 and ϵ, we have*

$$\boldsymbol{\mathcal{Q}}_{\rm d} \prec \mathbb{Q} \prec \boldsymbol{\mathcal{Q}}_{\rm d}\,. \tag{6.110}$$

Proof. The estimate from above. We will prove estimate (6.110) from above for case *i*) only, since cases *ii*) and *iii*) can be treated similarly. Obviously, matrix $\mathbb{Q}$ can be interpreted as assembled from the stiffness matrices of superelements on the rarefied grid nests $\delta_{i,j} := (\widetilde{\xi}_{1,i-1}, \widetilde{\xi}_{1,i}) \times (\widetilde{\xi}_{2,j-1}, \widetilde{\xi}_{2,j})$. We consider one of such nests, and denote the corresponding stiffness matrix by $\boldsymbol{\Theta}$. For simplicity, we omit the indices i,j in this and some other notations. Let $\nu_{2,j}$ denote the number of the transformed source grid intervals on the rarefied grid interval $(\widetilde{\xi}_{2,j-1}, \widetilde{\xi}_{2,j})$. The matrix $\boldsymbol{\Theta}$ can be represented in the form

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_1 \otimes \mathbb{M}_2 + \mathbb{M}_1 \otimes \boldsymbol{\Theta}_2, \tag{6.111}$$

where $\boldsymbol{\Theta}_k$ are the 1d stiffness matrices

$$\boldsymbol{\Theta}_1 = \frac{1}{h_1}\boldsymbol{\Delta}_1 \quad \text{and} \quad \boldsymbol{\Theta}_2 = \mathrm{diag}\,[\frac{1}{\hbar 2}\boldsymbol{\Delta}_2, \frac{1}{\mathfrak{h}_{1,i}}\boldsymbol{\Delta}_1], \tag{6.112}$$

with

$$\boldsymbol{\Delta}_1 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Delta}_2 = \mathrm{tridiag}\,[-1,2,-1]_1^{\nu_{2,j}-1},$$

and $\mathbb{M}_k$ are the corresponding 1d mass matrices. If we introduce $\mathbf{M}_1$ by the equality $\mathbb{M}_1 = h_1\mathbf{M}_1$, then the other mass matrix $\mathbb{M}_2$ may be represented in the form

$$\mathbb{M}_2 = \begin{pmatrix} \mathbb{M}_{\rm s} & \mathbb{M}_{\rm s,r} \\ \mathbb{M}_{\rm r,s} & \mathbb{M}_{\rm r} \end{pmatrix} = \begin{pmatrix} \hbar_2\mathbf{M}_2 & \mathbb{M}_{\rm s,r} \\ \mathbb{M}_{\rm r,s} & \mathfrak{h}_{2,j}\mathbf{M}_1 \end{pmatrix},$$

and it is clear that the inequality

$$\mathbb{M}_2 \le 2\,\mathrm{diag}\,[\hbar_2\mathbf{M}_2,\ \mathfrak{h}_{2,j}\mathbf{M}_1] \tag{6.113}$$

holds. From (6.112) and (6.113), we observe that

$$\begin{aligned} \mathbf{\Theta}_1 \otimes \mathbb{M}_2 &\le 2\,\mathrm{diag}\,[\hbar_2\mathbf{\Theta}_1 \otimes \mathbf{M}_2,\ \mathfrak{h}_{2,j}\mathbf{\Theta}_1 \otimes \mathbf{M}_1] \\ &\le 2\,\mathrm{diag}\,[\frac{2\hbar_2}{h_1}\mathbf{M}_2,\ \frac{2\hbar_2}{h_1}\mathbf{M}_2,\ \mathfrak{h}_{2,j}\mathbf{\Theta}_1 \otimes \mathbf{M}_1]\,. \end{aligned} \tag{6.114}$$

The first two matrices in the square brackets on the right are related to the vertical edges on the two lines $\xi_1 \equiv \xi_{1,i-1}, \xi_{1,i}$. We have also the inequality

$$\hbar_2\mathbf{M}_2 \le \frac{\mathfrak{h}_{2,j}^2}{\pi^2\hbar_2}\mathbf{\Delta}_2,$$

which is obviously the discrete form of the inequality

$$\int_{\widetilde{\xi}_{2,j-1}}^{\widetilde{\xi}_{2,j}} v^2(x)dx \le \frac{\mathfrak{h}_{2,j}^2}{\pi^2}\int_{\widetilde{\xi}_{2,j-1}}^{\widetilde{\xi}_{2,j}} (v')^2(x)dx$$

for piecewise linear functions. Combining this inequality with (6.114) and (6.101), we get

$$\frac{\hbar_2}{h_1}\mathbf{M}_2 \le \frac{\mathfrak{h}_{2,j}^2}{\pi^2 h_1\hbar_2}\mathbf{\Delta}_2 \le \frac{\overline{c}}{\pi^2}\,\frac{h_1}{\hbar_2}\mathbf{\Delta}_2, \tag{6.115}$$

and, therefore,

$$\mathbf{\Theta}_1 \otimes \mathbb{M}_2 \le 2\mathrm{diag}\,[\frac{2\overline{c}}{\pi^2}\,\frac{h_1}{\hbar_2}\mathbf{\Delta}_2,\ \frac{2\overline{c}}{\pi^2}\,\frac{h_1}{\hbar_2}\mathbf{\Delta}_2,\ \mathfrak{h}_{2,j}\mathbf{\Theta}_1 \otimes \mathbf{M}_1]\,. \tag{6.116}$$

Now we estimate the second term in (6.111). Using the right inequality of the obvious inequalities

$$\frac{h_1}{6}\mathbf{I} \le \mathbb{M}_1 \le \frac{h_1}{2}\mathbf{I}, \tag{6.117}$$

we can write

$$\begin{aligned} \mathbb{M}_1 \otimes \mathbf{\Theta}_2 &= \mathbb{M}_1 \otimes \mathrm{diag}\,[\frac{1}{\hbar_2}\mathbf{\Delta}_2,\ \frac{1}{\mathfrak{h}_{1,i}}\mathbf{\Delta}_1] \\ &= \mathrm{diag}\,[\mathbb{M}_1 \otimes \frac{1}{\hbar_2}\mathbf{\Delta}_2,\ \mathbb{M}_1 \otimes \frac{1}{\mathfrak{h}_{1,i}}\mathbf{\Delta}_1] \\ &\le \mathrm{diag}\,[\frac{h_1}{2\hbar_2}\mathbf{\Delta}_2,\ \frac{h_1}{2\hbar_2}\mathbf{\Delta}_2,\ \mathbb{M}_1 \otimes \frac{1}{\mathfrak{h}_{1,i}}\mathbf{\Delta}_1]\,. \end{aligned} \tag{6.118}$$

The preconditioner $\boldsymbol{\mathcal{Q}}_{\rm d}$ for one nest of the rarefied mesh has the form

$$\boldsymbol{\mathfrak{Q}} = \mathrm{diag}\,[\boldsymbol{\mathfrak{Q}}_{\rm s}^{(2)}, \boldsymbol{\Theta}_{\rm r}], \tag{6.119}$$

where $\boldsymbol{\mathfrak{Q}}_{\rm s}^{(2)} = h_1\hbar_2^{-1}\mathrm{diag}\,[\boldsymbol{\Delta}_2,\ \boldsymbol{\Delta}_2]$ and $\boldsymbol{\Theta}_{\rm r}$ is the 4×4 block of the matrix $\boldsymbol{\Theta}$ for the vertices of the nest $\delta_{i,j}$. Now, multiplying (6.118) by 2 and summing with (6.116), we get

$$\begin{aligned}\boldsymbol{\Theta} &\le 2\,\mathrm{diag}\,[c_1\,\frac{h_1}{\hbar_2}\boldsymbol{\Delta}_2,\ c_1\,\frac{h_1}{\hbar_2}\boldsymbol{\Delta}_2,\ \boldsymbol{\Theta}_{\rm r}]\\ &\le c_2\,\mathrm{diag}\,[\frac{h_1}{\hbar_2}\boldsymbol{\Delta}_2,\ \frac{h_1}{\hbar_2}\boldsymbol{\Delta}_2,\ \boldsymbol{\Theta}_{\rm r}] = c_2\boldsymbol{\mathfrak{Q}},\end{aligned} \tag{6.120}$$

with $c_2 = 2c_1$ and $c_1 = 1 + 2\overline{c}/\pi^2$. Assembling matrices $\boldsymbol{\mathfrak{Q}}$, which are defined for each nest, we come to $\mathbb{Q}$ on the left and to the block diagonal matrix $\boldsymbol{\mathcal{Q}}_{\rm d}$ on the right. Therefore,

$$\mathbb{Q} \le c_2\boldsymbol{\mathcal{Q}}_{\rm d}\,. \tag{6.121}$$

The estimate from below. The subdomain stiffness matrix $\boldsymbol{\Theta}$ can be represented in the similar to (6.104) block form

$$\boldsymbol{\Theta} = \begin{pmatrix}\boldsymbol{\Theta}_{\rm s} & \boldsymbol{\Theta}_{\rm sr}\\ \boldsymbol{\Theta}_{\rm rs} & \boldsymbol{\Theta}_{\rm r}\end{pmatrix},$$

and, at the same time, by the sum $\boldsymbol{\Theta} = \boldsymbol{\Theta}^{(1)} + \boldsymbol{\Theta}^{(2)}$, where $\boldsymbol{\Theta}^{(k)}$ is generated by that part of the integrals in $\widetilde{\alpha}(\cdot,\cdot)$ containing derivatives over ξ_k. From (6.111) and (6.112), it follows that second summand is block diagonal, so that

$$\boldsymbol{\Theta}^{(2)} = \mathrm{diag}\,[\boldsymbol{\Theta}_{\rm s}^{(2)}, \boldsymbol{\Theta}_{\rm r}^{(2)}] = \mathbb{M}_1 \otimes \boldsymbol{\Theta}_2 \le \boldsymbol{\Theta}. \tag{6.122}$$

Taking the left inequality of (6.117) into account, we get

$$\frac{1}{6}\boldsymbol{\mathfrak{Q}}_{\rm s}^{(2)} = \frac{h_1}{6\hbar_2}\mathrm{diag}\,[\,\boldsymbol{\Delta}_{2,j},\ \boldsymbol{\Delta}_{2,j}] \le \frac{h_1}{\hbar_2}\mathbf{M}_1 \otimes \boldsymbol{\Delta}_{2,j} = \boldsymbol{\Theta}_{\rm s}^{(2)} \le \boldsymbol{\Theta}\,. \tag{6.123}$$

Inequalities (6.114)–(6.116) imply that

$$\boldsymbol{\Theta}_{\rm s}^{(1)} \le \frac{4\overline{c}}{\pi^2}\,\frac{h_1}{\hbar_2}\mathrm{diag}\,[\boldsymbol{\Delta}_2,\ \boldsymbol{\Delta}_2] = \frac{4\overline{c}}{\pi^2}\,\boldsymbol{\mathfrak{Q}}_{\rm s}^{(2)} \le \frac{24\overline{c}}{\pi^2}\boldsymbol{\Theta}_{\rm s}^{(2)}\,.$$

These estimates, the preceding inequality and the representation $\boldsymbol{\Theta}_{\rm s} = \boldsymbol{\Theta}_{\rm s}^{(1)} + \boldsymbol{\Theta}_{\rm s}^{(2)}$ yield the estimates

$$c_3\boldsymbol{\Theta}_{\rm s} \le \boldsymbol{\Theta} \quad \text{and} \quad c_4\boldsymbol{\Theta}_{\rm r} \le \boldsymbol{\Theta}$$

with constants $c_3 = 1/(1 + 24\overline{c}/\pi^2)$ and $c_4 = c_3/2(1 + c_3)$. Therefore, we have

$$\mathrm{diag}\,[\boldsymbol{\Theta}_{\rm s},\ \boldsymbol{\Theta}_{\rm r}] \le 2(1 + 1/c_3)\boldsymbol{\Theta} \le (4/c_3)\boldsymbol{\Theta}\,. \tag{6.124}$$

Finally, by the use of (6.123) and (6.119), the bound (6.124) is transformed to

$$\frac{c_3}{24}\boldsymbol{\mathfrak{Q}} \le \boldsymbol{\Theta},$$

the direct concequence of which is the left bound in (6.110). □

6.3.2 *Schur Complement Preconditioner*

Lemma 6.6 directly leads to the block diagonal preconditioner spectrally equivalent to the Schur complement $\mathbb{Y}$ in (6.99). For illustrating this, we turn to case i). Similar to (6.57) and (6.58), the block representations for $\mathbb{Q}_{\rm r}$ and $\boldsymbol{\mathcal{Q}}_{\rm d}$ and their Schur complements, for which we use the notations $\mathbb{Y}_{\rm r}$ and $\boldsymbol{\mathcal{G}}_{\rm s}$, respectively, are

$$\begin{aligned} &\mathbb{Q}_{\rm r} = \begin{pmatrix} \mathbb{Q}^{\rm r}_{I} & \mathbb{Q}^{\rm r}_{I,B} \\ \mathbb{Q}^{\rm r}_{B,I} & \mathbb{Q}^{\rm r}_{B} \end{pmatrix}, \qquad && \boldsymbol{\mathcal{Q}}_{\rm s} = \begin{pmatrix} \boldsymbol{\mathcal{Q}}^{\rm s}_{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{Q}}^{\rm s}_{B} \end{pmatrix}, \\ &\mathbb{Y}_{\rm r} = \mathbb{Q}^{\rm r}_{B} - \mathbb{Q}^{\rm r}_{B,I}(\mathbb{Q}^{\rm r}_{I})^{-1}\mathbb{Q}^{\rm r}_{I,B}, \quad && \boldsymbol{\mathcal{G}}_{\rm s} = \boldsymbol{\mathcal{Q}}^{\rm s}_{B}. \end{aligned} \tag{6.125}$$

According to (6.106), the preconditioner $\boldsymbol{\mathcal{Q}}_{\rm s}$ is block diagonal, and the internal degrees of freedom are not coupled with the boundary ones. Therefore, for the Schur complement preconditioner, we can take

$$\boldsymbol{\mathcal{G}} = \operatorname{diag}[\boldsymbol{\mathcal{G}}_{\rm s}, \mathbb{Y}_{\rm r}], \tag{6.126}$$

with

$$\boldsymbol{\mathcal{G}}_{\rm s} = \frac{h_1}{h_2} \operatorname{diag}[\,\boldsymbol{\Delta}_{2,j}, \boldsymbol{\Delta}_{2,j}\,]_{j=1}^{m_2}, \tag{6.127}$$

where the two matrices in the square brackets correspond to the nodes in the interval $(\widetilde{\xi}_{2,j-1}, \widetilde{\xi}_{2,j})$ on the left and right vertical edges of $\partial\Omega_\xi$.

It is clear that, for any rectangle Ω and $\rho_k, h_k > 0$, the Schur complement preconditioner $\boldsymbol{\mathcal{G}}_{\rm s}$ is uniquely defined in the block diagonal form (6.127) with $\boldsymbol{\mathcal{G}}_{\rm s}$ and $\mathbb{Y}_{\rm r}$ defined as in (6.125). Explicit forms of $\boldsymbol{\mathcal{Q}}^{\rm s}_{B}$ follow from (6.106), (6.107) and (6.109).

Corollary 6.4. *For all positive h_1, h_2, ρ_1, ρ_2 and ϵ, we have*

$$\boldsymbol{\mathcal{G}} \prec \mathbb{Y} \prec \boldsymbol{\mathcal{G}}\,. \tag{6.128}$$

Proof. The proof follows from the definition of $\boldsymbol{\mathcal{G}}$ and Lemma 6.6. □

Preconditioning the Schur complement in the case of an arbitrary rectangle Ω, arbitrary positive parameters ρ_k and h_k, and quasiuniform discretization mesh was studied in the preceding Section 6.2. Since the Schur complement $\mathbb{Y}_{\rm r}$ exactly corresponds to this case, Corollary 6.4 allows us to apply the results obtained there to $\mathbf{B}$. Before doing this, we will make a remark reminding the spaces and decompositions, which underlie the preconditioners.

The described approach is based on the sequence of the FE spaces which is related to the sequence of the image spaces

$$\mathbb{V} = \mathbb{V}_{\rm r} \oplus \mathbb{W}, \qquad \mathbb{V}_{\rm r} = \mathbb{V}_{\rm c} \oplus \mathbb{W}_{\rm r}, \qquad \mathbb{V}_{\rm c} = \mathbb{V}_0 \oplus \mathbb{W}_{\rm c},$$
$$\mathbb{V} = \mathbb{W} \oplus \mathbb{W}_{\rm r} \oplus \mathbb{W}_{\rm c} \oplus \mathbb{V}_0, \qquad \mathbb{V}_0 \subseteq \mathbb{V}_{\rm c} \subseteq \mathbb{V}_{\rm r} \subseteq \mathbb{V}, \tag{6.129}$$

defined for the transformed problem and its discretization on the transformed subdomain $\Omega_\xi = \Omega_{j,\xi}$. In other words, the spaces $\mathbb{V}(\Omega_\xi)$, $\mathbb{V}_{\rm r}(\Omega_\xi)$, $\mathbb{V}_{\rm c}(\Omega_\xi)$, and $\mathbb{V}_0(\Omega_\xi)$ are induced by the transformed source FE mesh, by the rarefied and coarse meshes imbedded in it, and by the space of bilinear functions on Ω_ξ, respectively. The spaces $\mathbb{V}_{\rm r}(\Omega_\xi), \ldots, \mathbb{W}_{\rm c}(\Omega_\xi)$ define preimage spaces, for which the notations $\mathsf{V}_{\rm r}(\Omega), \ldots, \mathsf{W}_{\rm c}(\Omega)$ are used. The spaces $\mathcal{V}(\Omega)$ of preceding sections correspond to the space $\mathbb{V}_{\rm r}(\Omega_\xi)$ induced by the rarefied mesh.

The decomposition (6.129) of the transformed FE space $\mathbb{V}(\Omega_\xi)$ implies the corresponding decomposition of the source FE space

$$\mathcal{V}(\Omega) = \mathsf{W}(\Omega) \oplus \mathsf{W}_{\rm r}(\Omega) \oplus \mathsf{W}_{\rm c}(\Omega) \oplus \mathsf{V}_0(\Omega) \tag{6.130}$$

and the corresponding basis, which can be termed hierarchical. In what follows, we prefer to use special notations $\mathbb{Y}_{\rm Hi}$, $\mathbb{Y}_{\rm Hi,r}$, $\boldsymbol{\mathcal{G}}_{\rm Hi}$ for the Schur complements $\mathbb{Y}$, $\mathbb{Y}_{\rm r}$ and the preconditioner $\boldsymbol{\mathcal{G}}$ written in the hierarchical basis. Obviously, the space $\mathbb{V}_{\rm r}(\Omega_\xi)$ in the corresponding decomposition (6.129) possesses the same properties as the space $\mathcal{V}(\Omega)$ in Section 6.2. In particular, the decomposition $\mathbb{V}_{\rm r} = \mathbb{W}_{\rm r} \oplus \mathbb{W}_{\rm c} \oplus \mathbb{V}_0$ is the decomposition (6.74). Thus, for $\mathbb{Y}_{\rm Hi,r}$, we can use the preconditioners constructed there for $\mathbf{B}_{\rm Hi}$, *e.g.*, the preconditioner $\boldsymbol{\mathcal{C}}_{\rm Hi}$ of Lemma 6.3.

Theorem 6.4. *Let*

$$\boldsymbol{\mathcal{G}}_{\rm Hi} = \mathrm{diag}\,[\boldsymbol{\mathcal{G}}_{\rm s}, \boldsymbol{\mathcal{C}}_{\rm Hi}] = \mathrm{diag}\,[\boldsymbol{\mathcal{G}}_{\rm s}, \boldsymbol{\mathcal{B}}_W, \mathbf{B}^{(\rm w)}, \mathbf{B}^{(\rm v)}]\,. \tag{6.131}$$

Then the inequalities

$$\frac{1}{1+\log \underline{n}}\,\min\Big(\frac{1}{n_\epsilon}, \frac{1}{1+\log \underline{n}}\Big)\,\boldsymbol{\mathcal{G}}_{\rm Hi} \prec \mathbb{Y}_{\rm Hi} \prec \boldsymbol{\mathcal{G}}_{\rm Hi} \tag{6.132}$$

hold for all $\epsilon \in (0,1]$ *and* $h \le (0,\epsilon]$.

Proof. The proof immediately follows from the estimates given in Lemmas 6.3 and 6.6. □

Other preconditioners of the two preceding subsections, *e.g.*, those obtained by means of the overlapping decompositions, can be used for peconditioning the matrix $\mathbb{Y}_{\rm Hi}$. We mention that they can provide slightly asymptotically better bounds of the relative spectrum, compare (6.132) with second bound of Lemma 6.5.

The construction of the preconditioner and the bounds (6.132) can be easily adjusted to the more general case of $\Omega = H_1 \times H_2$. In the same way, we introduce the rarefied source and the coarsest meshes whereas the role of ϵ is played by $\min_k (H_k/H_{3-k})$. The above described form of the preconditioner is retained, if, after the transformation to the isotropic problem, the shortest edge is directed along the axis x_2. In the general case, for all positive H_k, $h_k \le H_k$, ρ_k, the counterpart of (6.132) reads as follows

$$\underline{\mu}\, \boldsymbol{\mathcal{G}}_{\mathrm{Hi}} \prec \mathbb{Y}_{\mathrm{Hi}} \prec \boldsymbol{\mathcal{G}}_{\mathrm{Hi}}, \tag{6.133}$$

where

$$\underline{\mu} = \frac{1}{(1+\log \underline{m})} \min\,(\frac{1}{\theta}, \frac{1}{(1+\log \underline{m})}),$$

with

$$\theta = \max_k \min\,(m_k, \frac{H_k\sqrt{\rho_{3-k}}}{H_{3-k}\sqrt{\rho_k}})\,.$$

Solving systems of linear algebraic equations with the preconditioner $\boldsymbol{\mathcal{G}}_{\mathrm{Hi}}$ as system matrix is cheap. Indeed, using FDFT, we need not more than $\mathcal{O}(\overline{n}\log \underline{n})$ a.o. The relative condition depends on the ratio θ of the maximal to the minimal lengths of the edges of the rectangle Ω_ξ. However, this dependence will not prevent us to contrive an almost optimal DD preconditioner-solver for piecewise variable orthotropic discretizations.

6.3.3 *Compatible Schur Complement Preconditioner-Solver for Subdomains*

Turning to problem (6.95), we are going to justify the Schur complement preconditioner-solver, in which the d.o.f. corresponding to the vertices of the domain are split from the remaining d.o.f. Additionally, we will decouple d.o.f. of any pair of adjacent edges and d.o.f. of the pair of short parallel edges. In distinction with the previously considered preconditioners, which provide the same type of block diagonalization, the new one, when incorporated in the solver for the problem (6.4) as a subdomain preconditioner, will be compatible with the same type preconditioners of neighboring subdomains Ω_j. This compatibility property is important for assembling subdomain preconditioners in an efficient global preconditioner-solver.

It is convenient to introduce some component preconditioners, entering the DD preconditioner, as defined by FE piecewise liner functions, because they are simpler and only slightly worsen the relative condition number.

We call by the source triangulation the one obtained by subdivision of each element of the source rectangular mesh in two triangles by one of the two diagonals, for simplicity, of the same direction for all elements. In this subsection, the notation $\mathcal{U}(\Omega)$ is used for the space of continuous functions which are linear on each triangle of the source triangulation. This space can represented by the direct sums

$$\mathcal{U}(\Omega) = \mathcal{U}_I(\Omega) \oplus \mathcal{U}^{\mathrm{B}}(\Omega), \quad \mathcal{U}^{\mathrm{B}}(\Omega) = \mathcal{U}^{\mathrm{E}}(\Omega) \oplus \mathcal{U}^{\mathrm{V}}(\Omega),$$

spanned over internal, boundary, edge and vertex FE functions, which are the nodal basis for the source triangulation. In correspondence to these representations of the space $\mathcal{U}(\Omega)$, the stiffness matrix $\mathbf{L}$ arising from the bilinear form in (6.95) on the space $\mathcal{U}(\Omega)$ can be represented in the block forms

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_I & \mathbf{L}_{IB} \\ \mathbf{L}_{BI} & \mathbf{L}_B \end{pmatrix} = \begin{pmatrix} \mathbf{L}_I & \mathbf{L}_{IE} & \mathbf{L}_{IV} \\ \mathbf{L}_{EI} & \mathbf{L}_E & \mathbf{L}_{EV} \\ \mathbf{L}_{VI} & \mathbf{L}_{VE} & \mathbf{L}_V \end{pmatrix}, \tag{6.134}$$

with zero blocks $\mathbf{L}_{IV} = \mathbf{L}_{VI}^{\top} = \mathbf{0}$. Therefore, the Schur complement $\mathbf{Z}_B = \mathbf{L}_B - \mathbf{L}_{BI}\mathbf{L}_I^{-1}\mathbf{L}_{IB}$ has the form

$$\mathbf{Z}_B = \begin{pmatrix} \mathbf{Z}_E & \mathbf{L}_{EV} \\ \mathbf{L}_{VE} & \mathbf{L}_V \end{pmatrix}, \quad \text{with } \mathbf{Z}_E = \mathbf{L}_E - \mathbf{L}_{EI}\mathbf{L}_I^{-1}\mathbf{L}_{IE},$$

and, due to the spectral equivalence $\mathbf{L} \prec \mathbf{Q} \prec \mathbf{L}$, we have

$$\mathbf{Z}_B \prec \mathbf{Y} \prec \mathbf{Z}_B\,.$$

Here it is assumed that $\mathbf{Y}$ is written in the nodal basis for the source rectangular grid.

Let us represent $\mathbf{Z}_E$ in the 4×4 block form

$$\mathbf{Z}_E = \left\{\mathbf{Z}^{\mathrm{E}}_{k,l}\right\}_{k,l=0}^{3}$$

with the blocks corresponding to the edges γ_k. We note that all 16 blocks are nonzero, see, *e.g.*, [Korneev (2002b)] and [Rytov (2006)] for more details. For the diagonal blocks, we will also use the simplified notations $\mathbf{Z}^{\mathrm{E}}_k = \mathbf{Z}^{\mathrm{E}}_{k,k}$. If θ in (6.133) is not too large, say $\theta \le 2$, then it is possible to use the preconditioner

$$\boldsymbol{\mathcal{Z}}^{\mathrm{E}} = \mathrm{diag}\left[\mathbf{Z}^{\mathrm{E}}_0, \mathbf{Z}^{\mathrm{E}}_1, \mathbf{Z}^{\mathrm{E}}_2, \mathbf{Z}^{\mathrm{E}}_3\right].$$

If $\theta > 2$, then any of the two preconditioners

$$\boldsymbol{\mathcal{Z}}^{\mathrm{E}} = \begin{pmatrix} \mathbf{Z}^{\mathrm{E}}_0 & \mathbf{Z}^{\mathrm{E}}_{0,1} & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}^{\mathrm{E}}_{1,0} & \mathbf{Z}^{\mathrm{E}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}^{\mathrm{E}}_2 & \mathbf{Z}^{\mathrm{E}}_{2,3} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}^{\mathrm{E}}_{3,2} & \mathbf{Z}^{\mathrm{E}}_3 \end{pmatrix} \quad \text{or} \quad \boldsymbol{\mathcal{Z}}^{\mathrm{E}} = \begin{pmatrix} \mathbf{Z}^{\mathrm{E}}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}^{\mathrm{E}}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}^{\mathrm{E}}_2 & \mathbf{Z}^{\mathrm{E}}_{2,3} \\ \mathbf{0} & \mathbf{0} & \mathbf{Z}^{\mathrm{E}}_{3,2} & \mathbf{Z}^{\mathrm{E}}_3 \end{pmatrix} \tag{6.135}$$

can be used. The first one is obtained by decoupling adjacent edges, whereas the second preconditioner additionally assumes decoupling of the pair of parallel edges which became shortest after mapping to Ω_ξ, *i.e.,* the edges γ_0 and γ_1 in the above expression. The right preconditioner in (6.135) was introduced by [Khoromskij and Wittum (1999)], see also [Khoromskij and Wittum (2004)].

We note that edge Schur complement preconditioners can similarly be defined by means of the matrices $\mathbf{Q}$ or $\widetilde{\mathbf{Q}}$, where the latter one is obtained from $\mathbf{Q}$ by setting $\mathbf{Q}_{IV} = \mathbf{Q}_{VI}^\top = \mathbf{0}$.

The Schur complement preconditioner-solver

$$\boldsymbol{\mathcal{Z}}_B = \mathrm{diag}\,[\boldsymbol{\mathcal{Z}}^{\mathrm{E}}, \boldsymbol{\mathcal{Z}}_0], \quad \text{with} \quad \boldsymbol{\mathcal{Z}}_0 = \sqrt{\rho_1\rho_2}\,\mathbf{B}_0, \tag{6.136}$$

corresponds to the two-level decomposition $\mathcal{U}^{\mathrm{B}}(\partial\Omega) = \mathcal{U}^{\mathrm{E}}(\partial\Omega) \oplus \mathcal{U}_0(\partial\Omega)$ of the boundary FE space that is defined by the traces of the spaces entering the decomposition $\mathcal{U}^{\mathrm{B}}(\Omega) = \mathcal{U}^{\mathrm{E}}(\Omega) \oplus \mathcal{U}_0(\Omega)$. Here, $\mathbf{B}_0$ is the matrix generated by the space $\mathcal{U}_0(\Omega)$ of continuous functions which are linear on each of the two triangles having vertices at the vertices of Ω. This matrix may be also generated by the subspace $\mathcal{V}_0(\Omega)$ of bilinear polynomials on Ω.

Theorem 6.5. *For all positive ρ_k, H_k, and $h_k \le H_k$, the preconditioners $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ and $\boldsymbol{\mathcal{Z}}_B$ satisfy the spectral equivalence inequalities*

$$\underline{\beta}_E\, \boldsymbol{\mathcal{Z}}^{\mathrm{E}} \prec \mathbf{Y}^E \prec \boldsymbol{\mathcal{Z}}^{\mathrm{E}}, \tag{6.137}$$

and

$$\underline{\mu}\boldsymbol{\mathcal{Z}}_B \prec \mathbf{Y} \prec \boldsymbol{\mathcal{Z}}_B, \tag{6.138}$$

respectively, with $\underline{\mu}$, defined in (6.133), and $\underline{\beta}_E = 1/(1+\log \underline{m})^2$.

Proof. We consider the case of $\theta > 2$ and the preconditioner $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ defined by the second expression (6.135). Let us suppose that there exists a preconditioner $\boldsymbol{\Upsilon}_B = \mathrm{diag}\,[\boldsymbol{\Upsilon}^{\mathrm{E}}, \boldsymbol{\Upsilon}_0]$ which has the structure similar to $\boldsymbol{\mathcal{Z}}_B$, *i.e.*,

$$\boldsymbol{\Upsilon}^{\mathrm{E}} = \begin{pmatrix} \boldsymbol{\Upsilon}_0^{\mathrm{E}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Upsilon}_1^{\mathrm{E}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Upsilon}_2^{\mathrm{E}} & \boldsymbol{\Upsilon}_{23}^{\mathrm{E}} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Upsilon}_{32}^{\mathrm{E}} & \boldsymbol{\Upsilon}_3^{\mathrm{E}} \end{pmatrix}, \tag{6.139}$$

and satisfies the spectral inequalities

$$\underline{\mu}\,\boldsymbol{\Upsilon}_B \le \mathbf{Y} \le \overline{\mu}\,\boldsymbol{\Upsilon}_B, \tag{6.140}$$

with some positive $\underline{\mu}$ and $\overline{\mu}$. Then, we have

$$(\underline{\mu}/\overline{\mu})\,\boldsymbol{\mathcal{Z}}_B \prec \mathbf{Y} \prec 4\overline{\mu}\,\boldsymbol{\mathcal{Z}}_B\,. \tag{6.141}$$

Indeed, let V_0, V_{E_0}, V_{E_1} and V_{E_2,E_3} be the vector spaces of d.o.f. corresponding to the independent blocks on the diagonal of the matrix $\boldsymbol{\mathcal{Z}}_B$. From (6.140), restricted to these subspaces, (6.135) and (6.139), it follows that

$$\underline{\mu}\,\boldsymbol{\Upsilon}^{\mathrm{E}} \prec \boldsymbol{\mathcal{Z}}^{\mathrm{E}} \prec \overline{\mu}\,\boldsymbol{\Upsilon}^{\mathrm{E}} \quad \text{and} \quad \underline{\mu}\,\boldsymbol{\Upsilon}_0 \prec \boldsymbol{\mathcal{Z}}_0 \prec \overline{\mu}\,\boldsymbol{\Upsilon}_0\,. \tag{6.142}$$

Combining (6.140) and (6.142), we get the estimates

$$\mathbf{Y} \geq \underline{\mu}\,\mathrm{diag}\,[\boldsymbol{\Upsilon}^{\mathrm{E}}, \boldsymbol{\Upsilon}_0] \geq (\underline{\mu}/\overline{\mu})\,\mathrm{diag}\,[\boldsymbol{\mathcal{Z}}^{\mathrm{E}}, \boldsymbol{\mathcal{Z}}_0] = (\underline{\mu}/\overline{\mu})\,\boldsymbol{\mathcal{Z}}_B\,.$$

This proves the left inequality (6.141). The right inequality (6.141) follows from the Cauchy inequality and the last inequality in (6.142).

Let us turn now to the block

$$\boldsymbol{\mathcal{G}}^{\mathrm{E}}_{\mathrm{Hi}} = \mathrm{diag}\,[\boldsymbol{\mathcal{G}}_{\mathrm{s}}, \boldsymbol{\mathcal{B}}_W, \mathbf{B}^{(\mathrm{w})}],$$

of the preconditioner $\boldsymbol{\mathcal{G}}_{\mathrm{Hi}}$ in Theorem 6.4. It has the same structure as $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$. If we repeat the derivation of (6.132), however, omitting the steps related to splitting vertices, then we come to the bounds

$$\underline{\beta}_E\,\boldsymbol{\mathcal{G}}^{\mathrm{E}}_{\mathrm{Hi}} \prec \mathbb{Y}^{\mathrm{E}} \prec \boldsymbol{\mathcal{G}}^{\mathrm{E}}_{\mathrm{Hi}}\,. \tag{6.143}$$

Therefore, one can pick up $\sqrt{\rho_1\rho_2}\,\boldsymbol{\mathcal{G}}^{\mathrm{E}}_{\mathrm{Hi}}$ for $\boldsymbol{\Upsilon}^{\mathrm{E}}$ and obtain (6.137).

Similarly, on the basis of (6.141) and Theorem 6.4, we prove inequalities (6.138) under the assumption of the use of any of the preconditioners (6.135). □

For each subdomain, we have introduced discretization, transformed rarefied and transformed coarse imbedded meshes, and, in general, these meshes are nonuniform. However, in order to reduce computational costs, we replace these meshes by the respective uniform orthogonal non-imbedded meshes when we are going to construct the preconditioners. Therefore, without loss of generality, we can consider the first block in $\boldsymbol{\mathcal{Z}}_B = \mathrm{diag}\,[\boldsymbol{\mathcal{Z}}^{\mathrm{E}}, \boldsymbol{\mathcal{Z}}_0]$ as generated on the uniform rectangular mesh topologically equivalent to the discretization mesh. The FDFT, applied edgewise to such $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ defined, *e.g.*, by first expression (6.135), turns the preconditioner into a block-diagonal matrix with 2×2 blocks. Obviously, the FDFT matrix, which is designated $\boldsymbol{\mathcal{F}}_E$, is the block diagonal matrix with identical blocks for the opposite among the edges γ_k, $k = 0, 1, 2, 3$, *i.e.*,

$$\boldsymbol{\mathcal{F}}_E = [\,\boldsymbol{\mathcal{F}}_0,\, \boldsymbol{\mathcal{F}}_0,\, \boldsymbol{\mathcal{F}}_2,\, \boldsymbol{\mathcal{F}}_2\,]\,. \tag{6.144}$$

For problem (6.95), the block diagonal matrix $\mathbf{\Lambda} := \boldsymbol{\mathcal{F}}_E^{\top} \boldsymbol{\mathcal{Y}}^{\mathrm{E}} \boldsymbol{\mathcal{F}}_E$ takes, after an appropriate perturbations, the forms

$$\mathbf{\Lambda} = \mathrm{diag}\left[\, \mathrm{diag}\,[\,\mathbf{\Lambda}_i^{(0)}\,]_{i=1}^{n_2-1},\ \mathrm{diag}\,[\,\mathbf{\Lambda}_k^{(2)}\,]_{k=1}^{n_1-1}\,\right] \quad \text{and}$$
$$\mathbf{\Lambda} = \mathrm{diag}\left[\, \mathrm{diag}\,[\,\mathrm{diag}\,[\mathbf{\Lambda}_{0,i}, \mathbf{\Lambda}_{0,i}\,]]_{i=1}^{n_2-1},\ \mathrm{diag}\,[\,\mathbf{\Lambda}_k^{(2)}\,]_{k=1}^{n_1-1}\,\right], \tag{6.145}$$

which are related to the first and second expressions for $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ in (6.135), respectively. Each block $\mathbf{\Lambda}_i^{(0)}$ couples a pair of opposite nodes $(0, x_{2,i})$ and $(1, x_{2,i})$ on the vertical edges, and each block $\mathbf{\Lambda}_k^{(2)}$ couples a pair of opposite nodes $(x_{1,k}, 0)$ and $(x_{1,k}, \epsilon)$ of the horizontal edges. For the second preconditioner (6.135), the 2×2 matrix $\mathbf{\Lambda}_i^{(0)}$ is diagonal with two equal nonzero entries. Due to the pointed out property, the system with the system matrix $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ can be solved by means of $\mathcal{O}((n_1 + n_2)\log \overline{n})$ a.o. The Schur complement $\mathbf{Y}_E$ and the preconditioners $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ can be calculated in the trigonometric basis via $n_1 \times n_2$ a.o., whereas the matrix-vector multiplication with the matrix $\boldsymbol{\mathcal{Z}}^{\mathrm{E}}$ requires $\mathcal{O}((n_1+n_2)\log \overline{n})$ a.o., see, *e.g.*, [Korneev (2002b)] and [Rytov (2006)].

The costs of some of these operations can still be reduced, provided that some up to date data-sparse matrix techniques like $\mathcal{H}$- or $\mathcal{H}^2$-matrices and tensor-train decompositions, see, *e.g.*, [Khoromskij and Wittum (2004)] [Hackbusch *et al.* (2005)] [Bebendorf (2008)], [Hackbusch (2009)], [Dolgov *et al.* (2011)], [Hackbusch (2012)], and the references therein. For instance, the $\mathcal{H}$-matrix approximation technique provides the cost $\mathcal{O}(n_{\partial\Omega} \log^g n_{\partial\Omega})$ for matrix-vector multiplications by the matrix $\mathbf{Y}^{\mathrm{E}}$, where $n_{\partial\Omega} = 2(n_1+n_2)$.

6.4 Discretizations with Piecewise Variable Orthotropism on Domains Composed of Shape Irregular Rectangles

We now turn to the piecewise orthotropic discretization (6.2)-(6.4), introduced at the beginning of this section, and remind its description. We consider the domain $\Omega = (0,1) \times (0,1)$, the closure of which is the assemblage $\overline{\Omega} = \cup\, \overline{\Omega}_j,$, $j = (j_1, j_2)$, $j_i = 1, \ldots, J_i$, $i = 1, 2$, of closures of rectangular subdomains $\Omega_j = (z_{1,j_1-1}, z_{1,j_1}) \times (z_{2,j_2-1}, z_{2,j_2})$, which are the elements of the rectangular *decomposition grid* $z_{k,0} = 0, \quad x_k = z_{k,j_k}$, $j_k = 0, 1, \ldots, J_k, \quad z_{k,J_k} = 1$, with $H_{k,j_k} = z_{k,j_k} - z_{k,j_k-1} > 0$. The decomposition grid is imbedded in the nonuniform orthogonal denser *discretization grid*

$$x_{k,0} = 0, \quad x_k = x_{k,i_k}, \quad i_k = 0, 1, \ldots, N_k, \quad x_{k,N_k} = 1,$$

i.e., $x_{k,\gamma_k} = z_{k,j_k}$ for some numbers $\gamma_k = \varkappa_k(j_k)$. This grid is also called the *source grid*. For simplicity, we assume that it is uniform on each subdomain Ω_j, and has sizes $h_{k,j_k} = H_{k,j_k}/n_{k,j_k}$, where n_{k,j_k} is the number of the fine mesh intervals on the decomposition grid interval (z_{k,j_k}, z_{k,j_k-1}), see Fig. 6.1 on page 146.

The notations $\mathcal{V}(\Omega)$ and $\mathring{\mathcal{V}}(\Omega)$ stand for the space FE functions and its subspace of FE functions vanishing on $\partial\Omega$, respectively. We mostly assume that these spaces are spaces of FE functions that are continuous on $\overline{\Omega}$ and bilinear on each rectangular element. Therefore, we sometimes also use the more specific notations $\mathcal{V}_\square(\Omega)$ and $\mathring{\mathcal{V}}_\square(\Omega)$.

We again consider the variational problem: given $f \in H^{-1}(\Omega)$, find $u \in \mathring{H}^1(\Omega)$ such that

$$\alpha_\Omega(u, v) = \langle f, v\rangle, \quad \forall\ v \in \mathring{H}^1(\Omega), \tag{6.146}$$

where the bilinear form is given by the expression

$$\alpha_\Omega(u, v) = \int_\Omega \nabla u(x) \cdot \boldsymbol{\rho}(x)\, \nabla v(x)\, dx,$$

with the 2×2 diagonal matrix $\boldsymbol{\rho} = \operatorname{diag}[\rho_1, \rho_2]$, and positive piecewise constant functions $\rho_k(x) = \rho_{k,j}$ for $x \in \Omega_j$. Let $\mathbf{K}$ be the FE stiffness matrix generated by the bilinear form $\alpha_\Omega(\cdot,\cdot)$ on $\mathring{\mathcal{V}}(\Omega)$ and let $\mathbf{S} = \mathbf{K}_B - \mathbf{K}_{B,I}\mathbf{K}_I^{-1}\mathbf{K}_{I,B}$ be its Schur complement.

In this section, we will conclude the construction and analysis of fast and robust solvers for the systems

$$\mathbf{K}\mathbf{u} = \mathbf{f} \quad \text{and} \quad \mathbf{S}\mathbf{u}_B = \mathbf{F}_B\,. \tag{6.147}$$

We can also subdivide each rectangular element of the orthogonal source grid into two right-angled triangles, and arrive at the triangulation, which is termed *source or fine triangulation*. Let us suppose that $\mathcal{V}(\Omega) = \mathcal{V}_\triangle(\Omega)$ and $\mathring{\mathcal{V}}(\Omega) = \mathring{\mathcal{V}}_\triangle(\Omega)$ are the spaces containing continuous, piecewise linear functions defined on the fine triangulation, and (6.147) are the corresponding systems of FE equations. DD preconditioner-solvers, which will be considered here, can be applied to the both types of FE systems (6.147), and have the same computational cost up to some constant. This immediately follows from the fact that the matrices $\mathbf{K} = \mathbf{K}_\square$, $\mathbf{K}_\triangle$ and $\mathbf{S} = \mathbf{S}_\square$, $\mathbf{S}_\triangle$, corresponding to rectangular and triangular elements, are spectrally equivalent, *i.e.,*

$$\frac{1}{3}\,\mathbf{K}_\triangle \le \mathbf{K}_\square \le \mathbf{K}_\triangle \quad \text{and} \quad \frac{1}{3}\,\mathbf{S}_\triangle \le \mathbf{S}_\square \le \mathbf{S}_\triangle, \tag{6.148}$$

independently of numbers $\rho_{k,j}$, $j_k = 1, 2, \ldots, J_k$, $k = 1, 2$, the orthogonal decomposition mesh and the orthogonal piecewise uniform fine mesh. For definiteness, our DD solvers will be derived for systems (6.147) generated by piecewise bilinear FE functions. At the same time, some component preconditioners are based on the discretization by piecewise linear FE functions.

For systems (6.147), algorithms of the same computational complexity can be derived. We start with the Schur complement algorithm. It uses two Schur complement preconditioners $\boldsymbol{\mathcal{S}}_1$ and $\boldsymbol{\mathcal{S}}_1$, relative to which the following properties hold. There are some positive $\underline{\gamma}_1$, $\overline{\gamma}_1$, $\underline{\gamma}_2$ and $\overline{\gamma}_2$ such that the spectral inequalities

$$\underline{\gamma}_1 \boldsymbol{\mathcal{S}}_1 \prec \mathbf{S} \prec \overline{\gamma}_1 \boldsymbol{\mathcal{S}}_1 \quad \text{and} \quad \underline{\gamma}_2 \boldsymbol{\mathcal{S}}_2 \prec \boldsymbol{\mathcal{S}}_1 \prec \overline{\gamma}_2 \boldsymbol{\mathcal{S}}_2 \tag{6.149}$$

are valid. Apart from these generally assumed properties, we expect that the preconditioner $\boldsymbol{\mathcal{S}}_1$ is much cheaper for the matrix-vector multiplications than $\mathbf{S}$, whereas, for $\boldsymbol{\mathcal{S}}_2$, there exists a solver which is much cheaper than presumable solvers for $\mathbf{S}$ and $\boldsymbol{\mathcal{S}}_1$. The last property means that the computational cost for solving SLAE with the system matrix $\boldsymbol{\mathcal{S}}_2$ is much less than the cost of the DD solver. This, in turn, allows us to relax the requirements for $\underline{\gamma}_2$ and $\overline{\gamma}_2$, and, as a consequence, to simplify the construction of the key preconditioner $\boldsymbol{\mathcal{S}}_2$.

The resulting solver for the second system in (6.147) is a two-stage iterative procedure. In this procedure, we solve the system $\mathbf{S}\mathbf{u}_B = \mathbf{F}_B$ via the PCGM with the preconditioner $\boldsymbol{\mathcal{S}}_1$, while systems $\boldsymbol{\mathcal{S}}_1 \mathbf{v}_B = \boldsymbol{\mathcal{F}}_B$ arising at each step of the PCGM iteration are solved inexactly by means of another iterative processes. We can use the Chebyshev iteration process

$$\mathbf{v}^{k+1} = \mathbf{v}^k - \sigma_k \boldsymbol{\mathcal{S}}_2^{-1} (\boldsymbol{\mathcal{S}}_1 \mathbf{v}^k - \boldsymbol{\mathcal{F}}_B), \quad k = 1, 2, \ldots, k_{\mathrm{s}}, \tag{6.150}$$

with the Chebyshev iteration parameters σ_k, the fixed number $k_{\mathrm{s}} = \mathrm{int}\lfloor c_{\mathrm{s}} \sqrt{\overline{\gamma}_2 / \underline{\gamma}_2} \rfloor_+$ of iterations and a positive constant c_{s} that is fixed for all PCGM iterations. This procedure can be viewed as solving the system $\mathbf{S}\mathbf{u}_B = \mathbf{F}_B$ by means of the PCGM with the preconditioner $\boldsymbol{\mathcal{S}}_{1,\,\mathrm{it}}$, the inverse of which is given by the expression

$$\boldsymbol{\mathcal{S}}_{1,\,\mathrm{it}}^{-1} = [\mathbf{I} - \prod_{k=1}^{k_{\mathrm{s}}} (\mathbf{I} - \sigma_k \boldsymbol{\mathcal{S}}_2^{-1} \boldsymbol{\mathcal{S}}_1)] \, \boldsymbol{\mathcal{S}}_1^{-1} \, .$$

Proposition 6.1. *Let us suppose that*

ι) the preconditioners $\boldsymbol{\mathcal{S}}_1$ and $\boldsymbol{\mathcal{S}}_2$ satisfy the spectral inequalities (6.149),

ιι) the matrix-vector multiplications by $\mathbf{S}$ and $\boldsymbol{\mathcal{S}}_1$ cost $\mathcal{N}_{\mathrm{s}}$ and $\mathcal{N}_1$ a.o., respectively, and

ιι) solving the system $\boldsymbol{\mathcal{S}}_2\mathbf{v}_B = \boldsymbol{\mathcal{F}}_B$ *requires* $\mathcal{N}_2$ *a.o., and* $k_{\rm s} = \mathcal{O}((\overline{\gamma}_2/\underline{\gamma}_2)^{1/2})$.

Then the computational cost for solving the system $\mathbf{S}\mathbf{u}_B = \mathbf{F}_B$ *is*

$$\mathcal{O}\left(\sqrt{\overline{\gamma}_1/\underline{\gamma}_1}\,\left[\mathcal{N}_{\rm s} + \sqrt{\overline{\gamma}_2/\underline{\gamma}_2}\,(\mathcal{N}_1+\mathcal{N}_2)\right]\right)$$

arithmetic operations.

Proof. In view of the results presented in Subsection 2.2.3, the proof is straightforward. □

Now we are going to describe the preconditioners $\boldsymbol{\mathcal{S}}_1$ and $\boldsymbol{\mathcal{S}}_2$ in detail. `Preconditioner` $\boldsymbol{\mathcal{S}}_1$. The preconditioner $\boldsymbol{\mathcal{S}}_1$ is not inverted in the DD iterations, and the matrix-vector multiplications by $\boldsymbol{\mathcal{S}}_1$ can be implemented subdomainwise. Apart from that, for each subdomain, transformations from the basis, corresponding to any decompositions of the space $\mathcal{V}(\Omega)$ following from (6.129), to the nodal basis for the discretization mesh and back are cheap. For these reasons, it is sufficient to consider the preconditioner $\boldsymbol{\mathcal{S}}_{1,j}$ for one subdomain, defined on the basis of the space decomposition (6.102) and the subspace preconditioners of Subsections 6.1.3 and 6.3.1. More precisely, in the preconditioner

$$\boldsymbol{\mathcal{G}}_j = \mathrm{diag}\,[\boldsymbol{\mathcal{G}}_{{\rm s},j}, \mathbb{Y}_{{\rm r},j}],$$

which is the preconditioner $\boldsymbol{\mathcal{G}}$ of (6.126) for the domain $\Omega = \Omega_j$, we replace the block $\mathbb{Y}_{{\rm r},j}$ by the matrix $\mathbb{C}_{{\rm r},j}$, which is the preconditioner $\mathbf{C}$ of Lemma 6.2 defined for the rarefied mesh. This results in the subdomain preconditioners

$$\boldsymbol{\mathcal{S}}_{1,j} = \sqrt{\rho_{1,j}\,\rho_{2,j}}\;[\boldsymbol{\mathcal{G}}_{\rm s,j}, \mathbb{C}_{{\rm r},j}], \tag{6.151}$$

which, if written in compatible bases, allow us to assemble the preconditioner $\boldsymbol{\mathcal{S}}_1$. In compliance with Corollary 6.4 and Lemma 6.2, we have

$$\boldsymbol{\mathcal{S}}_1 \prec \mathbf{S} \prec \boldsymbol{\mathcal{S}}_1, \tag{6.152}$$

where we assumed that $\mathbf{S}$ and $\boldsymbol{\mathcal{S}}_1$ are written in a common basis.

`Preconditioner` $\boldsymbol{\mathcal{S}}_2$. For each subdomain Ω_j, we use the preconditioner $\boldsymbol{\mathcal{S}}_{2,j} = \boldsymbol{\mathcal{Z}}_{B,j}$, where

$$\boldsymbol{\mathcal{Z}}_{B,j} = \mathrm{diag}\,[\boldsymbol{\mathcal{Z}}_j^{\rm E}, \boldsymbol{\mathcal{Z}}_{0,j}], \quad \text{with } \boldsymbol{\mathcal{Z}}_{0,j} = \sqrt{\rho_{1,j}\rho_{2,j}}\,\mathbf{B}_{0,j}. \tag{6.153}$$

Here $\boldsymbol{\mathcal{Z}}_{B,j}$ stands for the preconditioner $\boldsymbol{\mathcal{Z}}_B$ from Theorem 6.5 for the rectangle $\Omega = \Omega_j$. Let $\mathbf{K}_V$ be that block of the FE matrix $\mathbf{K}$ that corresponds

to the nodes of the decomposition mesh and is generated by the subspace of functions which are continuous on Ω and bilinear on each subdomain Ω_j. Then $\boldsymbol{\mathcal{S}}_2$ is assembled from the subdomain preconditioners $\boldsymbol{\mathcal{S}}_{2,j}$, and has the block-diagonal form $\boldsymbol{\mathcal{S}}_2 = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \mathbf{K}_V]$ with $\boldsymbol{\mathcal{S}}_E$ assembled from the matrices $\boldsymbol{\mathcal{Z}}_j^{\mathrm{E}}$. According to Theorem 6.5, we have

$$(\min_j \underline{\mu}_j)\ \boldsymbol{\mathcal{S}}_2 \prec \mathbf{S} \prec \boldsymbol{\mathcal{S}}_2 \tag{6.154}$$

with

$$\underline{\mu}_j = \frac{1}{(1+\log \underline{m}_j)}\ \min\left(\frac{1}{\theta_j}, \frac{1}{1+\log \underline{m}_j}\right)$$

representing nothing but $\underline{\mu}$ from (6.133) for $\Omega = \Omega_j$.

Taking into account the representation of the FE stiffness matrix in the block form

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{IB} \\ \mathbf{K}_{BI} & \mathbf{K}_B \end{pmatrix}, \tag{6.155}$$

we define the inverse to the DD preconditioner $\boldsymbol{\mathcal{K}}_{\mathrm{DD}}$ by the expression

$$\boldsymbol{\mathcal{K}}_{\mathrm{DD}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \mathbb{P}\boldsymbol{\mathcal{S}}_{1,\,\mathrm{it}}^{-1}\mathbb{P}^{\top}. \tag{6.156}$$

Here the matrix $\boldsymbol{\mathcal{K}}_I$ is related to the d.o.f. which are internal for each subdomain, and, like the block $\mathbf{K}_I = \mathrm{diag}\,[\,\mathbf{K}_{I,j}\,]_{j_1,j_2=1}^{J_1,J_2}$, has the block-diagonal structure $\boldsymbol{\mathcal{K}}_I = \mathrm{diag}\,[\,\boldsymbol{\mathcal{K}}_{I,j}\,]_{j_1,j_2=1}^{J_1,J_2}$. Furthermore, $\mathbb{P} = (\mathbb{P}_I^{\top}, \mathbf{I})^{\top}$ is the prolongation matrix. We assume that

$$\boldsymbol{\mathcal{K}}_{I,j} \prec \mathbf{K}_{I,j} \prec \boldsymbol{\mathcal{K}}_{I,j} \quad \text{and} \quad \|\,\mathbb{P}_{B_j}\mathbf{v}_{B_j}\,\|_{\mathbf{K}_j} \prec \|\,\mathbf{v}_{B_j}\,\|_{\mathbf{S}_j}, \tag{6.157}$$

where $\mathbb{P}_{B_j}$ denotes the restriction of the prolongation operator $\mathbb{P}$ to the subdomain $\overline{\Omega}_j$. In addition to this, we asume

$$\mathrm{ops}\,[\boldsymbol{\mathcal{K}}_I^{-1}\mathbf{f}_I] = N_\Omega \quad \text{and} \quad \mathrm{ops}\,[\mathbb{P}\mathbf{v}_B] \prec N_\Omega\,. \tag{6.158}$$

There are a variety of such preconditioner-solvers and prolongation operators in the literature. We here only refer to [Nepomnyaschikh (1991c)], [Griebel and Oswald (1995b)], [Haase and Nepomnyaschikh (1997)], and [Oswald (1999a)] for some examples.

Theorem 6.6. *Let $H_{k,j_k} \in (0,1)$, $n_{k,j_k} \geq 1$, $\rho_{k,j} > 0$ be arbitrary in the pointed out ranges. Then the total arithmetical costs $Q_{\mathbf{K}}$ and $Q_{\mathbf{S}}$ of the DD and Schur complement algorithms, respectively, satisfy the bounds*

$$Q_{\mathbf{K}} \prec N_1 N_2 + [N_\Gamma(1+\log \overline{N}) + \Upsilon(J_1 J_2)]\sqrt{\overline{N}}(1+\log \overline{N})^{1/2},$$

$$Q_{\mathbf{S}} \prec N_\Gamma(1+\log N_\Gamma) + [N_\Gamma(1+\log \overline{N}) + \Upsilon(J_1 J_2)]\sqrt{\overline{N}}(1+\log \overline{N})^{1/2}, \tag{6.159}$$

where $\overline{N} = \max_k N_k$, $N_\Gamma = J_1 N_2 + J_2 N_1$, and $\Upsilon(J_1, J_2)$ stands for the cost of the solution of the subsystem with the system matrix $\mathbf{K}_V$.

Proof. First we list the main factors contributing to the complexity of the Schur complement algorithm, and, afterwards, add a brief commentary. These factors are:

- the number of external PCGM iterations $k_{\,\mathrm{PCG}} = \mathrm{const}$,
- the cost $\mathcal{N}_S \prec N_\Gamma \log N_\Gamma$ a.o. of one matrix-vector multiplications by Schur complement matrix $\mathbf{S}$,
- the number of secondary iterations (6.150):

$$k_s \prec 1/\sqrt{\underline{\mu}} \prec \sqrt{\max_j \overline{n}_j(1+\log \underline{n}_j)} \prec \sqrt{\overline{N}(1+\log \overline{N})},$$

  with $\underline{\mu} = \min \underline{\mu}_j$,
- the cost of the matrix-vector multiplication by $\boldsymbol{\mathcal{S}}_1$ at each secondary iteration step: $\mathcal{N}_{\mathcal{S}_1} \prec N_\Gamma(1+\log \overline{N})$,
- the cost of solving the system with the preconditioner $\boldsymbol{\mathcal{S}}_2$ at each secondary iteration step: $\mathcal{N}_{\mathcal{S}_2} \prec N_\Gamma(1+\log \overline{N}) + \Upsilon(J_1, J_2)$.

Taking into account (6.152) and (6.154), we conclude that

$$\underline{\mu}\,\boldsymbol{\mathcal{S}}_2 \prec \boldsymbol{\mathcal{S}}_1 \prec \boldsymbol{\mathcal{S}}_2, \tag{6.160}$$

and, therefore, the given number k_s of secondary iterations provides the spectral equivalence

$$\boldsymbol{\mathcal{S}}_{1,\,\mathrm{it}} \asymp \boldsymbol{\mathcal{S}}_1\,. \tag{6.161}$$

Now, the last relationship and (6.152) guarantee that $k_{\,\mathrm{PCG}} \log \varepsilon^{-1}$ PCGM iterations provide the relative error in the norm $||\cdot||_{\mathbf{S}}$ bounded by the prescribed $\varepsilon > 0$. The first term in the expression for $\mathcal{N}_{\mathcal{S}_2}$ bounds the arithmetical cost for solving the system with the system matrix $\boldsymbol{\mathcal{S}}_E$. Using the above bounds according to Proposition 6.1, we arrive at the bound (6.159) for $Q_{\mathbf{S}}$.

The DD peconditioner $\boldsymbol{\mathcal{K}}_{\mathrm{DD}}$ is spectrally equivalent to the FE matrix $\mathbf{K}$, what immediately follows from Lemma 6.2, Corollary 6.3, Corollary 6.4 (6.157), (6.161) and Corollary 4.1. The estimate of the DD solver cost is obtained by taking additionally into account the above costs of operations, related to the interface, and (6.158). □

Suppose that $N_1 = N_2 = N$, $J_1 = J_2 = J$, and the decomposition mesh is fixed. Then

$$Q_{\mathbf{K}} \prec N^2\,. \tag{6.162}$$

If the number of subdomains grows with the growth of the numbers N_k of the source mesh lines, the contribution $\Upsilon(J_1 J_2)$ of the solver for the vertex

subproblem can compromise this bound. Assuming that a direct solver for systems with the system matrix $\mathbf{K}_V$ is sufficiently fast, we can retain the bound (6.162) under the condition $J_k \leq N^{1/2}/(1+\log N)^{3/2}$.

Each system $\mathbf{K}_{I,j}\mathbf{v}_{I,j} = \mathbf{F}_{I,j}$ can efficiently be solved by means of the FDFT with $\mathcal{O}(n_{1,j_1} n_{2,j_2} \log \overline{n}_j)$ a.o., $\boldsymbol{\mathcal{K}}_{I,j} = \mathbf{K}_{I,j}$, $\mathbb{P} = (-\mathbf{K}_{BI}\mathbf{K}_I^{-1}, \mathbf{I})^\top$. Thus, we can use the stting $\boldsymbol{\mathcal{K}}_{I,j} = \mathbf{K}_{I,j}$ and $\mathbb{P} = (-\mathbf{K}_{BI}\mathbf{K}_I^{-1}, \mathbf{I})^\top$, which only leads to a slight loss in efficiency.

Clearly, the numerical complexity of the same DD preconditioner-solver $\boldsymbol{\mathcal{K}}_{\mathrm{DD}}$ for the problem (6.4) with the bilinear form replaced by (6.7) and (6.8) will differ only by a constant only depending on μ_1 and μ_2. However, the implementation of the DD Schur complement solver can differ noticeably, since it requires calculation of $\mathbf{S}$ and multiplications by $\mathbf{S}$. These operations obviously become more expensive. However, using the $\mathcal{H}$-matrix technique for the calculation of $\mathbf{S}$, we can reduce the complexity to $\mathcal{O}(N_\Omega \log N_\Omega)$, with $N_\Omega = N_1 N_2$.

At the same time, the complexity of the same operation and of the matrix-vector multiplication by $\mathbf{S}$ is $\mathcal{O}(N_\Gamma \log N_\Gamma)$, provided that $\wp = \boldsymbol{\rho}$, *cf., e.g.,* [Hackbusch (2003)] and [Hackbusch *et al.* (2005)].

It is also worth emphasizing that the above estimates are derived at quite general assumptions. Indeed, we practically made no restrictions on H_{k,j_k}, n_{k,j_k}, and $\rho_{k,j} > 0$, and their changes from subdomain to subdomain! Moreover, if the variation of these values can be characterized by some functions, it in turn could help to improve the bounds.

Chapter 7

Nonoverlapping DD Methods for *hp* Discretizations of 2d Elliptic Equations

The so-called *hp* methods have been gaining much attention for the reason of their ability to attain much faster convergence than *h* methods for both smooth problems and problems with singularities. However, this noticeable advantage is often compromised by a high cost of the setup procedure which is caused by a high fill-in of stiffness matrices and by complex algorithms for their calculation. The complexity of the setup is only one part of the problem of implementation. The factorization procedures can be rather expensive for stiffness matrices of *hp* methods as well. For large scale problems, the use of direct methods is especially uneconomical and accompanied by large memory and time requirements. Nevertheless, despite all these negative factors, the computational cost of solvers grows with the number of unknowns, at most, algebraically, whereas the rate of convergence can be exponential. Therefore, there is a strong incentive for attempts to improve the performance of solution algorithms for *hp* discretizations. At this moment, the existing toolkit of fast solvers for the systems arising from *hp* discretizations is still much less developed than for *h* discretizations. At the same time, one can say that the principal difficulties in creation of such a *hp*-toolkit has been passed, at least, for discretizations by means of square and cubic reference elements. A key fact underlying this achievement was establishment of an one-to-one correspondence between major classes of *hp* discretizations and their *h* counterparts, the matrices of which are spectrally equivalent. This opened the road for the adaptation of efficient *h*-version solvers to the corresponding *hp*-versions.

The progress in the area of fast solvers for *hp* discretizations has been made in the last three decades within the framework of DD approach. It is worth noting again that the basic features of *hp* DD algorithms for elliptic problems in respect of their structure and tools of their analysis basically

originated from those developed for the h versions. At the same time, the components of algorithms as well as their justification for h and hp versions differ significantly. This concerns the preconditioner-solvers for local Dirichlet subproblems on the subdomains of decomposition as well as the interface Schur complement subproblem and prolongation procedures. The pioneering analysis of the basic structure of Dirichlet-Dirichlet DD algorithms for the p-version in 2d was made by [Babuška *et al.* (1991)], who developed also important technical tools. For further advances in the development and analysis of DD algorithms for hp discretizations and their components, we refer to [Oden *et al.* (1994, 1997)], [Pavarino (1994)], [Ivanov and Korneev (1995, 1996)], [Ainsworth (1996)], [Pavarino and Widlund (1996)], [Widlund (1997)], [Ainsworth and Senior (1997)], [Casarin (1997)], [Korneev and Jensen (1997, 1999)], [Korneev (2001, 2002b)], [Korneev *et al.* (2002b,a)], [Beuchler (2002)], [Korneev *et al.* (2003b,a)], [Beuchler *et al.* (2004)] and [Kim and Kim (2009)]. This list of references is far from being complete. Many other authors should have been mentioned in connection with the development of the fast DD hp solvers, especially those belonging to the French school. The works by [Bernardi and Maday (1992a, 1997)] on stability of polynomial interpolations in Sobolev's spaces play a crucial role in the analysis of the preconditioners for stiffness and mass matrices of the reference p-elements. The same holds for the works by [Bernardi *et al.* (1993)] and [Belgacem (1994)] on the prolongation (lifting) operators in the corresponding polynomial spaces. In relation to prolongation operators in the polynomial finite element spaces, the contributions by [Munoz-Sola (1997)] and [Demkowicz *et al.* (2008, 2009, 2012)] as well as the work on cheap prolongations by [Beuchler and Schöberl (2006)] should also be mentioned.

The difference between DD algorithms for the h and the hp discretizations starts from the beginning, when one looks at solvers for the local Dirichlet problems on subdomains of decomposition. They usually represent the most time-consuming component of DD solver contributing the main term in its asymptotic computational cost. For obtaining these preconditioner-solvers, an additional intermediate step was to be done in distinction with the h-version. It was concerned with the preconditioning of hp discretizations by simpler finite-difference type preconditioners, which, in turn, was reduced to the preconditioning of reference elements stiffness matrices. Such preconditioners turned out to be quite different for different types of the reference p-elements.

Much attention in the literature was paid to the reference elements of the Lagrange interpolation type with GLL and GLC nodes on the reference

cube $\tau_0 = (-1,1)^3$. By these families of the nodes are understood the nodes generated by the tensor products of 1d grids of the Gauss-Lobatto-Legendre and Gauss-Lobatto-Chebyshev quadratures. Such reference elements are commonly called *spectral*. Not less popular and having several meaningful advantages are the reference elements, defined on $\tau_0 = (-1,1)^3$, with the shape functions produced by the tensor products of the integrated Legendre's polynomials. Such reference elements are termed *hierarchical* due to the hierarchical type of their shape functions. The use of FD preconditioners on the same set of nodes for the stiffness matrices of cubic spectral reference elements was introduced by [Orszag (1980)]. In 1d, spectral equivalence of these preconditioners to the stiffness and mass matrices of the reference elements with GLL nodes was proved by [Bernardi and Maday (1992a,b)], whereas the straightforward step to a greater dimension may be found in [Canuto (1994)] and [Casarin (1997)].

For hierarchical reference elements, FD-like preconditioners were introduced by [Ivanov and Korneev (1996)] and further studied by [Korneev and Jensen (1997, 1999)]. It is worth noting that Orszag's preconditioners are introduced for Lagrange (nodal) finite elements in a natural way. That means that they are FD operators defined on the nodal unknowns of FE functions and approximate the partial differential operator on the same set of nodes of the reference element. In contrast to them, the FD preconditioners for hierarchical reference elements are constructed artificially. They are defined on the FE unknowns which are not nodal values of FE functions and artificially related to the nodes of a uniform square mesh. The FD operator is an approximation of an elliptic operator, which is quite different from the differential operator corresponding to the reference p-element stiffness matrix. For instance, if the latter is simply the Laplace differential operator, then the former is a differential operator with varying orthotropism, specified by the coefficients deteriorating on the part of the boundary.

Some time after simply looking FD preconditioners were suggested, fast local preconditioner-solvers have been contrived in the way of designing fast solvers for these FD-type preconditioners. Fast solvers of different types for *hp* discretizations of the local Dirichlet problems have appeared relatively recently in the papers by [Korneev (2002b)], [Korneev *et al.* (2002c)], [Beuchler (2002)], [Beuchler *et al.* (2004)] and [Korneev and Rytov (2005a,b)]. For comparison, we note that, *e.g.,* the pioneering results of [Fedorenko (1961, 1964)] and [Bakhvalov (1966)] on the multigrid methods for the h-version appeared about 40 years earlier. Fast preconditioner-solvers

for the internal stiffness matrices of 2d and 3d finite p-elements, presented in Chapters 8 and 9, were primarily developed in the works cited above.

The condensation of the internal unknowns for finite elements is a natural and efficient strategy for solving systems of algebraic equations, resulting from discretizations, by direct methods. If we implement such a substructuring technique consistently in the spirit of DD approach, we come to DD Dirichlet–Dirichlet algorithms with the domains of finite elements taken for subdomains of decomposition. Apart from efficiency, such algorithms provide a high level of parallelization. For this reason, when considering nonoverlapping DD algorithms for *hp* discretizations, we do mostly not distinguish between subdomains of decomposition and domains of finite elements.

In this fashion, we discuss *hp* adaptive orthotropic discretizations with powers of polynomials, changing from element to element differently in different directions. Indeed, some h adaptivity is not prohibited from the start, due to our general assumption that the FE grid is only shape regular. When finite elements of an assemblage are associated only with a few reference elements, probably of one type, such versions of DD algorithms possess a definite advantage. In this case, independently of the number of finite elements, a fast DD solver may need solvers for one "global" subsystem and for subsystems governed only by a fixed number of standard matrices induced by these reference elements. For instance, if 2d finite elements are associated with a single p reference element, a fast DD algorithm can be arranged in such away that it involves preconditioner-solvers only for three matrices, see Subsection 7.1.1. One is the stiffness matrix of the reference element, another is the global matrix related to the unknowns associated with the vertices, and the third one is the Schur complement matrix acting on d.o.f. of the typical edge of the reference element.

7.1 Structure of DD Preconditioners and its Reflection in the Relative Condition Number

7.1.1 *Main Components of DD Preconditioners*

For *hp* discretizations, as for h discretizations, it is natural to distinguish *internal, interface, edge,* and *vertex* degrees of freedom, and, respectively, decompose the FE vector and function spaces. Although these FE function spaces are different from those in the h-version, we use notations which are similar to those introduced in Subsection 4.2.1, and come to the same, as

in (4.29), block representations of the stiffness matrix

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,B} \\ \mathbf{K}_{B,I} & \mathbf{K}_B \end{pmatrix} = \begin{pmatrix} \mathbf{K}_\Sigma & \mathbf{K}_{\Sigma,V} \\ \mathbf{K}_{V,\Sigma} & \mathbf{K}_V \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,E} & \mathbf{K}_{I,V} \\ \mathbf{K}_{E,I} & \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,I} & \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix} \tag{7.1}$$

with

$$\mathbf{K}_B = \begin{pmatrix} \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix} \quad \text{and} \quad \mathbf{K}_\Sigma = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,E} \\ \mathbf{K}_{E,I} & \mathbf{K}_E \end{pmatrix}.$$

The sub-matrix related to the internal element d.o.f. is again block-diagonal

$$\mathbf{K}_I = \operatorname{diag}\left[\mathbf{K}_{I_r}\right]_{r=1}^{\mathcal{R}},$$

but now each block $\mathbf{K}_{I_r}$ corresponds to the internal unknowns of one finite element τ_r. The notation E_k will be used for an edge of a finite element. In general, Dirichlet-Dirichlet DD preconditioners may also be represented by the formulas (4.30) and (4.32). Moreover, Theorem 4.1 retains in the same form, but adjusted to the assumed domain decomposition, *i.e.,* with $\Omega_j = \tau_j$. If the vertex subproblem in the DD preconditioner is split from the others, as, *e.g.*, in (4.30), the construction of an efficient vertex preconditioner-solver turns out to be the same as for the h-version with the split coarse subspace component. Therefore, for getting an efficient DD preconditioner-solver of type (4.30), we have to find good preconditioner-solvers $\boldsymbol{\mathcal{K}}_{I_r}$ for the local Dirichlet problems on finite elements and $\boldsymbol{\mathcal{S}}_{E_q}$ for the Schur complement living on the edges. In the case of DD preconditioners of the type (4.32), instead of $\boldsymbol{\mathcal{S}}_{E_q}$, efficient Schur complement preconditioner-solvers $\boldsymbol{\mathcal{S}}_{B_r}$ for the finite element boundaries are needed. The use of cheap and efficient prolongation operators $\boldsymbol{\mathcal{P}}_{B_r}$ from the boundaries of finite elements on their entire domains is not less important.

The derivation of the listed components of DD preconditioner-solvers for *hp*-version of the finite element method is more involved than their counterparts in the h-version. They directly depend on the specific type of reference elements used for discretization, the choice of which deeply influences the final algorithms. At the same time, there is a factor simplifying DD algorithms where the subdomains of the decomposition are the domains of the finite elements. For instance, if the finite elements are associated with one single reference element and the shape regularity conditions are fulfilled, then, for all $r = 1, 2, \ldots, \mathcal{R}$, the matrices $\boldsymbol{\mathcal{K}}_{I_r}$, $\boldsymbol{\mathcal{S}}_{E_q}$, $\boldsymbol{\mathcal{S}}_{B_r}$ and $\boldsymbol{\mathcal{P}}_{B_r}$ are defined by four matrices, generated by the reference element, up to obvious perturbations.

For brevity, the vertex coordinate functions will be termed *hierarchical*, if they are nodal (Lagrange) linear and bilinear functions on triangular and square reference elements, respectively. For spaces on a reference element, we use notations $U, U_I, \ldots, U_V$ and $\mathcal{U}, \mathcal{U}_I, \ldots, \mathcal{U}_V$ which are counterparts of vector spaces $V, V_I, \ldots, V_V$ and FE functions spaces $\mathcal{V}, \mathcal{V}_I, \ldots, \mathcal{V}_V$, respectively.

In general, as noted above, DD preconditioners can be represented in the form

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \boldsymbol{\mathcal{P}}_B \boldsymbol{\mathcal{S}}_B^{-1} \boldsymbol{\mathcal{P}}_B^T \,, \tag{7.2}$$

that is similar to (4.32), and each of the matrices can be defined by means of the reference element. Let $\mathbf{A}$ be the stiffness matrix of the reference element, which is the preimage of the finite elements of the *hp* discretization. For example, in the case of the model problem (2.5) – (2.6), the matrix $\mathbf{A}$ can be defined by the identity

$$\mathbf{v}^\top \mathbf{A} \mathbf{w} \equiv a_{\tau_0}(v, w) \,, \quad \forall\, \mathbf{v}, \mathbf{w} \leftrightarrow v, w \in \mathcal{U} \,, \tag{7.3}$$

where

$$a_{\tau_0}(v, w) \equiv \int_{\tau_0} \nabla v \cdot \nabla w \, dx \,, \quad \forall\, v, w \in \mathcal{U} \,. \tag{7.4}$$

In the factorization

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_I & \mathbf{A}_{I,B} \\ \mathbf{A}_{B,I} & \mathbf{A}_B \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{B,I}\mathbf{A}_I^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_B \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_I^{-1}\mathbf{A}_{I,B} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} ,$$

of the matrix $\mathbf{A}$, the block-matrix $\mathbb{S}_B : U_B \to U_B$ is the Schur complement

$$\mathbb{S}_B = \begin{pmatrix} \mathbb{S}_E & \mathbb{S}_{EV} \\ \mathbb{S}_{VE} & \mathbb{S}_V \end{pmatrix} = \mathbf{A}_B - \mathbf{A}_{B,I}\mathbf{A}_I^{-1}\mathbf{A}_{I,B} \,,$$

acting in the space of d.o.f. living on the boundary of the reference element.

Let us suppose that we have the two preconditioners $\boldsymbol{\mathcal{A}}_I$ and $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ satisfying the spectral equivalence inequalities

$$\underline{\beta}_I \boldsymbol{\mathcal{A}}_I \le \mathbf{A}_I \le \overline{\beta}_I \boldsymbol{\mathcal{A}}_I \quad \text{and} \quad \underline{\beta}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \le \mathbb{S}_B \le \overline{\beta}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \,, \tag{7.5}$$

with positive numbers $\underline{\beta}_I, \overline{\beta}_I, \underline{\beta}_B$ and $\overline{\beta}_B$, and a prolongation operator $\mathcal{P}_{B,\mathrm{ref}} : \mathcal{U}_B \to \mathcal{U}$ satisfying

$$|\mathcal{P}_{B,\mathrm{ref}}\, v_B|^{\,2}_{1,\tau_0} \le c_{0,\mathcal{P}}\, |v_B|^{\,2}_{1/2,\partial\tau_0} \,, \quad \forall\ v_B \in \mathcal{U}_B(\partial\tau_0) \,, \tag{7.6}$$

with a positive number $c_{0,\mathcal{P}}$. The notation $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}$ will be used for the matrix of the prolongation operator $\mathcal{P}_{B,\mathrm{ref}}$. A characterization of the quality

of the prolongation, that is more suited for DD algorithms, is given by the inequality

$$\|\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}\,\mathbf{v}_B\|^{\,2}_{\mathbf{A}} \le c_{0,\mathcal{P}}\,\|\mathbf{v}_B\|^{\,2}_{\mathbb{S}_B}\,, \quad \forall\ \mathbf{v}_B \in U_B\,, \tag{7.7}$$

which can also be used in the situations when the seminorms $|v_B|_{\,1/2,\partial\tau_0}$ and $\|\mathbf{v}_B\|_{\,\mathbb{S}_B}$, $\forall\,\mathbf{v}_B \leftrightarrow v$, are not equivalent. Obviously, (7.7) follows from (7.6) by the trace theorem for the Sobolev space $H^1(\tau_0)$, whereas the opposite is not true.

If the conditions of shape regularity are fulfilled, then the matrices entering (7.2) can be defined as follows:

α) *Preconditioner for the local Dirichlet problems.* We set

$$\boldsymbol{\mathcal{K}}_I = \mathrm{diag}\,[\varrho_r \boldsymbol{\mathcal{A}}_I]_{r=1}^{\mathcal{R}}\,. \tag{7.8}$$

β) *Schur complement preconditioner.* The Schur complement preconditioner can be defined by assembling of the matrices $\varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$:

$$\boldsymbol{\mathcal{S}}_B = \sum_{r=1}^{\mathcal{R}} \mathbb{T}_{B,B_r}\varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}\mathbb{T}^{\top}_{B,B_r} = \biguplus_{r=1}^{\mathcal{R}} \varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}\,. \tag{7.9}$$

γ) *Prolongation matrix.* The prolongation matrix $\boldsymbol{\mathcal{P}}_B$ is uniquely defined by the condition that its restriction to an element τ_r is

$$\boldsymbol{\mathcal{P}}_{B,r} := \boldsymbol{\mathcal{P}}_B\big|_{\tau_r} = \boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}\,, \tag{7.10}$$

and $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}$ satisfies (7.6) or (7.7).

We remind that in relations, like the three ones above, we always assume appropriate orderings of d.o.f., in particular, of the reference element, which may change within one relationship, as, *e.g.*, in (7.8).

Lemma 7.1. *Let inequalities (7.5) and (7.7) hold, and let the matrices* $\boldsymbol{\mathcal{K}}_I, \boldsymbol{\mathcal{S}}_B$ *and* $\boldsymbol{\mathcal{P}}_B$ *be defined according to* ***α***) – ***γ***). *Then the spectral equivalence inequalities*

$$\underline{\gamma}\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{\gamma}\boldsymbol{\mathcal{K}}\,, \tag{7.11}$$

are valid with the values $\underline{\gamma}$ *and* $\overline{\gamma}$ *satisfying the inequalities*

$$\begin{aligned} &\underline{\gamma} \ge \underline{c}\left(1 - \sqrt{1 - c_{0,\mathcal{P}}^{-1}}\right)\min(\underline{\beta}_I, \underline{\beta}_B) > 0 \quad \textit{and} \\ &\overline{\gamma} \le \overline{c}c_{0,\mathcal{P}}\underline{c}\left(1 + \sqrt{1 - c_{0,\mathcal{P}}^{-1}}\right)\max(\overline{\beta}_I, \overline{\beta}_B)\,, \end{aligned} \tag{7.12}$$

where the constants $\underline{c}$ *and* $\overline{c}$ *depend only on the generalized conditions of shape regularity.*

Proof. The assumptions made in the lemma immediately yield the inequalities

$$\underline{\beta}_I \boldsymbol{\mathcal{K}} \le \mathbf{K}_I \le \overline{\beta} \boldsymbol{\mathcal{K}}_I \,, \quad \underline{\beta}_B \boldsymbol{\mathcal{S}}_B \le \mathbf{S}_B \le \overline{\beta}_B \boldsymbol{\mathcal{S}}_B \,, \tag{7.13}$$
$$\| \boldsymbol{\mathcal{P}}_B \, \mathbf{v}_B \|^2_{\mathbf{K}} \le c_{0,\mathcal{P}} \, \| \mathbf{v}_B \|^2_{\mathbf{S}_B} \,, \quad \forall \ \mathbf{v}_B \in V_B \,,$$

hold up to the additional omitted constants depending only on the generalized conditions of the shape regularity. The bounds (7.12) are simply a consequence of Theorem 4.2, see also Corollary 4.1, and (7.13). □

7.1.2 *Preconditioners of Nepomnyaschikh's and BPS Types*

Apart from preconditioner-solvers for the internal Dirichlet problems on domains of finite elements, interface Schur complement preconditioner-solvers is another component of a crucial importance. In what follows, we will present several sufficiently simple and efficient Schur complement preconditioners $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ for p reference elements. Simplest preconditioners in a sense of solution procedures and their parallelization are those, in which vertices are not coupled with the edges and each edge is not coupled with others. They do not provide optimal relative condition numbers, but the losses in the condition are insignificant. The arrangement of such preconditioners follows BPS type preconditioners for h discretizations of 2d elliptic equations introduced by [Bramble *et al.* (1986, 1987, 1988, 1989)]. The pioneering and essential step in the approving this type preconditioner for hp discretizations was made by Babuška *et al.* [Babuška *et al.* (1991)].

An alternative way is to use a sufficiently simple preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$, in which vertices and edges are not split and which possess at least two properties. First one, is the spectral equivalence of $\boldsymbol{\mathcal{S}}_B$ and $\mathbf{S}_B$, meaning that the inequalities (7.5) for the matrices $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ and $\mathbb{S}_B$ hold with $\underline{\beta}_B, \overline{\beta}_B$ being constants independent of h and p. Second one assumes cheap matrix-vector multiplications by $\boldsymbol{\mathcal{S}}_B$. At that, such a global Schur complement preconditioner and its structure, as defined by (7.9), in general can be more complicated than for interface Schur complement preconditioner for h-version. For this reason, solution of the system of algebraic equations with $\boldsymbol{\mathcal{S}}_B$ for the matrix by direct methods is costly and can easily compromise the optimality of the DD solver. Therefore, some remedies should be implemented to improve the interface component. Since the preconditioner $\boldsymbol{\mathcal{S}}_B$ (7.9) is cheap for multiplications, the use of an inexact solver can resolve the problem in a way similar to what was presented earlier for h-version,

see (4.32), (4.27). One needs another Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{solv}}$, which provides, probably, worse than $\underline{\beta}_B, \overline{\beta}_B$ values of $\underline{\varkappa}_B, \overline{\varkappa}_B$ in the inequalities

$$\underline{\varkappa}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{solv}} \le \mathbb{S}_B \le \overline{\varkappa}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{solv}}\,, \qquad \underline{\varkappa}_B > 0\,, \tag{7.14}$$

but is cheap for solving systems with it for the matrix. Then the inexact solver implicitly defines the new interface Schur complement preconditioner

$$\begin{aligned}\boldsymbol{\mathcal{S}}^{-1}_{\mathrm{it},B} &= \boldsymbol{\mathcal{I}}_{\circ}[\boldsymbol{\mathcal{S}}_B, \boldsymbol{\mathcal{S}}_{B,solv}] = \boldsymbol{\mathcal{I}}\,[\boldsymbol{\mathcal{S}}_B, \boldsymbol{\mathcal{S}}_{B,solv}, n_{1/2}] \\ &= [\mathbf{I} - \prod_{k=1}^{n_{1/2}} (\mathbf{I} - \sigma_k \boldsymbol{\mathcal{S}}^{-1}_{B,\mathrm{solv}} \boldsymbol{\mathcal{S}}_B)]\, \boldsymbol{\mathcal{S}}^{-1}_B\,, \end{aligned} \tag{7.15}$$

and the new DD preconditioner

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}^{+}_I + \boldsymbol{\mathcal{P}}_B \boldsymbol{\mathcal{S}}^{-1}_{\mathrm{it},B} \boldsymbol{\mathcal{P}}^T_B\,. \tag{7.16}$$

In general, this way does not end with a solver of linear complexity for the interface Schur complement problem, but the optimality of DD solver can be guaranteed.

Even the simplest choice of the diagonal matrix $\boldsymbol{\mathcal{S}}_{B,\mathrm{solv}} = \boldsymbol{\mathcal{D}}_B$ with the main diagonal taken from $\boldsymbol{\mathcal{S}}_B$, may lead to an optimal DD preconditioner-solver $\boldsymbol{\mathcal{K}}$, *i.e.,* with $\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \mathcal{O}(1)$ and $\mathrm{ops}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{f}] = \mathcal{O}(N)$ for any $\mathbf{f}$, where N is the number of unknowns. In Subsection 7.5.4, we address this type of preconditioners for *hp* discretizations of two families, which are generated by means of the square and triangular reference p elements with the positions of the nodes on their edges defined according to the Gauss-Lobatto-Chebyshev and Gauss-Lobatto-Legendre quadratures rules. The results of this section reflect the papers [Korneev and Jensen (1999)] and [Korneev *et al.* (2002b,a)] on *hp* discretizations with finite elements associated with the square and triangular reference elements, respectively. For the *h*-version of the FEM, Schur complement preconditioners of this type were introduced by [Nepomnyaschikh (1991a)].

The BPS type preconditioners assume the following basic properties:

i) Finite elements of the assemblage are associated with the square/triangular reference elements, whose vertex coordinate functions are nodal bilinear/linear polynomials, corresponding lowest order reference elements. If initially these coordinate functions are different, as, *e.g.*, for the spectral reference elements, they need to be transformed to the pointed out ones.

ii) Vertex d.o.f. are split from the remaining d.o.f. such that

$$\boldsymbol{\mathcal{K}} = \mathrm{diag}\,[\boldsymbol{\mathcal{K}}_\Sigma, \boldsymbol{\mathcal{K}}_V]\,, \quad \text{with } \boldsymbol{\mathcal{K}}^{-1}_\Sigma = \boldsymbol{\mathcal{K}}^{+}_I + \boldsymbol{\mathcal{P}}_{\Sigma\leftarrow E} \boldsymbol{\mathcal{S}}^{-1}_E \boldsymbol{\mathcal{P}}^{\top}_{\Sigma\leftarrow E}\,. \tag{7.17}$$

iii) In the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_E$, d.o.f. of each edge are split from d.o.f. of other edges, and, therefore, this preconditioner is also block diagonal

$$\boldsymbol{\mathcal{S}}_E = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{E_q}]_{q=1}^{Q}\,, \tag{7.18}$$

where Q is the number of edges $E_k \subset \Omega$.

Let

$$\mathbb{S}_E = \mathbf{A}_E - \mathbf{A}_{E,I}\mathbf{A}_I^{-1}\mathbf{A}_{I,E}$$

be the Schur complement arising in the factorization

$$\mathbf{A}_\Sigma = \begin{pmatrix} \mathbf{A}_I & \mathbf{A}_{I,E} \\ \mathbf{A}_{E,I} & \mathbf{A}_E \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{E,I}\mathbf{A}_I^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_E \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_I^{-1}\mathbf{A}_{I,E} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

of the submatrix $\mathbf{A}_\Sigma$ of the reference element stiffness matrix. Further, let $\boldsymbol{\mathcal{A}}_I$ and $\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}$ be the preconditioners for $\mathbf{A}_I$ and $\mathbb{S}_E$ satisfying the inequalities

$$\underline{\beta}_I\boldsymbol{\mathcal{A}}_I \le \mathbf{A}_I \le \overline{\beta}_I\boldsymbol{\mathcal{A}}_I \quad \text{and} \quad \underline{\beta}_E\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \le \mathbb{S}_E \le \overline{\beta}_E\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \tag{7.19}$$

with positive values $\underline{\beta}_I, \overline{\beta}_I, \underline{\beta}_E$ and $\overline{\beta}_E$. In many practical applications, the polynomial spaces on the reference elements are invariant with respect to the numbering of the edges. Thus, for simplicity, this property is here assumed as well. More specifically, we here assume that square or unilateral triangular p reference element are used with the complete polynomial spaces $\mathcal{Q}_p$ or $\mathcal{P}_p$, respectively. A rather good preconditioner $\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}$ for the Schur complement $\mathbb{S}_E$ can be defined as a block diagonal matrix of the forms

$$\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}] \text{ or } \boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}] \tag{7.20}$$

with the identical (at the proper numberings of d.o.f.) blocks for the square and triangular reference elements, respectively. These blocks are denoted $\boldsymbol{\mathcal{S}}_{00}$, and each corresponds to one edge. As it will be shown, for different reference elements with different bases of the edge coordinate polynomials, there exist preconditioners $\boldsymbol{\mathcal{S}}_{00}$ with properties needed for efficient computations. In other words, solving systems of algebraic equations with these matrices is sufficiently cheap and, for instance, requires from $\mathcal{O}(p\log p)$ to $\mathcal{O}(p^2)$ a.o., which numbers are inside the zone of the optimality of the DD solver. At the use of such matrices in the Schur complement preconditioner (7.20), one has $\underline{\beta}_E \succ (1+\log p)^{-2}$ and $\overline{\beta}_E \prec 1$, and, therefore, losses in the relative condition numbers are tolerable. These properties can be guaranteed, if, *e.g.*, the preconditioner $\boldsymbol{\mathcal{S}}_{00}$ is spectrally equivalent to the matrix of the quadratic form ${}_{00}|\cdot|^2_{1/2,E_\circ}$ on the space of the polynomials $\mathring{\mathcal{P}}(E_\circ)$

which is the restriction of $\mathcal{U}_E$ to the typical reference element edge $E_\circ$. In other words, we have

$$\|\mathbf{v}\|_{\boldsymbol{\mathcal{S}}_{00}} \asymp {}_{00}|v|^2_{1/2,E_\circ}\,, \quad \forall\ \mathbf{v} \leftrightarrow v \in \mathcal{U}_E\big|_{E_\circ}\,. \tag{7.21}$$

Let $\mathcal{P}_{E,\mathrm{ref}} : \mathcal{U}_E \to \mathcal{U}$ and $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}} : U_E \to U$ be notations for the prolongation operator on the reference element $\overline{\tau}_0$ from its edges and for the matrix of this operator, respectively. The quality of prolongations is characterized by the value of $c_{0,\mathcal{P}} > 0$ in the inequality

$$|\mathcal{P}_{E,\mathrm{ref}}\, v|^{\,2}_{1,\tau_0} \le c_{0,\mathcal{P}}\, |v|^{\,2}_{1/2,\partial\tau_0}\,, \quad \forall\ v \in \mathcal{U}_E(\partial\tau_0)\,. \tag{7.22}$$

The operator $\mathcal{P}_{E,\mathrm{ref}}$ is usually the restriction of the prolongation operator $\mathcal{P}_{B,\mathrm{ref}} : \mathcal{U}_B \to \mathcal{U}$, whereas its matrix $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}} : U_B \to U$ is the restriction of the matrix $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}} : U_B \to U$ and $\mathcal{P}_{B,\mathrm{ref}}$ satisfies the same type inequality

$$|\mathcal{P}_{B,\mathrm{ref}}\, v|^{\,2}_{1,\tau_0} \le c_{0,\mathcal{P}}\, |v|^{\,2}_{1/2,\partial\tau_0}\,, \quad \forall\ v \in \mathcal{U}_B(\partial\tau_0)\,. \tag{7.23}$$

From (7.23) it follows that the operator $\mathcal{P}_{B,\mathrm{ref}}$ should exactly recreate constant functions from their traces. Indeed, often the prolongation operator provides that

$$\mathcal{P}_{B,\mathrm{ref}}\,(v|_{\partial\tau_0}) = v\,, \quad \forall v \in \mathfrak{S}_1\,, \tag{7.24}$$

where $\mathfrak{S}_1 = \mathcal{P}_1, \mathcal{Q}_1$, accordingly with the triangular or square shapes of the reference element. That is $\mathfrak{S}_1 = \mathcal{U}_V$ for the hierarchical reference elements.

Let $\boldsymbol{\mathcal{P}}_{I\leftarrow E_q,\mathrm{ref}}$ be the block of the matrix $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$, which maps vectors living on one edge E_q of the reference element into the space U_I. If the space $\mathcal{U}$ is isotropic, these blocks, corresponding to different edges, are identical up to numbering of degrees of freedom.

By means of the introduced standard preconditioners $\boldsymbol{\mathcal{A}}_I$ and $\boldsymbol{\mathcal{S}}_{00}$, the prolongation matrix $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$ and the matrix $\mathbf{A}_V$, we define all matrices entering the DD preconditioner $\boldsymbol{\mathcal{K}}$:

- *The preconditioner for local Dirichlet problems* is the same as in (7.8).
- *The Schur complement preconditioner for edges* can be defined as the block diagonal matrix (7.18) having for independent blocks the matrices

$$\boldsymbol{\mathcal{S}}_{E_q} = (\varrho_{r'_q} + \varrho_{r''_q})\boldsymbol{\mathcal{S}}_{00} \tag{7.25}$$

  as independent blocks, where $r = r'_q, r''_q$ are the numbers of elements adjacent to the edge E_q.
- *Prolongation matrix from edges* $\boldsymbol{\mathcal{P}}_E$ is uniquely defined by the condition that its restriction to an element τ_r is $\boldsymbol{\mathcal{P}}_{E,r} = \boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$.

- *The preconditioner for the vertex subproblem.* To define the vertex subproblem preconditioner $\boldsymbol{\mathcal{K}}_V$, we can simply set $\boldsymbol{\mathcal{K}}_V = \mathbf{K}_V$, or we can assemble $\boldsymbol{\mathcal{K}}_V$ from the stiffness matrices $\varrho_r \mathbf{A}_V$ of finite elements on subdomains τ_r, and get

$$\boldsymbol{\mathcal{K}}_V = \biguplus_{r=1}^{\mathcal{R}} \boldsymbol{\mathcal{K}}_{V_r}, \quad \text{with } \boldsymbol{\mathcal{K}}_{V_r} = \varrho_r \mathbf{A}_V . \tag{7.26}$$

  If the dimension of the vertex space V_V is large, then a secondary iterative method, like multigrid, PCGM with the precondioners BPX, MDS etc., is used as inexact solver for the systems with the matrix $\boldsymbol{\mathcal{K}}_V$, and results in a different, but spectrally equivalent preconditioner $\boldsymbol{\mathcal{K}}_{V,\mathrm{it}}$.

We introduce the matrices $\mathbb{T}_V^\top$, $\mathbb{T}_{E_q}^\top$, , $\mathbb{T}_{I_r}^\top$ which pick up the entries for all vertex d.o.f., the d.o.f. belonging to the edge E_q, and the internal for an element τ_r d.o.f., respectively, from any vector belonging to the space V. Let also g', g'' be the numbers of the reference element edges, corresponding to the edge $\overline{E}_q = \overline{\tau}_{r'} \cap \overline{\tau}_{r''}$ in elements $\tau_{r'}$ and $\tau_{r''}$, respectively. The preconditioners described above and the prolongation matrix allow us to define the DD preconditioner $\boldsymbol{\mathcal{K}}$, the resulting expression for which is similar to (4.30):

$$\begin{aligned}\boldsymbol{\mathcal{K}}^{-1} &= \boldsymbol{\mathcal{K}}_I^+ + \boldsymbol{\mathcal{P}}_E \boldsymbol{\mathcal{S}}_E^{-1} \boldsymbol{\mathcal{P}}_E^\top + \boldsymbol{\mathcal{K}}_V^+ \\ &= \sum_{r=1}^{\mathcal{R}} \mathbb{T}_{I_r} \varrho_r^{-1} \boldsymbol{\mathcal{A}}_I^{-1} \mathbb{T}_{I_r}^\top + \sum_{q=1}^{Q} \varrho_q^{-1} \mathbb{T} \boldsymbol{\mathcal{S}}_{00}^{-1} \mathbb{T}^\top + \mathbb{T}_V \boldsymbol{\mathcal{K}}_V^{-1} \mathbb{T}_V^\top , \end{aligned} \tag{7.27}$$

with $\varrho_q = \varrho_{r'_q} + \varrho_{r''_q}$ and $\mathbb{T} = \mathbb{T}_{I_{r''_q}} \boldsymbol{\mathcal{P}}_{I \leftarrow E_{g''},\mathrm{ref}} + \mathbb{T}_{I_{r'_q}} \boldsymbol{\mathcal{P}}_{I \leftarrow E_{g'},\mathrm{ref}} + \mathbb{T}_{E_q}$. The difference in the appearance of the last term is explained by the assumption that the vertex coordinate functions of triangular/square reference elements are linear/bilinear polynomials, respectively. Among the components of this preconditioner, only the vertex subproblem preconditioner $\boldsymbol{\mathcal{K}}_V$, the dimension of which is much smaller than the dimension of $\mathbf{K}$, is sensitive to the coefficient ϱ jumps.

7.1.3 *Relative Condition Number of DD Preconditioners*

In this subsection we study the DD preconditioner of BPS type. An essential part of the analysis is the estimation of the influence of two steps in the simplification of DD preconditioner, which are similar to those implemented at the derivation of the DD preconditioner for the h-version (4.30)

and the interface Schur complement preconditioner (4.17). One step is the estimation of the effect of the splitting of the vertex subproblem from the rest. Another step takes into account decoupling all edges in the edge Schur complement preconditioner.

7.1.3.1 *Basic conclusions*

Below we formulate conclusions on the influence of the choice of preconditioners $\boldsymbol{\mathcal{A}}_I$ and $\boldsymbol{\mathcal{S}}_{00}$, the prolongation matrix $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$, and the global vertex problem preconditioner $\boldsymbol{\mathcal{K}}_V$ on the relative condition number of the DD preconditioner $\boldsymbol{\mathcal{K}}$. We concentrate on the case of $\boldsymbol{\mathcal{K}}$ in (7.27) and $\mathcal{U}_V = \mathcal{P}_1, \mathcal{Q}_1$ respectively to the use of triangular and square reference elements. By $\underline{c}$ and $\overline{c}$ are denoted positive constants, depending only on the generalized conditions of the angular quasiuniformity.

Theorem 7.1. *Let $\boldsymbol{\mathcal{K}}$ be the DD preconditioner (7.27) for the finite element matrix $\mathbf{K}$, and let the inequalities (7.19) and (7.22) be valid. Then the inequalities*

$$\underline{\gamma}\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{\gamma}\boldsymbol{\mathcal{K}} \tag{7.28}$$

hold, where the values $\underline{\gamma}$ and $\overline{\gamma}$ satisfy the bounds

$$\underline{\gamma} \ge \frac{\underline{c}\,\min\{\underline{\beta}_I, \underline{\beta}_E\}}{(c_{0,\mathcal{P}}(1+\log p))} \quad \text{and} \quad \overline{\gamma} \le \overline{c}\, c_{0,\mathcal{P}} \max\{\overline{\beta}_I, \overline{\beta}_E\}\,.$$

If, besides, the prolongation $\mathcal{P}_{E,\mathrm{ref}}$ is the restriction of $\mathcal{P}_{B,\mathrm{ref}}$, such that (7.23) and (7.24) hold, and a positive number $\underline{\beta}_{EB}$ satisfies the inequality

$$\underline{\beta}_{EB}\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \le \mathbb{S}_B\,, \tag{7.29}$$

then

$$\underline{\gamma} \ge \underline{c}\, c_{0,\mathcal{P}}^{-1} \min\{\underline{\beta}_I, \underline{\beta}_{EB}, \frac{1}{1+\log p}\} \quad \textit{and} \quad \overline{\gamma} \le \overline{c}\, c_{0,\mathcal{P}} \overline{\beta}_I\,. \tag{7.30}$$

Under additional assumptions (7.20) and (7.21), we have

$$\underline{\beta}_{EB} \succ (1+\log^2 p)^{-1}. \tag{7.31}$$

Proof. We will consecutively prove (7.28), (7.30) and (7.31). As usual, in view of the model problem under consideration, the generalized conditions of shape regularity, and the assembly procedure, the analysis is reduced to the reference element. In particular, we can use the inequalities (7.47) which underpin the proof of Lemma 7.2, found in the next subsubsection.

Applying Corollary 4.1 to the matrix $\mathbf{A}_\Sigma$ and taking into account (7.19), (7.22) one comes to (7.28).

Turning to the proof of (7.30), for more definiteness we assume that the vertex component preconditioner $\boldsymbol{\mathcal{K}}_V$ is assembled according to (7.26) and use the notation $\mathbb{T}^\top_{B,B_r}$ for the matrix, which picks up the entries of a vector $\mathbf{v}_B \in V_B$ living on the boundary $\partial\tau_r$. Respectively, if to write the DD preconditioner in the form

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^+ + \boldsymbol{\mathcal{P}}_B \boldsymbol{\mathcal{S}}_B^{-1} \boldsymbol{\mathcal{P}}_B^\top \tag{7.32}$$

the inter-element boundary Schur complement preconditioner will be

$$\boldsymbol{\mathcal{S}}_B = \sum_{r=1}^{\mathcal{R}} \mathbb{T}_{B,B_r} \varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \mathbb{T}^\top_{B,B_r} = \biguplus_{r=1}^{\mathcal{R}} \varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \tag{7.33}$$

with

$$\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} = \begin{pmatrix} \boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_V \end{pmatrix} \tag{7.34}$$

playing the role of the preconditioner for the reference element Schur complement $\mathbb{S}_B$.

The proof of (7.30) can be obtained by virtue of Lemma 7.1 and bounds for $\underline{\beta}_B$, $\overline{\beta}_B$. For that purpose we need one of useful representations of the norm and the seminorm for the space $H^{1/2}(\partial\tau_0)$. Let us denote by $\gamma_i,\ i = 1,2,3,4$, edges of the square τ_0 ordered, *e.g.*, counter-clockwise and by $\wedge^{(i)}$ its vertices between the edges γ_i, γ_{i+1} with the index understood modulo 4. We consider a FE function $v_B = v_E + v_V$ assuming by v_i and $v_{E,i}$ restrictions of v_B and v_E, respectively, to γ_i and by $v_{j,i}(t)$ the value of v_j at the distance t from vertex $\wedge^{(i)}$. One of the equivalent characterizations of the norm $\|v_B\|_{1/2,\partial\tau_0}$ is the following, see, *e.g.*, [Maz'ya and Poborchi (1998)]:

$$\begin{aligned} &\|v_B\|_{1/2,\partial\tau_0} = \|v_B\|_{0,\partial\tau_0} + |v_B|_{1/2,\partial\tau_0}\,, \\ &|v_B|_{1/2,\partial\tau_0} \simeq \sum_{i=1}^4 |v_i|^2_{1/2,\gamma_i} + \sum_{i=1}^4 \int_0^1 \tfrac{(v_{i+1,i}(t) - v_{i,i}(t))^2}{|t|}\, dt\,. \end{aligned} \tag{7.35}$$

In order to prove that $\overline{\beta}_B = \mathrm{const}$ in (7.5), we note that on account of the trace theorem for Sobolev's spaces and the prolongation theorem in the reference element polynomial spaces, expressed by the bound (7.22), we have the equivalence

$$\mathbf{v}_B^\top \mathbb{S}_B \mathbf{v}_B \asymp |v_B|_{1/2,\partial\tau_0}\,, \quad \forall\, \mathbf{v}_B \leftrightarrow v_B \in \mathcal{U}_B(\partial\tau_0)\,,$$

and, clearly, it holds also for $\mathbf{v}_E \leftrightarrow v_E \in \mathcal{U}_E(\partial\tau_0)$. Therefore, it suffices to use the definition of $\boldsymbol{\mathcal{S}}_E$ in (7.20)–(7.21) and the inequality

$$\mathbf{v}_E^\top \mathbb{S}_E \mathbf{v}_E \asymp |v_E|_{1/2,\partial\tau_0} \prec \sum_{i=1}^{4} {}_{00}|v_{E,i}|^2_{1/2,\gamma_i}\,, \quad \forall\, \mathbf{v}_E \leftrightarrow v_E \in \mathcal{U}_E(\partial\tau_0)$$

which directly follows by application of Cauchy inequality to second term on the right of second relationship (7.35).

Clearly, the 4×4 matrices $\mathbf{A}_V$ and $\mathbb{S}_V$ are spectrally equivalent, *i.e.,*

$$\mathbf{A}_V \prec \mathbb{S}_V \le \mathbf{A}_V\,. \tag{7.36}$$

The above inequalities for $\mathbb{S}_E$ and $\mathbb{S}_V$ the Cauchy inequality $\mathbb{S}_B \le 2(\mathbb{S}_E + \mathbb{S}_V)$ confirm that $\overline{\beta}_B = \text{const}$.

It is left to bound $\underline{\beta}_B$ in the inequality (7.5). Obviously, simultaneously with (7.56), proved below, we have

$$\mathbf{v}_V^\top \mathbf{A}_V \mathbf{v}_V \prec (1+\log p)\mathbf{v}_B^\top \mathbb{S}_B \mathbf{v}_B\,, \quad \forall\ \mathbf{v}_B = (\mathbf{v}_E^\top, \mathbf{v}_V^\top)^\top \in U_B\,, \tag{7.37}$$

which combined with the left inequality (7.36) result in the bound

$$\underline{\beta}_B \succ \min(\underline{\beta}_{EB}, (1+\log p)^{-1})\,. \tag{7.38}$$

Now, (7.30) follows from Lemma 7.1 and bounds obtained for $\underline{\beta}_B$ and $\overline{\beta}_B$.

Let $\mathring{\mathcal{P}}(I)$ be the space of polynomials of order $p \ge 2$ vanishing at the ends of the interval $I = (-1,1)$, and $v = \mathring{v} + \hat{v}$ with $\hat{v} \in \mathcal{P}_1(I)$. In order to get the bound (7.31) for $\underline{\beta}_{EB}$, we need some additional tools and, in particular, the inequalities

$${}_{00}|\mathring{v}|^2_{1/2,I} \preceq |\mathring{v}|^2_{1/2,I} + (1+\log p)\|\mathring{v}\|^2_{L_\infty(I)}\,, \quad \forall\ \mathring{v} \in \mathring{\mathcal{P}}_p(I)\,, \tag{7.39}$$

and

$${}_{00}|\mathring{v}|_{1/2,I} \preceq (1+\log p)|v|_{1/2,I}\,, \quad \forall\, v \in \mathcal{P}(I)\,, \tag{7.40}$$

which can be found in [Babuška *et al.* (1991)]. Since the square of the seminorm on the left-hand side can be characterized by the expressions

$$\begin{aligned} {}_{00}|v|^2_{1/2,I} &\eqsim \int_I\int_I \frac{(v(x)-v(y))^2}{(x-y)^2}\,dxdy + \int_I \frac{v^2}{1-x}\,dx + \int_I \frac{v^2}{x+1}\,dx \\ &= |v|^2_{1/2,I} + \int_I \frac{v^2}{x+1}\,dx + \int_I \frac{v^2}{1-x}\,dx \asymp |v|^2_{1/2,I} + \int_I \frac{v^2}{1-x^2}\,, \end{aligned} \tag{7.41}$$

the proof of (7.39) is reduced to the proof of the bounds

$$\int_I \frac{v^2}{x+1}\,dx,\ \int_I \frac{v^2}{1-x}\,dx \prec (1+\log p)\|v\|^2_{L_\infty(I)}\,, \quad \forall\, v \in \mathring{\mathcal{P}}_p(I)\,. \tag{7.42}$$

Clearly, we have

$$\int_{-1}^{1} \frac{v^2}{1-x}\,dx = \int_{-1}^{1-1/p^2} \frac{v^2}{1-x}\,dx + \int_{1-1/p^2}^{1} \frac{v^2}{1-x}\,dx\,. \tag{7.43}$$

In view of Markov's inequality

$$\|v'\|_{L_\infty(I)} \prec p^2 \|v\|_{L_\infty(I)}$$

and the boundary conditions $v(-1) = v(1) = 0$, we get

$$|v(x)| \prec \min(1-x, 1+x) p^2 \|v\|_{L_\infty(I)}\,,$$

and by substituting into the second term of (7.43) we arrive at the bound

$$\int_{1-1/p^2}^{1} \frac{v^2}{1-x}\,dx \prec \|v\|^2_{L_\infty(I)} \int_{1-1/p^2}^{1} (1-x) p^4 dx = \frac{1}{2}\|v\|^2_{L_\infty(I)}\,.$$

The first term in (7.43) is estimated by integration:

$$\int_{-1}^{1-1/p^2} \frac{v^2}{1-x}\,dx \le (\log 2 + \log p^2)\|v\|^2_{L_\infty(I)}\,.$$

Combining the last two estimates results in the desirable bound

$$\int_{-1}^{1} \frac{v^2}{1-x}\,dx \prec (1+\log p)\|v\|^2_{L_\infty(I)}\,.$$

Since the term

$$\int_{-1}^{1} v^2/(1+x)\,dx$$

can be bounded in the same way, we conclude that (7.39) holds.

Turning to the proof of (7.40), we substitute $\mathring{v} = v - \hat{v}$ into (7.39), use the triangular inequality and the obvious inequalities

$$|\hat{v}|_{1/2,I} \prec \|\hat{v}\|_{L_\infty(I)} \le \|v\|_{L_\infty(I)}\,, \quad \forall\, \hat{v} \in \mathcal{P}_1(I)\,,$$

we get

$$\begin{aligned} {}_{00}|\mathring{v}|_{1/2,I} &\prec |v|_{1/2,I} + (1+\log p)^{1/2}\|v\|_{L_\infty(I)} + |\hat{v}|_{1/2,I} \\ &\quad + (1+\log p)^{1/2}\|\hat{v}\|_{L_\infty(I)} \\ &\prec |v|_{1/2,I} + (1+\log p)^{1/2}\|v\|_{L_\infty(I)} + (1+\log p)^{1/2}\|\hat{v}\|_{L_\infty(I)} \\ &\prec |v|_{1/2,I} + (1+\log p)^{1/2}\|v\|_{L_\infty(I)}\,. \end{aligned}$$

By virtue of (7.53) it follows that

$${}_{00}|\mathring{v}|_{1/2,I} \prec (1+\log p)\|v\|_{1/2,I}. \tag{7.44}$$

Looking at the functional $\Theta(v) = {}_{00}|\mathring{v}|_{1/2,I},\ \forall\, v \in \mathcal{P}_p$, we see that, for any $v = \mathring{v} + \hat{v}$ with a fixed $\mathring{v}$, it has the same value independently of $\hat{v}$, from where we are coming to (7.40).

The definition of the preconditioner $\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}$, characterization (7.35) of the norm $\|\cdot\|_{1/2,\partial\tau_0}$, inequality (7.40), Cauchy's inequality, and (7.56) lead us to the following inequalities:

$$\begin{aligned}\mathbf{v}_E^\top \boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \mathbf{v}_E &\asymp \sum_{i=1}^{4} {}_{00}|v_{E,i}|^2_{1/2,\gamma_i} \prec (1+\log p) \sum_{i=1}^{4} |v_{E,i}|^2_{1/2,\gamma_i} \\ &\prec (1+\log p)|v_E|^2_{1/2,\partial\tau_0} \\ &\prec (1+\log p)[\,|v_B|^2_{1/2,\partial\tau_0} + |v_V|^2_{1/2,\partial\tau_0}] \\ &\prec (1+\log p)^2 |v_B|^2_{1/2,\partial\tau_0} \asymp \mathbf{v}_B^\top \mathbb{S}_B \mathbf{v}_B\,. \end{aligned} \tag{7.45}$$

From here we come to the expression (7.31) for $\underline{\beta}_{EB}$ in (7.29), and to the specific form of $\underline{\beta}_B$ in (7.38). This completes the proof of Theorem 7.1. □

In accordance with what was noted earlier, the main part of the analysis was estimation of the influence of two steps in the simplification of DD preconditioner, which are similar to those in the derivation of the DD preconditioner for the *h*-version (4.30) and the interface Schur complement preconditioner (4.17). There is an alternative and in a sense more straightforward way to transfer the results of analysis of these steps for the *h*-version into respective results for the *hp*-version. It was introduced in the paper of [Casarin (1997)] on DD solvers for *hp* discretizations of 3d elliptic problems by spectral finite elements. However, it rests upon fundamental results of [Bernardi and Maday (1992b, 1997)] on the stability of polynomial interpolations, which we touch in Subsection 7.4.2. A number of tools for direct analysis of the DD preconditioners for *hp* discretizations, presented in this subsubsection, were developed in the pioneering work by [Babuška *et al.* (1991)].

7.1.3.2 *Subsidiary results*

Now, we come to Lemma 7.2, which for the reference element is equivalent to the inequalities (7.47) that has been used in the proof of Theorem 7.1.

Lemma 7.2. *Let $\mathcal{U}_V = \mathcal{P}_1$ or $\mathcal{U}_V = \mathcal{Q}_1$ respectively to the use of triangular or square reference elements and $\boldsymbol{\mathcal{K}}_v = \mathrm{diag}\,[\mathbf{K}_\Sigma, \mathbf{K}_V]$. Then*

$$\frac{\underline{c}\,v}{1+\log p}\boldsymbol{\mathcal{K}}_v \le \mathbf{K} \le 2\boldsymbol{\mathcal{K}}_v \tag{7.46}$$

with the constant $\underline{c}_v$ *depending only on the conditions of the generalized angular quasiuniformity.*

Proof. The right part of (7.46) is the Cauchy inequality. The bounds looking similarly to (7.46) with p replaced by H/h are known from the analysis of BPS preconditioners for h discretizations of 2d second-order elliptic boundary value problems. In fact, they can be transformed into (7.46). For the problem under consideration with the generalized conditions of angular quasiuniformity fulfilled, it is sufficient to prove the left part of (7.46) for the reference element, *i.e.*, the left part of the inequalities

$$\frac{1}{1+\log p}(\mathbf{A}_\Sigma + \mathbf{A}_V) \prec \mathbf{A} \prec (\mathbf{A}_\Sigma + \mathbf{A}_V)\,. \tag{7.47}$$

Suppose $\mathcal{T}_h = \mathcal{T}_h(\tau_0)$ is the partition of τ_0 by a quasiuniform orthogonal mesh, characterized by the mesh parameter h, and $\mathcal{V}_h(\tau_0)$ is the space of continuous piecewise bilinear functions, induced by $\mathcal{T}_h$. We represent this space by the direct sum $\mathcal{V}_h(\tau_0) = \mathcal{V}_V(\tau_0) \oplus \mathcal{V}_{\Sigma,h}(\tau_0)$, where $\mathcal{V}_V(\tau_0)$ the subspace of the bilinear polynomials and $\mathcal{V}_{\Sigma,h}(\tau_0)$ contains $v \in \mathcal{V}_h(\tau_0)$ vanishing at the vertices of τ_0. According to [Bramble *et al.* (1986)], we have

$$||v_V||_{1,\tau_0} \prec (1+\log h^{-1})^{1/2}||v||_{1,\tau_0} \tag{7.48}$$

for any $v = v_V + v_\Sigma \in \mathcal{V}_V(\tau_0)$, $v_\Sigma \in \mathcal{V}_{\Sigma,h}(\tau_0)$. Let $\Xi_{\mathrm{G},p}$ be the set of $(p+1)^2$ GLL nodes on $\overline{\tau}_0$ and Ξ_p be the set of the nodes of the quasiuniform orthogonal grid such that $\Xi_{\mathrm{G},p} \subset \Xi_p$ and its size is $\mathfrak{h} = p^{-2}$. From the definition of GLL nodes it follows that such quasiuniform orthogonal grid exists, see Lemma 7.6. By $\mathcal{V}_{\mathrm{G},p}(\tau_0) = \mathcal{V}_V(\tau_0) \oplus \mathcal{V}_{\Sigma,\mathrm{G},p}(\tau_0)$ and $\mathcal{V}_\mathfrak{h}(\tau_0) = \mathcal{V}_V(\tau_0) \oplus \mathcal{V}_{\Sigma,\mathfrak{h}}(\tau_0)$ we denote the spaces of continuous piecewise bilinear functions induced by the grids with the set of nodes $\Xi_{\mathrm{GLL},p}$ and Ξ_p, respectively. Since (7.48) holds on the space $\mathcal{V}_\mathfrak{h}(\tau_0)$ with $h = \mathfrak{h}$, it is also valid on its subspace $\mathcal{V}_{\mathrm{G},p}(\tau_0) \subset \mathcal{V}_\mathfrak{h}(\tau_0)$. Therefore, inequalities

$$||v_V||_{1,\tau_0} \prec (1+\log \mathfrak{h}^{-1})^{1/2}||v||_{1,\tau_0} \prec (1+\log p)^{1/2}||v||_{1,\tau_0} \tag{7.49}$$

hold for all $v \in \mathcal{V}_{\mathrm{G},p}(\tau_0)$. For any continuous function u, we can define the unique interpolation $I_{\mathrm{G},p}u \in \mathcal{Q}_p$ over the set $\Xi_{\mathrm{G},p}$ of GLL nodes, satisfying

$$(I_{\mathrm{G},p}v)(\hat{x}) = v(\hat{x})\,, \quad \forall\, \hat{x} \in \Xi_{\mathrm{G},p}\,.$$

For the interpolation, we can establish the equivalences

$$||v||_{0,\tau_0} \asymp ||I_{\mathrm{G},p}v||_{0,\tau_0} \quad \text{and} \quad |v|_{1,\tau_0} \asymp |I_{\mathrm{G},p}v|_{1,\tau_0}\,, \tag{7.50}$$

for which we refer to Lemma 7.8 and Theorem 7.4. Now, the bound

$$||v_V||_{1,\tau_0} \prec (1+\log p)^{1/2}||I_{\mathrm{G},p}v||_{1,\tau_0}\,, \quad \forall\, v \in \mathcal{V}_{\mathrm{G},p}(\tau_0)\,, \tag{7.51}$$

which is equivalent to the bound $(1+\log p)^{-1}\,\mathbf{A}_V \prec \mathbf{A}$, follows from (7.49) and (7.50). Triangle and Cauchy's inequalities now yield

$$\mathbf{A}_\Sigma + \mathbf{A}_V \prec 2\mathbf{A} + 3\mathbf{A}_V\,.$$

Combining this inequality with the above inequality for $\mathbf{A}_V$, we come to (7.47), from where, as was mentioned, follows (7.46).

An alternative, more direct proof of (7.47) is given in the paper of [Babuška *et al.* (1991)]. This proof is based on the equivalence

$$\|v\|_{1/2,\mathcal{I}_\pi} \prec \|u\|_{1/2,I} \prec \|v\|_{1/2,\mathcal{I}_\pi}\,, \tag{7.52}$$

where u is any function from $H^{1/2}(I)$, $v(\phi) = u(\cos\phi)$, $I = (-1,1)$ and $\mathcal{I}_\pi = (0,\pi)$. We present their proof at the end of this subsection. Any 1d polynomial of order p satisfies the inequality

$$|u(x)| \prec (1+\log^{1/2} p)\|u\|_{1/2,(-1,1)}\,, \quad \forall\, u \in \mathcal{P}_p\,. \tag{7.53}$$

Indeed, representing $u \in \mathcal{P}_p$ by the series

$$u(x) = v(\phi) = \sum_{k=0}^{p} a_k \cos k\phi\,,$$

we observe that

$$\|v\|^2_{1/2,\mathcal{I}_\pi} \asymp \sum_{k=0}^{p} a_k^2(k+1)\,. \tag{7.54}$$

By the Cauchy–Schwarz inequality, (7.54) and (7.52), we get

$$\begin{aligned}
|u(x)| = |v(\phi)| &\le \sum_{k=0}^{p} |\,a_k| = \sum_{k=0}^{p} \frac{|\,a_k|}{(k+1)^{1/2}}\,(k+1)^{1/2} \\
&\le \left(\sum_{k=0}^{p} a_k^2(k+1)\right)^{1/2} \left(\sum_{k=0}^{p} \frac{1}{k+1}\right)^{1/2} \\
&\prec (1+\log^{1/2} p)\,\|v\|_{1/2,\mathcal{I}_\pi} \asymp (1+\log^{1/2} p)\,\|u\|_{1/2,I}\,.
\end{aligned}$$

If we take $a_k = (k+1)^{-1}$ and $x = 1$, then we see that

$$\|v\|_{L_\infty(\mathcal{I}_\pi)}/\|u\|_{1/2,\mathcal{I}_\pi} \succ (1+\log^{1/2} p).$$

Therefore, we have

$$\sup_{u\in\mathcal{P}_p} \frac{\|u\|_{L_\infty(I)}}{\|u\|_{1/2,I}} \asymp (1+\log^{1/2} p)\,. \tag{7.55}$$

We represent any $u \in \mathcal{U}$ in the form $u = u_\Sigma + u_V$ with $u_\Sigma \in \mathcal{U}_\Sigma$ and $u_V \in \mathcal{U}_V$. Since u_V is linear for triangular and bilinear for square reference elements, its maximum is attained at one of the vertices $\wedge^{(i)}$ of τ_0 with the coordinates denoted $x = x_{\wedge^{(i)}}$, and at these points $u_V(x_{\wedge^{(i)}}) = u(x)$. Consequently, $a_{\tau_0}(u_V, u_V) \prec \max_{\wedge^{(i)}} |u(x_{\wedge^{(i)}})|^2$ over all vertices of the reference element, and application of (7.53), the definition of the norm $\|\cdot\|^2_{1/2,\partial\tau_0}$, and the trace theorem result in the inequalities

$$\begin{aligned} a_{\tau_0}(u_V, u_V) &\prec \max_{\wedge^{(i)}} |u(x_{\wedge^{(i)}})|^2 \prec (1 + \log p)\, \|u\|^2_{1/2,E_i} \\ &\prec (1 + \log p)\, \|u\|^2_{1/2,\partial\tau_0} \prec (1 + \log p)\, \|u\|^2_{1,\tau_0}\,. \end{aligned}$$

Here E_i denotes one of the edges of τ_0 adjacent to the vertex at which u_V achieves its maximum. Further, by the factor space type arguments, we get

$$a_{\tau_0}(u_V, u_V) \prec (1 + \log p)\, |u|^2_{1,\tau_0} = (1 + \log p)\, a_{\tau_0}(u, u)\,, \tag{7.56}$$

and this coincides with the bound (7.51). □

Lemma 7.3. *Let u belong to $H^{1/2}(I)$ or ${}_{00}H^{1/2}(I)$, and let*

$$\Psi : \Psi u = v\,, \quad v(\varphi) = u(\cos\varphi)\,, \quad \varphi \in \mathcal{I}_\pi = (0, \pi)\,. \tag{7.57}$$

Then Ψ is a linear homomorphism between $H^{1/2}(I)$ and $H^{1/2}(\mathcal{I}_\pi)$ and between ${}_{00}H^{1/2}(I)$ and ${}_{00}H^{1/2}(\mathcal{I}_\pi)$, respectively.

Proof. For $\Omega_\varphi = \mathcal{I}_\pi \times I$, we define the map

$$\Phi : \Omega_\varphi \to \Omega\,, \qquad (\varphi, \zeta) \to (x, y)\,, \quad x + iy = \cos(\varphi + i\zeta)\,.$$

In what follows, for brevity we use the notations $w_z = \partial w/\partial z, z = \varphi, \zeta, x, y$. Since the cosine is a conformal mapping, the Cauchy–Riemann conditions hold, *i.e.* $x_\varphi = y_\zeta$ and $x_\zeta = -y_\varphi$, and, since $\cos'(\varphi + i\zeta) = -\sin(\varphi + i\zeta) = (x_\varphi + iy_\varphi) \neq 0$ for all $(\varphi, \zeta) \in \Omega_\varphi$, the Jacobian of Φ fulfil the relationship $J = x_\varphi^2 + y_\varphi^2 \neq 0\,,\ \forall\ (\varphi, \zeta) \in \Omega_\varphi$. Now, let u be a smooth function on Ω and $\Psi u(x, y) := u(\Phi(\varphi, \zeta)) =: v(\varphi, \zeta)$ with v defined, obviously, on Ω_φ. As a consequence of the chain rule and the Cauchy–Riemann conditions, we have

$$v_\varphi = u_x x_\varphi + u_y y_\varphi\,, \quad v_\zeta = u_x x_\zeta + u_y y_\zeta = -u_x y_\varphi + u_y x_\varphi\,,$$

and, therefore,

$$v_\varphi^2 + v_\zeta^2 = (u_x^2 + u_y^2)(x_\varphi^2 + y_\varphi^2) = (u_x^2 + u_y^2)J\,.$$

In turn, these equalities yield the representations

$$|u|^2_{1,\Omega} = \int_\Omega (u_x^2 + u_y^2) dx dy = \int_{\Omega_\varphi} (u_x^2 + u_y^2) J d\varphi d\zeta = \int_{\Omega_\varphi} (v_\varphi^2 + v_\zeta^2) d\varphi d\zeta\,.$$

If

$$w \in L^\infty(\Omega)\,, \quad w \geq 0\,, \quad \int_\Omega w^2 dxdy > 0\,,$$

then the equivalence relation

$$\|\psi\|^2_{1,\Omega_\varphi} \approx |\psi|^2_{1,\Omega_\varphi} + \int_{\Omega_\varphi} w|\psi|^2 d\varphi d\zeta\,, \quad \forall\ \psi \in H^1(\Omega_\varphi)\,, \tag{7.58}$$

holds. Relation (7.58) and the equality

$$\|u\|^2_{0,\Omega} = \int_\Omega u^2 dxdy = \int_{\Omega_\varphi} v^2 J d\varphi d\zeta\,,$$

imply inequalities

$$\|\Psi u\|_{1,\Omega_\varphi} \prec \|u\|_{1,\Omega} \prec \|\Psi u\|_{1,\Omega_\varphi}. \tag{7.59}$$

Any function $u \in H^{1/2}(-1,1)$ may be extended to $u \in H^1(\Omega)$ such that

$$\|u\|_{1,\Omega} \prec \|u\|_{1/2,I}. \tag{7.60}$$

Inequality (7.60), relation (7.58) and the trace theorem allow us to write

$$\|\Psi u\|_{1/2,\mathcal{I}_\pi} \prec \|\Psi u\|_{1,\Omega_\varphi} \prec \|u\|_{1,\Omega} \prec \|u\|_{1/2,I}\,. \tag{7.61}$$

In the same way, the inverse inequality is obtained, if we replace Ψ by Ψ^{-1}.

In order to obtain

$${}_{00}\|\Psi u\|_{1/2,\mathcal{I}_\pi} \prec {}_{00}\|u\|_{1/2,I} \prec {}_{00}\|\Psi u\|_{1/2,\mathcal{I}_\pi}\,, \tag{7.62}$$

we should additionally take two facts into account. If $u(-1) = u(1) = 0$ and $u \in H^{1/2}_{00}(I)$, then the function u can be extended to u on Ω in such a way that $u = 0$ on $\partial\Omega$ except for one edge, from which it is extended, and, that instead of (7.60), we have

$$\|u\|_{1,\Omega} \prec \|u\|_{1/2,\partial\Omega}\,. \tag{7.63}$$

Besides, for $u \in H^{1/2}(\partial\Omega)$ distinct from zero only on one edge, we have

$$\|u\|_{1/2,\partial\Omega} \asymp {}_{00}\|u\|_{1/2,I}\,,$$

and the same is true for Ω_φ. Obviously, we have the relation

$$\frac{\sin\varphi}{|1-\cos^2\varphi|} \asymp \frac{1}{\varphi(\pi-\varphi)}\,, \quad \varphi \in \mathcal{I}_\pi\,,$$

and, therefore, we get

$$\int_I \frac{u^2}{|1-x^2|}\,dx = \int_{\mathcal{I}_\pi} \frac{u^2(\cos\varphi)}{|1-\cos^2\varphi|}(-\sin\varphi)\,d\varphi \asymp \int_{\mathcal{I}_\pi} \frac{v^2(\varphi)}{\varphi(\pi-\varphi)}\,d\varphi\,.$$

From the last relations, equivalence of additional terms in ${}_{00}\|\Psi u\|_{1/2,\mathcal{I}_\pi}$ and ${}_{00}\|u\|_{1/2,I}$ directly follows. This completes the proof. □

7.2 Prolongations and Bounded Extension Splitting

Bounded extension splitting of the FE space plays an essential role in designing efficient *hp* Dirichlet–Dirichlet DD algorithms. If the interface coordinate functions are low energy functions, then we can split the internal unknowns from the interface boundary unknowns without losses in the relative condition number of the DD preconditioner. This is especially important for *hp* discretizations since they result in much more complex systems of algebraic equations. However, for obvious reasons DD preconditioner-solvers, if properly arranged according to (7.2), as a rule will result in more efficient algorithms. It is clear that discrete harmonic and almost discrete harmonic prolongations play an exceptional role in analysis of DD algorithms. Let us emphasize that, as it has became accepted in the literature, the term *almost discrete harmonic prolongation* is used equally with the term *low energy prolongation*. If prolongation is performed from the set of a lesser dimension, then it is as well often called *lifting*.

7.2.1 *A Few Remarks on Low Energy Prolongations in Polynomial Spaces*

Let us suppose that the generalized conditions of shape regularity are fulfilled. Then the requirement for FE boundary coordinate functions to have low energy is equivalent to the requirement that the bound

$$\|\mathbf{v}_B\|_{\mathbf{A}} \le \beta_B \|\mathbf{v}_B\|_{\mathbb{S}_B}\,, \quad \forall\ \mathbf{v}_B \in U_B\,, \tag{7.64}$$

holds with a sufficiently good β_B that in the best case should be a constant independent of h and p. Low energy coordinate polynomials can be defined with the use of low energy prolongation operators in the corresponding polynomial subspaces of the Sobolev space $H^1(\tau_0)$. In this relation, the following result is important.

Theorem 7.2. *Let $d = 2, 3$ and τ_0 be the unit simplex or cube equipped with the polynomial spaces $\mathcal{U} = \mathcal{P}_p$ or $\mathcal{Q}_p$, respectively, and $\mathcal{U}_B(\partial\tau_0)$ be the restriction of $\mathcal{U}$ to the boundary $\partial\tau_0$. Then there exists the prolongation operator $\mathcal{P} : \mathcal{U}_B(\partial\tau_0) \to \mathcal{U}$ such that, for any $\phi \in \mathcal{U}_B(\partial\tau_0)$, the prolongation $v := \mathcal{P}\phi$ satisfies the inequalities*

$$|v|_{1,\tau_0} \le \mathbb{c}_{\mathcal{P}} |\phi|_{1/2,\partial\tau_0} \quad \textit{and} \quad \|v\|_{1,\tau_0} \le \mathbb{c}_{\mathcal{P}} \|\phi\|_{1/2,\partial\tau_0} \tag{7.65}$$

with an absolute constant $\mathbb{c}_{\mathcal{P}}$.

Note that first bound (7.65) implies that the prolongation operator must recover a constant function from its trace.

Omitting the proof, we will refer to a few of the many papers on this subject that contributed to this Theorem. Apparently, [Babuška and Suri (1987)] were first to construct the polynomial liftings for the unit triangle and square, satisfying (7.65). They used their liftings for the purpose of establishing optimal convergence rates for the *hp* finite element methods for two dimensional elliptic boundary value problems. Afterwards, several successful attempts were made to simplify proofs of the bound (7.65) and structures of lifting operators and to expand results to the case of the polynomial subspaces in weighted Sobolev spaces. We refer the reader to [Maday (1989)], [Bernardi and Maday (1989, 1997)] and [Babuška *et al.* (1991)] for the case of triangular and square elements, and to [Belgacem (1994)] and [Munoz-Sola (1997)] for cubic and tetrahedral elements, In these papers many additional references may be found. Active studies of polynomial prolongations were continued, in particular, in the direction of the expansions of the results to different special polynomial spaces, in particular, to the spaces $\mathbf{H}(\mathbf{curl}, \tau_0) \cap \mathcal{P}_p$ and $H(\mathrm{div}, \tau_0) \cap \mathcal{P}_p$ on tetrahedra τ_0, see, *e.g.,* the series of papers by [Demkowicz *et al.* (2008, 2009, 2012)].

Early studies of prolongation operators in polynomial spaces were made with the intention of their use in the approximation and convergence analysis of the *hp* finite element methods and iterative DD methods for the respective systems of finite element equations. Hypothetically the same prolongation operators could be used in practical DD algorithms in the procedures of prolongations from the interface inside finite elements and in the opposite procedures of restrictions to the interface. However, in this case the additional requirement arises: prolongation must be computationally cheap. From this point of view some analytical prolongation operators were considered in [Korneev and Jensen (1997, 1999)] and [Korneev *et al.* (2002b,a)]. It was shown that the use of analytical prolongations can be cheaper than prolongations, *e.g.*, by means of solving local Dirichlet problems by the direct methods.

Analysis and justification of uniform in p spectral equivalence and almost spectral equivalence of finite-difference ($\backsim$ first order finite element) preconditioners for cubic reference p-elements stiffness and mass matrices acquired a good state in the 90s. This eased the problem of obtaining prolongation operators, possessing both important for computations properties of boundness and cost, since it became possible to mimic prolongations in the counterpart spaces of continuous piecewise linear/bilinear functions, generating preconditioners. One opened option can be specially mentioned. On the basis of the finite-difference preconditioners, several fast solvers for

the internal Dirichlet problems on the reference elements were developed soon. This allowed to prove almost optimality of the most universal prolongation procedure by means of inexact Dirichlet iterative solvers. Such prolongations possess main wishful properties for efficient computation: they guarantee (7.65) with the constant independent of p for almost optimal computational work. In particular, as a rule for many types of p-elements they require only by the factor $\log(1+p)$ more iterations, than for solution of Dirichlet problems with the homogeneous boundary condition. Examples of such prolongations inside hierarchical and spectral cubic reference p-elements are given in [Korneev *et al.* (2003a,b)] and [Korneev and Rytov (2008)].

[Beuchler and Schöberl (2005)] provided the lifting operator $\mathcal{P} : \mathcal{U}_E \to \mathcal{U}_\Sigma$ for the traces of polynomials u_Σ from the trunk space $\mathcal{Q}_p$ vanishing at the vertices of $\tau_0 = (-1,1)\times(-1,1)$. The lifting is produced separately for each edge, *i.e.*, $\mathcal{P}u_E = \sum_i \mathcal{P}_i u_{E_i},\ i = 1,2,3,4$, where E_i is one edge, by the direct solver for the discrete Dirichlet problem. The latter is governed by the specific preconditioner for the reference element stiffness matrix, corresponding to the homogeneous Dirichlet boundary condition on the three edges $E_k,\ k \neq i$. The preconditioner is multilevel along the direction parallel to E_i. It is obtained from the first order finite element preconditioner by the transformation of the nodal basis along this direction to the special multilevel wavelet basis and by the replacement of the transformed preconditioner by the diagonal one. The lifting operator has linear complexity and satisfies the inequalities

$$|\mathcal{P}u_E|_{1,\tau_0} \prec c(p)|u_E|_{1/2,\partial\tau_0} \quad \text{and} \quad \|\mathcal{P}u_E\|_{1,\tau_0} \prec c(p)\|u_E\|_{1/2,\partial\tau_0}\,, \quad (7.66)$$

with $c(p) = 1 + \log p$.

An important immediate consequence of Theorem 7.2 is that, for the complete reference p-elements, the energy, induced by the Schur complement, is equivalent to $|\cdot|^2_{1/2,\partial\tau_0}$, *i.e.*,

$$|v|^2_{1/2,\partial\tau_0} \prec \mathbf{v}^\top \mathbb{S}_B \mathbf{v} \prec |v|^2_{1/2,\partial\tau_0}\,, \quad \forall\ \mathbf{v} \Leftrightarrow v \in \mathcal{U}_B. \quad (7.67)$$

Indeed, if $v = \mathcal{P}v_B$ and $\mathbf{v}_B \Leftrightarrow v_B$, then

$$\mathbf{v}^\top \mathbb{S}_B \mathbf{v} = \inf_{\phi\in\mathcal{U}_I} |v_B + \phi|^2_{1,\tau_0} \prec |v|^2_{1,\tau_0}\,,$$

and applying (7.65), we obtain the right inequality. The left inequality follows by the trace theorem for functions from $H^1(\tau_0)$.

Therefore, taking domains of finite elements for subdomains of decomposition, we get the characterization of the boundary Schur complement similar to one arising in the h-version, *cf.* (7.67) and (4.13).

In view of (7.67), the inequality (7.64) is equivalent to

$$|v_B|_{1,\tau_0} \prec \sqrt{\beta_B}\,|v_B|_{1/2,\partial\tau_0}\,, \quad \forall\ v \in \mathcal{U}_B\,. \tag{7.68}$$

If boundary coordinate polynomials of the reference element are obtained by means of a low energy prolongation operator for which (7.68) holds, then, for the block diagonal preconditioner $\boldsymbol{\mathcal{K}} = \mathrm{diag}\,[\mathbf{K}_I, \mathbf{K}_B]$, we have

$$\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] \le c\beta_B\,,$$

with some positive constant c. If the edge coordinate functions are obtained by means of a low energy prolongation operator and vertex coordinate functions are linear/bilinear for triangular/quadrangular elements, then the preconditioner $\boldsymbol{\mathcal{K}}_* = \mathrm{diag}\,[\mathbf{K}_I, \mathbf{K}_E, \mathbf{K}_V]$ provides the relative condition number bound

$$\mathrm{cond}\,[\boldsymbol{\mathcal{K}}_*^{-1}\mathbf{K}] \le c_*(1+\log p)\beta_B\,,$$

with c_*, as well as c, depending only of the generalized conditions of the angular quasiuniformity. These relative condition number estimates for the block diagonal preconditioners follow from Theorem 4.2, inequality (7.64), similar inequality for the reference element edge functions, and Lemma 7.2.

Clearly, the use of the transformation to low energy interface coordinate functions in computations does not suggest easy ways to the reduction of the computational cost. Although the block diagonal forms of DD preconditioners is an obvious advantage, the blocks $\mathbf{K}_{I,B}$ and $\mathbf{K}_B$, as a rule, become dense, if even before the transformation they were not. Therefore, computation of these matrices as well as matrix-vector multiplications by them can become expensive.

7.2.2 *Almost Discrete Harmonic Prolongations by Means of Inexact Subdomain Dirichlet Solvers*

To date, there exists a variety of specially designed direct prolongation operators, which do not require solution of any systems of algebraic equations. Moreover, many of them are optimal or almost optimal in the computational cost, but have limitations of being only applicable either to specific reference elements, or to specific domains and meshes, or under the condition of smooth coefficients of the elliptic equations. Another more universal option, which is considered in this section, is to produce prolongations by means of solution of local Dirichlet problems for Laplace equation. Here special attention should be paid to the choice of the minimal fixed number of iterations, that can guarantee the last inequality (7.13) with sufficiently

good value of $c_{0,\mathcal{P}}$. While inexact solvers are used for solving local Dirichlet problems with homogeneous boundary conditions, the number of iterations completely depends on the efficiency of the preconditioners. Contrary to this, the necessary number of iterations in prolongation procedures additionally depends on the type of the edge coordinate functions of reference elements and, more precisely, on how the edge coordinate functions are defined inside reference elements.

Suppose the domains of finite elements serve for subdomains of decomposition, the generalized conditions of the shape regularity are fulfilled, and there is no large variations of the equation coefficients inside each finite element. Then prolongation operators can be defined by such for the corresponding reference elements, and quality of the global prolongation operator can be well controlled by the last inequality (7.13). For definiteness, we turn to the case of the square reference element and assume that:

A1) for each vertex, the coordinate polynomial is bilinear and equals 1 at the vertex and 0 at other vertices,

A2) for each edge, the corresponding edge polynomials are linear in the direction orthogonal to the edge,

A3) the reference element polynomial space is $\mathcal{Q}_p,\ p \ge 1$.

Lemma 7.4. *Let assumptions A1) – A3) be fulfilled,* $\mathbf{A}$ *be the reference element stiffness matrix, and* $\mathbb{A}$ *and* $\boldsymbol{\mathcal{A}}_I$ *be symmetric nonnegative preconditioners for* $\mathbf{A}$ *and its block* $\mathbf{A}_I$*, respectively, satisfying the inequalities*

$$\underline{\beta}\,\mathbb{A} \le \mathbf{A} \le \overline{\beta}\,\mathbb{A} \quad \textit{and} \quad \underline{\gamma}_I\,\boldsymbol{\mathcal{A}}_I \le \mathbf{A}_I \le \overline{\gamma}_I\,\boldsymbol{\mathcal{A}}_I\,. \tag{7.69}$$

Let also $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}} : U_B \to U$ *be such that, in the vector* $\mathbf{u} = \boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}\,\mathbf{v}_B$, $\mathbf{u}^\top = (\mathbf{u}_I^\top \mathbf{u}_B^\top)$*, the second subvector is* $\mathbf{u}_B = \mathbf{v}_B$ *and the first subvector is found iteratively,* i.e., $\mathbf{u}_I := -\mathbb{A}_{I,\mathrm{it}}^{-1}\mathbb{A}_{I,B}\mathbf{v}_B$*, where* $\mathbb{A}_{I,\mathrm{it}} = \boldsymbol{\mathcal{I}}[\mathbb{A}_I, \boldsymbol{\mathcal{A}}_I, \nu]$*. Then the bound*

$$\|\mathbf{u}\|_{\mathbf{A}} = |u|_{1,\,\partial\tau_0} \prec |v_B|_{1/2,\,\partial\tau_0}\,, \quad u \leftrightarrow \mathbf{u}\,, \quad v_B \leftrightarrow \mathbf{v}_B\,, \tag{7.70}$$

is guaranteed for

$$\nu \succ [\,\log(\overline{\beta}/\underline{\beta}) + \log(1+p)\,]/\log\rho^{-1} \tag{7.71}$$

iterations, where $\rho \le (1-\theta)/(1+\theta)$ *and* $\theta = \sqrt{\underline{\beta}\,\underline{\gamma}_I/\overline{\gamma}_I\,\overline{\beta}}$.

Proof. First of all we note that so defined prolongation recovers a constant polynomial on the reference element from its trace on the boundary $\partial\tau_0$, and the same is true for any bilinear polynomial. Similarly to first

line of (4.55), an easy bound of the seminorm $|v_B|_{1,\tau_0}$, the factor space argument, and Markov's type inequality allow us to write:

$$\|\mathbf{v}_B\|^2_{\mathbf{A}} = |v_B|^2_{1,\tau_0} \prec \inf_c \|v_B + c\|^2_{1,\partial\tau_0} \asymp |v_B|^2_{1,\partial\tau_0} \prec p^2 |v_B|^2_{1/2,\,\partial\tau_0}\,. \quad (7.72)$$

The remaining steps are simpler than in the proof of Lemma 4.1, since no mappings and quasiuniformity conditions are involved. They result in the key bound

$$\|\mathbf{u}\|_{\mathbf{A}} \prec [1 + (\overline{\beta}/\underline{\beta})\rho^{\nu}(1+p^2)]\,|v_B|_{1/2,\,\partial\tau_0}\,. \quad (7.73)$$

Now, Lemma 7.4 follows from this inequality, (7.69), the definition of $\mathbb{A}_{I,\mathrm{it}}$, and Lemma 2.1. □

The bounds (7.70) and (7.71) indeed hold for a wider range of square and triangular reference elements. For triangular reference elements, the following sufficient conditions can be formulated: A1) basis polynomials include three nodal vertex polynomials, A2) derivatives of the boundary polynomials are not too big, so that $|v_B|_{1,\tau_0} \prec p^t \|v_B\|_{1,\partial\tau_0}$ with some fixed t, A3) the reference element space is $\mathcal{P}_p$, $p \geq 1$. The given here assumption A2 can be used in place of the one for square reference elements, formulated before Lemma 7.4. The power t influences the constant in (7.70).

For Lagrange and, in particular, spectral elements, the prolongation matrix of Lemma does not return a constant function from its trace, because A1) is not fulfilled. In such a case, we can use the trick described in Subsection 4.2.3, see (4.51), for a resembling situation in the h-version. As well, some preliminary prolongation may be employed. Let $\mathbb{T}_B$ specify some cheap prolongation, which returns constant function from its trace. Then the final prolongation matrix can be defined as

$$\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}} = \begin{pmatrix} -\mathbb{A}^{-1}_{I,\mathrm{it}}\,(\mathbb{A}_{I,B} + \mathbb{A}_I \mathbb{T}_{I,B}) + \mathbb{T}_{I,B} \\ \mathbf{I} \end{pmatrix}. \quad (7.74)$$

In this way, the number of iterations for the inequalities (7.70) to be fulfilled can be made the same as in (7.71). It is sufficient, for instance, to define $\mathbb{T}_{I,B}$ in such a way, that the polynomial $v \leftrightarrow \mathbb{T}_B \mathbf{v}_B$, $\forall\, \mathbf{v}_B \in U_B$, belongs to the subspace $\mathcal{U}_{\mathbb{B}}$, containing polynomials described in the assumptions A1) and A2). For the hierarchical reference elements on the unit cube, introduced in Subection 7.3.1, $\mathcal{U}_{\mathbb{B}} = \mathcal{U}_B$, where $\mathcal{U}_B$ is the subspace of the boundary polynomials.

Corollary 7.1. *Let assumptions of Lemma 7.4 be fulfilled except A1), and let the preliminary prolongation* $\mathbb{T}_B$ *be such that* $\mathbb{T}_B \mathbf{v}_B \leftrightarrow v \in \mathcal{U}_B$, $\forall\, \mathbf{v}_B$,

where $\mathcal{U}_B$ is the subspace of the boundary polynomials for the hierarchical reference element of the order p. Let also $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}$ be defined by (7.74). Then (7.70) holds for ν satisfying (7.71).

Proof. Clearly, the proof is the same as the proof of Lemma 7.4. □

Prolongation by means of inexact Dirichlet solvers allows us to use the same iterative solver in the two main components of the Dirichlet–Dirichlet DD algorithm: one solves internal Dirichlet problems on the subdomains of the decomposition, and another carries out the prolongations. Algorithms of such prolongations are less dependent on the problem to be solved and type of its discretization. For instance, in general, proofs of efficiency are based on the equivalence of the broken weighted $H^{1/2}$-seminorm to the interface Schur complement, see (4.10) and (4.11). It is obvious that the prolongations by means of inexact Dirichlet solvers can be also efficient when this equivalence does not take place.

7.3 Square Reference *p*-Elements, Their Stiffness and Mass Matrices

The computational properties of the *hp* stiffness matrices entirely depend on the choice of coordinate polynomials in the reference element functional space which tolerates a variety of options. Depending on these properties, condition numbers may grow too fast with the degree p of the polynomials The first choice, that seemed reasonable for high accuracy computations, was the Lagrange interpolation coordinate polynomials over the set of uniformly distributed nodes. Let us suppose that

$$x_i = i/p\,, \quad i = 0, 1, \ldots, p\,,$$

are the uniformly distributed nodes on $[0, 1]$, and $\hat{p}^{(i)}(x)$ are the corresponding Lagrange interpolation coordinate polynomials such that $\hat{p}^{(i)}(x_j) = \delta_{i,j}$ and $\mathcal{P}_p = \mathrm{span}\,[\hat{p}^{(i)}(x)]_{i=0}^{p}$. For the interval $(-1, 1)$, coordinate polynomials $\hat{p}^{(i)}$, $i = 0, 1, \ldots, 7$, of the degree $p = 7$ are pictured in Fig. 7.1.

Furthermore, let $\mathbf{K}_p$ be the stiffness matrix generated by 1d bilinear form

$$a(u, v) = \int_0^1 u'v'dx \tag{7.75}$$

on the space $\mathcal{P}_p$ with the basis $\{\hat{p}^{(i)}(x)\}$, and let $\boldsymbol{\Delta}$ be the stiffness matrix generated by the piecewise linear functions on the same grid. There is

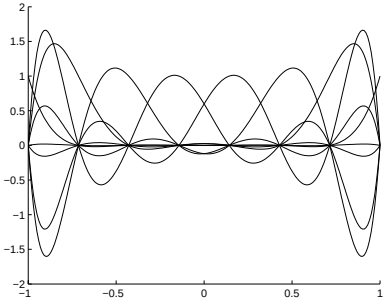

Fig. 7.1 Coordinate Lagrange interpolation polynomials of degree 7.

Table 7.1 Dependence of the condition number of $\mathbf{\Delta}^{-1}\mathbf{K}_p$ on p.

p	10	20	30	40	47
cond $[\mathbf{\Delta}^{-1}\mathbf{K}_p]$	$2.138 \cdot 10^5$	$9.703 \cdot 10^{10}$	$6.97 \cdot 10^{16}$	$2.43 \cdot 10^{21}$	$1.62 \cdot 10^{25}$

known the bound

$$m(p) := \max_{i,x} |\hat{p}^{(i)}(x)| \geq p^{-1}2^p\,, \quad i = 0, 1, \ldots, p\,, \quad x \in [0, 1], \tag{7.76}$$

for the Lebesgue function and the bound $\text{cond}\,[\mathbf{\Delta}^{-1}\mathbf{K}_p] \geq p^{-1}2^{2p}$ for the relative condition number, see, *e.g.,* [Ivanov and Korneev (1995)]. This bound is supported by numerical experiments, as can be seen from the Table 7.1, taken from the same paper. The condition number $\text{cond}\,[\mathbf{K}_p]$ grows even faster, especially in multi-dimensional case. Fig. 7.2 shows coordinate polynomials of the Lagrange interpolation of degrees $p = 6, 8, 10, 12, 14$ over the nodes $x^{(i)} = 2i/p - 1,\ i = 0, 1, \ldots, p$. Only coordinate polynomials for the node $x^{(p-2)} = 1 - 4/p$ are presented.

From estimate (7.76) and Table 7.1, we can obviously conclude that, for large p, Lagrange reference elements with the uniformly distributed nodes are not good for computations. Therefore, the right choice of the reference element is of crucial importance. In the modern computational practice, two families of square reference elements have became most popular due to their good properties, approved by computations. One family was introduced by [Szabo and Babuška (1987)]. It includes the elements with the coordinate functions defined by the tensor products of the integrated Legendre polynomials and termed *hierarchical*. Respectively the

corresponding reference elements are also termed *hierarchical*. An early description of another family, commonly termed as *spectral* elements, can be found in [Orszag (1980)]. This family comprises the Lagrange (nodal) reference elements with two types of distributions of the nodes induced by GLC (*Gauss-Lobatto-Chebyshev*) and GLL (*Gauss-Lobatto-Legendre*) quadrature rules. Hierarchical reference elements are more flexible in adaptive computations. This is for the reason that the orders of the internal and edge polynomials may be defined independently for each finite element interior and each edge of a finite element assemblage. The properties of spectral finite elements, important for fast computations, are well exposed in the book [Karniadakis and Sherwin (1999)]. A good choice of the coordinate polynomials for the triangular and tetrahedral reference p-elements is still under intensive research. For some advances in this direction, we refer to the papers by [Hesthaven (1998)] and [Beuchler and Schöberl (2006)].

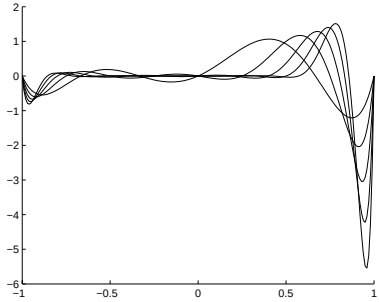

Fig. 7.2 Coordinate Lagrange interpolation polynomials of degrees $p = 6, 8, 10, 12, 14$ for the node at $x = 1 - 4/p$.

In this section, we describe typical reference elements together with the finite-difference/finite-element (FD/FE) preconditioners, which may be called *source preconditioners*. They deserved this term because they illuminated important properties of the p-element stiffness and mass matrices in a comprehensive and habitual form. In view of this, they served as a starting point for obtaining a number of other, slightly or more seriously modified preconditioners, which are better adapted to particular fast solvers of different types. Technically, efficient solvers for systems with the p-element stiffness matrices were created as solvers for the systems of finite-difference equations corresponding to the source FD/FE preconditioners. On this road, the DD-type, multigrid, iterative with BPX peconditioners and wavelet multiresolution solvers etc. have been developed.

It is worth noting that fast solvers for FD/FE preconditioners are interesting not only for p or hp discretizations, for which they were initially constructed, but also for first-order approximations to some deteriorating elliptic boundary value problems.

7.3.1 *Coordinate Polynomials of Square Hierarchical and Spectral Reference Elements*

Three major families of complete reference elements specified on the reference square $\tau_0 = (-1,1)\times(-1,1)$ are equipped with the trunk space $\mathcal{Q}_{p,x}$ of all polynomials of the order not greater $p\,,\ p \geq 1$, in each variable. They are the family of the hierarchical reference elements and two families of the spectral reference elements, denoted $\mathcal{E}_H$ and $\mathcal{E}_{Sp}$, respectively. For the bases in $\mathcal{Q}_{p,x}$, they use hierarchical and Lagrange interpolation polynomials, respectively, with the special choice of the nodes in the latter case.

Let us introduce the set $\mathcal{M}_{1,p} = \{\mathcal{L}_i(s)\,,\ i = 0,1,\ldots,p\}$ of polynomials of one variable defined by the formulas

$$\begin{aligned} &\mathcal{L}_0(s) \;=\; 0.5\,(1+s), \quad \mathcal{L}_1(s) = 0.5\,(1-s), \\ &\mathcal{L}_i(s) := \beta_i \int_{-1}^{s} P_{i-1}(t)\,dt = \gamma_i[P_i(s) - P_{i-2}(s)],\ i \geq 2, \end{aligned} \tag{7.77}$$

where P_i are Legendre's polynomials, and the numbers β_i and γ_i are given by the expressions $\beta_i = 0.5\sqrt{(2j-3)(2j-1)(2j+1)}$ and $\gamma_i = 0.5\sqrt{(2i-3)(2i+1)/(2i-1)}$, respectively. Thus, for $i \geq 2$, up to the multipliers β_i, $\mathcal{L}_i$ are the *integrated Legendre polynomials*. Legendre polynomials P_i for i=1,2,...,7 and integrated Legendre polynomials $\mathcal{L}_i$ for $i = 6, 7$ are shown in Fig. 7.3 and Fig. 7.4, respectively. The hierarchical basis in $\mathcal{Q}_{p,x}$ is given by the set

$$\mathcal{M}_p = \left\{L_{\boldsymbol{\alpha}}(x) = \mathcal{L}_{\alpha_1}(x_1)\mathcal{L}_{\alpha_2}(x_2)\ ,\ \boldsymbol{\alpha} \in \omega\right\},$$

where $\omega := \{\boldsymbol{\alpha} = (\alpha_1, \alpha_2) : 0 \leq \alpha_1, \alpha_2 \leq p\}$.

The multipliers β_i and γ_i provide the equality

$$\|\mathcal{L}_i(s)\|_{0,(-1,1)} = 1\,. \tag{7.78}$$

They were chosen in this way by [Ivanov and Korneev (1996)] in order to obtain simple FD-like preconditioners for the reference element stiffness and mass matrices.

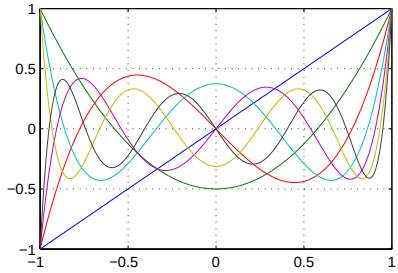

Fig. 7.3 Legendre polynomials P_i for i=1,2,...,7.

For the description of spectral elements $\mathcal{E}_{Sp}$, it is sufficient to specify the set of their nodes. For the interval [-1,1], GLL nodes are the zeroes η_k of the polynomial $(1-s^2)P_p'(s)$, *i.e.*

$$(1-\eta_k^2)P_p'(\eta_k) = 0\,, \quad k = 0, 1, \ldots, p\,. \tag{7.79}$$

The GLC nodes are explicitly defined as extremal points η_k of the Chebyshev polynomials

$$\eta_k = \cos(\frac{\pi(p-k)}{p}) = -\cos\frac{\pi k}{p}\,, \quad k = 0, 1, \ldots, p\,. \tag{7.80}$$

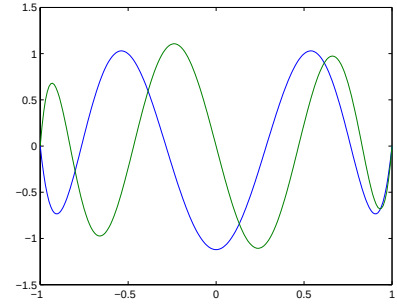

Fig. 7.4 Integrated Legendre polynomials $\mathcal{L}_i$ for $i = 6, 7$.

The two types of 2d meshes with the sets of nodes $x = \eta_{\boldsymbol{\alpha}} = (\eta_{\alpha_1}, \eta_{\alpha_2})$, $\boldsymbol{\alpha} \in \omega$, on the reference square τ_0 are the tensor products of the 1d GLL and GLC meshes on $(-1, 1)$. These meshes will be called *Gaussian or spectral meshes*. For the coordinate functions of spectral reference elements, we use the same notation $L_{\boldsymbol{\alpha}}$.

The coordinate functions $L_{\boldsymbol{\alpha}}$ for hierarchical and spectral elements are naturally subdivided in the subsets of the internal, boundary, edge and vertex coordinate functions. For the hierarchical element, these subsets correspond to the subsets $\omega_I = \{\boldsymbol{\alpha} : 2 \le \alpha_1, \alpha_2 \le p\}$, $\omega_B = \omega \setminus \omega_I$, $\omega_E = \omega_B \setminus \omega_V$ and $\omega_V = \{\boldsymbol{\alpha} : \alpha_k = 0, 1\}$, respectively. Obviously, for the spectral elements, we have $\omega_I = \{\boldsymbol{\alpha} : 1 \le \alpha_1, \alpha_2 \le p - 1\}$.

In the *hp*-adaptive computations, which are among the most efficient for problems with a lack of regularity, the orders of polynomials may vary from one element to another and, in each finite element, they may be different in different variables. In general, for each element, we may have a specific space $\mathcal{Q}_{\mathbf{p}_r}$, where multiindex $\mathbf{p}_r$ contains 6 entries, specifying the orders of polynomials in each direction inside τ_o and on each edge. Turning again to the hierarchical elements, in such a case we characterize the powers of polynomials by the vector $\mathbf{p}_r = (\mathbf{p}_I, \mathbf{p}_E)$, where $\mathbf{p}_I = (p_1, p_2)$ specifies the orders of the internal polynomials in each of the two variables and $\mathbf{p}_E = (p_{E_1}, \ldots, p_{E_4})$ specifies the orders of the polynomials on each edge E_k. We enumerate edges counter-clockwise with 1 corresponding to the edge on the line $x_2 \equiv -1$. Accordingly, the counterparts of the sets ω, ω_I, and ω_E become

$$\widetilde{\omega} = \widetilde{\omega}_I \cup \widetilde{\omega}_E \cup \omega_V\,, \quad \widetilde{\omega}_I = \{\boldsymbol{\alpha} : 2 \le \alpha_k \le p_{I,k}\}\,, \quad \widetilde{\omega}_E = \cup_{k=1}^{4} \widetilde{\omega}_{E_k}\,, \tag{7.81}$$

with $\widetilde{\omega}_{E_k} = \{\boldsymbol{\alpha} : \alpha_k = 2, \ldots, p_k,\, \alpha_{3-k} = 0\}$ and $\widetilde{\omega}_{E_{k+2}} = \{\boldsymbol{\alpha} : \alpha_k = 2, \ldots, p_{k+2},\, \alpha_{3-k} = 1\}$ for $k = 1, 2$, whereas for the reference element polynomial space, we get

$$\mathcal{Q}_{\mathbf{p}_r} = \mathrm{span}_{\widetilde{\omega}}\,[L_{\boldsymbol{\alpha}}(x)]\,. \tag{7.82}$$

For the element, which we refer to as *minimal incomplete* or *Serendipity element*, the polynomial space is

$$\mathcal{Q}_{\mathbf{p}_r} = \mathrm{span}_{\omega_{\min}}\,[L_{\boldsymbol{\alpha}}(x)]\,, \tag{7.83}$$

where $\omega_{\min} = \omega_{I,\min} \cup \omega_B$, with $\omega_{I,\min} = \{\boldsymbol{\alpha} : 2 \le \alpha_1, \alpha_2,\ \alpha_1 + \alpha_2 \le p)\}$.

We will call elements with the spaces, defined by (7.81) and (7.82), *adaptive*. As we will see later, the adaptive hierarchical elements can retain the properties necessary for obtaining fast solving procedures. The spectral elements are not so convenient for the adaptive computations, because, in general, the space $\mathcal{Q}_{\mathbf{p}_r}$ is not a tensor product space. Therefore, its polynomials can't be reproduced by means of one tensor product mesh. Only, orthotropic polynomial spaces, *i.e.*, $\mathcal{Q}_{\mathbf{p}_r}$ with $\mathbf{p}_I = (p_1, p_2)$ and $\mathbf{p}_E = (p_1, p_2, p_1, p_2)$, are an exception. Indeed, the coordinate polynomials of them can be introduced in a true spectral way.

7.3.2 *Stiffness and Mass Matrices of Hierarchical Elements*

The finite-difference Laplace operator on an orthogonal grid in a rectangle is represented by the sum of two Kronecker products of the unity matrix and the matrix representing the operator of the finite-difference second derivative. Due to the tensor product form of the hierarchical coordinate polynomials, the stiffness and mass matrices of the hierarchical reference element are conveniently represented in a similar form. The expansion

$$f(s) = \sum_{i=0}^{p} b_i \mathcal{L}_i(s)$$

and the quadratic forms $(f,f)_{(-1,1)}$ and $(f',f')_{(-1,1)}$ induce the matrices $\mathbb{K}_k,\ k=0,1$, satisfying the identities

$$(f,f)_{(-1,1)} = \mathbf{b}^\top \mathbb{K}_0 \mathbf{b} \quad \text{and} \quad (f',f')_{(-1,1)} = \mathbf{b}^\top \mathbb{K}_1 \mathbf{b}\,,$$

where $\mathbf{b} = \{b_i\}_{i=0}^p$ is the vector of the coefficients b_i. Let $\mathbf{A}$ and $\mathbb{M}$ be the stiffness and the mass matrices of the reference element, which are the Gram matrices induced by the basis $\mathcal{M}_p$ and the bilinear forms

$$a_{\tau_0}(v,u) = \int_{\tau_0} \nabla v \cdot \nabla u \, dx \quad \text{and} \quad (v,u)_{\tau_0} = \int_{\tau_0} v\, u \, dx\,.$$

Then we arrive at the representations

$$\mathbf{A} = \mathbb{K}_0 \otimes \mathbb{K}_1 + \mathbb{K}_1 \otimes \mathbb{K}_0 \quad \text{and} \quad \mathbb{M} = \mathbb{K}_0 \otimes \mathbb{K}_0\,. \tag{7.84}$$

In a view of importance of these matrices for constructing efficient preconditioner-solvers for the local Dirichlet problems in DD algorithms, we present them explicitly. By direct computation with the use of (7.77), we obtain

$$\mathbb{K}_1 = \operatorname{diag}\left[0, \sqrt{6}, \ldots, \frac{(2i-3)(2i+1)}{2}, \ldots, \frac{(2p-3)(2p+1)}{2}\right]. \tag{7.85}$$

Matrix $\mathbb{K}_0$ can be represented in the form $\mathbb{K}_0 = \boldsymbol{\mathcal{D}}^{-1}\mathbb{E}\,\boldsymbol{\mathcal{D}}^{-1}$ with the diagonal matrix $\boldsymbol{\mathcal{D}}$, which main diagonal is the square root of the main diagonal of the "mass" matrix $\mathbb{E}$. The latter is generated by the bilinear form $(\cdot,\cdot)_{(-1,1)}$ in the basis

$\{P_0, P_1, (P_2 - P_0), \ldots, (P_i - P_{i-2}), \ldots, (P_p - P_{p-2})\}$:

$$\mathbb{E} = \begin{pmatrix} 2 & 0 & -2 & & & & & & \\ 0 & \frac{2}{3} & 0 & -\frac{2}{3} & & & \mathbf{0} & & \\ -2 & 0 & (2+\frac{2}{5}) & 0 & -\frac{2}{5} & & & & \\ & \ldots & \ldots & \ldots & \ldots & & \ldots & & \\ & 0 & -\frac{2}{2i-3} & 0 & \frac{2}{2i-3}+\frac{2}{2i+1} & 0 & -\frac{2}{2i+1} & 0 & \ldots \\ & & & \ldots & \ldots & & \ldots & \ldots & \\ & \mathbf{0} & & \ldots & \ldots & & & & \\ & & & & & -\frac{2}{2p-3} & 0 & \frac{2}{2p-3} & \frac{2}{2p+1} \end{pmatrix}.$$

Consequently, for the 1d mass matrix $\bar{\mathbb{K}}_0$ generated in the basis $(\bar{P}_0, \bar{P}_1, \mathcal{L}_2, \ldots, \mathcal{L}_i, \ldots, \mathcal{L}_p))$, where $\bar{P}_k$ are the normalized Legendre polynomials, we have

$$\bar{\mathbb{K}}_0 = \begin{pmatrix} 1 & 0 & -\sqrt{5/6} & & & & & \\ & 1 & 0 & -\sqrt{7/10} & & & \mathbf{0} & \\ & & \ldots & \ldots & \ldots & \ldots & & \\ & & & 1 & 0 & -c_i & & \\ & & & & \ldots & \ldots & \ldots & \ldots \\ & & & & & & 1 & 0 & -c_{p-2} \\ & \mathbf{SYM} & & & & & 1 & 0 & \\ & & & & & & & 1 & \end{pmatrix} \tag{7.86}$$

with

$$c_i = \frac{1}{2}\left(\frac{(2i+3)(2i-5)}{(2i-3)(2i+1)}\right)^{1/2} = \frac{1}{2}\left(1 - \frac{12}{4i^2-4i-3}\right)^{1/2}. \tag{7.87}$$

Now, transforming it to the basis (7.77), we come to

$$\mathbb{K}_0 = \begin{pmatrix} 2/3 & 1/3 & -\sqrt{5/12} & \sqrt{7/60} & & & & \\ & 2/3 & -\sqrt{5/12} & -\sqrt{7/60} & & & \mathbf{0} & \\ & & \ldots & \ldots & \ldots & \ldots & & \\ & & & 1 & 0 & -c_i & & \\ & & & & \ldots & \ldots & \ldots & \ldots \\ & & & & & & 1 & 0 & -c_{p-2} \\ & \mathbf{SYM} & & & & & 1 & 0 & \\ & & & & & & & 1 & \end{pmatrix}. \tag{7.88}$$

The nonzero eigenvalues of $\bar{\mathbb{K}}_0$, $\mathbb{K}_1$ satisfy relations

$$\begin{gathered} \lambda_{\min}(\mathbb{K}_1) = \sqrt{6}, \quad \lambda_{\max}(\mathbb{K}_1) = (2p-3)(2p+1)/2\,, \\ \lambda_{\min}(\bar{\mathbb{K}}_0) \asymp 1/p^2, \quad \lambda_{\max}(\bar{\mathbb{K}}_0) \asymp 1\,. \end{gathered} \tag{7.89}$$

which are evident except that one for $\lambda_{\min}(\bar{\mathbb{K}}_0)$ that can be estimated as follows. For any $p < \infty$, according to (7.86) and (7.87), the matrix $\bar{\mathbb{K}}_0$ has a diagonal predominance, diminishing in the direction to the last row and deteriorating for $p \to \infty$. Indeed, simple calculations lead to the inequality

$$\frac{1}{2} - c_i \geq \frac{1}{2(i^2+1)}\,, \quad i \geq 4,$$

from which and from considering the first four rows of $\bar{\mathbb{K}}_0$, we have

$$1 - c_i - c_{i-2} \succ 1/(i^2+1)\,, \quad i = 0, 1, ..., 2N\,, \tag{7.90}$$

whence the bound for $\lambda_{\min}(\bar{\mathbb{K}}_0)$ follows.

In DD algorithms, the most important role is played by the blocks $\mathbf{A}_I$ and $\mathbb{M}_I$ of $\mathbf{A}$ and $\mathbb{M}$ related to the internal degrees of freedom. These blocks take even simpler block-diagonal forms, if we rearrange rows and columns in the order, which separates the basis polynomials in four groups each containing $L_{i,j}$ with the following indices:

$$\left.\begin{array}{ll} 1) & \text{all } i,j \text{ even}\,, \\ 2) & i \text{ even} \quad \text{and } j \text{ odd}\,, \\ 3) & i \text{ odd} \quad \text{and } j \text{ even}\,, \\ 4) & \text{all } i,j \text{ odd}\,. \end{array}\right\} \tag{7.91}$$

In view of (7.91), the space $\mathring{\mathcal{Q}}_{I,p}$ can be represented as the direct sum

$$\mathring{\mathcal{Q}}_{p,x} = \mathring{\mathcal{Q}}^{e,e}_{p,x} \oplus \mathring{\mathcal{Q}}^{e,o}_{p,x} \oplus \mathring{\mathcal{Q}}^{o,e}_{p,x} \oplus \mathring{\mathcal{Q}}^{o,o}_{p,x}\,, \tag{7.92}$$

and the ordering of d.o.f. according (7.91) splits the matrices $\mathbf{A}_I$ and $\mathbb{M}_I$ in four independent blocks. In the case $p = 2N+1$, that is, for simplicity, assumed in the following, they all have the same dimension N^2. In order to come to this representation, we first reorder the set of 1d coordinate polynomials to the form

$$\begin{aligned} \mathcal{M}_{1,p} = \{&\mathcal{L}_0(s), \mathcal{L}_2(s), \ldots, \mathcal{L}_{2k}(s), \ldots, \mathcal{L}_{2N}(s), \mathcal{L}_1(s), \mathcal{L}_3(s), \\ &\ldots, \mathcal{L}_{2k+1}(s), \ldots, \mathcal{L}_{2N+1}(s)\}\,, \end{aligned} \tag{7.93}$$

and, for $t = 0, 1$, we denote the blocks of the internal matrices $\mathbb{K}_{t,I}$ related to the coordinate polynomials of even and odd powers by $\mathbb{K}_{t,I,e}$ and $\mathbb{K}_{t,I,o}$, respectively. Now, we have

$$\mathbb{K}_{t,I} = \mathrm{diag}\,[\mathbb{K}_{t,I,e}, \mathbb{K}_{t,I,o}]\,,$$

where the matrices $\mathbb{K}_{1,I,a}$, with $a = e, o$, are obviously diagonal, whereas the matrices $\mathbb{K}_{0,I,a}$, are tridiagonal, *i.e.*

$$\begin{aligned} \mathbb{K}_{0,I,e} &= \mathrm{tridiag}\,[c_{2k-2}, 1, c_{2k}]_{k=1}^{N} \quad \text{and} \\ \mathbb{K}_{0,I,o} &= \mathrm{tridiag}\,[c_{2k-1}, 1, c_{2k+1}]_{k=1}^{N}\,. \end{aligned} \tag{7.94}$$

The stiffness and mass matrices take the forms

$$\begin{aligned}\mathbf{A}_I &= \mathrm{diag}\,[\mathbf{A}_{e,e}, \mathbf{A}_{o,e}, \mathbf{A}_{e,o}, \mathbf{A}_{o,o}] \quad \text{and} \\ \mathbb{M}_I &= \mathrm{diag}\,[\mathbb{M}_{e,e}, \mathbb{M}_{o,e}, \mathbb{M}_{e,o}, \mathbb{M}_{o,o}]\,, \end{aligned} \tag{7.95}$$

with the blocks

$$\mathbf{A}_{a,b} = \mathbb{K}_{1,I,a} \otimes \mathbb{K}_{0,I,b} + \mathbb{K}_{0,I,e} \otimes \mathbb{K}_{1,I,o} \quad \text{and} \quad \mathbb{M}_{a,b} = \mathbb{K}_{0,I,a} \otimes \mathbb{K}_{0,I,b}\,. \tag{7.96}$$

We remind that $\otimes$ denotes the Kronecker product.

7.4 Preconditioning of Stiffness and Mass Matrices by Finite-Difference Matrices

Apparently implementation of simple mesh operators for preconditioning finite element matrices was introduced by [Matsokin (1975)], who approved the use of the finite-difference Laplace operator on the uniform orthogonal grid as an efficient preconditioner for finite element $\mathcal{P}_1$-methods on a quasiuniform grid having special structure. In the spirit of the h-version, *i.e.*, with p fixed, the same type preconditioners are easily justified for higher order curvilinear Lagrange elements with the uniformly distributed nodes on the associated reference elements, see [Korneev (1977b,a)]. This immediately allowed to use a number of efficient iterative techniques for solution of FE equations, including fictitious component method and multigrid method. However, similarly to the simple example at the beginning of Section 7.3, at the uniform distribution of nodes on triangular or square reference p-elements, relative condition number retains the exponential growth with $p \to \infty$, see [Ivanov and Korneev (1995)]. FD/FE preconditioners, which are spectrally equivalent to the stiffness matrices of spectral and hierarchical reference p-elements uniformly in p, appeared in the literature in the 80s and 90s, respectively. It was observed that the $\mathcal{Q}_1$ and $\mathcal{P}_1$ finite element models on the rectangular and triangular grids with the nodes coinciding with the spectral reference element nodes can be efficiently used as preconditioners for spectral element stiffness matrices. Such an approach goes back to [Orszag (1980)] and [Deville and Mund (1990)]. However, the justification of the uniform in p spectral equivalence was produced much later. In relation to spectral elements with the GLL nodes, we refer to [Bernardi and Maday (1992a,b)] for the basic results in 1d, easily expandable to multidimensional reference elements, see, *e.g.*, [Canuto (1994)] and [Parter and Rothman (1995)]. For the $\mathcal{Q}_p$ hierarchical reference element with coordinate

functions, defined by tensor products of the integrated Legendre polynomials, the finite-difference preconditioner, spectrally equivalent to the stiffness matrix uniformly in p, was derived by [Ivanov and Korneev (1996)], see also [Korneev and Jensen (1997, 1999)]. It is worthwhile to note at once the principal difference between the finite-difference preconditioners for the stiffness matrices of spectral and hierarchical reference elements. In the former case, the internal stiffness matrix and its preconditioner result from two types of discretizations on the same grid of the same Dirichlet problem, governed by Laplace operator in the reference square. On the contrary, in the latter case the finite-difference preconditioner "approximates" deteriorating elliptic equation, which is quite different from Laplace operator.

7.4.1 *Preconditioners for Hierarchical Reference Elements*

For the blocks of the stiffness and mass matrices, related to internal, boundary, edge and vertex unknowns, we use the indexation similar to that one introduced for FE matrix $\mathbf{K}$, see (7.1). Hence, $\mathbf{A}_I$ and $\mathbb{M}_I$ are the blocks of $\mathbf{A}$ and $\mathbb{M}$ corresponding to internal coordinate polynomials. The preconditioners

$$\begin{aligned} \mathbf{\Lambda}_I &= \boldsymbol{\mathcal{D}} \otimes (\overline{\boldsymbol{\Delta}} + \boldsymbol{\mathcal{D}}^{-1}) + (\overline{\boldsymbol{\Delta}} + \boldsymbol{\mathcal{D}}^{-1}) \otimes \boldsymbol{\mathcal{D}} \quad \text{and} \\ \mathbf{\Xi}_I &= (\overline{\boldsymbol{\Delta}} + \boldsymbol{\mathcal{D}}^{-1}) \otimes (\overline{\boldsymbol{\Delta}} + \boldsymbol{\mathcal{D}}^{-1}) \,, \end{aligned} \tag{7.97}$$

that are spectrally equivalent to the internal stiffness and mass matrices $\mathbf{A}_I$ and $\mathbb{M}_I$, respectively, are completely defined by the following simple $(p-1) \times (p-1)$ matrices

$$\boldsymbol{\mathcal{D}} = \operatorname{diag}\left[i^2\right]_{i=1}^{p-1} \quad \text{and} \quad \overline{\boldsymbol{\Delta}} = \frac{1}{2} \begin{pmatrix} 2 & 0 & -1 & & & & \\ & 2 & & 0 & -1 & & \\ & & \cdots & \cdots & & & \\ & & & \cdots & \cdots & & \\ & \text{SYM} & & & 2 & 0 & -1 \\ & & & & & 2 & 0 \\ & & & & & & 2 \end{pmatrix} .$$

The stiffness and mass matrices as well as their preconditioners take even a simpler block-diagonal form, if we rearrange the rows and columns according to (7.91). Similarly to (7.95), this splits each of the matrices $\mathbf{\Lambda}_1$ and $\mathbf{\Xi}_I$ into four independent blocks, namely

$$\begin{aligned} \mathbf{\Lambda}_I &= \operatorname{diag}\left[\mathbf{\Lambda}_{e,e}, \mathbf{\Lambda}_{e,o}, \mathbf{\Lambda}_{o,e}, \mathbf{\Lambda}_{o,o}\right] \quad \text{and} \\ \mathbf{\Xi}_I &= \operatorname{diag}\left[\mathbf{\Xi}_{e,e}, \mathbf{\Xi}_{e,o}, \mathbf{\Xi}_{o,e}, \mathbf{\Xi}_{o,o}\right] . \end{aligned} \tag{7.98}$$

In turn, the $N^2 \times N^2$ blocks entering these matrices are represented by the expressions

$$\begin{aligned} \mathbf{\Lambda}_{a,b} &= \boldsymbol{\mathcal{D}}_a \otimes (\mathbf{\Delta} + \boldsymbol{\mathcal{D}}_b^{-1}) + (\mathbf{\Delta} + \boldsymbol{\mathcal{D}}_a^{-1}) \otimes \boldsymbol{\mathcal{D}}_b \quad \text{and} \\ \mathbf{\Xi}_{a,b} &= (\mathbf{\Delta} + \boldsymbol{\mathcal{D}}_a^{-1}) \otimes (\mathbf{\Delta} + \boldsymbol{\mathcal{D}}_b^{-1})\,, \end{aligned} \tag{7.99}$$

where $a, b = e, o$ and $\boldsymbol{\mathcal{D}}_e = \operatorname{diag}[4i^2]_{i=1}^N$, $\boldsymbol{\mathcal{D}}_o = \operatorname{diag}[4i^2 + 4i + 1]_{i=1}^N$,

$$\mathbf{\Delta} = \frac{1}{2} \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & 0 & \\ & \dots & \dots & & & \\ & 0 & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}. \tag{7.100}$$

The blocks of the matrix $\mathbf{\Lambda}_I$ are spectrally equivalent each to another. The same is true for the blocks of $\mathbf{\Xi}_I$. Therefore, for the matrices $\mathbf{\Lambda}_I^{(a)} = \operatorname{diag}[\mathbf{\Lambda}_{a,a}, \mathbf{\Lambda}_{a,a}, \mathbf{\Lambda}_{a,a}, \mathbf{\Lambda}_{a,a}]$ and $\mathbf{\Xi}_I^{(a)} = \operatorname{diag}[\mathbf{\Xi}_{a,a}, \mathbf{\Xi}_{a,a}, \mathbf{\Xi}_{a,a}, \mathbf{\Xi}_{a,a}]$, we have the spectral equivalence inequalities

$$\mathbf{\Lambda}_I^{(a)} \prec \mathbf{\Lambda}_I \prec \mathbf{\Lambda}_I^{(a)} \quad \text{and} \quad \mathbf{\Xi}_I^{(a)} \prec \mathbf{\Xi}_I \prec \mathbf{\Xi}_I^{(a)}$$

for $a = e, o$. Thus, if an iterative solver, say, for $\mathbf{\Lambda}_{e,e}$ is available, then similar solvers will be applicable to the other blocks of $\mathbf{\Lambda}_I$ with the same efficiency. Alternatively, we may use the preconditioner $\mathbf{\Lambda}_I^{(a)}$ with the same solver for all four blocks. For these reasons, it is sufficient to derive a fast solver for the block $\mathbf{\Lambda}_{e,e}$ only.

Further simplifications are possible without loss of the spectral equivalence (uniform in p). Namely, we can replace preconditioners $\mathbf{\Lambda}_{a,b}$ and $\mathbf{\Xi}_{a,b}$ from (7.99) by

$$\mathbf{\Lambda} = \boldsymbol{\mathcal{D}}_e \otimes \mathbf{\Delta} + \mathbf{\Delta} \otimes \boldsymbol{\mathcal{D}}_e \quad \text{and} \quad \mathbf{\Xi} = \mathbf{\Delta} \otimes \mathbf{\Delta}\,, \tag{7.101}$$

omitting the matrices $\boldsymbol{\mathcal{D}}_a^{-1}$ and $\boldsymbol{\mathcal{D}}_b^{-1}$ in brackets, and substituting, *e.g.,* $\boldsymbol{\mathcal{D}}_e$ for the rest diagonal blocks. In this way, we come to the preconditioners

$$\mathbf{\Lambda}_\triangle = \operatorname{diag}[\mathbf{\Lambda}, \mathbf{\Lambda}, \mathbf{\Lambda}, \mathbf{\Lambda}] \quad \text{and} \quad \mathbf{\Xi}_\triangle = \operatorname{diag}[\mathbf{\Xi}, \mathbf{\Xi}, \mathbf{\Xi}, \mathbf{\Xi}]\,. \tag{7.102}$$

Theorem 7.3. *The stiffness matrix $\mathbf{A}_I$ of the hierarchical reference element and its preconditioners $\mathbf{\Lambda}_I$, $\mathbf{\Lambda}_I^{(a)}$, $\mathbf{\Lambda}_\triangle$ are spectrally equivalent uniformly in p. The same is true for the mass matrix $\mathbb{M}_I$ and its preconditioners $\mathbf{\Xi}_I$, $\mathbf{\Xi}_I^{(a)}$, $\mathbf{\Xi}_\triangle$.*

Proof. Simple proofs of the spectral equivalence inequalities for the stiffness matrix $\mathbf{A}_I$ and its preconditioners $\mathbf{\Lambda}_I$, $\mathbf{\Lambda}_I^{(a)}$, as well as for the mass matrix $\mathbb{M}_I$ and its preconditioners $\mathbf{\Xi}_I$, $\mathbf{\Xi}_I^{(a)}$, can be found in [Ivanov and

Korneev (1996)], see also Lemma 4.2 and Remark 5.1 in [Korneev and Jensen (1997)]. For instance, for $\mathbf{A}_I$ and $\mathbf{\Lambda}_I$, the proof follows by the direct comparison of the 1d matrices $\mathbb{K}_0$ and $\mathbb{K}_1$ with the matrices $(\boldsymbol{\Delta}+\boldsymbol{\mathcal{D}}^{-1})$ and $\boldsymbol{\mathcal{D}}$, entering expressions (7.84) and (7.97), respectively, and by taking into account the diagonal predominance inequality (7.90). The uniform in p spectral equivalences $\mathbf{A}_I \asymp \mathbf{\Lambda}_\triangle$ and $\mathbf{M}_I \asymp \mathbf{\Xi}_\triangle$ are a direct consequence of Lemma 7.5 below, which was proved in [Korneev *et al.* (2003b)]. □

Lemma 7.5. *The spectral equivalence inequalities*

$$\boldsymbol{\Delta} \prec \boldsymbol{\Delta}_a = \boldsymbol{\Delta} + \boldsymbol{\mathcal{D}}_a^{-1} \prec \boldsymbol{\Delta} \tag{7.103}$$

are valid for $a = e$ and $a = o$.

Proof. Clearly, the left inequality is true. The right inequality is obtained from Hardy's inequality, if we take into account that the involved matrices are spectrally equivalent to the finite element matrices generated by the piecewise linear functions. Indeed, according Hardy's inequality, see, *e.g.*, [Maz'ya and Poborchi (1998)], we have

$$\|x^{-s}\mathcal{F}\|_{L_k(0,\infty)} \le \frac{1}{|s-\frac{1}{k}|}\,\|x^{1-s}f\|_{L_k(0,\infty)}\,,$$

where $\mathcal{F}$ is defined by the formula

$$\mathcal{F}(x) = \begin{cases} \int_0^x f(t)\,, & s > 1/k\,, \\ \\ \int_x^\infty f(t)\,, & s < 1/k\,, \end{cases}$$

with $k = 1, 2, ...\infty$ and $s \neq k$. We set $k = 2$, $s = 1$ and $\mathcal{F} \in \mathring{H}^1(0,1)$ such that $f = \mathcal{F}'$, continue the both functions outside the segment [0,1] by zero, and come to the inequality

$$\|x^{-1}v\|^{\,2}_{0,(0,1)} \le 4\,|\,v\,|^2_{1,(0,1)}\,, \quad \forall\, v \in \mathring{H}^1(0,1)\,. \tag{7.104}$$

On the interval $(0,1)$, we introduce the uniform mesh of the size $\hbar = (N+1)^{-1}$ and the space $\mathring{\mathcal{V}}(0,1)$ of the continuous piecewise linear functions vanishing at the ends of the interval. Let $\phi_i \in \mathring{\mathcal{V}}(0,1)$ be the nodal basis function for the node x_i, $i = 1, 2, \ldots, N$, *i.e.,* $\phi_i(x_i) = \delta_{i,j}$ and $\mathring{\mathcal{V}}(0,1) = \operatorname{span}[\phi_i]_{i=1}^N$. Since (7.104) holds for piecewise linear functions, we have

$$\boldsymbol{\mathcal{M}}_{\rm fem} \le\ \boldsymbol{\Delta}_{\rm fem}\,, \tag{7.105}$$

where $\boldsymbol{\Delta}_{\rm fem} = \{\langle(\phi_i)', (\phi_j)'\rangle_{\omega=1}\}_{i,j=1}^N$, $\boldsymbol{\mathcal{M}}_{\rm fem} = \{\langle\phi_i, \phi_j\rangle_{\omega=x^{-1}}\}_{i,j=1}^N$, and

$$\langle u, v\rangle_\omega = \int_0^1 \omega^2 u(x)v(x)\,dx\,.$$

But, evidently, $\hbar\boldsymbol{\Delta}_{\rm fem} = \boldsymbol{\Delta}$, and it is easy to find out that the matrices $\boldsymbol{\mathcal{D}}_e$ and $\hbar\boldsymbol{\mathcal{M}}_{\rm fem}$ are spectrally equivalent uniformly in p. In order to prove the latter, we construct the two piecewise constant functions $\hat{\varphi}$ and $\check{\varphi}$ as follows:

$$\hat{\varphi} = 0.5\hbar \;\text{ for }\; x \in (0,\hbar)\,; \quad \hat{\varphi} = i\hbar \;\text{ for }\; x \in ((i-1)\hbar, i\hbar)\,,\; i \geq 1\,;$$
$$\check{\varphi} = 0.5\hbar \;\text{ for }\; x \in (0,\hbar)\,; \quad \check{\varphi} = (i-1)\hbar \;\text{ for }\; x \in ((i-1)\hbar, i\hbar)\,,\; i \geq 1\,.$$

Clearly, they satisfy the inequalities $\check{\varphi} \leq x^2 \leq \hat{\varphi}$ for $\hbar \leq x \leq 1$. Then, the spectral equivalence is proved by the comparison of the matrices $\{\langle\phi_i,\phi_j\rangle_{\omega=\check{\varphi}}\}_{i,j=1}^N$ and $\{\langle\phi_i,\phi_j\rangle_{\omega=\hat{\varphi}}\}_{i,j=1}^N$ with $\boldsymbol{\mathcal{D}}_e$ and $\hbar\boldsymbol{\mathcal{M}}_{\rm fem}$, respectively. The consequence of the shown spectral equivalence and (7.105) is the inequality

$$\boldsymbol{\mathcal{D}}_a^{-1} \leq \boldsymbol{\Delta}\,, \tag{7.106}$$

approving the right inequality in (7.103). □

Theorem 7.3 and the generalized conditions of angular quasiuniformity allow us to formulate the following conclusion.

Corollary 7.2. *Let $\mathbf{K}_{I_r}$ and $\mathbf{M}_{I_r}$ be the finite element stiffness and mass matrices arising from the finite element dicretization of the boundary value problem given in Example 2.2 by means of a shape regular finite element assemblage and $\tau_r = \Omega_r$, $r = 1,2,\ldots,\mathcal{R} = J$. Further, let the preconditioner $\boldsymbol{\mathcal{K}}_{I_r}$ for $\mathbf{K}_{I_r}$ be one of the matrices $\boldsymbol{\mathcal{K}}_{I_r} = \varrho_r\boldsymbol{\Lambda}_I$, $\varrho_r\boldsymbol{\Lambda}_I^{(a)}$, $\varrho_r\boldsymbol{\Lambda}_\triangle$, whereas the preconditioner $\boldsymbol{\mathcal{M}}_{I_r}$ for $\mathbf{M}_{I_r}$ be one of the matrices $\boldsymbol{\mathcal{M}}_{I_r} = (h^{(r)})^2\,\boldsymbol{\Xi}_I$, $(h^{(r)})^2\,\boldsymbol{\Xi}_I^{(a)}$, $(h^{(r)})^2\,\boldsymbol{\Xi}_\triangle$. Here $h^{(r)}$ is the characteristic size of the finite element, see (3.4). Then $\mathbf{K}_{I_r}$ and $\boldsymbol{\mathcal{K}}_{I_r}$, as well as $\mathbf{M}_{I_r}$ and $\boldsymbol{\mathcal{M}}_{I_r}$, are spectrally equivalent uniformly with respect to p, r and $\mathcal{R}$. The constants in the corresponding spectral equivalence inequalities depend only on the constants in the generalized conditions of the angular regularity for the mappings $\mathcal{X}^{(r)}$.*

The use of the introduced preconditioners simplifies solution of the systems with the reference element and finite element internal stiffness matrices $\mathbf{A}_I$ and $\mathbf{K}_{I_r}$, respectively. This is obviously true in respect to the matrices $\mathbf{K}_{I_r}$, that may be completely filled in, if, for instance, coefficients of the elliptic problem are variable. First of all, these preconditioners in four times reduce dimensions of the systems to be solved at each iteration, so that, *e.g.,* for $p = 21$, instead of the system with the matrix 400×400, that may be completely filled in, it is necessary to solve 4 systems with the matrices 100×100 having the five-point stencils. Due to the finite-difference

interpretation, it becomes immediately possible to point out Gaussian elimination procedure with the asymptotically optimal ordering, one of which is the nested dissection of [George and Liu (1981)]. However, designing optimal and almost optimal in p solvers for these preconditioners still remains a challenge. This becomes clear when we look at their differential counterparts, which are the deteriorating elliptic equations. We illustrate these partial differential equations below.

Introducing the notations $\hbar = 1/(N+1)$, $\eta_i = i\hbar$, $\varphi_i = \eta_i^2$, we see that equations of the system

$$\mathbf{\Lambda}_{e,e}\mathbf{u} = \mathbf{f} \ , \quad \mathbf{u} = \{u_{i,j}\} \ , \quad \mathbf{f} = \{f_{i,j}\} \ , \tag{7.107}$$

may be written in the form

$$-2\left(\varphi_i \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j-1}}{\hbar^2} + \varphi_j \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{\hbar^2}\right) + \varphi_{i,j} u_{i,j} = f_{i,j}, \tag{7.108}$$

where $\varphi_{i,j} = (\varphi_i/\varphi_j) + (\varphi_j/\varphi_i)$, $1 \le i,j \le N$, and $u_{i,j} = 0$, if one of the indices is out of the pointed out range. Formally, this looks exactly as the finite-difference approximation on the uniform square mesh of the size $\hbar$ of the elliptic boundary value problem

$$\begin{aligned} \mathcal{L}u \equiv -2\left(\xi_1^2 \frac{\partial^2 u}{\partial \xi_2^2} + \xi_2^2 \frac{\partial^2 u}{\partial \xi_1^2}\right) + \left(\frac{\xi_1^2}{\xi_2^2} + \frac{\xi_2^2}{\xi_1^2}\right) u = f \ \text{ in } \pi_1 \\ u = 0 \ \text{ on } \partial\pi_1, \end{aligned} \tag{7.109}$$

where $\pi_1 = (\xi = (\xi_1, \xi_2) : 0 < \xi_1, \xi_2 < 1)$ denotes the unit square. The coefficients of the equation at the main terms are deteriorating at the edges $\xi_1 = 0$ and $\xi_2 = 0$, and, at the same time, the coefficient before the minor term becomes unbounded at the same edges. It is also necessary to emphasize that the values $u_{i,j}$ do not represent nodal values of the unknown function of the source problem, for solving which the preconditioner is used. The mesh and the domain in (7.108) and (7.108) are introduced purely artificially, and each node corresponds to one coordinate polynomial $L_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} = (2i, 2j)$, as it is pictured in Fig. 7.5 for $N = 7$. This is in contrast to the finite-difference preconditioner for the spectral elements, see Subsection 7.4.3.

If we choose the preconditioner $\Lambda_\triangle$, then we have to solve the simpler looking system

$$\mathbf{\Lambda}\mathbf{u} = \mathbf{f} \ , \quad \mathbf{u} = \{u_{i,j}\} \ , \quad \mathbf{f} = \{f_{i,j}\} \ , \tag{7.110}$$

with equations of the form

$$-2\left(\varphi_i \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j-1}}{\hbar^2} + \varphi_j \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{\hbar^2}\right) = f_{i,j}, \tag{7.111}$$

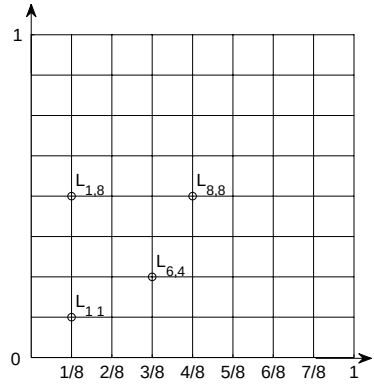

Fig. 7.5 Discretization for the preconditioner $\mathbf{\Lambda}_{e,e}$.

for $i, j = 1, 2, \ldots, N$, and the same boundary condition. Formally, this is a finite-difference approximation to the elliptic equation

$$\mathcal{L}u \equiv -2\left(\xi_1^2 \frac{\partial^2 u}{\partial \xi_2^2} + \xi_2^2 \frac{\partial^2 u}{\partial \xi_1^2}\right) = f \ \text{ in } \pi_1 \quad \text{and} \quad u = 0 \ \text{ on } \partial\pi_1. \tag{7.112}$$

Partial differential equations (7.109) and (7.112), have deteriorating coefficients expressed by the functions ξ_k^2. Clearly, for a given finite-difference operator, the choice of the corresponding differential operator is not unique. In particular, in (7.109) and (7.112), we could replace functions ξ_k^2, *e.g.,* by the functions $\varphi(\xi_k)$ defined by the formulas

$$\varphi(\zeta) = \begin{cases} \frac{1}{4}\hbar^2\,, & \zeta \in (0, \hbar)\,, \\ \zeta^2\,, & \zeta \in (\hbar, 1)\,. \end{cases} \tag{7.113}$$

Systems (7.107) and (7.110) are also the finite-difference approximations of the modified according to (7.113) differential equations (7.109) and (7.112). This observation is sometimes used in the following.

7.4.2 *One-Dimensional Spectral Elements*

Let $\mathbb{K}_{0,\mathrm{sp}}$ and $\mathbb{K}_{1,\mathrm{sp}}$ be the mass and stiffness matrices of a 1d spectral reference element $\mathcal{E}_{Sp}$. Preconditioners for these matrices are usually constructed in a more straightforward way. In an understandable sense, they are *true* finite-difference or finite element preconditioners. This is because they are derived as finite-difference or finite element approximations of the

same differential operator or of the corresponding bilinear form, respectively, and on the same grid which are associated with the spectral reference element $\mathcal{E}_{Sp}$. For instance, they can be FE matrices which are induced for the Neumann problem in τ_0 by the FE assemblages of $\mathcal{P}_1/\mathcal{Q}_1$ elements on the triangular/rectangular meshes, respectively, with the nodes at the nodes of a spectral reference p-element. Alternatively, they can be the finite-difference Laplacians, for instance, with the $(2d+1)$-points stencils, which, as it is well known, are spectrally equivalent to the FE preconditioners up to the multiplier in a form of a diagonal matrix.

Let $\mathcal{V}^p_{GL}(I)$ be the space of continuous piecewise linear functions related to the set of $p+1$ Gauss-Lobatto quadrature nodes. Under the assumption that $\mathbf{B}$ is a square matrix, the notation $\lfloor \mathbf{B} \rfloor_{\rm diag}$ will stand for the diagonal matrix with the same main diagonal as in $\mathbf{B}$. It is known that the mass matrix $\mathbb{M}$, induced by the bilinear form $(\cdot,\cdot)_I$ on $\mathcal{V}^p_{GL}(I)$, is spectrally equivalent to the diagonal matrix $\mathbb{K}_0 = \lfloor \mathbb{M} \rfloor_{\rm diag}$, which is given by

$$\mathbb{K}_0 = \mathrm{diag}\,[\widetilde{h}_i = \frac{1}{2}(\hbar_i + \hbar_{i+1})]_{i=0}^{p+1}\,, \tag{7.114}$$

with $\hbar_i = 0$ for $i = 0, p+1$. By $\mathbb{K}_1$, we denote the FE stiffness matrix, *i.e.,* the matrix corresponding to the bilinear form $(v', w')_I$ on the same space $\mathcal{V}^p_{GL}(I)$ with the nodal basis. In the case of 1d GLL grids, the equivalences underlying efficiency of such preconditioners are written as

$$\mathbb{K}_{0,\rm sp} \asymp \mathbb{K}_0 \quad \text{and} \quad \mathbb{K}_{1,\rm sp} \asymp \mathbb{K}_1 \tag{7.115}$$

with an easy extension to the multidimensional reference elements.

Since their appearance, FD/FE preconditioners for the spectral reference elements have successfully been used in many applications, and some advances have been made in a priory analysis of their efficiency. In this respect, we refer to the contributions by [Canuto and Quarteroni (1982)], [Deville and Mund (1985, 1990)], [Canuto *et al.* (1988)], [Quarteroni and Vali (1999)], [Karniadakis and Sherwin (1999)], and [Deville *et al.* (2002)]. It is also necessary to note that the preceding studies of properties of the Gauss and Gauss-Lobatto quadratures and orthogonal polynomials were essential for advances in this area. We refer to the book [Szego (1975)] for more information about these results

A breakthrough in the theoretical justification of the FD/FE preconditioners is due to [Bernardi and Maday (1992a, 1997)] who analyzed the $L_{2,\alpha}$- and H^1_α-stability of the Lagrange polynomial interpolations over the sets of nodes of the Gauss and Gauss-Lobatto quadratures. Here $L_{2,\alpha}(I)$ and $H^1_\alpha(I)$ are the weighted Sobolev spaces with the weight $(1-x^2)^\alpha$, *cf.*

the definitions given below. In particular, for the polynomial interpolations over the GLL nodes these authors proved the inequalities

$$\mathbb{K}_{0,\mathrm{sp}} \prec \mathbb{K}_0 \quad \text{and} \quad \mathbb{K}_{1,\mathrm{sp}} \prec \mathbb{K}_1 \,.$$

The transition from (7.115) to the respective equivalences for multidimensional spectral finite elements can be found in the papers of [Canuto (1994)] and [Casarin (1997)].

The results on the stability of the Lagrange polynomial interpolations over the sets of nodes of the Gauss and Gauss-Lobatto quadratures created a basement for subsequent advances in the area of preconditioning the FE matrices generated by the spectral element discretizations. In this section, we touch only a small part of them, directly related to the preconditioning. For more detailed presentations of properties of orthogonal polynomials, Gauss-Jacobi and Gauss-Lobatto-Jacobi quadratures, and the related interpolations, we refer the reader to [Szego (1975)], [Bernardi and Maday (1997)] and [Guo and Wang (2004)].

The following lemma shows that the step sizes $\hbar_k := \eta_k - \eta_{k-1}$ of the GLL and GLC grids (7.79) and (7.80) have the same asymptotic behaviour with respect to the power p.

Lemma 7.6. *The asymptote of the step sizes of the GLL and GLC grids is expressed by the relationship*

$$\hbar_k \asymp k/p^2 \,, \quad k \le N \,. \tag{7.116}$$

Proof. In the case of GLC grids, (7.116) can be proved easily. Indeed, according to (7.80), we have

$$\begin{aligned}\hbar_k &= -\frac{1}{2}\sin(\frac{2p-2k+1}{2p}\pi)\sin-\frac{\pi}{2p} = \frac{1}{2}[\sin(\frac{p-k}{p}+\frac{1}{2p})\pi]\sin\frac{\pi}{2p} \\ &= \frac{1}{2}[\sin(\frac{k}{p}-\frac{1}{2p})\pi]\sin\frac{\pi}{2p}\,,\end{aligned}$$

and since $(2/\pi)x \le \sin x \le x$ for all $x \in [0, \pi/2]$, we get

$$\frac{1}{p^2}(k-\frac{1}{2}) \le \hbar_k \le \frac{\pi^2}{4p^2}(k-\frac{1}{2})\,.$$

The proof of (7.116) for the GLL nodes is more involved, since there is no explicit expressions for the coordinates of the nodes. However, it can be completed with the use of the results, related to the Gauss and Gauss-Lobatto types guadratures for the integrals with the measure $(1-\xi^2)^\alpha d\xi$, $\alpha > -1$, called also Gauss-Jacobi and Gauss-Lobatto-Jacobi

quadratures. Whereas Gauss formula for the interval $I = (-1, 1)$ has only internal nodes, Gauss-Lobatto formula has the nodes at the ends of the interval I as well.

Let Γ be Euler's Gamma function and J_p^α be the Jacobi polynomial of degree $p \geq 0$ that satisfy the identity

$$\int_I J_m^\alpha(\xi) J_n^\alpha(\xi)(1-\xi^2)^\alpha d\xi = \delta_{m,n} \frac{2^{2\alpha+1}}{2n+2\alpha+1} \frac{\Gamma^2(n+\alpha+1)}{\Gamma(n+2\alpha+1)}\,. \quad (7.117)$$

The $p-1$ internal nodes of Gauss and Gauss-Lobatto formulas, here denoted by $\zeta_i^{\alpha,p-1}$ and $\xi_i^{\alpha,p}$, $i = 1, 2, \ldots, p-1$, respectively, are defined as roots of the polynomials J_{p-1}^α and $dJ_p^\alpha/d\xi$, *i.e.*

$$J_{p-1}^\alpha(\zeta_i^{\alpha,p-1}) = 0 \quad \text{and} \quad \frac{dJ_p^\alpha}{d\xi}(\xi_i^{\alpha,p}) = 0$$

for $i = 1, 2, \ldots, p-1$. Gauss and Gauss-Lobatto quadrature formulas, which are exact for the polynomials $\phi \in \mathcal{P}_{2p-1}$, have the form

$$\int_{-1}^{1} (1-\xi^2)^\alpha \phi(\xi)\, d\xi \simeq \sum_{i=1}^{p} \phi(\zeta_i^{\alpha,p}) \omega_i^{\alpha,p} \quad (7.118)$$

and

$$\int_{-1}^{1} (1-\xi^2)^\alpha \phi(\xi)\, d\xi \simeq \sum_{i=1}^{p-1} \phi(\xi_i^{\alpha,p}) \rho_i^{\alpha,p} + \phi(-1)\rho_0^{\alpha,p} + \phi(-1)\rho_p^{\alpha,p}\,, \quad (7.119)$$

respectively, where $\omega_i^{\alpha,p}$, $\rho_i^{\alpha,p}$ are their weights. GLL and GLC quadratures are particular cases of Gauss-Lobatto formula for $\alpha = 0$ and $\alpha = -1/2$, respectively.

Jacobi polynomials satisfy the differential equation

$$\big((1-\xi^2)^{\alpha+1} J_n^{\alpha\,\prime}\big)' + n(n+2\alpha+1)(1-\xi^2)^\alpha J_n^\alpha = 0\,.$$

In view of this equation and (7.117), we conclude that the derivatives $J_n^{\alpha\,\prime}$ are orthogonal with respect to the weight $(1-\xi^2)^{\alpha+1}$. In turn, this leads to the conclusion that the interior nodes of the Gauss-Lobatto formula with $n+2$ nodes and the weight $(1-\xi^2)^{\alpha+1}$ coincide with the nodes of Gauss formula with n nodes and the the weight $(1-\xi^2)^\alpha$. In other words, we have

$$\xi_k^{\alpha,n+1} = \zeta_k^{\alpha+1,n}\,, \quad k = 1, 2, \ldots, n\,, \quad (7.120)$$

whereas the weights satisfy the equality

$$\rho_k^{\alpha,n+1} = (1-(\zeta_k^{\alpha+1,n})^2)^{-1} \omega_k^{\alpha+1,n}\,, \quad k = 1, 2, \ldots, n\,. \quad (7.121)$$

The relations (7.120) and (7.121) are a consequence of the following fact. If $\phi(\xi) = (1-\xi^2)\psi(\xi)$, then (7.119) takes the form

$$\int_{-1}^{1}(1-\xi^2)^{\alpha}\phi(\xi)\,d\xi = \int_{-1}^{1}(1-\xi^2)^{\alpha+1}\psi(\xi)\,d\xi \simeq \sum_{i=1}^{p-1}[1-(\xi_i^{\alpha,p})^2]\psi(\xi_i^{\alpha,p})\rho_i^{\alpha,p}\,.$$

Since the Gauss-Lobatto quadrature formula is exact for $\phi \in \mathcal{P}_{2p-1}$, the formula in the right part is exact for all $\psi \in \mathcal{P}_{2p-3}$. Therefore, it is the Gauss formula (7.118) with α and p replaced by $\alpha+1$ and $p-1$, *i.e.*,

$$\sum_{i=1}^{p-1}[1-(\xi_i^{\alpha,p})^2]\psi(\xi_i^{\alpha,p})\rho_i^{\alpha,p} = \sum_{i=1}^{p-1}\psi(\zeta_i^{\alpha+1,p-1})\omega_i^{\alpha+1,p-1}\,, \quad \forall\,\psi \in \mathcal{P}_{2p-3}\,,$$

from the uniqueness of which we come to (7.120) and (7.121).

One important property of the Gauss quadrature nodes, that is described by the inequalities

$$\frac{\partial\zeta_k^{\alpha,n}}{\partial\alpha} < 0\,, \quad k = 1,2,\ldots,\lfloor\frac{n}{2}\rfloor\,, \tag{7.122}$$

was established by Szego and can be found in Theorem 6.21.31 of [Szego (1975)]. For a given α, the Gauss-Lobatto nodes are between the Gauss nodes by their definitions. Using this property, (7.120) and (7.122), we get the inequalities

$$\theta_{k-1}^{\alpha-1,n} < \theta_{k-1}^{\alpha,n} < \theta_k^{\alpha-1,n+1} \tag{7.123}$$

for the values $\theta_k^{\alpha,n} = \arccos(-\zeta_k^{\alpha,n})\,,\ k = 1,2,\ldots,n$.

The properties of the Gauss and Gauss-Lobatto quadrature nodes described above allow us to characterize their positions quantitatively. The corresponding inequalities, that were proved by [Bernardi and Maday (1992a)], are given in Lemma 7.7 below. By means of these inequalities, the proof of (7.116) for the GLL grids proceeds in the same way as the proof for the GLC grids. This completes the proof of Lemma 7.6. □

Lemma 7.7. *For $\alpha \geq -0.5$ and $\lambda = \lfloor 0.5-\alpha\rfloor$, we have*

$$(k-1)\frac{\pi}{n+1} \leq \theta_k^{\alpha,n} \leq (k-\lambda)\frac{\pi}{n+1}\,, \quad k = 1,2,\ldots,\lfloor\frac{n}{2}\rfloor\,, \tag{7.124}$$

and, for $-1 < \alpha < -0.5$,

$$(k-2)\frac{\pi}{n+1} \leq \theta_k^{\alpha,n} \leq k\frac{\pi}{n+1}\,, \quad k = 1,2,\ldots,\lfloor\frac{n}{2}\rfloor\,. \tag{7.125}$$

Proof. By means of (7.120), (7.122) and (7.123), these inequalities are obtained recursively, see Corollary 2.6 in [Bernardi and Maday (1992a)]. □

For $-\frac{1}{2} \le \alpha \le \frac{1}{2}$, the inequalities

$$(k-\frac{1}{2})\frac{\pi}{n} \le \theta_k^{\alpha,n} \le k\frac{\pi}{n+1}\,, \quad k=1,2,\ldots,\lfloor\frac{n}{2}\rfloor\,, \tag{7.126}$$

can be found in a form close to (7.124) in [Szego (1975)], see Theorem 6.21.3.

A key result for the analysis of efficiency of the finite-difference preconditioners for spectral elements with the GLL nodes is the interpolation error bound for the polynomial Lagrange interpolations over Gauss-Lobatto quadrature sets of nodes in 1d. We introduce the weighted L_2 space $L_{2,\alpha}(I)$ equipped with the norm $\|\phi\|_{L_{2,\alpha}(I)}$,

$$\|\phi\|^2_{L_{2,\alpha}(I)} = \int_I (1-\xi^2)^\alpha\, \phi^2(\xi)\, d\xi\,,$$

the weighted Sobolev spaces

$$H^k_\alpha(I) = \{\phi \in L_{2,\alpha}(I) : \frac{d^l\phi}{d\xi^l} \in L_{2,\alpha}(I),\ l=1,2,\ldots,k\}\,,$$

equipped with the norm $\|\phi\|_{H^k_\alpha(I)}$,

$$\|\phi\|^2_{H^k_\alpha(I)} = \|\phi\|_{L_{2,\alpha}(I)} + \sum_{l=1}^{k} |\phi|^2_{l,\alpha,I}\,, \quad |\phi|^2_{l,\alpha,I} = \int_I (1-\xi^2)^\alpha (\frac{d^l\phi}{d\xi^l})^2 d\xi\,,$$

and the the interpolation operator $\pi_{\alpha,p} : C(\overline{I}) \to \mathcal{P}_p$ over the Gauss-Lobatto nodes $\xi_i^{\alpha,p},\ i=0,1,\ldots,p$.

Theorem 7.4. *For any real $s \in [1, p+1]$, and any function $\phi \in H^s_\alpha(I)$, the interpolation error estimate*

$$|\phi - \pi_{\alpha,p}\phi|_{H^1_\alpha(I)} \le c p^{1-s}\, |\phi|_{H^s_\alpha(I)} \tag{7.127}$$

holds, where $c = c(\alpha, s)$ is a positive generic constant.

Proof. For the rather involved proof of this theorem, see Theorem 4.5 and Corollary 4.6 in [Bernardi and Maday (1992a)]. □

An immediate consequence of (7.127) is one of the most important estimates characterizing the stability of polynomial interpolations over the GLL nodes. Let $\mathcal{V}^{\alpha,p}_{GL}(I)$ be the space of continuous piecewise linear functions related to the set of Gauss-Lobatto quadrature nodes, and let $\hat{\pi}_{\alpha,p} : C(\overline{I}) \to \mathcal{V}^{\alpha,p}_{GL}(I)$ be the interpolation operator over this set of nodes, which satisfies the equations

$$(\hat{\pi}_{\alpha,p}\phi)(\xi_i^{\alpha,p}) = \phi(\xi_i^{\alpha,p})\,, \quad i=0,1,\ldots,p\,.$$

The inequality

$$|\psi|_{H^1_\alpha(I)} \prec |\hat{\pi}_{\alpha,p}\psi)|_{H^1_\alpha(I)}\,, \quad \forall\, \psi \in \mathcal{P}_p\,,$$

directly follows from the interpolation error estimate (7.127), if we write it for $\phi = \hat{\pi}_{\alpha,p}\psi \in \mathcal{V}^{\alpha,p}_{GL}(I)$ and apply the triangle inequality. For $\alpha = 0$, it turns into the inequality

$$|\psi|_{1,I} \prec |\hat{\pi}_{0,p}\psi)|_{1,I}\,, \quad \forall\, \psi \in \mathcal{P}_p\,. \tag{7.128}$$

For $\alpha = 0$, to which GLL nodes correspond, we simplify notations by omitting index α and setting $\xi_i^{(p)} = \xi_i^{0,p},\ \rho_i^{(p)} = \rho_i^{0,p},\ \pi_p = \pi_{0,p},\ \hat{\pi}_p = \hat{\pi}_{0,p}$. We recall that for the GLL nodes even more simple notations $\xi_i^{0,p} = \eta_i$ were used earlier, see (7.79), to which we will return after this section.

Theorem 7.5. *For any polynomial $\phi \in \mathcal{P}_p$ the equivalences*

$$\|\hat{\pi}_p\phi\|_{0,I} \asymp \|\phi\|_{0,I} \quad \textit{and} \quad |\hat{\pi}_p\phi|_{1,I} \asymp |\phi|_{1,I} \tag{7.129}$$

are valid.

Proof. For approving the second relation in (7.129), we need to derive only the inequality that is opposite to (7.128). However, it holds for any function $\phi \in H^1(I)$ and its piecewise linear interpolation $\hat{\phi}$ over arbitrary set of nodes $\xi_i,\ i = 0, 1, \ldots, p,\ \xi_0 = -1,\ \xi_p = 1,\ \xi_{i+1} > \xi_i$. Indeed, since $\hat{\phi}' = \text{const}$ at any interval $(\xi_i, \xi_{i+1}) \subset I$ and $(\phi - \hat{\phi}) = 0$ at $\xi = \xi_i, \xi_{i+1}$, we have

$$\int_{\xi_i}^{\xi_{i+1}} (\phi - \hat{\phi})'\hat{\phi}'\, d\xi = 0\,, \quad i = 0, 1, \ldots, p-1\,.$$

Therefore, we can write

$$\int_{\xi_i}^{\xi_{i+1}} (\hat{\phi}')^2\, d\xi = \int_{\xi_i}^{\xi_{i+1}} \phi'\hat{\phi}'\, d\xi \le \frac{1}{2}\int_{\xi_i}^{\xi_{i+1}} [(\hat{\phi}')^2 + (\phi')^2]\, d\xi\,.$$

This immediately results in the inequality

$$\int_{\xi_i}^{\xi_{i+1}} (\hat{\phi}')^2\, d\xi \le \int_{\xi_i}^{\xi_{i+1}} (\phi')^2\, d\xi\,. \tag{7.130}$$

We turn now to the proof of first relation (7.129). The Gauss-Lobatto quadrature is not exact for polynomials $\phi \in \mathcal{P}_{2p}$. However, it provides an important equivalence relation

$$\sum_{i=0}^{p} \phi^2(\xi_i^{\alpha,p})\rho_i^{\alpha,p} \asymp \int_I (1-\xi^2)^\alpha \phi^2(\xi) d\xi\,, \quad \forall\ \phi \in \mathcal{P}_p\,, \tag{7.131}$$

proved by [Canuto and Quarteroni (1982)]. From (7.114) and (7.131), it follows that, in order to prove first equivalence of (7.129), it is sufficient to establish spectral equivalence of the matrices $\mathbf{R} := \mathrm{diag}\,[\rho_i^{(p)}]_{i=0}^{p}$ and $\mathbb{K}_0$. We recall that we use the same notations for the step sizes $\hbar_i$ of GL grids, *e.g.*, GLL grids, related to different α and p.

In [Szego (1975)], see (15.3.14), one can found the bounds

$$\underline{c}\frac{(1-(\xi_i^{(p)})^2)^{1/2}}{p} \le \rho_i^{(p)} \le \overline{c}\frac{(1-(\xi_i^{(p)})^2)^{1/2}}{p}\,, \tag{7.132}$$

and, if we take into consideration (7.120) and (7.123), we can prove inequalities for the values

$$\vartheta_i^{\alpha,p} = \arccos\,(-\xi_i^{\alpha,p}) \tag{7.133}$$

that are similar to (7.124) and (7.125). For the values $\vartheta_i^{(p)} := \vartheta_i^{0,p}$, they take the form

$$\frac{i\pi}{p} \le \vartheta_i^{(p)} \le \frac{(i+1)\pi}{p}\,, \quad i = 1, 2, \ldots, \lfloor\frac{p}{2}\rfloor\,. \tag{7.134}$$

Now, on one hand, by substituting $\xi_i^{(p)} = -\cos\vartheta_i^{(p)}$ in (7.132) and taking into account (7.134), we come to the inequalities

$$\underline{c}\,\frac{(i+1)\pi}{p^2} \le \rho_i^{(p)} \le\, \overline{c}\,\frac{(i+1)\pi}{p^2} \tag{7.135}$$

with some generic positive constants $\underline{c}$ and $\overline{c}$.

On the other hand, in view of (7.116), we conclude that $\hbar_i$ satisfy similar inequalities of the form

$$\underline{c}\,\frac{(i+1)\pi}{p^2} \le \hbar_i \le\, \overline{c}\,\frac{(i+1)\pi}{p^2}\,, \quad i = 0, 1, 2, \ldots, \lfloor\frac{p}{2}\rfloor\,. \tag{7.136}$$

This completes the proof of the relation (7.129) for the L_2-norm. □

Corollary 7.3. *Let $\mathbb{K}_{1,\mathrm{sp}}$ and $\mathbb{K}_{0,\mathrm{sp}}$ be the stiffness and mass matrices for the one-dimensional spectral reference element with the GLL nodes. Then the equivalences (7.115) hold.*

Proof. The equivalences (7.115) are the matrix forms of the equivalences (7.129). □

7.4.3 *Two-Dimensional Spectral Reference Elements*

Finite-difference and finite element preconditioners for the spectral reference elements stiffness matrices possess close properties. While the former are simpler in the implementation, the latter are more convenient for the analysis. The finite element preconditioners are usually defined by means of the finite element assemblages of $\mathcal{P}_1$ or $\mathcal{Q}_1$ elements on the triangular or rectangular nonuniform meshes, respectively, with the nodes in the nodes of the spectral mesh. For the reason of similarity of properties and the spectral equivalence, in what follows we usually take the preconditioners of one type, *e.g.*, based on FE model with triangular elements, for a typical representative. Without restriction of generality, it is assumed $p = 2N$ that provides some simplifications due to the four-fold symmetry of 2d spectral reference elements.

The step sizes $\hbar_k := \eta_k - \eta_{k-1}$ of the GLC grid (7.80) satisfy the relationship $\hbar_k \asymp k/p^2,\ k \le N$. The step sizes of the GLL grid have the same asymptote. Therefore, for the reference elements with GLL nodes one can use preconditioners defined by means of the both meshes. Some proofs related to the spectral elements preconditioning became even easier for the finite-difference preconditioners induced by more general meshes having the pointed out asymptote. We will below discuss one family of such meshes.

For some fixed $c_1, c_2 > 0$ and$\gamma \ge 0$, we introduce the grid, which is defined by

$$\eta_0 = -1, \eta_i = \eta_{i-1} + \hbar_i, \eta_N = 0, \quad c_1 \frac{i^\gamma}{\aleph} \le \hbar_i \le c_2 \frac{i^\gamma}{\aleph},\ \aleph = \sum_{i=1}^{N} i^\gamma\,, \quad (7.137)$$

on $[-1, 0]$ and is continued on [0,1] by the symmetry. In the particular case of $c_1 = c_2 = 1$, we have the mesh sizes $\hbar_i = i/\aleph = 2i/(N^2 + N) = \beta_i\, i/p^2$, where $\beta_i \in [4, 8]$. Such meshes will be called *pseudospectral*.

For the stiffness matrices of the spectral reference elements, we use the notation $\mathbf{A}_{\rm sp}$ and assume that these matrices are induced by the Dirichlet integral

$$a_{\tau_0}(u, v) = \int_{\tau_0} \nabla u \cdot \nabla v \, dx\,.$$

Let an orthogonal mesh $x_k \equiv \eta_i,\ i = 0, 1, \ldots, p,\ k = 1, 2$, be given on τ_0, and let $\mathcal{H}(\tau_0)$ be the space of functions continuous on τ_0 and belonging to $\mathcal{Q}_{1,x}$ on each nest. The same notation $\mathcal{H}(\tau_0)$ will usually stand for the space of piecewise linear functions on the triangulation, which is obtained when each nest of the rectangular mesh is subdivided in two triangles by

one of the nest diagonals. The special notations $\mathcal{H}_{\rm sp}(\tau_0), \mathcal{H}_{\rm p/s}(\tau_0)$ are used for the space $\mathcal{H}(\tau_0)$ defined by means of the spectral and the pseudospectral meshes, respectively. The preconditioners $\boldsymbol{\mathcal{A}}_\circ$, $\boldsymbol{\mathcal{A}}_{\rm sp}$, $\boldsymbol{\mathcal{A}}_{\rm p/s}$ for $\mathbf{A}_{\rm sp}$ are understood as the FE matrices, induced by the Dirichlet integral a_{τ_0} on the spaces $\mathcal{H}(\tau_0)$, $\mathcal{H}_{\rm sp}(\tau_0)$, $\mathcal{H}_{\rm p/s}(\tau_0)$, respectively. That means that $\boldsymbol{\mathcal{A}}_\circ$ may imply one or both preconditioners $\boldsymbol{\mathcal{A}}_{\rm sp}$, $\boldsymbol{\mathcal{A}}_{\rm p/s}$. By means of the same finite element spaces, we introduce also the preconditioners $\boldsymbol{\mathcal{M}}_\circ$, $\boldsymbol{\mathcal{M}}_{\rm sp}$, $\boldsymbol{\mathcal{M}}_{\rm p/s}$ for the mass matrix $\mathbb{M}_{\rm sp}$.

Simpler finite-difference preconditioners for the stiffness and mass matrices can be introduced. Let

$$\mathbb{K}_0 = \mathrm{diag}\,[\widetilde{h}_i = \frac{1}{2}(\hbar_i + \hbar_{i+1})]_{i=0}^{p}\,, \quad \hbar_i = 0 \ \text{ for } \ i = 0, p+1\,, \tag{7.138}$$

and $\mathbb{K}_1$ be the FE stiffness matrix defined by the quadratic form $(v', w')_{(-1,1)}$ on the space of continuous on [-1,1] piecewise linear functions induced by the fine grid η_i. The explicit form of $\mathbb{K}_1$ is given by the equations

$$\begin{aligned}
(\mathbb{K}_1\mathbf{u})|_{i=0} &= \tfrac{1}{\hbar_1}(u_1 - u_0)\,,\\
(\mathbb{K}_1\mathbf{u})|_{i} &= -\tfrac{1}{\hbar_i}u_{i-1} + k_{ii}u_i - \tfrac{1}{\hbar_{i+1}}u_{i+1}, \quad i = 1, 2, \ldots, p\,,\\
(\mathbb{K}_1\mathbf{u})|_{i=p+1} &= \tfrac{1}{\hbar_p}(u_{p+1} - u_p),
\end{aligned} \tag{7.139}$$

where $k_{ii} = \hbar_i^{-1} + \hbar_{i+1}^{-1}$. The finite-difference preconditioners $\boldsymbol{\mathcal{A}}^\hbar := \boldsymbol{\mathcal{A}}^\hbar_{\rm sp}, \boldsymbol{\mathcal{A}}^\hbar_{\rm p/s}$ and $\boldsymbol{\mathcal{M}}^\hbar := \boldsymbol{\mathcal{M}}^\hbar_{\rm sp}, \boldsymbol{\mathcal{M}}^\hbar_{\rm p/s}$, can be represented in the Kronecker product forms

$$\boldsymbol{\mathcal{A}}^\hbar = \mathbb{K}_0 \otimes \mathbb{K}_1 + \mathbb{K}_1 \otimes \mathbb{K}_0 \quad \text{and} \quad \boldsymbol{\mathcal{M}}^\hbar = \mathbb{K}_0 \otimes \mathbb{K}_0\,, \tag{7.140}$$

with the matrices $\mathbb{K}_0$ and $\mathbb{K}_1$ defined for the corresponding mesh.

Lemma 7.8. *Let $\mathbf{A}_{\rm sp}$ and $\mathbb{M}_{\rm sp}$ be the stiffness and mass matrices for the spectral reference element with the GLL nodes. Then the equivalences*

$$\mathbf{A}_{\rm sp} \asymp \boldsymbol{\mathcal{A}}_\circ \asymp \boldsymbol{\mathcal{A}}^\hbar \quad \text{and} \quad \mathbb{M}_{\rm sp} \asymp \boldsymbol{\mathcal{M}}_\circ \asymp \boldsymbol{\mathcal{M}}^\hbar \tag{7.141}$$

hold. Thus, the matrices $\boldsymbol{\mathcal{A}}_\circ$ and $\boldsymbol{\mathcal{A}}^\hbar$ as well as and $\boldsymbol{\mathcal{M}}_\circ$ and $\boldsymbol{\mathcal{M}}^\hbar$ are spectrally equivalent preconditioners for the spectral element stiffness and mass matrices $\mathbf{A}_{\rm sp}$ and $\mathbb{M}_{\rm sp}$, respectively, uniformly with respect to p.

Proof. We note that

$$\mathbf{A}_{\rm sp} = \mathbb{K}_{1,\rm sp} \otimes \mathbb{K}_{0,\rm sp} + \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{1,\rm sp} \quad \text{and} \quad \mathbb{M}_{\rm sp} = \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{0,\rm sp}\,, \tag{7.142}$$

and turn to the first group of equivalences in (7.141). According to (7.140) and (7.142), the matrices $\mathbf{A}_{\rm sp}$ and $\boldsymbol{\mathcal{A}}_\circ$ are defined by the identical sums of

the Kronecker products of the matrices related to the 1d spectral element on the interval I. Therefore, by Lemma 7.9, given below, it is sufficient to establish the equivalences $\mathbb{K}_{1,\mathrm{sp}} \asymp \mathbb{K}_1$ and $\mathbb{K}_{0,\mathrm{sp}} \asymp \mathbb{K}_0$. These relations follow from Corollary 7.3, uniform spectral equivalence of the matrices $\mathbb{M}$ and $\mathbb{K}_0$, and the uniform equivalence of the spectral and pseudo-spectral meshes. Similar arguments approve the second group of equivalences in (7.141). □

Lemma 7.9. *Let us suppose that* $\mathbb{A}^{(k)}$ *and* $\mathbb{B}^{(k)}$, $k = 1, 2$, *are* $n \times n$ *matrices,* $\ker[\mathbb{B}^{(k)}] \subseteq \ker[\mathbb{A}^{(k)}]$, *matrices* $\mathbb{A}$ *and* $\mathbb{B}$ *are defined by*

$$\mathbb{A} = \mathbb{A}^{(1)} \otimes \mathbb{A}^{(2)} \quad \textit{and} \quad \mathbb{B} = \mathbb{B}^{(1)} \otimes \mathbb{B}^{(2)}\,,$$

and $\lambda^{(k)}$ *and* $\mathbf{x}^{(k)}$ *are eigenvalues and eigenvectors defined by the eigenvalue problem*

$$\mathbb{A}^{(k)}\mathbf{x}^{(k)} = \lambda^{(k)}\mathbb{B}^{(k)}\mathbf{x}^{(k)}\,. \tag{7.143}$$

Then $\lambda = \lambda^{(1)}\lambda^{(2)}$ *and* $\mathbf{z} = \mathbf{x}^{(1)} \otimes \mathbf{x}^{(2)}$ *are an eigenvalue and eigenvector of the eigenvalue problem*

$$\mathbb{A}\mathbf{z} = \lambda\mathbb{B}\mathbf{z}\,. \tag{7.144}$$

If $\mathbb{A}^{(k)}$ *and* $\mathbb{B}^{(k)}$ *are symmetric nonnegative matrices, then, on the subspace* $V := R^{n^2} \setminus \ker[\mathbb{A}]$, $\ker[\mathbb{A}] = \ker[\mathbb{A}^{(1)}] \times \ker[\mathbb{A}^{(2)}]$, *the inequalities*

$$\lambda^{(1)}_{\min}\lambda^{(2)}_{\min}\mathbb{B} \le \mathbb{A} \le \lambda^{(1)}_{\max}\lambda^{(2)}_{\max}\mathbb{B} \tag{7.145}$$

hold, where $\lambda^{(k)}_{\min}$ *and* $\lambda^{(k)}_{\max}$ *are the lowest nonzero and the maximal eigenvalues of the eigenvalue problem(7.143), respectively.*

Proof. It is easy to notice that the following relations hold:

$$\begin{aligned}\mathbb{A}\mathbf{z} &= \left(\mathbb{A}^{(1)} \otimes \mathbb{A}^{(2)}\right)\left(\mathbf{x}^{(1)} \otimes \mathbf{x}^{(2)}\right) = \left(\mathbb{A}^{(1)}\mathbf{x}^{(1)}\right) \otimes \left(\mathbb{A}^{(2)}\mathbf{x}^{(2)}\right)\\ &= \lambda^{(1)}\lambda^{(2)}\left(\mathbb{B}^{(1)}\mathbf{x}^{(1)}\right) \otimes \left(\mathbb{B}^{(2)}\mathbf{x}^{(2)}\right)\\ &= \lambda^{(1)}\lambda^{(2)}(\mathbb{B}^{(1)} \otimes \mathbb{B}^{(2)})(\mathbf{x}^{(1)} \otimes \mathbf{x}^{(2)}) = \lambda^{(1)}\lambda^{(2)}\mathbb{B}\mathbf{z}\,,\end{aligned}$$

which prove (7.144). Obviously, if $\mathbb{A}^{(k)}$ and $\mathbb{B}^{(k)}$ are s.p.d matrices, then $\mathbb{A}$ and $\mathbb{B}$ are also s.p.d matrices. Besides, for each $k = 1, 2$, we have n eigenvectors $\mathbf{x}^{(k)}$ which are orthogonal in the scalar products $(\mathbf{C}\cdot,\cdot)$, induced by the matrices $\mathbf{C} = \mathbf{A}^{(k)}, \mathbf{B}^{(k)}$, and have real eigenvalues $\lambda^{(k)} > 0$. Now, from the first statement of Lemma, just proved, and the fact that $\mathbb{A}$ and $\mathbb{B}$ are s.p.d. matrices, we conclude that the matrix $\mathbb{A}$ has n^2 eigenvectors of the form $\mathbf{z} = \mathbf{x}^{(1)} \otimes \mathbf{x}^{(2)}$, eigenvectors are orthogonal in the scalar products,

induced by the matrices $\mathbb{A}$ and $\mathbb{B}$, and $\lambda = \lambda^{(1)}\lambda^{(2)}$ are the corresponding eigenvalues. Therefore, the minimal and maximal eigenvalues of the matrix eigenvalue problem (7.144) are $\lambda_{\min} = \lambda^{(1)}_{\min}\lambda^{(2)}_{\min}$ and $\lambda_{\max} = \lambda^{(1)}_{\max}\lambda^{(2)}_{\max}$, respectively. Moreover, for the s.p.d. matrices $\mathbb{A}$ and $\mathbb{B}$, we have the inequalities $\lambda_{\min}\mathbb{B} \le \mathbb{A} \le \lambda_{\max}\mathbb{B}$, which are nothing but (7.145). In the more general case when $\ker[\mathbb{A}] \ne \varnothing$, the proof of (7.145) on the subspace V is the same. □

The equivalences $\mathbf{A}_{\rm sp} \asymp \boldsymbol{\mathcal{A}}_{\rm sp}$ and $\mathbb{M}_{\rm sp} \asymp \boldsymbol{\mathcal{M}}_{\rm sp}$ contained in (7.141) can also be rewritten in the form

$$|v|_{1,\tau_0} \asymp |\hat{w}|_{1,\tau_0} \quad \text{and} \quad \|v\|_{0,\tau_0} \asymp \|\hat{w}\|_{0,\tau_0} \tag{7.146}$$

that is valid for any $v \in \mathcal{Q}_p$ and $\hat{w} \in \mathcal{H}_{\rm sp}(\tau_0)$ coinciding at the spectral nodes. They are the counterparts of (7.129) for two dimensions.

7.4.4 *Factored Preconditioners for Spectral Elements*

We mention that the finite-difference and finite element preconditioners for the spectral reference elements look noticeably different from those for the hierarchical elements. In particular, the asymptotes of their condition numbers are different. Let us consider, for instance, the blocks $\boldsymbol{\mathcal{A}}_{I,\circ} = \boldsymbol{\mathcal{A}}_{I,{\rm sp}}, \boldsymbol{\mathcal{A}}_{I,{\rm p/s}}$ of the preconditioners $\boldsymbol{\mathcal{A}}_{\circ} = \boldsymbol{\mathcal{A}}_{\rm sp}, \boldsymbol{\mathcal{A}}_{\rm p/s}$ in Lemma 7.8, that are related to the internal degrees of freedom. They result from the finite element approximations of Laplace operator by means of the $\mathcal{Q}_1$ or $\mathcal{P}_1$ finite elements on essentially nonuniform orthogonal grids. In general, they are not block-diagonal matrices. At the same time, the preconditioner $\boldsymbol{\Lambda}_{\triangle}$ in (7.102), see p. 235, for the hierarchical reference element is a block-diagonal matrix with 4 independent blocks. Each block $\boldsymbol{\Lambda}$ of it is the finite-difference approximation to a deteriorating elliptic differential operator on the uniform grid. The preconditioners of each of the two types are grid operators which are irregular in their own way. This creates specific problems in designing fast solvers for them. Therefore, at first glance, fast solvers should be different for the preconditioners of different types. However, it has been found out that simple transformations of the matrices $\boldsymbol{\mathcal{A}}_{I,\circ}$, as well as $\boldsymbol{\mathcal{A}}^{\hbar}_{I} = \boldsymbol{\mathcal{A}}^{\hbar}_{I,{\rm sp}}, \boldsymbol{\mathcal{A}}^{\hbar}_{I,{\rm p/s}}$, as quadratic forms by diagonal matrices transform them into the matrices with the properties identical in essence to the properties of $\boldsymbol{\Lambda}$. As we shall see later, this practically means that any fast solver for $\boldsymbol{\Lambda}$ can be relatively easily adapted into a fast solver for the preconditioners $\boldsymbol{\mathcal{A}}^{\hbar}_{I} = \boldsymbol{\mathcal{A}}^{\hbar}_{I,{\rm sp}}, \boldsymbol{\mathcal{A}}^{\hbar}_{I,{\rm p/s}}$. This approach, which is illustrated below, opens a

direct way to the class of efficient factorized preconditioner-solvers for the stiffness matrices of the spectral reference elements.

We scale the nodal variables according to the relations

$$\widetilde{\mathbf{v}} = \mathbf{C}^{-1}\mathbf{v}\,, \quad \text{with} \quad \mathbf{C} = p^2\,\mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2}\,,$$

and consider the transformed matrix $\widetilde{\boldsymbol{\mathcal{A}}} = \mathbf{C}\boldsymbol{\mathcal{A}}^{\hbar}\mathbf{C}$. Taking into account the representation (7.140) by the sum of Kronecker products, we obtain

$$\begin{aligned}\widetilde{\boldsymbol{\mathcal{A}}} &= p^4\,\mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2}\,\boldsymbol{\mathcal{A}}^{\hbar}\,\mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} \\ &= p^4\,\mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2}\,[\,\mathbb{K}_0 \otimes \mathbb{K}_1 + \mathbb{K}_1 \otimes \mathbb{K}_0]\;\mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} \\ &= p^4\,\left[\,\mathbb{K}_0^2 \otimes \widetilde{\mathbb{K}}_1 + \widetilde{\mathbb{K}}_1 \otimes \mathbb{K}_0^2\right], \end{aligned} \tag{7.147}$$

with $\widetilde{\mathbb{K}}_1 = \mathbb{K}_0^{1/2}\mathbb{K}_1\mathbb{K}_0^{1/2}$.

In order to construct fast local Dirichlet solvers, we should primarily take care of preconditioning the block of the reference element stiffness matrix related to inner unknowns. For this reason, we consider the block $\widetilde{\boldsymbol{\mathcal{A}}}_I = \mathbf{C}_I^{\top}\boldsymbol{\mathcal{A}}^{\hbar,I}\mathbf{C}_I$ of the matrix $\widetilde{\boldsymbol{\mathcal{A}}}$, where $\boldsymbol{\mathcal{A}}^{\hbar,I}$ and $\mathbf{C}_I$ are the corresponding blocks of $\boldsymbol{\mathcal{A}}^{\hbar}$ and $\mathbf{C}$. Clearly, if (7.141) holds, then $\widetilde{\boldsymbol{\mathcal{A}}}_I$ is the spectrally equivalent preconditioner to the transformed internal stiffness matrix $\widetilde{\mathbf{A}}_I = \mathbf{C}_I^{\top}\mathbf{A}_I\mathbf{C}_I$.

Lemma 7.10. *Let the matrix $\boldsymbol{\mathcal{A}}^{\hbar}$ be generated on one of the spectral or on the pseudospectral mesh, and let the $(p-1)\times(p-1)$ matrices $\boldsymbol{\Delta}_{\diamond}$, $\boldsymbol{\mathcal{D}}_{\diamond}$ and $\boldsymbol{\Lambda}_{\diamond}$ be defined as follows:*

$$\begin{aligned}\boldsymbol{\Delta}_{\diamond} &= \operatorname{tridiag}[-1,2,-1]\,, \\ \boldsymbol{\mathcal{D}}_{\diamond} &= \operatorname{diag}[1,4,\ldots,(N-1)^2,N^2,(N-1)^2,\ldots,4,1]\,, \\ \boldsymbol{\Lambda}_{\diamond} &= \boldsymbol{\Delta}_{\diamond}\otimes\boldsymbol{\mathcal{D}}_{\diamond} + \boldsymbol{\mathcal{D}}_{\diamond}\otimes\boldsymbol{\Delta}_{\diamond}\,. \end{aligned}$$

Then $\boldsymbol{\Lambda}_{\diamond}$ and the preconditioner for the transformed internal stiffness matrix are spectrally equivalent, i.e.,

$$\boldsymbol{\Lambda}_{\diamond} \asymp \widetilde{\boldsymbol{\mathcal{A}}}_I\,. \tag{7.148}$$

Proof. It is sufficient to consider only the case of the pseudospectral mesh having the sizes $\hbar_i = \beta_i i/p^2$ for $i = 1,2,\ldots,N$, where, as was noted earlier, $\beta_i \in [4,8]$. As usual, we do not take care about absolute constants, entering the bounds. For this reason, we omit β_i from our considerations. In particular, in the proof of inequalities (7.149) below, we accept $\hbar_i = i/p^2$ for convenience. Thus, we write

$$\widetilde{h}_i = \frac{1}{2}(\hbar_i + \hbar_{i+1}) = \frac{1}{2p^2}(2i+1),$$

and, therefore, we have

$$\frac{i^2}{2p^4} \leq \widetilde{h}_i^2 \leq \frac{5i^2}{2p^4}\,, \quad i = 1, 2, \ldots, N\,,$$

with the consequence that $\boldsymbol{\mathcal{D}}_\diamond \asymp p^4 \mathbb{K}_{0,I}^2$.

It remains to prove that $\boldsymbol{\Delta}_\diamond \asymp \widetilde{\mathbb{K}}_{1,I}$. Let us introduce the notations $k_{i,j}$ for the entries of $\widetilde{\mathbb{K}}_1$, *i.e.,* $\widetilde{\mathbb{K}}_1 = \text{tridiag}\,[k_{i,i-1}, k_{i,i}, k_{i,i+1}]$, and estimate the entries from below and from above. For $i = 1, 2, \ldots, N$, we have

$$k_{i,i} = \frac{1}{2}(\hbar_i + \hbar_{i+1})(\frac{1}{\hbar_i} + \frac{1}{\hbar_{i+1}}) = 2 + \frac{1}{2i(i+1)}\,,$$

$$1 \geq |k_{i,i-1}| = \frac{1}{\hbar_i}\sqrt{(\hbar_{i-1} + \hbar_i)(\hbar_i + \hbar_{i+1})} = \sqrt{1 - \frac{1}{4i^2}} \geq 1 - \frac{1}{4i^2}\,,$$

$$1 \geq |k_{i,i+1}| \geq 1 - \frac{1}{4(i+1)^2}\,.$$

The matrix $\widetilde{\mathbb{K}}_{1,I}$ may be represented as

$$\widetilde{\mathbb{K}}_{1,I} = \mathbf{K}_{1,I} + \mathbf{D}\,,$$

where $\mathbf{K}_{1,I} = \text{tridiag}\,[k_{i,i-1}, (k_{i,i-1} + k_{i,i+1}), k_{i,i+1}]$ and $\mathbf{D} = \text{diag}\,[k_{i,i} - k_{i,i-1} - k_{i,i+1}]$. Due to the above relationships for $k_{i,j}$, we have

$$\frac{3}{4}\boldsymbol{\Delta}_\diamond \leq \widetilde{\mathbf{K}}_{1,I} \leq \boldsymbol{\Delta}_\diamond \quad \text{and} \quad \frac{1}{4}\boldsymbol{\mathcal{D}}_\diamond^{-1} \leq \mathbf{D} \leq \boldsymbol{\mathcal{D}}_\diamond^{-1}\,.$$

Therefore, we arrive at the spectral inequalities

$$\frac{1}{4}\boldsymbol{\Lambda}_\dagger \leq \widetilde{\boldsymbol{\mathcal{A}}}_I \leq \boldsymbol{\Lambda}_\dagger\,, \tag{7.149}$$

with

$$\boldsymbol{\Lambda}_\dagger = (\boldsymbol{\Delta}_\diamond + \boldsymbol{\mathcal{D}}_\diamond^{-1}) \otimes \boldsymbol{\mathcal{D}}_\diamond + \boldsymbol{\mathcal{D}}_\diamond \otimes (\boldsymbol{\Delta}_\diamond + \boldsymbol{\mathcal{D}}_\diamond^{-1})\,. \tag{7.150}$$

In order to come to the inequalities (7.148), we will prove that

$$\boldsymbol{\mathcal{D}}_\diamond^{-1} \leq \boldsymbol{\Delta}_\diamond\,. \tag{7.151}$$

The pair of the inequalities below, which is a direct consequence of Hardy's inequality, can be used for this purpose. Indeed, similar to (7.104), Hardy's inequality allows us to derive the inequalities

$$\|\frac{v}{1+x}\|_{0,(-1,1)}^2 \leq 8\,|\,v\,|_{1,(-1,1)}^2 \quad \text{and} \quad \|\frac{v}{1-x}\|_{0,(-1,1)}^2 \leq 8\,|\,v\,|_{1,(-1,1)}^2\,,$$

which are valid for all $v \in \mathring{H}^1(-1,1)$. From these inequalities, it follows that, for the function

$$\phi = \begin{cases} 1+x\,, & x \in [-1,0]\,, \\ 1-x\,, & x \in [0,1]\,, \end{cases} \tag{7.152}$$

we have

$$\|\phi^{-1}v\|^{\,2}_{0,(-1,1)} \le 8\,|\,v\,|^2_{1,(-1,1)}\,, \quad \forall\, v \in \mathring{H}^1(-1,1)\,. \tag{7.153}$$

The derivation of inequality (7.151) from inequality (7.153) is analogous to the derivation of inequality $\boldsymbol{\mathcal{D}}_a^{-1} \le \boldsymbol{\Delta}$ from inequality (7.104) in the proof of Lemma 7.5. □

Corollary 7.4. *Let the factored preconditioners*

$$\boldsymbol{\mathfrak{K}}_\diamond := \mathbf{C}_I^{-1}\boldsymbol{\Lambda}_\diamond\mathbf{C}_I^{-1} \quad \textit{and} \quad \boldsymbol{\mathfrak{K}}_\dagger := \mathbf{C}_I^{-1}\boldsymbol{\Lambda}_\dagger\mathbf{C}_I^{-1}$$

be defined for one of the spectral or for the pseudospectral mesh. Then they are spectrally equivalent to the preconditioners $\boldsymbol{\mathcal{A}}_{I,\circ} = \boldsymbol{\mathcal{A}}_{I,\mathrm{sp}}, \boldsymbol{\mathcal{A}}_{I,\mathrm{p/s}}$ of the internal stiffness matrix, i.e.,

$$\boldsymbol{\mathcal{A}}_{I,\circ} \asymp \boldsymbol{\mathfrak{K}}_\diamond\,, \quad \text{and} \quad \boldsymbol{\mathcal{A}}_{I,\circ} \asymp \boldsymbol{\mathfrak{K}}_\dagger\,, \tag{7.154}$$

uniformly in p. Moreover, we also have the spectral equivalences

$$\mathbf{A}_{\mathrm{sp}} \asymp \boldsymbol{\mathfrak{K}}_\diamond\,, \quad \text{and} \quad \mathbf{A}_{\mathrm{sp}} \asymp \boldsymbol{\mathfrak{K}}_\dagger \tag{7.155}$$

for the stiffness matric $\mathbf{A}_{\mathrm{sp}}$ of the spectral reference element with GLL nodes.

Proof. The Corollary is the direct consequence of Lemmas 7.8, 7.10, and (7.149). □

The matrix $\mathbf{C}$ is diagonal. Hence, the arithmetical costs of solving systems with the preconditioners $\boldsymbol{\mathfrak{K}}_\diamond$ and $\boldsymbol{\mathfrak{K}}_\dagger$ coincide in the order with the arithmetical costs of solving systems with the matrices $\boldsymbol{\Lambda}_\diamond$ and $\boldsymbol{\Lambda}_\dagger$.

Let us introduce the notations $\hbar = 2p^{-1}$ and $\phi_i = \phi(-1+i\hbar)$. The matrix $\boldsymbol{\Lambda}_\diamond$ may be interpreted as the finite-difference approximation

$$\boldsymbol{\Lambda}_\diamond\mathbf{u}|_{i,j} = -\phi_i^2\,\frac{u_{i,j-1} - 2u_{i,j} + u_{i,j-1}}{\hbar^2} - \phi_j^2\,\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{\hbar^2} = f_{i,j} \tag{7.156}$$

to the PDE with the differential operator

$$\mathcal{L}_{\mathrm{sp}}u \equiv -\phi^2(\xi_1)\frac{\partial^2 u}{\partial\xi_2^2} - \phi^2(\xi_2)\frac{\partial^2 u}{\partial\xi_1^2} \tag{7.157}$$

in τ_0 under homogeneous Dirichlet boundary condition on a uniform grid. In each quarter of τ_0, differential expression of this operator coincides with the differential operator (7.112) up to the multiplier and the translation and rotation of the axes. As a consequence, we will later show that in essence the same solution procedures may be used for solving systems of linear algebraic equations with the matrices $\mathbf{\Lambda}_\diamond$, $\mathbf{\Lambda}_\dagger$ the matrix $\mathbf{\Lambda}$ in (7.101) for the hierarchical reference element. This is in spite of the difference in their dimensions, which are $(p-1)^2\times(p-1)^2$ and $(\mathrm{int}\,[p/2])^2\times(\mathrm{int}\,[p/2])^2$, respectively.

The analysis of fast methods becomes simpler, if the preconditioner is a finite element matrix. We subdivide each square nest of the size $\hbar$ in the two triangles by one of the diagonals and introduce in this way the triangulation of τ_0. Let us now introduce the space $\mathring{\mathcal{H}}^{\hbar}_{\Delta}(\tau_0)$ of continuous in τ_0 piecewise linear functions vanishing on $\partial\tau_0$. Then we denote by $\mathbf{B}_\diamond$ the FE stiffness matrix generated by the bilinear form

$$a^{(1)}_{\tau_0}(u,v)=\int_{\tau_0}\left(\phi^2(\xi_2)u_{\xi_1}v_{\xi_1}+\phi^2(\xi_1)u_{\xi_2}v_{\xi_2}\right)\,d\xi\,,$$

i.e., $\mathbf{u}^\top\mathbf{B}_\diamond\mathbf{v}=a^{(1)}_{\tau_0}(u,v)$ for all $\mathbf{u},\mathbf{v}\leftrightarrow u,v\in\mathring{\mathcal{H}}^{\hbar}_{\Delta}(\tau_0)$, where $w_{\xi_k}=\partial w/\partial\xi_k$.

The FE preconditioner, for which we retain the notation $\mathbf{B}_\diamond$, can be defined by means of the square bilinear elements as well. The function $\phi(s)$, $s\in(-1,1)$, can be replaced by a reasonably chosen piecewise constant function or by $\phi(s)=\max\left[\,c\hbar,\min\left(1+s,1-s\right)\right]$ with $c=\mathrm{const}$.

Theorem 7.6. *For the matrices* $\mathbf{B}_\diamond$ *and* $\mathfrak{B}_\diamond=\hbar^{-2}\mathbf{C}_I^{-1}\mathbf{B}_\diamond\mathbf{C}_I^{-1}$*, the spectral equivalence relations*

$$\hbar^{-2}\mathbf{B}_\diamond\asymp\mathbf{\Lambda}_\diamond\quad\textit{and}\quad\mathfrak{B}_\diamond\asymp\boldsymbol{\mathcal{A}}^{\hbar}_I$$

hold. If $\mathbf{A}_{I,\mathrm{sp}}$ *is generated with the help of the mesh with the GLL nodes, then we have the spectral equivalence*

$$\mathfrak{B}_\diamond\asymp\mathbf{A}_{I,\mathrm{sp}}\,.$$

Proof. The proof of the first of the equivalences, which we only outline, is produced in the standard finite element fashion. The matrix $\mathbf{\Lambda}_\dagger$ can be viewed as a finite element matrix. It can be assembled from the element stiffness matrices for the artificial finite elements specified on the triangles of the triangulation, which is used for defining $\mathbf{B}_\diamond$. The trianglewise comparison of the pares of the finite elements stiffness matrices, whose assemblages are the matrices $\mathbf{\Lambda}_\diamond$ and $\mathbf{B}_\diamond$, yields the result. While doing this, the Dirichlet boundary condition on $\partial\tau_0$ is taken into account. The second equivalence follows from the first one, Lemma 7.10 and Corollary 7.4. The step to the last equivalence is approved by Lemma 7.8. □

7.5 Schur Complement Preconditioners for Reference Elements

7.5.1 *Introductory Remarks*

In the case of the complete spaces $\mathcal{U} = \mathcal{P}_p, \mathcal{Q}_p$ on the triangular and square reference elements, respectively, the boundary Schur complement preconditioners for the both geometries can be defined with a good share of similarity. Let the stiffness matrices of reference elements are induced by the Dirichlet integral. There are at least two alternatives, which in general features resemble those developed in the h-version for subdomains of decompositions, defined by the triangular or quadrangular coarse meshes. In one, the Schur complement preconditioner is designed as a block diagonal preconditioner with $n_V + 1$ independent blocks, where n_V is the number of vertices of the reference element. Among these blocks, one is related to the degrees of freedom of n_V vertices and each of other n_V blocks is related to the degrees of freedom of one of the edges. At the best choice of these blocks, the growth of the relative condition number, bounded by $1+\log^2 p$, is insignificant, see Proposition 7.1. This bound is a consequence of the basic facts on the existence of the lifting operators $(\mathcal{U}|_{\partial\tau_0} \cap H^{1/2}(\partial\tau_0)) \to (\mathcal{U} \cap H^1(\tau_0))$, uniformly bounded in p, the trace theorem and some inequalities for polynomials from the space $\mathcal{U}$. In particular, these basic facts imply that a good boundary Schur complement preconditioner must be equivalent in the spectrum to the matrix of the quadratic form $|\cdot|^2_{1/2,\partial\tau_0}$, see, *e.g.,* the inequalities (7.67). In turn, one edge block of the Schur complement preconditioner can be defined up to a constant multiplier as the matrix spectrally equivalent to the matrix of the quadratic form ${}_{00}|\cdot|^2_{1/2,E_o}$, where E_o is a representative edge of the reference element.

Obviously, the preconditioner described above has the structure as the Schur complement preconditioner introduced by [Bramble *et al.* (1986)] for the 2d h-version BPS preconditioner. Another alternative is an adaptation to hp-version of the approach, described for the h-version in Chapter 4 and initially appeared in [Nepomnyaschikh (1991a)]. In it, the Schur complement preconditioner for the whole interface boundary results from an inexact iterative solver for the subsidiary preconditioner, which must be close in the spectrum to the Schur complement preconditioner and provide cheap matrix-vector multiplications. This preconditioner can be defined, *e.g.*, by the square of the broken weighted $H^{1/2}$-seminorm, see (4.10), in which domains of p-finite elements play the role of subdomains of decomposition.

In what follows, we will discuss the both alternatives and their computational costs. It is needless to stress that the Schur complement preconditioner is the component of DD algorithm of significant importance. Schur complement does not depend on the choice of coordinate functions in the subspace $\mathcal{U}_I \subset \mathcal{U}_I$. In the case of the complete polynomial spaces $\mathcal{U}$, from this and from the trace-continuation theorems it follows, that for construction of the preconditioner, we need only to know the traces of the boundary coordinate functions.

The problem of good reference element Schur complement preconditioning was studied by many authors, see, *e.g.*, [Ainsworth (1996)], [Ivanov and Korneev (1996)], [Guo and Cao (1997)], [Korneev and Jensen (1997, 1999)], [Korneev *et al.* (2002b,a)] and others. This subsection, primarily reflects approaches developed in [Ivanov and Korneev (1996)], [Korneev and Jensen (1999)] and [Korneev *et al.* (2002b)]. We also present results of numerical experiments which support the results of analysis.

For the illustration of the approaches to the Schur complement preconditioning, it is sufficient to consider, *e.g.*, the triangular reference elements. In what follows, $\wedge^{(k)}$, E_k, and s_k, respectively, denote the vertices of the reference triangle ordered in the counter-clockwise direction, the edge joining the vertices $\wedge^{(k-1)}$, $\wedge^{(k)}$, and the variable $s_k = (L_k - L_{k-1}) \in (-1, 1)$, where $L_k = L_k(x)$ for $k = 1, 2, 3$ are the baricentric coordinates. It is assumed that index k is changing modulo 3.

7.5.2 *Polynomial Bases on the Edges*

The bases in the spaces of traces of the finite element coordinate polynomials on the edges fall into two groups containing hierarchical and Lagrange interpolation type coordinate polynomials.

1. Boundary coordinate polynomials with hierarchical traces. The vertex coordinate functions for a hierarchical basis are linear polynomials

$$p_k(x) = L_k(x)\,, \quad k = 1, 2, 3\,, \tag{7.158}$$

which are barycentric coordinates in the triangle. The traces of the edge coordinate functions, which we order with the index $i = 1, 2, ..., p-1$, may be defined by different sets of polynomials. One of the choices is the set of the integrated Legendre polynomials $\mathcal{L}_i$ according to the relation

$$p_{k,i}(x)|_{E_l} = \begin{cases} \mathcal{L}_{i+1}(s_k(x))\,, & \text{if} \quad l = k\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{7.159}$$

For another choice, we may consider the polynomials

$$p_{k,i}(x)|_{E_l} = \begin{cases} (1 - s_k^2(x))s_k^{i-1}(x) \text{ , if } l = k \text{ ,} \\ 0 \text{ ,} \qquad\qquad\qquad\quad \text{otherwise .} \end{cases} \tag{7.160}$$

2. Lagrange polynomial interpolation basis. Let $s_k = s_{k,i} = \eta_i$, $i = 0, 1, \ldots, p$, be the GLL or GLC or other sets of nodes, see, *e.g.* (7.79) and (7.80), which are specified on the edge E_k, and

$$\phi_i(s_k) = \prod_{m=0,\, m\neq i}^{p} \frac{s_k - s_{k,m}}{s_{k,i} - s_{k,m}} \tag{7.161}$$

be the corresponding Lagrange interpolation basis. The traces of the reference element boundary coordinate functions may be defined by these polynomials, *i.e.,* for the edge coordinate functions, we set

$$p_{k,i}(x)|_{E_l} = \begin{cases} \phi_i(s_k(x)), \text{ if } l = k \\ 0, \qquad\quad \text{otherwise ,} \end{cases} \tag{7.162}$$

where $i = 1, 2, \ldots, p-1$. For the the vertex coordinate functions, we can adopt linear polynomials (7.158) or Lagrange interpolation coordinate polynomials

$$p_k(x)|_{E_l} = \begin{cases} \phi_p(s_k(x)) \text{ ,} & \text{if } l = k \text{ ,} \\ \phi_0(s_{k+1}(x)) \text{ ,} & \text{if } l = k+1 \text{ ,} \\ 0 \text{ ,} & \text{otherwise .} \end{cases} \tag{7.163}$$

7.5.3 *Edge Schur Complement Preconditioning for Hierarchical Coordinate Polynomials*

We will consider edge Schur complement preconditioners for the two described above types of traces of the hierarchical coordinate polynomials on edges. Both preconditioners can be represented in the factored form with simple matrices as factors which are sufficiently cheap in computations.

1. Simple polynomial basis. We look for the edge Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{00}$ in (7.20) in the case of the edge coordinate polynomials

with traces (7.160). Let us introduce the bases Ψ_k and Ψ^k, where

$$\Psi_1 = \begin{pmatrix} 1 \\ s_k \\ (1-s_k^2) \\ (1-s_k^2)s_k \\ \vdots \\ (1-s_k^2)s_k^{p-2} \end{pmatrix}, \quad \Psi^1 = \begin{pmatrix} 1 \\ s_k \\ s_k^2 \\ s_k^3 \\ \vdots \\ s_k^p \end{pmatrix}, \tag{7.164}$$

$$\Psi_2 = \begin{pmatrix} 1 \\ \cos\phi \\ (\cos\phi)^2 \\ \vdots \\ (\cos\phi)^p \end{pmatrix}, \quad \Psi^2 = \begin{pmatrix} 1 \\ \cos\phi \\ \cos 2\phi \\ \vdots \\ \cos p\phi \end{pmatrix},$$

the matrices $\mathbf{C_1}$ and $\mathbf{C}_2$ of the transformations $\Psi_1 = \mathbf{C}_1\Psi^1$, $\Psi_2 = \mathbf{C}_2\Psi^2$, and the diagonal matrix $\boldsymbol{\mathcal{D}}_{1/2} = \mathrm{diag}\,[0,\, 1, \ldots, p]$. Obviously, for $\mathbf{C}_1$, we have

$$\mathbf{C}_1 = \begin{pmatrix} 1 & & & & & \\ 0 & 1 & & \mathbf{0} & & \\ 1 & 0 & -1 & & & \\ & & \cdots & & & \\ & \mathbf{0} & 1 & 0 & -1 & \\ & & & 1 & 0 & -1 \end{pmatrix}.$$

Furthermore, $\mathbf{C}_2$ is a known lower triangular matrix. We define $\boldsymbol{\mathcal{S}}_{00}$ by crossing out two first rows and columns from the $(p+1)\times(p+1)$ matrix

$$\mathbb{E} = \mathbf{C}_1\mathbf{C}_2\boldsymbol{\mathcal{D}}_{1/2}\mathbf{C}_2^\top\mathbf{C}_1^\top. \tag{7.165}$$

We note that, while $\mathbb{E}$ is singular, $\boldsymbol{\mathcal{S}}_{00}$ is not.

The cost of solving the systems with the matrices $\mathbf{C}_1$ and $\mathbf{C}_1^\top$ is $\mathcal{O}(p)$. Since $\mathbf{C}_2$ is a lower triangular matrix, solving systems with the matrices $\mathbf{C}_2$ and $\mathbf{C}_2^\top$ without use of any special algorithms, but by the general simple elimination procedure requires not more than $\mathcal{O}(p^2)$ operations. Matrix $\boldsymbol{\mathcal{S}}_{00}$ itself is not represented in the factored form similar to (7.165). However, (7.165) may be used in the solution procedure for the system $\boldsymbol{\mathcal{S}}_{00}\mathbf{v} = \mathbf{f}$ in such a way that the computational cost of solving will not exceed the cost of solving the system with the system matrix $\mathbf{C}_2$ with respect to the order. This simple procedure is described below.

The use of (7.165) slightly simplifies, if we set $\boldsymbol{\mathcal{D}}_{1/2} = \mathrm{diag}\,[\,(1+k)\,]_{k=0}^{p}$. Such a choice is justified by the estimate

$$ {}_{00}|v_{E_k}|^2_{1/2,E_k} \succ \|v_{E_k}\|^2_{0,E_k}\,, \quad \forall\, v \in H^{1/2}_{00}(E_k)\,, $$

from which the inequalities

$$ {}_{00}\|v_{E_k}\|^2_{1/2,E_k} \prec {}_{00}|v_{E_k}|^2_{1/2,E_k} \prec {}_{00}\|v_{E_k}\|^2_{1/2,E_k}\,, \quad \forall\, v \in H^{1/2}_{00}(E_k)\,, $$

can be derived. Now, the matrix $\mathbb{E}$ is positive definite. Therefore, instead of the system $\boldsymbol{\mathcal{S}}_{00}\mathbf{v}_E = \mathbf{f}_E,, \mathbf{v}_E = \{v_i\}_{i=2}^{p}$, we solve the system

$$ \mathbb{E}\mathbf{v} = \mathbf{f}\,, \tag{7.166} $$

where the solution vector has the form $\mathbf{v}_E = \{v_i\}_{i=0}^{p}$, and the right-hand side can be represented in the form $\mathbf{f} = (f_0, f_1, \mathbf{f}_E^{\top})^{\top}$, where the components f_0 and f_1 are unknown. Then, we find f_0 and f_1 from the 2×2 system

$$ v_0 = \{\mathbb{E}^{-1}\mathbf{f}\}_0 = 0 \quad \text{and} \quad v_1 = \{\mathbb{E}^{-1}\mathbf{f}\}_1 = 0\,, $$

expressing the boundary conditions. The computation of vector $\mathbf{v}_E$ is concluded by substituting the values of f_0 and f_1 in the expressions for $v_i = \{\mathbb{E}^{-1}\mathbf{f}\}_i$, $i = 2,3,\ldots,p$, that were found at the first step by the solution of the system (7.166). For the choice $\boldsymbol{\mathcal{D}}_{1/2} = \mathrm{diag}\,[0,\,1,\ldots,p\,]$, the computations are slightly different. An additional equation arises, expressing the solvability condition for systems with singular matrix $\boldsymbol{\mathcal{D}}_{1/2}$, accompanied by appearing of a third parameter to be defined from the boundary conditions. This parameter takes into account that the solution of the system with the matrix $\boldsymbol{\mathcal{D}}_{1/2}$ is not unique.

2. Integrated Legendre polynomials basis for edge traces. In a similar way, the preconditioner $\boldsymbol{\mathcal{S}}_{00}$ is defined, when traces of edge coordinate functions are the integrated Legendre polynomials as in (7.159). We use the bases

$$ \Psi_1 = \begin{pmatrix} 1 \\ s_k \\ \mathcal{L}_2(s_k) \\ \vdots \\ \mathcal{L}_p(s_k) \end{pmatrix}, \quad \Psi^1 = \begin{pmatrix} 1 \\ s_k \\ P_2(s_k) \\ \vdots \\ P_p(s_k) \end{pmatrix}, \quad \Psi_2 = \begin{pmatrix} 1 \\ \cos\phi \\ P_2(\cos\phi) \\ \vdots \\ P_p(\cos\phi) \end{pmatrix}, \tag{7.167} $$

and the matrices $\mathbf{C}_k$ of the respective transformations

$$ \Psi_1 = \mathbf{C}_1\Psi^1, \quad \Psi_2 = \mathbf{C}_2\Psi^2\,, \tag{7.168} $$

where Ψ^2 is the same as in (7.164). The preconditioner $\boldsymbol{\mathcal{S}}_{00}$ is obtained, if we cross out the two first rows and columns from the matrix of the form

(7.165) with the same matrix $\boldsymbol{\mathcal{D}}_{1/2}$ and the matrices $\mathbf{C}_k$, defined above. As in the previous case, the systems of algebraic equations governed by the matrices $\mathbf{C}_k$ are easy to solve. Indeed, the first matrix has the form

$$\mathbf{C}_1 = \begin{pmatrix} 1 & & & & \\ 0 & 1 & & \mathbf{0} & \\ -\frac{\beta_2}{3} & 0 & \frac{\beta_2}{3} & & \\ 0 & -\frac{\beta_3}{5} & 0 & \frac{\beta_3}{5} & \\ & & \cdots & & \\ & & \mathbf{0} & -\frac{\beta_p}{2p-1} \; 0 & \frac{\beta_p}{2p-1} \end{pmatrix},$$

whereas the second one is a lower triangular matrix with the entries specified by the expressions for the Legengre polynomials $P_j(\cos\phi)$ in elements of the basis Ψ^2. These expressions may be found, *e.g.,* in [Gradshteyn and Ryzhik (1979)], cf. p. 1025. Clearly, other ways of transformation from the basis Ψ_1 to Ψ^2 of (7.167) can be used. For Ψ^1 and Ψ_2 in (7.168), we can take the vectors of (7.164) under the same notations and use the respective transformations.

Traces of the edge coordinate polynomials on the edges of the square reference elements may have the same form (7.159) or (7.160). In these cases the edge Schur complement preconditioners $\boldsymbol{\mathcal{S}}_{00}$ may be exactly the same as described above. For instance, (7.159) are the traces of edge polynomials $L_{\boldsymbol{\alpha}}(x) = \mathcal{L}_{\alpha_1}(x_1)\mathcal{L}_{\alpha_2}(x_2)$, $\boldsymbol{\alpha} \in \omega_E$, of the hierarchical reference element $\mathcal{E}_H$, introduced in Subsection 7.3.1. Therefore, the matrix $\boldsymbol{\mathcal{S}}_{00}$, defined above for the traces (7.159), may be used for the preconditioner with the same — up to some generic constant — efficiency in computational work, as for the triangular elements.

Proposition 7.1. *Let traces of the edge coordinate polynomials of a complete triangular or square reference element be defined by (7.159) or (7.160), $n_E = 3, 4$ be number of edges of the reference element. Let also the one edge Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{00}$ be defined as above, and $\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}$ be defined as in (7.20). Then the spectral inequalities*

$$\frac{\underline{c}_E}{(1+\log p)^2}\,\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \le \mathbb{S}_E \le \overline{\beta}\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \;\; \textit{and} \;\; \frac{1}{(1+\log p)^2}\,\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}} \prec \mathbb{S}_B \quad (7.169)$$

are valid, where $\underline{c}_E$ and $\overline{\beta}$ are some positive generic constants.

Proof. The method of K-interpolation allows us to conclude that the matrix $\boldsymbol{\mathcal{D}}_{1/2}$ represents the norm ${}_{00}|v|^2_{1/2,E_k}$ in the basis Ψ^2 for polynomials v vanishing at the ends of E_k, *i.e.*,

$${}_{00}|v_{E_k}|^2_{1/2,E_k} \asymp \mathbf{v}^T_{E_k}\,\boldsymbol{\mathcal{S}}_{00}\mathbf{v}_{E_k}\,, \quad \forall \mathbf{v}_{E_k} \leftrightarrow v_{E_k} \in \mathcal{U}_E|_{E_k}\,, \qquad (7.170)$$

that corresponds to (7.21). Therefore, the last inequality (7.169) follows from the equivalence relation (7.170) and from the inequalities (7.29) and (7.31) of Theorem 7.1, p. 209. The definitions of the matrices $\mathbb{S}_E$ and $\boldsymbol{\mathcal{S}}_E$, Theorem 7.2, the Cauchy inequality and (7.170) result in the inequalities

$$\begin{aligned}\mathbf{v}_E^T\,\mathbb{S}_E\mathbf{v}_E &\le |\mathcal{P}_{B,\mathrm{ref}}\,(v_E|_{\partial\tau_0})|^2_{1,\tau_0} \prec |v_E|^2_{1/2,\partial\tau_0}\\ &\le 3\sum_{k=1}^{n_e} {}_{00}|v_E|^2_{1/2,E_k} \asymp 3\mathbf{v}_E^\top\,\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}\mathbf{v}_E\,, \end{aligned} \tag{7.171}$$

proving the right inequality in the first pair of inequalities (7.169). By virtue of (7.40), the trace theorem for functions from the space $H^1(\tau_0)$, and the definition of $\mathbb{S}_E$, we also get

$$\begin{aligned}\mathbf{v}_E^\top\,\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}\,\mathbf{v}_E &\asymp \sum_{k=1}^{n_e} {}_{00}|v_E|^2_{1/2,E_k} \prec (1+\log p)^2\sum_{k=1}^{n_e}|v_E|^2_{1/2,E_k}\\ &\le (1+\log p)^2|v_E|^2_{1/2,\partial\tau_0} \prec (1+\log p)^2\mathbf{v}_E^T\,\mathbb{S}_E\,\mathbf{v}_E\,,\end{aligned} \tag{7.172}$$

completing the proof. □

The use of the Schur complement preconditioners, introduced in this section, results in the DD preconditioners (7.27), whose preconditioning efficiency is estimated in the following corollary.

Corollary 7.5. *Let $\boldsymbol{\mathcal{K}}$ be the DD preconditioner (7.27) with the single edge reference element Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{00}$ defined above. Let also the prolongation $\mathcal{P}_{E,\mathrm{ref}}$ be defined by the restriction of $\mathcal{P}_{B,\mathrm{ref}}$ such that (7.22)–(7.24) hold. Then the inequalities*

$$\underline{\gamma}\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{\gamma}\boldsymbol{\mathcal{K}} \tag{7.173}$$

hold, with the bounds of the relative spectrum

$$\underline{\gamma} = \underline{c}\,c_{0,\mathcal{P}}^{-1}\min\,(\underline{\beta}_I, \frac{\underline{c}_E}{(1+\log p)^2}) \quad \textit{and} \quad \overline{\gamma} = \overline{c}\,c_{0,\mathcal{P}}\max(\overline{\beta}_I, \overline{\beta}_E)\,. \tag{7.174}$$

Proof. Theorem 7.2 and inequalities (7.170) make the assumptions of Theorem 7.1 fulfilled, under which inequalities (7.173) are equivalent to (7.28) with the values $\underline{\gamma}$ and $\underline{\beta}_{EB}$ from (7.30) and (7.31), respectively. □

3. Numerical experiments.

For a significant range $10 \le p \le 200$ of powers of polynomials in the edge subspace of traces of the reference element, numerical results are in a good agreement with the estimate

$$\mathrm{cond}\,[\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}\,\mathbb{S}_E^{-1}] \prec (1+\log p)^2$$

of Lemma 7.1. Apparently, this range covers a considerable part of the asymptotic zone of the value cond $[\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}\,\mathbb{S}_E^{-1}]$. In support of this conclusion, we reproduce two graphs which were obtained for the hierarchical square reference element $\mathcal{E}_H$ with the coordinate functions forming the set $\mathcal{M}_p$. They show dependence on p of condition numbers for

- the Schur complement $\mathbb{S}_E$, Figure 7.6, and for
- the preconditioned Schur complement, *i.e.,* for the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}^{-1}\mathbb{S}_E$, Figure 7.7.

It is worth emphasizing that in these two graphs scales along vertical axes differ in 100 times. If we scale the second graph as the first one, then it will be hardly distinct from the abscissa. The graphs are taken from the papers by [Anufriev *et al.* (2003)] and [Anufriev and Korneev (2005)], where numerical results testing other components of the DD preconditioner of finite element methods with p-elements and DD preconditioner itself are also discussed.

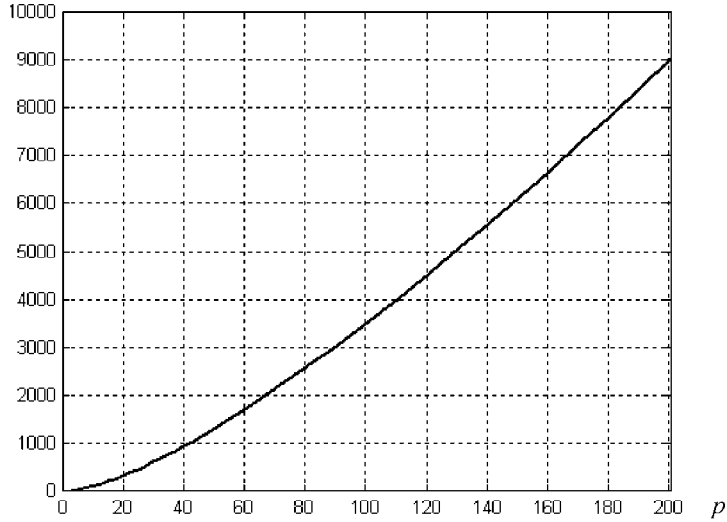

Fig. 7.6 Schur complement condition number cond $[\mathbb{S}_E]$ in dependence on p.

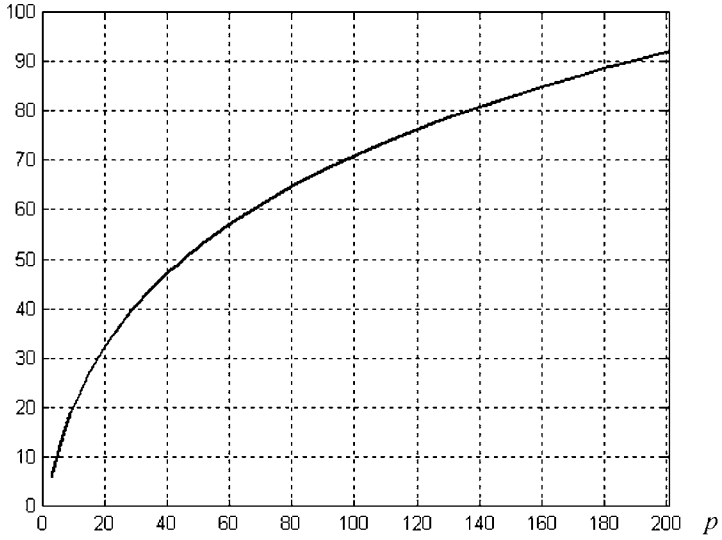

Fig. 7.7 Relative Schur complement condition number cond $[\boldsymbol{\mathcal{S}}_{E,\mathrm{ref}}^{-1}\mathbb{S}_E]$ in dependence on p.

7.5.4 *Inter-element Boundary Schur Complement Preconditioning, Lagrange Interpolation Edge Coordinate Functions*

As was shown in the preceding subsection, splitting vertex degrees of freedom from the rest in DD preconditioner results in the factor $(1 + \log p)$ in the relative condition numbers of the Schur complement and the DD preconditioners. It is replaced by the worse factor $(1 + \log p)^2$, if different edges are split in the edge Schur complement preconditioner. There are a few ways to avoid these losses. In this subsection, one of them is implemented to the finite element assemblage generated with the use of the reference elements, whose edge coordinate polynomials have traces on the edges, represented in the *nodal* basis. More precisely, it is assumed that these traces are Lagrange interpolation basis polynomials over the sets of the GLC nodes. Such discretizations afford the possibility to derive efficient Schur complement preconditioners which look quite similar to those being commonly used the h-version.

Let us assume that the traces of the reference element are defined by relationships (7.161)–(7.163), and enumerate the $n_{_E}p$ nodes on the boundary of the reference element counter-clockwise, starting from the lower left

vertex ($\wedge_1$, if τ_0 is the triangle). For the boundary Schur complement preconditioner, we take the $n_{_E}p \times n_{_E}p$ matrix

$$\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} := \boldsymbol{\Delta}^{1/2}_{\partial\tau_0} = \begin{pmatrix} 2 & -1 & & & -1 \\ -1 & 2 & -1 & & \\ & \cdots & & & \\ & & \cdots & & \\ & & -1 & 2 & -1 \\ -1 & & & -1 & 2 \end{pmatrix}^{1/2} . \tag{7.175}$$

Theorem 7.7. *Let, for the triangular or square reference element, the polynomial space be $\mathcal{U} = \mathcal{P}_p$ or $\mathcal{Q}_p$, respectively, and the traces on $\partial\tau_0$ of the boundary coordinate functions be defined by (7.161)–(7.163) for GLC nodes (7.80). Then the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ of (7.175) satisfies the spectral equivalence inequalities*

$$\underline{\beta}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \le \mathbb{S}_B \le \overline{\beta}_B \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} , \tag{7.176}$$

with positive generic spectral equivalence constants $\underline{\beta}_B$ and $\overline{\beta}_B$.

Proof. The matrix $\mathbb{S}_{1/2,j} = \boldsymbol{\Delta}^{1/2}_{\partial\Omega_j}$ in (4.20), p. 80, and the matrix $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} = \boldsymbol{\Delta}^{1/2}_{\partial\tau_0}$ coincide, if the numbers of nodes on their boundaries are equal. Besides, according to the inequalities (4.19) and the inequalities (7.177) of Lemma 7.11, proved below, the seminorms induced by these matrices are equivalent to the seminorm $|\cdot|_{1/2,\partial\tau_0}$ for continuous piecewise linear functions on $\partial\tau$ over a uniform mesh. Application of the equivalence inequalities (7.67), p. 220, concludes the proof. □

For definiteness, we assume below that τ_0 is the reference triangle. We now retain the notation η_α for GLC nodes , where $\alpha = (\alpha_1, \alpha_2) \in \omega_B$ and ω_B is the "boundary" subset of the set $\omega := \{\alpha : 0 \le \alpha_1, \alpha_2, (\alpha_1 + \alpha_2) \le p\}$. Let also $\widehat{\eta}_\alpha$ be the nodes, uniformly distributed on $\partial\tau_0$ and dividing each edge in p equal parts, and $\mathcal{H}^\hbar(\partial\tau_0)$ be the space of functions continuous on $\partial\tau_0$ and linear between these nodes. The following lemma can be found in [Korneev and Jensen (1999)].

Lemma 7.11. *For any $v \in \mathcal{P}_p$ and $w \in \mathcal{H}^\hbar(\partial\tau_0)$ such that*

$$v(\eta_\alpha) = w(\widehat{\eta}_\alpha) , \quad \forall\ \eta_\alpha \in \partial\tau_0 ,$$

the norms $\|\cdot\|_{1/2,\partial\tau_0}$ as well as seminorms $|\cdot|_{1/2,\partial\tau_0}$ are equivalent uniformly in p, i.e.,

$$\|w\|_{1/2,\partial\tau_0} \asymp \|v\|_{1/2,\partial\tau_0} \quad \textit{and} \quad |w|_{1/2,\partial\tau_0} \asymp |v|_{1/2,\partial\tau_0} . \tag{7.177}$$

Proof. For each side, we define a mapping $E_k \to E_k$ of the form

$$s_k = -\cos\frac{\pi}{2}(t_k+1)\,, \quad s_k, t_k \in [-1,1]\,, \tag{7.178}$$

with the coordinate s_k for the points on the side E_k introduced at the beginning of Section 7.5, see p. 256. The corresponding mapping $\partial\tau_0 \to \partial\tau_0$, which is denoted as $x = \mathcal{Y}(y)$ and is defined by (7.178) on each edge, is continuous on $\partial\tau_0$, and, as we show below, preserves the norm $\|\cdot\|_{1/2,\partial\tau_0}$. This means that, if $v(x) \in \mathcal{P}_p|_{\partial\tau_0}$ and $v_{\cos}(y) = v(\mathcal{Y}(y))$, then

$$\|v\|_{1/2,\partial\tau_0} \asymp \|v_{\cos}\|_{1/2,\partial\tau_0}\,. \tag{7.179}$$

In order to prove this, we first of all note that the equivalence

$$\|v\|_{1/2,E_k} \asymp \|v_{\cos}\|_{1/2,E_k}$$

is a consequence of Lemma 7.3. Thus, it remains to consider only the terms in $\|\cdot\|_{1/2,\partial\tau_0}$ that are additional to the sum of $\|\cdot\|_{1/2,E_k}$ over $k = 1,\dots,n_e$. These terms correspond to the vertices of the reference element, and it is sufficient to consider one vertex, *e.g.*, $\wedge^{(1)}$. The additional integral related to this vertex is

$$\int_{-1}^{0} \frac{(v_1(-s_2) - v_2(s_2))^2}{(1+s_2)} ds_2 = \int_0^1 \frac{(v_1(1-t) - v_2(t-1))^2}{t} dt\,,$$

where $v_k(s_k)$ is the restriction of the function $v(x)$ to the side E_k. According to (7.178),

$$t = 1 + s_2 = 2(\sin\frac{\pi}{4}(t_2+1))^2 = 2(\sin\frac{\pi}{4}\bar{t})^2,$$

where t and $\bar{t}$ are the distances from the vertex $\wedge^{(1)}$ after and before the mapping $\mathcal{Y}(y)$. Taking into account the elementary inequalities

$$\frac{1}{2}\bar{t} \le \sin\frac{\pi}{4}\bar{t} \le \frac{\pi}{4} \quad \text{and} \quad \frac{1}{\sqrt{2}} \le \cos\frac{\pi}{4}\bar{t} \le 1\,,$$

we obtain

$$\begin{aligned}\int_0^1 \frac{(v_1(1-t) - v_2(t-1))^2}{t}\, dt &\le \int_0^1 \frac{(\tilde{v}_1(1-\bar{t}) - \tilde{v}_2(\bar{t}-1))^2}{\sin\frac{\pi}{4}\bar{t}}(\cos\frac{\pi}{4}\bar{t})\, d\bar{t} \\ &\asymp \int_0^1 \frac{(\tilde{v}_1(1-t) - \tilde{v}_2(t-1))^2}{t}\, dt\,,\end{aligned}$$

with the abbreviation $\tilde{v}_1 = v_{\cos,1}$ and $\tilde{v}_2 = v_{\cos,2}$. Consequently, the equivalence (7.179) takes place.

Since the mapping $x = \mathcal{Y}(y)$ maps the nodes $\widehat{\eta}_\alpha$ of the uniform grid onto GLC nodes η_α, we have

$$v_{\cos}(\widehat{\eta}_\alpha) = v(\eta_\alpha) = w(\widehat{\eta}_\alpha)\,. \tag{7.180}$$

For any pair $v_{\cos}$ and w specified on $\partial\tau_0$ and coinciding at the nodes of the uniform grid, the inequalities

$$\|w\|_{1/2,\partial\tau_0} \prec \|v_{\cos}\|_{1/2,\partial\tau_0} \prec \|w\|_{1/2,\partial\tau_0} \tag{7.181}$$

hold. Indeed, it is easy to derive the inequalities

$$\|w\|_{k,\partial\tau_0} \prec \|v_{\cos}\|_{k,\partial\tau_0} \prec \|w\|_{k,\partial\tau_0} \tag{7.182}$$

for $k = 0, 1$. For the 1d case, *i.e.,* for any side of τ_0, they can be found in [Ivanov and Korneev (1995)], see also [Bahlmann and Korneev (1993)]. Inequalities (7.181) are a consequence of (7.182), and can be derived by means of the K-interpolation technique between the spaces $L_2(\partial\tau_0) = H^0(\partial\tau_0)$ and $H^1(\partial\tau_0)$. Now, (7.179) and (7.181) lead to the first relationship in (7.177). The proof of the second one is the same, if we take into account that the equivalences, used above, also hold for the seminorms. □

7.5.5 *Numerical Complexity of Schur Complement Algorithms*

As in the preceding subsection, we assume that on each edge of the boundary $\partial\tau_0$ traces of the boundary coordinate polynomials are Lagrange interpolation coordinate polynomials over the set of GLC quadrature nodes, see (7.161)–(7.163). If the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_B$ is assembled from the matrices

$$\boldsymbol{\mathcal{S}}_r = \varrho_r \boldsymbol{\mathcal{S}}_{B,\mathrm{ref}} \tag{7.183}$$

with $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ defined by (7.175), then, according to Theorem 7.7, it will be spectrally equivalent to the Schur complement $\mathbf{S}_B$, *i.e.*,

$$\boldsymbol{\mathcal{S}}_B \asymp \mathbf{S}_B\,. \tag{7.184}$$

The preconditioner $\boldsymbol{\mathcal{S}}_B$, which is defined in the same way as in the h version, is efficient for the matrix-vector multiplications. However, special algorithms for solving systems of algebraic equations with the system matrix $\boldsymbol{\mathcal{S}}_B$ are required. A simple choice is to use the inexact solver of simple iterations with Chebyshev iteration parameters in order to obtain the preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$ like the one in (4.22) and (4.27). As such a preconditioner, we can consider $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}} = \boldsymbol{\mathcal{I}}\,[\boldsymbol{\mathcal{S}}_B, \mathbb{D}_B, \nu_{1/2}]$, where $\mathbb{D}_B = \lfloor \boldsymbol{\mathcal{S}}_B \rfloor_{\mathrm{diag}}$ is the diagonal matrix with the diagonal entries of $\boldsymbol{\mathcal{S}}_B$, or another matrix $\boldsymbol{\mathcal{D}}_B$, which is spectrally equivalent to $\mathbb{D}_B$, *i.e.,* $\boldsymbol{\mathcal{D}}_B \asymp \mathbb{D}_B$.

Let us suppose, for simplicity, that

$$0 < \underline{\mu} \le \varrho(x) \le \overline{\mu}\,, \quad \forall\, x \in \Omega\,,$$

with the positive constants $\underline{\mu}$ and $\overline{\mu}$ sufficiently close to unity and that the triangulation is quasiuniform. Under the conditions above, ϱ_r can be understood as the mean value of $\varrho_r(x)$ over τ_r. For some types of discretizations, we are able to get the bound

$$\operatorname{cond}\,[\mathbb{D}_B^{-1}\boldsymbol{\mathcal{S}}_B] = \mathcal{O}(p)$$

with the constant depending on Ω and the constants in the generalized conditions of the quasiuniformity (3.2) or (3.3). For instance, in view of the definition of $\boldsymbol{\mathcal{S}}_B$ by (7.183), (7.175), it holds, if the discretization mesh is topologically equivalent to the orthogonal mesh. From this bound, it follows that it is necessary to perform $n_{1/2} \le c\sqrt{p}$ iterations in order to produce the preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$ which is spectrally equivalent to $\boldsymbol{\mathcal{S}}_B$ and $\mathbf{S}_B$. The preconditioners $\boldsymbol{\mathcal{S}}_B$ and $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$ have the forms identical to the preconditioners (4.22) and (4.27) for the h-version on the corresponding mesh of nodes on the interface boundary, *i.e.,* τ_r, h and p correspond to Ω_j, H_j and $(N_{\partial\Omega_j}/4) \sim H_j/h$ for the h-version, respectively. As a consequence, proofs of the above bounds can be produced in a similar way to the proofs of their counterparts in the case of the h-version.

Suppose now that the internal unknowns have been eliminated, and we need to solve the system with the Schur complement

$$\mathbf{S}_B\mathbf{u}_B = \mathbf{f}\,. \tag{7.185}$$

If we use the PCGM with the preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$, the total arithmetical cost can be estimated by

$$c_1\mathcal{R}(p^2 + p^{3/2}\log p)\log\varepsilon^{-1} \tag{7.186}$$

with the constant c_1 depending on the domain and the values of $\alpha^{(1)}, \theta$ in the quasiuniformity conditions for the FE mesh. In this bound, the first and second terms account for the matrix-vector multilication by the Schur complement $\mathbf{S}_B$ and for the operation of the inexact solution of the system with the preconditioner $\boldsymbol{\mathcal{S}}_B$, respectively. The latter operation realizes the multiplication of a vector by the implicitly defined $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}^{-1}$, in which the matrix-vector multiplications by $\boldsymbol{\mathcal{S}}_B$ are performed by means of FDFT applied elementwise. The notation ε stands for the prescribed accuracy of the PCG iteration with respect to the norm induced by $\mathbf{S}_B$.

In the DD algorithm, multiplications by $\mathbf{S}$ are absent. If one uses the described Schur complement preconditioner and optimal solvers for Dirichlet problems on subdomains of decomposition, the estimate (7.186) of the total computational work retains. Therefore, the DD algorithm is optimal,

but now the main term in (7.186) corresponds to the solution the local Dirichlet problems.

In view of the similarity with the h-version, established above, a few other techniques more developed in the frame of the h-version can be attracted for the efficient inter-subdomain Schur complement preconditioning and solution of the inter-subdomain Schur complement problem. For example, in the latter algorithm, the computation of Schur complement $\mathbf{S}_B$ and the matrix-vector multiplications by this matrix are executed and can be expensive. These multiplications are costly even if performed elementwise, *i.e.* by means of the multiplications by Schur complements $\mathbf{S}_{B_r}$ for finite elements. Cost reduction of these operations can be achieved with the help of algorithms which were originally developed for the boundary element methods, such as $\mathcal{H}$-matrices and tensor-train decomposition techniques. We refer to the papers by [Khoromskij and Wittum (1999)], [Hsiao *et al.* (2001)], [Hackbusch *et al.* (2005)], [Bebendorf (2008)], [Hackbusch (2009, 2012)], [Khoromskij (2011)], and [Dolgov *et al.* (2011, 2012)], which are a few representatives of this vast area of research.

7.5.6 *Schur Complement Preconditioner with Alternative Inexact Solvers*

The preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$ was defined by means of the preconditioner-solver $\mathbb{D}_B$ and the preconditioner-multiplicator $\boldsymbol{\mathcal{S}}_B$. Obviously, instead of $\mathbb{D}_B$ a more efficient preconditioner-solver can be used. One of them results from the approach traditionally used for the discretizations by means of the hierarchical reference elements. After transformation of vertex degrees of freedom to the bilinear coordinate polynomials, this preconditioner becomes block diagonal with the decoupled blocks related to the subspace of the vertex degrees of freedom and subspaces of the degrees of freedom of each edge. The major bulk of computations is not influenced by jumps of the coefficient ϱ on the inter-element boundaries, if it is elementwise constant or varies not too much over each finite element. Only the vertex component, of much smaller dimension than the edge component, bears full responsibility for taking into account the influence of the jumps.

Let us use the notations Φ and Ψ for the basis (7.161)–(7.163) and for the basis with the same edge polynomials, as in Φ, but with vertex polynomials which are the traces of the linear/bilinear vertex polynomials for triangular/rectangular reference elements, respectively. By $\mathbf{C}_B$, we denote the transformation matrix such that $\mathbf{S}_B = \mathbf{C}_B^\top \widetilde{\mathbf{S}}_B \mathbf{C}_B$ with the Schur com-

plements $\mathbf{S}_B$ and $\widetilde{\mathbf{S}}_B$ corresponding to the bases Φ and Ψ, respectively. Let $\widetilde{\mathbf{A}}_V$ be the vertex block of the reference element stiffness matrix related to Ψ, matrix $\widetilde{\boldsymbol{\mathcal{K}}}_V$ be assembled from the element matrices $\widetilde{\boldsymbol{\mathcal{K}}}_{V_r} = \varrho_r \widetilde{\mathbf{A}}_V$, $\boldsymbol{\mathcal{S}}_E$ be defined as in (7.18), (7.20) and (7.25) with the $(p-1)\times(p-1)$ matrix

$$\boldsymbol{\mathcal{S}}_{00} = (\mathrm{tridiag}\,[-1,2,-1])^{1/2}\,.$$

Finally, we introduce the matrices

$$\begin{gathered}\boldsymbol{\mathcal{S}}_{B,1} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \widetilde{\boldsymbol{\mathcal{K}}}_V]\,, \qquad \boldsymbol{\mathcal{S}}_{B,2} = \mathbf{C}_B^\top \boldsymbol{\mathcal{S}}_{B,1}\mathbf{C}_B\,,\\ \boldsymbol{\mathcal{S}}_{B,\mathrm{it}} = \boldsymbol{\mathcal{I}}\,[\boldsymbol{\mathcal{S}}_B, \boldsymbol{\mathcal{S}}_{B,2}, \nu_{1/2}]\,.\end{gathered} \tag{7.187}$$

For the iterative solution of the system (7.185), we can use the preconditioner-solvers $\boldsymbol{\mathcal{S}}_{B,2}$ or $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$. However, the latter will be cheaper, especially if it is used in the DD solver. In comparison with the algorithm presented in the previous subsection it diminishes only the second term in (7.186), *c.f.*, Proposition 7.2.

An equivalent, but a little more costly procedure is the following. We transform (7.185) to the system

$$\widetilde{\mathbf{S}}_B \widetilde{\mathbf{u}}_B = \widetilde{\mathbf{f}}\,, \tag{7.188}$$

where

$$\widetilde{\mathbf{S}}_B = \mathbb{C}_B^\top \mathbf{S}_B \mathbb{C}_B\,, \quad \mathbf{u}_B = \mathbb{C}_B \widetilde{\mathbf{u}}_B\,, \quad \widetilde{\mathbf{f}} = \mathbb{C}_B^\top \mathbf{f}\,, \quad \mathbb{C}_B = \mathbf{C}_B^{-1}\,. \tag{7.189}$$

For iterative solution of (7.188), we can use the preconditioner-solvers $\boldsymbol{\mathcal{S}}_{B,1}$ or $\widetilde{\boldsymbol{\mathcal{S}}}_{B,\mathrm{it}} = \boldsymbol{\mathcal{I}}\,[\widetilde{\boldsymbol{\mathcal{S}}}_B, \boldsymbol{\mathcal{S}}_{B,1}, \nu_{1/2}]$ with

$$\widetilde{\boldsymbol{\mathcal{S}}}_B = \mathbb{C}_B^\top \boldsymbol{\mathcal{S}}_B \mathbb{C}_B\,. \tag{7.190}$$

Now, it is easy to conclude that

$$\frac{1}{1+\log^2 p}\underline{c}_B\, \boldsymbol{\mathcal{S}}_{B,2} \le \boldsymbol{\mathcal{S}}_B \simeq \mathbf{S}_B \le \overline{c}_B\, \boldsymbol{\mathcal{S}}_{B,2} \tag{7.191}$$

with the positive constants $\underline{c}_B$ and $\overline{c}_B$, depending only on the generalized conditions of shape regularity. In particular, in view of (7.187), (7.189), (7.190) and the fact that $\mathbf{C}_B$ and $\mathbf{C}^{-1}$ are uniquely defined, the spectral equivalence inequalities (7.191) are a direct consequence of the inequalities

$$\frac{1}{1+\log^2 p}\underline{c}_B\, \boldsymbol{\mathcal{S}}_{B,1} \le \widetilde{\boldsymbol{\mathcal{S}}}_B \simeq \widetilde{\mathbf{S}}_B \le \overline{c}_B\, \boldsymbol{\mathcal{S}}_{B,1}\,, \tag{7.192}$$

to which they are equivalent. The equivalence $\boldsymbol{\mathcal{S}}_B \asymp \mathbf{S}_B$ is given in (7.184), and $\widetilde{\boldsymbol{\mathcal{S}}}_B \asymp \widetilde{\mathbf{S}}_B$ is the same equivalence, but written in different basis. The preconditioner $\boldsymbol{\mathcal{S}}_{B,1}$ is the same as the Schur complement preconditioner

$\boldsymbol{\mathcal{S}}_B$ in (7.33) and (7.34) on page 210. Thus, the left inequality in (7.192) follows from (7.29), (7.31), the definition of $\boldsymbol{\mathcal{S}}_{B,1}$ similarly with (7.33), (7.34) and the generalized conditions of shape regularity. The right inequality (7.192) is the consequence of Cauchy's inequality, the inequality $\mathbf{S}_E \prec \boldsymbol{\mathcal{S}}_E$, see (7.19) and Theorem 7.1, and the spectral equivalences of $\widetilde{\mathbf{K}}_V$, $\widetilde{\boldsymbol{\mathcal{K}}}_V$ and the vertex block of $\widetilde{\boldsymbol{\mathcal{S}}}_B$.

Inequalities (7.191) and (7.192) allow us to conclude that, for $\nu_{1/2} = \mathcal{O}(\log p)$, we have the spectral equivalences inequalities

$$\begin{aligned} \underline{\gamma}_{B,\mathrm{it}}\, \widetilde{\boldsymbol{\mathcal{S}}}_{B,\mathrm{it}} \le \widetilde{\mathbf{S}}_B \asymp \widetilde{\boldsymbol{\mathcal{S}}}_B \le \overline{\gamma}_{B,\mathrm{it}}\, \widetilde{\boldsymbol{\mathcal{S}}}_{B,\mathrm{it}}\,, \\ \underline{\gamma}_{B,\mathrm{it}}\, \boldsymbol{\mathcal{S}}_{B,\mathrm{it}} \le \mathbf{S}_B \asymp \boldsymbol{\mathcal{S}}_B \le \overline{\gamma}_{B,\mathrm{it}}\, \boldsymbol{\mathcal{S}}_{B,\mathrm{it}}\,, \end{aligned} \tag{7.193}$$

with positive spectral equivalence constants $\underline{\gamma}_{B,\mathrm{it}}$ and $\overline{\gamma}_{B,\mathrm{it}}$, only depending on the generalized conditions of shape regularity.

Let us count the number of arithmetic operations spent by the algorithm for solving the system (7.185) by the use of the preconditioner $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$:

- One matrix-vector multiplication by $\mathbf{S}_B$ requires $\mathcal{O}(\mathcal{R}p^2)$ a.o.
- One matrix-vector multiplication by $\boldsymbol{\mathcal{S}}_B$ with the use FDFT requires $\mathcal{O}(\mathcal{R}p\log p)$ a.o., and $\mathcal{O}(\log p)$ multiplications in the secondary iteration process spend $\mathcal{O}(\mathcal{R}p\log^2 p)$ a.o.
- The cost of $\mathcal{O}(\log p)$ solutions of systems of algebraic equations with the matrix $\boldsymbol{\mathcal{S}}_{B,2}$ is $\mathcal{O}(\mathcal{R}p\log^2 p)$. It is assumed that the factorized form the preconditioner $\boldsymbol{\mathcal{S}}_{B,2} = \mathbf{C}_B^\top \boldsymbol{\mathcal{S}}_{B,1}\mathbf{C}_B$ is taken into account as well as the block diagonal structure of the preconditioner $\boldsymbol{\mathcal{S}}_{B,1}$, allowing one to use FDFT for each edge block.
- The number of a.o. required for solution of the system of algebraic equations with the matrix $\widetilde{\boldsymbol{\mathcal{K}}}_V$ is denoted by $\mathit{\Gamma}(\mathcal{R})$. If optimal elimination or nested dissection procedures, see [George and Liu (1981)], will be applicable, then $\mathit{\Gamma}(\mathcal{R}) \le c\mathcal{R}^{1.5}$, with some positive constant c. In general, the coefficient ϱ may have values chaotically distributed among finite elements with considerable jumps, and, therefore, special solving procedures must be employed.

Finally, we can formulate the following conclusion.

Proposition 7.2. *Let $\varrho = \varrho_r = \mathrm{const} > 0$ in τ_r $r = 1, 2, \ldots, \mathcal{R}$, and let $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$ be defined by (7.187). Then, for $\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}$, the spectral equivalence inequalities (7.193) hold,*

$$\mathrm{ops}[\boldsymbol{\mathcal{S}}_{B,\mathrm{it}}^{-1}\mathbf{v}_B] \le c\,(\mathcal{R}p\log^2 p + \mathit{\Gamma}(\mathcal{R})\log p)\,, \quad \forall\, \mathbf{v}_B \in V_B\,,$$

and the calculation of the solution $\mathbf{u}_B$ *to the system (7.185) requires not more than* $c(\mathcal{R}(p^2 + p\log^2 p) + F(\mathcal{R})\log p)$ *a.o., where the positive constant* c *depends only on the generalixed conditions of shape regularity.*

Chapter 8

Fast Dirichlet Solvers for 2d Reference Elements

It is very clear that fast solvers for the local Dirichlet problems on the finite elements are of utmost importance for total efficiency of the DD solvers. They represent the main component of the DD algorithms, especially when they are simultaneously used for the prolongations from the interface boundary inside finite elements. For large p, the use of direct solvers is often prohibitively expensive, because, *e.g.,* for varying coefficients of the elliptic problem or curvilinear finite elements, the element stiffness matrices can be completely filled in or very dense. This is especially true, since only general direct solution procedures will be applicable in such situations. Thus, iterative solution algorithms became preferable, and, under the shape regularity conditions, the construction of them is reduced to the construction of efficient preconditioner-solvers for the Dirichlet problems on the associated reference elements.

In view of the properties of the finite-difference and low order finite element preconditioners for the reference element stiffness matrices, discussed in Section 7.4, creation of fast local Dirichlet iterative solvers is an intricate problem. For the spectral discretizations, a pioneering fast multigrid-type solver was reported in [Shen *et al.* (2000)]. The numerical experiments presented demonstrated a very good performance, and were accompanied by the theoretical justification for 1d case. An attempt to design a fast solver for the hierarchical reference element stiffness matrix was made in [Korneev (2001)]. It was conceived as a special Dirichlet–Dirichlet DD preconditioner-solver [1] for the finite-difference-like preconditioner of Subsection 7.4.1, suggested earlier by [Ivanov and Korneev (1996)]

[1] The edge inter-subdomain Schur complement $\boldsymbol{\mathcal{S}}$ derived in [Korneev (2001)], see (4.3), has indeed a more complicated form. However, it is an almost optimal preconditioner for the Schur complement.

for the reference element stiffness matrix. The number of subdomains in it is $\mathcal{O}(\log^2 p)$, and, therefore, grows alongside with p. This DD solution technique was improved in [Korneev (2002b)] and then in [Korneev (2013b)]. In the latter paper, the algorithm of almost linear complexity, obtained previously, was modernized into one providing linear complexity, *i.e.*, spending $\mathcal{O}(p^2)$ a.o. Following this paper, we present these results in Subsection 8.1.2.

A simple and elegant algebraic multigrid preconditioner-solver of almost optimal complexity was suggested by [Beuchler (2002)] for the same type 2d p-reference element. It was designed as a solver for the preconditioner, which is the spectrally equivalent finite element version of the finite-difference preconditioner of [Ivanov and Korneev (1996)]. Due to one simplification of the finite-difference/finite element preconditioner, independently approved by [Korneev *et al.* (2003b)], [Beuchler *et al.* (2004)], and [Beuchler and Braess (2006)], the cost of the multigrid preconditioner-solver was reduced to the optimal order, *i.e.* $\mathcal{O}(p^2)$. This solver is presented in Subsection 8.1.1. In the cited paper of [Beuchler *et al.* (2004)], the authors also developed a fast solver based on a multilevel wavelet decompositions of the 1d first order FE spaces, involved in the definition of the finite element preconditioner. We postpone the introduction of this important, but more complex fast solver to the presentation of solvers for 3d reference p-elements in Subsection 9.3.1.

The above listed DD, algebraic multigrid and multilevel wavelet solvers represent three core types of fast solvers for the Dirichlet problems on hierarchical p-elements. They illuminate some basic general properties of the finite-difference/finite element-like, but still complicated preconditioners, on the basis of which other fast solvers can be created, *e.g.*, based on BPX and MDS types of multilevel preconditioning, algebraic multilevel preconditioning like those proposed by [Axelsson and Vassilevski (1989, 1990)], overlapping DD preconditioning etc. We also mention generalizations to some special reference elements, like incomplete and adaptive ones, see, *e.g.,* [Korneev *et al.* (2003a)], which are important for attaining exponential convergence by means of hp-adaptive discretizations.

For expanding the toolkit of fast Dirichlet solvers to the spectral p-reference elements, a substantial step has been made only recently. It was established that, for the stiffness and mass matrices of these reference elements, there exists a diagonal transformation to the forms which are similar to that of the hierarchical reference elements in essential properties, see [Korneev and Rytov (2005a, 2008)]. This allows us easily to adapt all fast

solvers existing for the hierarchical reference elements to the spectral reference elements and vice versa. Fast solvers for the latter reference elements, constructed in this way, are presented in Sections 8.3 and 9.3.

Creation of fast solvers directly for the finite-difference precondtioners for the spectral reference elements, *i.e.*, without resorting to the transformation of variables, is also possible. One of such solvers, belonging DD type, is developed in Section 8.3.

8.1 Fast Dirichlet Solvers for Hierarchical Reference Elements

8.1.1 *Algebraic Multigrid Solvers*

8.1.1.1 *Description of the Algorithm*

The finite element type preconditioners are more suitable for the analysis. For this reason, we will introduce the finite element preconditioner $\mathbf{B}$ in such a way that it will be spectrally equivalent to the finite-difference preconditioner $\mathbf{\Lambda}$ of (7.101), p. 235, uniformly in p. The preconditioner will be defined by means of the same square mesh of the size $\hbar$ on π_1 with the finite-difference preconditioner. It will be called the *discretization, fine or source mesh* in distinction with a *decomposition, coarse* or a *sequence of coarser meshes* used in different solvers.

Let us subdivide each square nest of the fine mesh in two triangles by one of its diagonals, and let $\mathring{\mathcal{H}}(\pi_1)$ be the space of the continuous piecewise linear functions on π_1, vanishing on $\partial\pi_1$. We introduce the matrix $\mathbf{B}$ as the FE matrix generated by the bilinear form

$$a^{(1)}_{\pi_1}(u,v) := \int_{\pi_1} [\varphi(x_2)u_{x_1}v_{x_1} + \varphi(x_1)u_{x_2}v_{x_2}]\, dx\,,\ \forall\ u,v \in \mathring{\mathcal{H}}(\pi_1)\,, \quad (8.1)$$

with the choice of the nodal basis in $\mathring{\mathcal{H}}(\pi_1)$. Clearly, the bilinear form corresponds to the differential operator $\mathcal{L}$ of the equation (7.112) up to multiplier 2.

Lemma 8.1. *There hold the inequalities*

$$\mathbf{B} \prec \hbar^2 \mathbf{A}_{e,e} \prec \mathbf{B}\,. \quad (8.2)$$

Proof. In view of Theorem 7.3, it is sufficient to prove simpler inequalities

$$\mathbf{B} \prec \hbar^2 \mathbf{\Lambda} \prec \mathbf{B}\,. \quad (8.3)$$

It is easy to notice that the finite element matrix $\mathbf{B}$ has the 5-point stencil and structure identical to those of the finite-difference preconditioner $\mathbf{\Lambda}$. Namely, by easy calculations, we obtain

$$\mathbf{B} = \hbar^2 \mathbb{A}\,, \quad \text{with } \mathbb{A} = (\boldsymbol{\mathcal{D}}_e + \frac{2}{3}) \otimes \mathbf{\Delta} + \mathbf{\Delta} \otimes (\boldsymbol{\mathcal{D}}_e + \frac{2}{3})\,, \tag{8.4}$$

and, comparing with (7.101), we conclude that $\mathbb{A}$ only slightly differs from $\mathbf{\Lambda}$ by the additional summand $2/3$ in brackets. □

Without loss of generality, we assume $N+1 = 2^{l_0}$, additionally to the earlier assumption $p = 2N+1$. This allows us to introduce the sequence of l_0 imbedded orthogonal meshes of the sizes $\hbar_l = 2^{-l}$. The subsets of the nodes $x = \hbar_l(i,j)$ in π_1 of these meshes are denoted by X_l, each containing $N_l \times N_l$ internal nodes, where obviously $N_l = 2^l - 1$. This sequence induces the sequence of the finite element spaces $\mathring{\mathcal{H}}_l(\pi_1)$ and the sequence of the finite element matrices $\mathbf{B}_l$, such that $\mathring{\mathcal{H}}_{l_0}(\pi_1) = \mathring{\mathcal{H}}(\pi_1)$ and $\mathbf{B}_{l_0} = \mathbf{B}$. For simplicity, we write $\mathring{\mathcal{H}}_l$ instead of $\mathring{\mathcal{H}}_l(\pi_1)$ and also use the following notations:

- The subspace $\mathcal{W}_l$ is such that $\mathring{\mathcal{H}}_l = \mathcal{W}_l \oplus \mathring{\mathcal{H}}_{l-1}$ and will be called the *wavelet subspace of the l-th level*. We denote by $\mathbf{B}_{l,w}$ the block on the diagonal of the matrix $\mathbf{B}_l$ corresponding to the subspace $\mathcal{W}_l$.

- U_l and W_l are the spaces of vectors with the entries which are nodal values of functions from $\mathring{\mathcal{H}}_l$ and from $\mathcal{W}_l$ at the nodes of the sets X_l and $X_{l,w} := X_l \setminus X_{l-1}$, respectively.

- $\mathbb{P}_{l-1} : U_{l-1} \to U_l$ is the usual interpolation matrix, such that, if $v \in \mathring{\mathcal{H}}_{l-1}$ and $\mathbf{v}^{(l-1)} \in U_{l-1}$ is its vector representation, then $\mathbf{v}^{(l)} := \mathbb{P}_{l-1}\mathbf{v}^{(l-1)}$ is the vector representation of v as a function of $\mathring{\mathcal{H}}_l$.

- $\mathbb{R}_l : U_l \to W_l$ is the matrix defining the restrictions $\mathbf{v}^{(l-1)} := \mathbb{R}_l \mathbf{v}^{(l)}$ of the vector $\mathbf{v}^{(l)}$ to the set of nodes $X_{l,w}$.

From the above definitions, it follows the multilevel representation of the FE space

$$\mathring{\mathcal{H}} = \mathcal{W}_{l_0} \oplus \mathcal{W}_{l_0-1} \oplus \ldots \oplus \mathcal{W}_1\,, \quad \mathcal{W}_1 := \mathring{\mathcal{H}}_1\,. \tag{8.5}$$

Suppose, we have good preconditioners $\boldsymbol{\mathcal{B}}_{l,w}$ for the matrices $\mathbf{B}_{l,w}$. We are going to approve the efficiency of the multigrid solver for the system

$$\mathbf{B}_l \mathbf{u} = \mathbf{f}\,, \tag{8.6}$$

one iteration of which produces $\mathbf{u}^{k+1} := \mathbf{mgm}(l, \mathbf{u}^k, \mathbf{f})$ from a given approximation $\mathbf{u}^k$ according to the procedure outlined below.

For simplicity of description of the multigrid method, we assume that the system matrix $\mathbf{B}_l$ is written in the basis corresponding to the representation (8.5) of the finite element space.

- **If** $l \geq 1$, then **do**

  (1) Presmoothing in the space W_l:

  $\mathbf{v} := \mathbf{u}^k$;

  do ν **times** $\quad \mathbf{v} := \mathbf{v} - \tau^{-1}\mathbb{R}_l^T \boldsymbol{\mathcal{B}}_{l,w}^{-1}\mathbb{R}_l(\mathbf{B}_l\mathbf{v} - \mathbf{f})$;

  (2) Coarse grid correction in the space V_{l-1}:

  $\mathbf{d}_{l-1} := \mathbb{P}_{l-1}^T(\mathbf{f} - \mathbf{B}_l\mathbf{v})$; $\mathbf{w}^0 = 0$;

  do μ_{l-1} **iterations** $\mathbf{w}^k = \mathbf{Mgm}(l-1, \mathbf{w}^{k-1}, \mathbf{d}_{l-1})$;

  $\mathbf{v} := \mathbf{v} + \mathbb{P}_{l-1}\mathbf{w}^{\mu_{l-1}}$;

  (3) Postsmoothing in the space W_l:

  do ν **times** $\quad \mathbf{v} := \mathbf{v} - \tau^{-1}\mathbb{R}_l^T \boldsymbol{\mathcal{B}}_{l,w}^{-1}\mathbb{R}_l(\mathbf{B}_l\mathbf{v} - \mathbf{f})$;

  $\mathbf{u}^{i+1} = \mathbf{v}$

- **else**

  solve (8.6) **by some direct method**

- **endif**

Here τ are iteration parameters, which in the simplest case are taken constant and independent of l. Parameters ν and μ_k define the number of pre- and post-smoothing iterations and the number of multigrid iterations on the level k, respectively.

The choice of the preconditioner $\boldsymbol{\mathcal{B}}_{l,w}$ is very important, and an optimal one provides the spectral equivalence to the matrix $\mathbf{B}_{l,w}$ and existence of an optimal solver for the system with the matrix $\boldsymbol{\mathcal{B}}_{l,w}$. One of such preconditioners is described below.

We denote by $\Im_l^i$ the line, which passes through the nodes $(\overline{i}, \overline{j})$ such that $\max(\overline{i}, \overline{j}) = i$. This line is composed of the two orthogonal segments of equal length: the segment $[0, i\hbar_l]$ of the line $x_2 \equiv i\hbar_l$ and the segment $[0, i\hbar_l]$ of the line $x_1 \equiv i\hbar_l$, see Figure 8.1. It may be noticed that $\mathbf{B}_{l,w}$ may be represented in the block form

$$\mathbf{B}_{l,w} = \text{tridiag}\,[\mathbf{B}_{l,w}^{(i,i-1)},\ \mathbf{B}_{l,w}^{(i)},\ \mathbf{B}_{l,w}^{(i,i+1)}]_{i=1,2,\ldots,N_l}\ , \tag{8.7}$$

where each block $\mathbf{B}_{l,w}^{(i)}$ is related to those unknowns from the subspace W_l, which correspond to the nodes on the line $\Im_l^i$. As well, one can see that the blocks $\mathbf{B}_{l,w}^{(i)}$ with the even numbers $i = 2k$ are diagonal and the blocks $\mathbf{B}_{l,w}^{(i)}$ with the odd numbers $i = 2k+1$ are tridiagonal. In Figure 8.1, the nodes corresponding to the blocks $\mathbf{B}_{l,w}^{(2k)}$ and $\mathbf{B}_{l,w}^{(2k+1)}$ are shown by crosses and circles, respectively. The preconditioner $\boldsymbol{\mathcal{B}}_{l,w}$ is defined simply by decoupling in $\mathbf{B}_{l,w}$ the blocks $\mathbf{B}_{l,w}^{(i)}$, *i.e.,* by setting

$$\boldsymbol{\mathcal{B}}_{l,w} = \text{diag}\,[\mathbf{B}_{l,w}^{(i)}]_{i=1,2,\ldots,N_l}\,. \tag{8.8}$$

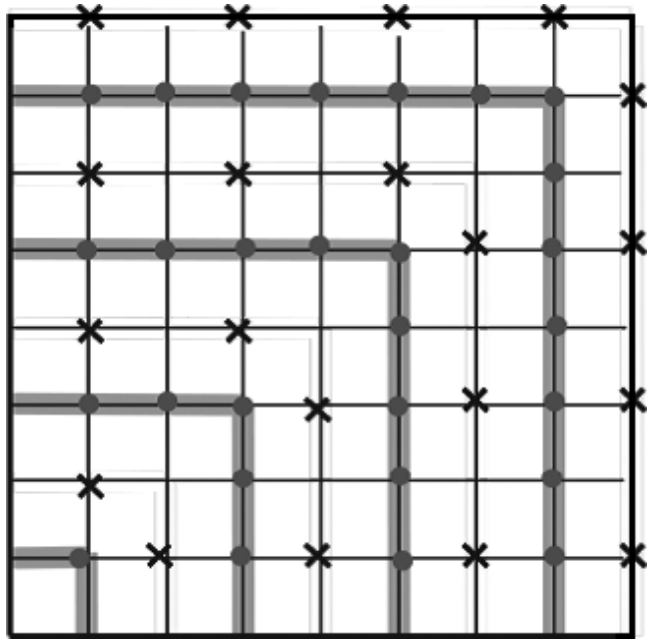

Fig. 8.1 Meshes corresponding to the algebraic multigrid method.

8.1.1.2 *Rate of Convergence and Influencing Factors*

The most decisive factor for convergence of the described multigrid method is the strengthened Cauchy inequality

$$(a^{(1)}_{\pi_1}(u,w))^2 \le c_0\, a^{(1)}_{\pi_1}(u,u)\, a^{(1)}_{\pi_1}(w,w)\ , \quad \forall u \in \mathring{\mathcal{H}}_{l-1}\ , \quad \forall w \in \mathcal{W}_l\ , \tag{8.9}$$

hold with some $c_0 < 1$ independent of p. Another factor of crucial importance is the quality of preconditioning of the matrices $\mathbf{B}_{l,w}$ expressed by the inequalities

$$c_1 \boldsymbol{\mathcal{B}}_{l,w} \le \mathbf{B}_{l,w} \le c_2 \boldsymbol{\mathcal{B}}_{l,w} \tag{8.10}$$

hold, as we show in the next subsubsection with positive constants independent of p, l and l_0. Under the condition that sufficiently accurate values of c_1, c_2 are found, it is reasonable to assume the use of the iteration parameter $\tau = 2/(c_1 + c_2)$ in the smoothing process. Since even the structure of the efficient algorithm depends on specific values of c_k, $k = 0, 1, 2$, we should get sufficiently sharp bounds for these constants.

Theorem 8.1. *Let the sequences of the subspaces $\mathcal{W}_l, \mathring{\mathcal{H}}_l$, the related matrices $\mathbf{B}_{l,w}$ and their preconditioners $\boldsymbol{\mathcal{B}}_{l,w}$ be as defined above. Then the*

inequalities (8.10) and (8.9) hold with such positive constants $c_0 < 2/3$ *and* c_1, c_2 *independent of* p, l *and* l_0*, that for* $\tau = 2/(c_1 + c_2)$*,* $\mu \geq 3$ *and* $\nu \geq \nu_o(c_0, c_1, c_2) = 3$ *the convergence factor of the multigrid iterations*

$$\rho_{l,\mathrm{mult}} := \sup_{\mathbf{u}^k \in U_l} \frac{\|\mathbf{u}^{k+1} - \mathbf{u}\|_{\mathbf{B}_l}}{\|\mathbf{u}^k - \mathbf{u}\|_{\mathbf{B}_l}}$$

is estimated by the constant $\rho < 1$ *uniformly in* p, l, l_0*.*

Proof. The proof follows from Theorem 8.3, Lemma 8.2, and Theorem 8.4. At that, Theorem 8.3 and Lemma 8.2 provide the theoretical basis for defining the convergence factor $\rho_{l,\mathrm{mult}}$ as a function of c_0, c_1, c_2, whereas the specific values of these constants are estimated in Theorem 8.4. To add a few details, we note that the bounds of constants c_0, c_1, c_2 in the latter theorem are produced by the standard superelement technique. In other words, for the discretization of the lth level, these bounds are produced by means of estimating respective values for the restrictions of the subspaces $\mathring{\mathcal{H}}_{l-1}$ and $\mathcal{W}_l$ on the domains of the triangular or square superelements. Triangles of the triangulation of $(l-1)$-th level and the nests of the $(l-1)$-th level orthogonal mesh serve for the domains of superelements, which are denoted $\delta^{l-1}_{\triangledown,\boldsymbol{\gamma}}$, $\delta^{l-1}_{\triangle,\boldsymbol{\gamma}}$, and $\delta^{l-1}_{\boldsymbol{\gamma}}$, respectively, where $\overline{\delta}^{l-1}_{\boldsymbol{\gamma}} = \overline{\delta}^{l-1}_{\triangledown,\boldsymbol{\gamma}} \cup \overline{\delta}^{l-1}_{\triangle,\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, and γ_k are the numbers of the mesh lines $x_k \equiv \hbar_{l-1}\gamma_k$ crossing the upper right vertex of the superelement, see, *e.g.*, (A.1). For $\mathbf{B}$ induced by the bilinear form (8.1) with $\varphi \equiv 1$, we have $c_0 = 1/2$, see [Maitre and Musy (1982)]. In the case under consideration $\varphi(1) = 1$, the value of c_0 for the superelement $\delta^{l-1}_{\boldsymbol{\gamma}}$, $\gamma = (N_{l-1}, N_{l-1})$ tends to $1/2$ at $l \to \infty$, while bounds $c_0 \geq 1/2$ can be supported for come other superelements, *i.e.*, by direct computation. For the reason that $\varkappa \geq \theta^2 = c_0$ according to (8.31) in the proof of Theorem 8.3, we come to the conclusion that $\varkappa \geq 1/2$. Therefore, Lemma 8.2 can guarantee convergence only at $\mu \geq 3$ with the convergence factor $\sigma_l = \rho_{l,\mathrm{mult}}$ in the last line of (8.33), but only under condition that c_0, c_1, c_2 and ν are such that $\varkappa < 2/3$. According to Theorem 8.4, for smoothings implemented in the algorithm we have $\kappa = \varrho_\circ^\nu$ and $\varrho_\circ \leq \sqrt{0.35}$, whereas the condition $\varkappa \leq 2/3$ of Lemma 8.2 takes the form

$$\nu_o > \left[\ln\left(\sqrt{2/3} - \max_{\boldsymbol{\gamma}} \theta_\circ(\boldsymbol{\gamma})\right) - \ln\left(1 - \max_{\boldsymbol{\gamma}} \theta_\circ(\boldsymbol{\gamma})\right)\right] / \ln \varrho_\circ \,.$$

In particular, if (A.39) is fulfilled, then one should accept $\nu_o = 3$. □

Counting the number of arithmetic operations, we conclude that, for $\mu = 3$, the method is optimal with respect to the computational work.

We note that the best value $c_0 = 95/176 = 0,5327(72)$ was obtained by [Beuchler (2001)], see Theorem 2.2 in this paper, with the use of a computer code for symbolic linear algebra computations. At that, the angle between the subspaces $\mathring{\mathcal{H}}_{l-1}, \mathcal{W}_l$ was estimated by the minimal angle between the restrictions of these subspaces on the domains of the triangular and square superelements. Triangles of the triangulation of $(l-1)$-th level, having no vertices on the axes, and the nests of the $(l-1)$-th level orthogonal mesh adjacent to the axes served for the domains of the former and latter superelements, respectively. The problem for one superelement, formulated similarly to (4.41), was reduced to the 2×2 matrix eigenvalue problem. The computer code delivered rather lengthy expressions for the coefficients of the corresponding quadratic equation. These expressions are rational functions having polynomials of 6th power in each variable γ_1, γ_2 in the numerator and denominator. However, they allowed rather easily to locate the triangle $\delta^{l-1}_{\triangle,\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} = (2,2)$, for which the angle is minimal. Then for this triangle the corresponding eigenvalue problem was solved numerically. We implement a sufficiently standard technique for the analytical estimation of c_0 by means of the consistent application of the Cauchy inequality with epsilon, which resulted in $c_0 = 0.5694$, see the proof of Theorem 8.4 in the next subsubection and Appendix A.1.

Obviously, for $c_{0,\triangledown}(\boldsymbol{\gamma})$, $c_{0,\triangle}(\boldsymbol{\gamma})$ and $c_0(\boldsymbol{\gamma})$ standing for the values of c_0 for the respective superelements, we have

$$c_{0,\triangledown}(k,k) = c_{0,\triangle}(k,k) = c_0(k,k)\,, \qquad k = 1, \ldots, \overline{N}\,. \tag{8.11}$$

This follows by the symmetry properties of the FE discretization with respect to the diagonal of π_1 passing through the origin.

8.1.1.3 *Theoretical Basis*

The basis for the proof is the result of [Schieweck (1986)] on the so called *Approximate projection algorithm*, which is presented first. Let H be a Hilbert space and $a : H \times H \to R$ a continuous symmetric strongly elliptic bilinear form specifying the scalar product $a(u,v)$, $\forall\, u, v \in H$, and the norm

$$\|\, v \,| = (a(v,v))^{1/2}\,, \qquad \forall\, v \in H\,.$$

It is assumed that $H = H_1 \oplus H_2$ and that a strengthened Cauchy inequality

$$a(u,v) \le \theta \,\|\, u \,|\, |\, v \,\|\,, \qquad \forall\, u \in H_1,\ \forall\, v \in H_2\,, \tag{8.12}$$

holds with $\theta \in (0,1)$. In approximate projection algorithm with some initial approximation $u^{(1)} \in H$ given, finding the next approximation $u^{(2)} \in H$ to the solution u of the equation

$$a(u,v) = f(v)\,, \qquad f \in H'\,, \qquad \forall\, v \in H\,, \tag{8.13}$$

consists of the following steps:

1. **Find**

$$\widetilde{u}_1 \in u^{(1)} + H_1 : \quad |\,\widetilde{u}_1 - u_1\,| \le \rho_1 |\,u^{(1)} - u_1\,|, \quad \rho_1 \in (0,1)\,,$$

where $u_1 \in u^{(1)} + H_1$ satisfies $a(u_1, v) = f(v),\ \forall\, v \in H_1$.

2. **Find**

$$\widetilde{u}_2 \in \widetilde{u}_1 + H_2 : \quad |\,\widetilde{u}_2 - u_2\,| \le \rho_2 |\,\widetilde{u}_1 - u_2\,|, \quad \rho_2 \in (0,1)\,,$$

where $u_2 \in \widetilde{u}_1 + H_2$ satisfies $a(u_2, v) = f(v),\ \forall\, v \in H_2$.

3. **Find**

$$\widetilde{u}_3 \in \widetilde{u}_2 + H_1 : \quad |\,\widetilde{u}_3 - u_3\,| \le \rho_3 |\,\widetilde{u}_2 - u_3\,|, \quad \rho_3 \in (0,1)\,,$$

where $u_3 \in \widetilde{u}_2 + H_1$ satisfies $a(u_3, v) = f(v),\ \forall\, v \in H_1$. Set $u^{(2)} = \widetilde{u}_3$.

Convergence of the approximate projection algorithm is approved by the following theorem.

Theorem 8.2. *Let (8.12) be fulfilled and* $\rho_k \in (0,1)$, $k = 1,2,3$. *Then*

$$|\,u^{(2)} - u\,| \le \sigma |\,u^{(1)} - u\,|,$$

with

$$\sigma = \rho_2 + (1-\rho_2)[\rho_1 + \theta(1-\rho_1)][\rho_3 + \theta(1-\rho_3)] \in (0,1)\,.$$

Proof. Let

$$e^{(k)} = u^{(k)} - u\,, \quad \widetilde{e}_k = \widetilde{u}_k - u\,, \quad e_k = u_k - u\,,$$

denote the errors and P_k, $Q_k = I - P_k$ denote the orthogonal projectors on H_k and $H_k^{\perp}$. From definitions of u_1 and u follows $a(u_1 - u, v) = 0,\ \forall\, v \in H_1$, *i.e.,*

$$e_1 \perp H_1\,, \quad e_1 = Q_1 e^{(1)}\,, \tag{8.14}$$

with the last equality following from $e_1 \in e^{(1)} + H_1$. Since $\widetilde{u}_1 \in u^{(1)} + H_1$, one has

$$e_1 = e^{(1)} - \phi_1\,, \quad \phi_1 \in H_1\,, \tag{8.15}$$

and, therefore, $z_1 := \rho_1^{-1}(\phi_1 - P_1 e^{(1)}) \in H_1$. The inequality in step 1 is equivalent to

$$|\,\widetilde{e}_1 - e_1\,| \le \rho_1 |\,e^{(1)} - e_1\,|\,.$$

Putting (8.14) and (8.15) in this inequality and taking into account that $z_1 \in H_1$, we get

$$|\,z_1\,| = |\,P_1 e^{(1)}\,|\,. \tag{8.16}$$

The same facts result in the expression $\widetilde{e}_1 = e^{(1)} - (\rho_1 z_1 - P_1 e^{(1)})$ or

$$\widetilde{e}_1 = \rho_1(Q_1 e^{(1)} - z_1) + (1-\rho_1) Q_1 e^{(1)}\,, \tag{8.17}$$

which consequence is

$$|\,\widetilde{e}_1\,| \le \rho_1 |\,Q_1 e^{(1)} - z_1\,| + (1-\rho_1) |\,Q_1 e^{(1)}\,|\,. \tag{8.18}$$

Now, in view of $Q_1 e^{(1)} \perp H_1, z_1 \in H_1$ and (8.16), we can write

$$\begin{aligned} |\,Q_1 e^{(1)} - z_1\,|^2 &= |\,Q_1 e^{(1)}\,|^2 + |\,z_1\,|^2 \\ &\le |\,Q_1 e^{(1)}\,|^2 + |\,P_1 e^{(1)}\,|^2 = |\,Q_1 e^{(1)} + P_1 e^{(1)}\,|^2, \end{aligned}$$

implying

$$|\,Q_1 e^{(1)} - z_1\,| \le |\,e^{(1)}\,|\,. \tag{8.19}$$

The last inequality allows us to rewrite (8.18) as

$$|\,\widetilde{e}_1\,| \le \rho_1 |\,e^{(1)}\,| + (1-\rho_1) |\,Q_1 e^{(1)}\,| \le |\,e^{(1)}\,|\,. \tag{8.20}$$

Exactly the same way of judgement allows us to obtain

$$\begin{aligned} &\widetilde{e}_2 = \rho_2(Q_2\widetilde{e}_1 - z_2) + (1-\rho_1) Q_2 \widetilde{e}_1\,, \quad z_2 \in H_2\,, \\ &|\,Q_2\widetilde{e}_1 - z_2\,| \le |\,\widetilde{e}_1\,|\,, \\ &|\,\widetilde{e}_2\,| \le \rho_2 |\,\widetilde{e}_1\,| + (1-\rho_2) |\,Q_2\widetilde{e}_1\,|\,, \\ &|\,\widetilde{e}_3\,| \le \rho_3 |\,\widetilde{e}_2\,| + (1-\rho_3) |\,Q_1\widetilde{e}_2\,|\,. \end{aligned} \tag{8.21}$$

Clearly,

$$|\,Q_2\widetilde{e}_1\,| = \min_{w \in H_2} |\,\widetilde{e}_1 + w\,|\,, \tag{8.22}$$

and, for $w = \rho_1 P_2 Q_1 e^{(1)} \in H_2$ by means of (8.17), we obtain

$$|\,\widetilde{e}_1 - w\,| \le \rho_1 |\,Q_1 e^{(1)} - z_1\,| + (1-\rho_1) |\,Q_2 Q_1 e^{(1)}\,|\,. \tag{8.23}$$

If $P_k v = 0$ for $v \in H$, then $|\!|\!| Q_{3-k} v |\!|\!| \le \theta |\!|\!| v |\!|\!|$. By means of this fact and (8.19) we arrive at the inequality

$$|\!|\!| \widetilde{e}_1 - w |\!|\!| \le \rho_1 |\!|\!| e^{(1)} |\!|\!| + (1-\rho_1)\theta |\!|\!| Q_1 e^{(1)} |\!|\!| ,$$

which, in view of (8.22) and $|\!|\!| Q_1 e^{(1)} |\!|\!| \le |\!|\!| e^{(1)} |\!|\!|$, takes the form

$$|\!|\!| Q_2 \widetilde{e}_1 |\!|\!| \le [\rho_1 + (1-\rho_1)\theta] |\!|\!| e^{(1)} |\!|\!| . \tag{8.24}$$

With the use of two first lines of (8.21) in a similar way it is proved that

$$|\!|\!| Q_1 \widetilde{e}_2 |\!|\!| \le \rho_2 |\!|\!| \widetilde{e}_1 |\!|\!| + (1-\rho_2) |\!|\!| Q_2 \widetilde{e}_1 |\!|\!| . \tag{8.25}$$

Turning to the right part of the last line of (8.21), we can substitute the third line of (8.21) and (8.25), then use the inequalities (8.24) and (8.20). This concludes the proof. □

Theorem 8.2 allows us to analyze the convergence of algebraic multigrid methods under a rather general assumptions. Suppose, we have a sequence of imbedded FE spaces

$$\mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}_{l_0} = \mathcal{V} \tag{8.26}$$

with each subspace being the sum of two subspaces

$$\mathcal{V}_l = \mathcal{V}_{l-1} + \mathcal{W}_l , \quad \mathcal{V}_1 = \mathcal{W}_1 ,$$

satisfying the strengthened Cauchy inequality

$$a(u,v) \le \theta |\!|\!| u |\!|\!| \, |\!|\!| v |\!|\!| , \quad \forall\, u \in \mathcal{V}_{l-1},\ \forall\, v \in \mathcal{W}_l , \quad l = 2, \dots l_0 . \tag{8.27}$$

Let us introduce some smoothing operator S_l such that

$$|\!|\!| S_l v_l - y_l |\!|\!| \le \kappa |\!|\!| v_l - y_l |\!|\!| , \quad \kappa \in (0,1) , \tag{8.28}$$

where $y_l \in z_l + W_l$ is the solution of

$$a(y_l, v) = f(v) , \quad \forall\, v \in \mathcal{W}_l .$$

Suppose, we need to find the FE solution $u \in \mathcal{V}_l$ to the equation (8.13) in the space $\mathcal{V}_l$

$$a(u,v) = f(v) , \quad \forall\, v \in \mathcal{V}_l . \tag{8.29}$$

One iteration of the algebraic multigrid method produces $u^{k+1} := \mathrm{mgm}(l, u^k, f) \in \mathcal{V}_l$ from a given approximation $u^k \in \mathcal{V}_l$ as follows.

- **If** $l \ge 1$, then **do**
  (1) Presmoothing in the space $\mathcal{W}_l$: **compute** $v_1 = S_l u^k$.

(2) Coarse grid correction in the space $\mathcal{V}_{l-1}$: **do** $w^{k+1} = \mu_{l-1}$ **iterations** $\mathrm{mgm}(l-1, w^k, g)$, $w^0 = 0$, **for solving equation**

$$a(v_2, v) = g(v)\,, \quad g(v) = f(v) - a(v_1, v)\,, \quad \forall\, v \in \mathcal{V}_{l-1}\,.$$

Put $v_2 = v_1 + w^{\mu_{l-1}}$.

(3) Postsmoothing in the space $\mathcal{W}_l$: **compute** $u^{k+1} = S_l v_2$.

- **else**

  solve (8.29) **by a direct method**
- **endif**

Theorem 8.3. *Let (8.27) and (8.28) hold and $u \in \mathcal{V}_l$ solve (8.29). Then*

$$|\,u^{k+1} - u\,| \le \sigma_l\,|\,u^k - u\,|,$$

where σ_l defined by the recursion

$$\sigma_l = \sigma_{l-1}^{\mu_{l-1}} + (1 - \sigma_{l-1}^{\mu_{l-1}})[\kappa + \theta(1-\kappa)]^2 \in (0,1)\,,\ l = 2, 3, \dots\,,\ \sigma_1 = 0\,. \quad (8.30)$$

Proof. The proof follows from Theorem 8.2, since the algebraic multigrid algorithm can be treated as the approximate projection algorithm with $H = \mathcal{V}_l$, $H_1 = \mathcal{W}_l$, $H_2 = \mathcal{V}_{l-1}$, $\rho_1 = \rho_3 = \kappa$, and $\rho_2 = \sigma_{l-1}^{\mu_{l-1}}$. □

The following lemma, typical for analysis of algebraic multigrid methods, supplies us with the lacking important details on the behavior of the sequence $\{\sigma_l\}$. Introducing the notation

$$\varkappa = [\kappa + \theta(1-\kappa)]^2\,, \quad (8.31)$$

we rewrite (8.30) in a shorter form

$$\sigma_l = \sigma_{l-1}^{\mu_{l-1}} + (1 - \sigma_{l-1}^{\mu_{l-1}})\varkappa \in (0,1)\,, \quad l = 2, 3, \dots\,, \quad \sigma_1 = 0\,. \quad (8.32)$$

Lemma 8.2. *For $\mu_l \equiv \mu \ge 1$ and $\varkappa \in (0,1)$, the sequence (8.32) is monotonically increasing and bounded from above by 1. For $\mu = 1$ it always converges to 1. If $\mu \ge 2$ and $\varkappa < 1 - 1/\mu$, then the equation*

$$\sigma = \varkappa + \sigma^\mu(1 - \varkappa)$$

has a solution $\sigma \in [0,1)$, and elements of the sequence (8.32) are contained in the interval $[0, \sigma]$. For $\varkappa \in (0,1)$ and $\mu = 2$, we have

$$\sigma_l \to_{l\to\infty} \begin{cases} 1\,, & \text{for } \varkappa \ge 1/2\,, \\ \varkappa/(1-\varkappa)\,, & \text{for } \varkappa < 1/2\,, \end{cases}$$

whereas the case $\varkappa \in (0,1)$ and $\mu = 3$ yields

$$\sigma_l \to_{l\to\infty} \begin{cases} 1\,, & \text{for } \varkappa \ge 2/3\,, \\ \sqrt{\frac{1}{4} + \frac{\varkappa}{1-\varkappa}} - \frac{1}{2}\,, & \text{for } \varkappa < 2/3\,. \end{cases} \quad (8.33)$$

Proof. The proof can be found in several papers, see, *e.g.*, Lemma 3.2 in [Schieweck (1986)]. □

From Theorem 8.3 and Lemma 8.2, it follows that the rate of convergence of the multigrid iterations presented in Subsubsection 8.1.1.1 is controlled by three values $c_k > 0,\ k = 0, 1, 2$, entering (8.9) and (8.10), which, in the case under consideration, are constants estimated in the corollary and theorem below.

Corollary 8.1. *Let* $\nu = 3$ *and*

$$\varrho_\circ = \frac{c_2 - c_1}{c_2 + c_1}\,.$$

Then, in order to provide $\varkappa < 2/3$, *the constant in the strengthened Cauchy inequality (8.9) must satisfy the inequality*

$$c_0 < c^\circ = \left(\frac{\sqrt{2/3} - \varrho_\circ^3}{1 - \varrho_\circ^3}\right)^2\,, \tag{8.34}$$

which is satisfied for $c_0 \le 0.5694$ *in (A.39) and* c_1, c_2 *in* (8.35).

Theorem 8.4. *Under the conditions of Theorem 8.1, the constant* c_0 *satisfies (8.34), and there hold the bounds*

$$c_1 \ge 1 - \varrho_\circ\,, \quad c_2 \le 1 + \varrho_\circ\,, \quad \varrho_\circ \le \sqrt{0.35}\,. \tag{8.35}$$

Proof. The original proof was made by [Beuchler (2002)] by means of the superelement technique, which in a slightly simplified form is reproduced in Appendix A.1. Constant $c_0(\boldsymbol{\gamma})$ for square superelements is estimated analytically by a standard technique based on consistent application of the Cauchy inequality with epsilon. It results in the quadratic equation with the explicitly given and relatively simply looking coefficients and in the sufficiently good value $c_0 = 0.5694$, see (A.39), approving Theorem 8.1 as confirmed by Corollary 8.1. □

8.1.1.4 *Some Remarks and Numerical Experiment*

1. The use of multigrid solver for preconditioning. Let $\mathbf{M}_\mu$ be the linear error transition operator defined by one multigrid cycle. Then $\Bbbk \ge 1$ multigrid iterations applied to solving the system (8.6) implicitly define the multigrid preconditioner $\mathbf{M}\mathbf{g}^{-1} = (\mathbf{I} - \mathbf{M}_\mu^\Bbbk)\mathbf{B}^{-1}$ for $\mathbf{B}^{-1}$, which according to (8.2) is also a preconditioner for $\mathbf{A}_{e,e}^{-1}$.

Corollary 8.2. *Let $\mu = 3$, $\nu \geq 3$ and $\Bbbk \geq 1$. Then*

$$\underline{c}\,\mathbf{Mg}^{-1} \leq \mathbf{A}_{e,e}^{-1} \leq \overline{c}\,\mathbf{Mg}^{-1}\,, \tag{8.36}$$

with the positive spectral equivalence constants $\underline{c}$ and $\overline{c}$ which are independent of p.

PCGM with the preconditioner $\mathbf{Mg}$ for systems with the system matrix $\mathbf{A}_{e,e}$ is also optimal, *i.e.,* at $\Bbbk = 1$ has the computational cost $\mathcal{O}(N^2)$.

The above results rest on the facts, reflected in the inequalities (8.10) and (8.9), that the coupling between the unknowns inside of each line $\Im_i$ is stronger than between unknowns of different lines, and that the angle between the subspaces $\mathring{\mathcal{H}}_{l-1}, \mathcal{W}_l$ is bounded by the constant $c_0 < 1$ well separated from unity uniformly in l and p. These basic facts allow us to derive different fast multilevel solvers-preconditioners, *e.g.,* of algebraic multilevel type as proposed by [Axelsson and Vassilevski (1989, 1990)], MDS type etc. One of them is found in [Beuchler (2001)]. There are also some close versions of multilevel solvers, which have not yet been approved theoretically, but demonstrated very good performance in the practice. Above, we applied smoothings only in the subspaces W_l with the smoothers $\mathbf{T}_{l,w} = I - \tau \boldsymbol{\mathcal{B}}_{l,w}^{-1}\mathbf{B}_{l,w}$. A competing algorithm is obtained, if smoothings are performed in the subspaces U_l with the smoothers $\mathbf{T}_l = I - \tau \boldsymbol{\mathcal{B}}_l^{-1}\mathbf{B}_l$, where the preconditioner is defined in a way close to the definition of $\boldsymbol{\mathcal{B}}_{l,w}$. Namely, we set

$$\boldsymbol{\mathcal{B}}_l = \text{diag}\,[\mathbf{B}_l^{(i)}]_{i=1,2,\ldots,N}\,, \tag{8.37}$$

with the blocks $\mathbf{B}_l^{(i)}$ appearing in the similar to (8.7) representation

$$\mathbf{B}_l = \text{tridiag}\,[\mathbf{B}_l^{(i,i-1)},\ \mathbf{B}_l^{(i)}, \mathbf{B}_l^{(i,i+1)}]_{i=1,2,\ldots,N}\,. \tag{8.38}$$

Each block $\mathbf{B}_l^{(i)}$ on diagonal corresponds to the unknowns at the nodes of one line $\Im_i$, but now all the nodes of this line participate, and for this reason all the blocks $\mathbf{B}_l^{(i)}$ are tridiagonal.

2. *Numerical experiment.* Systems with the independent blocks $\mathbf{A}_{a,b},\ a,b = e,o$, of the reference element stiffness matrix $\mathbf{A}_I$ were solved by the PCGM with the multigrid preconditioner $\mathbf{Mg}$ at $\mu = 3,\ \nu = 3$. The stopping criteria for PCGM iterations was the reduction of the preconditioned energy norm $(\mathbf{e}^\top \mathbf{A}_{a,b}\mathbf{Mg}^{-1}\mathbf{A}_{a,b}\mathbf{e})^{1/2}$ of the error $\mathbf{e}$ in 10^9 times. The notations in Table 8.1, reflecting results of the numerical experiment, are the following:

$p = 2N + 1$ — the highest degree in each variable of the polynomials in the hierarchical basis,

Table 8.1 Multigrid Preconditioned Conjugate Gradient Method

l	p	n	i_1	i_2	i_3	M	t_1	t_2	t_3
2	7	9	6	9	6	448	0	0	0
3	15	49	15	14	12	2,032	0	0	0
4	31	225	18	17	14	8,768	0.01	0.01	0.01
5	63	961	21	18	15	36,560	0.05	0.05	0.04
6	127	3,969	22	20	17	149,472	0.37	0.34	0.29
7	255	16,129	23	21	18	604,656	1.92	1.76	1.52
8	511	65,025	24	22	19	2,432,512	10.26	9.47	8.22
9	1,023	261,121	24	22	20	9,758,224	47.69	43.98	40.57
10	2,047	1,046,529	24	23	21	39,089,696	210.07	202.23	185.79

$n = N^2$ — the number of the unknowns of the systems with the matrices $\mathbf{A}_{a,b},\ a, b = e, o$,

l — number of the fine grid, so that $N = 2^l$,

i_1,i_2 and i_3 — numbers of PCGM-iterations for solving the systems with the matrices $\mathbf{A}_{e,e},\ \mathbf{A}_{e,o}$ and $\mathbf{A}_{o,o}$, respectively,

M — the volume of the computer dynamic memory in bytes, used at solution of the system with the block $\mathbf{A}_{e,e}$,

t_1, t_2 and t_3 — times in seconds, spent for solving the systems $\mathbf{A}_{e,e}$, $\mathbf{A}_{e,o}$ and $\mathbf{A}_{o,o}$, respectively. The computer code was run on PENTIUM-III, 750MHz with 256MB of operative memory.

We see that the number of PCGM iterations grows slower than $(\log N)^{1/2}$ from $l = 4$ on. In the asymptotic range, the computer time is growing slower, than $N^2(\log N)^{1/2}$. Also, the results clearly show, that all blocks $\mathbf{A}_{e,e},\ \mathbf{A}_{e,o}$ and $\mathbf{A}_{o,o}$ and the preconditioner $\mathbf{Mg}$ are indeed spectrally equivalent, although the chosen particular form of $\mathbf{B}$ seems to be more close to $\mathbf{A}_{o,o}$, than to $\mathbf{A}_{e,e}$. Apparently the performance of the computer code can be improved, since the stopping criteria for multigrid iterations at each PCGM iteration required high accuracy, *i.e.,* the multigrid solver was practically used as the exact one. Use of the optimized k for the exit from multigrid iterations apparently can result in additional reduction of the computer time.

The numerical experiment was performed by M.V. Nikitin [2] and described in [Korneev *et al.* (2002d,c)], from where Table 8.1 is borrowed. Thorough testing of the algebraic multigrid solver, demonstrating its high efficiency, can be found in [Beuchler (2002)].

[2] *Multigrid method for problems of heat conduction in the domains with curvilinear boundaries.* Master thesis. St. Petersburg State Polytechnical university, Dept. of Physics and Mechanics, 2001. (In Russian)

3. The case of incomplete elements. The multigrid solver can be adjusted to the important case of the incomplete finite p-elements, the use of which usually serves to reduce the number of unknowns at large p. For simplicity we look at the one type of incomplete elements namely at the minimal incomplete element with the internal subspace $\mathcal{U}_I = \mathcal{U}_{I,\min} := \mathrm{span}_{\omega_{I,\min}}[L_{\boldsymbol{\alpha}}]$, where the set $\omega_{I,\min}$ is defined in (7.83). Again, the internal stiffness matrix has the form $\mathbf{A}_I = \mathrm{diag}\,[\mathbf{A}_{e,e}, \mathbf{A}_{e,o}, \mathbf{A}_{o,e}, \mathbf{A}_{o,o}]$, but now blocks $\mathbf{A}_{a,b}$ have different dimensions. Nevertheless, they possess important common properties allowing to apply an optimal multigrid solver of the same type to all of them.

We retain the notations $\mathbf{\Lambda}_{e,e}$ and $\mathbf{B} = \mathbf{B}_{e,e}$ for the preconditioners of the matrix $\mathbf{A}_{e,e}$. Matrices $\mathbf{\Lambda}_{e,e}$ and $\mathbf{B}$ are obtained respectively from $\mathbf{\Lambda}_{e,e}$ and $\mathbf{B}$ for the complete reference element by crossing out the rows and columns, corresponding to the coordinate functions $L_{\boldsymbol{\alpha}}$, $\alpha_1 + \alpha_2 > p$. For convenience, we can assume that the square nests of the fine mesh are divided in pairs of triangles by diagonals joining the upper left and the lower right vertices. Let π_2 be the subdomain of π_1, which is below the straight line, joining the points $(\hbar, 1)$ and $(1, \hbar)$, $a^{(1)}_{\pi_2}(\cdot,\cdot)$ be the restriction of the bilinear form $a^{(1)}_{\pi_1}(\cdot,\cdot)$ to π_2, and $\mathring{\mathcal{H}}(\pi_2)$ be the space of functions from $\mathring{\mathcal{H}}(\pi_1)$, vanishing outside of π_2. It is easy to notice that the matrix $\mathbf{B}$ is induced by the bilinear form $a^{(1)}_{\pi_2}(\cdot,\cdot)$ on the subspace $\mathring{\mathcal{H}}(\pi_2)$.

Since $\mathcal{U}_{I,\min}$ is a subspace of the space $\mathcal{U}_I$ for the complete reference element, it is not necessary to formulate a new multigrid method and the results on its convergence. The matrices $\mathbf{B}_{l,w}$ and $\boldsymbol{\mathcal{B}}_{l,w}$ retain their forms (8.7),(8.8), but now the line $\Im^i_l \bigcap \pi_2$ has two disjoint segments, if $i > 2^{l-1}$. Accordingly, each block $\mathbf{B}^{(i)}_{l,w}$ with an odd $i = 2k+1 > 2^{l-1}$ contains two independent tridiagonal blocks each corresponding to one segment. Therefore, the implementation of the multigrid becomes even easier, also see [Korneev *et al.* (2002c,d)] on this issue. Clearly, the results on the convergence of the multigrid method for the reference elements with the complete spaces $\mathcal{U}_I$ are directly applicable to the multigrid method for the reference elements with the incomplete spaces $\mathcal{U}_{I,\min}$. This is because the analysis of the convergence is based on the bounds (8.9) and (8.10), which are retained for the subspaces of $\mathcal{U}_I$.

8.1.2 *Secondary DD Solver for the Local Dirichlet Problems*

As will be seen later, a variety of approaches can be employed for the construction of fast solvers for the systems governed by the finite-difference preconditioners for the internal stiffness matrices of the hierarchical reference p-elements, representatives of which are the systems (7.107) and (7.110). In this subsubsection we consider the domain decomposition BPS type solver. It is constructed with the use of the special orthogonal nonuniform deteriorating decomposition mesh, which has the number of subdomains proportional to $\log^2 p$. The decomposition is adapted to the behavior of the coefficients of the differential equations (7.109) and (7.112). It allows us to replace the coefficients of these differential equations by such subdomain-wise constant ones, which provide the equivalence of the energy integrals induced by the differential operators in (7.109) and (7.112). Therefore, discrete local Dirichlet problems on the nests of the decomposition mesh (secondary in relation to the local Dirichlet problems on the finite elements in the primal DD) result from the problems for the differential equations with the constant coefficients in the rectangular subdomains. This immediately allows one to optimize or almost optimize the main part of the computational work, since there is a variety of fast solvers for such boundary value problems. Thus, the construction of an efficient preconditioner-solver for the related interface Schur complement problem, the most involved part of the algorithm, is crucial for the efficiency of the entire solver. This is done with the help of the techniques developed in Chapter 6.

It is worth to remind that the geometrical objects of the outlined DD algorithm, like the decomposition and discretization meshes etc., have nothing in common with the geometry of the domain of the source boundary value problem and its decomposition. It is for the reason that the finite-difference interpretation of preconditioners Λ and $\Lambda_{e,e}$ is artificial. In order to distinguish between the primal DD solver for hp finite element systems (7.107) and (7.110) and the subsidiary domain decomposition solver for the local problems, arising in the primal solver, we use the abbreviation DD_{loc} for the subsidiary one.

The discretization, to which DD_{loc} corresponds, is an example of the family of the more general discretizations, studied in Chapter 6. As well the DD_{loc} preconditioner-solver itself for this special discretization will be basically a particular case of the ones considered in Chapter 6. At the same time particular properties of this important example of discretization make pos-

sible at least one essential improvement of the general algorithm. The main conclusion of this section is that the suggested $DD_{\rm loc}$ preconditioner-solver for systems (7.107) and (7.110) has total linear complexity for asymptotically optimal number $\mathcal{O}(p^2)$ of a.o. Additionally we note that $DD_{\rm loc}$ solvers are examples of efficient application of DD technique to the deteriorating and singular elliptic equations with variable in the magnitude and direction strong orthotropism.

8.1.2.1 *Finite-Difference Preconditioner Adapted to Local DD Solver for Hierarchical Reference Element*

Following [Korneev (2002b)], we introduce the rectangular deteriorating *decomposition mesh*, imbedded in the square *discretization mesh* of the size $\hbar = 1/(N+1)$, which will be also termed *fine mesh*. It subdivides π_1 in rectangular subdomains δ_{l_1,l_2}, such that in either of them the coefficients of equation (7.109) differ not too much from appropriate constants. The mesh lines of the fine and decomposition meshes are $\xi_k \equiv \eta_i$ and $\xi_k \equiv \overline{\xi}_{k,l} := \zeta_l$, respectively, where $k = 1, 2, \quad i = 0, 1, \ldots, N+1, \quad l = 0, 1, \ldots, \ell_0$, and the numbers ℓ_0 and ζ_l are as specified below. Indices l and i count numbers of lines of the coarse and fine meshes, respectively.

Two parameters $q > 1$ and $n_0 \geq 1$ can be used in order to generate the decomposition mesh as follows:

$$\begin{aligned} &\zeta_0 = 0, \quad \zeta_{l_0} = 1, \\ &\zeta_l = \eta_i, \quad \text{for} \quad i = \gamma(l) := \mathrm{int}\lfloor (q^l - 1)n_0 \rfloor, \quad l = 1, 2, \ldots, \ell_0 - 2, \\ &\zeta_{\ell_0 - 1} = \begin{cases} \eta_{\gamma(\ell_0-1)}, & \text{if } \exists\, l = \ell_0 : \ \eta_{\gamma(\ell_0)} = 1, \\ \eta_{i(\ell_0)}, & \text{if } \exists\, l = \ell_0 : \ \eta_{\gamma(\ell_0-1)} < 1 \text{ and } \eta_{\gamma(\ell_0)} > 1\,, \end{cases} \end{aligned} \tag{8.39}$$

where $i(\ell_0) = \mathrm{int}\lfloor (\gamma(\ell_0 - 2) + N + 1)/2 \rfloor$, and $\mathrm{int}\lfloor a \rfloor$ stands for the integer part of a.

One example of the decomposition mesh of such type is described as follows. It is easy to note that at the choice $q, n_0 = 2$, we need to satisfy $N = 2^{\ell_0+1} - 3$ in order $\eta_{\gamma(\ell_0)} = 1$ to be fulfilled. For $N = 30$, implying $p = 2N + 1 = 61$ and $\ell_0 = 4$, such fine and decomposition meshes are shown in Figure 8.2.

In what follows, for simplicity, we assume that for $l = 1, 2, \ldots, \ell_0$ all $\gamma(l)$ are different, which is true, *e.g.,* for $q \geq 1.5,\ n_0 \geq 2$. If $\gamma(\ell_0) = N+1$, then, from (8.39), we obtain

$$\ell_0 = \Im(q, n_0, N) = \mathrm{int}\lfloor \ln\left(\frac{N+1}{n_0} + 1\right) / \ln q \rfloor\,.$$

If $\gamma(\ell_0) \neq N+1$, then

$$\Im(q, n_0, N) \leq \ell_0 \leq \Im(q, n_0, N) + 1\,, \tag{8.40}$$

and, therefore, for q, n_0 fixed, we have

$$\ell_0 = \mathcal{O}(\ln N)\,. \tag{8.41}$$

Other formulas, retaining basic properties of (8.39), can be implemented for generation of the decomposition mesh. For instance, one can use the function $\gamma(l) = \mathrm{int}\langle (q^l - 1)n_0 \rangle$, where $\mathrm{int}\langle a \rangle$ is the integer closest to a. The same bound holds for ℓ_0. We introduce the piecewise constant functions

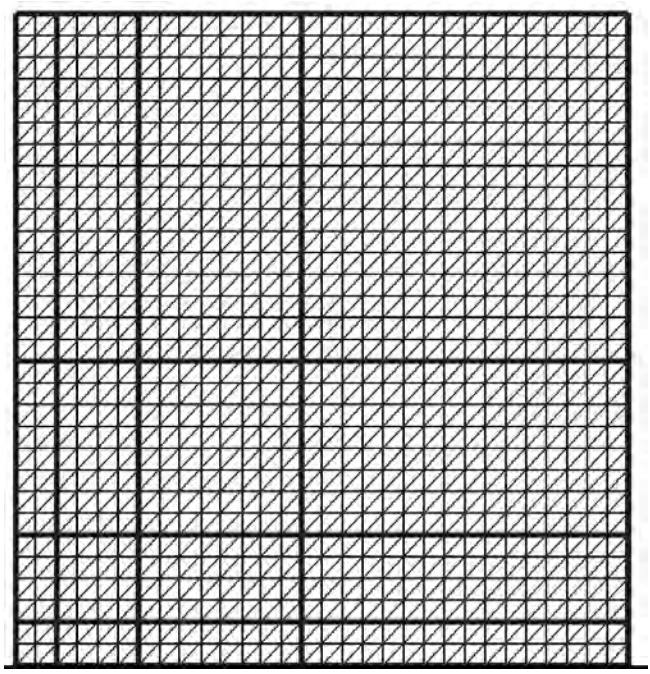

Fig. 8.2 Nonuniform domain decomposition and uniform discretization mesh.

$\psi(\zeta)$ and $b(\xi)$ of one and two variables, respectively:

$$\begin{array}{lll} \psi(\zeta) := \psi_k := \frac{1}{2}[\overline{\varphi}_{k-1} + \overline{\varphi}_k]\,, & \text{for} \quad \zeta \in (\zeta_{k-1}, \zeta_k)\,, \\ b(\xi) := b_{k,k} := 1 + \frac{1}{2}[\overline{\varphi}_k \overline{\varphi}_{k-1}^{-1} + \overline{\varphi}_{k-1}\overline{\varphi}_k^{-1}]\,, & \text{for} \quad \xi \in \delta_{k,k}\,, \\ b(\xi) := b_{k,l} := \frac{1}{2}\overline{\varphi}_{k,l} & \text{for} \quad \xi \in \delta_{k,l}\,, k \neq l\,, \end{array} \tag{8.42}$$

where $\overline{\varphi}_{k,l} = \overline{\varphi}_l \overline{\varphi}_{k-1}^{-1} + \overline{\varphi}_{k-1}\overline{\varphi}_l^{-1} + \overline{\varphi}_k \overline{\varphi}_{l-1}^{-1} + \overline{\varphi}_{l-1}\overline{\varphi}_k^{-1}$ for $k \neq l$, $\overline{\varphi}_k = \varphi_{\gamma(k)}$ for $k = 1, 2, \ldots, l_0$, $\overline{\varphi}_0 = \hbar^2$, and $\delta_{k,l} := \{\xi : \zeta_{k-1} < \xi_1 < \zeta_k\,,\ \zeta_{l-1} < \xi_2 < \zeta_l\}$.

Easy calculations lead us to the following property that is important for the sequel: the functions $\psi(\zeta)$ and $b(\xi)$ defined above satisfy the inequalities

$$\begin{gathered} c_1\psi(\zeta) \le \varphi(\zeta) \le c_2\psi(\zeta)\,, \quad \zeta \in (\hbar,1)\,, \\ c_1 b(\xi) \le \left[\frac{\varphi(\xi_1)}{\varphi(\xi_2)} + \frac{\varphi(\xi_2)}{\varphi(\xi_1)}\right] \le c_2 b(\xi)\,, \quad \xi \in (\hbar,1)\times(\hbar,1) \subset \pi_1\,, \end{gathered} \tag{8.43}$$

with $\varphi(\zeta)$, defined in (7.113), and positive constants $c_k = c_k(n_0, q)$ that is independent of p.

There are of course alternative ways of defining functions $\psi(\zeta)$ and $b(\xi)$ for obtaining a $DD_{\rm loc}$ preconditioner with the relative condition bounded by an absolute constant. For instance, $b_{k,l}$ can be defined as arithmetic means of $(\xi_1^2/\xi_2^2 + \xi_2^2/\xi_1^2)$ over $\delta_{k,l}$.

For the purpose of analysis of efficiency, it is convenient to define a version of the preconditioner as FE matrix. We subdivide each nest of the fine mesh by one of the diagonals in two triangles and introduce the space $\mathring{\mathcal{H}}(\pi_1)$ of the piecewise linear functions, which are continuous on π_1 and vanish on $\partial\pi_1$. We also introduce the bilinear forms

$$\begin{gathered} a_{\pi_1}(v,w) = a^{(1)}_{\pi_1}(v,w) + \hbar^2 \textstyle\sum_{i,j=1}^{N} \overline{b}_{i,j} v(\eta_i,\eta_j) w(\eta_i,\eta_j)\,, \\ a^{(1)}_{\pi_1}(v,w) = 2\int_{\pi_1} \left(\psi(\xi_2)\frac{\partial v}{\partial\xi_1}\frac{\partial w}{\partial\xi_1} + \psi(\xi_1)\frac{\partial v}{\partial\xi_2}\frac{\partial w}{\partial\xi_2}\right) d\xi, \end{gathered} \tag{8.44}$$

where

$$\overline{b}_{i,j} = \frac{1}{4}[b(\eta_i-0,\eta_j-0)+b(\eta_i+0,\eta_j-0)+b(\eta_i-0,\eta_j+0)+b(\eta_i+0,\eta_j+0)]\,.$$

As a counterpart of the preconditioners $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_{ee}$, introduced in Subsection 7.4.1, we define $\mathbf{B}^{(\hbar)} = \mathbf{B}^{(\hbar)}_{e,e}$ as the matrix of the bilinear form $\hbar^{-2}a^{(1)}_{\pi_1}(v,w)$ or bilinear form $\hbar^{-2}a_{\pi_1}(v,w)$ on the subspace $\mathring{\mathcal{H}}(\pi_1)$ with the choice of the nodal basis.

Lemma 8.3. *The matrices $\mathbf{B}^{(\hbar)}$ and $\mathbf{\Lambda}_{e,e}$ are spectrally equivalent uniformly in p,* i.e.*,*

$$c_1\mathbf{B}^{(\hbar)} \le \mathbf{\Lambda}_{e,e} \le c_2\mathbf{B}^{(\hbar)}\,, \tag{8.45}$$

with positive constants c_1 and c_2 depending only on n_0 and q.

Proof. Due to Lemma 7.5 and Theorem 7.3, the term with the multiplier h^2 in the bilinear form $a^{(1)}_{\pi_1}(v,w)$ can be omitted from consideration, and it is sufficient to consider only $\mathbf{B}^{(\hbar)}$ induced by the bilinear form $a^{(1)}_{\pi_1}(v,w)$. In turn, Lemma 8.1, in particular the inequalities (8.3), reduce the proof of (8.45) to the proof of the spectral equivalence of the matrices $\mathbf{B}$ and $\hbar^2\mathbf{B}^{(\hbar)}$, which is a direct consequence of inequalities (8.43). □

In view of their spectral equivalence, where it does not produce the confusion, we use the notation $a_{\pi_1}(\cdot,\cdot)$ for both bilinear forms (8.44) on the space $\mathring{\mathcal{H}}(\pi_1)$.

The discrete operators $\mathbf{B}^{(\hbar)}, \mathbf{\Lambda}$ and $\mathbf{\Lambda}_{ee}$ are defined on the vector space R^{N^2}, for which the notations $\mathbb{U} = \mathbb{U}^{e,e}$ will be used. Clearly, $\mathbb{U} \subset U_I$ and U_I is the space of vectors, corresponding to the subspace of internal polynomials $\mathring{\mathcal{Q}}_{p,x}$, whereas $\mathbb{U}$ corresponds to the subspace $\mathring{\mathcal{Q}}^{e,e}_{p,x}$, see (7.92). Obviously, we can represent $\mathbb{U}$ by the direct sums

$$\mathbb{U} = U_{\mathbb{I}} \oplus U_{\mathbb{III}} \quad \text{and} \quad \mathbb{U} = U_1 \oplus U_2 \oplus U_3\,,$$

where U_k, $k = 1, 2, 3$, are the subspaces of subvectors with entries related to the internal for subdomains $\delta_{k,l}$, edge and vertex degrees of freedom, respectively. The summands in the first sum are the subspaces $U_{\mathbb{I}} = U_1$ and $U_{\mathbb{III}} = U_2 \oplus U_3$.

These subspaces imply the respective block representations

$$\mathbf{B}^{(\hbar)} = \begin{pmatrix} \mathbf{B}_{\mathbb{I}} & \mathbf{B}_{\mathbb{I},\mathbb{III}} \\ \mathbf{B}_{\mathbb{III},\mathbb{I}} & \mathbf{B}_{\mathbb{III}} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{21} & \mathbf{B}_2 & \mathbf{B}_{23} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_3 \end{pmatrix}, \quad \mathbf{B}_{\mathbb{I}} = \mathbf{B}_1\,, \tag{8.46}$$

where, obviously, $\mathbf{B}_{31} = \mathbf{B}_{13} = \mathbf{0}$, and the dimensions of the blocks on the diagonal are

$$N_1 = (N + 1 - l_0)^2\,, \quad N_2 = 2(N + 1 - l_0)(l_0 - 1)\,, \quad N_3 = (l_0 - 1)^2\,.$$

The block $\mathbf{B}_1$ itself is the block-diagonal matrix

$$\mathbf{B}_1 = \mathrm{diag}\,[\,\mathbf{B}_1^{k,l}\,]_{k,l=1}^{\ell_0},$$

where each $(n_k - 1)(n_l - 1) \times (n_k - 1)(n_l - 1)$ block $\mathbf{B}_1^{k,l}$ corresponds to the internal d.o.f. of the subdomain $\delta_{k,l}$. Here, $n_k = \gamma(k) - \gamma(k-1)$.

It is clear that due to the discussed earlier spectral equivalences, structure of the reference element internal stiffness matrix $\mathbf{A}_I$, and the conditions of the generalized quasiuniformity, a fast solver for the system

$$\mathbf{B}^{(\hbar)}\mathbf{v} = \mathbf{f} \tag{8.47}$$

can be easily transformed into a fast solving procedure for the systems with the matrices $\mathbf{A}_{a,b}$, $a, b = e, o$, which are the blocks of the reference element internal stiffness matrix $\mathbf{A}_I$, and, therefore, to the matrix $\mathbf{A}_I$ itself, and to the finite element stiffness matrices $\mathbf{K}_{I_r}$. This is the reason that, according to Subsection 7.3.2 and Lemma 8.3, the matrix

$$\boldsymbol{\mathcal{A}}_I := \mathrm{diag}\,[\mathbf{B}^{(\hbar)}, \mathbf{B}^{(\hbar)}, \mathbf{B}^{(\hbar)}, \mathbf{B}^{(\hbar)}]$$

is spectrally equivalent to $\mathbf{A}_I$ for $p = 2N + 1$.

For restrictions of the subspace $\mathbb{U}$ and other vector subspaces to $\delta_{k,l}$, we use upper indices k, l. For instance, $U_i^{k,l}$ is the restriction of U_i to $\delta_{k,l}$. Additional lower indices i are used in the notations of subspaces $\mathcal{H}_i(\delta_{k,l})$ and $\mathcal{H}_i(\pi_1) \subset \mathring{\mathcal{H}}(\pi_1)$, where the former is the restriction of the latter to $\delta_{k,l}$ and the latter corresponds to U_i. Therefore, $\mathcal{H}_2(\delta_{k,l})$ is the restriction of $\mathring{\mathcal{H}}_2(\pi_1)$ to $\delta_{k,l}$, where $\mathring{\mathcal{H}}_2(\pi_1)$ is the subspace of FE functions, nodal values of which are distinct from zero only on the edges of the inter-subdomain boundary. At the same time $\mathcal{H}_1(\delta_{k,l})$ is the restriction of $\mathring{\mathcal{H}}_1(\pi_1)$ to $\delta_{k,l}$, and $\mathring{\mathcal{H}}_1(\pi_1)$ is the subspace of FE functions vanishing on the inter-subdomain boundary. Due to the finite element interpretation of $\mathbf{B}^{(\hbar)}$, this matrix as well as some others, used in the algorithm, can be assembled from the stiffness matrices for subdomains. Notations for subdomain stiffness matrices are supplied with upper indices k, l replacing the upper index $\hbar$. For instance, $\mathbf{B}^{(\hbar)}$ is assembled from the matrices $\mathbf{B}^{k,l}$, see the next paragraph for details.

8.1.2.2 *Outline of Schur Complement and DD Preconditioning*

Since we have several fast algorithms for solving the local Dirichlet problems, including direct ones, substructuring technique can efficiently be implemented. By eliminating the internal (for subdomains $\delta^{k,l}$) d. o. f., the system (8.47) is reduced to the system

$$\mathbb{S}_{\mathbb{II}} \mathbf{v}_{\mathbb{II}} = \mathbf{f}_{\mathbb{II}}^{\mathbb{I}}, \tag{8.48}$$

with the interface Schur complement matrix

$$\mathbb{S}_{\mathbb{II}} = \mathbf{B}_{\mathbb{II}} - \mathbf{B}_{\mathbb{II},\mathbb{I}} \mathbf{B}_{\mathbb{I}}^{-1} \mathbf{B}_{\mathbb{I},\mathbb{II}} = \begin{pmatrix} \mathbb{S}_E & \mathbf{B}_{2,3} \\ \mathbf{B}_{3,2} & \mathbf{B}_3 \end{pmatrix}, \tag{8.49}$$

where $\mathbb{S}_E = \mathbf{B}_2 - \mathbf{B}_{2,1} \mathbf{B}_I^{-1} \mathbf{B}_{1,2}$.

In general, explicit calculation of Schur complement can immediately kick the solver out of the range of optimality, because this matrix is densely populated. For the same reason, at the iterative solution of (8.48) by PCGM, multiplcations by Schur complement at each iteration may be also costly, especially when the preconditioner is not spectrally equivalent to the Schur complement and, therefore, the number of iterations grows with p. In the case under consideration matrices $\mathbb{S}_{\mathbb{II}}$ and $\mathbb{S}_E$ can be calculated with the use of the special trigonometric basis and FDFT for $\mathcal{O}(N^2)$ a.o., and the computational cost of the matrix-vector multiplication by $\mathbb{S}_{\mathbb{II}}$ is the same. The situation is aggravated by the fact that until now the preconditioner-solver for $\mathbb{S}_{\mathbb{II}}$, which is spectrally equivalent to this matrix uniformly in p, has the total computational cost cN^2, $c = \text{const}$, and is explicitly given,

has not been invented. From this it follows that the computational effort can be diminished only in the way presented in Chapter 6, *i.e.,* by the use of the preconditioner implicitly defined by the inexact solver for some cheap intermediate preconditioner-multiplicator.

Namely, for the Schur complement $\mathbb{S}_{\mathbb{II}}$, we design the preconditioner-solver

$$\boldsymbol{\mathcal{S}}_{\mathbb{II},\,\mathrm{it}} := \boldsymbol{\mathcal{I}}_{\circ}[\boldsymbol{\mathcal{S}}_{\mathbb{II},1}, \boldsymbol{\mathcal{S}}_{\mathbb{II},2}]\,, \tag{8.50}$$

which results from the inexact iterative solver, based on two explicitly defined preconditioners $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}$ and $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}$ for $\mathbb{S}_{\mathbb{II}}$. The preconditioners $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}$ and $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}$ will be chosen in such a way that the inequalities

$$\begin{aligned}
\underline{\beta}_1 \boldsymbol{\mathcal{S}}_{\mathbb{II},1} \le \mathbb{S}_{\mathbb{II}} \le \overline{\beta}_1 \boldsymbol{\mathcal{S}}_{\mathbb{II},1},\\
\underline{\beta}_2 \boldsymbol{\mathcal{S}}_{\mathbb{II},2} \le \boldsymbol{\mathcal{S}}_{\mathbb{II},1} \le \overline{\beta}_2 \boldsymbol{\mathcal{S}}_{\mathbb{II},2},\\
\underline{\beta}_{\mathrm{it}} \boldsymbol{\mathcal{S}}_{\mathbb{II},\,\mathrm{it}} \le \boldsymbol{\mathcal{S}}_{\mathbb{II},1} \le \overline{\beta}_{\mathrm{it}} \boldsymbol{\mathcal{S}}_{\mathbb{II},\,\mathrm{it}}
\end{aligned} \tag{8.51}$$

hold with sufficiently good betas. More specifically, $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}$ will be cheap for matrix-vector multiplications and spectrally equivalent to $\mathbb{S}_{\mathbb{II}}$ uniformly in p. The preconditioner-solver $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}$ will be cheap for operations $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}^{-1}\mathbf{v}$, $\forall \mathbf{v} \in U_{\mathbb{II}}$, while having as good as possible values $\underline{\beta}_2$ and $\overline{\beta}_2$, nevertheless dependent on p. The pointed out properties of the two Schur complement preconditioners will provide constant $\underline{\beta}_{\mathrm{it}}$ and $\overline{\beta}_{\mathrm{it}}$, independent of p, and cheap operations $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}^{-1}\boldsymbol{\mathcal{S}}_{\mathbb{II},1}\mathbf{v}$ for any $\mathbf{v} \in U_{\mathbb{II}}$.

The Dirichlet–Dirichlet-type DD preconditioner-solver for the system (8.47) is denoted by $\boldsymbol{\mathcal{B}}^{\hbar}$ and is defined through its inverse in the standard form

$$(\boldsymbol{\mathcal{B}}^{\hbar})^{-1} = \begin{pmatrix} \boldsymbol{\mathcal{B}}_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \mathbf{P}_1 \\ \mathbf{I} \end{pmatrix} \boldsymbol{\mathcal{S}}_{\mathbb{II}}^{-1} \begin{pmatrix} \mathbf{P}_1^{\top} & \mathbf{I} \end{pmatrix}, \tag{8.52}$$

where

- $\boldsymbol{\mathcal{B}}_1 = \mathrm{diag}\,[\,\boldsymbol{\mathcal{B}}_1^{k,l}\,]_{k,l=1}^{\ell_0}$ and each $\boldsymbol{\mathcal{B}}_1^{k,l}$ is the preconditioner-solver for the corresponding block $\mathbf{B}_1^{k,l}$ on the diagonal of $\mathbf{B}_1$,
- $\mathbf{P} = (\mathbf{P}_1^{\top}, \mathbf{I})^{\top} : U_{\mathbb{II}} \mapsto \mathbb{U}$ is an appropriate prolongation matrix,
- $\boldsymbol{\mathcal{S}}_{\mathbb{II}}$ is the preconditioner-solver for the interface Schur complement matrix, for which we will consider $\boldsymbol{\mathcal{S}}_{\mathbb{II}} = \boldsymbol{\mathcal{S}}_{\mathbb{II},\,\mathrm{it}}$.

If $\alpha_{k,l}(u,v)$ is the restriction to $\delta_{k,l}$ of the bilinear form $a_{\pi_1}^{(1)}(\ ,\ )$ in (8.44) multiplied by $\hbar^{-2}$, *i.e.*,

$$\alpha_{k,l}(u,v) = \frac{2}{\hbar^2} \int\limits_{\delta_{k,l}} \left(\psi_l \frac{\partial u}{\partial \xi_1}\frac{\partial v}{\partial \xi_1} + \psi_k \frac{\partial u}{\partial \xi_2}\frac{\partial v}{\partial \xi_2} \right) d\xi\,, \tag{8.53}$$

then the subdomain stiffness matrices are defined by the identities

$$\alpha_{k,l}(u,v) = \mathbf{u}^T \mathbf{B}_1^{k,l} \mathbf{v}\,, \quad \forall\ \mathbf{u},\mathbf{v} \leftrightarrow u, v \in \mathring{\mathcal{H}}_1(\delta_{k,l})\,.$$

Therefore, solving the system with the system matrix $\mathbf{B}_1$ is equivalent to solving systems

$$\mathbf{B}_1^{k,l}\,\mathbf{v} = \mathbf{f}\,, \tag{8.54}$$

the equations of which have the form

$$-2\psi_l \frac{v_{i-1,j} - 2v_{i,j} + v_{i+1,j}}{\hbar^2} - 2\psi_k \frac{v_{i,j-1} - 2v_{i,j} + v_{i,j+1}}{\hbar^2} = f_{i,j}\,, \tag{8.55}$$

with i,j in the range $\gamma(k-1) < i < \gamma(k)$, $\gamma(l-1) < j < \gamma(l)$, and the boundary conditions

$$v_{g,p} = 0 \quad \text{if } g = \gamma(k-1), \gamma(k) \ \text{ or } \ p = \gamma(l-1), \gamma(l). \tag{8.56}$$

Since ψ_k, ψ_l are constants, the solution of the system (8.55),(8.56) poses no problem. There are direct methods, combining, *e.g.*, 1d FDFT (Fast Discrete Fourier Transform) in the shorter direction and the Gauss elimination procedure in another. The latter is usually implemented in a special form termed as *Thomas algorithm*, named after Llewellyn Thomas, in [Conte and Boor (1972)], and *progonka* [Samarskii and Nikolayev (1989)]. Such a direct method for the system (8.55) and (8.56) spends $\mathcal{O}(n_k n_l \log \min(n_k, n_l))$ a.o. If there is no need in further optimization of the solver for systems (8.55), then it assumes $\boldsymbol{\mathcal{B}}_1^{k,l} = \mathbf{B}_1^{k,l}$ and $\boldsymbol{\mathcal{B}}_1 = \mathbf{B}_1$ in (8.52).

There is also a number of robust in respect to ratios ψ_k/ψ_l asymptotically optimal iterative algorithms, *i.e.,* requiring $\mathcal{O}(n_k n_l)$ a.o., based, for instance, on the multilevel preconditioners. The latter can be used in the fashion of inexact iterative solvers, *i.e.*, if $\mathbb{B}^{k,l}$ is such a multilevel preconditioner, then we set $\boldsymbol{\mathcal{B}}_1 = \mathbf{B}_{1,\mathrm{it}} := \boldsymbol{\mathcal{I}}_\circ[\mathbf{B}_1, \mathbb{B}_1]$. Two preconditioner-solvers of BPX type appropriate for achievement of optimal computational cost are described in Subsubsections 8.1.3.2 and 8.1.3.3.

In accordance with the discussion above, one may be sure to find such subdomain preconditioner-solvers $\boldsymbol{\mathcal{B}}_1^{k,l}$ that

$$\begin{aligned} &\underline{\gamma}_1 \boldsymbol{\mathcal{B}}_1 \prec \mathbf{B}_1 \prec \overline{\gamma}_1 \boldsymbol{\mathcal{B}}_1\,, \quad 0 < \overline{\gamma}_1, \overline{\gamma}_1 = \mathrm{const}\,, \\ &\mathrm{ops}[\boldsymbol{\mathcal{B}}_1^{-1}\mathbf{v}] = \mathcal{O}(N^2)\,, \quad \forall\ \mathbf{v} \in U_{\mathbb{I}}\,. \end{aligned} \tag{8.57}$$

The prolongation matrix $\mathbf{P}$ is "assembled" from the prolongation matrices $\mathbf{P}_{k,l}$ for the subdomains. We assume that they satisfy the inequalities

$$\begin{aligned} &\|\mathbf{P}_{k,l}\mathbf{v}\|_{\mathbf{B}_{\mathbb{II}}^{k,l}} \le c_{\mathrm{P}} \|\mathbf{v}\|_{\mathbb{S}_{\mathbb{II}}^{k,l}}\,, \\ &\mathrm{ops}\,[\mathbf{P}\mathbf{v}] = cN^2\,, \quad \forall\ \mathbf{v} \in U_{\mathbb{III}}\,, \end{aligned} \tag{8.58}$$

with constants independent of k, l, and p. In the literature, there is a number of suggestions for prolongation operators for uniformly orthotropic discretizations on rectangles, which satisfy these assumptions, see, *e.g.*, [Griebel and Oswald (1995b)], [Grauschopf *et al.* (1997)], and [Oswald (1999a)]. It was mentioned above that there are also the almost optimal direct solvers and optimal iterative solvers for the Dirichlet problems (8.55)–(8.56). Simultaneously with them, we can perform the prolongation operators by direct solvers and by inexact iterative solvers, which satisfy the first inequality (8.58), while the second inequality is satisfied with the additional multiplier $\log N$ on the right.

At the definition of the interface Schur complement preconditioner-solvers $\boldsymbol{\mathcal{S}}_{\mathbb{III},k},\ k = 1, 2$, we closely follow the approach of Section 6.4, however, with some simplifications permitted by the specific form of the problem.

Let us introduce the notations $\mathring{\mathcal{H}}_{V,\square}(\pi_1)$ and $\mathring{\mathcal{H}}_{V,\triangle}(\pi_1)$ for the spaces of continuous on π_1 piecewise bilinear on the decomposition mesh and, respectively, piecewise linear on the coarse triangulation functions. The corresponding finite element matrices are denoted by $\mathbf{B}_{V,\square}$ and $\mathbf{B}_{V,\triangle}$. In compliance with the conclusions of Section 6.4, we can use either of the matrices

$$\boldsymbol{\mathcal{S}}_{\mathbb{III},2} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \mathbf{B}_{V,\triangle}], \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \mathbf{B}_{V,\square}]\,.$$

However, for the discrete problem (8.47), we can also approve the uniform in p spectral equivalence of the preconditioner

$$\boldsymbol{\mathcal{S}}_{\mathbb{III},2} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \boldsymbol{\mathcal{D}}_V]\,, \tag{8.59}$$

with the diagonal matrix $\boldsymbol{\mathcal{D}}_V$, for which one can pick up $\boldsymbol{\mathcal{D}}_{V,\triangle} = \lfloor \mathbf{B}_{V,\triangle} \rfloor_{\mathrm{diag}}$ or $\boldsymbol{\mathcal{D}}_{V,\square} = \lfloor \mathbf{B}_{V,\square} \rfloor_{\mathrm{diag}}$, see Lemma 8.4.

The matrix $\boldsymbol{\mathcal{S}}_E$ is defined as in Subsection 6.3.3, meaning that it is assembled from the subdomain matrices

$$\mathbb{S}_E^{k,l} = \mathbf{B}_2^{k,l} - \mathbf{B}_{2,1}^{k,l} (\mathbf{B}_{\mathbb{I}}^{k,l})^{-1} \mathbf{B}_{1,2}^{k,l}\,,$$

similarly to $\mathbb{S}_{\mathbb{III}}$, which is assembled from the matrices

$$\mathbb{S}_{\mathbb{III}}^{k,l} = \mathbf{B}_{\mathbb{III}}^{k,l} - \mathbf{B}_{\mathbb{III},\mathbb{I}}^{k,l} (\mathbf{B}_{\mathbb{I}}^{k,l})^{-1} \mathbf{B}_{\mathbb{I},\mathbb{III}}^{k,l}\,.$$

We introduce the notations $E_\gamma^{k,l}$ for the edges of a subdomain $\delta_{k,l}$ on the lines $\xi_1 \equiv \zeta_{k-1}, \zeta_k$ and lines $\xi_2 \equiv \zeta_{l-1}, \zeta_l$, respectively. Each matrix $\mathbb{S}_E^{k,l}$ is represented in 4×4 block form

$$\mathbb{S}_E^{k,l} = \begin{pmatrix} \mathbb{S}_{E,1}^{k,l} & \mathbb{S}_{E,12}^{k,l} & \mathbb{S}_{E,13}^{k,l} & \mathbb{S}_{E,14}^{k,l} \\ \mathbb{S}_{E,21}^{k,l} & \mathbb{S}_{E,2}^{k,l} & \mathbb{S}_{E,23}^{k,l} & \mathbb{S}_{E,24}^{k,l} \\ \mathbb{S}_{E,13}^{k,l} & \mathbb{S}_{E,32}^{k,l} & \mathbb{S}_{E,3}^{k,l} & \mathbb{S}_{E,34}^{k,l} \\ \mathbb{S}_{E,14}^{k,l} & \mathbb{S}_{E,42}^{k,l} & \mathbb{S}_{E,34}^{k,l} & \mathbb{S}_{E,4}^{k,l} \end{pmatrix}, \tag{8.60}$$

where blocks $\mathbb{S}^{k,l}_{E,\gamma},\ \gamma = 1,2,3,4,$ on the diagonal correspond to the edges $E^{k,l}_{\gamma}$. The preconditioner $\boldsymbol{\mathcal{S}}^{k,l}_{E}$ for $\mathbb{S}^{k,l}_{E}$ is defined by decoupling in $\mathbb{S}^{k,l}_{E}$ the adjacent and shorter edges. In other words, we adopt

$$\boldsymbol{\mathcal{S}}^{k,l}_{E} = \begin{pmatrix} \mathbb{S}^{k,l}_{E,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbb{S}^{k,l}_{E,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}^{k,l}_{E,3} & \mathbb{S}^{k,l}_{E,34} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}^{k,l}_{E,34} & \mathbb{S}^{k,l}_{E,4} \end{pmatrix}, \qquad \text{for } k < l\,, \tag{8.61}$$

and use the symmetry for $l < k$, whereas for $k = l$ all edges are decoupled:

$$\boldsymbol{\mathcal{S}}^{k,k}_{E} = \operatorname{diag}[\mathbb{S}^{k,k}_{E,i}]_{i=1,2,3,4}\,. \tag{8.62}$$

We note that, instead of (8.61), it is possible to use the preconditioner

$$\boldsymbol{\mathcal{S}}^{k,l}_{E} = \begin{pmatrix} \mathbb{S}^{k,l}_{E,1} & \mathbb{S}^{k,l}_{E,12} & \mathbf{0} & \mathbf{0} \\ \mathbb{S}^{k,l}_{E,21} & \mathbb{S}^{k,l}_{E,2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}^{k,l}_{E,3} & \mathbb{S}^{k,l}_{E,34} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}^{k,l}_{E,34} & \mathbb{S}^{k,l}_{E,4} \end{pmatrix}, \qquad \text{for } k \neq l\,, \tag{8.63}$$

which leads to the same with (8.61) asymptotic estimates of the computational work of the Schur complement and domain decomposition algorithms.

The preconditioner $\boldsymbol{\mathcal{S}}_E$ is assembled from the subdomain preconditioners $\boldsymbol{\mathcal{S}}^{k,l}_{E}$. Several factors, which in part have been established earlier and in part will be approved in what follows, are important for its efficiency:

i) $\operatorname{cond}[\boldsymbol{\mathcal{S}}^{-1}_{E}\mathbb{S}_E] \prec \mathcal{O}(\log^2 N)$ is sufficiently good, see Theorem 6.5, p. 187,

ii) after the application of FDFT to the subvectors of degrees of freedom of each edge, the system $\boldsymbol{\mathcal{S}}_E\mathbf{x} = \mathbf{y}$, $\forall \mathbf{y} \in U_{\mathbb{III}}$ is transformed to another, solution of which requires $\mathcal{O}(N \log N)$ a.o.,

iii) matrices $\mathbb{S}_{\mathbb{III}}, \mathbb{S}_E$ and their preconditioners can be calculated for $\mathcal{O}(N^2)$ a.o., see in relation to *ii*) and *iii*) Subsubection 8.1.3.1,

iv) there are techniques like H-matrices and tensor-train decompositions which allow further reduce the computational cost of the implementation of such preconditioners, see, *e.g.*, [Khoromskij and Wittum (1999)], [Hsiao *et al.* (2001)], [Hackbusch *et al.* (2005)], and [Dolgov *et al.* (2011)].

Among a few options, one of the simplest efficient candidate for $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$ is the preconditioner $\boldsymbol{\mathcal{S}}_1$ assembled from subdomain preconditioners (6.151), p. 192. According to (6.152) it is spectrally equivalent to $\boldsymbol{\mathcal{S}}_{\mathbb{III}}$ uniformly in p and is cheap for matrix-vector multiplications. Since the orthotropism of the discrete problems (8.54) does not vary chaotically, but in a systematic fashion, there are less types of the subdomain preconditioners (6.151)

contributing to $\boldsymbol{\mathcal{S}}_{\mathbb{III},1} = \boldsymbol{\mathcal{S}}_1$, and its description is easier. Moreover, some important simplifications of the preconditioner of this type became possible. We consider this preconditioner-multiplicator with more details in the next subsubsection.

8.1.2.3 *Spectrally Equivalent Preconditioner Cheap for Multiplications*

The subdomains $\delta_{k,k}$ are squares, the bilinear forms $\alpha_{k,k}(\cdot,\cdot)$ are isotropic, and the spectrally equivalent preconditioners $\mathbb{S}^{k,k}_{\mathbb{III},1} := \boldsymbol{\mathcal{S}}_{1,(k,k)}$ such that $\mathrm{ops}\,[\mathbb{S}^{k,k}_{\mathbb{III},1}\mathbf{v}] = n_k \log n_k,\ k = 1, 2, \dots \ell_0$, are well known. Therefore, in view of the symmetry of the discrete operators in (8.55) with respect to the diagonal of π_1 connecting the vertices $(0,0), (1,1)$, it is necessary to consider, *e.g.*, only the subdomains $\delta_{k,l}$ for $l < k$. Schur complement preconditioners for these subdomains can be partitioned in two types, depending on the subsidiary coarser meshes, imbedded in the subdomain discretization mesh. We first remind the arrangements of the coarser meshes.

The change of variables $y_1 = (\xi_1 - \zeta_k),\ y_2 = \sqrt{\psi_l/\psi_k}\,(\xi_2 - \zeta_l)$ transforms the domain $\delta_{k,l}$ into $\pi^{k,l} = a_k \times \widetilde{a}_l$, and defines the *transformed discretization mesh* on $\pi^{k,l}$ with the nodes $y^{(i,j)} = (y_1^{(i)}, y_2^{(j)})$, where

$$y_1^{(i)} = (\eta_i - \zeta_k) \quad \text{and} \quad y_2^{(j)} = \sqrt{\psi_l/\psi_k}\,(\eta_j - \zeta_l)\,.$$

The steps of this mesh are denoted $\hbar_s$, $s = 1, 2$, with $\hbar_1 = \hbar$ and $\hbar_2 = \sqrt{\psi_l/\psi_k}\hbar$. Obviously, for $l < k$, we have

$$\hbar_2 < \hbar \quad \text{and} \quad a_k > a_l > \widetilde{a}_l\,. \tag{8.64}$$

Let $\widetilde{m}_k = \widetilde{m}_k(l) := \mathrm{int}\langle a_k/\widetilde{a}_l\rangle$. For rectangles $\pi^{k,l}$ satisfying

$$2 \le \widetilde{m}_k \le n_k\,, \tag{8.65}$$

we introduce the auxiliary rectangular quasiuniform (shape and size regular) grid, which is imbedded in the transformed discretization mesh and has the nests $\pi^{k,l}_{\varkappa},\ \varkappa = 1, 2, \dots, \widetilde{m}_k$, sufficiently close in sizes to the square $\pi^{k,l}_{\circ} = \widetilde{a}_l \times \widetilde{a}_l$. We call it the *subdomain coarse mesh* and use for it and its mesh lines the notations $\omega^{k,l}$ and $y_1^{(i_\varkappa)} = z^{(\varkappa)}$, respectively. The closeness of the nests to the square $\pi^{k,l}_{\circ}$ can be characterized by the positive constants $\alpha^{(1)}$ and $\alpha^{(2)}$ in the inequalities

$$\alpha^{(1)}\widetilde{a}_l \le a_{k,\varkappa} \le \alpha^{(2)}\widetilde{a}_l\,, \tag{8.66}$$

where $a_{k,\varkappa} = z^{(\varkappa)} - z^{(\varkappa-1)}$. We can easily show that subdomain coarse meshes can be constructed such that $\alpha^{(1)}$, $\alpha^{(2)}$ are absolute constants, *i.e.*, independent of p and k, l.

In the case $\widetilde{m}_k \geq n_k$ or, equivalently,

$$\widetilde{a}_l < \hbar \frac{n_k}{n_k + 0.5}\,,$$

implying condition (8.65) is not fulfilled, the local coarse mesh is introduced as the rectangular mesh with the nests $\pi_i^{k,l} = (y^{(i-1)}, y^{(i)}) \times (0, \widetilde{a}_l)$.

The *rarefied transformed mesh* at

$$\mathrm{int}\langle \widetilde{a}_l/\hbar \rangle > 1 \tag{8.67}$$

is the finest quasiuniform mesh imbedded in the transformed discretization mesh on $\pi^{k,l}$. In view of (8.64), the mesh parameter of such a mesh is $\hbar$. For subdomains with $\mathrm{int}\langle \widetilde{a}_l/\hbar \rangle \leq 1$ the rarefied and coarse meshes coincide.

It is worth noting that, for some arrangements of the decomposition mesh, the coarse and rarefied meshes are uniform, as is shown in Figure 6.3. One of them is obtained for the setting

$$\begin{gathered} p = 2N + 1\,, \quad N + 1 = 2^{\ell_0}\,, \quad \zeta_0 = 0\,, \quad \zeta_{\ell_0} = 1\,, \\ \gamma(k) = 2^k\,, \quad \zeta_k = 2^{k-\ell_0}\,, \quad \text{for} \quad k = 1, 2, \ldots, \ell_0. \end{gathered} \tag{8.68}$$

One difference from (8.39) is that the first step size in each direction is repeated twice and

$$n_1 = 2\,, \quad n_j = 2^{j-1}\,, \quad \text{for} \quad j = 2, 3, \ldots, \ell_0\,.$$

For more definiteness, jointly with (8.68) the piecewise constant functions ψ_k are accepted to be

$$\psi_k = \frac{1}{2}(\zeta_k^2 + \zeta_{k-1}^2) = \frac{1}{2}(2^{2(k-\ell_0)} + 2^{2(k-\ell_0-1)}) = \frac{5}{2}\,2^{2(k-\ell_0-1)}\,. \tag{8.69}$$

It is sufficiently clear that the *DD* algorithm efficient for the latter decomposition will be efficient for the decomposition mesh described earlier, and the asymptotic numerical complexities of both algorithms will be the same.

The preconditioner $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}^{k,l}$ gets the most simple form when the rarefied mesh coincides with the coarse mesh, *i.e.*, condition (8.67) is not fulfilled and $\widetilde{m}_k = n_k$. It corresponds to the representation of the subdomain finite element space by the direct sum similar to (6.102)

$$\mathcal{H}(\delta_{k,l}) = \mathbb{V}_{\mathrm{c}}(\delta_{k,l}) \oplus \mathbb{W}(\delta_{k,l}), \tag{8.70}$$

where $\mathbb{V}_{\mathrm{c}}(\delta_{k,l})$ is the space of the continuous functions bilinear on each nest $\pi_{\varkappa}^{k,l}$, $\varkappa = 1, 2, \ldots, n_k$.

Let $\mathbb{C}_{\mathrm{c}}^{k,l}$ be the matrix, induced by the bilinear form (8.53) on the space $\mathbb{V}_{\mathrm{c}}(\delta_{k,l})$. Note that nodes of the coarse mesh are on the boundary $\partial\pi^{k,l}$ and

this matrix is defined on their degrees of freedom. Let $\mathsf{X}_{\rm s}^{k,l} = \mathsf{X}_{\rm ls}^{k,l} \cup \mathsf{X}_{\rm rs}^{k,l}$ be the union of the discretization nodes on the left and right edges of the rectangle $\delta_{k,l}$, which are on the lines $y_1 \equiv 0$ and $y_1 \equiv a_k$, respectively, and do not include vertices of the rectangle $\pi^{k,l}$. The notations $U_{\rm s}^{k,l} = U_{\rm ls}^{k,l} \cup U_{\rm rs}^{k,l}$ stand for the corresponding spaces of the degrees of freedom. We introduce the block diagonal matrix $\boldsymbol{\mathcal{G}}_{\rm s}^{k,l} : U_{\rm s}^{k,l} \to U_{\rm s}^{k,l}$ by the relation

$$\boldsymbol{\mathcal{G}}_{\rm s}^{k,l} = \frac{\hbar_1}{\hbar_2} \operatorname{diag}\left[\, \boldsymbol{\Delta}_{2,\rm s}, \boldsymbol{\Delta}_{2,\rm s} \,\right], \tag{8.71}$$

where

$$\boldsymbol{\Delta}_{2,\rm s} = \operatorname{tridiag}\left[-1, 2, -1\right]_1^{n_l - 1}. \tag{8.72}$$

Each of two blocks in (8.71) is defined on the consecutively enumerated degrees of freedom of the nodes in the sets $\mathsf{X}_{\rm ls}^{k,l}$ and $\mathsf{X}_{\rm rs}^{k,l}$. According to the problem under consideration, the definitions introduced above and (6.151), the subdomain preconditioner-multiplicator is defined by the formula

$$\boldsymbol{\mathcal{S}}_{\rm III,1}^{k,l} = \sqrt{\psi_k\, \psi_l}\ [\boldsymbol{\mathcal{G}}_{\rm s}^{k,l}, \mathbb{C}_{\rm c}^{k,l}], \tag{8.73}$$

Under assumptions (8.68) and (8.69), we have

$$\begin{aligned} &a_k = 2^{k-\ell_0-1}, \quad \sqrt{\psi_l \psi_k} = \tfrac{5}{2}\, 2^{l+k-2(\ell_0-1)}, \quad \sqrt{\psi_l/\psi_k} = 2^{l-k}, \\ &\widetilde{a}_l/\hbar = \sqrt{\psi_l/\psi_k}(\zeta_l - \zeta_{l-1})/\hbar = 2^{l-k} \cdot 2^{l-\ell_0-1} \cdot 2^{\ell_0} = 2^{2l-k-1}, \end{aligned} \tag{8.74}$$

and the condition $\operatorname{int}\langle \widetilde{a}_l/\hbar \rangle \le 1$ can be reformulated as

$$l < 0.5(k+1) + \frac{\log 1.5}{\log 2}.$$

Therefore, for the considerable part of subdomains, the preconditioner-multiplicator is rather simple.

Now we turn to the case when condition (8.67) is fulfilled. As a consequence, the rarefied mesh does not coincide with the subdomain coarse mesh and has nodes inside $\delta_{k,l}$. The preconditioner-multiplicator is defined exactly as $\mathbf{C}$ in (6.151), *i.e.,* it can be written in the block form

$$\boldsymbol{\mathcal{S}}_{\rm III,1}^{k,l} = \sqrt{\psi_k\, \psi_l}\ [\boldsymbol{\mathcal{G}}_{\rm s}^{k,l}, \mathbb{C}_{\rm r}^{k,l}], \tag{8.75}$$

similar to (8.73), but with more complicated blocks $\boldsymbol{\mathcal{G}}_{\rm s}^{k,l}$ and $\mathbb{C}_{\rm r}^{k,l}$. Indeed, we have $\mathbb{C}_{\rm r}^{k,l} : U_{\rm B,r}^{k,l} \to U_{\rm B,r}^{k,l}$, where $U_{\rm B,r}^{k,l}$ is the space of degrees of freedom of the nodes of the rarefied mesh on the boundary $\partial\delta_{k,l}$.

Let m_k, m_l, where $m_k = n_k$, be the numbers of the intervals of the rarefied mesh on $\delta_{k,l}$, $\nu_{2,i}$ be the number of intervals of the transformed discretization mesh on the i-th interval of the rarefied mesh, $i = 1, 2, \ldots, m_2$,

and $\mathsf{X}^{k,l}_{\rm ls,i}$, $\mathsf{X}^{k,l}_{\rm rs,i}$ be the sets of nodes inside the i-th interval of the rarefied mesh on the left and right edge, respectively. The sets of the corresponding degrees of freedom of the nodes in $\mathsf{X}^{k,l}_{\rm ls,i}$, $\mathsf{X}^{k,l}_{\rm rs,i}$ are denoted $U^{k,l}_{\rm ls,i}$, $U^{k,l}_{\rm rs,i}$. We also introduce the matrices

$$\boldsymbol{\Delta}_{2,\rm s,i} = \mathrm{tridiag}\,[-1,2,-1]_1^{\nu_{2,i}-1} \quad \text{and} \quad \boldsymbol{\Delta}_{2,\rm rs} = \mathrm{diag}\,[\boldsymbol{\Delta}_{2,\rm s,i}]_{i=1}^{m_l}\,. \tag{8.76}$$

The block $\boldsymbol{\mathcal{G}}^{k,l}_{\rm s} : U^{k,l}_{\rm s} \to U^{k,l}_{\rm s}$ is defined by (8.71), but with $\boldsymbol{\Delta}_{2,\rm s}$ given in (8.76), and

$$U^{k,l}_{\rm s} = U^{k,l}_{\rm ls} \cup U^{k,l}_{\rm rs}\,, \quad U^{k,l}_{\rm ls} = \cup_{i=1}^{m_l} U^{k,l}_{\rm ls,i}\,, \quad U^{k,l}_{\rm rs} = \cup_{i=1}^{m_l} U^{k,l}_{\rm rs,i}\,.$$

If (8.68) and (8.69) hold, then

$$\nu_{2,i} = \frac{\hbar}{\hbar_2} \equiv 2^{k-l} \quad \text{and} \quad m_l = \max(1, \frac{\widetilde{a}_l}{\hbar}) = \max(1, 2^{2l-k-1})\,.$$

The definition of $\mathbb{C}^{k,l}_{\rm r}$ literally follows (6.60)–(6.62) with differences caused only by the specific forms of the continuous problems and their discretizations.

The subdomain coarse grid $\omega^{k,l}$ defines overlapping intervals $\tau_0 = (0, z^{(1)})$, $\tau_\varkappa = (z^{(\varkappa-1)}, z^{(\varkappa+1)})$, $\varkappa = 1, \ldots, \widetilde{m}_k - 1$, $\tau_{\widetilde{m}_k} = (z^{(\widetilde{m}_k-1)}, 1)$, and the subsets of the nodes of the rarefied grid $\mathsf{Z}^{k,l}_{B,i} = \overline{\tau}_i \cap \mathsf{Z}^{k,l}_B$, $i = 0, 1, \ldots, \widetilde{m}_k$, where $\mathsf{Z}^{k,l}_B$ is the subset of the nodes of the rarefied mesh living on the boundary $\partial\delta_{k,l}$. We also introduce the subsets $\mathsf{Z}^{k,l}_{E_j}$, $j = 2, 3$, of the nodes of the rarefied grid on the lines $y_2 \equiv 0, \widetilde{a}_l$ and the sets $\mathsf{Z}^{k,l}_{i,j} = \mathsf{Z}^{k,l}_{B,i} \cap \mathsf{Z}^{k,l}_{E_j}$. The subspaces of the degrees of freedom of the nodes from the introduced sets are denoted by $U^{k,l}_{B,i}$, $U^{k,l}_{E_j}$, $U^{k,l}_{i,j}$, respectively. Besides, we set $U^{k,l}_{E_=} = U^{k,l}_{E_2} \cup U^{k,l}_{E_3}$, where each subset $U^{k,l}_{E_\gamma}$ corresponds to the nodes on the closure of the edge $E^{k,l}_\gamma$.

The preconditioner-multiplicator will be defined by means of the matrices

$$\begin{aligned}
&\boldsymbol{\Delta}_{1/2,i} : U^{k,l}_{B,i} \to U^{k,l}_{B,i}\,, \qquad \text{for } i = 0, \widetilde{m}_k\,,\\
&\boldsymbol{\Delta}_{1/2,i,j} : U^{k,l}_{i,j} \to U^{k,l}_{i,j}\,, \qquad \text{for } i = 1, 2, , \ldots, \widetilde{m}_k - 1,\ j = 2, 3\,,\\
&\boldsymbol{\nabla} : U^{k,l}_{E_=} \to U^{k,l}_{E_=}\,.
\end{aligned} \tag{8.77}$$

In particular, we have

$$\boldsymbol{\Delta}_{1/2,i} = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & \mathbf{0} & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & \mathbf{0} & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}^{1/2}, \qquad i = 0, \widetilde{m}_k\,, \tag{8.78}$$

and the matrices $\boldsymbol{\Delta}_{1/2,i,j}$ have the same appearence and differ from (6.61) only in dimension. The unity blocks on the diagonal of the $2(n_k+1)\times 2(n_k+1)$ matrix

$$\boldsymbol{\nabla}=\frac{h}{\epsilon}\begin{pmatrix}\mathbf{I} & -\mathbf{I}\\ -\mathbf{I} & \mathbf{I}\end{pmatrix}$$

correspond to the degrees of freedom $U^{k,l}_{E_2}$ and $U^{k,l}_{E_3}$ of the nodes on the edges $\overline{\gamma}_2$ and $\overline{\gamma}_3$, respectively.

For the decomposition mesh (8.68) and ψ_k defined in (8.69), we have

$$\widetilde{m}_k=\tfrac{a_k}{\widetilde{a}_l}=2^{2(k-l)}\,,\quad \dim\left[\boldsymbol{\Delta}_{1/2,i}\right]=3\tfrac{\widetilde{a}_l}{\hbar}+1=3m_l+1=3\cdot 2^{2l-k-1}+1\,,$$
$$\dim\left[\boldsymbol{\Delta}_{1/2,i,j}\right]=2\tfrac{\widetilde{a}_l}{\hbar}+1=2m_l+1=2^{2l-k}+1\,,$$

with two last numbers denoting the dimensions of the matrices in the first and second rows of (8.77).

Now, the second block in the preconditioner-multiplicator (8.75) is defined by the topological sum of the matrices

$$\mathbb{C}^{k,l}_{\mathrm{r}}=\boldsymbol{\nabla}+\boldsymbol{\Delta}_{1/2,0}+\boldsymbol{\Delta}_{1/2,\widetilde{m}_k}+\sum_{j=2,3}\sum_{i=1}^{\widetilde{m}_k-1}\boldsymbol{\Delta}_{1/2,i,j}\,. \tag{8.79}$$

Multiplications $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}\mathbf{v}$ are produced by means of subdomainwise multiplications $\boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{III},1}\mathbf{v}^{k,l}$, while subdomain preconditioners $\boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{III},1}$ can be based, *e.g.*, on different for different subdomains types of multilevel decompositions of FE spaces $\mathcal{H}(\delta_{k,l})$. This simplifying factor results in several additional options for choosing $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$ suitable for our purposes. Among them boundary Schur complement preconditioners resulting from multilevel FE space decomposition techniques, considered by [Griebel and Oswald (1995b)] and [Oswald (1999a)], can be considered. The techniques listed in the item *iii*) of Subsubsection 8.1.2.2 can be also attracted for efficient boundary Schur complement preconditioning in the case under consideration. In particular, a sparse and accurate approximation of $\mathbb{S}^{k,l}_{\mathbb{III}}$ of the complexity $\mathcal{O}(n_{k,l}\log^2 n_{k,l})$, $n_{k,l}=n_k+n_l$, and the same complexity for the matrix-vector multiplications was considered in [Khoromskij and Wittum (1999)].

8.1.2.4 *Computational Complexity*

The fast DD_{loc} and Schur complement solvers, presented in the previous subsubection, were constructed following the approach of Chapter 6. On this way, the uncertainty with the solver for the vertex subproblem and its computational cost was removed, since the diagonal matrix was suggested for the preconditioner. It is left to bound its relative condition number.

Lemma 8.4. *Matrices* $\mathbf{B}_V = \mathbf{B}_{V,\square}, \mathbf{B}_{V,\triangle}$ *and* $\boldsymbol{\mathcal{D}}_V$ *are spectrally equivalent uniformly in* p, i.e.,

$$\underline{\gamma}_V \boldsymbol{\mathcal{D}}_V \le \mathbf{B}_V \le \overline{\gamma}_V \boldsymbol{\mathcal{D}}_V \,, \tag{8.80}$$

with positive constants $\underline{\gamma}_V$ *and* $\overline{\gamma}_V$*, depending only on* n_0 *and* q.

Proof. Alongside with (6.148), we have the spectral equivalence inequalities

$$\frac{1}{3}\mathbf{B}_{V,\triangle} \le \mathbf{B}_{V,\square} \le \mathbf{B}_{V,\triangle} \quad \text{and} \quad \frac{1}{3}\boldsymbol{\mathcal{D}}_{V,\triangle} \le \boldsymbol{\mathcal{D}}_{V,\square} \le \boldsymbol{\mathcal{D}}_{V,\triangle} \,, \tag{8.81}$$

and, therefore, it is sufficient to prove (8.80) for the matrices $\mathbf{D}_{V,\triangle}$ and $\mathbf{B}_{V,\triangle}$ only. Matrix $\mathbf{B}_{V,\triangle}$ is represented by the sum of Kronecker products

$$\mathbf{B}_{V,\triangle} = \boldsymbol{\Delta}_H \otimes \mathbf{T}_H + \mathbf{T}_H \otimes \boldsymbol{\Delta}_H \,, \tag{8.82}$$

with the matrices

$$\boldsymbol{\Delta}_H = \operatorname{tridiag}\,[-\frac{1}{H}_i, \frac{1}{H}_i + \frac{1}{H}_{i+1}, -\frac{1}{H}_{i+1}]_{i=1}^{\ell_0-1} \quad \text{and}$$
$$\mathbf{T}_H = \operatorname{diag}\,[\frac{1}{2}(\psi_i H_i + \psi_{i-1} H_{i+1})]_{i=1}^{\ell_0-1} \,.$$

To simplify further considerations, we assume that the decomposition mesh and the coefficients ψ_k are defined as in (8.68) and (8.69). In view of (8.82) and these assumptions, it is sufficient to show the spectral equivalence

$$\mathbb{D}_H \prec \boldsymbol{\Delta}_H \prec \mathbb{D}_H \,, \tag{8.83}$$

where

$$\mathbb{D}_H = \operatorname{diag}\,[\frac{1}{H}_i + \frac{1}{H}_{i+1}]_{i=1}^{\ell_0-1} \,.$$

The proof of the inequalities of the same type as (8.83) is known in the literature for more general discretizations with the sizes of finite elements growing proportionally to the distance from the boundary. For (8.83), it is especially simple. The matrix $\boldsymbol{\Delta}_H$ can be factorized in the form

$$\boldsymbol{\Delta}_H = \boldsymbol{\mathcal{D}}_\circ \boldsymbol{\Delta}_\circ \,,$$

where

$$\boldsymbol{\mathcal{D}}_\circ = \operatorname{diag}\,[\,\frac{2^i}{\hbar}\,]_{i=1}^{\ell_0-1} \,, \quad \boldsymbol{\Delta}_\circ = \operatorname{tridiag}\,[-1\,, \frac{1}{2}(3+\delta_{1,i})\,, -\frac{1}{2}(1+\delta_{1,i})]_{i=1}^{\ell_0-1} \,,$$

and $\delta_{1,i}$ Kronecker's symbol. For $\epsilon = 1/\sqrt{2}$, we have

$$-\epsilon + \frac{1}{2}(3+\delta_{1,i}) - \frac{1+\delta_{1,i}}{2\epsilon} \ge \frac{1}{12} \,.$$

The use of the Cauchy inequality with epsilon

$$2ab \le \epsilon a^2 + \epsilon^{-1} b^2 \,, \quad \forall\, \epsilon > 0 \,,$$

for "eliminating" the off-diagonal entries of the matrix $\boldsymbol{\Delta}_H$ has the straightforward consequence

$$\boldsymbol{\Delta}_H \ge \frac{1}{18}\, \mathbb{D}_H \,.$$

This is the left inequality (8.83). Since inequality $\boldsymbol{\Delta}_H \le 2\mathbb{D}_H$ is obvious, the proof is complete. □

A smoother, than in the bilinear form(6.146), change of the coefficients of the bilinear form $a_{\pi_1}(\cdot,\cdot)$ and $a^{(1)}_{\pi_1}(\cdot,\cdot)$ in (8.44) might give the hope that the relative condition number of the interface Schur complement preconditioner and, as a consequence, of the $DD_{\rm loc}$ preconditioner could be improved. Indeed, let $\underline{n}_{k,l} = \min(n_k, n_l)$. For about a half of subdomains $\delta_{k,l}$, we can establish the energy inequalities

$$\frac{1}{\overline{n}_{k,l}} \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II},2} \prec \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II}} \prec \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II},2} \,, \quad \text{for } l \le \frac{1}{2}k + 1 \text{ or } k \le \frac{1}{2}l + 1 \,,$$

and

$$\frac{1}{1 + \log^2 \underline{n}_{k,l}} \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II},2} \prec \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II}} \prec \boldsymbol{\mathcal{S}}^{k,l}_{\mathbb{II},2} \,, \quad \text{for } |l - k| \le 2 \,,$$

which are better than those for the remaining subdomains.

Theorem 8.5. *Let $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}$ and $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}$ be the preconditioners introduced above in (8.73),(8.75),(8.79) and (8.59), (8.61), respectively. Let also the preconditioner $\boldsymbol{\mathcal{S}}_{\mathbb{II},\,\rm it}$ be defined by the inexact iterative solver (8.50) for $k_s = \nu_{1/2}$ iterations. Then, in the inequalities (8.51), $\underline{\beta}_1, \overline{\beta}_1, \overline{\beta}_2 = \text{const}$,*

$$\underline{\beta}_2^{-1} \succ N(1 + \log N) \,, \tag{8.84}$$

and, therefore,

$$\nu_{1/2} \asymp \sqrt{\overline{\beta}_1 (1 + \log N) N / \underline{\beta}_1}, \tag{8.85}$$

$\underline{\beta}_{\rm it}, \overline{\beta}_{\rm it} = \text{const}$. *The constants in the presented relations depend only on n_0 and q.*

Proof. Since the preconditioner $\boldsymbol{\mathcal{S}}_{\mathbb{II},1}$ is a particular case of the preconditioner $\boldsymbol{\mathcal{S}}_1$, one has $\underline{\beta}_1, \overline{\beta}_1 = \text{const}$ in agreement with (6.152). The preconditioner $\boldsymbol{\mathcal{S}}_{\mathbb{II},2}$ differs from $\boldsymbol{\mathcal{S}}_2$ only by the choice of the preconditioner for the vertex subproblem. Therefore, (6.152)–(6.154) and (8.80) approve the values of $\underline{\beta}_2, \overline{\beta}_2$. The conclusion on $\underline{\beta}_{\rm it}, \overline{\beta}_{\rm it}$ follows by Lemma 2.1 and the bounds for $\underline{\beta}_2, \overline{\beta}_2$ and $\nu_{1/2}$. □

Let us turn to the estimates of the arithmetical work for solving systems (8.48) and (8.47), denoting the corresponding numbers of arithmetic operations by Q_S and Q_B, respectively. It is easy to see that, for the matrices entering (8.71), (8.79) and corresponding vectors $\mathbf{v}$, we obtain the bounds

$$\begin{aligned}
&\mathrm{ops}[\boldsymbol{\mathcal{G}}_{\mathrm{s}}^{k,l}\mathbf{v}] \prec n_l\,,\\
&\mathrm{ops}[\boldsymbol{\nabla}^{k,l}\mathbf{v}] \prec n_k\,,\\
&\mathrm{ops}[\boldsymbol{\Delta}_{1/2,i}^{k,l}\mathbf{v}] \prec (1+\log m_l)m_l\,,\\
&\mathrm{ops}[\boldsymbol{\Delta}_{1/2,i,j}^{k,l}\mathbf{v}] \prec (1+\log m_l)m_l\,,
\end{aligned}$$

where the matrices from the expression (8.79) are supplied with additional upper indices k, l. We first sum these bounds for one subdomain $\delta_{k,l}$ and come to

$$\mathrm{ops}[\boldsymbol{\mathcal{S}}_{\mathrm{III},1}^{k,l}\mathbf{v}] \prec (1+\log m_l)(n_l+n_k)+n_k \prec (1+\log m_l)n_k\,.$$

Then the summation over $k, l = 1, 2, \ldots, \ell_0$ yields

$$\mathrm{ops}[\boldsymbol{\mathcal{S}}_{\mathrm{III},1}\mathbf{v}] \prec (1+\log N)^2 N\,, \quad \forall\, \mathbf{v} \in U_{\mathrm{III}}\,,$$

and, in view of (8.85), we get the estimates

$$\begin{aligned}
k_s \cdot \mathrm{ops}[\boldsymbol{\mathcal{S}}_{\mathrm{III},1}\mathbf{v}_B] &\prec k_s(1+\log N)^2 N\\
&\prec \sqrt{\overline{\beta}_1/\underline{\beta}_1}(1+\log N)^{5/2}N^{3/2}\\
&\prec (1+\log N)N_{\mathrm{III}}^{3/2}\,,
\end{aligned} \tag{8.86}$$

where $N_{\mathrm{III}} = N_2 + N_3$ is the number of unknowns in the Schur complement problem.

Results of the next paragraph approve the estimate

$$\mathrm{ops}\,[\boldsymbol{\mathcal{S}}_{\mathrm{III},2}^{-1}\mathbf{v}] \prec (1+\log N)^2 N + \log^2 N\,, \quad \forall\ \mathbf{v} \in U_{\mathrm{III}}\,,$$

under assumption that the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$, see (8.92), of the quadratic form $\boldsymbol{\mathcal{S}}_E$ after applying FDFT to the vectors of unknowns of each edge, is known. In this estimate first term is related to the operation $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{-1}\mathbf{v}_{E,\mathrm{tr}}$ and FDFT to the initial basis. Second term estimates the arithmetical cost of $\boldsymbol{\mathcal{D}}_V^{-1}\mathbf{v}_V$. However, the very calculation of $\boldsymbol{\mathcal{S}}_E$, which can be done once, requires $\mathcal{O}(N^2)$ a.o. Therefore, combining with (8.85) and (8.86), we get

$$\mathrm{ops}\,[\boldsymbol{\mathcal{S}}_{\mathrm{III},\,\mathrm{it}}^{-1}\mathbf{v}] \prec N^2 + \aleph\,, \quad \forall\ \mathbf{v} \in U_{\mathrm{III}}\,, \tag{8.87}$$

where

$$\aleph = \sqrt{(1+\log N)N}((1+\log N)^2 N + \log^2 N)$$

and first term in (8.87) reflects the calculation of $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$.

In the Schur complement and DD algorithms under consideration, the systems (8.48) and (8.47) are solved by PCGM with the preconditioners $\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\mathrm{it}}$ and $\boldsymbol{\mathcal{B}}^{\hbar}$, respectively. Let Q_S and Q_{DD} denote the corresponding numbers of arithmetic operations for these solvers. Since, according to (8.51)

$$\underline{\beta}_1\underline{\beta}_{\mathrm{it}}\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\mathrm{it}} \le \mathbb{S}_{\mathbb{III}} \le \overline{\beta}_1\overline{\beta}_{\mathrm{it}}\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\mathrm{it}}$$

and according to Theorem 8.5 all betas in (8.1.2.4) are constants depending only on n_0 and q, we get

$$Q_S \prec N^2 + \aleph\sqrt{\overline{\beta}_1\overline{\beta}_{\mathrm{it}}/\underline{\beta}_1\underline{\beta}_{\mathrm{it}}} \prec N^2 + \aleph\,.$$

Taking additionally into consideration assumptions (8.57) on the preconditioner-solvers for subdomain Dirichlet problems and assumptions (8.58) on the prolongation operator results in the bound

$$Q_{DD} \prec c_{\mathrm{P}}\sqrt{\max(\overline{\beta}_1\overline{\beta}_{\mathrm{it}},\overline{\gamma}_1)/\min(\underline{\beta}_1\underline{\beta}_{\mathrm{it}},\underline{\gamma}_1)}\;(N^2+\aleph) \prec (N^2+\aleph)\,, \quad (8.88)$$

approving the optimality of the DD algorithm. Thus, we have proved

Corollary 8.3. *Let the assumptions of the Theorem 8.5 and the inequalities (8.57), (8.58) be fulfilled. Then the* DD_{loc} *solver has linear computational complexity and the number of arithmetic operations is estimated by (8.88).*

8.1.3 *Two Main Modulus of DD Solver for Local Dirichlet Problems*

In this subsection, we more elaborately consider the preconditioner-solvers for the edge subproblem and for the internal Dirichlet problems on the subdomains of domain decomposition.

8.1.3.1 *Compatible Preconditioner for the Edge Problem*

First of all, we present additional details on the edge block $\mathbb{S}_E^{k,l}$ of the interface Schur complement $\mathbb{S}_{\mathbb{III}}^{k,l}$ and its preconditioner $\boldsymbol{S}_E^{k,l}$ found in (8.60) and (8.61), respectively, and on the computation of these matrices.

Here it is convenient to use another notations for the edges of subdomains. Let $T_{s,t}$ be the edge of the rectangle $\delta_{k,l}$ on the line $\xi_{3-s} \equiv \zeta_{i_s-1+t}$ of the decomposition mesh, where $s = 1, 2$, $t = 0, 1$ and $i_1 = k$, $i_2 = l$, and $\boldsymbol{\mu}_{s,t}^{(\alpha)}$ be the corresponding trigonometric discrete harmonic basis vectors. They are defined by the two basic properties:

a) the traces of $\boldsymbol{\mu}_{s,t}^{(\alpha)}$ at the nodes on $\partial\delta_{k,l}$ are zeroes, except for the edge $T_{s,t} \subset \pi_1$, where their components are

$$\begin{aligned} &\mu_{1,0}^{(\alpha)}(i,0) = \mu_{1,1}^{(\alpha)}(i,n_l) = \sin\alpha\,\tfrac{i\pi}{n_k}\,, \quad \alpha, i = 1,2,\ldots,n_k-1\,, \\ &\mu_{2,0}^{(\beta)}(0,j) = \mu_{2,1}^{(\beta)}(n_k,j) = \sin\beta\,\tfrac{j\pi}{n_l}\,, \quad \beta, j = 1,2,\ldots,n_l-1\,, \end{aligned} \tag{8.89}$$

b) vectors $\boldsymbol{\mu}_{s,t}^{(\alpha)}$ are discrete harmonic, *i.e.*,

$$(\boldsymbol{\mu}_{s,t}^{(\alpha)})^T \mathbf{B}^{k,l}\mathbf{v} = 0\,, \quad \text{for all } \mathbf{v} \in U_{I,1}^{(k,l)}\,. \tag{8.90}$$

It is easy to note that the basis $\{\boldsymbol{\mu}_{s,t}^{(\alpha)}\}$ is compatible with similar bases for other subdomains $\delta_{k,l}$. This allows us to define the discrete harmonic trigonometric basis with the identical to (8.89) traces on the edges of subdomains $\delta_{k,l}$, and, in turn, to define Schur complement $\mathbb{S}_{E,\mathrm{tr}}$ in this basis and its preconditioner $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ by assembling matrices $\mathbb{S}_{E,\mathrm{tr}}^{k,l}$ and $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}$. First of all, we discuss the structure of preconditioners $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}, \boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$.

Lemma 8.5. *For an appropriate ordering of the unknowns, the matrix* $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ *is the block diagonal matrix with* $2(N-l_0+1)$ *independent blocks, each of which is a* $(l_0-1)\times(l_0-1)$ *three-diagonal matrix.*

Proof. A direct consequence of (8.89), (8.90) and symmetry properties of subdomains $\delta_{k,l}$ and bilinear forms $a_{k,l}(u,v)$ are the following equalities defining structure of $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$:

$$\left.\begin{aligned} &b_{k,l}(s,\alpha) = (\boldsymbol{\mu}_{s,t}^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_{s,t}^{(\alpha)})_{k,l}\,, \\ &c_{k,l}(s,\alpha) = (\boldsymbol{\mu}_{s,0}^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_{s,1}^{(\alpha)})_{k,l}\,, \\ &(\boldsymbol{\mu}_{s,t}^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_{s,m}^{(\beta)})_{k,l} = 0\,, \end{aligned}\right\} \quad \begin{aligned} &t \neq m\,, \quad \alpha \neq \beta\,, \\ &\alpha,\beta = 1,2,\ldots,n^{(s)}-1\,, \end{aligned} \tag{8.91}$$

where $n^{(1)} = n_k$, $n^{(2)} = n_l$ and $(\mathbf{v}, \mathbf{B}^{k,l}\mathbf{v})_{k,l} := \mathbf{v}^\top \mathbf{B}^{k,l}\mathbf{w}$. In particular, these equalities follow from orthogonality relations (8.99) and representations (8.100) of the involved vectors. Therefore, in the discrete harmonic trigonometric basis, (8.63) turns into the matrix

$$\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l} = \begin{pmatrix} \mathbb{S}_{E,1,\mathrm{tr}}^{k,l} & \mathbb{S}_{E,12,\mathrm{tr}}^{k,l} & \mathbf{0} & \mathbf{0} \\ \mathbb{S}_{E,21,\mathrm{tr}}^{k,l} & \mathbb{S}_{E,2,\mathrm{tr}}^{k,l} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}_{E,3,\mathrm{tr}}^{k,l} & \mathbb{S}_{E,34,\mathrm{tr}}^{k,l} \\ \mathbf{0} & \mathbf{0} & \mathbb{S}_{E,34,\mathrm{tr}}^{k,l} & \mathbb{S}_{E,4,\mathrm{tr}}^{k,l} \end{pmatrix},$$

all nonzero blocks of which are diagonal.

For the special local ordering of d.o.f., we have

$$\boldsymbol{\mathcal{S}}^{k,l}_{E,\mathrm{tr}} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}^{k,l}(s,\alpha)]^{2;n^{(s)}-1}_{s=1;\alpha=1} \;\text{ and }\; \boldsymbol{\mathcal{S}}^{k,l}(s,\alpha) = \begin{pmatrix} b_{k,l}(s,\alpha) & c_{k,l}(s,\alpha) \\ c_{k,l}(s,\alpha) & b_{k,l}(s,\alpha) \end{pmatrix},$$

with the rows and columns in the 2×2 matrix $\boldsymbol{\mathcal{S}}^{k,l}(s,\alpha)$ corresponding to $t=0,1$. In this ordering, matrix $\boldsymbol{\mathcal{S}}^{k,l}$ is block diagonal with only 2×2 blocks, the number of which is n_k+n_l-2. This results in a simple structure of the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$. We can easily enumerate interface unknowns in such a way that this matrix will have $2(N-\ell_0+1)$ independent $(\ell_0-1)\times(\ell_0-1)$ blocks on its diagonal, which are denoted $\boldsymbol{\mathcal{S}}(s,k,\alpha)$, *i.e.*,

$$\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}(s,k,\alpha)]\,, \tag{8.92}$$

with $s=1,2$, $k=1,2,\ldots,\ell_0$ and $\alpha=1,2,\ldots,n^{(s)}-1$.

Each block $\boldsymbol{\mathcal{S}}(s,k,\alpha)$ is the tridiagonal matrix

$$\boldsymbol{\mathcal{S}}(s,k,\alpha) = \begin{pmatrix} b_1 & c_1 & & & & \\ & b_2 & c_2 & & 0 & \\ & & \cdots & \cdots & & \\ & \mathrm{SYM} & & \cdots & \cdots & \\ & & & & b_{\ell_0-2} & c_{\ell_0-2} \\ & & & & & b_{\ell_0-1} \end{pmatrix}, \tag{8.93}$$

in which $b_l = b_{k,l}(s,\alpha) + b_{k,l+1}(s,\alpha)$, $c_l = c_{k,l+1}(s,\alpha)$. We see that the system with the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ is the union of $2(N-\ell_0+1)$ independent subsystems of ℓ_0-1 algebraic equations, each with a tridiagonal matrix. Therefore, the total fill in of the matrix of this system is $2(\ell_0-1)(N-\ell_0+1)=\mathcal{O}(N\log N)$, and it can be solved for not more than $\mathcal{O}(N\log N)$ a.o. To attain the latter bound, it is sufficient to implement the *progonka* respectively Thomas method, see [Samarskii and Nikolayev (1989)], for solving each independent system with the tridiagonal matrix. The lemma is proved □

Let us remark that although $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ is the major part of $\mathbb{S}_{E,\mathrm{tr}}$, the fill in of the latter is $\mathcal{O}(N^2)$, *i.e.,* much larger than of the former. It is for the reason that off-diagonal blocks of $\mathbb{S}_{E,\mathrm{tr}}$, coupling the adjacent edges, are densely filled in, although they play a secondary role and can be neglected in the preconditioner without serious damage to the relative condition number.

Theorem 8.6. *Let $Q^{(1)},Q^{(2)},Q^{(3)}$ be respectively the arithmetical costs of*

1) computation of matrices $\mathbb{S}_{E,\mathrm{tr}}$ and $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$,

2) solving systems with the matrices $\boldsymbol{\mathcal{S}}_E, \boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$, *and*
3) solving the systems

$$\mathbb{S}_E \mathbf{v}_E = \mathbf{f}_E \quad \textit{and} \quad \mathbb{S}_{E,\mathrm{tr}} \mathbf{v}_{E,\mathrm{tr}} = \mathbf{f}_{E,\mathrm{tr}} \tag{8.94}$$

by the PGCM with the preconditioner $\boldsymbol{\mathcal{S}}_E$ *for the matrix* $\mathbb{S}_E$ *and the preconditioner* $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ *for the matrix* $\mathbb{S}_{E,\mathrm{tr}}$.

Then we have

$$Q^{(1)} = \mathcal{O}(N^2),\ Q^{(2)} = \mathcal{O}(N(\log N)) \ \textit{and}\ Q^{(3)} = \mathcal{O}(N^2 \log N)\,. \tag{8.95}$$

Proof. The arithmetical cost of solving systems (8.94) is influenced by costs of calculation of the involved matrices, their fill in and the relative condition number according to the inequalities

$$\frac{1}{(1+\log p)^2}\,\boldsymbol{\mathcal{S}}_E \prec \mathbb{S}_E \prec \boldsymbol{\mathcal{S}}_E\,, \tag{8.96}$$

which are a consequence of Theorem 6.5. First we will approve the bounds for the arithmetical costs of the matrices $\mathbb{S}_{E,\mathrm{tr}}$, $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ and with this purpose consider fill in of $\mathbb{S}_{E,\mathrm{tr}}$. Since it is clear that the cost of assembling of these matrices is minor relatively to the cost of calculation of all subdomain matrices $\mathbb{S}_E^{k,l}$, we start from the latter.

To simplify notations, we omit the indices k, l in the notations of the subdomain subspaces and other objects. Let us introduce the subsidiary basis of vectors $\{\boldsymbol{\nu}_{s,t}^{(\alpha)} :\ s = 1,2,\ t = 0,1,\ \alpha = 1,2,\ldots,n^{(s)}-1\}$ in the subspace of $\mathbb{U}$, isomorphic to U_2 and having for components

$$\nu_{1,t}^{(\alpha)}(i,j) = \mu_1^{(\alpha)}(i)\mu_2^{(tn_l)}(j) \quad \text{and} \quad \nu_{2,t}^{(\alpha)}(i,j) = \mu_1^{(tn_k)}(i)\mu_2^{(\alpha)}(j)\,, \tag{8.97}$$

where $i,j = 0,1,\ldots,n(s)$ and

$$\mu_s^{(\alpha)}(i) = \sin\alpha\frac{i\pi}{n(s)}\,, \quad \mu_s^{(0)}(i) = \cos\frac{i\pi}{2n(s)}\,, \quad \mu_s^{(n(s))}(i) = \sin\frac{i\pi}{2n(s)}\,.$$

The vectors with the components $\mu_s^{(\beta)}(i)$ are denoted $\boldsymbol{\mu}_s^{(\beta)}$. Sometimes, without change of the notations, these vectors will be considered as elements of $R^{(n_k+1)(n_l+1)}$ with the components changing only in one of the indices i,j.

We will also need the trigonometric basis in the subspace $U_1 = U_{\mathbb{I}}$, which restriction to a subdomain $\overline{\delta}_{k,l}$ is

$$\mathcal{M}_{tr}^{(\mathbb{I})} = \{\boldsymbol{\mu}^{(\alpha,\beta)} = \boldsymbol{\mu}_1^{(\alpha)} \otimes \boldsymbol{\mu}_2^{(\beta)},\ \alpha = 1,2,\ldots,n_k-1,\ \beta = 1,2,\ldots,n_l-1\}\,,$$

whereas the components of the basis vectors $\boldsymbol{\mu}^{(\alpha,\beta)}$ are

$$\mu^{(\alpha,\beta)}(i,j) = \sin\alpha\frac{i\pi}{n_k}\sin\beta\frac{j\pi}{n_l}\,. \tag{8.98}$$

Clearly, in $\mathcal{M}_{tr}^{(\mathbb{I})}$ and (8.98) a local ordering of the basis vectors, nodes and components of the basis vectors are used.

The following orthogonality properties hold:

$$\begin{gathered}(\boldsymbol{\mu}_1^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_1^{(\gamma)})_{k,l} = 0, \quad (\boldsymbol{\mu}_2^{(\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_2^{(\varkappa)})_{k,l} = 0,\\ (\boldsymbol{\nu}_{1,t}^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_1^{(\gamma)})_{k,l} = 0, \quad (\boldsymbol{\nu}_2^{(\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}_2^{(\varkappa)})_{k,l} = 0,\\ (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}^{(\gamma,\varkappa)})_{k,l} = 0\ , \ \text{for}\ \ \alpha \neq \gamma\ \ \text{or}\ \ \beta \neq \varkappa\ . \end{gathered} \tag{8.99}$$

Taking into account (8.90) and (8.99), we conclude that

$$\boldsymbol{\mu}_{1,t}^{(\alpha)} = \boldsymbol{\nu}_{1,t}^{(\alpha)} + \sum_{\beta=1}^{n_l-1} a_{1,t}{}^{(\alpha,\beta)}\boldsymbol{\mu}^{(\alpha,\beta)}\ , \quad \boldsymbol{\mu}_{2,t}^{(\beta)} = \boldsymbol{\nu}_{2,t}^{(\beta)} + \sum_{\alpha=1}^{n_k-1} a_{2,t}{}^{(\alpha,\beta)}\boldsymbol{\mu}^{(\alpha,\beta)}\ , \tag{8.100}$$

with the coefficients

$$\begin{aligned} a_{1,t}{}^{(\alpha,\beta)} &= (\boldsymbol{\nu}_{1,t}^{(\alpha)}, \mathbf{B}^{k,l}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l} \Big/ (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l}\ ,\\ a_{2,t}{}^{(\alpha,\beta)} &= (\boldsymbol{\nu}_{2,t}^{(\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l} \Big/ (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}^{k,l}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l}\ . \end{aligned} \tag{8.101}$$

Let us estimate the cost of the calculation of the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$. Assembling the matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ from known matrices $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}$ is cheap and costs less than $\mathcal{O}(2N\log N)$ a.o. This is the reason that, at the special ordering of d.o.f., this matrix is tridiagonal and only diagonal elements of this matrix are really assembled from two numbers. The main part is the calculation of the matrices $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}$, which is considered below.

Obviously, it is sufficient to consider $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}$ for $2 \le k,l \le (l_0-1)$ when $\partial\pi_1 \cap \partial\delta_{k,l} = \varnothing$, and each such matrix has the entries related to all four edges. Let $\mathbb{I}_k$ and $\boldsymbol{\Delta}_k$ be the $(n_k+1)\times(n_k+1)$ matrices

$$\mathbb{I}_k = \mathrm{diag}[1/2, 1, \ldots, 1, 1/2] \quad \text{and} \quad \boldsymbol{\Delta}_k = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & 0 & \\ & \ddots & \ddots & \ddots & & \\ & & & -1 & 2 & -1 \\ & 0 & & & -1 & 1 \end{pmatrix},$$

respectively. Vectors (8.98) and (8.97) can be represented as $\boldsymbol{\mu}^{(\alpha,\beta)} = \boldsymbol{\mu}_1^{(\alpha)} \otimes \boldsymbol{\mu}_2^{(\beta)}$, $\nu_{1,t}^{(\alpha)} = \boldsymbol{\mu}_1^{(\alpha)} \otimes \boldsymbol{\mu}_2^{(tn_l)}$, $\boldsymbol{\nu}_{2,t}^{(\beta)} = \boldsymbol{\mu}_1^{(tn_k)} \otimes \boldsymbol{\mu}_2^{(\beta)}$, and, consequently,

$$\mathbf{B}^{k,l} = 2\hbar^{-2}(\psi_k\mathbb{I}_k \otimes \boldsymbol{\Delta}_l + \psi_l\boldsymbol{\Delta}_k \otimes \mathbb{I}_l) + b_{k,1}\mathbb{I}_k \otimes \mathbb{I}_l\ .$$

Since the relationships used for calculations of all diagonal elements $b_{k,l}(s,\alpha)$ of matrix $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}^{k,l}$ are similar, we will present them for $b_{k,l}(1,\alpha)$,

which are related to the edge $T_{1,1}$. We take into account (8.100), (8.101), and that elements of the basis $\{\boldsymbol{\mu}_{s,t}^{(\alpha)}\}$ are orthogonal to the internal basis vectors $\boldsymbol{\mu}^{(\alpha,\beta)}$ in the scalar product, induced by the matrix $\mathbf{B}^{k,l}$. As a consequence, we have

$$\begin{aligned} b(1,\alpha) &= (\boldsymbol{\mu}_{1,1}^{(\alpha)}, \mathbf{B}\boldsymbol{\mu}_{1,1}^{(\alpha)})_{k,l} \\ &= (\boldsymbol{\nu}_{1,1}^{(\alpha)}, \mathbf{B}\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l} + \sum_{\beta=1}^{n_l-1} {a_{1,1}}^{(\alpha,\beta)} (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l} \qquad (8.102) \\ &= (\boldsymbol{\nu}_{1,1}^{(\alpha)}, \mathbf{B}\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l} + \sum_{\beta=1}^{n_l-1} ((\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l}^2 / (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l}, \end{aligned}$$

where indices k, l are omitted. The scalar products involved in (8.102) are known numbers, for some of which we prefer to introduce simple special notations:

$$\begin{aligned} &(\boldsymbol{\mu}_1^{(\alpha)})^\top \boldsymbol{\mu}_1^{(\alpha)} = n_k/2 \,, \\ &(\boldsymbol{\mu}_2^{(n_l)})^\top \mathbb{I}_l \, \boldsymbol{\mu}_2^{(n_l)} = n_l/2 \,, \qquad (\boldsymbol{\mu}_2^{(\beta)})^\top \mathbb{I}_l \, \boldsymbol{\mu}_2^{(n_l)} = \gamma_2(\beta) \,, \\ &(\boldsymbol{\mu}_1^{(\alpha)})^\top \boldsymbol{\Delta}_k \, \boldsymbol{\mu}_1^{(\alpha)} =: \varkappa_1(\alpha) \,, \quad (\boldsymbol{\mu}_2^{(\beta)})^\top \boldsymbol{\Delta}_l \, \boldsymbol{\mu}_2^{(\beta)} =: \varkappa_2(\beta) \,, \\ &(\boldsymbol{\mu}_2^{(\beta)})^\top \boldsymbol{\Delta}_l \, \boldsymbol{\mu}_2^{(n_l)} =: \sigma_2(\beta) \,, \quad (\boldsymbol{\mu}_2^{(n_l)})^\top \boldsymbol{\Delta}_l \, \boldsymbol{\mu}_2^{(n_l)} =: \sigma_2(n_l) \,. \end{aligned} \qquad (8.103)$$

Therefore, we have

$$(\boldsymbol{\nu}_{1,1}^{(\alpha)}, \mathbf{B}\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l} = \hbar^{-2} \left(\psi_k n_k \sigma_2(n_l) + \psi_l \sigma_1(\alpha) n_l\right) + \frac{1}{4} b_{k,l} n_k n_l \,. \qquad (8.104)$$

From (8.103) and (8.104), we conclude that, in order to calculate scalar products (8.104) for all $\alpha = 1, 2, \ldots, n_k - 1$, we have to make $c(n_k - 1)$ a.o., with $c = \text{const}$.

Also, we need the values $\mathcal{A}_{\alpha,\beta} := (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}\,\boldsymbol{\nu}_{1,1}^{(\alpha)})_{k,l}$ and $\mathcal{C}_{\alpha,\beta} := (\boldsymbol{\mu}^{(\alpha,\beta)}, \mathbf{B}\boldsymbol{\mu}^{(\alpha,\beta)})_{k,l}$ given by

$$\begin{aligned} \mathcal{A}_{\alpha,\beta} &= \hbar^{-2} \left(\psi_k n_k \sigma_2(\beta) + \psi_l \varkappa_1(\alpha) n_l\right) + \tfrac{1}{2} b_{k,l} n_k \gamma_2(\beta), \\ \mathcal{C}_{\alpha,\beta} &= \hbar^{-2} \left(\psi_k n_k \varkappa_2(\beta) + \psi_l \varkappa_1(\alpha) n_l\right) + \tfrac{1}{4} b_{k,l} n_k n_l. \end{aligned} \qquad (8.105)$$

For calculation of $b_{k,l}(2,\beta)$, $c_{k,l}(1,\alpha)$, $c_{k,l}(2,\beta)$, we come to quite similar expressions. In the matrix $\mathbb{S}_{E,\mathrm{tr}}^{k,l}$, the blocks coupling adjacent edges are also distinct from zero. For the entry of this block we use the notation $d_{k,l}(s;t,m;\beta,\gamma)$, where s, t, m stand for a couple of adjacent edges, characterized by the numbers s, t and $s := 3 - s, t := m$, according to the ordering of the edges introduced at the beginning of this subsubsection. According to [Rytov (2006)], we have, *e.g.*, the following resulting expressions:

$$
\begin{aligned}
b_{k,l}(1,\alpha) &= \frac{n_k}{2}\left[4\frac{\psi_k}{h^2}\sin^2\tfrac{\pi}{4n_l} + \lambda^{(\alpha)}\tfrac{\psi_l}{h^2} + \frac{b_{k,l}}{2}\right] + \\
&+ \frac{2\psi_k n_k}{h^2 n_l}\sum_{\beta=1}^{n_l-1}\sin^2\frac{\beta\pi}{n_l}\left[\frac{1}{2(\cos\frac{\pi}{2n_l}-\cos\frac{\beta\pi}{n_l})} - \frac{2\psi_k}{\lambda^{\alpha,\beta}h^2}\right], \\
c_{k,l}(1,\alpha) &= -\frac{\psi_k h_k}{h^2}\left[\sin\frac{\pi}{n_l} + \right. \\
&\left. + \frac{2}{n_l}\sum_{j=1}^{n_l-1}(-1)^j\sin^2\frac{j\pi}{n_l}\left[\tfrac{1}{2c_j(l)} - \frac{2\psi_k}{\lambda^{\alpha,j}h^2}\right]\right], \\
d_{k,l}(1;0,1;\alpha,\beta) &= \frac{4\psi_k\psi_l}{\lambda^{\alpha,\beta}h^4}\sin\frac{\alpha\pi}{n_k}\sin\frac{\beta\pi}{n_l},
\end{aligned}
\tag{8.106}
$$

with $c_j(l) = \cos(\pi/2n_l) - \cos(j\pi/n_l)$. Appropriate replacements of indices with the help of the symmetry properties allow us to write down the remaining entries of the edge block $\mathbb{S}^{k,l}_{E,\mathrm{tr}}$ of the Schur complement.

From (8.102)–(8.106), it follows that arithmetical costs of the matrices $\mathbb{S}^{k,l}_{E,\mathrm{tr}}$ and $\mathbb{S}_{E,\mathrm{tr}}$ are $\mathcal{O}(n_k n_l)$ and $\mathcal{O}(N^2)$, respectively. Obviously, the arithmetical costs of preconditioners $\boldsymbol{\mathcal{S}}^{k,l}_{E,\mathrm{tr}}$ and $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$ are also $\mathcal{O}(n_k n_l)$ and $\mathcal{O}(N^2)$. Let us note, that if FDFT is used for solving local Dirichlet problems on subdomains $\delta_{k,l}$, then numbers $\mathcal{C}_{\alpha,\beta}$ are known before calculation of the matrix $\boldsymbol{\mathcal{S}}^{k,l}_{E,\mathrm{tr}}$.

The estimate of the cost for solving system (8.48) by PCGM is the consequence of Theorem 6.5, reflected in the inequalities (8.96), the estimate of the cost for solving the system with the matrix $\boldsymbol{\mathcal{S}}$, and the fact that in the matrices $\mathbb{S}^{k,l}$ additional to $\boldsymbol{\mathcal{S}}^{k,l}_2$ nonzero blocks are completely filled in. The last obstacle makes the matrix-vector multiplications by $\mathbb{S}$ the most expensive operation of the iterative procedure. □

8.1.3.2 *BPX-type Preconditioner-Solver of Griebel and Oswald*

Here $a_{k,l}(u,v)$ is defined as the restriction of the bilinear form $\alpha_{\pi_1}(\ ,\ )$ in (8.44) to the subdomain $\delta_{k,l}$, and namely

$$
\begin{aligned}
a_{k,l}(u,v) = 2\int\limits_{\delta_{k,l}}\left(\psi_l\frac{\partial u}{\partial\xi_1}\frac{\partial v}{\partial\xi_1} + \psi_k\frac{\partial u}{\partial\xi_2}\frac{\partial v}{\partial\xi_2}\right)d\xi \\
+ b_{k,l}\hbar^2\sum_{i,j}\sigma_i\varkappa_j u(\eta_i,\eta_j)v(\eta_i,\eta_j),
\end{aligned}
$$

with

$$\sigma_i = \varkappa_j = 1\,, \quad \text{for} \quad \gamma(k-1) < i < \gamma(k)\,,\ \gamma(l-1) < j < \gamma(l)\,,$$
$$\sigma_i = \varkappa_j = \tfrac{1}{2}\,, \quad \text{for} \quad i = \gamma(k-1), \gamma(k)\,, \quad j = \gamma(l-1), \gamma(l)\,,$$

and the sum over $\gamma(k-1) \le i \le \gamma(k)\,,\ \gamma(l-1) \le j \le \gamma(l)$. Therefore, ψ_k is the constant equal to the value of ψ on the interval (ζ_{k-1}, ζ_k) and $b_{k,l} = b(\xi)$ for $\xi \in \delta_{k,l}$, as it is defined in (8.42).

We will consider the optimal preconditioner-solver for the system

$$\mathbf{B}_1^{k,l}\,\mathbf{v} = \mathbf{f} \tag{8.107}$$

with the $(n_k - 1)(n_l - 1) \times (n_k - 1)(n_l - 1)$ matrix $\mathbf{B}_1^{k,l}$, induced by to the identity

$$\hbar^{-2} a_{k,l}(u, v) = \mathbf{u}^T \mathbf{B}_1^{k,l} \mathbf{v}\,, \quad \forall\ \mathbf{u}, \mathbf{v} \leftrightarrow u, v \in \mathcal{H}_1(\delta_{k,l})\,.$$

This matrix is slightly different from the matrix of the system (8.55), at that the system (8.107) takes the form

$$-2\psi_l \frac{v_{i-1,j} - 2v_{i,j} + v_{i+1,j}}{\hbar^2} - 2\psi_k \frac{v_{i,j-1} - 2v_{i,j} + v_{i,j+1}}{\hbar^2} + b_{k,l} v_{i,j} = f_{i,j}\,,$$
$$\gamma(k-1) < i < \gamma(k)\,,\ \gamma(l-1) < j < \gamma(l)\,,$$
$$v_{m,n} = 0 \quad \text{for}\ \ m = \gamma(k-1), \gamma(k),\ \ n = \gamma(l-1), \gamma(l)\,. \tag{8.108}$$

There are several options for choosing a fast solver for this system. As an example, we will consider the BPX type solver presented in [Griebel and Oswald (1995b)]. First, we formulate one result of these authors.

For simplicity, we consider the Dirichlet problem in the unit square: find $u \in \mathring{H}^1(\pi_1)$ such that

$$\hat{a}(u, v) \equiv \int_{\pi_1} \left(c_1 \frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_1} + c_2 \frac{\partial v}{\partial x_2} \frac{\partial w}{\partial x_2} + uv \right) dx = (f, v) \tag{8.109}$$

for all $v \in \mathring{H}^1(\pi_1)$. Further, we assume that, for a fixed $d > 1$, the unit square is subdivided by the quasiuniform mesh with the step sizes $h_{1,i} \approx d^{-J_1},\ h_{2,i} \approx d^{-J_2},\ i = 1, 2, \ldots, \hat{J}_1,\ j = 1, 2, \ldots, \hat{J}_2$ and $\hat{J}_k \approx d^{J_k}$. We introduce the sequence of imbedded meshes, characterized by vectors $\mathbf{t} = (t_1, t_2)$ with the steps $\hbar_{1,i} \approx d^{-t_1},\ \hbar_{2,i} \approx d^{-t_2},\ i = 1, 2, \ldots, \hat{t_1},\ j = 1, 2, \ldots, \hat{t_2}$ and $\hat{t_k} \approx d^{t_k}$, for which the sets of internal nodes are denoted $X_{\mathbf{t}}$. These meshes induce the FE spaces $\mathring{\mathcal{V}}_{\mathbf{t}}(\pi_1)$, which are the spaces of continuous piecewise bilinear functions, vanishing on $\partial\pi_1$. In particular, for $\mathbf{t} = \mathbf{J} = (J_1, J_2)$, we have the space $\mathring{\mathcal{V}}(\pi_1) := \mathring{\mathcal{V}}_{\mathbf{J}}(\pi_1)$ on the finest mesh. Let $\phi_{\mathbf{t},i,j}(x) \in \mathring{\mathcal{V}}_{\mathbf{t}}(\pi_1)$ be the coordinate function, which is equal to

unity at the node $P_{\mathbf{t},i,j}$ of the set $X_{\mathbf{t}}$ and zero at the rest of nodes of $X_{\mathbf{t}}$, and $\mathcal{V}_{\mathbf{t},i,j}(\pi_1) = \mathrm{span}\,[\phi_{\mathbf{t},i,j}(x)]$ be the 1d space spanned by this coordinate function. Then

$$\mathring{\mathcal{V}}_{\mathbf{t}}(\pi_1) = \sum_{i,j=1}^{\hat{t}_1-1,\,\hat{t}_2-1} \mathcal{V}_{\mathbf{t},i,j}(\pi_1)\,.$$

Obviously, in the case of the Numann boundary condition on a part of the boundary the spaces, spanned by the coordinate functions for the nodes on on that part of the boundary, should be included into the sum. Let us introduce the notation $\{H; b\}$ for the Hilbert space H with the scalar product defined by some bilinear form b.

Theorem 8.7. *Let*

$$(i_0, j_0) = \begin{cases} (0, \lfloor \log_{d^2}(c_1/c_2)\rfloor)\,, & \text{if } \ c_1 = \max(c_1, c_2) \geq 1\,, \\ (\lfloor \log_{d^2}(c_2/c_1)\rfloor, 0)\,, & \text{if } \ c_2 = \max(c_1, c_2) \geq 1\,, \\ (\lfloor \log_{d^2}(c_1^{-1})\rfloor, \lfloor \log_{d^2}(c_2^{-1}\rfloor)\,, & \text{if } \ \max(c_1, c_2) < 1\,, \end{cases} \tag{8.110}$$

and $\mathbf{t}^m = (t_1^m, t_2^m) := (\min(m + i_0, J_1), \min(m + j_0, J_2))$, $0 \leq m \leq \overline{m} := \max(J_1 - i_0, J_2 - j_0)$. *Then the decompositions*

$$\{\mathring{\mathcal{V}}; \hat{a}(\,,\,)\} = \sum_{m=0}^{\overline{m}} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} \{\mathcal{V}_{\mathbf{t}^m,i,j}\,; d^{2m}(\,,\,)_{\pi_1}\}\,, \tag{8.111}$$

and

$$\{\mathring{\mathcal{V}}; \hat{a}(\,,\,)\} = \sum_{m=0}^{\overline{m}} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} \{\mathcal{V}_{\mathbf{t}^m,i,j}\,; \hat{a}(\,,\,)\} \tag{8.112}$$

are stable in the sense that the equivalences

$$\hat{a}(v, v) \asymp |||v|||_0 \quad \textit{and} \quad \hat{a}(v, v) \asymp |||v|||_1$$

hold uniformly with respect to c_k *and* J_k, $k = 1, 2$, *where*

$$|||v|||_0 := \inf_{v_{m,i,j} \in \mathcal{V}_{\mathbf{t}^m,i,j} : \sum v_{m,i,j} = v} \sum_{m=0}^{\overline{m}} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} d^{2m}\,(v_{m,i,j}, v_{m,i,j})_{\pi_1}$$

and

$$|||v|||_1 := \inf_{v_{m,i,j} \in \mathcal{V}_{\mathbf{t}^m,i,j} : \sum v_{m,i,j} = v} \sum_{m=0}^{\overline{m}} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} \hat{a}(v_{m,i,j}, v_{m,i,j}).$$

Proof. We refer to [Griebel and Oswald (1995b)], in particular, to Theorems 1 and 2 for the proof. □

Returning to the local Dirichlet problems on subdomains $\delta_{k,l}$, due to the symmetry properties, we need again to study only the case $l < k$. It is convenient to consider the matrices $\mathbf{B}_1^{k,l}$, generated by the bilinear forms

$$a_{k,l}(v,w) = \int_{\delta_{k,l}} \left(\psi_l \frac{\partial v}{\partial \xi_1}\frac{\partial w}{\partial \xi_1} + \psi_k \frac{\partial v}{\partial \xi_2}\frac{\partial w}{\partial \xi_2} + \epsilon b v w \right) d\xi\,. \tag{8.113}$$

According to Lemma 8.3, one can insert any multiplier $\epsilon \in [0, c],\ c \approx 1$, in (8.113) without compromising the spectral equivalence. To simplify further the description of the algorithm, we consider subdomains $\delta_{k,l}$ generated by the decomposition mesh (8.68).

In order to adapt to the conditions of Theorem 8.7, we map each nest $\delta_{k,l}$ onto the square π_1 by means of the change of variables $\overline{\xi}_k = \xi_k/a_k$ and rewrite the bilinear form $a_{k,l}$ as

$$a_{k,l}(v,w) = \int_{\pi_1} \left(\widehat{\psi}_l \frac{\partial v}{\partial \overline{\xi}_1}\frac{\partial w}{\partial \overline{\xi}_1} + \widehat{\psi}_k \frac{\partial v}{\partial \overline{\xi}_2}\frac{\partial w}{\partial \overline{\xi}_2} + \epsilon b v w \right) a_k a_l \, d\overline{\xi}\,,$$

where $\widehat{\psi}_k = \psi_k/a_l^2$, $\widehat{\psi}_l = \psi_l/a_k^2$ and $\overline{\xi} = (\overline{\xi}_1, \overline{\xi}_2)$. We can introduce on π_1 the sequence of the embedded meshes of sizes $2^{-t_1} \times 2^{-t_2}$, characterized by vectors $\mathbf{t} = (t_1, t_2)^\top$ and the corresponding FE spaces

$$\mathring{\mathcal{V}}_{\mathbf{t}} = \sum_{i,j=1}^{\hat{t}_1-1,\,\hat{t}_2-1} \mathcal{V}_{\mathbf{t},i,j}\,, \quad \mathcal{V}_{\mathbf{t},i,j} = \operatorname{span}[\phi_{\mathbf{t},i,j}]\,, \quad \hat{t}_k = 2^{t_k}\,,$$

with the 1d subspaces $\mathcal{V}_{\mathbf{t},i,j}$, each spanned by a single nodal coordinate function $\phi_{\mathbf{t},i,j}$. In the case under consideration, space $\mathring{\mathcal{V}}_{\mathbf{t}}$ for $\mathbf{t} = (k-1, l-1)^\top$ is the transformed FE space $\mathcal{H}_1(\delta_{k,l})$.

Let us assume ϵ to be sufficiently small such that, for $l \le k$ and after dividing by ϵb, the bilinear form $a_{k,l}(v,w)$ satisfies the condition in the second row of (8.110). Since $n_k = 2^{k-1}$, $J_k = k-1$, and for $\mathbf{t}^m$ defining the subsequence of spaces $\mathring{\mathcal{V}}_{\mathbf{t}^m}$, we have

$$\mathbf{t}^m = (t_1^m, t_2^m) = (\min(m + i_0, k-1), \min(m, l-1)),$$

where $i_0 = \log_4 \widehat{\psi}_k/\widehat{\psi}_l = \log_4 2^{4(k-l)} = 2(k-l)$ and $0 \le m \le \overline{m} := \max(k-1-i_0, l-1) = l-1$. In turn, the pointed out subsequence of spaces induces the decomposition

$$\{\mathring{\mathcal{V}}^{(k,l)}; a_{k,l}(\,,\,)\} = \sum_{m=0}^{l-1} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} \{\mathcal{V}_{\mathbf{t}^m,i,j}\,; 2^{2m} a_k a_l\,(\,,\,)_{\pi_1}\}. \tag{8.114}$$

Let us denote by $\boldsymbol{\mathcal{D}}_m$ the diagonal matrix with the entries defined by the scalar products with weights in (8.114), by V_m the spaces of nodal values

of the FE functions from $\mathring{\mathcal{V}}_{\mathbf{t}^m}$, and by $\mathbb{T}_m : V_m \to V_l$ the matrix, which for any $v_m \in \mathring{\mathcal{V}}_{\mathbf{t}^m}$ transforms its vector representation $\mathbf{v}_m \in V_m$ into its vector representation $\mathbf{v}_l \in V_l$. Now we are in a position to define the BPX like preconditioner $\mathbb{B}_1^{k,l}$ for $\mathbf{B}_1^{k,l}$ which inverse is

$$(\mathbb{B}_1^{k,l})^{-1} = \hbar^2 \sum_{m=1}^{l-1} \mathbb{T}_m \boldsymbol{\mathcal{D}}_m^{-1} \mathbb{T}_m^\top \,. \tag{8.115}$$

Corollary 8.4. *Uniformly in k, l and p, the decomposition (8.114) is stable in a sense of the equivalences*

$$a_{k,l}(v,v) \asymp |||v|||_0 \tag{8.116}$$

and

$$a_{k,l}(v,v) \asymp |||v|||_1 \tag{8.117}$$

where

$$|||v|||_0 := \inf_{v_{m,i,j} \in \mathcal{V}_{\mathbf{t}^m,i,j} : \sum v_{m,i,j} = v} \sum_{m=0}^{l-1} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} 2^{2m} a_k a_l \, (v_{m,i,j}, v_{m,i,j})_{\pi_1},$$

$$|||v|||_1 := \inf_{v_{m,i,j} \in \mathcal{V}_{\mathbf{t}^m,i,j} : \sum v_{m,i,j} = v} \sum_{m=0}^{l-1} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} a_{k,l} \, (v_{m,i,j}, v_{m,i,j}).$$

Furthermore, the matrices $\mathbf{B}_1^{k,l}$ and $\mathbb{B}_1^{k,l}$ are spectrally equivalent.

Proof. The stability follows from Theorem 8.7, whereas the spectral equivalence of the matrices is a direct consequence of the stability. □

Evidently, at $k = l$ we have usual coarsening simultaneously in two directions, and $\mathbb{B}_1^{k,k}$ is a classical BPX preconditioner. On the other hand, for $2l - 1 \le k$, we have only semi-coarsening, *i.e.,* coarsening in the one direction ξ_2. Indeed, in this case, for the first component of $\mathbf{t}^m = (\min(m + i_0, k-1), \min(m, l-1))$, we have

$$\begin{aligned} \min(m + i_0, k-1) &= \min(m + 2(k-l), k-1) \\ &\ge \min(m + k - 1, k-1) \ge k - 1 \,, \quad \forall\, m \ge 0. \end{aligned}$$

At $k \le 2l - 2$, coarsening is produced $k - l$ times in the one direction ξ_2 with m taking the values $m = l-1, l-2, \ldots, (l-1-(k-l)) = 2l-k-1$. Note that in this case $k - l \le l - 2$ and that for $m = 2l - k - 1$ we have $\mathbf{t}^m = (k-1, 2l-k-1)$. Then coarsening simultaneously in the both directions is done $2l-k-1$ times, involving $m = 2l-k-1, 2l-k-2, \ldots, 1, 0$.

Let us note that the space $\mathcal{V}_{\mathbf{t}^0}$ is empty in the case of the Dirichlet problem on subdomain $\delta_{k,l}$ and sufficiently small ϵ. Therefore, in (8.114) and in the equivalence relations (8.116),(8.117) of Corollary 8.4 summation can be started from $m = 1$. However, at $\epsilon = 1$ constants in the spectral equivalence of matrices $\mathbf{B}_1^{k,l}$ and $\mathbb{B}_1^{k,l}$ can be improved. In this case, $\psi_k, \psi_l < b$ for most of subdomains $\delta_{k,l}$, and the derivation of the BPX type preconditioner $\mathbb{B}_1^{k,l}$ is completed under conditions of the third row in (8.110). As a consequence, one can have $i_0 \neq 0$, $j_0 \neq 0$, and the nonempty subspace $\mathcal{V}_{\mathbf{t}^0}$.

Taking into account that $\mathcal{V}_{\mathbf{t}^0}$ is empty and the matrix, generated by the space $\mathcal{V}_{\mathbf{t}^1}$, is tridiagonal, we can replace (8.114) by the decomposition

$$\{\mathring{\mathcal{V}}^{(k,l)}; a_{k,l}(\,,\,)\} = \{\mathcal{V}_{\mathbf{t}^1}; a_{k,l}(\,,\,)\} + \sum_{m=2}^{l-1} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} \{\mathcal{V}_{\mathbf{t}^m,i,j}\,; 2^{2m} a_k a_l\,(\,,\,)_{\pi_1}\}, \tag{8.118}$$

which is also stable, and, *e.g.,*

$$a_{k,l}(v,v) \asymp |||v|||_1, \tag{8.119}$$

where

$$|||v|||_1 := \inf_{v_1 \in \mathcal{V}_{\mathbf{t}^1},\, v_{m,i,j} \in \mathcal{V}_{\mathbf{t}^m,i,j} :\, v_1 + \sum_2^{l-1} v_{m,i,j} = v} \{a_{k,l}\,(v_1, v_1) + \\ + \sum_{m=2}^{l-1} \sum_{P_{\mathbf{t}^m,i,j} \subset X_{\mathbf{t}^m}} a_{k,l}\,(v_{m,i,j}, v_{m,i,j})\}\,.$$

Efficient ways of the multiplication $(\mathbb{B}_1^{k,l})^{-1}\mathbf{v} =: \mathbf{w}$ with an arbitrary vector $\mathbf{v}$, which is understood as solving the system $\mathbb{B}_1^{k,l}\mathbf{w} = \mathbf{v}$, are well known. Its computational cost is $\mathcal{O}(n_k n_l)$ a.o. for one subdomain $\delta_{k,l}$ and, therefore, $\mathcal{O}(N^2)$ a.o. for all subdomains.

8.1.3.3 *Multilevel Wavelet Solver*

We assume that the decomposition mesh is defined as in (8.68). On $(0,1)$ for each $\ell = 1, 2, \ldots, \ell_0$, we introduce the uniform mesh $\eta_i^{(\ell)}$ of the size $\hbar_\ell = 2^{-\ell}$ and the space $\mathcal{V}_\ell(0,1)$ of the continuous piecewise linear functions, vanishing at the ends of the interval (0,1). The space of $\mathcal{V}_\ell(0,1)$ is a subspace of $\mathcal{V}_{\ell+1}(0,1)$, its dimension equals $N_\ell = 2^\ell - 1$ with $N_{\ell_0} = N$. Let $\phi_i^\ell \in \mathcal{V}_\ell(0,1)$ be the the nodal basis function for the node $\eta_i^{(\ell)}$, *i.e.*, such that $\phi_i^\ell(\eta_j^{(\ell)}) = \delta_{i,j}$ and $\mathcal{V}_\ell(0,1) = \mathrm{span}[\phi_i^\ell]_{i=0}^{N_\ell+1}$. This basis induces

the Gram matrices

$$\mathbf{\Delta}_\ell = \hbar_\ell \left\{ \langle \frac{d\phi_i^\ell}{dx}, \frac{d\phi_i^\ell}{dx} \rangle_{\omega=1} \right\}_{i,j=1}^{N_\ell} \quad \text{and} \quad \mathbf{M}_{\psi,\ell} = \hbar_\ell^{-1} \left\{ \langle \phi_i^\ell, \phi_j^\ell \rangle_{\omega=\sqrt{\psi}} \right\}_{i,j=1}^{N_\ell} ,$$

where

$$\langle u, v \rangle_\omega := \int_0^1 \omega^2 u\, v\, dx\,.$$

Obviously, $\mathbf{\Delta}_{\ell_0} = \mathbf{\Delta}$. We can easily establish that $\mathbf{M}_\psi = \mathbf{M}_{\psi,\ell_0}$ is spectrally equivalent to $\boldsymbol{\mathcal{D}}_e^{-1}$ uniformly in N. As a consequence, the preconditioner $\mathbf{\Lambda}$ is spectrally equivalent to the matrix

$$\mathbf{B}_{\psi,1} := \mathbf{M}_\psi \otimes \mathbf{\Delta} + \mathbf{\Delta} \otimes \mathbf{M}_\psi \tag{8.120}$$

uniformly in p, and we can construct fast solvers for systems with the matrices $\mathbf{A}_{ee}$, $\mathbf{\Lambda}_{ee}$, $\mathbf{\Lambda}$ on the basis of the preconditioner $\mathbf{B}_{\psi,1}$. This matrix can be also assembled from the local matrices $\mathbf{B}_{\psi,1}^{k,l}$, which are counterparts of the matrices $\mathbf{B}_1^{k,l}$.

The matrices $\mathbf{B}_{\psi,1}^{k,l}$ are determined locally. Under the assumption of the decomposition mesh (8.68), we introduce the Dirichlet restriction $\mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k)$ of the space $\mathcal{V}_\ell(0,1)$ to each interval (ζ_{k-1}, ζ_k) with $k \geq \ell_0 - \ell + 2$ and adopt that $\mathring{\mathcal{V}}(\zeta_{k-1}, \zeta_k) = \mathring{\mathcal{V}}_{\ell_0}(\zeta_{k-1}, \zeta_k)$. *I.e.,* the subspace $\mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k)$ contains functions from the space $\mathcal{V}_\ell(0,1)$, vanishing outside the interval (ζ_{k-1}, ζ_k), and, therefore, has the dimension $n_{k,\ell} = 2^{k-(\ell_0-\ell)-1} - 1$ with $n_{k,\ell_0} = 2^{k-1}$. By $\mathbf{\Delta}_k^{(\ell)}$ and $\mathbf{M}_k^{(\ell)}$, we denote the matrices of the bilinear forms

$$(u', v')_{(k)} \quad \text{and} \quad (u, v)_{(k)}, \quad u, v \in \mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k)\,, \tag{8.121}$$

respectively, where

$$(u, v)_{(k)} := \int_{\zeta_{k-1}}^{\zeta_k} u\, v\, dx\,.$$

By setting $\mathbf{\Delta}_k = \mathbf{\Delta}_k^{(\ell_0)}$ and $\mathbf{M}_k = \mathbf{M}_k^{(\ell_0)}$, from (8.120), we get the representation

$$\mathbf{B}_{\psi,1}^{k,l} = \psi_k \mathbf{M}_k \otimes \mathbf{\Delta}_l + \psi_l \mathbf{\Delta}_k \otimes \mathbf{M}_l\,. \tag{8.122}$$

Each subspace $\mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k)$ may be represented by the direct sum

$$\mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k) = \mathring{\mathcal{V}}_{\ell-1}(\zeta_{k-1}, \zeta_k) \oplus \mathring{\mathcal{W}}_\ell(\zeta_{k-1}, \zeta_k)\,,$$

and, therefore,

$$\mathring{\mathcal{V}}_\ell(\zeta_{k-1}, \zeta_k) = \mathring{\mathcal{W}}_\ell(\zeta_{k-1}, \zeta_k) \oplus \ldots \oplus \mathring{\mathcal{W}}_{\ell_0-k+2}\,,$$

where $\mathring{\mathcal{W}}_{\ell_0-k+2}(\zeta_{k-1}, \zeta_k) := \mathring{\mathcal{V}}_{\ell_0-k+2}(\zeta_{k-1}, \zeta_k)$. In relation with these space decompositions, we introduce the notations:

- $\mathbb{V}^k = \left\{\phi_i : \phi_i \in \mathring{\mathcal{V}}(\zeta_{k-1}, \zeta_k)\right\}$,
- $\mathbb{W}^k_{\rm mult} = \left\{\chi^\ell_i : \chi^\ell_i \in \mathring{\mathcal{W}}_\ell(\zeta_{k-1}, \zeta_k),\ \ell_0 - k + 2 \le \ell \le \ell_0\right\}$ is the multiscale wavelet basis in the subspace $\mathring{\mathcal{V}}(\zeta_{k-1}, \zeta_k)$,
- $\boldsymbol{\Delta}_{k,\rm mult}$ and $\boldsymbol{\mathcal{M}}_{k,\rm mult}$ are Gramm's matrices of the bilinear forms (8.121) in the basis $\mathbb{W}^k_{\rm mult}$,
- $\mathbb{D}_{1,k}$ and $\mathbb{D}_{0,k}$ are the diagonal matrices having common main diagonals with $\boldsymbol{\Delta}_{k,\rm mult}$ and $\boldsymbol{\mathcal{M}}_{k,\rm mult}$, respectively.

By $\mathbf{Q}_k$, we denote the transformation matrix from the multiscale basis $\mathbb{W}^k_{\rm mult}$ to the finest FE basis $\mathbb{V}^k$, *i.e.,* if $\mathbf{v}$ and $\mathbf{v}_{\rm mult}$ are the vectors of coefficients representing a function from $\mathring{\mathcal{V}}(\zeta_{k-1}, \zeta_k)$ in the bases $\mathbb{V}^k$ and $\mathbb{W}^k_{\rm mult}$, respectively, then $\mathbf{v}_{\rm mult} = \mathbf{Q}_k\,\mathbf{v}$.

Theorem 8.8. *For each N, there exists such a multiscale wavelet basis $\mathbb{W}^k_{\rm mult}$ that the matrices $\boldsymbol{\Delta}_{k,\rm mult}$ and $\boldsymbol{\mathcal{M}}_{k,\rm mult}$ are simultaneously spectrally equivalent to the matrices $\mathbb{D}_{1,k}$ and $\mathbb{D}_{0,k}$, respectively, and*

$$\mathrm{ops}[\mathbf{Q}_k\mathbf{v}], \mathrm{ops}[(\mathbf{Q}_k)^\top\mathbf{v}], \mathrm{ops}[\mathbf{Q}_k^{-1}\mathbf{v}] \prec n_k\,, \quad \forall\,\mathbf{v} \in R^{n_k-1}\,,$$

uniformly in k and p.

Proof. Existence of the multiscale wavelet bases satisfying the conditions sufficient for the theorem to be valid and construction of such bases were the subject of a number of papers. We refer to the papers of [Griebel and Oswald (1995b)], [Dahmen and Schneider (1999)], [Cohen (2000)] and [Beuchler *et al.* (2004)], where multiscale wavelet preconditioners for more general matrices can be also found, see in this relation Subsections 9.3.1 and 9.3.2. □

Corollary 8.5. *The matrix $\mathbf{B}^{k,l}_{\rm mult,1}$ defined by its inverse*

$$(\mathbf{B}^{k,l}_{\rm mult,1})^{-1} := (\mathbf{Q}_k^\top \otimes \mathbf{Q}_l^\top)\left[\psi_k^{-1}\mathbb{D}_{0,k} \otimes \mathbb{D}_{1,l} + \psi_l^{-1}\mathbb{D}_{1,k} \otimes \mathbb{D}_{0,l}\right]^{-1}(\mathbf{Q}_k \otimes \mathbf{Q}_l),$$

is spectrally equivalent to the matrices $\mathbf{B}^{k,l}_1$ and $\mathbf{B}^{(k,l)}_{\psi,1}$, and the bound

$$\mathrm{ops}[(\mathbf{B}^{k,l}_{\rm mult,1})^{-1}\mathbf{v}] \prec n_k\,n_l\,, \quad \forall\mathbf{v} \in R^{(n_k-1)(n_l-1)}\,,$$

holds uniformly in k, l and p.

There is quite a number of other iterative algorithms for solving systems (8.55) optimal in the computational work and robust in respect to the aspect ratios ψ_k/ψ_l. Among them, we can find multilevel preconditioners-solvers of BPX, MDS and other types, based on stable FE and wavelet

decompositions of the tensor product FE spaces, some multigrid and Schur complement algorithms. In this relation we refer to [Schieweck (1986)], [Pflaum (2000)], [Khoromskij and Wittum (2004)], where additional references can be found. It is worth to mention also the approaches from other directions like $\mathcal{H}$-matrices and tensor-train decompositions, see, *e.g.*, [Hackbusch *et al.* (2005)], [Bebendorf (2008)], [Hackbusch (2009)], [Dolgov *et al.* (2011)], [Khoromskij (2011)] and [Hackbusch (2012)], which are only a few representatives of this vast area of research.

8.1.4 *Numerical Experiment with DD Solver for Local Dirichlet Problems, Some Generalizations*

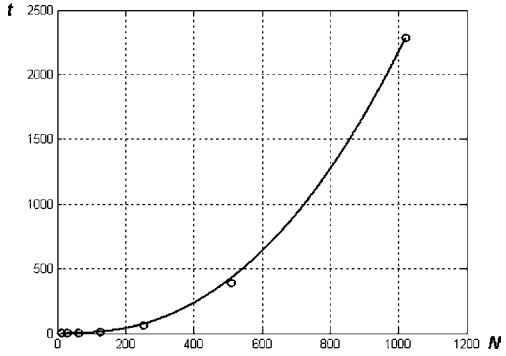

Fig. 8.3 Computational cost of the $DD_{\rm loc}$ solver for local Dirichlet problems.

System (8.47) with the system matrix $\mathbf{B}^{(\hbar)}$ and the right-hand side vector with constant entries were solved numerically by means of the PCGM with the preconditioner $DD_{\rm loc}$ in [Anufriev *et al.* (2003)], see also [Anufriev and Korneev (2005)]. $DD_{\rm loc}$ type preconditioner-solver, implemented in these experiments, had basically the same structure and major components as $\boldsymbol{\mathcal{B}}^{(\hbar)}$ of (8.52), some of which were slightly less efficient. In particular, the FDFT was used for the exact solution of the Dirichlet problems on subdomains of the decomposition and for the prolongations from their interface. The a priory estimate of computational work for that preconditioner is $\mathcal{O}(N^2(\ln N)^{3.5})$, $N = \lfloor p/2 \rfloor_+$. The dependence of the computational time upon N in seconds is shown in Figure 8.3, where circular markers correspond to a priory predicted values of the computational cost. As is seen from Figure 8.3, numerical and theoretical results are in a good agreement.

8.2 Numerical Testing of DD Solver for Dirichlet Problem in a *L*-Shaped Domain

The DD preconditioner-solver for the Dirichlet problem for the Poisson equation

$$-\Delta u = f \ \text{ in } \Omega \quad \text{and} \quad u = 0 \ \text{ on } \partial\Omega, \tag{8.123}$$

where the domain Ω is L-shaped, was numerically tested. Figure 8.4 depicts the domain and the finite element grid, which is the decomposition grid at the same time. The system of linear algebraic equations was generated by means of the p discretization with the hierarchical reference element $\mathcal{E}_H$ described in Subsection 7.3.1, *i.e.,* having polynomials of the set $\mathcal{M}_p$ as coordinate functions.

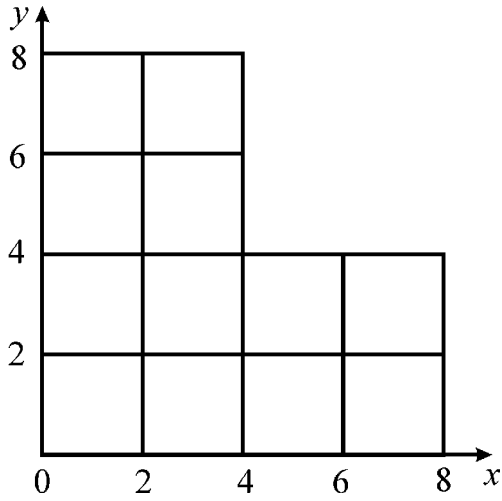

Fig. 8.4 L-shaped domain and domain decomposition mesh.

The DD preconditioner of the form (7.27) with vertex d.o.f. split from the rest of d.o.f. was implemented in the computer code. Blocks on the diagonal of the preconditioner-solver $\boldsymbol{\mathcal{K}}_I = \text{diag}\,[\boldsymbol{\mathcal{K}}_{I_r}]_{r=1}^{\mathcal{R}}$ for the internal unknowns were defined as DD_{loc} preconditioner-solvers, the same, which were used in the experiments described in Subsection 8.1.4. For the Schur complement preconditioner appearing in (7.27), the one described in Subsection 7.5.3, see Lemma 7.1, was used. Prolongation operations were performed according to Subsection 7.2.2, *i.e.,* by means of the inexact solution of local Dirichlet problems by PCGM with the alternative to use one of the

two preconditioners. One is the same preconditioner-solver $DD_{\rm loc}$, which is pointed out above, and another is the multigrid preconditioner **Mg**, described in Subsection 8.1.1.

Therefore, the implemented DD algorithm is a three-stage iteration process with the PCGM as for the leading iteration process, which solves the global system of equations with $\boldsymbol{\mathcal{K}}$ for the preconditioner, so for the second-stage iterations with preconditioners $\boldsymbol{\mathcal{K}}_{I,0}$ in the procedures of prolongations inside finite elements from their boundaries. In both PCGM processes iteration parameters β_k were calculated according to (2.40). Experiments were executed on PC PentiumIII-733Mhz with 768Mb memory.

The number n_g of leading PCGM iterations, needed to reduce the Euclidean norm of the residual in 10^4 times, is plotted in Figure 8.5, whereas computational time t in seconds is plotted in Figure 8.6. The growth of the iteration number and the computer cost are in a good agreement with the predictions by a priori estimates.

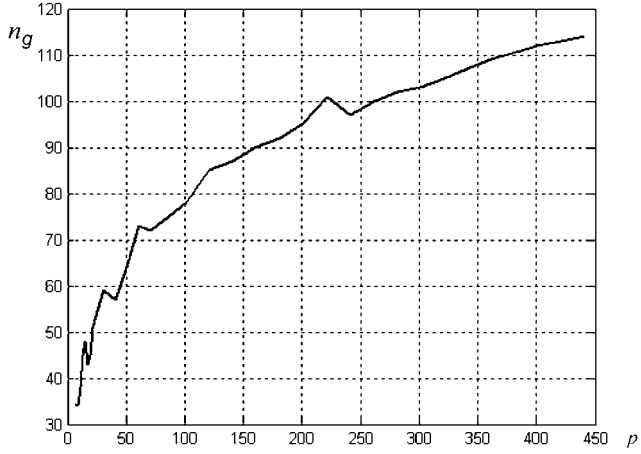

Fig. 8.5 Number of PCGM iterations against polynomial degree p.

The numbers of secondary iterations were also optimized experimentally in a sense of the fastest achievement of the prescribed accuracy of the numerical solution. As it was noted above, prolongations in (7.27) from FE boundaries inside finite elements and restrictions were carried out by means of the secondary inexact iterative process PCGM with iterative parameters β_k, defined by (2.40). In a form of a linear two-level iteration method such prolongations are defined in Lemma 7.4 of Subsection 7.2.2. In these

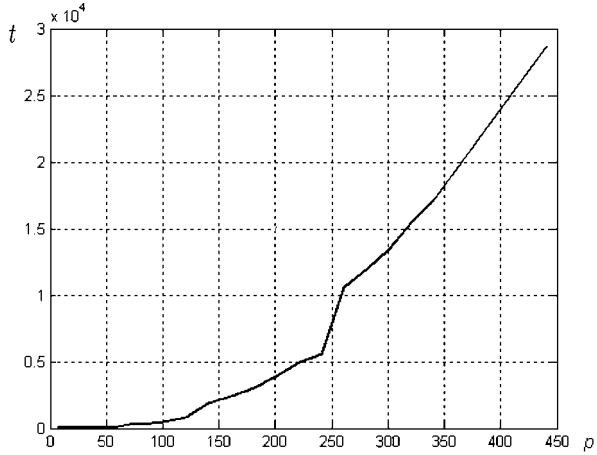

Fig. 8.6 CPU time in seconds against polynomial degree p.

experiments, the master prolongation matrix (7.74) was

$$\boldsymbol{\mathcal{P}}_B = \begin{pmatrix} -\mathbf{A}_{I,\mathrm{it}}^{-1}\left(\mathbf{A}_{I,B} + \mathbf{A}_I \mathbb{T}_{I,B}\right) + \mathbb{T}_{I,B} \\ \mathbf{I} \end{pmatrix}, \tag{8.124}$$

where

$$\mathbf{A}_{I,\mathrm{it}}^{-1} = \left[\mathbf{I} - \prod_{k=1}^{\nu} \left(\mathbf{I} - \sigma_k \boldsymbol{\Lambda}_{\triangle,\mathrm{it}}^{-1} \mathbf{A}_I\right)\right] \mathbf{A}_I^{-1}, \tag{8.125}$$

$\mathbf{A}_I, \mathbf{A}_{I,B}$ are the blocks of the reference element stiffness matrix $\mathbf{A}$, and $\boldsymbol{\Lambda}_{\triangle,\mathrm{it}}^{-1}$ was produced by the inexact multigrid solver for the preconditioner $\boldsymbol{\Lambda}_{\triangle}$ of (7.102). In other words, the preconditioner $\boldsymbol{\Lambda}_{\triangle,\mathrm{it}} = \mathbf{Mg}$, see Corollary 8.2, was implemented. Moreover, the number of the secondary iterations n_{pr} was varied. This enabled experimental finding the optimal (in a sense of the arithmetical cost of the whole DD solver) number n_{pr} of secondary iterations in dependence on p. The corresponding graph is given in Fig. 8.7, it is in a complete agreement with the theoretical estimate $n_{\mathrm{pr}}(p) = \mathcal{O}((\ln(1+p))/\rho)$. If the number of the secondary iterations is increased beyond optimal, then the number n_g of global iterations becomes lesser, whereas their total computational cost increases. Figure 8.8 shows the reduction of the Euclidean norm of the relative residual

$$\gamma_k = \|\mathbf{K}\mathbf{u}^{(k)} - \mathbf{f}\|/\|\mathbf{f}\|$$

for solving the FE global system of algebraic equations for $p = 127$ in dependence on the number $k = n_g$ of PCGM iterations of the global DD

solver and values of n_{pr}. Similar results were obtained for a number of other values of p. We refer to [Anufriev *et al.* (2003)] and [Anufriev and Korneev (2005)] for additional information about the described numerical experiments.

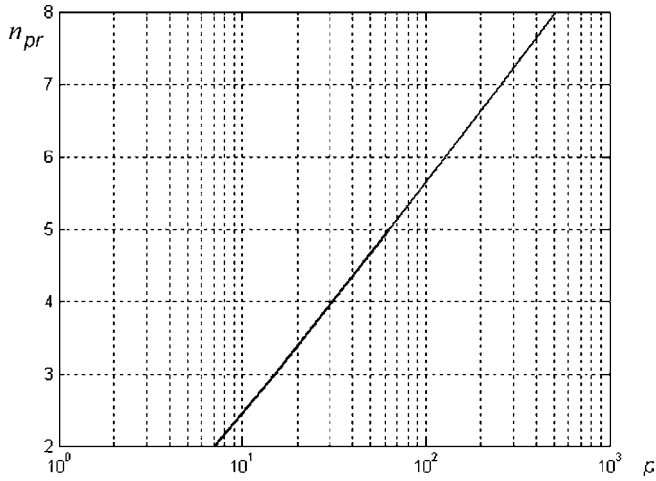

Fig. 8.7 Optimized numbers of secondary iterations against polynomial degree p.

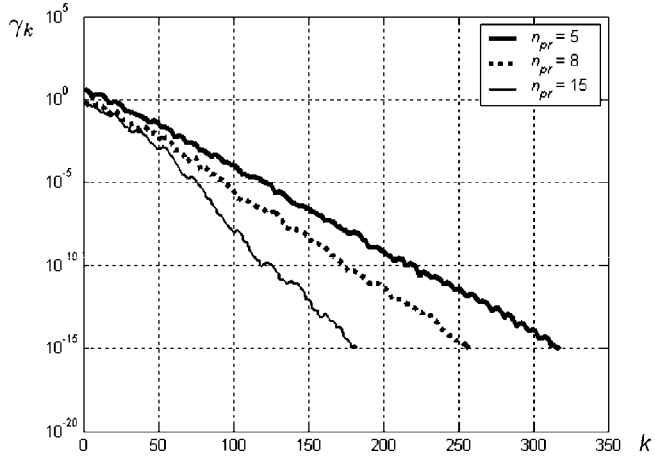

Fig. 8.8 Relative residuals against number of iterations of the global solver.

8.3 Fast Dirichlet Solvers for 2d Spectral Reference Elements

Although the appearances as well as some important properties of the finite-difference preconditioners for the hierarchical and spectral reference elements seem quite different, the same ideas can be used at the construction of fast solvers in both cases. This judgement is supported by the fact that, as was shown earlier, a diagonal transformation of the internal matrix of the spectral reference element produces a matrix with properties which are quite similar to those of the hierarchical reference element internal stiffness matrix. The consequence of the latter fact is that any solver designed for the hierarchical elements can indeed be adapted to the spectral elements. The example of solvers derived with the use of the aforementioned transformation of variables is the multilevel solver presented in Subsection 8.3.2. The properties of the preconditioners for the spectral reference elements stiffness matrices, which are finite-difference preconditioners in a strict sense, can also be exploited in a more direct way. The fast $DD_{\rm loc}$ solver presented in the following subsection illustrates this approach.

8.3.1 *Fast Domain Decomposition Preconditioner-Solver*

We turn to the finite-difference preconditioners which are spectrally equivalent to the spectral reference elements stiffness matrices and are described in Subsection 7.4.2. First of all, we will construct the $DD_{\rm loc}$ preconditioner-solver similar to that one considered in Subsection 8.1.2 for the hierarchical reference elements. This solver will be derived in a more direct way, *i.e.,* without resort to the factorized preconditioners $\mathfrak{K}_\diamond$ and $\mathfrak{K}_\dagger$ from Corollary 7.4. The difference between the preconditioner-solvers $DD_{\rm loc}$ of Subsection 7.4.2 for the hierarchical reference elements and those, which we are going to derive, is, in particular, in the definition of the decomposition mesh. In the former case the decomposition mesh, imbedded in the uniform square mesh, was defined in such a way that, in each subdomain, variable coefficients of elliptic equation (7.109) could be replaced by the constant ones, which preserve the equivalence of the energy integral. Now, we will take care that, in each subdomain, variable mesh steps in the FD preconditioner (7.140) could be replaced by special constant ones. These piecewise constant mesh steps are chosen such that the ratios of the old steps to the new ones are bounded from below and above by positive constants uniformly over all subdomains. This implies that in both cases,

i.e., at the replacements by the piecewise constant coefficients and by the piecewise constant mesh steps, the important property is preserved: the new modified prconditioners are spectrally equivalent to the source FE/FD prconditioners.

The most important part of the $DD_{\rm loc}$ preconditioner-solver is the Shcur complement preconditioner-solver. Except for some formal differences, the path of its construction in Subsubsection 8.3.1.3 as well as related results resemble those for the hierarchical reference element.

8.3.1.1 *Domain Decomposition*

Without loss of generality, we assume $p = 2N$. Since any of the preconditioners $\boldsymbol{\mathcal{A}}_\circ = \boldsymbol{\mathcal{A}}_{\rm sp}, \boldsymbol{\mathcal{A}}_{\rm p/s}$ is spectrally equivalent to the spectral reference element with GLL nodes, any of them can be adapted to the $DD_{\rm loc}$ solver. Accordingly, by η_i, we can understand the coordinates of the GGL, GLC, or the peudospectral meshes. The tensor product nonuniform mesh, defined by means of one of those 1d meshes will be termed the *fine (spectral) mesh* in the contrast to the imbedded *decomposition or coarse mesh*, which is described below.

The coordinates of the mesh lines of the decomposition mesh are denoted by ζ_k, $k = 0, 1, \ldots, 2\ell_0$, with $\zeta_0 = -1$, $\zeta_{\ell_0} = 0$, $\zeta_{2\ell_0} = 1$, and we also introduce the notations $\overline{h}_k = a_k = \zeta_k - \zeta_{k-1}$ and $n_k = \widetilde{i}(k) - \widetilde{i}(k-1)$. Here the discrete function $\widetilde{i}(k)$ is such that, for $i = \widetilde{i}(k)$, we have $\eta_i = \zeta_k$. For the interval $(-1, 0)$ they may be determined in accordance with the conditions

$$\frac{\hbar_{\widetilde{i}(k)}}{\hbar_{\widetilde{i}(k-1)+1}} = \max_{\widetilde{i}(k-1)<i\le\widetilde{i}(k)} \frac{\hbar_i}{\hbar_{\widetilde{i}(k-1)+1}} \le c_0 = \text{const}, \ 1 \le k \le \ell_0\,, \quad (8.126)$$

with some generic constant c_0. The construction of decomposition mesh can be done in the following way. We choose some $n_1 \in \mathbb{N}^+$, $n_1 > 1$, and then, by induction, for every known $\widetilde{i}(k-1)$, the number $\widetilde{i}(k)$ can uniquely be found by conditions (8.126) for $c_0 \ge n_1$. Since $|\zeta_{\ell_0-1}|$, found in this way, can be small, we can redefine ζ_{ℓ_0-1} by setting $\zeta_{\ell_0-1} = \eta_j$, where $\eta_j < 1$ and is the fine mesh node closest to $\zeta_{\ell_0-2}/2$. The coordinates of the decomposition mesh found for the interval $(-1, 0)$ are expanded by symmetry on $(-1, 1)$.

Fine meshes, used for generation of the preconditioners $\boldsymbol{\mathcal{A}}_\circ = \boldsymbol{\mathcal{A}}_{\rm sp}, \boldsymbol{\mathcal{A}}_{\rm p/s}$, are geometrically equivalent. Therefore, it is sufficient to establish the existence of the decomposition mesh and the bound $\ell_0 \le c \log N$, $c =$ const, only for one of these meshes. For this purpose, we turn to the more

general mesh (7.137) for $\gamma > 0$, assuming for simplicity that $\hbar_{i+1} > \hbar_i$ for $i = 1, 2, \dots \ell_0$.

Lemma 8.6. *Let conditions (7.137) be fulfilled and $n_1 > 1$. Then, for any $c_0 \geq n_1^\gamma c_2/c_1$, there exist $\zeta_k = \eta_{\widetilde{i}(k)}$ for $k = 1, 2, \dots, \ell_0 - 1$ satisfying (8.126) with*

$$\ell_0 \leq \mathrm{int}\, \lfloor \frac{\gamma(\log N - \log n_1)}{\log(c_0 c_2/c_1)} \rfloor_+ .$$

Proof. If $n_1 > 1$ is chosen, then

$$\max_{0<i,j\leq n_1} \hbar_i/\hbar_j \leq n_1^\gamma c_2/c_1 ,$$

and, at $c_0 = n_1^\gamma c_2/c_1$, the sequence of coordinates ζ_k is defined with growing n_k. From (7.137) and (8.126), it follows that

$$\frac{\widetilde{i}(k)}{\widetilde{i}(k-1)+1} \leq (\frac{c_0 c_2}{c_1})^{1/\gamma},$$

and, obviously, $\widetilde{i}(k-1)+1 \leq 3\widetilde{i}(k-1)/2$ and $\widetilde{i}(k) = \widetilde{i}(k-1)+n_k$. Therefore,

$$n_k \leq [\frac{3}{2}(\frac{c_0 c_2}{c_1})^{1/\gamma} - 1]\,\widetilde{i}(k-1) .$$

Introducing the notation $c = 1.5(c_0 c_2/c_1)^{1/\gamma} - 1$, we see that

$$\widetilde{i}(0) = 0\,,\ \widetilde{i}(1) = n_1\,,\ \widetilde{i}(2) = n_1(1+c)\,,\ \dots\ ,\ \widetilde{i}(k) = n_1(1+c)^{k-1}\,,$$

and, therefore, $\ell_0 - 1 \leq \mathrm{int}\, \lfloor a \rfloor_+$, where a is found from the equality $n_1(1+c)^a = N$. Hence, the estimate for ℓ_0 directly follows. □

In view of the relationship $\hbar_k \asymp k/p^2$, $k \leq N$, which holds for the mesh (7.137) and GLL or GLC mesh, we come to the following conclusion.

Corollary 8.6. *Let $n_1 > 1$ be fixed and $c_0 \geq n_1$. For the spectral GLL, GLC or pseudospectral meshes, there exist coarse imbedded tensor product meshes with the coordinates of the mesh lines satisfying (8.126) and $\ell_0 = \mathcal{O}(\log N)$.*

Proof. The pseudospectral mesh is a particular case of the mesh in the above lemma at $\gamma = c_1 = c_2 = 1$, from where the statement follows for the pseudospectral meshes. The GLL and GLC meshes are geometrically equivalent to the pseudospectral meshes, their step sizes on $(-1, 0)$ are monotonically growing, but the ratios $\hbar_{i+1}/\hbar_i$ are monotonically diminishing. This approves Corollary and, in particular, the choice of c_0 for these meshes. □

The nests of the decomposition tensor product grid are denoted by $\delta_{k,l} = \{x : \zeta_{k-1} < x_1 < \zeta_k\,,\, \zeta_{l-1} < x_2 < \zeta_l\}$. The piecewise uniform *shifted fine mesh* $x_1, x_2 = \vartheta_i,$ is introduced in such a way that it

- contains the same number of mesh lines with the fine mesh, *i.e.,* for $i = 0, 1, \ldots, p$,
- covers the coarse mesh, *i.e.,* $\vartheta_{\widetilde{i}(k)} = \eta_{\widetilde{i}(k)} = \zeta_k$ for $k = 0, 1, \ldots, 2\ell_0$, and
- on each nest $\delta_{k,l}$ of the coarse mesh is the uniform mesh with the steps $h_k = a_k/n_k$.

Having divided each nest of the shifted fine mesh in two triangles, we define the triangulation of τ_0, the corresponding space of piecewise linear continuous functions, and the related FE matrices, denoted by $\mathcal{H}_D(\tau_0)$ and $\mathbf{B}^h_{\mathrm{D}}$, $\mathbf{M}^h_{\mathrm{D}}$, respectively.

Lemma 8.7. *Let* $\boldsymbol{\mathcal{A}}^{\hbar}$ *and* $\boldsymbol{\mathcal{M}}^{\hbar}$ *be the finite-element preconditioners (7.140), p. 248, defined on spectral or pseudospectral meshes. Then the matrices* $\boldsymbol{\mathcal{A}}^{\hbar}$, $\mathbf{B}^h_{\mathrm{D}}$ *as well as the matrices* $\boldsymbol{\mathcal{M}}^h$, $\mathbf{M}^h_{\mathrm{D}}$ *are spectrally equivalent uniformly in p,* i.e.*,*

$$\boldsymbol{\mathcal{A}}^{\hbar} \asymp \mathbf{B}^h_{\mathrm{D}} \quad \textit{and} \quad \boldsymbol{\mathcal{M}}^h \asymp \mathbf{M}^h_{\mathrm{D}}\,. \tag{8.127}$$

Proof. Since the spectral and pseudospectral meshes are geometrically equivalent, it is sufficient to compare preconditioners corresponding to one type of the fine mesh, *e.g.*, the preconditioners $\boldsymbol{\mathcal{A}}^{\hbar} = \boldsymbol{\mathcal{A}}^{\hbar}_{\mathrm{sp}}$ and $\mathbf{B}^h_{\mathrm{D}} = \mathbf{B}^h_{\mathrm{D,sp}}$ for the GLL fine mesh. The consequence of Corollary 8.6 is that the GLL mesh and the shifted GLL mesh are geometrically equivalent uniformly in p, *i.e.*,

$$c_0^{-1} h_k \le \hbar_i \le c_0 h_k\,,$$

for $i = \widetilde{i}(k-1)+1, \widetilde{i}(k-1)+2, \ldots, \widetilde{i}(k-1)+n_k = \widetilde{i}(k)\,,\ k = 1, 2, \ldots, \ell_0$. Simple calculations with the use of these inequalities yield the spectral equivalence inequalities

$$c_0^{-2}\mathbf{B}^h_{\mathrm{D,sp}} \le \boldsymbol{\mathcal{A}}^{\hbar}_{\mathrm{sp}} \le c_0^2 \mathbf{B}^h_{\mathrm{D,sp}}\,.$$

Similarly, the preconditioners for the reference element mass matrix satisfy the similar inequalities

$$c_0^{-2}\mathbf{M}^h_{\mathrm{D,sp}} \le \boldsymbol{\mathcal{M}}^{\hbar}_{\mathrm{sp}} \le c_0^2 \mathbf{M}^h_{\mathrm{D,sp}}\,.$$

The proof of Lemma now immediately follows from the two equivalences given above and the geometrical equivalence of the spectral and pseudospectral meshes. □

Corollary 8.7. *Let* $\mathbf{A}$ *and* $\mathbb{M}$ *be the stiffness and mass matrices of the spectral reference element with GLL nodes, and let* $\mathbf{B}_{\mathrm{D}}^{h}$ *and* $\mathbf{M}_{\mathrm{D}}^{h}$ *be the same matrices as in Lemma 8.7. Then the matrices* $\mathbf{A}$ *and* $\mathbf{B}_{\mathrm{D}}^{h}$ *as well as the matrices* $\mathbb{M}$ *and* $\mathbf{M}_{\mathrm{D}}^{h}$ *are spectrally equivalent uniformly in* p, i.e.,

$$\mathbf{A} \asymp \mathbf{B}_{\mathrm{D}}^{h} \quad \textit{and} \quad \mathbb{M} \asymp \mathbf{M}_{\mathrm{D}}^{h}\,.$$

Proof. For the proof, it is sufficient to take into account Lemmas 7.8 and 8.7. □

In what follows, we consider the fast local Dirichlet solver based on the preconditioner which is the block $\mathbf{B}_{\mathrm{D},I}^{h}$ of $\mathbf{B}_{\mathrm{D}}^{h}$ and for which we will use also the simpler notation $\mathbf{B}$ with all indices omitted. It is generated by means of the subspace $\mathring{\mathcal{H}}_D(\tau_0) = \{v \in \mathcal{H}_D(\tau_o) : v|_{\partial\tau_0} = 0\}$ of internal FE functions over the triangulation with the vertices in the nodes of the shifted GLL mesh. The DD_{loc} algorithm, that is quite similar to the one presented in Subsection 8.1.2, can be successfully applied to systems with the system matrix $\mathbf{B}$ directly. Moreover, the resulting algorithm has the same asymptotical computational cost. In order to illuminate the similarity of the algorithms, the same notations $\delta_{k,l}$, $\mathbf{B}_i^{k,l}$, $\mathbf{B}_i^{k,l}$, as in the DD_{loc} algorithm for the hierarchical reference element, will be used for the nests of the decomposition mesh and for the related finite element matrices and their blocks. Although, the objects with the same notations do not coincide, their computational properties are similar.

In the case under consideration, one can not separate degrees of freedom in four groups, which allowed to represent the hierarchical reference element $\mathcal{E}_H$ internal stiffness matrix in the block diagonal form (7.95) : $\mathbf{A}_I = \mathrm{diag}\,[\mathbf{A}_{ee}, \mathbf{A}_{eo}, \mathbf{A}_{oe}, \mathbf{A}_{oo}]$. Therefore, here $\mathbf{B}$ is the preconditioner for the whole internal stiffness $(p-1)^2 \times (p-1)^2$ matrix $\mathbf{A}_I$, whereas in the case of the reference element $\mathcal{E}_H$ the notation $\mathbf{B}$ stood for the preconditioner of any of the independent blocks $\mathbf{A}_{ee}, \mathbf{A}_{eo}, \mathbf{A}_{oe}, \mathbf{A}_{oo}$ of the dimension $\tilde{N}^2 \times \tilde{N}^2$, $(p-1)/2 \le \tilde{N} \le \lfloor p/2 \rfloor_+$. This causes some other differences in the components of DD_{loc} solver, increases its cost, but, nevertheless, do not damage its asymptotic optimality.

The fine spectral, the coarse and the shifted fine meshes have four lines of symmetry: two axes and two diagonals of the square τ_0. This is reflected in the shapes of the decomposition subdomains $\delta_{k,l}$, matrices of local Dirichlet problems in subdomains, subdomain Schur complements etc. For instance, in the case of the reference element $\mathcal{E}_H$, for each $k = 1, 2, \ldots, \ell_0$ we have one square subdomain $\delta_{k,k}$ and the corresponding subdomain stiffness matrix $\mathbf{B}_1^{k,k}$. In the case under consideration, for each square subdomain

$\delta_{k,k},\ k = 1, 2, \ldots, \ell_0,$ and matrix $\mathbf{B}_1^{k,k}$ there are three other square subdomains $\delta_{l,m},\ l, m = k, \ell_0 + k,$ of the same size and matrices $\mathbf{B}_1^{l,m}$ which are the same as $\mathbf{B}_1^{k,k}$ up to perturbations. For each rectangular subdomain $\delta_{k,l}$ there are 7 other subdomains of the same sizes to all of which correspond stiffness matrices $\mathbf{B}_1^{k,l}$ the same up to perturbations. Besides, in the cases of the hierarchical and spectral reference elements, the very matrices $\mathbf{B}_1^{k,l},\ k, l = 1, 2, \ldots, \ell_0,$ are different, see, *e.g.*, (8.129), (8.108).

Clearly, the aforementioned symmetry properties are worth to be taken into account for improving the efficiency of the local solver.

8.3.1.2 *Local Dirichlet Problems and Prolongations*

The local discrete Dirichlet problems

$$\mathbf{B}_1^{k,l}\mathbf{v} = \mathbf{f}\,, \quad 1 \le k, l \le 2\ell_0\,, \tag{8.128}$$

which are counterparts of the local discrete Dirichlet problems (8.54), became now the system of the finite element equations

$$-h_l\frac{v_{i-1,j} - 2v_{i,j} + v_{i+1,j}}{h_k} - h_k\frac{v_{i,j-1} - 2v_{i,j} + v_{i,j+1}}{h_l} = f_{i,j}\ ,$$

$$\widetilde{i}(k-1) < i < \widetilde{i}(k)\ ,\ \ \widetilde{i}(l-1) < j < \widetilde{i}(l)\ , \tag{8.129}$$

$$v_{i,j} = 0 \quad \text{when } i = \widetilde{i}(k-1), \widetilde{i}(k) \text{ or } j = \widetilde{i}(l-1), \widetilde{i}(l)\ .$$

Systems (8.129) may be solved by the same fast algorithms with the same subdomain preconditioners $\boldsymbol{\mathcal{B}}_1^{k,l}$, which were pointed out for solving systems (8.55), (8.108). For instance, by 2d FDFT or a combination of 1d FDFT and Gauss elimination for tridiagonal systems, both requiring $\mathcal{O}(p^2 \log p)$ a.o. Moreover, optimal PCGM solvers with multilevel preconditioners of BPX and MDS types for orthotropic discretizations, discussed in Subsubsection 8.1.2.2 can be implemented. We note that algebraic multigrid methods for orthotropic discretizations derived on the basis of the strengthened Cauchy inequality by [Axelsson and Vassilevski (1989)], as well as other types, *e.g.,* independently considered by [Schieweck (1986)] and [Pflaum (2000)], are also applicable to (8.129).

As well the same prolongation operators $\mathbf{P}_{k,l} : U_{\mathbb{III}}^{k,l} \mapsto \mathbb{U}^{k,l}$, as in Subsubsection 8.1.2.2, can be used in the $DD_{\rm loc}$ algorithm under consideration. This and above discussion of the local Dirichlet preconditioner-solvers $\boldsymbol{\mathcal{B}}_1^{k,l}$ allow us to retain the assumptions (8.57) and (8.58).

8.3.1.3 *Schur Complement and $DD_{\rm loc}$ Preconditioning*

With a few exceptions, we now use the same notations for similar objects in the Schur complement preconditioning algorithm as in Subsection 8.1.2.

Let $\mathbf{B}$ be represented in the block forms

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\mathbb{I}} & \mathbf{B}_{\mathbb{I},\mathbb{III}} \\ \mathbf{B}_{\mathbb{III},\mathbb{I}} & \mathbf{B}_{\mathbb{III}} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{21} & \mathbf{B}_2 & \mathbf{B}_{23} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_3 \end{pmatrix}, \tag{8.130}$$

where $\mathbf{B}_{\mathbb{I}} = \mathbf{B}_1$ and $\mathbf{B}_{31} = \mathbf{B}_{13} = \mathbf{0}$, and let the Schur complement $\mathbb{S}_{\mathbb{III}}$ and its block $\mathbb{S}_E$ on the diagonal be defined by the same expressions as in (8.49).

The subdomain Schur complement $\mathbb{S}_E^{k,l}$ and its preconditioner $\boldsymbol{\mathcal{S}}_E^{k,l}$ can be again represented in the forms (8.60) and(8.61), whereas the global preconditioner $\boldsymbol{\mathcal{S}}_E$ results from the assembling procedure. In the same way as in Subsubsection 8.1.3.1, we can introduce the piecewise trigonometric basis in the subspace $U_{12} := U_1 \bigcup U_2$, which, for brevity, will be called *trigonometric basis*. The notation $\boldsymbol{\mathcal{S}}_{E,\rm tr}$ will stand for the preconditioner $\boldsymbol{\mathcal{S}}_E$, transformed to this basis, and similar changes in the notations are assumed for other matrices.

Lemma 8.8. *For the appropriate ordering of the unknowns, matrix $\boldsymbol{\mathcal{S}}_{E,\rm tr}$ is the block diagonal matrix with $2(p-2l_0)$ independent blocks, each being a $(2l_0-1)\times(2l_0-1)$ tridiagonal matrix. The arithmetical cost of computation of the matrices $\mathbb{S}_{E,\rm tr}$ and $\boldsymbol{\mathcal{S}}_{E,\rm tr}$, and the arithmetical cost of solving systems with the matrices $\boldsymbol{\mathcal{S}}_E, \boldsymbol{\mathcal{S}}_{E,\rm tr}$ are estimated as $\mathcal{O}(p^2)$ and $\mathcal{O}(p(\log p))$, respectively.*

Proof. The structures of the matrices $\mathbb{S}_E^{k,l}$, $\boldsymbol{\mathcal{S}}_E^{k,l}$ and $\mathbb{S}_{E,\rm tr}^{k,l}$, $\boldsymbol{\mathcal{S}}_{E,\rm tr}^{k,l}$ are similar to the structures of the analogous matrices described in Subsubsection 8.1.3.1 for the case the reference element $\mathcal{E}_H$. In view of this, proofs of the first part of Lemma 8.8 and Lemma 8.5 differ only in dimensions and numbers of involved subdomains and matrices. The proof of the second part of Lemma 8.8 follows the proof of the corresponding bounds of the arithmetical costs in Theorem 8.6. □

Theorem 8.9. *Let $\mathbb{S}_E$ and $\boldsymbol{\mathcal{S}}_E$ be the matrices defined above for the block $\mathbf{B}$. Then the inequalities*

$$\frac{1}{(1+\log p)^2}\,\boldsymbol{\mathcal{S}}_E \prec \mathbb{S}_E \prec \boldsymbol{\mathcal{S}}_E \tag{8.131}$$

hold, and the arithmetical cost for solving each of the systems

$$\mathbb{S}_E \mathbf{v}_E = \mathbf{f}_E \quad \textit{and} \quad \mathbb{S}_{E,\mathrm{tr}} \mathbf{v}_{E,\mathrm{tr}} = \mathbf{f}_{E,\mathrm{tr}}$$

by means of the PGC with the preconditioners $\boldsymbol{\mathcal{S}}_E$ *and* $\boldsymbol{\mathcal{S}}_{E,\mathrm{tr}}$*, respectively, is*

$$\mathcal{O}(p^2(\log p))\,. \tag{8.132}$$

Proof. The proof of (8.131) and (8.132) is the same as the proof of (8.96) and (8.95). □

In order to obtain an optimal DD_{loc} preconditioner-solver, we need the preconditioners $\boldsymbol{\mathcal{S}}_{\mathbb{III},k}$, $k = 1, 2$, for the Schur complement $\mathbb{S}_{\mathbb{III}}$ with the properties similar to those described for their counterparts in Subsubsections 8.1.2.3 and 8.1.3.1.

Let $\mathring{\mathcal{H}}_V(\pi_1) = \mathring{\mathcal{H}}_{V,\square}(\tau_0), \mathring{\mathcal{H}}_{V,\triangle}(\tau_0)$ be the notations for the spaces of continuous on τ_0 piecewise bilinear on the decomposition mesh and piecewise linear on the coarse triangulation functions, respectively. Let also $\mathbf{B}_V$ be the finite element matrix, generated by the space $\mathring{\mathcal{H}}_V(\tau_0)$, $\boldsymbol{\mathcal{B}}_V$ be a preconditioner for $\mathbf{B}_V$, and $\boldsymbol{\mathcal{D}}_V = \lfloor \mathbf{B}_V \rfloor_{\mathrm{diag}}$. It is reasonable to set

$$\boldsymbol{\mathcal{S}}_{\mathbb{III},2} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \boldsymbol{\mathcal{B}}_V]$$

with a good option $\boldsymbol{\mathcal{B}}_V = \boldsymbol{\mathcal{D}}_V$.

Lemma 8.9. *The following statements are valid:*

1) The matrices $\mathbf{B}_V$ *and* $\boldsymbol{\mathcal{D}}_V$ *are spectrally equivalent uniformly in p,* i.e., *inequalities*

$$\underline{\gamma}_V \boldsymbol{\mathcal{D}}_V \le \mathbf{B}_V \le \overline{\gamma}_V \boldsymbol{\mathcal{D}}_V \tag{8.133}$$

hold with generic positive constants $\underline{\gamma}_V$ *and* $\overline{\gamma}_V$*.*

2) For the Schur complement preconditioner

$$\boldsymbol{\mathcal{S}}_{\mathbb{III},2} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_E, \boldsymbol{\mathcal{D}}_V]\,,$$

we have the inequalities

$$\frac{1}{(1+\log p)p}\, \boldsymbol{\mathcal{S}}_{\mathbb{III},2} \prec \mathbb{S}_{\mathbb{III}} \prec \boldsymbol{\mathcal{S}}_{\mathbb{III},2}\,. \tag{8.134}$$

Proof. It is not difficult to notice that sizes of the nests of the decomposition meshes are changing more rapidly, than in the case in the DD_{loc} algorithm for the hierarchical reference element. Therefore, in order to prove 1) it is sufficient to follow the outline of the proof of Lemma 8.4. For the proof of 2), we turn to the energy bounds (6.154), a particular case

of which are the inequalities (8.134). On the basis of the definitions of the imbedded decomposition meshes and the fine shifted piecewise uniform meshes, we come to the conclusion that

$$\max_j \underline{\mu}_j^{-1} \prec \max_j \left\{(1+\log \underline{m}_j)\cdot \max(\theta_j, 1+\log \underline{m}_j)\right\} \le (1+\log p)p\,.$$

This completes the proof of the lemma. □

Our aim is to obtain a $DD_{\rm loc}$ preconditioner which is optimal with respect to the arithmetical complexity of the preconditioning operation. For attaining this, any preconditioner-multiplicator $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$, satisfying the spectral equivalence inequalities

$$\underline{\gamma}_1 \boldsymbol{\mathcal{S}}_{\mathbb{III},1} \le \mathbb{S}_{\mathbb{III}} \le \overline{\gamma}_1 \boldsymbol{\mathcal{S}}_{\mathbb{III},1} \tag{8.135}$$

with some positive generic constants $\underline{\gamma}_1$ and $\overline{\gamma}_1$ and the arithmetical complexity bound

$$\mathrm{ops}\,[\boldsymbol{\mathcal{S}}_{\mathbb{III},1}\mathbf{v}] \le \Psi(p) := p^{3/2}[\underline{\gamma}_1/\overline{\gamma}_1(1+\log p)]^{1/2}, \tag{8.136}$$

is suitable. We recall that the existence of a good solver for $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$ is not assumed. Some choices are suggested by the results of Subsection 8.1.2 for the counterpart of this preconditioner in $DD_{\rm loc}$ algorithm for the reference element $\mathcal{E}_H$. The preconditioner-multiplicator of the same type as $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$ in Subsubsection 8.1.2.3 seems to be among the best choises. In the case under consideration, we have

$$\Psi(p) = \mathcal{O}(p\log^2 p)\,,$$

and $\underline{\gamma}_1$ and $\overline{\gamma}_1$ are the constants depending only on c_0 in (8.126).

Let (8.135) and (8.136) be fulfilled. According to (8.134), (2.50) and (2.52), in order to obtain an iterative Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\rm it} := \boldsymbol{\mathcal{I}}_\circ\,[\boldsymbol{\mathcal{S}}_{\mathbb{III},1}, \boldsymbol{\mathcal{S}}_{\mathbb{III},2}]$ satisfying the inequalities

$$0.5\,\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\rm it} \le \boldsymbol{\mathcal{S}}_{\mathbb{III},1} \le 1.5\,\boldsymbol{\mathcal{S}}_{\mathbb{III},\,\rm it}\,,$$

it is sufficient to make

$$\nu_{1/2} = \mathcal{O}\left(\sqrt{\overline{\gamma}_1 p(1+\log p)/\underline{\gamma}_1}\right)$$

iterations. In this case, the arithmetical cost of the matrix-vector multiplication by $\boldsymbol{\mathcal{S}}^{-1}_{\mathbb{III},\,\rm it}$ obeys the bound

$$\mathrm{ops}\,[\boldsymbol{\mathcal{S}}^{-1}_{\mathbb{III},\,\rm it}\mathbf{v}] \le c\left[p^2 + p(1+\log p)^2\sqrt{p(1+\log p)\overline{\gamma}_1/\underline{\gamma}_1}\right]\,, \quad \forall\ \mathbf{v}\in U_{\mathbb{III}}\,,$$

where c is a positive generic constant.

With the Schur complement preconditioner defined, descriptions of the Schur complement and $DD_{\rm loc}$ solvers do not differ from given in Subsubsection 8.1.2.2 for the case of the hierarchical reference element. There is also no difference in the ways of deriving the bounds of total arithmetic complexity, especially if to take into account similarity of the formulated subsidiary results. For this reason, we present only final conclusions, omitting the proofs. For the total numbers of arithmetic operations for the Schur complement and $\mathrm{DD}_{\rm loc}$ solvers, we use the notations $Q_{S,\rm sp}$ and $Q_{DD,\rm sp}$, respectively.

Theorem 8.10. *Let the conditions (8.135) and (8.136) be fulfilled with* $\overline{\gamma}_1, \overline{\gamma}_1 = \mathrm{const} > 0$. *Then we have*

$$Q_{S,\rm sp} = \mathcal{O}(N^2)\,. \tag{8.137}$$

If additionally (8.57),(8.58) hold, then the arithmetical complexity of the $DD_{\rm loc}$ *solver is bounded by*

$$Q_{DD,\rm sp} = \mathcal{O}(N^2)\,. \tag{8.138}$$

In distinction from the case of the hierarchical reference element, the described $DD_{\rm loc}$ fast algorithm is with no difficulties adjustable to the reference element stiffness matrix preconditioner $\mathbf{B}_{\rm D}^{h}$, which can be used in the DD Neumann-Dirichlet and other algorithms.

8.3.2 *Fast Multilevel Solver*

In this section we consider an alternative approach to designing fast solvers for local Dirichlet problems governed by the internal stiffness matrices of the spectral finite or reference elements. It is based on the use of the factorized preconditioners of Subsection 7.4.3 and the closeness in the important properties of the main factor in the factorized preconditioner to the properties of the finite-difference/finite element preconditioners for the hierarchical reference element.

According to Lemma 7.10, p. 251, Corollary 7.4, and Theorem 7.6, the matrix $\widetilde{\boldsymbol{\mathcal{A}}}_I$ is spectrally equivalent to the matrix $\boldsymbol{\Lambda}_\diamond$ and to the matrix $\hbar^{-2}\mathbf{B}_\diamond$ uniformly in p. Moreover, *e.g.*, for the spectral elements with the GLL nodes, the stiffness matrices $\mathbf{A}_I$, $\mathbf{K}_{I_r}$ are spectrally equivalent to $\mathfrak{B}_\diamond = \hbar^{-2}\mathbf{C}_I^{-1}\mathbf{B}_\diamond\mathbf{C}_I^{-1}$ uniformly in p. For coming to this conclusion, it is sufficient to take additionally into account Lemma 7.8 and the generalized conditions of the shape regularity. As a consequence, in order to get a fast solver for all these matrices it is sufficient to derive a fast solver for $\boldsymbol{\Lambda}_\diamond$

or $\mathbf{B}_\diamond$. The properties of the latter two matrices are quite similar to the properties of the preconditioners $\mathbf{\Lambda}$ and $\mathbf{B}$, respectively, figuring in (7.110) and (8.2). This similarity allows to contemplate the multilevel solver for the matrix $\mathbf{B}_\diamond$, which has much in common with the solver for $\mathbf{B}$, presented in Subsection 8.1.1. The choice of the preconditioner $\mathbf{B}_\diamond$ is explained by its convenience for the analysis, whereas for the practice $\mathbf{\Lambda}_\diamond$ seem more suitable.

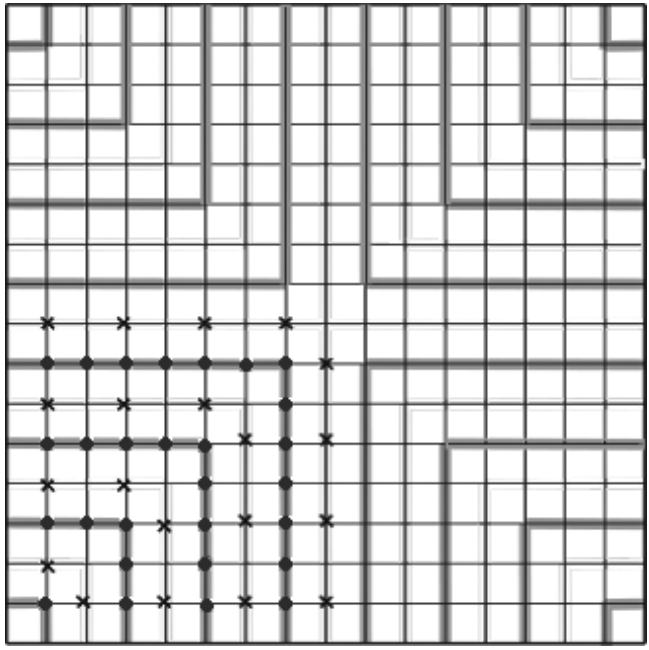

Fig. 8.9 Multigrid meshes.

To avoid confusion, we note that below we switch to the same notations for the objects playing the same role in the structure of the multigrid algorithm of this subsection and in the algorithm of Subsection 8.1.1.

Without loss of generality, we may assume $p = 2N,\ N = 2^{\ell_0 - 1}$, and introduce the sequence of ℓ_0 imbedded square meshes of the sizes $\hbar_l = 2^{1-l},\ l = 1, 2, \ldots, \ell_0$. The sets of the nodes $\xi = \hbar_l(i, j)$ in τ_0 of these meshes are denoted by X_l. On this sequence of meshes, we define the sequence of the finite element spaces $\mathcal{H}^0_l(\tau_0)$ and the sequence of the finite element matrices $\mathbf{B}_l$, such that $\mathcal{H}^0_{\ell_0}(\tau_0) = \mathcal{H}^0(\tau_0)$ and $\mathbf{B}_{\ell_0} = \mathbf{B}_\diamond$. For simplicity, we

write $\mathcal{H}^0_l$ instead of $\mathcal{H}^0_l(\tau_0)$. Moreover, we use the following notations:

- V_l and W_l are the spaces of vectors with the entries, specified on the sets of nodes X_l and $X_{w,l} := X_l \setminus X_{l-1}$, such that

$$V_l = V_{l-1} \oplus W_l = W_l \oplus W_{l-1} \oplus ... \oplus W_2 \oplus W_1 , \quad W_1 := V_1 .$$

- $\mathbf{P}_{l-1} : V_{l-1} \to V_l$ is the usual interpolation operator from the grid of the level $l-1$ to the next level finer grid. If $v \in \mathcal{H}_{l-1}(\tau_o)$ and $\mathbf{v}_{l-1} \in V_{l-1}$ is its vector representation, then $\mathbf{v}_l = \mathbf{P}_{l-1}\mathbf{v}_{l-1}$ is the vector representation of v in V_l.
- $\mathbf{R}_l : V_l \to W_l$ is the restriction matrix such that $\mathbf{v}_{l-1} := \mathbf{R}_l\mathbf{v}_l$ contains the entries of $\mathbf{v}_l$ for the nodes of $X_{w,l}$.
- $\mathbf{B}_{v,l} = \mathbf{B}_{l-1}$ and $\mathbf{B}_{w,l}$ are the blocks on the diagonal of $\mathbf{B}_l$ corresponding to the subspaces V_{l-1} and W_l, respectively.

Let us suppose that we have a sufficiently good preconditioner $\boldsymbol{\mathcal{B}}_{w,l}$ for the wavelet block $\mathbf{B}_{w,l}$ of the matrix $\mathbf{B}_l$. Then, one multigrid iteration for the system $\mathbf{B}_l\mathbf{u} = \mathbf{F}$ producing $\mathbf{u}^{k+1,l} := \mathbf{mgm}(l, \mathbf{B}_l, \mathbf{F}, \mathbf{u}^{k,l})$ from a given approximation $\mathbf{u}^{k,l}$ is described as follows:

- **If** $l \geq 1$, then **do**

  (1) Presmoothing in the space W_l:
  $\mathbf{v} := \mathbf{u}^{k,l}$;
  do ν **times** $\quad \mathbf{v} := \mathbf{v} - \sigma_l^{-1}\mathbf{R}_l^t\boldsymbol{\mathcal{B}}_{w,l}^{-1}\mathbf{R}_l(\mathbf{B}_l\mathbf{v} - \mathbf{F})$;

  (2) Coarse grid correction in the space V_{l-1}:
  $\mathbf{d}_{l-1} := \mathbf{P}_{l-1}^t(\mathbf{F} - \mathbf{B}_l\mathbf{v})$; $\mathbf{w} = 0$;
  do μ_{l-1} **iterations** $\mathbf{w} = \mathbf{mgm}(l-1, \mathbf{B}_{l-1}, \mathbf{d}_{l-1}, \mathbf{w})$;
  $\mathbf{v} := \mathbf{v} + \mathbf{P}_{l-1}\mathbf{w}$;

  (3) Postsmoothing in the space W_l:
  do ν **times** $\quad \mathbf{v} := \mathbf{v} - \sigma_l^{-1}\mathbf{R}_l^t\boldsymbol{\mathcal{B}}_{w,l}^{-1}\mathbf{R}_l(\mathbf{B}_l\mathbf{v} - \mathbf{F})$;
  $\mathbf{u}^{k+1,l} = \mathbf{v}$

- **else**
  solve $\mathbf{B}_l\mathbf{u} = \mathbf{F}$ **some direct method**
- **endif**

The optimal constant iteration parameter for the convergence in the norm induced by the matrix $\mathbf{B}_{l,w}$ is $\sigma_l = 2/(\lambda_{min} + \lambda_{max})$, where $\lambda_{min} = \lambda_{min}(\boldsymbol{\mathcal{B}}_{w,l}^{-1}\mathbf{B}_{l,w})$ and $\lambda_{max} = \lambda_{max}(\boldsymbol{\mathcal{B}}_{w,l}^{-1}\mathbf{B}_{l,w})$ are minimal nonzero and maximal eigenvalues of the generalized eigenvalue problem $\mathbf{B}_{l,w}\mathbf{v} = \lambda\boldsymbol{\mathcal{B}}_{w,l}\mathbf{v}$. In the case of the uniform in p and l spectral equivalence of $\mathbf{B}_{l,w}$ and $\boldsymbol{\mathcal{B}}_{w,l}$,

which will be guaranteed by the choice of the preconditioner $\boldsymbol{\mathcal{B}}_{w,l}$, the iteration parameter may be the same for all levels. However, due to the use of the iterations in items 1 and 3 as inexact solvers, the practically optimal constant iteration parameter can differ from the pointed out one and can be adjusted, *e.g.*, on the basis of numerical experiment.

In the accepted notations, the described algorithm is identical to the one considered in Subsection 8.1.1. This is in spite of the differences between the preconditioners $\mathbf{B}_{\diamond}$ and $\mathbf{B}$ for the internal stiffness matrices of the spectral and hierarchical reference elements, respectively, and the differences between the FE spaces and the bilinear forms, by means of which these preconditioners are generated. The distinction from the algorithm of Subsection 8.1.1 is more showing when we look at the specific structures of the involved matrices and, in particular, at the structure of the preconditioner $\boldsymbol{\mathcal{B}}_{w,l}$ for the pre- and post-smoothing iterations. Clearly, the choice of the preconditioner $\boldsymbol{\mathcal{B}}_{w,l}$ is very important for the efficiency of the multilevel solver.

Let us turn to the quarter $(-1,0)\times(-1,0)$ of the square τ_0. For each l-th mesh, we denote by $\Im_{l,\kappa}$, where $\kappa = 1, 2, \ldots, 2^{l-1}$, the mesh line passing through the nodes with the indices (i,j) satisfying $\max(i,j) \equiv \kappa$. For each κ, the line $\Im_{l,\kappa}$ has two parts: the segment $-1 \le \xi_1 \le (-1+\hbar_l\kappa)$ of the line $\xi_2 \equiv -1+\hbar_l\kappa$ and the segment $-1 \le \xi_2 \le (-1+\hbar_l\kappa)$ of the line $\xi_1 \equiv -1+\hbar_l\kappa$. On the rest three subsquares of τ_0, we define each line $\Im_{l,\kappa}$ by the symmetry with respect to the axes and the diagonals of τ_0. Therefore, each line $\Im_{l,\kappa}$ consists of four disjoint parts in four quarters of the coordinate system. For all l, the lines $\Im_{l,N_l}$ with $N_l = 2^{l-1}$ coincide with the intersection of the axes with τ_0.

We can reorder the rows and columns, crossing the diagonal of the block $\mathbf{B}_{w,l}$, in such a way that we obtain sub-blocks $\mathbf{B}_{w,l}^{(\kappa)}$ on diagonal, each related to the nodes on the line $\Im_{l,\kappa}$. Having in mind the pointed out block structure, the preconditioner $\boldsymbol{\mathcal{B}}_{w,l}$ is obtained from $\mathbf{B}_{w,l}$ by making zero all off-diagonal blocks. For the blocks on diagonal of $\boldsymbol{\mathcal{B}}_{w,l}$, we set $\boldsymbol{\mathcal{B}}_{w,l}^{(\kappa)} = \mathbf{B}_{w,l}^{(\kappa)}$. It is easy to see, that for even numbers κ, the blocks $\boldsymbol{\mathcal{B}}_{w,l}^{(\kappa)}$ are diagonal. In the blocks $\boldsymbol{\mathcal{B}}_{w,l}^{(\kappa)}$ with odd κ, only the neighbouring nodes on the line $\Im_{l,\kappa}$ are coupled, so that these blocks are tridiagonal. Indeed, for $\kappa < N_l$, the blocks $\boldsymbol{\mathcal{B}}_{w,l}^{(\kappa)}$ contain four independent subblocks, each related to only one connected part of the line $\Im_{l,\kappa}$.

Nodes of the set $X_{w,l}$ are depicted in Fig. 8.9. If to accept that the fine mesh corresponds to the level "l", then circles and crosses mark the

nodes of the set $X_{w,l}$, which live on the lines $\Im_{l,\kappa}$ with odd and even κ, respectively.

The analysis of the convergence consists of the three steps, which are the same as in the case of the hierarchical reference element. The first step is the proof of the spectral equivalence

$$c_1 \boldsymbol{\mathcal{B}}_{l,w} \leq \mathbf{B}_{l,w} \leq c_2 \boldsymbol{\mathcal{B}}_{l,w} \tag{8.139}$$

with the positive constants c_1 and c_2 independent of p and l. The second step is the estimation of the constant in the strengthed Cauchy inequality

$$\left(a^{(1)}_{\tau_0}(u,w)\right)^2 \leq c_0\, a^{(1)}_{\tau_0}(u,u)\, a^{(1)}_{\tau_0}(w,w) \quad \forall u \in \mathcal{H}_{l-1}\,,\ \forall w \in \mathcal{W}_l\,, \tag{8.140}$$

with $c_0 < 2/3$ and $\mathcal{W}_l \leftrightarrow W : \mathcal{H}_l = \mathcal{W}_l \oplus \mathcal{H}_{l-1}$. The third step is the application of the results by [Schieweck (1986)] and [Pflaum (2000)], which together with (8.139) and (8.140) allow us to formulate Theorem 8.11 below, establishing the rate of the convergence of the multilevel solver in the norm induced by the matrix $\mathbf{B}_l$.

Theorem 8.11. *Let us choose the parameters σ, μ and ν in the multigrid solver for the system $\mathbf{B}_l\mathbf{u} = \mathbf{F}$ as follows: $\sigma = 2/(c_1 + c_2)$, $\mu \geq 3$, and ν greater than some $\nu_o(c_0, c_1, c_2)$. Then the convergence factor*

$$\widetilde{\rho}_{l,\mathrm{mult}} := \sup_{\mathbf{u}^k \in U_l} \frac{\|\mathbf{u}^{k+1} - \mathbf{u}\|_{\mathbf{B}_l}}{\|\mathbf{u}^k - \mathbf{u}\|_{\mathbf{B}_l}}$$

is estimated by some constant $\widetilde{\rho} < 1$ which does not depend on p and l.

Proof. The basic constants c_0, c_1, c_2 completely control the efficiency of the multigrid method. Their estimation can be done with the use of the same superelement technique with the same superelements as in the proof of Theorem 8.4. In view of the symmetry of the involved discrete problems with respect to the both axes, it is sufficient to consider superelements of one quoter, *e.g.*, $(-1,0) \times (-1,0)$. The finite element functions from $\mathcal{H}_l(\tau_0)$ are not fixed on the upper and right edges of this quoter, and it is the only difference which, however, practically does not influence the implementation of the superelement technique. It yields the same bounds for the aforementioned constants, which approve the theorem. □

It is worth noting that we are interested in the convergence of the vector of our source unknowns in the norm which is equivalent to the norm induced by the reference element internal stiffness matrix $\mathbf{A}_I$ or by the internal stiffness matrix $\mathbf{K}_{I_r}$ of some finite element τ_r. Under the generalized

conditions of quasiuniformity conditions, the former norm for any vector $\mathbf{v} \in U_I$ is equivalent to the seminorm $|\cdot|_{1,\tau_r}$ for the FE functions living in the interior of τ_r. For example, let us take the discretization by the spectral finite elements with the GLL nodes. If $\mathbb{B}_\hbar$ is used as the preconditioner for the Dirichlet problems on finite elements, we solve the system $\mathbb{B}_\hbar \mathbf{v} = \mathbf{f}$ by transforming it to the system $\mathbf{B}_+ \mathbf{u} = \mathbf{F}$, $\mathbf{F} = \hbar^2 \mathbf{C}_I \mathbf{f}$, and applying k multigrid iterations to the latter system. Then, after transforming the kth iterate to the initial variables, we arrive at the convergence estimate

$$\|\mathbf{v}^k - \mathbf{v}\|_{\mathbb{B}_\hbar} \le \widetilde{\rho}^k_{\ell_0,\mathrm{mult}} \|\mathbf{v}^0 - \mathbf{v}\|_{\mathbb{B}_\hbar}, \tag{8.141}$$

which follows from Theorem 8.11 and the relation $\mathbb{B}_\hbar = \hbar^{-2} \mathbf{C}_I^{-1} \mathbf{B}_{\mathrm{sp}} \mathbf{C}_I^{-1}$. According to Lemmas 7.8 and 7.10, Theorem 7.6, Corollary 7.4, and the generalized conditions of quasiuniformity, the matrix $\mathbb{B}_\hbar$ is spectrally equivalent to the stiffness matrices $\mathbf{A}_I$ and $\mathbf{K}_{I_r}$ (at $\rho = 1$ on τ_r) uniformly in p. Therefore, we obtain convergence estimates in the norms $||\cdot||_{\mathbf{A}_I}$, $||\cdot||_{\mathbf{K}_{I_r}}$, $|\cdot|_{1,\tau_0}$, $|\cdot|_{1,\tau_r}$ with the same order as in (8.141).

In the computational practice, the most efficient approach often consists in using the multigrid method for constructing spectrally equivalent multigrid preconditioner by a few multigrid iterations. Let $\mathbf{M}_\mu$ be the linear error transition operator defined by one multigrid cycle applied to the system $\mathbf{B}_{\mathrm{sp}} \mathbf{u} = \mathbf{F}$. Then $\Bbbk$ multigrid iterations applied to solving this system implicitly define the preconditioner $\mathbf{Mg}_{\mathrm{sp}}^{-1} = \hbar^{-2} \mathbb{D}_I^{-1} (\mathbf{I} - \mathbf{M}_\mu^\Bbbk) \mathbf{B}_{\mathrm{sp}}^{-1} \mathbb{D}_I^{-1}$ for $\mathbf{C}_I^{-1}$, which is also a spectrally equivalent preconditioner for $\mathbf{A}_I^{-1}$.

Theorem 8.12. *Let $\mu = 3$, $\nu = 3$ and $\Bbbk \ge 1$. Then the spectral equivalence inequalities*

$$\underline{c}\, \mathbf{Mg}_{\mathrm{sp}}^{-1} \le \mathbf{A}_I^{-1} \le \overline{c}\, \mathbf{Mg}_{\mathrm{sp}}^{-1} \tag{8.142}$$

hold with the positive spectral equivalence constants $\underline{c}$ and $\overline{c}$ independent of p (and $\Bbbk$). Moreover, for $\Bbbk = 1$ the procedure of the matrix-vector multiplication by $\mathbf{Mg}_{\mathrm{sp}}^{-1}$ requires not more than $\mathcal{O}(p^2)$ arithmetic operations.

Proof. We use the spectral equivalence $(\mathbf{I} - \mathbf{M}_\mu^\Bbbk) \mathbf{B}_{\mathrm{sp}}^{-1} \asymp \mathbf{B}_{\mathrm{sp}}^{-1}$, which is the consequence of Theorem 8.11, the spectral equivalence $\mathbf{A}_I \asymp \mathbb{B}_\hbar$, and the relation between $\mathbb{B}_\hbar$ and $\mathbf{B}_{\mathrm{sp}}$. □

Obviously, for growing ν and $\Bbbk$, the quality of the preconditioning only improves, and the constants $\underline{c}$ and $\overline{c}$ approach the unity. At the same time, a good strategy with the respect to the overall complexity of the DD_{loc} preconditioner is to use relatively small ν and $\Bbbk$ in the range of their admissible values.

8.4 The Numerical Complexity of DD Methods in Two Dimensions

8.4.1 *Hierarchical Discretizations*

Here we look at the case when all finite elements are associated with a single complete hierarchical reference element $\mathcal{E}_H$ and the DD preconditioner $\boldsymbol{\mathcal{K}}$ is defined for the FE stiffness matrix $\mathbf{K}$ by (7.27). In this preconditioner, the vertex d.o.f. are split from the others. It contains four main components, which include three preconditioner-solvers and one prolongation operator from the edges inside the finite elements. The preconditioner-solvers are aimed at solving internal Dirichlet problems on finite elements, internal problems on edges of finite elements and one global problem related to the d.o.f. at the vertices of finite elements. Evidently, each component, except for the vertex solver, suggests natural elementwise, or edgewise parallelization of computations.

We remind that $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$ is defined as the restriction of the prolongation matrix $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}$ to the prolongation operator from one edge E of the reference element, see Section 7.1. Let us also note that the stiffness matrix $\mathbf{A}$ of the reference element $\mathcal{E}_H$ is sparse and matrix-vector multiplication by $\mathbf{A}$ is cheap. Therefore, in this case, the choice $\mathbb{A} = \mathbf{A}$ in the prolongation $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$ is reasonable, see Lemma 7.4 on page 222.

Theorem 8.13. *Let the generalized conditions of the shape quasiuniformity be fulfilled, and let the DD preconditioner $\boldsymbol{\mathcal{K}}$ for the FE stiffness matrix $\mathbf{K}$ be defined by (7.27). Let also the following conditions hold:*

a) for any $\mathbf{v}_V$, $\operatorname{ops}[\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{v}_I] = \mathcal{O}(p^2\log^\theta(1+p))$, $\operatorname{ops}[\boldsymbol{\mathcal{K}}_V^{-1}\mathbf{v}_V] = \mathcal{O}(\mathcal{R})$, and $\operatorname{cond}[\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{A}_I] = \mathcal{O}(1)$ and $\operatorname{cond}[\boldsymbol{\mathcal{K}}_V^{-1}\mathbf{K}_V] = \mathcal{O}(1)$;

b) the single edge preconditioner $\boldsymbol{\mathcal{S}}_{00}$ is defined as in Subsection 7.5.3;

c) the prolongation matrix $\boldsymbol{\mathcal{P}}_{E,\mathrm{ref}}$ is defined according Lemma 7.4 with $\mathbb{A} = \mathbf{A}$ and $\nu \succ \log(1+p)$.

Then the following estimates hold:

$$\begin{aligned} &\operatorname{cond}[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \mathcal{O}(\log^2(1+p))\,, \\ &\operatorname{ops}[\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}] = \mathcal{O}(\mathcal{R}p^2\log^{\theta+1}(1+p))\,, \quad \forall\ \mathbf{v}\in V\,. \end{aligned} \tag{8.143}$$

Proof. The proof follows directly from assumptions $a)-c)$, Corollary 7.5 and Lemma 7.4. □

The block $\mathbf{K}_V$ is the FE stiffness matrix generated by the assemblage of the shape-quasiuniform finite elements associated with the square bilinear

reference element. If the dimension of this block is large, the problem of designing of a good preconditioner-solvers $\boldsymbol{\mathcal{K}}_V$ arises. Preconditioner-solvers $\boldsymbol{\mathcal{K}}_V$, satisfying the assumption *a*), can be designed with the use of the techniques studied in Chapters 3–6. They are efficient in the case of a sufficiently smooth ϱ varying not too much from the constant, *e.g.*, $\varrho \approx 1$, $\forall\, x \in \Omega$. Cases of ϱ having significant relative jumps on the inter-element boundary, changing chaotically from one finite element to another and other paths of singular behavior require special consideration. An example of a fast solver for the case of the diagonal coefficient matrix $\boldsymbol{\varrho} > 0$ changing arbitrarily between subdomains of a nonuniform coarse orthogonal grid was considered in Chapter 6. A variety of other algorithms can be found in the vast literature on this subject.

Several preconditioners $\boldsymbol{\mathcal{A}}_I$, satisfying assumptions *a)* with $\theta = 0$ and 1, were considered in Section 8.1. They are based on inexact solvers of the multilevel and $DD_{\rm loc}$ types for the preconditioners which were obtained by the adaptation of the source preconditioners, derived in Section 7.4, to the particular type fast solver. Therefore, taking into account results of Section 8.1 and the theorem above, we are able to contemplate DD solvers, which in general require $\mathcal{O}(\log(1+p))$ PCGM iterations, while on each iteration $\mathcal{O}(p^2 \log(1+p))$ arithmetic operations are spent for solving systems with the DD preconditioners. Logarithms in the computational costs are due to the splitting the vertex unknowns from the rest, splitting the degrees of freedom living on different edges in the Schur complement preconditioner, and the prolongations by means of the inexact Dirichlet iterative solvers. However, we also considered spectrally equivalent inter-element boundary Schur complement preconditioners, in which splitting vertex d.o.f. is avoided. According to Theorem 7.7, they do not compromise the optimality of the DD preconditioner. In order to come to an optimal DD preconditioner with respect to the arithmetic operation counts, it is necessary to use additionally optimal prolongations from the inter-element boundary. We refer to [Beuchler and Schöberl (2005)] for an example of such a prolongation procedure.

8.4.2 *Discretizations by Spectral Elements*

Now we turn to FE discretizations where all finite elements are associated with one single spectral reference element $\mathcal{E}_{Sp}$ of the order p in each variable, and to the DD preconditioner of the form (7.16). Such preconditioner contains only three components — two preconditioner-solvers and

one prolongation operator from element boundaries inside finite elements. The component preconditioner-solvers are aimed at solving internal Dirichlet problems on finite elements and the global problem on the interface boundary. Such DD preconditioner can be slightly more efficient, more robust and more suitable for parallelization in comparison with the one considered in the previous subsection. However, discretizations by spectral elements are obviously less flexible for adaptive computations.

Theorem 8.14. *Let us assume that the hp discretization by the spectral finite elements fulfil the conditions of the generalized shape regularity. Further, let* $\boldsymbol{\mathcal{K}}$ *be the DD preconditioner of the form (7.16) and let the following conditions hold:*

a) there are valid the bounds $\mathrm{cond}\,[\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{A}_I] = \mathcal{O}(1)$ *and*

$$\mathrm{ops}\,[\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{v}_I] = \mathcal{O}(p^2(\log(1+p))^{\theta})\,, \quad \forall\, \mathbf{v}_I \in U_I;$$

b) the Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{it},B}$ *is defined in (7.15) and (7.9), with* $\boldsymbol{\mathcal{S}}_{B,\mathrm{ref}}$ *from (7.175) and* $\boldsymbol{\mathcal{S}}_{B,\mathrm{solv}} = \mathbb{D}_B := \lfloor \boldsymbol{\mathcal{S}}_B \rfloor_{\mathrm{diag}}$ *is the diagonal matrix with the main diagonal common with* $\boldsymbol{\mathcal{S}}_B$*,*

c) the prolongation operator $\boldsymbol{\mathcal{P}}_B$ *is defined by (7.10) with* $\boldsymbol{\mathcal{P}}_{B,\mathrm{ref}}$ *satisfying (7.6) or (7.7) and* $\mathrm{ops}\,[\boldsymbol{\mathcal{P}}_B\mathbf{v}_B] = \mathcal{O}(p^2\log^{\theta_1}(1+p))$ *for* $\forall\, \mathbf{v}_B \in U_B$*.*

Then, we have

$$\begin{aligned} &\mathrm{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \mathcal{O}(1)\,, \\ &\mathrm{ops}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}] = \mathcal{O}(\mathcal{R}p^2\log^{\theta+\theta_1}(1+p))\,, \quad \forall\, \mathbf{v} \in V\,. \end{aligned} \tag{8.144}$$

Proof. The proof is a direct consequence of the assumptions, Theorem 7.7 and Lemma 7.1. □

In Section 8.3, we have constructed preconditioner-solvers $\boldsymbol{\mathcal{A}}_I$ for the spectral reference elements satisfying the assumption *a*) with $\theta = 0$. Therefore, in order to obtain an optimal preconditioner-solver $\boldsymbol{\mathcal{K}}$ with respect to the arithmetic work, it is additionally necessary to use a prolongation operator $\boldsymbol{\mathcal{P}}_B$ from the inter-element boundary, satisfying the conditions c) with $\theta_1 = 0$.

Chapter 9

Nonoverlapping Dirichlet–Dirichlet DD Methods for *hp* Discretizations of 3d Elliptic Equations

We now consider the variational formulation of the Dirichlet problem from Example 2.1 in a 3d domain Ω: given some function $f \in L_2(\Omega)$, find the function $u \in \mathbb{V} = \mathring{H}^1(\Omega)$ such that the variational equation

$$a(u,v) = \langle f, v\rangle\,, \quad \forall\, v \in \mathbb{V}\,, \tag{9.1}$$

holds, with the bilinear and linear forms

$$a(u,v) = \int_\Omega \varrho(x)\,\nabla u(x)\cdot\nabla v(x)\,dx \quad \text{and} \quad \langle f, v\rangle = \int_\Omega f(x)v(x)\,dx\,, \tag{9.2}$$

respectively, see also variational formulation (2.4)–(2.5).

As usual, it is assumed that Ω coincides with the domain of the finite element assemblage. This is for the reasons that we do not bother with the approximation error of FE discretization and that the accuracy of approximation of the boundary and boundary condition does not influence the structure and efficiency of the DD method, except for special situations. At the same time, for sufficiently smooth domains, there are techniques to provide the approximation of the boundary and the boundary condition, in particular nonhomogeneous, with a required accuracy, see, *e.g.*, [Korneev (1970, 1979b,a)] and [Ciarlet (1978)].

In respect of the coefficient $\varrho(x)$, we distinguish two situations. One is when $\varrho(x)$ is uniformly positive and bounded, *i.e.,* there exist two positive constants $\underline{\mu}$ and $\overline{\mu}$ such that

$$0 < \underline{\mu} \le \varrho(x) \le \overline{\mu}\,, \quad x \in \Omega\,. \tag{9.3}$$

In the other case, the domain is the union (1.12) of nonoverlapping subdomains $\Omega_j,\ j = 1, 2, \ldots, J$, and

$$\varrho(x) = \varrho_j = \text{const} > 0\,, \quad \forall\, x \in \Omega_j\,, \quad j = 1, 2, \ldots, J\,, \tag{9.4}$$

i.e., in each subdomain, the coefficient ϱ is an arbitrary positive constant ϱ_j. Clearly, the latter situation models a little more general one when in each subdomain the coefficient ϱ is variable, but not too much. Namely,

$$0 < \underline{\mu}_j \le \varrho(x) \le \overline{\mu}_j\,, \quad \forall\, x \in \Omega_j\,, \quad j = 1, 2, \ldots, J\,,$$

with numbers $\underline{\mu}_j$ and $\overline{\mu}_j$, for which the relation $\overline{\mu}_j/\underline{\mu}_j$ is bounded uniformly in j and is sufficiently close to unity. This implies that, for each j, there exists the number ϱ_j satisfying

$$\underline{c}_\varrho\, \varrho_j \le \varrho(x) \le \overline{c}_\varrho\, \varrho_j\,, \quad \forall\, x \in \Omega_j\,, \quad j = 1, 2, \ldots, J\,,$$

with the positive constants $\underline{c}_\varrho$ and $\overline{c}_\varrho$ independent of j. In what follows, domains τ_r of finite elements are taken for subdomains of decomposition, and, therefore, we have $\tau_r = \Omega_r,\ r = 1, 2, \ldots, \mathcal{R} = J$. Such a strategy will be good, when the powers of polynomials $p = p_r$ are relatively high and do not vary much from one finite element to the other.

Let us remind that we always assume that the stiffness matrix of the reference element is generated by the Dirichlet integral over the domain of the reference element.

9.1 General Structure of DD and Schur Complement Preconditioners

The outline of this chapter is the following. We start with the discussion of the general structure of DD Dirichlet–Dirichlet preconditioner-solvers and the ideas behind it. Then we look at the main components of DD algorithms more attentively adding important details on their functioning and analyze contributions to the relative condition numbers of the DD preconditioners and to the overall numerical complexity of the DD solvers.

9.1.1 *Main Components of DD Preconditioners*

In 3d, assuming the ordering of d.o.f. adapted to DD, we usually operate with the block representation of FE stiffness matrix

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,B} \\ \mathbf{K}_{B,I} & \mathbf{K}_B \end{pmatrix} = \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,F} & \mathbf{K}_{I,W} \\ \mathbf{K}_{F,I} & \mathbf{K}_F & \mathbf{K}_{F,W} \\ \mathbf{K}_{W,I} & \mathbf{K}_{W,F} & \mathbf{K}_W \end{pmatrix} = \tag{9.5}$$

$$= \begin{pmatrix} \mathbf{K}_I & \mathbf{K}_{I,F} & \mathbf{K}_{I,E} & \mathbf{K}_{I,V} \\ \mathbf{K}_{F,I} & \mathbf{K}_F & \mathbf{K}_{F,E} & \mathbf{K}_{F,V} \\ \mathbf{K}_{E,I} & \mathbf{K}_{E,F} & \mathbf{K}_E & \mathbf{K}_{E,V} \\ \mathbf{K}_{V,I} & \mathbf{K}_{V,F} & \mathbf{K}_{V,E} & \mathbf{K}_V \end{pmatrix},$$

looking in the same way as (5.13). Since the subdomains of decomposition coincide with domains τ_r of finite elements, the indices in the above representations correspond to the sets of d.o.f. living on the following subsets of the FE discretization:

I – interiors of finite elements,
B – inter-element boundary,
F – interiors of faces of finite elements,
W – wire basket of FE mesh,
E – interiors of edges of finite elements, and
V – vertices.

The matrix $\mathbf{K}_I = \operatorname{diag}[\mathbf{K}_{I_r}]_{r=1}^{\mathcal{R}}$ is again block diagonal with each block $\mathbf{K}_{I_r}$ corresponding to the internal d.o.f. of the finite element τ_r.

This structure of the FE matrix equally fits two families of the FE discretizations, on which we concentrate in this chapter. These discretizations are obtained by means of the hierarchical and spectral cubic reference elements. By efforts of many researches spanning several decades, rather efficient DD Dirichlet–Dirichlet preconditioner-solvers have been developed for both families. We will consider those, which in general have five major components and are naturally realized in the five major modules of the computer codes. Three of them are the preconditioner-solvers for the subsystems of the algebraic equations to be solved at each DD iteration. They include the subsystems for the local internal Dirichlet problems on finite elements, for the local internal problems on finite element faces, and for the global problem, related to the wire basket of the FE grid. The major components include also two types of prolongations: the prolongations inside elements from finite element boundaries and the prolongations inside faces from their boundaries (edges and vertices). In the case of the elementwise

constant coefficient ϱ, it enters the matrices of the systems to be solved by the aforementioned two local solvers only through multipliers before these matrices. As a consequence, their computational properties are not influenced by jumps of ϱ. In this case the prolongations do not depend on ϱ as well. Only the wire basket problem is a single global problem of the algorithm and essentially depends on the coefficient ϱ and its jumps on the inter-element boundary. However, this component has much smaller dimension in the comparison with others.

In general, the procedure, defining the inverse to the DD preconditioner-solver $\boldsymbol{\mathcal{K}}$ for the FE stiffness matrix $\mathbf{K}$, can be expressed by the formulas

$$\boldsymbol{\mathcal{K}}^{-1} = \boldsymbol{\mathcal{K}}_I^{+} + \mathbf{P}_{V_B \to V} \boldsymbol{\mathcal{S}}_B^{-1} \mathbf{P}_{V_B \to V}^{\top} , \tag{9.6}$$

with

$$\boldsymbol{\mathcal{S}}_B^{-1} = \boldsymbol{\mathcal{S}}_F^{+} + \mathbf{P}_{V_W \to V_B} (\boldsymbol{\mathcal{S}}_W^{F})^{-1} \mathbf{P}_{V_W \to V_B}^{\top} . \tag{9.7}$$

The procedure, represented by the inverse to the wire basket Schur complement preconditioner-solver $\boldsymbol{\mathcal{S}}_W^F$, assumes a special algorithm as well and can be subjected to some additional decomposition. In (9.6) and (9.7), the notations $\boldsymbol{\mathcal{K}}_I^{+}$ and $\boldsymbol{\mathcal{S}}_F^{+}$ stand for the procedures realizing actions of the pseudo-inverses to the preconditioners $\boldsymbol{\mathcal{K}}_I$ and $\boldsymbol{\mathcal{S}}_F$ for the internal Dirichlet problems on finite elements and the internal problems on faces, respectively. Expressions (9.6) and (9.7) expose the three upper hierarchical levels of the algorithm, which are related to finite elements, faces and wire basket, respectively. It is important that procedures $\boldsymbol{\mathcal{K}}_I^{-1}\mathbf{v}_I$, $\mathbf{P}_{V_B \to V}\mathbf{v}_B$, $\mathbf{P}_{V_B \to V}^{\top}\mathbf{v}$ can be performed elementwise while the procedures $\boldsymbol{\mathcal{S}}_F^{-1}\mathbf{v}_F, \mathbf{P}_{V_W \to V_B}\mathbf{v}_W$ and $\mathbf{P}_{V_W \to V_B}^{\top}\mathbf{v}_B$ allow facewise completition. Thus, DD solver suggests a natural way to the parallelization of computations.

The DD preconditioner $\boldsymbol{\mathcal{K}}$ in (9.6) and (9.7) is the analogue of the finite element stiffness matrix factorized according to the exact substructuring elimination algorithm. This factorization is

$$\mathbf{K} = \mathbf{C} \begin{pmatrix} \mathbf{K}_I & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_B \end{pmatrix} \mathbf{C}^{\top} , \tag{9.8}$$

with

$$\mathbf{S}_B = \begin{pmatrix} \mathbf{S}_F & \mathbf{S}_{FW} \\ \mathbf{S}_{WF} & \mathbf{S}_W \end{pmatrix} = \mathbf{C}_B \begin{pmatrix} \mathbf{S}_F & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_W^F \end{pmatrix} \mathbf{C}_B^{\top} , \tag{9.9}$$

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K}_{BI}\mathbf{K}_I^{-1} & \mathbf{I} \end{pmatrix} , \quad \mathbf{C}_B = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{S}_{WF}\mathbf{S}_F^{-1} & \mathbf{I} \end{pmatrix} ,$$

and the Schur complement matrices

$$\mathbf{S}_B = \mathbf{K}_B - \mathbf{K}_{BI}\mathbf{K}_I^{-1}\mathbf{K}_{IB} \quad \text{and} \quad \mathbf{S}_W^F = \mathbf{S}_W - \mathbf{S}_{WF}\mathbf{S}_F^{-1}\mathbf{S}_{FW}\,. \tag{9.10}$$

The factorization of the inverse matrix $\mathbf{K}^{-1}$, corresponding to (9.8), is given by the formulas

$$\mathbf{K}^{-1} = \mathbf{C}^{-1}\begin{pmatrix}\mathbf{K}_I^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_B^{-1}\end{pmatrix}\mathbf{C}^{-\top}\,, \tag{9.11}$$

where

$$\mathbf{S}_B = \mathbf{C}_B^{-1}\begin{pmatrix}\mathbf{S}_F^{-1} & \mathbf{0}\\ \mathbf{0} & (\mathbf{S}_W^F)^{-1}\end{pmatrix}\mathbf{C}_B^{-\top}\,, \tag{9.12}$$

and

$$\mathbf{C}^{-1} = \begin{pmatrix}\mathbf{I} & -\mathbf{K}_I^{-1}\mathbf{K}_{IB}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\,, \quad \mathbf{C}_B^{-1} = \begin{pmatrix}\mathbf{I} & -\mathbf{S}_F^{-1}\mathbf{S}_{FW}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\,.$$

If we introduce the prolongation matrices

$$\mathbf{P}_{V_B\to V} = \begin{pmatrix}-\mathbf{K}_I^{-1}\mathbf{K}_{IB}\\ \mathbf{I}\end{pmatrix}\,, \qquad \mathbf{P}_{V_W\to V_B} = \begin{pmatrix}-\mathbf{S}_F^{-1}\mathbf{S}_{FW}\\ \mathbf{I}\end{pmatrix}\,, \tag{9.13}$$

the inverse to $\mathbf{K}$, expressed by (9.11) and (9.12), can also be written in the similar to (9.6), (9.7) way

$$\mathbf{K}^{-1} = \mathbf{K}_I^{+} + \mathbf{P}_{V_B\to V}\mathbf{S}_B^{-1}\mathbf{P}_{V_B\to V}^{\top}\,, \tag{9.14}$$

with

$$\mathbf{S}_B^{-1} = \mathbf{S}_F^{+} + \mathbf{P}_{V_W\to V_B}(\mathbf{S}_W^F)^{-1}\mathbf{P}_{V_W\to V_B}^{\top}\,. \tag{9.15}$$

Since the global DD preconditioner is usually created by means of the respective preconditioners for the reference element stiffness matrix, we turn to the latter. In the compliance with (9.5), the stiffness matrix of the 3d reference p-element can be written in the block forms

$$\begin{aligned}\mathbf{A} = \begin{pmatrix}\mathbf{A}_I & \mathbf{A}_{I,B}\\ \mathbf{A}_{B,I} & \mathbf{A}_B\end{pmatrix} = \begin{pmatrix}\mathbf{A}_I & \mathbf{A}_{I,F} & \mathbf{A}_{I,W}\\ \mathbf{A}_{F,I} & \mathbf{A}_F & \mathbf{A}_{F,W}\\ \mathbf{A}_{W,I} & \mathbf{A}_{W,F} & \mathbf{A}_W\end{pmatrix} =\\ = \begin{pmatrix}\mathbf{A}_I & \mathbf{A}_{I,F} & \mathbf{A}_{I,E} & \mathbf{A}_{I,V}\\ \mathbf{A}_{F,I} & \mathbf{A}_F & \mathbf{A}_{F,E} & \mathbf{A}_{F,V}\\ \mathbf{A}_{E,I} & \mathbf{A}_{E,F} & \mathbf{A}_E & \mathbf{A}_{E,V}\\ \mathbf{A}_{V,I} & \mathbf{A}_{V,F} & \mathbf{A}_{V,E} & \mathbf{A}_V\end{pmatrix}\,.\end{aligned} \tag{9.16}$$

Accordingly, the factorizations of the same type as (9.8), (9.9) can be introduced

$$\mathbf{A} = \mathbb{C} \begin{pmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_B \end{pmatrix} \mathbb{C}^\top , \quad \mathbb{C} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{BI}\mathbf{A}_I^{-1} & \mathbf{I} \end{pmatrix} , \tag{9.17}$$

$$\mathbb{S}_B = \begin{pmatrix} \mathbb{S}_F & \mathbb{S}_{FW} \\ \mathbb{S}_{WF} & \mathbb{S}_W \end{pmatrix} = \mathbb{C}_B \begin{pmatrix} \mathbb{S}_F & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_W^F \end{pmatrix} \mathbb{C}_B^\top , \quad \mathbb{C}_B = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbb{S}_{WF}\mathbb{S}_F^{-1} & \mathbf{I} \end{pmatrix} , \tag{9.18}$$

with the Schur complements

$$\mathbb{S}_B = \mathbf{A}_B - \mathbf{A}_{BI}\mathbf{A}_I^{-1}\mathbf{A}_{IB} \quad \text{and} \quad \mathbb{S}_W^F = \mathbb{S}_W - \mathbb{S}_{WF}\mathbb{S}_F^{-1}\mathbb{S}_{FW} . \tag{9.19}$$

In the rest of this section, we briefly comment on the preconditioner-solvers $\boldsymbol{\mathcal{K}}_I^+$, $\boldsymbol{\mathcal{S}}_B^{-1}$, $\boldsymbol{\mathcal{S}}_F^+$ $(\boldsymbol{\mathcal{S}}_W^F)^{-1}$, the prolongation matrices $\mathbf{P}_{V_B \to V}$, $\mathbf{P}_{V_W \to V_B}$ and on their relations to the corresponding matrices of the reference element and their preconditioners. While doing this, we will also fix notations for the main objects which are considered in the sequel.

1) *The preconditioner-solver for the local Dirichlet problems* has a typical block diagonal form, so that

$$\boldsymbol{\mathcal{K}}_I^+ := \begin{pmatrix} \boldsymbol{\mathcal{K}}_I^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} , \quad \text{with} \quad \boldsymbol{\mathcal{K}}_I = \operatorname{diag}[\,\boldsymbol{\mathcal{K}}_{I_1}, \boldsymbol{\mathcal{K}}_{I_2}, \ldots, \boldsymbol{\mathcal{K}}_{I_\mathcal{R}}\,] . \tag{9.20}$$

The conditions of generalized angular quasiuniformity allow us to choose $\boldsymbol{\mathcal{K}}_{I_r} = h_r \varrho_r \boldsymbol{\mathcal{A}}_I$ for all $r = 1, 2, \ldots, \mathcal{R}$, where $\boldsymbol{\mathcal{A}}_I$ is some preconditioner-solver for the reference element internal stiffness matrix $\mathbf{A}_I$.

2) *The preconditioner-solvers for the Schur complements* $\mathbf{S}_B$, $\mathbf{S}_W^F$ are denoted $\boldsymbol{\mathcal{S}}_B$, $\boldsymbol{\mathcal{S}}_W^F$, respectively. Similarly to (9.20), but now with some damage to the relative condition number, the preconditioner-solver $\boldsymbol{\mathcal{S}}_F$ may be adopted in the block diagonal form

$$\boldsymbol{\mathcal{S}}_F = \operatorname{diag}[\boldsymbol{\mathcal{S}}_{F^1}, \boldsymbol{\mathcal{S}}_{F^2}, \ldots, \boldsymbol{\mathcal{S}}_{F^Q}] , \tag{9.21}$$

where Q is the number of the FE faces inside the computational domain Ω. Each block $\boldsymbol{\mathcal{S}}_{F^k}$ corresponds to the degrees of freedom of one face F^k.

Decoupling faces in the preconditioner $\boldsymbol{\mathcal{S}}_F$ worsen the relative condition numbers

$$\operatorname{cond}[\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{S}_B] \quad \text{and} \quad \operatorname{cond}[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] ,$$

but not significantly. For rather general, but sufficiently regular discretizations, the growth of these numbers is characterized by the multiplier $\mathcal{O}(\log^2 p)$. In Subsubsection 9.1.3.2, such bound is proved for the case of a single reference element used for generating all finite elements of the assemblage. In Subsubsection 9.1.3.3, we discuss a more general situation when

finite elements of the assemblage are associated with multiple reference elements providing more flexible approximations. The aforementioned bound was first proved by [Pavarino and Widlund (1996)] and independently by [Casarin (1997)]. We remind that the same cost is paid for decoupling edges in the interface Schur complement preconditioner in the 2d case.

Suppose, $\varrho \equiv 1$, the FE mesh is a uniform square mesh, and all finite elements are associated with a single reference element. Then all blocks $\mathbf{S}_{F^k}$ on the diagonal of $\mathbf{S}_F$ are identical up to the appropriate ordering of the unknowns on each face. Therefore, all face preconditioners $\boldsymbol{\mathcal{S}}_{F^k}$ may be well defined by one matrix.

Turning to a more general situation, we consider a face $\overline{F}^k = \overline{\tau}_{r'} \cap \overline{\tau}_{r''}$ common for two finite elements $\tau_{r'}$ and $\tau_{r''}$, whose characteristic sizes are $h_{r'}$ and $h_{r''}$, and $\varrho_{r'}$, $\varrho_{r''}$ are the mean values of ϱ on these finite elements. Let also the p-reference element on $\tau_0 = (-1,1)^3$ be complete and invariant of the numbers and directions of the axes and let $\boldsymbol{\mathcal{F}}_\diamond$ denote the preconditioner-solver for the one-face Schur complement of the reference element stiffness matrix. Then, assuming the generalized conditions of the shape regularity are fulfilled, we arrive at the representation

$$\boldsymbol{\mathcal{S}}_{F^k} = (h_{r'}\varrho_{r'} + h_{r''}\varrho_{r''})\boldsymbol{\mathcal{F}}_\diamond\,, \quad k = 1, 2, \ldots, Q\,. \tag{9.22}$$

An alternative way is to define the face Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{F_r}$ for each finite element and use the assembling procedure for defining global preconditioner

$$\boldsymbol{\mathcal{S}}_F = \bigoplus_r \boldsymbol{\mathcal{S}}_{F_r}\,. \tag{9.23}$$

Let $\boldsymbol{\mathcal{F}}$ be the preconditioner for the reference element Schur complement $\mathbb{S}_F$. Then, for the appropriate order of the d.o.f., we have

$$\boldsymbol{\mathcal{S}}_{F_r} = h_r\varrho_r\boldsymbol{\mathcal{F}}\,. \tag{9.24}$$

In general, in the preconditioner $\boldsymbol{\mathcal{S}}_{F_r}$ and, therefore, in $\boldsymbol{\mathcal{S}}_F$, faces can be coupled.

The preconditioner $\boldsymbol{\mathcal{S}}_B$ (and/or $\boldsymbol{\mathcal{S}}_F$) can be implicitly defined by an inexact iterative solver for $\mathbf{S}_B$ (and/or $\mathbf{S}_F$). Sometimes, even at the iterative solution of the system with the Schur complement $\mathbf{S}_B$ as system matrix, $\mathbf{S}_B$ is not explicitly calculated: at each iteration, the multiplication by $\mathbf{S}_B$ is completed by performing operations in the right part of first relation (9.10) with the use of some fast solver for $\mathbf{K}_I$. Clearly, in this case the computational cost of the operation $\boldsymbol{\mathcal{S}}_B^{-1}\mathbf{v}_B$ is not less, than the cost of the operation $\boldsymbol{\mathcal{K}}_I^{-1}\mathbf{v}_I$. In the DD solvers, as a rule Schur complements (9.10)

are also not calculated explicitly or implicitly. However, such algorithms require a fast preconditioner-solver for the block $\mathbf{S}_F$ or the blocks $\mathbf{S}_{F^k}$ of the interface Schur complement. The problem of contemplating such preconditioners for *hp* discretizations of 3d elliptic equations is much more difficult, than in 2d case, discussed in Section 7.5. Only recently examples of such face preconditioner-solvers appeared in the literature, one of which is presented in Subsection 9.3.2.

3) *The prolongation matrices*

$$\begin{aligned} \mathbf{P}_{V_B \to V} &: V_B \to V\,, \\ \mathbf{P}_{V_W \to V_B} &: V_W \to V_B \end{aligned}$$

can also be generated by means of a few standard matrices, if, *e.g.*, finite elements of the assemblage are associated with one reference element invariant of the numbering of the axes and their directions. In this case, we can set

$$\mathbf{P}_{V_B \to V}|_{\overline{\tau}_r} = \mathbb{P}_B : U_B \to U \quad \text{and} \quad \mathbf{P}_{V_W \to V_B}|_{\partial\tau_r} = \mathbb{P}_W : U_W \to U_B, \tag{9.25}$$

where $\mathbb{P}_B$ and $\mathbb{P}_W$ are the corresponding prolongation matrices for the reference element with the numbers of rows and columns matching those for the finite element on τ_r. In the case of the reference element $\mathcal{E}_H$, the prolongation $\mathbb{P}_B$ incorporates six identical (geometrically) prolongations from faces, twelve identical prolongations from edges and the prolongation from vertices of the reference element. Therefore, only three standard matrices, appropriately perturbed, are needed for defining prolongations $\mathbf{P}_{V_B \to V}|_{\overline{\tau}_r}$, $r = 1, 2 \ldots, \mathcal{R}$, for all finite elements. A similar block structure of $\mathbb{P}_W$ is related to edges and vertices.

4) *The wire basket preconditioner* $\boldsymbol{\mathcal{S}}^F_W$ is assembled from the matrices $\boldsymbol{\mathcal{S}}^F_{W_r} := h_r \varrho_r \boldsymbol{\mathcal{W}}$, *i.e.,*

$$\boldsymbol{\mathcal{S}}^F_W = \biguplus_r \boldsymbol{\mathcal{S}}^F_{W_r}, \tag{9.26}$$

where the standard $(12p-4)\times(12p-4)$ matrix $\boldsymbol{\mathcal{W}}$ is induced by a specially chosen bilinear form, which is defined on the wire basket of the reference element. The solution procedure for the system

$$\boldsymbol{\mathcal{S}}^F_W \mathbf{v}_W = \mathbf{f}_W \tag{9.27}$$

can be additionally somehow parallelized. However, for simplicity, we mainly consider system (9.27) as a single global problem of DD algorithm. One reason for this is the dimension of $\boldsymbol{\mathcal{S}}^F_W$, which is $\mathcal{O}(p\mathcal{R})$, and, therefore, is relatively small. As a consequence, the DD algorithm can retain

optimality in p, if even $\boldsymbol{\mathcal{S}}_W^F$ would have been denser and the systems (9.27) were solved by a direct elimination procedure. Indeed, for the spectral and for the hierarchical discretizations, special more efficient direct elimination wire basket solvers have been developed. Some of them are discussed in Subsection 9.1.3.

5. *Additional comments.* We summarize that the DD Dirichlet–Dirichlet preconditioner-solver $\boldsymbol{\mathcal{K}}$ can be defined by five standard matrices

$$\boldsymbol{\mathcal{A}}_I, \quad \boldsymbol{\mathcal{F}}_\diamond, \quad \boldsymbol{\mathcal{W}}, \quad \mathbb{P}_B, \quad \mathbb{P}_W, \tag{9.28}$$

induced by the reference element, geometric parameters h_r of finite elements, coefficients ϱ_r and the information on the connectivity of finite elements. It is also worth to underline, that the procedure $\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}$ does not require assembling any of the matrices, figuring in (9.6) and (9.7), except for $\boldsymbol{\mathcal{S}}_W^F$. Assembling and subassembling operations in DD iterations are performed only for some subvectors and resulting vector of the discrete solution, see Algorithm 9.1.

In adaptive discretizations, especially by hierarchical finite elements, each finite element τ_r can be image of a specific reference element with specific spaces $\mathcal{U} = \mathcal{U}_r$ and $U = U_r$ on the reference cube τ_0. Moreover, in every such reference element, the subspaces of face and edge FE functions $\mathcal{U}_{F_r^k}, \mathcal{U}_{E_r^l} \in \mathcal{U}_r$ and, respectively, $U_{F_r^k}, U_{E_r^l} \in U_r$ may be different for different faces F_r^k and edges E_r^l. Under the assumptions, discussed below, this does not change the general structure of DD algorithm, if, instead of (9.28), we use their own reference matrices for each structural subspace of degrees of freedom. In other words, for each element τ_r, we may have to use the matrices

$$\boldsymbol{\mathcal{A}}_{r,I}, \quad \boldsymbol{\mathcal{F}}_{r,k}, \quad \boldsymbol{\mathcal{W}}_r, \quad \mathbb{P}_{B_r}, \quad \mathbb{P}_{W_r}, \tag{9.29}$$

generated by means of the associated reference element. The matrix $\boldsymbol{\mathcal{F}}_r$, which is the face preconditioner for the respective reference element, can be defined as a block diagonal matrix

$$\boldsymbol{\mathcal{F}}_r = \operatorname{diag}\left[\boldsymbol{\mathcal{F}}_{r,1}, \boldsymbol{\mathcal{F}}_{r,2}, \ldots, \boldsymbol{\mathcal{F}}_{r,6}\right]$$

with six independent blocks, each corresponding to one face. Indeed, at a proper local ordering of unknowns of compatible discretizations, each block depends only on the global number of the face, and in compliance with this, we have

$$\boldsymbol{\mathcal{S}}_{F_k} = \left(h_{r'}\varrho_{r'}\boldsymbol{\mathcal{F}}_{r',l'} + h_{r''}\varrho_{r''}\boldsymbol{\mathcal{F}}_{r'',l''}\right), \tag{9.30}$$

where

$$\overline{F}^k = \overline{F}^{l'}_{r'} = \overline{F}^{l''}_{r''} = \overline{\tau}_{r'} \cap \overline{\tau}_{r''}\,, \quad k = 1, 2, \ldots, Q\,,$$

and r', r'' are numbers of the finite elements having a common face F^k. It is assumed that the face F^k is the image of the two faces with local numbers l' and l'' in the two associated with $\tau_{r'}$ and $\tau_{r''}$ reference elements, respectively.

As was noted in [Korneev *et al.* (2003a)], under sufficiently mild conditions, adaptive features of *hp* discretizations do not worsen much the order of the computational complexity of DD solvers in respect to the total number of unknowns, when compared with the discretizations induced by a single complete reference p-element.

9.1.2 *Abstract Bounds of Relative Spectrum of DD Preconditioners*

The quality of a DD preconditioner is predetermined by the quality of the reference element preconditioners and prolongation operators (9.28), or (9.29), or other and the generalized conditions of shape regularity. The influence of the last factor can be improved by the change of the finite element mesh or mappings from the reference elements, *i.e.*, by the change of the discretization suitable for the particular elliptic problem. For estimating the impact of the rest factors, we start from the bounds of the relative spectrum of the DD preconditioner in general form.

Lemma 9.1. *Let the reference element preconditioners $\boldsymbol{\mathcal{A}}_I$, $\boldsymbol{\mathcal{F}}$, $\boldsymbol{\mathcal{W}}$ and the prolongation matrices $\mathbb{P}_B$, $\mathbb{P}_W$, used for generating the ensemble of finite p-elements, satisfy the inequalities*

$$\underline{\gamma}_I \boldsymbol{\mathcal{A}}_I \le \mathbf{A}_I \le \overline{\gamma}_I \boldsymbol{\mathcal{A}}_I,\ \ \underline{\gamma}_F \boldsymbol{\mathcal{F}} \le \mathbb{S}_F \le \overline{\gamma}_F \boldsymbol{\mathcal{F}},\ \ \underline{\gamma}_W \boldsymbol{\mathcal{W}} \le \mathbb{S}^F_W \le \overline{\gamma}_W \boldsymbol{\mathcal{W}}, \tag{9.31}$$

and

$$\|\mathbb{P}_B \mathbf{v}_B\|^2_{\mathbf{A}} \prec \beta_B \|\mathbf{v}_B\|^2_{\mathbb{S}_B}\,, \quad \|\mathbb{P}_W \mathbf{v}_W\|^2_{\mathbb{S}_B} \prec \beta_W \|\mathbf{v}_W\|^2_{\mathbb{S}^F_W} \tag{9.32}$$

for some positive $\underline{\gamma}_L, \overline{\gamma}_L$, $L = I, F, W$, and β_B, β_W independent of r. Then, for

$$\begin{aligned}
\underline{\gamma} &\ge (1 - \sqrt{1 - \beta_B^{-1}}) \min\left(\underline{\gamma}_I, (1 - \sqrt{1 - \beta_W^{-1}}) \min(\underline{\gamma}_F, \underline{\gamma}_W)\right) \\
&\ge \frac{1}{2\beta_B} \min\left(\underline{\gamma}_I, \frac{\underline{\gamma}_F}{2\beta_W}, \frac{\underline{\gamma}_W}{2\beta_W}\right)\,, \\
\overline{\gamma} &\le (1 + \sqrt{1 - \beta_B^{-1}}) \max\left(\overline{\gamma}_I, (1 + \sqrt{1 - \beta_W^{-1}}) \max(\overline{\gamma}_F, \overline{\gamma}_W)\right) \\
&\le 4 \max(\overline{\gamma}_I, \overline{\gamma}_F, \overline{\gamma}_W)\,,
\end{aligned} \tag{9.33}$$

and positive $\underline{c},\overline{c}$*, depending only on the constants in the generalized conditions of shape regularity, the DD preconditioner satisfies the inequalities*

$$\underline{c}\,\underline{\gamma}\,\boldsymbol{\mathcal{K}} \le \mathbf{K} \le \overline{c}\,\overline{\gamma}\,\boldsymbol{\mathcal{K}}\,. \tag{9.34}$$

Proof. We note that the matrices entering (9.31) and (9.32) can depend on r. However for simplicity, we omitted index r in their notations. Lemma is proved by the application of Corollary 4.1 two times, in accordance with the representation (9.6) and (9.7) of the DD preconditioner $\boldsymbol{\mathcal{K}}$, and by taking into account the generalized conditions of the shape regularity and assumptions (9.31)–(9.32) elementwise. An essential part of the proof is modelled by the proof of the bounds

$$\underline{\gamma}\,\boldsymbol{\mathcal{A}}_{DD} \le \mathbf{A} \le \overline{\gamma}\,\boldsymbol{\mathcal{A}}_{DD} \tag{9.35}$$

for the preconditioner with the pseudo-inverse of the form

$$\boldsymbol{\mathcal{A}}^{+}_{DD} = \boldsymbol{\mathcal{A}}^{+}_{I} + \mathbb{P}_B \widetilde{\boldsymbol{\mathcal{S}}}^{+}_{B} \mathbb{P}^{\top}_{B}\,, \tag{9.36}$$

where

$$\widetilde{\boldsymbol{\mathcal{S}}}^{+}_{B} = \boldsymbol{\mathcal{F}}^{+} + \mathbb{P}_W \boldsymbol{\mathcal{W}}^{+} \mathbb{P}^{\top}_{W}\,. \tag{9.37}$$

This preconditioner for the reference element stiffness matrix $\mathbf{A}$ is introduced by expressions (9.36) and (9.37)) similar to (9.6) and (9.7). After the application of Corollary 4.1 to the preconditioner $\widetilde{\boldsymbol{\mathcal{S}}}_B$ of (9.37) for the Schur complement $\mathbb{S}_B$, we get

$$\underline{\gamma}_B \widetilde{\boldsymbol{\mathcal{S}}}_B \le \mathbb{S}_B \le \overline{\gamma}_B \widetilde{\boldsymbol{\mathcal{S}}}_B\,, \quad \text{with}$$

$$\underline{\gamma}_B \ge (1 - \sqrt{1 - \beta_W^{-1}}) \min(\underline{\gamma}_F, \underline{\gamma}_W) \ge \tfrac{1}{2\beta_W} \min(\underline{\gamma}_F, \underline{\gamma}_W)\,, \tag{9.38}$$

$$\overline{\gamma}_B \le (1 + \sqrt{1 - \beta_W^{-1}}) \max(\overline{\gamma}_F, \overline{\gamma}_W) \le 2 \max(\overline{\gamma}_F, \overline{\gamma}_W)\,.$$

If to take into account the bounds above and assumptions of lemma on the matrices $\boldsymbol{\mathcal{A}}_I$ and $\mathbb{P}_B$, the application of Corollary 4.1 to the preconditioner $\boldsymbol{\mathcal{A}}_{DD}$ of (9.36) results in the bounds (9.35) with $\underline{\gamma}$ and $\overline{\gamma}$ given in (9.33).□

Remark 9.1. Clearly, the estimates (9.34), (9.33) hold for adaptive discretizations by the elements associated with cubic or tetrahedral reference elements. In this case, as was mentioned above, the matrices entering (9.31) and (9.32) depend on r.

In what follows, it will be shown that the standard preconditioners and prolongations (9.28) can be chosen rather efficiently. At the proper choice, they will provide the computational cost of the operation $\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}$, which is optimal up to the multiplier $(1 + \log p)$.

9.1.3 *Interface Boundary Schur Complement Preconditioning*

The choice of the interface Schur complement preconditioner is crucial for the efficiency of the Schur complement algorithms as well as DD Dirichlet–Dirichlet algorithms. The arrangements of such preconditioners for the hierarchical and the spectral *hp* discretizations have much in common with *h* discretizations, but specific forms of the underlying element preconditioners are more complex, especially in 3d. As a rule in spite of having smaller by the order of p dimension, than the FE system, the interface problem still remains rather complex. To coup with the complexity, an additional decomposition is usually introduced. In 2d, the interface problem is well decomposed in the edge and the vertex components, while in 3d their counterparts are the face and the so called wire basket components. The development and analysis of the *wire basket perconditioning* for the *h*-version is well reflected in the fundamental papers of [Bramble *et al.* (1986, 1987, 1988, 1989)] and [Smith (1991, 1993)], see also [Smith *et al.* (1996)], [Pavarino and Widlund (1996, 1999a,b)] and [Casarin (1997)] for spectral discretizations. A version of this approach for the hierarchical *hp* discretizations was discussed in [Korneev *et al.* (2003b,a)]. In spite of these contributions, for some time the important question about fast solvers for the face subproblem in the interface Schur complement preconditioner remained open. For *hp* discretizations by the finite elements, associated with the cubic hierarchical and spectral reference elements, fast preconditioner-solvers for the face subproblems, formulated in the spaces $H_{00}^{1/2}(F^k)$, were suggested in [Korneev *et al.* (2003b,a)] and [Korneev and Rytov (2005a, 2008)]. These suggestions bear on the solver of [Beuchler *et al.* (2004)] for the 2d discrete Dirichlet problem in $\mathring{H}^1(\tau_0)$ governed by the reference element $\mathcal{E}_H$ internal stiffness matrix. By means of the K-interpolation technique, it was transformed into the solver for the face problem on a face of the 3d reference element $\mathcal{E}_H$. A more involved path led to the fast face solver for the spectral reference element. The interface preconditioners, which we consider in this subsection, are based on the results of the cited works.

9.1.3.1 *An Outline of the Approach*

The preconditioner $\boldsymbol{\mathcal{S}}_B$ for the global interface Schur complement $\mathbf{S}_B$, see (9.10), in the factorization (9.8) is assembled from Schur complement preconditioners $\boldsymbol{\mathcal{S}}_{B_r}$ for each finite element. The derivation of the

preconditioners $\boldsymbol{\mathcal{S}}_{B_r}$ is mimicked by and includes the Schur complement preconditioning for the reference element, which is considered below.

The factorization (9.9) for the reference element stiffness matrix is

$$\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{A}_{BI}\mathbf{A}_I^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_I & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_B \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{A}_I^{-1}\mathbf{A}_{IB} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{9.39}$$

with the Schur complement $\mathbb{S}_B = \mathbf{A}_B - \mathbf{A}_{BI}\mathbf{A}_I^{-1}\mathbf{A}_{IB}$ having the block form

$$\mathbb{S}_B = \begin{pmatrix} \mathbb{S}_F & \mathbb{S}_{FW} \\ \mathbb{S}_{WF} & \mathbb{S}_W \end{pmatrix}.$$

The following three steps, which can be found, *e.g.*, in [Pavarino and Widlund (1999a)], are sometimes used for the interpretation of the approach to the Schur complement preconditioning:

i) Changing the basis in the subspace V_W by means of the transformation matrix

$$\boldsymbol{\mathcal{C}}_B = \begin{pmatrix} \mathbf{I} & -\boldsymbol{\mathcal{C}}_{FW} \\ \mathbf{0} & \mathbf{I} \end{pmatrix},$$

so that the Schur complement becomes the matrix

$$\mathbb{S}_{\mathbb{C}} = \boldsymbol{\mathcal{C}}_B^T \mathbb{S}_B \boldsymbol{\mathcal{C}}_B = \begin{pmatrix} \mathbb{S}_F & \text{nonzero} \\ \text{nonzero} & \widetilde{\mathbb{S}}_W \end{pmatrix}.$$

ii) The transformation matrix has to be chosen in such a way that, with minor losses in the relative condition number, it were possible to eliminate in $\mathbb{S}_C$ the coupling between faces and the wire basket and, additionally, the coupling between different faces. Having these couplings removed, one comes to the 2×2 block diagonal tentative preconditioner

$$\widetilde{\mathbb{S}}_{\mathbb{C}} = \begin{pmatrix} \widetilde{\mathbb{S}}_F & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbb{S}}_W \end{pmatrix}$$

for $\mathbb{S}_{\mathbb{C}}$. The first block is 6×6 block diagonal matrix $\widetilde{\mathbb{S}}_F = \text{diag}\,[\mathbb{S}_{F_1}, \mathbb{S}_{F_2}, \ldots, \mathbb{S}_{F_6}]$ with each block related to one of the six faces F_i. Obviously, with $\mathcal{U}^{(r)} = \mathcal{Q}_p$ for $p \in \mathbb{N}$, these blocks are identical up to appropriate ordering of the face unknowns. They can also be identical, if the space $\mathcal{U}^{(r)}$ and its basis are invariant at changes of numbers of the axes x_k and their directions.

iii) Replacing blocks $\widetilde{\mathbb{S}}_F$ and $\widetilde{\mathbb{S}}_W$ by some scaled and simpler matrices $\boldsymbol{\mathcal{F}}$ and $\boldsymbol{\mathcal{W}}$, of which the former is block diagonal, *i.e.*, $\boldsymbol{\mathcal{F}} = \text{diag}\,[\boldsymbol{\mathcal{F}}_1, \boldsymbol{\mathcal{F}}_2, \ldots, \boldsymbol{\mathcal{F}}_6]$.

Having completed these three steps, we get the Schur complement preconditioner

$$\widetilde{\boldsymbol{\mathcal{S}}}_B = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \boldsymbol{\mathcal{C}}_{FW}^T & \mathbf{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{F}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{W}} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \boldsymbol{\mathcal{C}}_{FW} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} . \tag{9.40}$$

The factorization (9.18) of the Schur complement

$$\mathbb{S}_B = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbb{S}_{WF}\mathbb{S}_F^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbb{S}_F & \mathbf{0} \\ \mathbf{0} & \mathbb{S}_W^F \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbb{S}_F^{-1}\mathbb{S}_{FW} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} , \tag{9.41}$$

with $\mathbb{S}_W^F = \mathbb{S}_W - \mathbb{S}_{WF}\mathbb{S}_F^{-1}\mathbb{S}_{FW}$, allows us to represent its pseudo-inverse as

$$\mathbb{S}_B^+ = \mathbb{S}_F^+ + \mathbb{P}_{WB}\ (\mathbb{S}_W^F)^+\ \mathbb{P}_{WB}^T ,$$

where

$$\mathbb{P}_{WB} = \begin{pmatrix} -\mathbb{S}_F^{-1}\mathbb{S}_{FW} \\ \mathbf{I} \end{pmatrix} . \tag{9.42}$$

Therefore, in order to split faces and wire basket without considerable losses in the relative condition number, it is necessary to choose the transformation matrix $\boldsymbol{\mathcal{C}}_B$ such that the prolongation operator

$$\mathbb{P}_{\mathrm{a}} = \begin{pmatrix} -\boldsymbol{\mathcal{C}}_{FW} \\ \mathbf{I} \end{pmatrix} , \tag{9.43}$$

is a sufficiently good approximation of the prolongation operator $\mathbb{P}_{WB} : U_W \to U_B$. It is also necessary that $\ker[\boldsymbol{\mathcal{W}}] = \ker[\mathbb{S}_W^F]$, where, evidently, $\ker[\mathbb{S}_W^F]$ contains constant vectors, which similarly to $\mathbb{P}_{WB}$ should be exactly reproduced by the operator $\mathbb{P}_{\mathrm{a}}$. If these basic properties are violated, than $\ker[\widetilde{\boldsymbol{\mathcal{S}}}_B] \neq \ker[\mathbb{S}_B]$, and, therefore, $\widetilde{\boldsymbol{\mathcal{S}}}_B$ cannot be a good preconditioner for $\mathbb{S}_B$.

The steps i)-iii) lead to the DD preconditioner of the same form, which is described in Subsection 9.1.1. However, they suggest an alternative of a different set of assumptions from used in Lemma 9.1, which also allow to bound the relative condition number.

Lemma 9.2. *Let $\mathbb{P}_B$ satisfy (9.32) and*

$$\underline{\gamma}_F \boldsymbol{\mathcal{F}} \le \mathbb{S}_{\mathbb{C}} , \quad \mathbb{S}_F \le \overline{\gamma}_F \boldsymbol{\mathcal{F}} , \quad \underline{\gamma}_W \boldsymbol{\mathcal{W}} \le \mathbb{S}_W^F , \quad \widetilde{\mathbb{S}}_W \le \widetilde{\gamma}_w \boldsymbol{\mathcal{W}} . \tag{9.44}$$

Then inequalities (9.34) hold with

$$\underline{\gamma} \ge \frac{1}{2\beta_B} \min\left(\underline{\gamma}_I , \frac{1}{\underline{\gamma}_F^{-1} + \underline{\gamma}_W^{-1}} \right) \quad \textit{and} \quad \overline{\gamma} \le 4 \max\left(\overline{\gamma}_I , \overline{\gamma}_F , \widetilde{\gamma}_W \right) . \tag{9.45}$$

Proof. The proof is the same as of Lemma 9.1, but instead of (9.38) we use (4.44). □

Let us note that the conditions imposed on the prolongation operator $\mathbb{P}_W$ are reflected in the last two inequalities (9.44).

Obviously, the described approach resembles in a way the one presented earlier for 2d. In the interface preconditioner for 2d discretizations, d.o.f. at the vertices are decoupled from those at the edges, and d.o.f. at different edges are also decoupled. In 3d, the wire basket d.o.f. are decoupled from those at the faces and different faces are also decoupled. Since for the tetrahedral and cubic reference elements, equipped with the complete polynomial spaces $\mathcal{P}_p$ and $\mathcal{Q}_p$, respectively, the prolongation Theorem 7.2 holds uniformly in p, we conclude that

$$|v_B|^2_{1/2,\partial\tau_0} \asymp \mathbf{v}_B^\top \mathbb{S}_B \mathbf{v}_B\,, \quad \forall\, \mathbf{v}_B \leftrightarrow v_B \in \mathcal{U}_B(\partial\tau_0)\,. \tag{9.46}$$

Thus, as in the h-version, the preconditioners for Schur complements of a similar to the presented above structure can be derived for finite elements of different types by means of the preconditioning the matrix of the quadratic form $|v_B|^2_{1/2,\partial\tau_0}$, $v_B \in \mathcal{U}_B(\partial\tau_0)$, instead of a direct preconditioning the Schur complement $\mathbb{S}_B$.

9.1.3.2 *Face Preconditioners*

We look first at the assemblage of finite elements associated with a single reference element, which has $\mathcal{Q}_p$ for the space spanned by the coordinate polynomials. Suppose $\mathbb{S}_{00}$ is the symmetric matrix of the typical one-face quadratic form ${}_{00}|v|^2_{1/2,F_l}$, $\forall v \in \mathcal{U}_{F_l}$. In this section, the simplified notation F_l is primarily used for a face F_0^l of the reference cube τ_0. If $\boldsymbol{\mathcal{S}}_{00}$ is the matrix close in the spectrum to $\mathbb{S}_{00}$, *i.e.,* satisfying the inequalities

$$\underline{\gamma}_0 \boldsymbol{\mathcal{S}}_{00} \le \mathbb{S}_{00} \le \overline{\gamma}_0 \boldsymbol{\mathcal{S}}_{00} \tag{9.47}$$

with sufficiently good $\overline{\gamma}_0, \underline{\gamma}_0 > 0$, then we can set

$$\boldsymbol{\mathcal{F}} = \mathrm{diag}\,[\boldsymbol{\mathcal{S}}_{00}, \boldsymbol{\mathcal{S}}_{00}, \ldots, \boldsymbol{\mathcal{S}}_{00}]\,, \tag{9.48}$$

assuming that the face d.o.f. are correspondingly ordered.

For approving such a choice, one can use the following representation of the $H^{1/2}(\partial\tau_0)$-seminorm:

$$|\,v\,|^2_{1/2,\partial\tau_0} = \int_{\partial\tau_0} \frac{(v(x)-v(y))^2}{|x-y|^3} ds(x) ds(y)\,. \tag{9.49}$$

The one face F_i seminorm $|\, v \,|_{1/2,F_i}$ is defined by the same expression (9.49), but with the integral taken over the face. For definiteness, it is accepted that F_i is in the plane $x_i \equiv -1$ for $i = 1, 2, 3$ and in the plane $x_{i-3} \equiv 1$ for $i = 4, 5, 6$. Therefore, we have

$$|\, v \,|^2_{1/2,F_i} = \int_{F_i} \frac{(v(x) - v(y))^2}{|x - y|^3} dx_k dx_l dy_k dy_l \tag{9.50}$$

with i, k, l or $i - 3, k, l$ all different for $i \le 3$ or $i \ge 4$, respectively.

One of the seminorms equivalent to (9.49) is written in the form

$$\begin{aligned} |\, v \,|^2_{1/2,\partial\tau_0} = &\sum_{i=1}^{6} |\, v \,|^2_{1/2,F_i} \\ &+ \sum_{k,l:k>l,k\neq l+3} \int_{-1}^{1} \int_{0}^{2} \Big(v|_{x_{\kappa(k)}=\theta(k),\, x_{\kappa(l)}=\theta(l)(1-t)} \\ &\qquad - v|_{x_{\kappa(k)}=\theta(k)(1-t),\, x_{\kappa(l)}=\theta(l)} \Big)^2 \frac{dt}{t} dx_i \,, \end{aligned} \tag{9.51}$$

where $i \neq k, l$ and $\kappa(n) = n - 3\,\mathrm{int}\,\lfloor (n-1)/3 \rfloor$, $\theta(n) = 2\,\mathrm{int}\,\lfloor (n-1)/3 \rfloor - 1$, for $n = 1, 2, \ldots, 6$. We refer the reader to [Lions and Magenes (1973)] and [Belgacem (1994)] for the proofs of the equivalence of the characterizations (9.49) and (9.51) of the seminorm $|\, v \,|_{1/2,\partial\tau_0}$.

We also need the norm ${}_{00}|\cdot|_{1/2,F_i}$, which can be defined by the formula

$${}_{00}|\, v \,|^2_{1/2,F_i} = |\, v \,|^2_{1/2,F_i} + \int_{F_i} \frac{u^2}{\mathrm{dist}((x_k, x_l), \partial F_i)} dx_k dx_l \,. \tag{9.52}$$

Lemma 9.3. *Let $\mathcal{U} = \mathcal{Q}_p$ and $\mathcal{U}_F(\partial\tau)$ be the subspace of traces on $\partial\tau_0$ of polynomials from $\mathcal{U}$ vanishing on edges. Then the inequalities*

$$\frac{1}{1 + \log^2 p} \sum_{i=1}^{6} {}_{00}|\, v \,|^2_{1/2,F_i} \prec |\, v \,|^2_{1/2,\partial\tau_0} \prec \sum_{i=1}^{6} {}_{00}|\, v \,|^2_{1/2,F_i} \tag{9.53}$$

hold for all $v \in \mathcal{U}_F(\partial\tau)$ uniformly in p.

Proof. The right inequality (9.53) is the consequence of characterizations (9.51) and (9.52) and Cauchy inequality. The left inequality (9.53) is equivalent to the inequality

$$\frac{1}{1 + \log^2 p}\, {}_{00}|v|^2_{1/2,F_i} \prec |v|^2_{1/2,\partial\tau_0} \,, \quad \forall\, v \in \mathcal{U}_F(\partial\tau) \,, \tag{9.54}$$

which can be derived by the analogy with the h-version. Let $\mathcal{H}_{\mathrm{sp},F}(\partial\tau_0)$ be the space of traces on $\partial\tau_0$ of functions from $\mathcal{H}_{\mathrm{sp}}(\tau_0)$ vanishing on the edges.

Due to Lemma 7.8, see also (7.146), and the K-interpolation method between the spaces $L_2(F_i)$, $\mathring{H}^1(F_i)$ and between the spaces $L_2(\partial\tau_0)$, $H^1(\partial\tau_0)$, we have

$$ {}_{00}|v|_{1/2,F_i}, \asymp {}_{00}|\hat{v}|_{1/2,F_i} \quad \text{and} \quad |v|_{1/2,\partial\tau_0} \asymp |\hat{v}|_{1/2,\partial\tau_0} $$

for any $v \in \mathcal{U}_F(\partial\tau)$ and $\hat{v} \in \mathcal{H}_{\mathrm{sp},F}(\partial\tau)$, coinciding at the GLL nodes. Therefore, the proof is reduced to the proof of the inequality

$$ \frac{1}{1+\log^2 p}\, {}_{00}|\hat{v}|^2_{1/2,F_i} \prec |\hat{v}|^2_{1/2,\partial\tau_0}\,, \quad \forall\, \hat{v} \in \mathcal{H}_{\mathrm{sp},F}(\partial\tau_0)\,. \tag{9.55} $$

The spectral mesh can be covered by the quasiuniform orthogonal mesh having mesh parameter $\hbar$ comparable with the smallest mesh size of the spectral mesh, which order is p^{-2}. Let $\mathcal{V}_{\hbar,F}(\partial\tau_0)$ be the space of continuous piecewise bilinear functions induced by this mesh and vanishing at the edges. For functions from this space, the similar to (9.55) inequality

$$ \frac{1}{1+\log^2 \hbar^{-1}}\, {}_{00}|\hat{v}|^2_{1/2,F_i} \prec |\hat{v}|^2_{1/2,\partial\tau_0}\,, \quad \forall\, \hat{v} \in \mathcal{V}_{\hbar,F}(\partial\tau_0)\,, \tag{9.56} $$

holds, see, *e.g.,* (5.50). Since $\mathcal{H}_{\mathrm{sp},F}(\partial\tau_0) \subset \mathcal{V}_{\hbar,F}(\partial\tau_0)$ and $\log\hbar^{-1} \asymp \log p^2 = 2\log p$, the inequality (9.55) holds as well. □

Due to the trace theorem and Cauchy inequality, the inequalities

$$ \frac{1}{1+\log^2 p}\, {}_{00}|v|^2_{1/2,F_i} \prec |v|^2_{1/2,F_i} \text{ and } \frac{1}{1+\log^2 p}\, {}_{00}|v|^2_{1/2,F_i} \prec |v|^2_{1,\tau_0}, \tag{9.57} $$

which hold for all $v \in \mathcal{U}_F(\tau_0)$, directly follow from (9.54).

For several types of cubic and tetrahedral reference elements, the proof can be found in a number of papers, with the earliest one given by [Pavarino and Widlund (1996)] for cubic elements, see Lemma 5.9. [Ainsworth and Guo (2000)] proved this inequality for tetrahedrons. Some results for the cubic p-reference elements, like Lemma 9.3 and Lemma 9.6 below, can be produced on the basis of the results of Chapter 5 for the h-version and the results of Chapter 7 on the stability of the polynomial interpolations over the nodes of the GLL mesh. Such an approach was introduced and used consistently by [Casarin (1997)].

Theorem 9.1. *Let a single reference element, hierarchical or spectral, with the coordinate polynomials spanning the space $\mathcal{Q}_p$ be used for the FE discretization. Let also $\boldsymbol{\mathcal{F}}_\diamond = \boldsymbol{\mathcal{S}}_{00}$ and the preconditioners $\boldsymbol{\mathcal{S}}_{F_k}$ for faces be defined by (9.22). Then we have*

$$ \frac{\underline{\gamma}_0}{(1+\log p)^2}\, \boldsymbol{\mathcal{F}} \prec \mathbb{S}_F \prec \overline{\gamma}_0 \boldsymbol{\mathcal{F}}\,, \quad \frac{\underline{c}_F \underline{\gamma}_0}{(1+\log p)^2}\, \boldsymbol{\mathcal{S}}_F \le \mathbf{S}_F \le \overline{c}_F \overline{\gamma}_0 \boldsymbol{\mathcal{S}}_F\,, \tag{9.58} $$

with the constants $\underline{c}_F$,$\overline{c}_F > 0$ depending only on the constants in the generalized conditions of the shape regularity.

Proof. The first pair of the inequalities are the key ones, since another pair is obtained by means of the assembling procedure and the scaling, which takes into account the generalized conditions of the shape regularity and local values of ϱ. Turning to the bound $\mathbb{S}_F \prec \overline{\gamma}_0 \boldsymbol{\mathcal{F}}$, we note that it is proved by the chain of the inequalities

$$\mathbf{v}^\top \mathbb{S}_F \mathbf{v} \prec |\, v \,|^2_{1/2,\partial\tau_0} \prec \sum_{k=1}^{6} {}_{00}|\, v \,|^2_{1/2,F_k} \asymp \sum_{k=1}^{6} \mathbf{v}_k^\top \mathbb{S}_{00} \mathbf{v}_k \prec \overline{\gamma}_0 \mathbf{v}^\top \boldsymbol{\mathcal{F}}\, \mathbf{v}\,, \quad (9.59)$$

which hold for all $\mathbf{v} \leftrightarrow v \in \mathcal{U}_F$, where the first step requires the proof of the respective prolongation theorem such as Theorem 7.2 or Theorem 9.2 below. The following three steps are clearly due to (9.53), the characterizations (9.51), (9.52) of the seminorms $|\, v \,|^2_{1/2,\partial\tau_0}$, ${}_{00}|\cdot|_{1/2,F_k}$, respectively, the definition of the preconditioner $\boldsymbol{\mathcal{F}}$, and assumption (9.47). The proof of the first inequality (9.58) is completed by the application of the trace theorem, the left bound (9.54) and the left inequality (9.47). □

Let us represent the matrix $\mathbb{S}_F$ in the block form $\mathbb{S}_F = \{\mathbb{S}_{k,l}\}_{k,l=1}^6$ with each block $\mathbb{S}_{k,k}$ corresponding to one face F_k of the reference element. At the appropriate ordering of the coordinate polynomials, all the blocks on diagonal are similar, *i.e.,* are equal to the same matrix, which we denote $\mathbb{S}_{F,0}$. Simultaneously with the first pair of inequalities (9.58), we have

$$\frac{1}{1+\log^2 p}\, \mathbb{S}_{00} \prec \mathbb{S}_{F,0} \prec \mathbb{S}_{00}\,.$$

Later, we will give examples of fast preconditioner-solvers for the matrices $\mathbb{S}_{00}$ for the p reference elements $\mathcal{E}_H, \mathcal{E}_L$.

Theorem 9.2. *Let $\mathcal{Q}_p(\partial\tau_0)$ be the space of traces of polynomials from the space $\mathcal{Q}_p$ on the boundary of the cube $\tau_0 = (-1,1)^3$. There exists the linear lifting operator $P : \mathcal{Q}_p(\partial\tau_0) \to \mathcal{Q}_p$ such that for any $\phi \in \mathcal{Q}_p(\partial\tau_0)$ the polynomial $v = P\phi$ satisfies*

$$|\, v \,|_{1,\tau_0} \le c_{\mathbb{P}} |\, \phi \,|_{1/2,\,\partial\tau_0} \quad (9.60)$$

with an absolute constant $c_{\mathbb{P}}$.

Proof. This theorem is a particular case of a more general result proven in [Belgacem (1994)]. □

9.1.3.3 *Face Preconditioners in p-Adaptive Computations*

In the adaptive *hp* discretizations, the pre-image space of polynomials on the corresponding reference face may be different for each face of the FE discretization. This means that different finite elements $\mathcal{E}_r$ may be associated

with different reference elements $\mathcal{E}_{H,r}$ or $\mathcal{E}_{L,r}$, which are specified on the same reference cube τ_0, but equipped with different spaces $\mathcal{U}^{(r)} \subseteq \mathcal{Q}_p$ subordinate to the conditions of compatibility. Every such space may have different spaces of traces $\mathcal{U}^{(r)}(F_i)$ on different faces F_i, $i = 1, 2, \ldots, 6$, of the cube τ_0. In such a situation, p_r denotes the *maximal power* in one variable of the coordinate polynomials of the reference element and $p = \max_r p_r$. Let $\mathbb{S}^{(r)}_{F_i,00}$ be the symmetric matrix of the quadratic form ${}_{00}|v|^2_{1/2,F_i}$, $\forall v \in \mathcal{U}^{(r)}(F_i)$, which is the square of the seminorm in the space $H^{1/2}_{00}(F_i)$. Under some conditions, similarly with the considered above case of a single complete reference element, efficient preconditioning may be achieved by setting

$$\boldsymbol{\mathcal{F}}^{(r)} = \operatorname{diag}\left[\boldsymbol{\mathcal{S}}^{(r)}_{F_1,00}, \boldsymbol{\mathcal{S}}^{(r)}_{F_2,00}, \ldots, \boldsymbol{\mathcal{S}}^{(r)}_{F_6,00}\right]$$

with $\boldsymbol{\mathcal{S}}^{(r)}_{F_i,00}$ being a matrix that is close in the spectrum [1] to the matrix $\mathbb{S}^{(r)}_{F_i,00}$, *i.e.*,

$$\underline{\gamma}_0 \boldsymbol{\mathcal{S}}^{(r)}_{F_i,00} \le \mathbb{S}^{(r)}_{F_i,00} \le \overline{\gamma}_0 \boldsymbol{\mathcal{S}}^{(r)}_{F_i,00} \tag{9.61}$$

with sufficiently good $\overline{\gamma}_0, \underline{\gamma}_0 > 0$ independent of i, r and $\mathcal{R}$. Let vector $\mathbf{p}_{I_r} = (p_{I_r,1}, p_{I_r,2}, p_{I_r,3})$ characterizes the powers of the internal polynomials in each variable on the reference cube and $p_{\,r,F}$ is the maximal power in each variable of 2d polynomials on the faces. A simple sufficient condition for providing that losses in the condition will not be significant is

$$\min\left(p_{I_r,1}, p_{I_r,2}, p_{I_r,3}\right) \ge p_{\,r,F}\,. \tag{9.62}$$

Lemma 9.4. *Let the condition (9.62) be fulfilled. Then inequalities*

$$\frac{1}{(1+\log p_{r,F})^2}\,\boldsymbol{\mathcal{F}}^{(r)} \prec \mathbb{S}^{(r)}_F \prec \boldsymbol{\mathcal{F}}^{(r)}\,,\ \frac{1}{(1+\log p_F)^2}\,\underline{c}_F\,\boldsymbol{\mathcal{S}}_F \le \mathbf{S}_F \le \overline{c}_F \boldsymbol{\mathcal{S}}_F\,, \tag{9.63}$$

hold, where $p_F = \max_r p_{r,F}$ *and* $\underline{c}_F$, $\overline{c}_F$ *are the positive constants depending on* $\underline{\gamma}_0, \overline{\gamma}_0$ *and the generalized conditions of the angular quasiuniformity.*

Proof. In the case under consideration, bounds of the type (9.53) take the form

$$\frac{1}{(1+\log p_{r,F})^2} \sum_{i=1}^{6} {}_{00}|\,v\,|^2_{1/2,F_i} \prec |\,v\,|^2_{1/2,\partial\tau_0} \prec \sum_{i=1}^{6} {}_{00}|\,v\,|^2_{1/2,F_i} \tag{9.64}$$

[1] Suppose, $\mathbf{A}$ and $\mathbf{B}$ are s.p.d. matrices depending on some parameter, *e.g.,* order of polynomials p, and the inequalities $\underline{\beta}\mathbf{B} \le \mathbf{A} \le \overline{\beta}\mathbf{B}$ hold with $\underline{\beta}, \overline{\beta} > 0$ not necessarily independent of p. We call $\mathbf{A}$ and $\mathbf{B}$ close in the spectrum, if $\overline{\beta}/\underline{\beta}$ grows with p slowly, *e.g.*, as $\mathcal{O}(\log^k p)$, $p = \text{const}$.

that hold for all $v \in \mathcal{U}_F^{(r)}$. The remainer of the proof is the same as in Theorem 9.1. □

9.1.3.4 *Numerical Polynomial Prolongations from Faces and the Interface Boundary for Adaptive Finite Elements*

We turn to the prolongation matrices $\mathbb{P}_B$, or $\mathbb{P}_{B_r}$ for a more general situation of discretizations with p-adaptive features. In this relation, it is worth to remind that estimate (9.60) of Theorem 9.2 and other estimates of this type, discussed, *e.g.,* in Subsection 7.2.2 for 2d, were proved with the use of specially designed prolongation operators in explicit form. Clearly, as it was done for 2d, some of them might be put in a convenient matrix form and used in DD algorithms for computer codes. At their use in DD algorithms, we will call such prolongation operators *analytic*, although in this case it is expected that the prolongations can be realized in sufficiently fast computational procedures. Discussions of the use of the analytic prolongations in DD algorithms for discretizations of 2d elliptic problems by quadrilateral and triangular finite elements of several types can be found in [Ivanov and Korneev (1995, 1996)], [Korneev and Jensen (1997)] and [Korneev *et al.* (2002a)]. In respect of the numerical complexity, such prolongations can be advantageous at least before the prolongations by means of direct elimination solvers for the corresponding local Dirichlet problems. However, in the 3d case, analytic prolongations became considerably more complex, see, *e.g.,* [Belgacem (1994)] and [Munoz-Sola (1997)]. At the same time, there are several suggestions on the computationally cheap polynomial prolongations — indeed almost optimal and optimal in the arithmetic operations count — which allow to guarantee (9.60), if it holds for the discrete-harmonic functions. In particular, in the case of 3d, such prolongations may be based on iterative solvers for the corresponding Dirichlet problem on the reference element. In relation with this, it is worth to remark that for a good convergence of a DD algorithm we need not (9.60), but the inequality (4.37), which for the case under consideration is written as

$$\|\mathbb{P}_B \mathbf{v}_B\|_{\mathbf{A}} \le c_{\mathbb{P}_B} \|\mathbf{v}_B\|_{\mathbb{S}_B} . \tag{9.65}$$

For polynomial spaces induced by some reference elements, such as incomplete, anisotropic and other specific types, the analytic prolongation operators providing (9.60) may not exist, see *e.g.,* [Bernardi *et al.* (1993)], whereas there are some which provide (9.65). In particular in many of such specific situations, typical for adaptive computations, prolongatioins

by inexact Dirichlet solvers may retain their efficiency in a sense of the inequality (9.65).

Often, for the prolongatioins by inexact Dirichlet solvers, it is reasonable to use two preconditioners, which we denote here $\mathbf{G}$ and $\boldsymbol{\mathcal{A}}_I$. They must be sufficiently close in the spectrum to the reference element stiffness matrix $\mathbf{A}$ and to the internal stiffness matrix $\mathbf{A}_I$, respectively, *i.e.,* for sufficiently good positive $\underline{\gamma}_G, \ldots, \overline{\beta}_I$, there should hold inequalities

$$\underline{\gamma}_G \mathbf{G} \le \mathbf{A} \le \overline{\gamma}_G \mathbf{G} \quad \text{and} \quad \underline{\beta}_I \boldsymbol{\mathcal{A}}_I \le \mathbf{G}_I \le \overline{\beta}_I \boldsymbol{\mathcal{A}}_I \,. \tag{9.66}$$

It is also assumed, firstly, that for the preconditioner $\mathbf{G}$ matrix-vector multiplications are cheap and, secondly, that there exists a fast solving procedure for the systems of equations with the matrix $\boldsymbol{\mathcal{A}}_I$. Then, for some fixed ν, we may set $\mathbf{G}_{I,\mathrm{it}} = \boldsymbol{\mathcal{I}}\,[\mathbf{G}_I, \boldsymbol{\mathcal{A}}_I, \nu_\epsilon]$ and

$$\mathbb{P}_B = \begin{pmatrix} \mathbb{P}_{I,B} \\ \mathbb{P}_{B,B} \end{pmatrix} = \begin{pmatrix} -\mathbf{G}_{I,\mathrm{it}}^{-1} \mathbf{G}_{I,B} \\ \mathbf{I} \end{pmatrix} .$$

It is important that $\mathbf{G}_I$ and $\mathbf{G}_{I,B}$ are the blocks of the same matrix. If, acting in a more direct way, we replace them by the corresponding blocks of the finite element stiffness matrix $\mathbf{K}^{(r)}$, the necessary number of iterations for a good prolongation will be the same in the order. However, the cost of the prolongation can be high due to the costly matrix-vector multiplications by $\mathbf{K}^{(r)}$. This, *e.g.*, takes place when coefficients of the elliptic equation are variable and the matrix $\mathbf{K}^{(r)}$ is completely filled in. Expediency of the implementation of two preconditioners for the use of inexact solvers for prolongations is typical for *hp* FE methods.

Efficiency of the iterative low energy prolongations are analyzed similarly to the 2d case, see Lemmas 7.4 and 4.1. At the same time some adjustments, including those allowing to avoid unnecessary restrictions on the adaptive discretizations, should be made. Also, the quality of the initial prolongation, which can be the cause of the main contribution to the computational cost, is specific for each particular case. At first, we use the assumptions that a cubic or a tetrahedral reference element is equipped with the complete polynomial space, *i.e.,* $\mathcal{Q}_p$ or $\mathcal{P}_p$, respectively, and inequality

$$\|v_B\|_{1,\tau_0} \asymp \|\mathbf{v}_B\|_{\mathbf{A}} \prec \phi(p)\, |v_B|_{1/2,\,\partial\tau_0} \tag{9.67}$$

holds with some function $\phi(p)$ for any vector $\mathbf{v}_B \in U_B$ and the corresponding function $v_B \in \mathcal{U}_B$, $v_B \leftrightarrow \mathbf{v}_B$,. From the trace and prolongation theorems it also follows that Schur complement $\mathbb{S}_B$ is equivalent to the matrix of the quadratic form $|v_B|^2_{1/2,\,\partial\tau_0}$ uniformly in p, and all these facts

result in the similar to (7.73), see p. 223, inequality

$$| u |_{1,\tau_0} \le [c_{\mathbb{P}} + \frac{\overline{\gamma}_G}{\underline{\gamma}_G}\rho^{\nu}(c_{\mathbb{P}} + \phi(p))]| u |_{1/2,\tau_0}\,, \tag{9.68}$$

where $\rho \le (1-\theta)/(1+\theta)$, $\theta = \sqrt{\underline{\beta}_I/\overline{\beta}_I}$, $c_{\mathbb{P}}$ is the absolute constant from the inequality (9.60), and u corresponds to the vector $\mathbf{u}$, which is the result of the prolongation $\mathbf{u} = \mathbb{P}_B\mathbf{u}_B$, $\forall\ \mathbf{u}_B \in U_B$. Clearly, the inequality (9.67) may hold only when the subspace $\mathcal{U}_B$ exactly represents constant functions, as, *e.g,* in the case of the hierarchical coordinate polynomials.

In the case of p-adaptive discretizations, *i.e.,* $\mathcal{U} = \mathcal{U}^{(r)}$, all the matrices entering (9.66) as well as the extension operator may depend on r, *i.e.,* $\mathbb{P}_B = \mathbb{P}_B^{(r)}$, and it is reasonable to replace (9.67) by the inequality

$$\|\mathbf{v}_B\|_{\mathbf{G}} \le \Theta(p)\,|v_B|_{1,\partial\tau_0}\,. \tag{9.69}$$

Obviously, in general Θ depends on the way of definition of the polynomials from $\mathcal{U}_B^{(r)}$ inside the reference element. Below, $\boldsymbol{\mathcal{G}}$ denotes the Schur complement of $\mathbf{G}$, and we remind that $\mathbb{S}_B$ is the Schur complement of the reference element matrix $\mathbf{A}$ and that mostly we do not attach index r to the values depending on it.

Lemma 9.5. *Let the domain τ_0 of the reference element be the unit cube or tetrahedron equipped with the polynomial space $\mathcal{U} = \mathcal{U}^{(r)}$, the bounds (9.66) and (9.69) hold, and u and u_I, u_B be the polynomials corresponding to the vector $\mathbf{u} = \mathbb{P}_B\mathbf{u}_B$ and its components $\mathbf{u}_I$, $\mathbf{u}_B$, respectively. Let also p_B be the maximal power (in each variable or total) of the 2d polynomials, which are the restrictions of the polynomials from the subspace $\mathcal{U}_B$ to the faces F_k. Then the inequality*

$$\|\mathbf{u}\|_{\mathbf{G}} \prec [1 + \rho^k(1 + p_B^2\Theta)]\,\|\mathbf{u}_B\|_{\boldsymbol{\mathcal{G}}}\,, \tag{9.70}$$

holds, where $\rho \le (1-\theta)/(1+\theta)$, $\theta = \sqrt{\underline{\beta}_I/\overline{\beta}_I}$, and the inequalities

$$\|\mathbf{u}\|_{\mathbf{G}} \prec \|\mathbf{u}_B\|_{\boldsymbol{\mathcal{G}}} \quad \text{and} \quad \|\mathbf{u}\|_{\mathbf{A}} \prec \sqrt{\frac{\overline{\gamma}_G}{\underline{\gamma}_G}}\|\mathbf{u}_B\|_{\mathbb{S}_B} \tag{9.71}$$

are attained for

$$\nu \succ \operatorname{int}\lfloor \frac{\log(1 + p_B^2\Theta)}{\log\rho^{-1}}\rfloor \succ \frac{1}{\theta}\,(\log p_B + \log\Theta)$$

iterations.

Proof. Let $\boldsymbol{\varphi}_I$ be the solution of the system $\mathbf{G}_I\boldsymbol{\varphi}_I = \mathbf{G}_{IB}\mathbf{u}_B$ and $\boldsymbol{\varphi} = \boldsymbol{\varphi}_I + \mathbf{u}_B$. The polynomial corresponding to $\boldsymbol{\varphi}$ is denoted by $\varphi \leftrightarrow \boldsymbol{\varphi}$. We have the convergence estimate

$$\|\mathbf{u}_I^k - \boldsymbol{\varphi}_I\|_{\mathbf{G}_I} \le \rho^k \|\boldsymbol{\varphi}_I\|_{\mathbf{G}_I},$$

from which, for $\mathbf{u}^k = \mathbf{u}_I^k + \mathbf{u}_B$, it follows

$$\|\mathbf{u}^k - \boldsymbol{\varphi}\|_{\mathbf{G}} \le \rho^k (\|\boldsymbol{\varphi}\|_{\mathbf{G}} + \|\mathbf{u}_B\|_{\mathbf{G}_B})$$

and

$$\|\mathbf{u}^k\|_{\mathbf{G}} \le \|\boldsymbol{\varphi}\|_{\mathbf{G}} + \rho^k (\|\boldsymbol{\varphi}\|_{\mathbf{G}} + \|\mathbf{u}_B\|_{\mathbf{G}_B}). \tag{9.72}$$

For $\mathbf{u}_B \leftrightarrow u_B \in \mathcal{U}_B$, inequality (9.69), the Markov type inequality and the trace theorem allow us to proceed as follows

$$\begin{aligned}\|\mathbf{u}_B\|_{\mathbf{G}_B} &\le \Theta\, |u_B|_{1,\partial\tau_0} \prec p_B^2 \Theta\, |u_B|_{1/2,\partial\tau_0} \\ &\prec p_B^2 \Theta \inf_{\psi_I \in \mathbb{V}_I} |u_B + \psi_I|_{1,\tau_0} \asymp p_B^2 \Theta\, \|\mathbf{u}_B\|_{\mathbb{S}_B},\end{aligned} \tag{9.73}$$

where $\mathbf{G}_B$ is the block of $\mathbf{G}$ related to the d.o.f. on the boundary.

Evidently, $\boldsymbol{\varphi}$ is discrete harmonic in a sense of the discrete operator $\mathbf{G}$. Therefore, we have

$$\|\boldsymbol{\varphi}\|_{\mathbf{G}} = \|\mathbf{u}_B\|_{\boldsymbol{\mathcal{G}}}.$$

Now, (9.72), (9.73) and the above equality imply

$$\|\mathbf{u}^k\|_{\mathbf{G}} \prec [1 + \rho^k (1 + p_B^2 \Theta)]\, \|\mathbf{u}_B\|_{\boldsymbol{\mathcal{G}}}.$$

Taking into account the first pair of inequalities (9.66), we get

$$\|\mathbf{u}^k\|_{\mathbf{A}} \le \sqrt{\frac{\overline{\gamma}_G}{\underline{\gamma}_G}} [1 + \rho^k (1 + p_B^2 \Theta)]\, \|\mathbf{u}_B\|_{\mathbb{S}_B}. \tag{9.74}$$

Now, we would like to find $k = \nu$ such that the term in brackets became not greater, than some $c^\circ > 1$. Therefore, ν must provide the inequality $\rho^\nu (1 + p_B^2 \Theta) \le c_1$ with $c_1 = c^\circ - 1$, from where it follows

$$\nu \ge \frac{\log(1 + p_B^2 \Theta) - \log c_1}{\log \rho^{-1}}.$$

The expression for ν in Lemma is obtained by setting $c_1 \ge 1$ and estimating $\log \rho^{-1}$ like $\log q^{-1}$ in (2.51) and (2.52). □

Note, that no restrictions on the structure of the space of polynomials $\mathcal{U} = \mathcal{U}^{(r)}$ on the reference element were imposed, except for (9.69). In particular, Lemma is applicable to the incomplete elements, for which estimates of the type (9.60) can't be proved, see, *e.g.,* [Bernardi *et al.* (1993)].

Let $\tau_0 = (-1,1)^3,\ \omega = \{\alpha = (\alpha_1,\alpha_2,\alpha_3) : 0 \le \alpha_1,\alpha_2,\alpha_3 \le p\}$,

$$\begin{aligned} &\mathcal{M}_p = \{L_\alpha(x) = \mathcal{L}_{\alpha_1}(x_1)\mathcal{L}_{\alpha_2}(x_2)\mathcal{L}_{\alpha_3}(x_3)\ ,\ \alpha \in \omega\}\,, \\ &\mathcal{M}_V = \{L_\alpha(x),\ \alpha_k = 0,1,\ k = 1,2,3\}\,, \end{aligned} \tag{9.75}$$

and, for the basis in the reference element space $\mathcal{U}$, a subset $\mathcal{M}_\star \subseteq \mathcal{M}_p$ be used which contains the subset $\mathcal{M}_V$. Then we can find that

$$\Theta \prec 1\,. \tag{9.76}$$

Clearly, (9.76) also holds for any cubic reference element with the same subspace $\mathcal{U}_B$, spanned by the boundary coordinate polynomials $L_\alpha(x)$. For the incomplete hierarchical reference element on the cube τ_0 with any basis $\mathcal{M}_\star$, the stiffness matrix $\mathbf{A}$ is sparse, and the matrix-vector multiplications by $\mathbf{A}$ are cheap. For instance, for the case $\mathcal{U} = \mathrm{span}[\mathcal{M}_p]$ with the basis $\mathcal{M}_p$, one has $\mathrm{ops}\,[\mathbf{Av}] = \mathcal{O}(p^3)$. Therefore, if $\boldsymbol{\mathcal{A}}_I$ is an efficient preconditioner-solver, such that for $\mathbf{G}_I = \mathbf{A}_I$ the numbers $\underline{\beta}_I, \overline{\beta}_I$ are constants, then an efficient component $\mathbb{P}_{I,B}$ of the prolongation matrix $\mathbb{P}_B$ can be defined by the expressions

$$\mathbb{P}_{I,B} := -\mathbf{A}_{I,\mathrm{it}}^{-1}\mathbf{A}_{I,B}\,, \quad \mathbf{A}_{I,\mathrm{it}} := \boldsymbol{\mathcal{I}}\,[\mathbf{A}_I, \boldsymbol{\mathcal{A}}_I, \nu]\,, \quad \nu = \mathcal{O}(1+\log p)\,. \tag{9.77}$$

In the case of the Lagrange and in particular spectral reference elements, for $p \ge 2$ the subspace $\mathcal{Q}_1$ does not belong to $\mathcal{U}_B$, and, as a consequence, the seminorms should be replaced by the norms in the right parts of (9.67) and (9.69). Besides, necessary adjustments should be made in the prolongation procedures with the aim to provide returning constant functions from their traces. Let $\mathbb{T}_{I,B}$ be a tentative prolongation operator, such that if $\mathbf{v}_B$ is a constant function on the boundary, then $\mathbb{T}_{I,B}\mathbf{v}_B$ represents the function equal to the same constant on $\overline{\tau}_0$. We introduce also the operator $\mathbf{T}_B : U_B \to U$, which for any $\mathbf{v}_B$ produces the vector $\mathbf{v} = \mathbf{T}_{I,B}\mathbf{v}_B$ representing the constant function equal to the average

$$\overline{v}_B = \frac{1}{24}\int_{\partial\tau_0} v_B ds\,, \quad v_B \in \mathcal{U}_B\,,\ v_B \leftrightarrow \mathbf{v}_B\,,$$

of the trace on $\partial\tau_0$ of the function, represented by the vector $\mathbf{v}_B$. Now, the

similar to (4.51) and (7.74) prolongation operators can be introduced:

$$\mathbb{P}_B = \begin{pmatrix} -\mathbf{A}_{I,\mathrm{it}}^{-1}\mathbf{A}_{I,B}(\mathbf{I} - \mathbf{T}_{B,B}) + \mathbf{T}_{I,B} \\ \mathbf{I} \end{pmatrix}, \qquad \mathbb{P}_B = \begin{pmatrix} -\mathbf{A}_{I,\mathrm{it}}^{-1}\left(\mathbf{A}_{I,B} + \mathbf{A}_I\mathbb{T}_{I,B}\right) + \mathbb{T}_{I,B} \\ \mathbf{I} \end{pmatrix}. \tag{9.78}$$

In general, *i.e.,* when $\mathcal{U}_B \nsubseteq \mathcal{Q}_1$, instead of (9.69), it is possible to prove

$$\|\mathbf{v}_B\|_{\mathbf{G}} \le \Theta \, \|v_B\|_{1,\partial\tau_0}, \quad \forall\ \mathbf{v}_B \in U_B. \tag{9.79}$$

Suppose, one uses the prolongation operators (9.78), and (9.79) holds. The bound of the number of iterations, sufficient for providing (9.71), looks similarly to the given in Lemma 9.5, and it is obtained without significant changes in the proof.

9.1.4 *Wire Basket Component*

The wire basket preconditioner $\boldsymbol{\mathcal{S}}_W^F$ and the prolongation operator $\mathbf{P}_{V_W \to V_B}$ from the wire basket on the interface compose the wire basket component of DD algorithms. In turn, according to the structure of DD preconditioner, presented in Subsection 9.1.1, they are assembled from the scaled matrices $\boldsymbol{\mathcal{W}}_r$ and $\mathbb{P}_{W_r}$, respectively, defined for the associated reference elements. Their design has much in common with the h-version and stems from the work of [Bramble *et al.* (1989)]. The adaptation to discretizations by spectral elements was provided by [Pavarino and Widlund (1996, 1999b)] and [Casarin (1997)], and to the hierarchical and hp adaptive discretizations by [Korneev *et al.* (2003a,b)].

We note that the dimension of the matrices $\mathbf{S}_W^F, \boldsymbol{\mathcal{S}}_W^F$ is $\mathcal{O}(\mathcal{R}p)$ and, therefore, this component does not cause much trouble even without optimization, if the problem and its discretzation are sufficiently regular. Indeed, at a small number of subdomains, the asymptotic numerical complexity of DD preconditioner-solver with the respect to p would not be increased by this component even at solution of systems with the matrix $\boldsymbol{\mathcal{S}}_W^F$ by the direct elimination procedure. Nevertheless, this preconditioner-solver requires attentive consideration, because it is a single global preconditioner in the DD solver, and at considerable jumps of the coefficient ϱ numerical solution can became not easy. The dimension $\dim\left[\mathbf{P}_{V_W \to V_B}\right] = \mathcal{O}(\mathcal{R}p) \times \mathcal{O}(\mathcal{R}p^2)$ of the prolongation operator from the wire basket into faces is larger and its construction requires more care. In

this subsubsection we present wire basket components as for hierarchical so spectral *hp* discretizations.

9.1.4.1 *Wire Basket Preconditioners*

Let W be the wire basket of the reference cube τ_0, $\mathbf{z}_1$ be the vector living on W and corresponding to the polynomial, which is identically equal to unity. We remind that by definition the seminorm in the space $L_2(W)$ is

$$|\, v \,|_{L_2(W)} \stackrel{\text{def}}{=} \inf_{c \in R} \| \, v - c \, \|_{L_2(W)} \, .$$

Similarly to the case of DD algorithms for h-version, see Subsection 5.1.5, it plays an important role in the construction of the wire basket preconditioner for *hp* discretizations. We have $\ker [\mathbb{S}_W^F] = \text{span}\, [\mathbf{z}_1]$, and, therefore, a good preconditioner $\boldsymbol{\mathcal{W}}$ for $\mathbb{S}_W^F$ should possess the same property. In general, a sufficiently good preconditioner $\boldsymbol{\mathcal{W}}$ may be defined as the matrix $\mathbf{W}$ of the quadratic form

$$b_W(v_W, v_W) := c_W \, |v_W|^2_{L_2(W)} \equiv \mathbf{v}_W^\top \mathbf{W} \, \mathbf{v}_W \, , \quad \mathbf{v}_W \leftrightarrow v_W, \tag{9.80}$$

where c_W is a scaling factor, or as the matrix which is sufficiently close in the spectrum to $\mathbf{W}$, *i.e.,* such that

$$\underline{\gamma}_W \boldsymbol{\mathcal{W}} \le \mathbf{W} \le \overline{\gamma}_W \, \boldsymbol{\mathcal{W}} \tag{9.81}$$

for sufficiently good positive $\underline{\gamma}_W$ and $\overline{\gamma}_W$. In the case of adaptive discretizations, we have $\mathbf{W} = \mathbf{W}_r$ and $\boldsymbol{\mathcal{W}} = \boldsymbol{\mathcal{W}}_r$, but $\underline{\gamma}_W$ and $\overline{\gamma}_W$ are assumed to be fixed, *i.e.*, independent of r.

Let us introduce the mass matrix $\mathbf{M}_{W_r}$ defined by the identity

$$\mathbf{v}^T \mathbf{M}_{W_r} \mathbf{v} = ||v||^2_{L_2(W)} \, , \quad \forall \, \mathbf{v} \leftrightarrow v \in \mathcal{U}_W^{(r)} \, ,$$

for the wire basket of the reference element associated with finite element τ_r. Then a simple computation gives

$$\begin{aligned} |v|^2_{L_2(W)} &= \int_W v^2 \, ds - \frac{(\int_W v \, ds)^2}{\int_W \, ds} \\ &= \mathbf{v}^T \left(\mathbf{M}_{W_r} - \frac{(\mathbf{M}_{W_r} \mathbf{z}_1)(\mathbf{M}_{W_r} \mathbf{z}_1)^T}{\mathbf{z}_1^T \mathbf{M}_{W_r} \mathbf{z}_1} \right) \mathbf{v} \, . \end{aligned} \tag{9.82}$$

Hence, we have

$$\mathbf{W}_r = c_W \left(\mathbf{M}_{W_r} - \frac{1}{24} (\mathbf{M}_{W_r} \mathbf{z}_1)(\mathbf{M}_{W_r} \mathbf{z}_1)^T \right) \, . \tag{9.83}$$

In the case of the trilinear vertex polynomials, the matrix

$$(\mathbf{M}_{W_r} \mathbf{z}_1)(\mathbf{M}_{W_r} \mathbf{z}_1)^T$$

does not depend on r, and has only one nonzero block related to the vertices with the entries equal to 9.

The wire basket preconditioner $\boldsymbol{\mathcal{S}}_W^F$ is assembled from the matrices $h_r \varrho_r \boldsymbol{\mathcal{W}}_r$. However, the algorithms of solving systems of equations with $\boldsymbol{\mathcal{S}}_W^F$ for the matrix, used in the practice and described below, do not require assembling it or the wire basket mass matrix.

Following the described outline, the wire basket preconditioners for cubic and tetrahedral elements can be defined similarly. Below, we only present them for the former. The justification of the choice of the bilinear form for the wire basket preconditioner, as usual, comes from the trace and prolongation theorems in the respective polynomial spaces.

Lemma 9.6. *Let the polynomial space $\mathcal{U} = \mathcal{Q}_p$ be specified on the reference cube τ_0, and for any $v_W \in \mathcal{U}_W$ the polynomial $u \in \mathcal{U}$ be such that*

$$|u|_{1,\tau_0} = \inf_{\phi \in \mathcal{U}_{IF}} |v_W + \phi|_{1,\tau_0}\,, \quad \mathcal{U}_{IF} := \mathcal{U}_I \cup \mathcal{U}_F\,.$$

Then we have

$$(1+\log p)^{-1} |v_W|^2_{L_2(W)} \prec |u|^2_{1,\tau_0} \prec |v_W|^2_{L_2(W)}\,. \tag{9.84}$$

Proof. The proof, as the proof of Lemma 9.3, can be produced by the analogy with the h-version.

Let $\mathcal{H}_{\mathrm{sp},W}(\tau_0)$ be the wire basket subspace of the space $\mathcal{H}_{\mathrm{sp}}(\tau_0)$ of piecewise $\mathcal{Q}_1$ functions on the orthogonal mesh with GLL nodes. For any $v \in \mathcal{Q}_p$ and $\hat{\pi}_p v \in \mathcal{H}_{\mathrm{sp}}(\tau_0)$ coinciding at the spectral nodes, one has the inequalities

$$\begin{gathered} \|v\|_{0,\tau_0} \asymp \|\hat{\pi}_p v\|_{0,\tau_0}\,, \quad |v|_{1,\tau_0} \asymp |\hat{\pi}_p v|_{1,\tau_0}\,, \\ ||v||_{L_2(W)} \asymp ||\hat{\pi}_p v||_{L_2(W)}\,, \quad |v|_{L_2(W)} \asymp |\hat{\pi}_p v|_{L_2(W)}\,. \end{gathered} \tag{9.85}$$

The inequalities of the first row are proved by means of Theorem 7.5, p. 245, in the same way as the equivalences (7.146) in 2d. The first inequality in the second row follows by application of the first inequality in (7.129) to each edge of the cube τ_0, whereas the second inequality in this row is a consequence of the first. If (9.84) holds, then, by virtue of (9.85), we arrive at the inequalities

$$(1+\log p)^{-1} |\hat{v}_W|^2_{L_2(W)} \prec |\hat{u}|^2_{1,\tau_0} \prec |\hat{v}_W|^2_{L_2(W)}\,, \tag{9.86}$$

that hold for all $\hat{v}_W \in \mathcal{H}_{\mathrm{sp},W}(\tau_0)$, where $\hat{u}$ satisfies

$$|\hat{u}|_{1,\tau_0} = \inf_{\hat{\phi}} |\hat{v}_W + \hat{\phi}|_{1,\tau_0}\,, \quad \hat{\phi}\big|_W = 0\,,\ \hat{\phi} \in \mathcal{H}_{\mathrm{sp}}(\tau_0)\,. \tag{9.87}$$

Indeed, in order to obtain the right inequality, we apply (9.85) to the right inequality in (9.84), and get

$$|\hat{\pi}_p u|^2_{1,\tau_0} \prec |\hat{v}_W|^2_{L_2(W)}\,, \quad \hat{v}_W = \hat{\pi}_p v_W \in \mathcal{H}_{\rm sp}(\tau_0)\,,\ \forall\, v_W \in \mathcal{U}_W\,.$$

The definition of $\hat{u}$ in (9.87) allows us to replace the function under the norm on the left by $\hat{u}$. It is left to take into account that the set of all $\hat{v}_W$ in the above inequality is the subspace $\mathcal{H}_{{\rm sp},W}(\tau_0)$. The left inequality in (9.86) is proved in the same way.

The opposite is also true, and, therefore, it is sufficient to prove (9.86). Suppose, that the reference cube τ_0 is discretized by means of $\mathcal{Q}_1$ finite elements on the quasiuniform orthogonal mesh with the mesh parameter $\hbar$, and $\mathcal{V}_\hbar(\tau_0)$ is the corresponding FE space. Then, according to (5.95), see p. 131, and the trace theorem, we have

$$(1+\log\hbar^{-1})^{-1}|\,\hat{v}\,|^2_{L_2(W)} \prec |\,\hat{v}\,|^2_{1,\tau_0}\,, \quad \forall\ \hat{v} \in \mathcal{V}_\hbar(\tau_0)\,. \tag{9.88}$$

The spectral mesh can be covered by the quasiuniform orthogonal mesh with the mesh parameter $\hbar$ comparable with the smallest mesh size of the spectral mesh, which order is p^{-2}. Since $\mathcal{H}_{\rm sp}(\tau_0)$ is a subspace of such a space $\mathcal{V}_h(\tau_0)$, the inequality (9.88) holds for all $\hat{v} \in \mathcal{H}_{\rm sp}(\tau_0)$. Now the left inequality (9.86) follows from (9.88) and (9.85), if to take into account that $\log\hbar^{-1} \asymp \log p^2 = 2\log p$.

We turn to the right inequality (9.86). Functions from the subspace $\mathcal{H}_{{\rm sp},W}(\tau_0)$ are distinct from zero only on the two types of nests of the spectral mesh: 8 corner cubic nests, each having a vertex in the vertex of τ_0, and the rest nests, which are rectangular parallelepipeds with one edge of the largest size lying on one of the edges of τ_0. Smaller edges of the nest are equal and lie in the two faces converging at the edge, which is situated on the edge of τ_0. Direct calculations for each nest with the use of these simple properties show that

$$|\,\hat{v}\,|^2_{1,\tau_0} \prec \|\,\hat{v}\,\|^2_{L_2(W)}\,, \quad \forall\ \hat{v} \in \mathcal{H}_{{\rm sp},W}(\tau_0)\,, \tag{9.89}$$

which is similar to the inequality (5.97) for the case of the quasiuniform finite element meshes. Applying the quotient space argument, then (9.85), and minimizing the seminorm on the left among polynomials having the same value on $\partial\tau_0$ with $v \leftrightarrow \hat{v}$, $v \in \mathcal{U}$, we come to the right inequality (9.84). □

The proof of the inequalities (9.84) by virtue of the results for the h version belongs to [Casarin (1997)].

Clearly, the left inequality also implies that, if $\mathcal{U} \subset \mathcal{Q}_p$, then, for any $v \in \mathcal{U}$ and $v|_W = v_W$, we have

$$(1 + \log p)^{-1} |v_W|^2_{L_2(W)} \prec |v|^2_{1,\tau_0} .$$

In the case of the hierarchical basis $\mathcal{M}_p$, the second term in the brackets in (9.83) is the matrix in which the only 8×8 block, related to the vertex degrees of freedom, is nonzero. Since $\mathbf{M}_{W_r}$ is assembled from the mass matrices for edges, simplification of the wire basket preconditioners may be achieved, *e.g.,* by replacement of these edge matrices by the preconditioners for 1d mass matrices, which were introduced in Section 7.4. In particular, the tridiagonal blocks of matrix $\mathbf{M}_{W_r}$, related to each of the edges, may be preconditioned by the matrix $\overline{\Delta}$.

In the case of the spectral reference element, for the representative of which we take the element with GLL nodes, several close preconditioners may be defined with the same asymptots of the generalized condition numbers and computational complexity. One of the simplest one results from the implementation the GLL quadrature rule to compute the L_2-norm in (9.80). Namely, it is induced by the bilinear form

$$b_{W,\,\mathrm{Sp}}(v_W, v_W) = c_W \inf_c \sum_{\eta_{\boldsymbol{\alpha}} \in W} \sigma_{\boldsymbol{\alpha}} (v_W(\eta_{\boldsymbol{\alpha}}) - c)^2 , \tag{9.90}$$

where $\eta_{\boldsymbol{\alpha}}$ are the quadrature nodes and $\sigma_{\boldsymbol{\alpha}}$ are the corresponding weights. The choice of this bilinear form is underpinned by the equivalence relationship

$$\|v_W\|^2_{L_2(W)} \asymp \sum_{\eta_{\boldsymbol{\alpha}} \in W} \sigma_{\boldsymbol{\alpha}} (v_W(\eta_{\boldsymbol{\alpha}}))^2 , \tag{9.91}$$

which follows by the application of (7.131) to each of the 12 edges of the cube τ_0. From (9.91), we also obtain the equivalence

$$|v_W|^2_{L_2(W)} \asymp \inf_c \sum_{\eta_{\boldsymbol{\alpha}} \in W} \sigma_{\boldsymbol{\alpha}} (v_W(\eta_{\boldsymbol{\alpha}}) - c)^2 , \tag{9.92}$$

which is (9.90) up to the factor c_W. We come to the similar to (9.83) expression

$$\boldsymbol{\mathcal{W}}_r = c_W \left(\mathbf{D}_{W_r} - \frac{1}{24} (\mathbf{D}_{W_r} \mathbf{z}_1)(\mathbf{D}_{W_r} \mathbf{z}_1)^\top \right) , \tag{9.93}$$

in which $\mathbf{D}_{W_r}$ is the diagonal matrix of the quadratic form (9.91) with the weights of the quadrature rule for diagonal entries. However, in distinction from (9.83), the matrix $(\mathbf{D}_{W_r} \mathbf{z}_1)(\mathbf{D}_{W_r} \mathbf{z}_1)^\top$ is fully populated.

9.1.4.2 *Some Remarks on the Implementation*

We remind that here N_W, N_E and N_V are the notations for the numbers of the wire basket, edge and vertex unknowns, respectively. In the case of the discretization by the finite elements with the hierarchical coordinate polynomials, the system

$$\boldsymbol{\mathcal{S}}_W \mathbf{v} = \mathbf{f}\,, \quad \mathbf{v}, \mathbf{f} \in V_W, \tag{9.94}$$

of the size N_W may be easily condensed to the system

$$\boldsymbol{\mathcal{S}}_V^E \mathbf{v}_V = \overline{\mathbf{f}}_V\,, \quad \mathbf{v}, \mathbf{f} \in V_V, \tag{9.95}$$

with respect to the vertex unknowns, which has the much smaller size N_V. The condensation by elimination of the internal for the edges d. o. f. can be performed for $\mathcal{O}(N_E)$ arithmetic operations. In order to approve this conclusion, we consider the matrix $\mathbb{M}_W$ obtained by assembling the matrices $\mathbf{M}_W^{(r)}$, where by definition

$$\mathbf{M}_W^{(r)} = c_W h_r \varrho_r \mathbf{M}_{W_r}. \tag{9.96}$$

Having represented $\boldsymbol{\mathcal{S}}_W$ and $\mathbb{M}_W$ in the block forms

$$\boldsymbol{\mathcal{S}}_W = \begin{pmatrix} \boldsymbol{\mathcal{S}}_E & \boldsymbol{\mathcal{S}}_{E,V} \\ \boldsymbol{\mathcal{S}}_{V,E} & \boldsymbol{\mathcal{S}}_V \end{pmatrix} \quad \text{and} \quad \mathbb{M}_W = \begin{pmatrix} \mathbb{M}_E & \mathbb{M}_{E,V} \\ \mathbb{M}_{V,E} & \mathbb{M}_V \end{pmatrix},$$

we see that $\boldsymbol{\mathcal{S}}_E = \mathbb{M}_E$. This follows from (9.83) and the mentioned above fact that the matrix $(\mathbf{M}_{W_r}\mathbf{z}_1)(\mathbf{M}_{W_r}\mathbf{z}_1)^T$ does not depend on r and has only one nonzero block related to the vertices with the entries equal to 9. The global mass matrix $\mathbb{M}_W$ may also be viewed as assembled from the mass matrices $\overline{\mathbf{M}}_{E_j}$ of 1d "elements" $\overline{E}^j$, each being the edge "j" together with its vertices. Clearly, we have

$$\overline{\mathbf{M}}_{E^j} = c_W \left(\sum_i h_{E^j} \varrho_{r_i(j)} \right) \overline{\mathbf{M}}_{E^j,0}\,,$$

where the matrix $\overline{\mathbf{M}}_{E^j,0}$ is induced by the quadratic form $\|\cdot\|^2_{L^2(-1,1)}$ on the correspondimg to this edge space of polynomials, and $r_i(j)$ are the numbers of elements $\overline{\tau}_r$, having E^j for the edge. For the hierarchical reference elements, the block $\mathbf{M}_{E^j,0}$, which is the restriction of $\overline{\mathbf{M}}_{E^j,0}$ to the subspace of the internal for the edge E^j unknowns, is a tridiagonal matrix, see Section 7.4. Hence, $\mathbb{M}_E$ is the block diagonal matrix

$$\mathbb{M}_E = c_W \operatorname{diag}\left[\left(\sum_i h_{E^j} \varrho_{r_i(j)}\right) \mathbf{M}_{E^j,0} \right]_{j=1}^{Q_E},$$

and, indeed, the d.o.f. internal for edges may be easily eliminated for $\mathcal{O}(N_E)$ arithmetic operations. This part of the two-step procedure for calculation of $\boldsymbol{\mathcal{S}}_V^E$ can be done edgewise in parallel:

1) condense edge d.o.f. freedom in the matrices $\mathbf{W}_r$, defined in (9.83), and obtain 8×8 matrices, which we denote $\mathbf{V}_r$,

2) assemble the matrix $\boldsymbol{\mathcal{S}}_V^E = \boldsymbol{\mathcal{S}}_V - \boldsymbol{\mathcal{S}}_{V,E}\boldsymbol{\mathcal{S}}_E^{-1}\boldsymbol{\mathcal{S}}_{E,V}$ from the matrices $h_r\varrho_r\mathbf{V}_r$.

Step 2) is unnecessary, if the system with the matrix $\boldsymbol{\mathcal{S}}_V^E$ is solved iteratively.

Some simplifications are obvious reflecting properties of finite element assemblages. In the case of the quasiuniform finite element mesh, the length h_{E^j} of the edge E^j may be replaced by the mesh parameter h. Suppose, the powers of the 1d reference edge polynomials are the same for all edges, *i.e.,* $p_{E^j} \equiv p_E$. Then $\mathbf{M}_{E^j,0} \equiv \mathbf{M}_{E,0}$ and $\mathbf{V}_r \equiv \mathbf{V}_0$ up to perturbations. For simplicity, suppose additionally that $\varrho = \text{const}$, FE mesh is quasiuniform and topologically equivalent to the orthogonal one, and at the definition of the matrix $\boldsymbol{\mathcal{S}}_V^E$ the values h_{E^j}, h_r are replaced by the mesh parameter h. Then the matrix $\mathbb{M}_E$ contains repeated scaled blocks $\mathbf{M}_{E,0}$ on its diagonal, *i.e.,*

$$\mathbb{M}_E = 4c_W h\varrho \,\mathrm{diag}\,[\,\mathbf{M}_{E,0}\,]_{j=1}^{Q_E}\,,$$

the matrix $\boldsymbol{\mathcal{S}}_V^E$ has a regular 27-point stencil, and, hence, fast solvers for it can most likely be devised.

In the case of the spectral discretizations, the matrix $\boldsymbol{\mathcal{S}}_V^E$ is the same or possesses close properties, *e.g.*, if the bilinear form (9.90) is used for defining the wire basket preconditioner. However, the condensation of the edge unknowns requires more operations.

For solving (9.94), another algorithm can be used, which was initially introduced for h discretizations, see Subsection 5.1.5. Instead of solving system (9.95), it assumes three steps, which require solving three systems: one with the wire basket mass matrix, another with the matrix resulting from $\mathbb{M}_W$ after condensation of the edge unknowns, and an additional system of the dimension $\mathcal{R}$. Below, we present this algorithm briefly.

Solving (9.94) is equivalent to the minimization of the quadratic functional

$$\min_{\mathbf{v}} \left(\frac{1}{2} \sum_r \min_{c_r} (\mathbf{v}^{(r)} - \mathbf{z}_1 c_r)^T \mathbf{M}_W^{(r)} (\mathbf{v}^{(r)} - \mathbf{z}_1 c_r) - \mathbf{v}^T \mathbf{f} \right), \tag{9.97}$$

with $\mathbf{v}^{(r)}$ being the restriction of $\mathbf{v}$ to the wire basket of an element $\overline{\tau}_r$.

From (9.97), we get system

$$\mathbb{M}_W \mathbf{v} - \sum_r \mathbf{M}_W^{(r)} \mathbf{z}_1 c_r = \mathbf{f}\,, \quad \mathbf{z}_1^T \mathbf{M}_W^{(r)} (\mathbf{v}^{(r)} - \mathbf{z}_1 c_r) = 0\,. \tag{9.98}$$

For the initial step, we can express $\mathbf{v}$ from first subsystem and substitute it into second subsystem. This gives the subsystem for c_k's

$$(\mathbf{z}_1^\top \mathbf{M}_W^{(r)} \mathbf{z}_1) c_r - \mathbf{z}_1^\top \mathbf{M}_W^{(r)} \mathbb{M}_W^{-1} \sum_i \mathbf{M}_W^{(i)} \mathbf{z}_1 c_i = \mathbf{z}_1^\top \mathbf{M}_W^{(r)} \mathbb{M}_W^{-1} \mathbf{f}\,. \tag{9.99}$$

Having it solved, we return to the first subsystem (9.98) in order to find $\mathbf{v}$. For convenience, we consider vectors $\mathbf{M}_W^{(r)} \mathbf{z}_1$ in (9.98) and (9.99) as expanded by zero entries to match the dimension of $\mathbf{v}$.

The system (9.99) may be simplified, if to take into account properties of the involved matrices and vectors for a particular type of the discretization. We turn first to the hierarchical one. For the systems $\mathbb{M}_W \boldsymbol{\psi} = \boldsymbol{\phi}$, the procedure similar to solving the system (9.95) may be used, but now it becomes simpler. At the first step, we reduce it to the system $\mathbb{M}_V^E \boldsymbol{\psi}_V = \overline{\boldsymbol{\phi}}_V$, $\mathbb{M}_V^E = \mathbb{M}_V - \mathbb{M}_{V,E} \mathbb{M}_E^{-1} \mathbb{M}_{E,V}$, with respect to the vertex component of the vector $\boldsymbol{\psi}$ by the condensation of edge unknowns. The condensation and the backward substitution can be done in parallel for each 1d element $\overline{E}_j$. The properties of the matrix $\mathbb{M}_V^E$ are better than of $\boldsymbol{\mathcal{S}}_V^E$. Indeed, it is spectrally equivalent to its diagonal, and, therefore, the system with this matrix may be solved by PCGM with the diagonal preconditioner for $\mathcal{O}(N_E)$ arithmetic operations. We conclude that $\mathrm{ops}\,[(\mathbb{M}_W)^{-1} \boldsymbol{\phi}] = \mathcal{O}(N_W)$ for any $\boldsymbol{\phi} \in V_W$.

Since, the block of the matrix $\mathbf{M}_W^{(i)}$, related to the vertices, is generated by the trilinear polynomials, we have

$$\mathbf{M}_W^{(i)} \mathbf{z}_1 = 3\, c_W\, h_i\, \varrho_i\, \mathbf{z}^{(i)}\,, \quad \mathbf{z}^{(i)} = (0, 0, \ldots, 0, \underbrace{1, 1, \ldots, 1}_{8 \text{ vertices of } \tau_i}, 0, 0, \ldots, 0)^\top\,,$$

where we simply used another notation $\mathbf{z}^{(i)}$ for the vector $\mathbf{z}_1$ specified at the vertices of τ_i and continued by zeroes to the other vertices of the FE assemblage. Besides, we can notice that $(\mathbf{z}^{(r)})^T \mathbb{M}_W^{-1} \mathbf{z}^{(i)} = (\mathbf{z}^{(r)})^T (\mathbb{M}_V^E)^{-1} \mathbf{z}^{(i)}$, and in the result the system (9.99) may be rewritten as

$$c_r - \frac{3}{8} (\mathbf{z}^{(r)})^T (\mathbb{M}_V^{(W)})^{-1} \sum_i c_W\, h_i \varrho_i\, \mathbf{z}^{(i)} c_i = \frac{1}{8} (\mathbf{z}^{(r)})^T \mathbb{M}_W^{-1} \mathbf{f}\,. \tag{9.100}$$

When the wire basket preconditioner for the spectral discretizations is designed on the basis of the bilinear form (9.90), the system (9.99) takes the form

$$24\, c_r - \mathbf{d}_r^\top \mathbf{D}_W^{-1} \sum_i c_W\, h_i \varrho_i\, \mathbf{d}_i\, c_i = \mathbf{d}_r^\top \mathbf{D}_W^{-1} \mathbf{f}\,, \tag{9.101}$$

where $\mathbf{d}_r := \mathbf{D}_{W_r}\mathbf{z}_1$ is, evidently, the vector with the entries equal to the diagonal entries of the matrix $\mathbf{D}_{W_r}$ and continued by zeroes on the whole subspace V_W. Although vectors $\mathbf{d}_r$ in (9.101) are completely filled in — opposite to vectors $\mathbf{z}_1$ in (9.100), which have zeroes for all entries except ones living at the vertices — the former system is simpler and more sparse.

9.1.4.3 *Prolongations from Wire Basket by Inexplicit Solvers for Faces*

The choice of the prolongation operator $\mathbb{P}_W$ from the wire basket on the interface boundary in (9.28), (9.29) is controlled by the two contradictory conditions: to be computationally cheap and to be as close as possible to the prolongation operator $\mathbb{P} : \mathcal{U}_W \to \mathcal{U}_B$ in (9.42). The latter implies that the inequality

$$\|\mathbb{P}_W\mathbf{v}\|_{\mathbb{S}_B} \le c_{\mathbb{P}}\|\mathbf{v}\|_{\mathbb{S}_W^F}\,, \quad \forall\ \mathbf{v} \in V_W\,, \tag{9.102}$$

holds for sufficiently good $c_{\mathbb{P}}$. We will model the situation when different finite elements of the discretization are associated with the reference elements equipped with different polynomial spaces. For this purpose, we assume that $\mathcal{U} \subseteq \mathcal{Q}_p$ for some $p > 1$, but omit index r in the notations.

The two ways of designing such prolongation operators can be considered. One is based on inexact solvers for the internal problems on faces and assumes existence of efficient preconditioner-solvers for the face Schur complements. In another, the prolongation is explicitly defined by sufficiently simple expressions. Although the definition of the former follows the standard path, it deserves some additional attention. One of the reasons is that the invertible trace theorems for the traces on the wire basket have not been proved for functions from the space $H^1(\tau_0)$, as well from the spaces $\mathcal{Q}_p$ equipped with the same norm.

For the reference element polynomial space $\mathcal{U} \subseteq \mathcal{Q}_p$, by p_B, p_F, p_E we denote the maximal powers in each variable of the polynomials from the trace spaces $\mathcal{U}_B(\partial\tau_0)$,$\mathcal{U}_F(F)$ and $\mathcal{U}_E(E)$, respectively. In other words, they are the powers in each variable of the traces of the boundary, face and edge polynomials, respectively, on the boundary, faces and edges of the reference element. Similar notations are used for maximal powers in each variable of the traces on separate faces, edges etc. Most often we have the situation, when coordinate polynomials of the reference element possess the following property: if $v_W \in \mathcal{U}_W$, then for some $\varkappa \ge 0$

$$\| v_W \|_{1,\tau_0} \prec p_\circ^{\varkappa} \| v_W \|_{1,W}\,, \tag{9.103}$$

and $\varkappa$ is easily evaluated. The value $p_\circ$ depends on the powers of polynomials $v_W \in \mathcal{U}_W$ in the directions orthogonal to the edges. Suppose, *e.g.,* that τ_0 is a cube, the vertex polynomials are trilinear and the edge polynomials are bilinear in the plane orthogonal to the edge. Then, we can find out that $\varkappa = 0$. In particular, this takes place for the reference elements with the hierarchical coordinate polynomials. For the isotropic spectral elements, we have $p_\circ = p$.

Lemma 9.7. *Let (9.103) hold and*

$$\underline{\gamma}_F \boldsymbol{\mathcal{F}} \le \mathbb{S}_F \le \overline{\gamma}_F \boldsymbol{\mathcal{F}}$$

with positive $\underline{\gamma}_F, \overline{\gamma}_F$. *Let also for any* $\mathbf{v}_W \in \mathcal{U}_W$ *the prolongation* $\mathbf{u} = \mathbb{P}_W \mathbf{v}_W$ *be defined in such a way that* $\mathbf{u}_W = \mathbf{v}_W$ *and*

$$\mathbf{u}_F = \mathbf{u}_F^{k_o} = -\boldsymbol{\mathcal{I}}^{-1}[\mathbb{S}_F, \boldsymbol{\mathcal{F}}, k_o]\mathbb{S}_{FW}\mathbf{v}_W\,,$$

i.e., *the latter subvector is produced by the iteration process*

$$\mathbf{u}_F^{k+1} = \mathbf{u}_F^k - \sigma_{k+1}\boldsymbol{\mathcal{F}}^{-1}(\mathbb{S}_F\mathbf{u}_F^k + \mathbb{S}_{FW}\mathbf{v}_W)\,, \quad k = 0, 1, \dots, k_o - 1\,, \tag{9.104}$$

with the initial guess $\mathbf{u}_F^0 = \mathbf{0}$ *and Chebyshev iteration parameters* σ_k *for a fixed number* k_o *of iterations. Then, for any*

$$k_o \succ [1 + \varkappa \log p_\circ + \log p_W + \log(1 + \log p_B)]/(\log \rho^{-1})\,,$$

the inequality (9.102) holds with an absolute constant $c_{\mathbb{P}}$, *where* $\rho = (1 - \theta)/(1 + \theta)$, $\theta = \sqrt{\underline{\gamma}_F/\overline{\gamma}_F}$. *For the hierarchical cubic reference elements, this estimate is simplified to*

$$k_o \succ [1 + \log p_W + \log(1 + \log p_B)]/(\log \rho^{-1})\,.$$

Proof. Note that, if $\mathbf{v}_W$ is a constant vector, then $\mathbb{S}_{FW}\mathbf{v}_W = \mathbf{0}$ and $\mathbf{u}_F^k = \mathbf{0}$ for any $k \ge 1$.

Let $\boldsymbol{\varphi}_F$ be the solution of the system $\mathbb{S}_F\boldsymbol{\varphi}_F = -\mathbb{S}_{FW}\mathbf{v}_W$ and vector $\boldsymbol{\varphi}_B$ have $\boldsymbol{\varphi}_F$ and $\boldsymbol{\varphi}_W = \mathbf{v}_W$ for the components. By means of the convergence estimate $\|\mathbf{u}_F^k - \boldsymbol{\varphi}_F\|_{\mathbb{S}_F} \le \rho^k \|\boldsymbol{\varphi}_F\|_{\mathbb{S}_F}$, and the equality $\|\boldsymbol{\varphi}_B\|_{\mathbb{S}_B} = \|\mathbf{v}_W\|_{\mathbb{S}_W^F}$, for $\mathbf{u}_B^k = \mathbf{u}_F^k + \mathbf{v}_W$, we get

$$\begin{aligned}\|\mathbf{u}_B^k\|_{\mathbb{S}_B} &\le \|\boldsymbol{\varphi}_B\|_{\mathbb{S}_B} + \rho^k(\|\boldsymbol{\varphi}_B\|_{\mathbb{S}_B} + \|\mathbf{v}_W\|_{\mathbb{S}_B}) \\ &= \|\mathbf{v}_W\|_{\mathbb{S}_W^F} + \rho^k(\|\mathbf{v}_W\|_{\mathbb{S}_W^F} + \|\mathbf{v}_W\|_{\mathbb{S}_B})\,. \end{aligned} \tag{9.105}$$

As usual, vectors, belonging to different subspaces, are summed assuming that they are continued by zero entries outside of the sets of their initial definition. Let $v_W \in \mathcal{U}_W$ correspond to $\mathbf{v}_W$. By the definition of $\mathbb{S}_B$ as

the stiffness matrix for discrete harmonic functions, (9.103) and Markov's inequality, we have

$$\begin{aligned}\|\mathbf{v}_W\|^{\,2}_{\mathbb{S}_B} &= \inf_{w\in\mathcal{U}_I} |v_W + w|^2_{1,\tau_0} \le |v_W|^2_{1,\tau_0} \\ &\prec p_\circ^{2\varkappa}\|v_W\|^2_{1,W} \prec p_\circ^{2\varkappa} p_W^4 \|v_W\|^2_{0,W}\,, \end{aligned} \tag{9.106}$$

where

$$\|v\|^2_{l,\partial\tau_0} := \left(\sum_{k=1}^{6}\|v\|^2_{l,F_k}\right)^{1/2} \quad \text{and} \quad |v|^2_{l,\partial\tau_0} := \left(\sum_{k=1}^{6}|v|^2_{l,F_k}\right)^{1/2}$$

for a nonnegative integer l. Applying the quotient space argument, we come to the bound

$$\|\mathbf{v}_W\|_{\mathbb{S}_B} \prec p_\circ^{\varkappa}\, p_W^2 \inf_c \|v_W - c\|_{0,W}\,. \tag{9.107}$$

Suppose, w is any function from the space $\mathcal{U}_B(\partial\tau_0)$ of the traces on $\partial\tau_0$ of the polynomials from $\mathcal{U}$. According to Theorem 7.2, there exists such a linear prolongation operator $\mathcal{P} : \mathcal{U}_B(\partial\tau_0) \to \mathcal{Q}_{p_B}$ in the space $\mathcal{Q}_{p_B}$ from $\partial\tau_0$ on τ_0 that

$$|\mathcal{P}w|_{1,\tau_0} \prec |w|_{1/2,\partial\tau_0}\,.$$

Now, by virtue of Lemma 9.6 and the above inequality, we conclude that, for the traces of polynomials $w \in \mathcal{U}_B(\partial\tau_0)$ on the wire basket, inequality

$$||w - \overline{w}||^{\,2}_{0,W} \prec (1+\log p_B)|w|^2_{1/2,\partial\tau_0} \tag{9.108}$$

holds, where $\overline{w}$ is the average of w over the wire basket. It is fulfilled as well for the polynomial $\varphi_B \in \mathcal{U}_B$ corresponding to vector $\boldsymbol{\varphi}_B$. Therefore, substituting φ_B in (9.108) applying the trace theorem and equality $\|\boldsymbol{\varphi}_B\|_{\mathbb{S}_B} = \|\mathbf{v}_W\|_{\mathbb{S}^F_W}$, we get

$$\begin{aligned}||v_W - \overline{v}_W||^{\,2}_{0,W} &\prec (1+\log p_B)|\boldsymbol{\varphi}_B|^2_{1/2,\partial\tau_0} \\ &\prec (1+\log p_B)\|\boldsymbol{\varphi}_B\|^{\,2}_{\mathbb{S}_B} \\ &= (1+\log p_B)\|\mathbf{v}_W\|^{\,2}_{\mathbb{S}^F_W}\,. \end{aligned} \tag{9.109}$$

Substitution of (9.109) in (9.107) results in the inequality

$$\|\mathbf{v}_W\|_{\mathbb{S}_B} \prec p_\circ^{\varkappa} p_W^2 \sqrt{(1+\log p_B)}\,\|\mathbf{v}_W\|_{\mathbb{S}^F_W}\,,$$

which allows us to rewrite (9.105) as

$$\|\mathbf{u}^k_B\|_{\mathbb{S}_B} \le c\,(1+\rho^k + \rho^k p_\circ^{\varkappa} p_W^2\sqrt{(1+\log p_B)})\|\mathbf{v}_W\|_{\mathbb{S}^F_W}\,. \tag{9.110}$$

From (9.110), we conclude that, if $k_o \ge [1+\varkappa\log p_\circ + 2\log p_W + 0.5\log(1+\log p_B)]/(\log\rho^{-1})$, then

$$\|\mathbb{P}_a\mathbf{v}_W\|_{\mathbb{S}_B} = \|\mathbf{u}^{k_o}_B\|_{\mathbb{S}_B} \le 3c\,\|\mathbf{v}_W\|_{\mathbb{S}^F_W}\,.$$

As was noted, for the hierarchical elements, $\varkappa = 0$. Lemma is proved. □

We did not imposed any conditions on the powers of the traces on edges and faces as well as inside the reference element. If to require that $\mathcal{U} \supseteq \mathcal{Q}_{p_B}$, then by virtue of prolongation Theorem 7.2 the powers of $p_\circ$ and p_W in (9.106) can be lowered. In the case of spectral elements the number of iterations depends on $p_\circ$. However, one can use for the initial prolongation the polynomial defined by the vector $\mathbf{v}_W$, as for the hierarchical reference element. This removes the dependence on $p_\circ$.

Lemma 9.7 allows us to choose the prolongation operator from the wire basket on faces more reasonably. For the efficiency it is important to have a good preconditioner-solver $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{F}}_r$. In contrast to the edge problem preconditioning for 2d and 3d reference elements, the options for a choice of a fast preconditioner-solvers $\boldsymbol{\mathcal{F}}_r$ are much more restricted. For the complete and orthotropic hierarchical and spectral reference elements, examples of the fast preconditioner-solvers for faces will be considered in Subsection 9.3.2, following the suggestions by [Korneev *et al.* (2003b,a)].

The matrix-vector multiplications by $\mathbb{S}_F$ in (9.104) can increase the cost of a DD solver as well. One of the schemes of reducing this cost, discussed earlier, requires an additional preconditioner $\widetilde{\mathbb{S}}_B$ close in the spectrum to $\mathbb{S}_B$ and such that multiplications $\widetilde{\mathbb{S}}_B\mathbf{v}_B$ are cheap. Assuming $\widetilde{\mathbb{S}}_B$ and $\mathbb{S}_B$ are spectrally equivalent, we can replace $\mathbb{S}_F$ and $\mathbb{S}_{F,W}$ in (9.104) by $\widetilde{\mathbb{S}}_F$ and $\widetilde{\mathbb{S}}_{F,W}$ with the increase in the number of iterations only by a constant. Such a preconditioner may be defined implicitly, *e.g.,* as $\widetilde{\mathbb{S}}_B = \boldsymbol{\mathcal{A}}_B - \boldsymbol{\mathcal{A}}_{BI}\boldsymbol{\mathcal{A}}_I^{-1}\boldsymbol{\mathcal{A}}_{IB}$, if a fast solver exists for the realization of the operation $\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{v}_I$. In this case, *e.g.,* for each muliplcation by $\widetilde{\mathbb{S}}_F$ one can use formula $\mathbb{S}_F\mathbf{v}_F = (\boldsymbol{\mathcal{A}}_F - \boldsymbol{\mathcal{A}}_{FI}\boldsymbol{\mathcal{A}}_I^{-1}\boldsymbol{\mathcal{A}}_{IF})\mathbf{v}_F$, which will not damage the optimality of a whole DD solver.

9.1.4.4 *Prolongations from Wire Basket by Explicit Operators, Hierarchical and Spectral Elements*

Cheap prolongation operators from the wire baskets to the faces of finite elements can be defined explicitly by means of the same prolongation matrices as for the reference elements. Turning to the reference element with the coordinate polynomials (9.75), we at first introduce the tentative prolongation operator. Then it is adjusted in order to satisfy the requirements for the low energy prolongations.

For definiteness, we consider the face in the plane $x_3 \equiv -1$, which is denoted F_3, and suppose that the subspace of the internal functions for this

face is

$$\mathcal{U}_{_{F_3}}(F_3) = \operatorname{span}\left[L_{i,j}(x) = \mathcal{L}_i(x_1)\mathcal{L}_j(x_2)\right]_{i,j=2}^{p_1,p_2}, \tag{9.111}$$

where the notations $p_k = p_{_{F_3,k}}$ are used for brevity. We also suppose that the restrictions of the space $\mathcal{U}(\tau_0)$ to the edges belong to the spaces, spanned by the polynomials $\mathcal{L}_i(x_k)$, $0 \le i \le p_{E_l}$, $l = k, k+2$, and $k = 1, 2$. It is accepted that, the edges E_l, $l = k, k+2$, lie on the lines $x_{3-k} \equiv -1, 1$, respectively, and we require that

$$p_{_{F_3,k}} \ge \max_{l=k,k+2} p_{_{E_l}}\,. \tag{9.112}$$

For $t = 0, 1$, let us denote by $\mathcal{L}_{t,k}(x_k)$ the polynomials obtained by the orthogonalization of the polynomial $\mathcal{L}_t$ to the internal polynomials $\mathcal{L}_i$, $i = 2, 3, \ldots, p_{_{F_3,l}}$, in the space $L_2(-1, 1)$. The evaluation of the coefficients $c_{t,i}^{(l)}$ in

$$\mathcal{L}_{t,k}(x_k) = \mathcal{L}_t(x_k) + \sum_{i=2}^{p_{_{F_3,l}}} c_{t,i}^{(l)} \mathcal{L}_i(x_k)\,, \tag{9.113}$$

requires $\mathcal{O}(p_{_{F_3,l}})$ a.o. This is for the reason that the matrix of the corresponding system of algebraic equations is the tridiagonal block $\mathbb{K}_{0,I}$ of the matrix $\mathbb{K}_0$, introduced in Subsection 7.3.2, see also (7.88). For convenience, we use the notation $v_{i,j}$ with the same i, j, as in (9.111), for the components of the restriction of a vector $\mathbf{v}_W$ to ∂F_3. If the restriction ϕ of a function v_W to the wire basket and the corresponding vector $\mathbf{v}_W$ are given, then the tentative prolongation $u = \widetilde{I}_{W\to B}\phi$ of ϕ is defined on the face F_3 as

$$\begin{aligned} u = &\sum_{t,s=0}^{1} \mathcal{L}_{t,1}(x_1)\mathcal{L}_{s,2}(x_2) v_{k,l} + \sum_{t=0}^{1} \sum_{j=2}^{p_{_{E_{2+2t}}}} \mathcal{L}_{t,1}(x_1)\mathcal{L}_j(x_2) v_{k,j} \\ &+ \sum_{s=0}^{1} \sum_{i=2}^{p_{_{E_{1+2s}}}} \mathcal{L}_i(x_1)\mathcal{L}_{k,2}(x_2) v_{i,k}\,. \end{aligned} \tag{9.114}$$

Implying that the tentative operator $\widetilde{I}_{W\to B}$ is similarly defined for other faces, we can introduce the function $\mathcal{F} = 1 - \widetilde{I}_{W\to B}\,1$ which vanishes at the wire basket. Clearly, it can be represented as

$$\mathcal{F} = \sum_{k=1}^{6} \mathcal{F}_k\,,$$

where each $\mathcal{F}_k$ is nonzero only on one face F_k. The prolongation operator $I_{W\to B} : \mathcal{U}_W(W) \to \mathcal{U}_B(B)$ is defined according to

$$u = I_{W\to B}\,\phi := \widetilde{I}_{W\to B}\,\phi + \sum_{k=1}^{6} \mathcal{F}_k\,\overline{\phi}_{\partial F_k}\,, \tag{9.115}$$

with $\overline{\phi}_{\partial F_k}$ being the average value of ϕ on the boundary ∂F_k of a face F_k. We use for the matrix of this prolongation operator the notations $\mathbb{P}_W = \mathbf{I}_{W\to B}$ or $\mathbb{P}_W^{(r)} = \mathbf{I}_{W\to B}^{(r)}$ when it is used for the prolongations for element "r", which is the image of the reference element in the FE discretization.

In the case of the spectral element, the tentative prolongation operator $\widetilde{I}_{W\to B}$ can be defined in a much simpler way. If ϕ is the restriction of a function v_W to the wire basket and $\mathbf{v}_W$ is the corresponding vector, then by definition the tentative prolongation $v_B = \widetilde{I}_{W\to B}\phi$ has zero values at all internal for faces spectral nodes.

Lemma 9.8. *Let, for the hierarchical or spectral 3d reference p-elements, $\mathcal{U} = \mathcal{Q}_p$, and let $I_{W\to B} : \mathcal{U}_W(W) \to \mathcal{U}_B(B)$ be the prolongation operator. Then the estimate*

$$|I_{W\to B}\,\phi|^2_{1/2,\partial\tau_0} \prec (1+\log p)|\,\phi\,|^2_{0,W}\,, \quad \forall\ \phi \in \mathcal{U}_W(W), \tag{9.116}$$

is valid.

Proof. Due to (7.142) and (7.146), the prolongation operators for the hierarchical and spectral elements have the essential properties in common, and, as a consequence, it is sufficient to prove Lemma only, *e.g.*, for the spectral reference elements. The proof is based on inequalities

$$|\widetilde{I}_{W\to B}\,\phi|_{1/2,\partial\tau_0} \prec \|\,\phi\,\|_{0,W}\,, \quad \forall\ \phi \in \mathcal{U}_W(W)\,, \tag{9.117}$$

and

$$|\mathcal{F}_k|^2_{1/2,\partial\tau_0} \prec (1+\log p)\,. \tag{9.118}$$

Their proofs are the same as the proofs of the right and left inequalities in (9.84), respectively, if we take into account that $|u_W|_{1,\tau_0} \asymp |u_W|_{1/2,\partial\tau_0}$ for u_W in (9.84) by the trace theorem and Theorem 9.2. □

Obviously, since the prolongation for each face is independent of other faces, it is compatible for adjacent finite elements and for a FE assemblage can be implemented facewise.

The outline of the proof is the same as for the similar prolongations from the wire basket to the faces of subdomains of decomposition in h-version, see Subsections 5.1.5 and 5.2.3, and goes back to the papers by [Bramble *et al.* (1986)]-[Bramble *et al.* (1989)] and [Dryja *et al.* (1994)]. Its adaptation to the spectral discretizations was done by [Pavarino and Widlund (1996)] and [Casarin (1997)].

9.1.4.5 *Flow-Chart of the Dirichlet–Dirichlet DD Algorithm*

Algorithm 9.1 `Solution procedure for` $\mathcal{K}\mathbf{v}=\mathbf{f}$

for $r = 1, 2, \ldots, \mathcal{R}$ **do**

1) $\mathbf{v}_{I_r} := \mathcal{K}_{I_r}^{-1}\mathbf{f}_{I_r}$, {solving local Dirichlet problems in parallel}

2) $\mathbf{f}_{B_r}^{(1)} := \mathbf{P}_{V_{B_r}\to V_{I_r}}^{\top}\mathbf{v}_{I_r}$, {restrictions to element boundaries in parallel}

end for

3) $\mathbf{f}_B^{(1)} := \biguplus \mathbf{f}_{B_r}^{(1)}$, {correction of the right-hand side}

4) $\mathbf{f}_B := \mathbf{f}_B + \mathbf{f}_B^{(1)}$, {updating the interface subvector}

for $q = 1, 2, \ldots, \mathcal{Q}$ **do**

5) $\mathbf{v}_{F_q} := \mathcal{S}_{F_q}^{-1}\mathbf{f}_{F_q}$, {facewise solving}

6) $\mathbf{f}_{W_q}^{(1)} := \mathbf{P}_{V_{W_q}\to V_{F_q}}^{\top}\mathbf{v}_{F_q}$, { facewise restrictions to the boundaries}

end for

7) $\mathbf{f}_W^{(1)} := \biguplus \mathbf{f}_{W_q}^{(1)}$, { assembling the correction to the current wire basket subvector of the right part}

8) $\mathbf{f}_W := \mathbf{f}_W + \mathbf{f}_W^{(1)}$, { updating the current wire basket subvector of the right part}

9) $\mathbf{v}_W := \mathcal{S}_W^{-1}\mathbf{f}_W$, { solving the system with the wire basket Schur complement preconditioner for wire basket subvector of $\mathbf{v}_W$}

for $q = 1, 2, \ldots, \mathcal{Q}$ **do**

10) $\mathbf{v}_{F_q}^{(1)} := \mathbf{P}_{V_{W_q}\to V_{F_q}}\mathbf{v}_{W_q}$, { facewise prolongations inside faces from their boundaries in parallel}

11) $\mathbf{v}_{F_q} := \mathbf{v}_{F_q} + \mathbf{v}_{F_q}^{(1)}$, {facewise updating current face subvector in parallel}

end for

for $r = 1, 2, \ldots, \mathcal{R}$ **do**

12) $\mathbf{v}_{I_r}^{(1)} := \mathbf{P}_{V_{B_r}\to V_{I_r}}\mathbf{v}_{B_r}$, { elementwise prolongations inside elements from their boundaries in parallel}

13) $\mathbf{v}_{I_r} := \mathbf{v}_{I_r} + \mathbf{v}_{I_r}^{(1)}$, { elementwise updating of current internal subvectors in parallel}

end for

14) **Set** $\mathbf{v}_{I_r}$, $\mathbf{v}_{F_q}$, $\mathbf{v}_W$ **for components of** $\mathbf{v}$, ***i.e.*, set**

$$\mathbf{v}^{\top} := (\mathbf{v}_{I_1}^{\top}, \ldots, \mathbf{v}_{I_{\mathcal{R}}}^{\top}, \mathbf{v}_{F_1}^{\top}, \ldots, \mathbf{v}_{F_{\mathcal{Q}}}^{\top}, \mathbf{v}_W^{\top})$$

9.1.5 *Numerical Complexity*

First let us turn to the case when the FE discretization is generated by means of a single cubic hierarchical reference element with the space $\mathcal{U}(\tau_0) = \mathcal{Q}_p$, $p \geq 1$, and let us assume that there is an efficient preconditioner-solver $\boldsymbol{\mathcal{A}}_I$ for the systems with the system matrix $\mathbf{A}_I$. Then a DD preconditioner-solver $\boldsymbol{\mathcal{K}}$ can be defined by (9.6), (9.7) and (9.28) according to the assumptions "A" listed below:

i) $\boldsymbol{\mathcal{A}}_I$ is a preconditioner-solver satisfying (9.31),
ii) $\boldsymbol{\mathcal{F}}_\diamond = \boldsymbol{\mathcal{S}}_{00}$ satisfies (9.47),
iii) $\boldsymbol{\mathcal{W}}$ satisfies (9.81) with the matrix $\mathbf{W}$ defined by (9.80) for $c_W = 1 + \log p$,
iν) the inequality (9.79) holds with $\Theta(p) = p^t$, $t = \text{const}$, and the submatrix $\mathbb{P}_{I,B}$ of the prolongation matrix $\mathbb{P}_B$ is defined for $\nu = \mathcal{O}(1 + \log p)$ iterations according (9.77), if $\mathcal{Q}_1 \subseteq \mathcal{U}_B$, and by one of the expressions (9.78), if $\mathcal{Q}_1 \nsubseteq \mathcal{U}_B$,
ν) the prolongation $\mathbb{P}_W$ is defined
  ν$_\text{a}$) by $k_o = (1 + \log p)^2$ iterations according to Lemma 9.7, if (9.103) holds, or
  ν$_\text{b}$) by setting $\mathbb{P}_W = \mathbf{I}_{W \to B}$ with the latter matrix as in (9.115).

We start from the set of the assumptions "A" with the option *ν*$_\text{a}$). Under these assumptions, the condition number $\text{cond}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}]$, can be estimated by means of the bounds (9.33), but first of all it is necessary to estimate some values entering these bounds.

From *ii*) and the proof of Theorem 9.1, it follows that $\underline{\gamma}_F \succ \underline{\gamma}_0/(1 + \log p)^2$ and $\overline{\gamma}_F \prec \overline{\gamma}_0$. Besides, taking into account *iii*), (9.81) and Lemma 9.6, we have

$$(1 + \log p)^{-2}|\mathbf{v}_W|^2_{\boldsymbol{\mathcal{W}}} \prec |\mathbf{v}_W|^2_{\mathbb{S}^F_W} \prec (1 + \log p)^{-1}|\mathbf{v}_W|^2_{\boldsymbol{\mathcal{W}}}\,. \tag{9.119}$$

Obviously, the assumption *iν*) and Lemma 9.5 approve the bound $\beta_B \prec 1$.

Finally, we consider the consequences of the assumptions *ν*). In the case *ν*$_\text{a}$) together with the condition (9.103) that is assumed to be fulfilled, Lemma 9.7 allows us to get the bound

$$||\mathbb{P}_W\mathbf{v}_W||^2_{\mathbb{S}_B} \prec (1 + \log p)^\kappa ||\mathbf{v}_W||^2_{\mathbb{S}^F_W}\,, \quad \kappa = 0\,. \tag{9.120}$$

Therefore, $\beta_W \prec (1 + \log p)^\kappa$. With the help of the above estimates and (9.33), we conclude that (9.34) holds with

$$\begin{aligned} &\underline{\gamma} \geq \tfrac{1}{2\beta_B} \min\left(\underline{\gamma}_I, \tfrac{\underline{\gamma}_F}{2\beta_W}, \tfrac{1}{2\beta_W}\right) \succ \min\left(\underline{\gamma}_I, \tfrac{\min(1,\underline{\gamma}_0)}{(1+\log p)^2}\right)\,, \\ &\overline{\gamma} \leq \max\left(\overline{\gamma}_I, \overline{\gamma}_F, \overline{\gamma}_W\right) \prec \max\left(\overline{\gamma}_I, \overline{\gamma}_0, 1/(1 + \log p)\right)\,. \end{aligned} \tag{9.121}$$

It can be noticed that, in the case of the iterative prolongations $\mathbb{P}_W$, there is no special need to choose $c_W = 1+\log p$, since any choice, satisfying $1 \prec c_W \prec (1+\log p)$, results in the condition number cond $[\boldsymbol{\mathcal{K}}^{-1}\mathbf{K}] = \overline{\gamma}/\underline{\gamma}$ of the same order.

Estimates (9.33) do not allow to guarantee the best possible condition number, when the prolongations $\mathbb{P}_W = \mathbf{I}_{W\to B}$ in (9.115) are implemented. In such a case Lemma 9.2 is more suitable for the use. The same values of β_B, $\overline{\gamma}_F \prec \overline{\gamma}_0$, and $\overline{\gamma}_W \succ (1+\log p)^{-2}$, as estimated above, are also present in the assumptions (9.44) of this Lemma. However, it is necessary to estimate $\widetilde{\gamma}_W$ and $\underline{\gamma}_F$, which now have different meaning.

Taking into account estimate (9.116) and the definitions of the matrices $\widetilde{\mathbb{S}}_B$, $\mathbb{S}_B$, $\widetilde{\mathbb{S}}_W$, we can write

$$||\mathbf{v}_W||^2_{\widetilde{\mathbb{S}}_B} = ||\mathbb{P}_W\mathbf{v}_W||^2_{\mathbb{S}_B} \prec |\mathbf{v}_W|^2_{\boldsymbol{\mathcal{W}}}\,. \tag{9.122}$$

Hence, $\widetilde{\gamma}_W \prec 1$ in any case. The first estimate in (9.44) was proved by [Pavarino and Widlund (1996)]. In particular, for the prolongation operator (9.115), they proved the inequality

$$_{00}||u_F||^2_{1/2,F_k} = {}_{00}||u - I_{W\to B}u||^2_{1/2,F_k} \le c(1+\log p)^2|u|^2_{1,\tau_0}\,, \tag{9.123}$$

from where it directly follows that $\underline{\gamma}_F \succ \underline{\gamma}_0(1+\log p)^{-2}$. Now, inserting the estimated values of β_B, $\underline{\gamma}_F$, $\overline{\gamma}_F$, $\underline{\gamma}_W$ and $\widetilde{\gamma}_W$ in (9.45), we get

$$\underline{\gamma} \ge \frac{1}{2\beta_B}\min\left(\underline{\gamma}_I, \frac{1}{(\underline{\gamma}_0^{-1}+1)(1+\log p)^2}\right)\,, \quad \overline{\gamma} \le 4\max(\overline{\gamma}_I, \overline{\gamma}_0, 1)\,. \tag{9.124}$$

Let us suppose that all the values in (9.121) and (9.124), except for p, are absolute constants. We will see that the means to satisfy the assumptions i)-ν), considered earlier and to be considered in this section, allow us to approve this supposition for rather general DD preconditioner-solvers. In this case, in order to solve the system $\mathbf{Kv} = \mathbf{f}$ with the accuracy ϵ by PCG, we need $n_{\text{PCG}} = \boldsymbol{\mathcal{O}}((1+\log p)\log\epsilon^{-1})$ iterations. Let us introduce the notations

$$\text{ops}\,[\boldsymbol{\mathcal{A}}_I^{-1}\mathbf{v}_I] = n_I\,, \quad \forall\,\mathbf{v}_I \in U_I\,,$$

$$\text{ops}\,[\boldsymbol{\mathcal{S}}_{00}^{-1}\mathbf{v}_F] = n_F\,, \quad \forall\,\mathbf{v}_F \in U_F\,.$$

Using the iterative prolongation $\mathbb{P}_W$ according to ν_{a}) or ν_{b}), we arrive at the bounds

$$\text{ops}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}] \le c_1(1+\log p)^2(n_I+n_F)\mathcal{R}\,,$$

$$\text{ops}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}] \le c_1[(1+\log p)n_I+n_F]\mathcal{R}\,,$$

respectively, where minor terms are omitted. Logarithms appear due to the iterative prolongations $\mathbb{P}_B$ and $\mathbb{P}_W$ in the case $\nu_{\rm a}$) and due to the iterative prolongation $\mathbb{P}_B$ in the case $\nu_{\rm b}$). The consequence of these estimates and (9.121), (9.124) is that the respective arithmetical costs $\mathcal{N}_{DD} := n_{\rm PCG}{\rm ops}\,[\boldsymbol{\mathcal{K}}^{-1}\mathbf{v}]$ of preconditioning operations for solving the system $\mathbf{Kv} = \mathbf{f}$ with the accuracy ϵ can be estimated by

$$\begin{aligned} &\mathcal{N}_{DD} \le c\,(1+\log p)^3[n_I + n_F]\mathcal{R}\log\epsilon^{-1}\,,\\ &\mathcal{N}_{DD} \le c(1+\log p)[(1+\log p)n_I + n_F]\mathcal{R}\log\epsilon^{-1}, \end{aligned} \tag{9.125}$$

with the constant c only depending, on the constants in the generalized conditions of the shape regularity, as the constants $\underline{c}$ and $\overline{c}$ in the relative spectrum bounds (9.34).

Thus, we have proved the following theorem.

Theorem 9.3. *Let the assumptions "A" imposed on the DD preconditioner - solver be fulfilled. Then the inequalities (9.34) hold, where $\underline{\gamma}$ and $\overline{\gamma}$ satisfy (9.121) and (9.124), respectively to the assumptions $\nu_{\rm a}$) or $\nu_{\rm b}$), and $\underline{c}$ and $\overline{c}$ are positive constants depending only on the generalized conditions of the shape regularity. If additionally β_B and γ's in the left parts of (9.121) and (9.124) are constants, then the arithmetical complexities of the preconditioning in DD algorithms satisfy (9.125).*

In the rest of this chapter, we present examples of the preconditioner-solvers, which guarantee $n_I \prec p^3$ and $n_F \prec p^2$ and, therefore, allow to obtain the almost optimal in the arithmetical work DD preconditioner-solvers. For general unstructured finite element meshes, the bottleneck in reducing the complexity of a DD solver is the matrix-vector multiplications $\mathbf{Kv}$, the related to which arithmetical work can be significant. We refer to the paper by [Melenk *et al.* (2001)] devoted to the analysis and reduction of the cost of these multiplications. The obvious case, when they are relatively cheap and do not compromise total almost optimality of the DD preconditioner-solver, is the case of the discretisations of elliptic equations with constant coefficients by the rectangular hierarchical p-elements.

9.2 Reference Elements and Finite-Difference Preconditioners

The tensor product nature of the coordinate polynomials results in the representations of the reference element stiffness and mass matrices and

their finite-difference/finite element preconditioners by sums of Kronecker's products of the matrices, related to 1d discretizations. Due to this reason, expressions for the preconditioners are straightforwardly adjusted to the multidimensional case with any dimensions $d > 1$. However, the finite-difference and the differential operators, which respectively they formally represent and approximate, in the case of the hierarchical reference elements significantly differ from those for 2d. Therefore, they can have different computational properties, and this is exactly the case for the popular reference element with the hierarchical coordinate functions obtained by means of the integrated Legendre polynomials. When passing from 2d to 3d, the most important computational characteristic of the preconditioner, namely the order of the related differential operator, changes from 2nd to 4th. This makes some important types of fast solvers for the preconditioners in 2d not expandable to 3d, and suggests development of new solution procedures.

We start from the introduction of the stiffness and mass matrices of the reference elements and their finite-difference/finite element preconditioners.

9.2.1 *Hierarchical Reference Elements*

The basis in the space $\mathcal{Q}_p$ of the complete hierarchical 3d reference element is the set

$$\mathcal{M}_p = \left\{ L_\alpha(x) = \mathcal{L}_{\alpha_1}(x_1)\mathcal{L}_{\alpha_2}(x_2)\mathcal{L}_{\alpha_3}(x_3)\ ,\ \alpha \in \omega \right\}\ , \tag{9.126}$$

where $\omega = \{\alpha = (\alpha_1, \alpha_2, \alpha_3) : 0 \le \alpha_1, \alpha_2, \alpha_3 \le p\}$. We reorder the coordinate polynomials in $\mathcal{M}_p$ in a way similar to that one used for the square reference element, see (7.93), p. 232. It reflects the natural separation of the set $\mathcal{M}_{1,p}$ of 1d polynomials in two groups, containing polynomials of even and odd degrees, respectively. Accordingly, basis polynomials $L_{i,k,l}$ are separated in eight groups characterized by the values of the indices:

$$\left.\begin{array}{l} 1) - \text{ all } i,k,l \text{ even}, \\ 2) - i,k \text{ even and } l \text{ odd}, \\ 3) - i,l \text{ even and } k \text{ odd}, \\ 4) - i \text{ odd and } k,l \text{ even}, \\ 5) - i,l \text{ odd and } k \text{ even}, \\ 6) - i,l \text{ odd and } k \text{ even}, \\ 7) - i,k \text{ odd and } l \text{ even}, \text{ and} \\ 8) - \text{ all } i,k,l \text{ odd}. \end{array}\right\} \tag{9.127}$$

After these perturbations, the hierarchical reference element stiffness

and mass matrices $\mathbf{A}_I$ and $\mathbb{M}_I$ get the block diagonal structure:

$$\begin{aligned} \mathbf{A}_I &= \text{diag}\,[\mathbf{A}_{eee}, \mathbf{A}_{eeo}, \mathbf{A}_{eoe}, ..., \mathbf{A}_{ooe}, \mathbf{A}_{ooo}]\ , \\ \mathbb{M}_I &= \text{diag}\,[\mathbb{M}_{eee}, \mathbb{M}_{eeo}, \mathbb{M}_{eoe}, ..., \mathbb{M}_{ooe}, \mathbb{M}_{ooo}]\ , \end{aligned} \tag{9.128}$$

where

$$\begin{aligned} \mathbf{A}_{abc} &= \mathbb{K}_{1,I,a} \otimes \mathbb{K}_{0,I,b} \otimes \mathbb{K}_{0,I,c} + \mathbb{K}_{0,I,a} \otimes \mathbb{K}_{1,I,b} \otimes \mathbb{K}_{0,I,c} \\ &\quad + \mathbb{K}_{0,I,a} \otimes \mathbb{K}_{0,I,b} \otimes \mathbb{K}_{1,I,c}\ , \\ \mathbb{M}_{abc} &= \mathbb{K}_{0,I,a} \otimes \mathbb{K}_{0,I,b} \otimes \mathbb{K}_{0,I,c}\ , \quad a,b,c = e,o\,, \end{aligned} \tag{9.129}$$

and $\mathbb{K}_{l,I,e}, \mathbb{K}_{l,I,o}$, $l = 0, 1,$ are the matrices for 1d reference element, introduced in Subsection 7.3.2, see (7.94).

The finite-difference like preconditioners for the internal block $\mathbf{A}_I$ of the reference element stiffness matrix can be defined by the expressions

$$\begin{aligned} \mathbf{\Lambda}_I &= \text{diag}\,[\mathbf{\Lambda}_{eee}, \mathbf{\Lambda}_{eeo}, \mathbf{\Lambda}_{eoe}, ..., \mathbf{\Lambda}_{ooe}, \mathbf{\Lambda}_{ooo}]\ , \\ \mathbf{\Lambda}_I^{(a)} &= \text{diag}\,[\mathbf{\Lambda}_{aaa}, ..., \mathbf{\Lambda}_{aaa}]\ , \quad \mathbf{\Lambda}_\triangle = \text{diag}\,[\mathbf{\Lambda}, \mathbf{\Lambda}, ..., \mathbf{\Lambda}]\ , \end{aligned} \tag{9.130}$$

reflecting the block diagonal structure of the matrices in (9.128). As in 2d, we define the $N^3 \times N^3$ blocks on the diagonals by the use of the matrices

$$\mathbf{\Delta}_e = \mathbf{\Delta} + \boldsymbol{\mathcal{D}}_e^{-1} \quad \text{and} \quad \mathbf{\Delta}_o = \mathbf{\Delta} + \boldsymbol{\mathcal{D}}_o^{-1}\,,$$

introduced in (7.100) and (7.103) on p. 235. This yields the preconditioners

$$\begin{aligned} \mathbf{\Lambda}_{abc} &= \boldsymbol{\mathcal{D}}_a \otimes \mathbf{\Delta}_b \otimes \mathbf{\Delta}_c + \mathbf{\Delta}_a \otimes \boldsymbol{\mathcal{D}}_b \otimes \mathbf{\Delta}_c + \mathbf{\Delta}_a \otimes \mathbf{\Delta}_b \otimes \boldsymbol{\mathcal{D}}_c\ , \\ \mathbf{\Lambda} &= \boldsymbol{\mathcal{D}}_e \otimes \mathbf{\Delta} \otimes \mathbf{\Delta} + \mathbf{\Delta} \otimes \boldsymbol{\mathcal{D}}_e \otimes \mathbf{\Delta} + \mathbf{\Delta} \otimes \mathbf{\Delta} \otimes \boldsymbol{\mathcal{D}}_e\ , \end{aligned} \tag{9.131}$$

with $a, b, c = e, o$.

Spectrally equivalent preconditioners of the same type for the internal mass matrix are taken in the form

$$\begin{aligned} \mathbf{\Xi}_I &= \text{diag}\,[\mathbf{\Xi}_{eee}, \mathbf{\Xi}_{eeo}, \mathbf{\Xi}_{eoe}, ..., \mathbf{\Xi}_{ooe}, \mathbf{\Xi}_{ooo}]\ , \\ \mathbf{\Xi}_I^{(a)} &= \text{diag}\,[\mathbf{\Xi}_{aaa}, ..., \mathbf{\Xi}_{aaa}]\ , \quad \mathbf{\Xi}_\triangle = \text{diag}\,[\mathbf{\Xi}, \mathbf{\Xi}, ..., \mathbf{\Xi}]\ , \end{aligned} \tag{9.132}$$

where $\mathbf{\Xi}_{a,b,c} = \mathbf{\Delta}_a \otimes \mathbf{\Delta}_b \otimes \mathbf{\Delta}_c$ and $\mathbf{\Xi}_\triangle = \mathbf{\Delta} \otimes \mathbf{\Delta} \otimes \mathbf{\Delta}$.

Clearly, Theorem 7.3 and Corollary 7.2 on p. 237, are retained in the 3d case, and are reformulated below for convenience.

Theorem 9.4. *Let $\boldsymbol{\mathcal{A}}_I^\star$ be any of the preconditioners $\mathbf{\Lambda}_I$, $\mathbf{\Lambda}_I^{(a)}$, $\mathbf{\Lambda}_\triangle$ for the internal stiffness matrix $\mathbf{A}_I$, and let $\boldsymbol{\mathcal{M}}_I^\star$ be any of the preconditioners $\mathbf{\Xi}_I$, $\mathbf{\Xi}_I^{(a)}$, $\mathbf{\Xi}_\triangle$ for the internal mass matrix $\mathbb{M}_I$ of the 3d hierarchical reference element. Then we have*

$$\mathbf{A}_I \asymp \boldsymbol{\mathcal{A}}_I^\star \quad \textit{and} \quad \mathbb{M}_I \asymp \boldsymbol{\mathcal{M}}_I^\star\,.$$

Corollary 9.1. *Let* $\mathbf{K}_{I_r}$ *and* $\mathbf{M}_{I_r}$ *be the stiffness and mass matrices of the finite elements of a discretization of the 3d problem in Example 2.2, p. 20, by means of a shape regular finite element assemblage and* $\tau_r = \Omega_r$, $r = 1, 2, \ldots, \mathcal{R} = J$. *Let also* $\boldsymbol{\mathcal{K}}_{I_r}$, $\boldsymbol{\mathcal{M}}_{I_r}$ *be the preconditioners for* $\mathbf{K}_{I_r}$, $\mathbf{M}_{I_r}$, *respectively, defined by*

$$\boldsymbol{\mathcal{K}}_{I_r} = h^{(r)} \varrho_r \boldsymbol{\mathcal{A}}^{\star}_I \quad \textit{and} \quad \boldsymbol{\mathcal{M}}_{I_r} = (h^{(r)})^3 \boldsymbol{\mathcal{M}}^{\star}_I \,,$$

where $h^{(r)}$ *is the characteristic size of the finite element, see (3.4). Then* $\mathbf{K}_{I_r}$ *and* $\boldsymbol{\mathcal{K}}_{I_r}$, *as well as* $\mathbf{M}_{I_r}$ *and* $\boldsymbol{\mathcal{M}}_{I_r}$, *are spectrally equivalent uniformly in* p, r *and* $\mathcal{R}$, i.e.

$$\mathbf{K}_{I_r} \asymp \boldsymbol{\mathcal{K}}_{I_r} \quad \textit{and} \quad \mathbf{M}_{I_r} \asymp \boldsymbol{\mathcal{M}}_{I_r}$$

up to the positive constants depending only on the generalized conditions of the shape regularity.

In the analysis, it is more convenient to deal with the finite element preconditioners. Let the unit cube $\pi_1 := (0,1)^3$ be subdivided by the uniform square mesh of the size $\hbar = 1/(N+1)$, and $\mathring{\mathcal{H}}(\pi_1)$ be the space of functions continuous on π_1, trilinear on each cubic cell of the mesh, and vanishing on $\partial\pi_1$. The finite element preconditioner $\boldsymbol{\Lambda}^{(e)}_{\rm fem}$ is introduced the matrix of the symmetric bilinear form

$$a_{\pi_1}(u,v) = \frac{1}{\hbar} \sum_{k,l,m=1}^{3} \int_{\pi_1} [\, \varphi_k u_{,lm} v_{,lm} + (\frac{\varphi_k}{\varphi_l} + \frac{\varphi_l}{\varphi_k}) u_{,m} v_{,m} + \frac{\varphi_k}{\varphi_l \varphi_m} uv \,]\, d\xi \tag{9.133}$$

or the form

$$a_{\pi_1}(u,v) = \frac{1}{\hbar} \sum_{k,l,m=1}^{3} \int_{\pi_1} \varphi_k u_{,lm} v_{,lm} d\xi \tag{9.134}$$

on the space $\mathring{\mathcal{H}}(\pi_1)$ with the FE nodal basis. Here the notations $\varphi_k = \xi_k^2$, $u_{,m} = \partial u/\partial \xi_m$ and $u_{,lm} = \partial^2 u/\partial \xi_l \partial \xi_m$ are used and the summation over repeating indices is assumed.

Theorem 9.5. *The matrices*

$$\boldsymbol{\Lambda}_{I,\rm fem} = \mathrm{diag}\,[\boldsymbol{\Lambda}^{(e)}_{\rm fem}, ..., \boldsymbol{\Lambda}^{(e)}_{\rm fem}]$$

and $h^{(r)} \varrho_r \boldsymbol{\Lambda}_{I,\rm fem}$ *are spectrally equivalent to the internal stiffness matrix of the reference element and to the internal stiffness matrix of any finite element of the finite element assemblage, respectively,* i.e.,

$$\begin{gathered} \boldsymbol{\Lambda}_{I,\rm fem} \prec \mathbf{A}_I \prec \boldsymbol{\Lambda}_{I,\rm fem} \,, \\ \underline{\gamma}_I h^{(r)} \varrho_r \boldsymbol{\Lambda}_{I,\rm fem} \le \mathbf{K}_{I_r} \le \overline{\gamma}_I h^{(r)} \varrho_r \boldsymbol{\Lambda}_{I,\rm fem} \,, \end{gathered} \tag{9.135}$$

with the positive constants $\underline{\gamma}_I$ *and* $\overline{\gamma}_I$ *depending only on the constants in the generalized conditions of the shape regularity.*

Proof. The proof is reduced to the proof of the inequalities

$$\mathbf{\Lambda}_{\rm fem}^{(e)} \prec \mathbf{\Lambda}_{eee} \prec \mathbf{\Lambda}_{\rm fem}^{(e)}\,, \tag{9.136}$$

from which left pair of the inequalities (9.135) directly follows by the definitions of $\mathbf{\Lambda}_{I,\rm fem}$ and $\mathbf{\Lambda}_I^{(a)}$ and Theorem 9.4. The proof of the right pair of inequalities (9.135) involves additionally the use of Corollary 9.1.

The proof of (9.136) for the matrix $\Lambda_{\rm fem}^{(e)}$ of the bilinear form (9.134) is produced by the standard elementwise comparison. We only outline it. Let the notation $\pi_{\boldsymbol{\alpha}} = \{\xi : (\alpha_k - 1)\hbar < \xi_k < \alpha_k\hbar,\ k = 1,2,3\}$ stands for the domains of finite elements and $\mathbf{\Lambda}_{{\rm fem},\boldsymbol{\alpha}}$ denotes an element 8×8 stiffness matrix. The matrix $\mathbf{\Lambda}_{\rm fem}^{(e)}$ is the result of assembling of finite element stiffness matrices $\mathbf{\Lambda}_{{\rm fem},\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}$ with components, satisfying $1 \le \alpha_k \le N$, and subsequent imposing the Dirichlet boundary condition. The matrix $\mathbf{\Lambda}_{eee}$ can be also represented as the result of the assemblage of the 8×8 artificial finite element stiffness matrices, which are denoted $\mathbf{\Lambda}_{eee,\boldsymbol{\alpha}}$, and imposing the Dirichlet boundary condition. With such representations done, the proof of (9.136) is reduced to the proof of the spectral equivalences of the matrices $\mathbf{\Lambda}_{I,{\rm fem},\boldsymbol{\alpha}}$ and $\mathbf{\Lambda}_{eee,\boldsymbol{\alpha}}$ uniform in N and $\boldsymbol{\alpha}$. This proof is done without special technical difficulties. There are also alternative paths to prove (9.136).

The spectral equivalence of the matrices $\mathbf{\Lambda}_{{\rm fem},\boldsymbol{\alpha}}$, corresponding to both bilinear forms (9.133) and (9.134), is a consequence of Lemma 7.5. □

It is helpful to look at the finite-difference interpretation of the introduced preconditioners, which was among main initial intentions at their derivation. Let us consider the following boundary value problem:

$$\mathcal{L}u \equiv z^2 u_{,xxyy} + y^2 u_{,xxzz} + x^2 u_{,yyzz} = f \text{ in } \pi_1, \quad u = 0 \text{ on } \partial\pi_1, \tag{9.137}$$

where, *e.g.*, $u_{,xxyy} = \partial^4 u/\partial x^2 \partial y^2$. The differential operator of this problem (up to the multiplier) is induced by the bilinear form (9.134). Indeed, for any sufficiently smooth v, vanishing on the boundary $\partial\pi_1$, we have $(\mathcal{L}u, v)_{\pi_1} = \hbar a_{\pi_1}(u,v)$, since all boundary integrals vanish after the integration by parts. Now, if we introduce the uniform square mesh of size $\hbar = 1/(N+1)$ and approximate the differential operator in (9.137) by a nineteen-point stencil finite-difference operator, we come to the system

$$\hbar^{-2}\,\mathbf{\Lambda}_I \mathbf{u} = \mathbf{f}\,, \tag{9.138}$$

with $\mathbf{u}, \mathbf{f} \in \mathcal{V}_I = R^{(N-1)^3}$ and the same $\mathbf{\Lambda}_I$ as in (9.130). Let $u_{i,j,k}$ be the values of the approximate solution at the nodes $(x, y, z) = \hbar(i, j, k)$ and $f_{i,j,k} = f(i\hbar, j\hbar, k\hbar)$, for $i, j, k = 0, 1, \ldots, N+1$. With these notations, (9.138) take the form

$$\begin{aligned}
\hbar^{-2}[\, & k^2\,(4u_{i,j,k} - 2u_{i,j-1,k} - 2u_{i-1,j,k} - 2u_{i,j+1,k} - 2u_{i+1,j,k} \\
& + u_{i-1,j-1,k} + u_{i-1,j+1,k} + u_{i+1,j+1,k} + u_{i+1,j-1,k}) \\
& + j^2\,(4u_{i,j,k} - 2u_{i,j,k-1} - 2u_{i-1,j,k} - 2u_{i+1,j,k} - 2u_{i,j,k+1} \\
& + u_{i-1,j,k-1} + u_{i-1,j,k+1} + u_{i+1,j,k+1} + u_{i+1,j,k-1}) \\
& + i^2\,(4u_{i,j,k} - 2u_{i,j,k-1} - 2u_{i,j-1,k} - 2u_{i,j+1,k} - 2u_{i,j,k+1} \\
& + u_{i,j-1,k-1} + u_{i,j-1,k+1} + u_{i,j+1,k+1} + u_{i,j+1,k-1})] = f_{i,j,k}\,, \\
& 1 \le i, j, k \le N\,,
\end{aligned} \tag{9.139}$$

where $u_{i,j,k} = 0$, if at least one of the indices equals 0 or $N+1$.

The preconditioners introduced above can significantly ease solving local Dirichlet problems on subdomains of decomposition, what is especially important in 3d case. First of all they reduce the dimensions of the systems to be solved in 8 times. Besides, finite-difference/finite element interpretations immediately allow to apply, *e.g.,* some banded or the nested dissection type algorithms, which can provide almost optimal order of eliminations amongst direct Gauss elimination procedures. However further improvement of solvers does not lie on the surface. From (9.134), (9.137) and (9.139), we see that the preconditioners are the finite-difference/finite element approximations of the rather specific partial differential equation. This equation contains only the 4th order mixed derivatives with the coefficients, deteriorating on the part of the boundary $\partial\pi_1$ and, at the same time, highly orthotropic on the part of the domain π_1. These properties complicate designing optimal or almost optimal in p solvers for the systems with these preconditioners for the system matrices. Nevertheless, we can now find several suggestions on such solvers in the literature, a few of which will be discussed in what follows. As in the 2d case, these contributions assume some preliminary adaptation of the finite-difference/finite element preconditioners to the specific type of a fast solver.

9.2.2 *Hierarchical Elements for Adaptive Computations*

In p-adaptive implementation, we can have a specific space $\mathcal{Q}_{\mathbf{p}_r}$ for each element, where $\mathbf{p}_r = (\mathbf{p}_{I_r}, \mathbf{p}_{F_r}, \mathbf{p}_{E_r})$ is the multiindex with the entries which itselvs can be multiindices, *i.e.,* submultiindices, of different dimensions. The entries of $\mathbf{p}_r$ characterize the orders of the internal, face and edge polynomials, respectively, which in turn may be different for each direction. More precisely, in general, one may have

$$\mathbf{p}_{I_r} = (p_{I_r,1}, p_{I_r,2}, p_{I_r,3}), \quad \mathbf{p}_{r,F} = (\mathbf{p}_{r,F_1}, \mathbf{p}_{r,F_2}, \ldots, \mathbf{p}_{r,F_6}),$$
$$\mathbf{p}_{r,F_i} = (p_{r,F_i,1}, p_{r,F_i,2}), \quad \mathbf{p}_{r,E} = (p_{r,E_1}, p_{r,E_2}, \ldots, p_{r,E_{12}}),$$

with F_i and E_i denoting faces and edges of the reference cube. For definiteness, we assume that the faces in the planes $x_k \equiv -1, 1$ orthogonal to the axis x_k have numbers $i = k, k+3$, respectively. The edges with $i = 4(k-1)+l,\ l = 1,2,3,4$, are parallel to the axis x_k and, in the plane orthogonal to it, are ordered counter-clockwise, starting from the lower left. At this ordering, the degrees and the sets of indices α, for the internal, face and edge polynomials $L_\alpha(x)$ are defined as follows:

- $\widetilde{\omega}_I = \{\alpha : 2 \le \alpha_k \le p_{I_r,k}\}$ for the internal polynomials,
- $\widetilde{\omega}_{F_i} = \{\alpha : \alpha_k = 0$ for $i = k$ and $\alpha_k = 1$ for $i = k+3; 2 \le \alpha_{k-1} \le p_{r,F_i,1};\ 2 \le \alpha_{k+1} \le p_{r,F_i,2}\}$ for the face F_i polynomials, where $k \pm 1$ are understood modulo 3, and
- $\widetilde{\omega}_{E_i} = \{\alpha : 2 \le \alpha_k \le p_{r,E_i};\ \alpha_{k\pm1}(i)\}$ for the edge E_i polynomials, with $(\alpha_{k-1}(i), \alpha_{k+1}(i)) = (0,0,), (1,0), (1,1,), (0,1)$ respectively to $i = 4(k-1)+l$.

Besides, we have 8 vertex trilinear coordinate polynomials, with the set of indices $\widetilde{\omega}_V = (\alpha : \alpha_k = 0, 1)$. The sets $\widetilde{\omega}_I, \widetilde{\omega}_{F_i}$ and $\widetilde{\omega}_{E_i}$ depend on the number r of the element, but we do not attach the corresponding index to their notations.

Summarizing, we conclude that the reference element polynomial space for an element τ_r is

$$\mathcal{Q}_{\mathbf{p}_r} = \text{span}\,[L_\alpha(x)]_{\alpha\in\omega_r},$$

where $\omega_r = \widetilde{\omega}_I \cup \widetilde{\omega}_F \cup \widetilde{\omega}_E \cup \widetilde{\omega}_V$, with $\widetilde{\omega}_F = \cup_{i=1}^{6} \widetilde{\omega}_{F_i}$ and $\widetilde{\omega}_E = \cup_{i=1}^{12} \omega_{E_i}$. Hence, $\mathcal{Q}_{\mathbf{p}_r} \subset \mathcal{Q}_{\bar{\mathbf{p}}_r}$, where $\bar{\mathbf{p}}_r = (p_{r,1} p_{r,2}, p_{r,3})$, $p_{r,k} = \max_{\alpha\in\omega_r}(\alpha_k)$ and $\mathcal{Q}_{\bar{\mathbf{p}}_r}$ is the space of polynomials of degrees not greater $p_{r,k}$ in each variable x_k. Vectors $\mathbf{p}_r$ are assumed to be such that finite elements of the FE assemblage are compatible and generate the FE space belonging to $C(\overline{\Omega}) \bigcap H^1(\Omega)$.

Stiffness and mass matrices of the hierarchical reference element with different powers of polynomials for its interior, faces and edges and different directions can be easily written down explicitly. We suppose that $\mathcal{Q}_{\mathbf{p}} = \mathcal{Q}_{\bar{\mathbf{p}}}$, $\bar{\mathbf{p}} = (p_1, p_2, p_3)$, and for simplicity $p_k = 2N_k + 1$, where index r is omitted. Therefore, we have an orthotropic, in a sense of powers of polynomials, reference element. Then the stiffness matrix $\mathbf{A}_I$ and the preconditioner $\mathbf{\Lambda}_I$ are defined by the same expressions (9.128), (9.129) and (9.130), (9.131), respectively. However, the three matrices which are factors in each of the Kronecker products in (9.129) and (9.131) have dimensions N_1, N_2, N_3, respectively to first, second and third positions of factors. Particularly, in (9.129) the matrices with the indices a, b, c are of the dimensions N_1, N_2, N_3, respectively. If we have the reference element with the space $\mathcal{U} = \mathcal{Q}_{\mathbf{p}} \subset \mathcal{Q}_{\bar{\mathbf{p}}}$, then the stiffness matrix $\mathbf{A}_I$ for it is obtained from $\mathbf{A}_I$ for the element with the space $\mathcal{Q}_{\bar{\mathbf{p}}}$ by omitting all the rows and columns, corresponding to the absent degrees of freedom. The preconditioner $\mathbf{\Lambda}_I$ for such an element is obtained in the same way.

Polynomial spaces $\mathcal{Q}_{\mathbf{p}}$, described above, have special structure. Under mild conditions, they allow to design efficient DD solvers for discretizations by finite elements, which are images of reference elements equipped with such polynomial spaces. Some reference elements with other choices of subsets in $\mathcal{M}_p$ for the bases in polynomial spaces are used in the discretizations induced by a single reference element, *i.e.*, in the discretzations with no p-adaptation. In particular, the Serendipity family of incomplete reference elements can be defined as follows. Let us represent $\mathcal{M}_p$ as

$$\mathcal{M}_p = \mathcal{M}_{p,I} \cup \mathcal{M}_{p,B}\,,$$

where $\mathcal{M}_{p,I}$ is the set of the internal coordinate functions, and define

$$\mathcal{M}_I^p = \{L_{\boldsymbol{\alpha}}\}_{\boldsymbol{\alpha}\in\omega_I^p}\,, \quad \omega_I^p = \{\boldsymbol{\alpha} : 2 \le (\alpha_1 + \alpha_2 + \alpha_3), \alpha_1, \alpha_2, \alpha_3 \le p\}\,.$$

For the set of coordinate functions of the incomplete reference element we can use

$$\mathcal{M}^p = \mathcal{M}_I^p \cup \mathcal{M}_{p,B}\,.$$

9.2.3 *Spectral Reference Elements*

For the 3d spectral reference elements, the sets of nodes are defined by the tensor product of 1d meshes $x_k = \eta_i,\ k = 1, 2, 3$, (7.79) and (7.80), p. 228, whereas the pseudospectral mesh is the tensor product of the 1d mesh satisfying (7.137) at $\gamma = 1$. The notations $\mathbf{A} = \mathbf{A}_{\rm sp}, \mathbf{A}_{\rm p/s}$ and $\mathbb{M} = \mathbb{M}_{\rm sp}, \mathbb{M}_{\rm p/s}$

will stand for the reference element stiffness matrices, induced by the bilinear form

$$a_{\tau_0}(u,v) = \int_{\tau_0} \nabla u \cdot \nabla v \, dx \,,$$

and mass matrices, respectively. Suppose, that one of the pointed out above rectangular grids $x_k = \eta_i$ is given. Let $\mathcal{H}(\tau_0)$ denote the space of continuous on $\overline{\tau}_0$ and trilinear on each nest of the mesh functions vanishing on $\partial\tau_0$. We use the similar to the 2d case notations $\boldsymbol{\mathcal{A}}_\circ = \boldsymbol{\mathcal{A}}_{\rm sp}, \boldsymbol{\mathcal{A}}_{\rm p/s}$ for the matrices, generated by the bilinear form $a_{\tau_0}(u,v)$ on the spaces $\mathcal{H}(\tau_0)$. The former matrix corresponds to the GLL or GLC grid and the latter to the pseudospectral grid. The corresponding preconditioners for the mass matrices $\mathbb{M}$ are denoted by $\boldsymbol{\mathcal{M}}_\circ = \boldsymbol{\mathcal{M}}_{\rm sp}, \boldsymbol{\mathcal{M}}_{/\rm sp}$.

Lemma 9.9. *Let* $\mathbf{A}$ *and* $\mathbb{M}$ *be the stiffness and mass matrices for the 3d Lagrange reference element with the GLL nodes. Then we have* $\mathbf{A} \asymp \boldsymbol{\mathcal{A}}_\circ$ *and* $\mathbb{M} \asymp \boldsymbol{\mathcal{M}}_\circ$.

Proof. The stiffness and the mass matrices of the reference element are

$$\begin{aligned} \mathbf{A}_{\rm sp} &= \mathbb{K}_{1,\rm sp} \otimes \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{0,\rm sp} + \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{1,\rm sp} \otimes \mathbb{K}_{0,\rm sp} \\ &\quad + \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{1,\rm sp} \,, \\ \mathbb{M}_{\rm sp} &= \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{0,\rm sp} \otimes \mathbb{K}_{0,\rm sp} \,. \end{aligned} \tag{9.140}$$

The preconditioners may also be represented in the form of the sum of the Kronecker products, *e.g.*, of the 1d FE operators

$$\begin{aligned} \boldsymbol{\mathcal{A}}_\circ &= \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_1 + \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_1 \otimes \mathbb{K}_{0,\rm fem} \\ &\quad + \mathbb{K}_1 \otimes \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_{0,\rm fem}, \\ \boldsymbol{\mathcal{M}}_\circ &= \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_{0,\rm fem} \otimes \mathbb{K}_{0,\rm fem} \,, \end{aligned} \tag{9.141}$$

where $\mathbb{K}_1$ is the same matrix as in (7.139) and by $\mathbb{K}_{0,\rm fem}$ is understood the FE matrix, generated by the scalar product in $L_2(-1,1)$ on the space of the piecewise linear functions over the mesh $x = \eta_i$. Therefore, as in 2d, it is sufficient to establish the equivalences

$$\mathbb{K}_{1,\rm sp} \asymp \mathbb{K}_1 \quad \text{and} \quad \mathbb{K}_{0,\rm sp} \asymp \mathbb{K}_0 \,,$$

which are found in the proof of Lemma 7.8. We recollect that $\mathbb{K}_{1,\rm sp} = \mathbb{K}_1$, provided that both matrices are defined on the same mesh. □

Clearly, $\mathbb{K}_{0,\rm fem} \asymp \mathbb{K}_0$ with $\mathbb{K}_0$ the same as in (7.138), p. 248, and, therefore, the simpler preconditioners $\boldsymbol{\mathcal{A}}^\hbar = \boldsymbol{\mathcal{A}}^\hbar_{\rm sp}, \boldsymbol{\mathcal{A}}^\hbar_{\rm p/s}$ and $\boldsymbol{\mathcal{M}}^\hbar = \boldsymbol{\mathcal{M}}^\hbar_{\rm sp}, \boldsymbol{\mathcal{M}}^\hbar_{\rm p/s}$

spectrally equivalent to $\mathbf{A}$ and $\mathbb{M}$, respectively, may be introduced by the formulas

$$\begin{aligned} \boldsymbol{\mathcal{A}}^{\hbar} &= \mathbb{K}_0 \otimes \mathbb{K}_0 \otimes \mathbb{K}_1 + \mathbb{K}_0 \otimes \mathbb{K}_1 \otimes \mathbb{K}_0 + \mathbb{K}_1 \otimes \mathbb{K}_0 \otimes \mathbb{K}_0 \,, \\ \boldsymbol{\mathcal{M}}^{\hbar} &= \mathbb{K}_0 \otimes \mathbb{K}_0 \otimes \mathbb{K}_0 \,. \end{aligned} \tag{9.142}$$

9.2.4 *Factorized Preconditioners for Spectral Reference Elements*

Now, we are going to approve the preconditioner of the form $\boldsymbol{\mathfrak{K}}_{\diamond} = \mathbf{C}_I^{-1}\boldsymbol{\Lambda}_{\diamond}\mathbf{C}_I^{-1}$ for the internal stiffness matrix $\mathbf{A}_I$, in which $\mathbf{C}_I$ is the diagonal matrix and $\boldsymbol{\Lambda}_{\diamond}$ is the finite-difference (finite element) matrix on the uniform mesh. Although the matrix $\boldsymbol{\Lambda}_{\diamond}$ and the matrix $\boldsymbol{\Lambda}$, earlier defined for the hierarchical reference element, are significantly different, some of their properties essential for computations are similar. Due to this important observation, a variety of fast solvers of the same type, as designed for $\boldsymbol{\Lambda}$ in 2d case, can be expanded to 3d case, *i.e.*, can be adapted to the matrices $\boldsymbol{\Lambda}_{\diamond}$, $\boldsymbol{\mathfrak{K}}_{\diamond}$ and eventually to $\mathbf{A}_I$.

The transformation of variables $\mathbf{v} = \mathbf{C}\widetilde{\mathbf{v}}$ by the diagonal matrix

$$\mathbf{C} = p^3\, \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2}$$

transforms matrix (9.142) into the matrix

$$\begin{aligned} \widetilde{\mathbb{A}} := \mathbf{C}\boldsymbol{\mathcal{A}}^{\hbar}\mathbf{C} = p^6\, \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2}\boldsymbol{\mathcal{A}}^{\hbar}\, \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} \otimes \mathbb{K}_0^{1/2} = \\ = p^6 \left(\mathbb{K}_0^2 \otimes \mathbb{K}_0^2 \otimes \widetilde{\mathbb{K}}_1 + \mathbb{K}_0^2 \otimes \widetilde{\mathbb{K}}_1 \otimes \mathbb{K}_0^2 + \mathbb{K}_0^2 \otimes \mathbb{K}_0^2 \otimes \widetilde{\mathbb{K}}_1 \right) \end{aligned} \tag{9.143}$$

with $\widetilde{\mathbb{K}}_1 = \mathbb{K}_0^{1/2}\mathbb{K}_1\mathbb{K}_0^{1/2}$ and the same $\mathbb{K}_1$ as in (7.147). Further, with the help of the matrices $\boldsymbol{\Delta}_{\diamond}$ and $\boldsymbol{\mathcal{D}}_{\diamond}$, defined in Lemma 7.10, p. 251, we can introduce the matrices

$$\begin{aligned} \boldsymbol{\Lambda}_{\dagger} = \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} \otimes (\boldsymbol{\Delta}_{\diamond} + \boldsymbol{\mathcal{D}}_{\diamond}^{-1}) + \boldsymbol{\mathcal{D}}_{\diamond} \otimes (\boldsymbol{\Delta}_{\diamond} + \boldsymbol{\mathcal{D}}_{\diamond}^{-1}) \otimes \boldsymbol{\mathcal{D}}_{\diamond} + \\ + (\boldsymbol{\Delta}_{\diamond} + \boldsymbol{\mathcal{D}}_{\diamond}^{-1}) \otimes \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} \,, \\ \boldsymbol{\Lambda}_{\diamond} = \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\Delta}_{\diamond} + \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\Delta}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} + \boldsymbol{\Delta}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} \otimes \boldsymbol{\mathcal{D}}_{\diamond} \,, \end{aligned} \tag{9.144}$$

and, in the same way as in Lemma 7.10, we can prove that they are spectrally equivalent to the block $\widetilde{\mathbb{A}}_I$ of the matrix $\widetilde{\mathbb{A}}$, *i.e.,*

$$\boldsymbol{\Lambda}_{\diamond} \asymp \widetilde{\mathbb{A}}_I \asymp \boldsymbol{\Lambda}_{\dagger} \,. \tag{9.145}$$

The matrix $\boldsymbol{\Lambda}_{\diamond}$ may be interpreted as the usual 7-point finite-difference approximation on a cubic mesh to the differential operator

$$\mathcal{L}u = -\left[\phi^2(y)\phi^2(z)u_{,xx} + \phi^2(x)\phi^2(z)u_{,yy} + \phi^2(x)\phi^2(y)u_{,zz}\right] \tag{9.146}$$

in τ_0 under homogeneous Dirichlet boundary condition, where ϕ is the same function as in (7.157). Indeed, if we introduce the uniform orthogonal mesh of the size $\hbar = 1/N$, then, for each internal node, we can write

$$\mathbf{\Lambda}_\diamond \mathbf{u}|_{\mathbf{i}} = -\tfrac{1}{\hbar^2} \sum_{k=1,2,3} \phi^2_{i_{k-1}} \phi^2_{i_{k+1}} [u_{\mathbf{i}-\mathbf{e}_k} - 2u_{\mathbf{i}} + u_{\mathbf{i}+\mathbf{e}_k}] , \qquad (9.147)$$

with $1 \le i_k \le (p-1)$. We mention that $\mathbf{u}$ is assumed to have zero entries at the nodes on $\partial\tau_0$, and indices $k-1, k+1$ are understood modulo 3. The notations $\mathbf{e}_k = (e_{k,1}, e_{k,2}, e_{k,3})^\top$ stand for the unit vectors with the components equal to Kronecker's delta $e_{k,l} = \delta_{k,l}$.

Let $\mathbf{C}_I$ be the block of the matrix $\mathbf{C}$ related to internal unknowns. According to (9.145) and the definition of $\widetilde{\mathbb{A}}_I$, the preconditioners

$$\mathfrak{K}_\diamond = \mathbf{C}_I^{-1} \mathbf{\Lambda}_\diamond \mathbf{C}_I^{-1} \quad \text{and} \quad \mathfrak{K}_\dagger = \mathbf{C}_I^{-1} \mathbf{\Lambda}_\dagger \mathbf{C}_I^{-1}$$

are spectrally equivalent to $\mathbf{A}_I$ uniformly in p. Since $\mathbf{C}_I$ is the diagonal matrix, the efficiency of these preconditioners entirely depends on the quality of fast solvers for $\mathbf{\Lambda}_\diamond$ and $\mathbf{\Lambda}_\dagger$, respectively.

Comparing second expression (9.144), and (9.146), (9.147) with second expression (9.131) and (9.137), (9.139), we see differences between $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\diamond$, which are even more significant than in the 2d case. The most visible one is that the matrix $\mathbf{\Lambda}$ approximates a 4th order partial differential equation, whereas $\mathbf{\Lambda}_\diamond$ approximates a 2nd order equation. This is related to the representation of $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\diamond$ by the sums of the permutations of Kronecker's products $\boldsymbol{\mathcal{D}}_e \otimes \mathbf{\Delta} \otimes \mathbf{\Delta}$ and $\boldsymbol{\mathcal{D}}_\diamond \otimes \boldsymbol{\mathcal{D}}_\diamond \otimes \mathbf{\Delta}_\diamond$, respectively. We see that in the former and the latter Kronecker products, matrices $\mathbf{\Delta}$ and $\mathbf{\Delta}_\diamond$ of the second order finite-differences stand for two and one factors, respectively. However, the properties of 1d factors in each pair $\boldsymbol{\mathcal{D}}_e$, $\boldsymbol{\mathcal{D}}_\diamond$ and $\mathbf{\Delta}$, $\mathbf{\Delta}_\diamond$ obviously have much in common. This allows us to use, at least, some fast solvers of the same type for $\mathbf{\Lambda}$ and $\mathbf{\Lambda}_\diamond$. As an example, the wavelet multilevel solver for the preconditioner $\mathbf{\Lambda}_\diamond$ of the same type with the solver for $\mathbf{\Lambda}$, presented in Subsection 9.3.1, will be considered in Subsection 9.3.5. There is still another option. In particular, the $DD_{\rm loc}$ algorithms have simpler structure, if constructed directly for the preconditioners $\boldsymbol{\mathcal{A}}_{\diamond,I}\, \boldsymbol{\mathcal{A}}_I^{\hbar}$, which are the blocks of the preconditioners $\boldsymbol{\mathcal{A}}_\diamond\, \boldsymbol{\mathcal{A}}^{\hbar}$, related to the internal d.o.f.

9.3 Fast Preconditioner-Solvers for Internal and Face Problems

The major contributions to overall numerical complexity of DD algorithms are primarily due to solvers for local Dirichlet problems on finite elements

and solvers for local problems on faces. Although at hands there are spectrally equivalent sparse finite-difference preconditioners for the Dirichlet problems on finite elements, which we described in the preceding section, additional steps are to be made in order to obtain solvers fast uniformly in p. Face problem seems even more difficult, but in some cases it occurs possible to contemplate fast solvers for local problems on faces as well. This section contains some recent contributions first of all in the development of fast solvers for the local problems on finite elements, which were designed on the basis of the introduced preconditioners. Also we present the fast preconditioner-solvers for the internal problems on faces.

9.3.1 *Multilevel Wavelet Solver of Beuchler-Schneider-Schwab for Internal Dirichlet Problems, Hierarchical Elements*

In what follows, we use the notation $\mathbf{\Lambda}_e$ instead of $\mathbf{\Lambda}_{e,e}$ and $\mathbf{\Lambda}_{eee}$ and similar simplifications are made in the notations for the corresponding blocks of the stiffness and mass matrices of reference elements. As it was shown, the preconditioner $\mathbf{\Lambda}_e$ is represented by the sum of Kronecker products with three matrix multipliers among which only two matrices are different. In particular, they can be 1d stiffness and mass matrices, induced by 1d bilinear forms on the space of piecewise linear functions. This makes possible the following approach. Suppose, we are able to find a multilevel Riesz basis, in which the both matrices became simultaneously spectrally equivalent to their diagonals uniformly in p. Suppose also that the transformation from this basis to the source basis is cheap. This implies that we obtain two cheap factored preconditioners, which are spectrally equivalent to the 1d stiffness and mass matrices and correspond to one basis. Then, by the properties of the Kronecker product, we get a cheap transformation of $\mathbf{\Lambda}_e$ to the matrix spectrally equivalent to its diagonal. Additionally, the matrix, obtained by the backward transformation of this diagonal matrix, may be used as a preconditioner, which is spectrally equivalent to the matrix $\mathbf{\Lambda}_e$ in 3d case. The computational cost of the preconditioner depends on the cost of the forth and back transformations. A rather important for different applications fact of the *existence* of such a multiscale Riesz wavelet bases was established by Beuchler *et al.* [Beuchler *et al.* (2004)], who applied this fact on purpose of contriving an optimal solver for the matrix $\mathbf{\Lambda}_e$. Expansion of the multiscale wavelet preconditioner-solver to stiffness and mass matrices of spectral reference elements was produced in [Korneev

and Rytov (2005a)]-[Korneev and Rytov (2008)]. [Korneev *et al.* (2003a,b)] adapted fast solvers of this type to the problems on faces of the hierarchical reference elements.

Below, we formulate results on the existence of a multiscale Riesz wavelet basis with the expected properties and its application to the preconditioning in the case of complete tensor product internal subspace $\mathcal{Q}_{I,p}$ and hierarchical coordinate functions.

For simplicity, it is assumed $p = 2N + 1$ and $N + 1 = 2^{l_0}$. For each $l = 1, 2, \ldots, l_0$, we introduce the uniform mesh x_i^l of the size $\hbar_l = 2^{-l}$ on the interval $(0, 1)$ and the space $\mathcal{V}_l(0, 1)$ of the continuous piecewise linear functions on (0,1), vanishing at the ends of the interval. The dimension of $\mathcal{V}_l(0, 1)$ is $N_l = 2^l - 1$ with $N_{l_0} = N$.

Let $\phi_i^l \in \mathcal{V}_l(0, 1)$ be the the nodal basis function for the node x_i^l, so that $\phi_i^l(x_j^l) = \delta_{i,j}$ and $\mathcal{V}_l(0, 1) = \mathrm{span}[\phi_i^l]_{i=1}^{N_l}$. This basis induces the Gram matrices

$$\boldsymbol{\Delta}_l = \hbar_l \left\{ \langle (\phi_i^l)', (\phi_j^l)' \rangle_{\chi=1} \right\}_{i,j=1}^{N_l} , \quad \boldsymbol{\mathcal{M}}_l = \hbar_l^{-1} \left\{ \langle \phi_i^l, \phi_j^l \rangle_{\chi=x} \right\}_{i,j=1}^{N_l} ,$$

where

$$\langle \phi, \psi \rangle_\chi := \int_0^1 \chi^2 \phi, \psi dx \, .$$

Obviously, $\boldsymbol{\Delta}_{l_0} = \boldsymbol{\Delta}$ and we easily establish that $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}}_{l_0}$ is spectrally equivalent to $\boldsymbol{\mathcal{D}}_e^{-1}$ uniformly in N. For the reason that $\boldsymbol{\mathcal{M}}$ is a finite element matrix, as is $\boldsymbol{\Delta}$, it is more convenient for the multiresolution analysis.

The representation of each $\mathcal{V}_l$ by the sum $\mathcal{V}_l = \mathcal{V}_{l-1} \oplus \mathcal{W}_l$ results in the decomposition

$$\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \cdots \oplus \mathcal{W}_{l_0}$$

with the notations $\mathcal{V} = \mathcal{V}_{l_0}$ and $\mathcal{W}_1 = \mathcal{V}_1$.

Before to formulate the result on the existence of a special multilevel basis underlying the respective fast solver, we introduce some notations. Let $\{\psi_i^k\}$ be some *single scale basis* in $\mathcal{W}_k$ and $\Psi^l := \{\psi_i^k\}_{(i,k)\in\omega_l}$ be the *multiscale wavelet basis* composed of the single scale bases for $k = 1, 2, \ldots, l$, where ω_l is the set of the corresponding pairs of indices i, k. Since $\dim[\mathcal{W}_k] = 2^{k-1}$, we can accept that the single scale basis is $\{\psi_i^k, i = 1, 2, \ldots, 2^{k-1}\}$ and $\omega_l = \{(i, k) : i = 1, 2, \ldots, 2^{k-1},\ k = 1, 2, \ldots, l\}$. The multiscale wavelet basis induces the matrices

$$\begin{aligned} \boldsymbol{\Delta}_{\text{wavelet}} &= \left\{ \langle (\psi_i^k)', (\psi_j^l)' \rangle_1 \right\}_{(i,k),(j,l)\in\omega_{l_0}} , \\ \boldsymbol{\mathcal{M}}_{\text{wavelet}} &= \left\{ \langle \psi_i^k, \psi_j^l \rangle_x \right\}_{(i,k),(j,l)\in\omega_{l_0}} , \end{aligned} \tag{9.148}$$

and diagonal matrices

$$\begin{aligned}\mathbb{D}_1 &= \lfloor \boldsymbol{\Delta}_{\text{wavelet}} \rfloor_{\text{diag}} = \text{diag}\,[\langle (\psi_i^k)', (\psi_i^k)' \rangle_1]_{(i,k)\in\omega_{l_0}}\,,\\ \mathbb{D}_0 &= \lfloor \boldsymbol{\mathcal{M}}_{\text{wavelet}} \rfloor_{\text{diag}} = \text{diag}\,[\langle \psi_i^k, \psi_i^k \rangle_x]_{(i,k)\in\omega_{l_0}}\,.\end{aligned} \tag{9.149}$$

One of the desired properties of the wavelet bases, which can be contemplated at their construction, makes the bandwidth of the matrices $\boldsymbol{\Delta}_{\text{wavelet}}$ and $\boldsymbol{\mathcal{M}}_{\text{wavelet}}$ independent of p. It is helpful for providing that the transformations from the multiscale wavelet basis to the initial single scale finite element basis and back are fast. This assumes that the supports of the wavelet basis functions are local with respect to the corresponding scale, *i.e.,* $\text{diam}\,\text{supp}[\psi_j^l] \le c\,2^{-l}$ with an absolute positive constant c. Let us denote the transformation matrix from the multi-scale wavelet basis to the basis $\{\phi_i^{l_0}\}_{i=1}^N$ by $\mathbf{Q}$. If $\mathbf{v}$ and $\mathbf{v}_{\text{wavelet}}$ are the vectors of the coefficients of a function from $\mathcal{V}(0,1)$ in the single-scale nodal and multi-scale wavelet bases, respectively, then $\mathbf{v} = \mathbf{Q}\,\mathbf{v}_{\text{wavelet}}$.

Theorem 9.6. *There exist the wavelet bases* Ψ^{l_0} *such that the matrices* $\boldsymbol{\Delta}_{\text{wavelet}}$ *and* $\boldsymbol{\mathcal{M}}_{\text{wavelet}}$ *are simultaneously spectrally equivalent to the diagonal matrices* $\mathbb{D}_1$ *and* $\mathbb{D}_0$*, respectively, uniformly in* N*, and that the multiplications* $\mathbf{Q}\,\mathbf{v}_{\text{wavelet}}$*,* $\forall \mathbf{v} \in R^N$*, require* $\mathcal{O}(N)$ *arithmetic operations.*

Proof. We omit the proof of this important result, requiring the involved multiresolution wavelet analysis and completed in the paper by [Beuchler *et al.* (2004)]. □

This theorem immediately allows us to suggest the efficient multilevel preconditioner-solver for the typical block $\mathbf{A}_e$ of the reference element stiffness matrix.

Theorem 9.7. *Let*

$$\boldsymbol{\Lambda}_{\leftarrow w}^{-1} = \begin{cases} (\mathbf{Q}^T \otimes \mathbf{Q}^T)[\,\mathbb{D}_0 \otimes \mathbb{D}_1 + \mathbb{D}_1 \otimes \mathbb{D}_0\,]^{-1}(\mathbf{Q} \otimes \mathbf{Q}), & d = 2,\\ (\mathbf{Q}^T \otimes \mathbf{Q}^T \otimes \mathbf{Q}^T)[\mathbb{D}_0 \otimes \mathbb{D}_1 \otimes \mathbb{D}_1 + \\ \mathbb{D}_1 \otimes \mathbb{D}_0 \otimes \mathbb{D}_1 + \mathbb{D}_1 \otimes \mathbb{D}_1 \otimes \mathbb{D}_0]^{-1}(\mathbf{Q} \otimes \mathbf{Q} \otimes \mathbf{Q}), & d = 3\,, \end{cases} \tag{9.150}$$

then $\boldsymbol{\Lambda}_{\leftarrow w} \asymp \mathbf{A}_e$, i.e.,

$$\text{cond}\,[\boldsymbol{\Lambda}_{\leftarrow w}^{-1}\mathbf{A}_e] \asymp 1\,, \tag{9.151}$$

and the computational cost of the operation $\boldsymbol{\Lambda}_{\leftarrow w}^{-1}\mathbf{v}$ *for any* $\mathbf{v} \in U_I$ *is* $\mathcal{O}(N^d)$.

Proof. We consider the new preconditioner

$$\widehat{\mathbf{\Lambda}} = \begin{cases} \boldsymbol{\mathcal{M}} \otimes \boldsymbol{\Delta} + \boldsymbol{\Delta} \otimes \boldsymbol{\mathcal{M}}\,, & d=2\,, \\ \boldsymbol{\mathcal{M}} \otimes \boldsymbol{\Delta} \otimes \boldsymbol{\Delta} + \boldsymbol{\Delta} \otimes \boldsymbol{\mathcal{M}} \otimes \boldsymbol{\Delta} + \boldsymbol{\Delta} \otimes \boldsymbol{\Delta} \otimes \boldsymbol{\mathcal{M}}\,, & d=3\,, \end{cases}$$

which is the slightly modified preconditioner $\mathbf{\Lambda}$ for the sample block $\mathbf{A}_e$ of the reference element stiffness matrix. As it was mentioned above, $\boldsymbol{\Delta}_{l_0} = \boldsymbol{\Delta}$ and the matrices $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{M}}_{l_0}$ and $\boldsymbol{\mathcal{D}}_e^{-1}$ are spectrally equivalent uniformly in N. Taking additionally into account Theorem 7.3, see also the definitions of the matrices figuring in this theorem, we conclude that

$$c_1 \widehat{\mathbf{\Lambda}} \le \mathbf{A}_e \le c_2 \widehat{\mathbf{\Lambda}} \tag{9.152}$$

with absolute constants $c_k > 0$. On the other hand, $\mathbf{\Lambda}_{\leftarrow w} \asymp \widehat{\mathbf{\Lambda}}$ due to the definitions of these matrices by the Kronecker products and Theorem 9.6, whence and from (9.152) the proof follows. □

Suppose, the reference element stiffness matrix $\mathbf{A}_I$ is induced by the orthotropic subspace $\mathcal{Q}_{I,\mathbf{p}_I}$ of the internal polynomials of the powers in each variable characterized by the vector $\mathbf{p}_I = (p_{I,1}, p_{I,2}, p_{I,3})$, $p_{I,k} = 2N_k{+}1$, $N_k{+}1 = 2^{l_{0,k}}$, which is $\mathbf{p}_I = (p_{I,1}, p_{I,2})$ in the case $d = 2$. Without loss of generality, here we assume $p_{I,1} \le p_{I,2} \le p_{I,3}$ for a convenience. In the same way as $\mathbf{Q}, \mathbb{D}_0, \mathbb{D}_1$, now we obtain the matrices $\mathbf{Q} = \mathbf{Q}_k$, $\mathbb{D}_0 = \mathbb{D}_{0,k}$, $\mathbb{D}_1 = \mathbb{D}_{1,k}$, corresponding to the powers $p_{I,k}$. In the same way as Theorem 9.7, it is proved that for the preconditioner , which inverse is

$$\mathbf{\Lambda}_{\leftarrow w}^{-1} = \begin{cases} (\mathbf{Q}_1^T \otimes \mathbf{Q}_2^T)\,[\,\mathbb{D}_{0,1} \otimes \mathbb{D}_{1,2} + \mathbb{D}_{1,1} \otimes \mathbb{D}_{0,2}\,]^{-1}\,(\mathbf{Q}_1 \otimes \mathbf{Q}_2)\,, & d=2\,, \\ (\mathbf{Q}_1^T \otimes \mathbf{Q}_2^T \otimes \mathbf{Q}_3^T)\,[\,\mathbb{D}_{0,1} \otimes \mathbb{D}_{1,2} \otimes \mathbb{D}_{1,3} + \mathbb{D}_{1,1} \otimes \mathbb{D}_{0,2} \otimes \mathbb{D}_{1,3} + \\ \qquad \mathbb{D}_{1,1} \otimes \mathbb{D}_{1,2} \otimes \mathbb{D}_{0,3}\,]^{-1}(\mathbf{Q}_1 \otimes \mathbf{Q}_2 \otimes \mathbf{Q}_3)\,, & d=3\,, \end{cases}$$

one also has

$$\operatorname{cond}\,[\mathbf{\Lambda}_{\leftarrow w}^{-1}\mathbf{A}_e] \prec 1\,.$$

The authors of [Beuchler *et al.* (2004)] presented several examples of wavelets satisfying Theorem 9.6. In support of the conclusions of analysis, good results of numerical experiments with a number of wavelet bases were reported. Additionally the experiments showed that wavelets of a wider set, than restricted by the conditions arising in the theoretical analysis, can lead to fast peconditioner-solvers of the $\mathbf{\Lambda}_{\leftarrow w}$ type.

9.3.2 *Multiresolution Wavelet Solver for Faces*

For the Dirichlet problems on finite elements, some other solvers might be competitive to the wavelet multiresolution solver, associated with the preconditioners $\mathbf{\Lambda}_{\leftarrow w}$. An additional good feature of the latter is that it can be easily adapted to the problems on faces. Following to [Korneev *et al.* (2003a,b)], first we define the multiscale wavelet preconditioner (9.150) for the internal mass matrix of 2d reference element. Then we use the K-method of interpolation between the diagonal internal mass and stiffness matrices, expressed in the multiscale wavelet basis, in order to obtain the diagonal matrix of the quadratic form ${}_{00}|\cdot|^2_{1/2,F_k}$ in the multiscale wavelet basis. Finally we transform the obtained diagonal matrix to the initial basis and come to the anticipated face preconditioner-solver.

We consider the mass matrix $\mathbf{M}_I$ for the 2d isotropic reference element of the power p. According to Lemma 7.5, the matrix

$$\mathbb{M}_{\vartriangle} = \mathrm{diag}\left[\mathbb{M}_e, \mathbb{M}_e, \mathbb{M}_e, \mathbb{M}_e\right], \qquad \mathbb{M}_e = \mathbf{\Delta} \otimes \mathbf{\Delta},$$

is a good preconditioner for $\mathbf{M}_I$, which satisfies the inequalities

$$\mathbb{M}_{\vartriangle} \prec \mathbf{M}_I \prec \mathbb{M}_{\vartriangle}. \tag{9.153}$$

At the same time, an additional to Theorem 9.7 consequence of Theorem 9.6, is the spectral equivalence

$$\mathbb{M}_e^{-1} \asymp \left(\mathbf{Q}^T \otimes \mathbf{Q}^T\right)\left(\mathbb{D}_1 \otimes \mathbb{D}_1\right)^{-1}\left(\mathbf{Q} \otimes \mathbf{Q}\right). \tag{9.154}$$

Let us denote the entries of the diagonal matrices $\mathbb{D}_0$ and $\mathbb{D}_1$ via $d_{0,i}$ and $d_{1,i}$, respectively. Diagonal entries of the matrix $\mathbb{D}_{\Delta,e} := \mathbb{D}_0 \otimes \mathbb{D}_1 + \mathbb{D}_1 \otimes \mathbb{D}_0$ are $d^{(1)}_{i,j} = d_{1,j}d_{0,i} + d_{1,i}d_{0,j}$, the diagonal matrix $\mathbb{D}_{M,e} := \mathbb{D}_1 \otimes \mathbb{D}_1$ has for the entries on diagonal $d^{(o)}_{i,j} = d_{1,i}d_{1,j}$, and we define the diagonal matrix $\mathbb{D}_{1/2}$ with the entries on diagonal

$$d^{(1/2)}_{i,j} = d_{1,i}d_{1,j}\sqrt{\frac{d_{0,i}}{d_{1,i}} + \frac{d_{0,j}}{d_{1,j}}}.$$

Now for one face Schur complement preconditioner we may take $\boldsymbol{\mathcal{S}}_{00} = \boldsymbol{\mathcal{S}}_{\leftarrow w}$ defined by its inverse as

$$\boldsymbol{\mathcal{S}}^{-1}_{\leftarrow w} = \left(\mathbf{Q}^T \otimes \mathbf{Q}^T\right)\mathbb{D}^{-1}_{1/2}\left(\mathbf{Q} \otimes \mathbf{Q}\right). \tag{9.155}$$

In turn, at the appropriate identical ordering of degrees of freedom of each of the faces, the preconditioner $\boldsymbol{\mathcal{F}}$ in (9.48) may be defined as the block diagonal matrix

$$\boldsymbol{\mathcal{F}} = \mathrm{diag}\left[\boldsymbol{\mathcal{S}}_{\leftarrow w}, \boldsymbol{\mathcal{S}}_{\leftarrow w}, ..., \boldsymbol{\mathcal{S}}_{\leftarrow w}\right] \tag{9.156}$$

with 6 identical independent blocks $\boldsymbol{\mathcal{S}}_{\leftarrow w}$.

Theorem 9.8. *Let $\mathcal{U} = \mathcal{Q}_p$. Then, for the preconditioner $\boldsymbol{\mathcal{S}}_{00} = \boldsymbol{\mathcal{S}}_{\leftarrow w}$, the inequalities (9.47) hold with $\underline{\gamma}_F, \overline{\gamma}_F \asymp 1$, and, therefore, we have* $\mathrm{cond}\,[\boldsymbol{\mathcal{S}}_{00}^{-1}\mathbb{S}_{00}] = \mathrm{cond}\,[\boldsymbol{\mathcal{S}}_{\leftarrow w}^{-1}\mathbb{S}_{00}] \asymp 1$. *The computational cost of the multiplications $\boldsymbol{\mathcal{S}}_{\leftarrow w}^{-1}\mathbf{v}$ is optimal,* i.e., $\mathcal{O}(N^2)$.

Proof. We need to consider one face, *e.g.,* the face F_3, for which the space of the traces $\mathcal{U}_{F_3}(F_3)$ coincides with the space $\mathcal{U}_I$ for the 2d reference element. We consider the latter space, the corresponding vector space U_I, and the vector space $U_{I,w}$ of the coefficients of expansions in the basis, which is the tensor product of 1d multiscale Riesz wavelet bases on $\tau_0 = (-1,1)\times(-1,1)$. If $v \in \mathcal{U}_I$ and $\mathbf{v}_w \in U_{I,w}$ are such that $v \leftrightarrow \mathbf{v} = (\mathbf{Q}\otimes\mathbf{Q})\mathbf{v}_w$, then

$$\begin{aligned} &\mathbf{v}_w^\top \mathbb{D}_{M,e}\mathbf{v}_w \asymp \|v\|^2_{0,\tau_0} \asymp \mathbf{v}_w^\top \mathbb{D}_{M,e}\mathbf{v}_w\,, \\ &\mathbf{v}_w^\top \mathbb{D}_{\Delta,e}\mathbf{v}_w \asymp |v|^2_{1,\tau_0} \asymp \mathbf{v}_w^\top \mathbb{D}_{\Delta,e}\mathbf{v}_w\,. \end{aligned} \tag{9.157}$$

These inequalities are direct consequences of the definitions of the matrices $\mathbb{D}_{M,e}$ and $\mathbb{D}_{\Delta,e}$, see (9.150), (9.152), (9.153) and (9.154).

It is known that for functions vanishing on $\partial\tau_0$ the norm ${}_{00}|\cdot|^2_{1/2,\tau_0}$ can be obtained by the method of K-interpolation between the spaces $L_2(\tau_0)$ and $\mathring{H}^1(\tau_0)$. At the same time (9.157) means that $\|\cdot\|_{\mathbb{D}_{M,e}}$ and $\|\cdot\|_{\mathbb{D}_{\Delta,e}}$ are the norms in $L_2(\tau_0)$ and $\mathring{H}^1(\tau_0)$ for the FE functions represented in the tensor product multiscale wavelet basis. Therefore, the interpolation between the spaces is reduced to the interpolation between the matrices $\mathbb{D}_{M,e}$ and $\mathbb{D}_{\Delta,e}$, which results in the matrix $\mathbb{D}_{1/2}$ defined above. By its definition, we have

$$\mathbf{v}_w^\top \mathbb{D}_{1/2}\mathbf{v}_w \asymp {}_{00}|v|^2_{1/2,\tau_0} \asymp \mathbf{v}_w^\top \mathbb{D}_{1/2}\mathbf{v}_w\,. \tag{9.158}$$

Since $\mathbb{S}_{00}$ is the matrix of the quadratic form, standing in the middle, the spectral equivalence bounds of Theorem hold. The computational work for calculating the vector $\boldsymbol{\mathcal{S}}_{\leftarrow w}^{-1}\mathbf{v},\ \forall\,\mathbf{v}\in U_I,$ is evidently of the same order as for the operations $\boldsymbol{\Lambda}_{\leftarrow w}^{-1}\mathbf{v},\ \forall\,\mathbf{v}\in U_I,$ in 2d, *i.e.,* $\mathcal{O}(N^2)$. □

It is clear, that in the same way one can define the matrix $\boldsymbol{\mathcal{S}}_{\leftarrow w}$ spectrally equivalent to the matrix of quadratic form ${}_{00}|v|^2_{1/2,\tau_0}$ for the case of the orthotropic space $\mathcal{U}_I(\tau_0),\ \tau_0 = (-1,1)^2$. However, this matrix can be used for the face preconditioner, if the 3d space $\mathcal{U}_I(\tau_0),\ \tau_0 = (-1,1)^3,$ of the reference element is sufficiently reach, *e.g.,* minimal power $p_{I,1}$ for the internal polynomials is not less than maximal power for the face.

9.3.3 *Domain Decomposition Solver for Hierarchical Reference Elements*

The domain decomposition solver for the 2d reference element internal stiffness matrices, presented in Subsection 8.1.2, seems to be difficult for generalization on 3d problems. The reason is essential difference in the source FD/FE preconditioners. Whereas in 2d they are discretizations of 2nd order elliptic equations, in 3d they are discretizations of 4th order equations, containing only mixed derivatives, see for 3d (9.160)–(9.162) below. Nevertheless, numerical experiments support the conjecture, that preconditioner of DD type can be efficient for 3d problems. As earlier, we will use for such a preconditioner-solver an abbreviation $DD_{\rm loc}$, reckoning that in the context of this paper it is assumed for the use as a secondary solver in the leading DD solver for an *hp* discretization.

Let us also note that in the case of spectral reference element the source FD/FE preconditioners are discretizations of the Poisson equation. Therefore, DD preconditioner-solver is easier constructed and its theoretical justification seem more feasible.

9.3.3.1 *Decomposition Mesh and Finite-Difference Preconditioner with Piecewise Constant Coefficients*

Having obtained the FD preconditioner $\mathbf{\Lambda}$, we are able to make further simplifications for better adaptation of the FD/FE preconditioner to $DD_{\rm loc}$ preconditioner-solver, which we are going to describe very briefly. Like to 2d case, the matrix $\mathbf{B}$ of the $DD_{\rm loc}$ preconditioner, which satisfy the demand to be spectrally equivalent to $\mathbf{\Lambda}$, will be obtained by replacing coefficients in the finite-difference operator (9.137) by piecewise constant ones on a special imbedded *coarse decomposition mesh*. Obviously, such a mesh should reflect the rate of deterioration of coefficients and, therefore, have to be nonuniform. Indeed, it is defined as the tensor product of 1d nonuniform meshes described in Subsection 8.1.2. For the convenience, we remind the definition of the coarse mesh and use the both types of notations for the space variables $\xi = (\xi_1, \xi_2, \xi_3) = (x, y, z)$.

Let $\xi_{s,i} = \eta_i = i\hbar$ be the coordinates of the planes of the fine mesh, where $s = 1, 2, 3$ are the numbers of the corresponding axes ξ_s and $i = 0, 1, \ldots, N+1$. The coordinates of planes of the decomposition mesh, which are denoted by $\overline{\xi}_{s,l} = \zeta_l$, coincide with some planes of the fine mesh and are defined by the same as in the case of 2d formulas (8.39), depending on

two parameters q and n_0. For convenience, we give these formulas here:

$$\zeta_0 = 0, \quad \zeta_{l_0} = 1,$$

$$\zeta_l = \eta_i, \quad \text{for } i = \gamma(l) := \operatorname{int} \lfloor (q^l - 1) n_0 \rfloor,\ l = 1, 2, \ldots, \ell_0 - 2, \tag{9.159}$$

$$\zeta_{\ell_0 - 1} = \begin{cases} \eta_{\gamma(\ell_0 - 1)}, & \text{if } \exists\, l = \ell_0 :\ \eta_{\gamma(\ell_0)} = 1, \\ \eta_i, & \text{if } \exists\, l = \ell_0 :\ \eta_{\gamma(\ell_0 - 1)} < 1,\ \eta_{\gamma(\ell_0)} > 1\,, \end{cases}$$

where the index i in the very last formula is given by expression

$$i = \operatorname{int} \lfloor \frac{\gamma(\ell_0 - 2) + N + 1}{2} \rfloor.$$

For simplicity, we assume that $\gamma(1) \geq 2$, if a different setting is not specially mentioned. Under this assumption, it follows from the second line of (9.159) that $q, n_0 \geq 2$. Clearly, the bounds (8.40) and (8.41) remain valid.

The cells of the decomposition mesh are the parallelepipeds

$$\delta_{\Bbbk} = \{\xi : \zeta_{k_s - 1} < \xi_s < \zeta_{k_s},\ \ s = 1, 2, 3\}\,, \quad 1 \leq k_s \leq \ell_0\,,$$

faces $F_{\Bbbk, s, t}$ of which lie in the planes $\xi_s \equiv \overline{\xi}_{s, k_s + t}$, $t = 0, 1$, where $\Bbbk = (k_1, k_2, k_3)$. Let us introduce the piecewise constant function

$$\psi(\zeta) = \begin{cases} \psi_l \equiv \frac{1}{2}(\zeta_{l-1}^2 + \zeta_l^2), & \zeta \in (\zeta_{l-1}, \zeta_l)\,, \\ \psi_1 \equiv \frac{1}{2}(\hbar^2 + \zeta_1^2), & \zeta \in (\zeta_0, \zeta_1), \end{cases}$$

and, similarly to (9.137), the matrix $\hbar^{-2}\mathbf{B}$ arising from the finite-difference discretization of the differential equation

$$\psi(z) u_{xxyy} + \psi(y) u_{xxzz} + \psi(x) u_{yyzz} = f(x, y, z)\,, \quad (x, y, z) \in \pi_1\,, \tag{9.160}$$

under the homogeneous Dirichelet boundary conditions $u = 0$ on $\partial\pi_1$. Both matrices $\mathbf{B}$ and $\mathbf{\Lambda}$ have a nineteen points stencils and are spectrally equivalent uniformly in p, as stated in Lemma 9.10 below.

In the analysis and sometimes in applications, it is more convenient to deal with the FE matrix for the problem (9.160), which is denoted by $\mathbf{B}_{\,\mathrm{fem}}$. Let $\mathring{\mathcal{H}}(\pi_1)$ be the space of functions, which are continuous on π_1, zero on $\partial\pi_1$, and trilinear on each nest of the fine mesh. Then $\mathbf{B}_{\,\mathrm{fem}}$ is the FE matrix generated by the bilinear form

$$a_{\pi_1}(u, v) = \frac{1}{\hbar} \int\limits_{\pi_1} [\, \psi(z) u_{,xy} v_{,xy} + \psi(y) u_{,xz} v_{,xz} + \psi(x) u_{,yz} v_{,yz}\,]\, dx\, dy\, dz \tag{9.161}$$

on the space $\mathring{\mathcal{H}}(\pi_1)$ by means of the nodal basis. It is easy to see that this matrix is well defined. At the same time, it is clear that, in the contrast with the 2d case, the space of the piecewise linear functions cannot be used.

Lemma 9.10. *Any of the eight independent blocks* $\mathbf{A}_{a,b,c}$, $a,b,c = o,e$, *of the internal stiffness matrix* $\mathbf{A}_I$, *which are defined in (9.128) and (9.129), is spectrally equivalent to matrices* $\mathbf{B}$ *and* $\mathbf{B}_{\rm fem}$. *In particular,*

$$\underline{\gamma}_B \mathbf{B} \le \mathbf{A}_e \le \overline{\gamma}_B \mathbf{B} \quad \text{and} \quad \underline{\gamma}_B \mathbf{B}_{\rm fem} \le \mathbf{A}_e \le \overline{\gamma}_B \mathbf{B}_{\rm fem}$$

with the positive constants $\underline{\gamma}_B$ *and* $\overline{\gamma}_B$ *depending only on* n_0 *and* q.

Proof. The proof will be completed by comparing the matrices $\mathbf{B}$ and $\mathbf{B}_{\rm fem}$ with the matrices $\mathbf{\Lambda}$ and $\mathbf{\Lambda}^{(e)}_{\rm fem}$, respectively, and by applying Theorem 9.5. For instance, we have the expression

$$\mathbf{B} = \widetilde{\boldsymbol{\mathcal{D}}}_e \otimes \boldsymbol{\Delta} \otimes \boldsymbol{\Delta} + \boldsymbol{\Delta} \otimes \widetilde{\boldsymbol{\mathcal{D}}}_e \otimes \boldsymbol{\Delta} + \boldsymbol{\Delta} \otimes \boldsymbol{\Delta} \otimes \widetilde{\boldsymbol{\mathcal{D}}}_e \,, \tag{9.162}$$

which is similar to the expression (9.131) for $\mathbf{\Lambda}$. It may be obtained from the latter by replacing $\boldsymbol{\mathcal{D}}_e$ by the matrix

$$\widetilde{\boldsymbol{\mathcal{D}}}_e = \frac{1}{2\hbar^2} \operatorname{diag} \left[\psi(\eta_i - 0) + \psi(\eta_i + 0)\right]_{i=1}^{N} .$$

Therefore, by virtue of Theorem 9.5 and the representations (9.162) and (9.131) of the matrices $\mathbf{B}$ and $\mathbf{A}_e$, respectively, the proof for these matrices is reduced to the comparison of the matrices $\boldsymbol{\mathcal{D}}_e$ and $\widetilde{\boldsymbol{\mathcal{D}}}_e$, which are obviously spectrally equivalent uniformly in p.

Since the first pair of the inequalities in Lemma 9.10 holds, the proof of the other pair is reduced to the proof of the spectral equivalence of $\mathbf{B}_{\rm fem}$ and $\mathbf{B}$. In this regard it may be observed that the latter matrix can be interpreted as a FE matrix assembled from the matrices for some finite elements specified on the square nests of the fine mesh. Thus, the spectral equivalence of $\mathbf{B}_{\rm fem}$ and $\mathbf{B}$ can be proved by the elementwise comparison of the stiffness matrices of the respective finite elements. We omit the details of these considerations. □

For the preconditioners of the internal stiffness matrices of an *hp* discretization, we may take

$$\mathbf{B}_I = \operatorname{diag} [\underbrace{\mathbf{B}, ..., \mathbf{B}}_{8 \text{ times}}] \quad \text{and} \quad \mathbf{B}_{I,\rm fem} = \operatorname{diag} [\underbrace{\mathbf{B}_{\rm fem}, ..., \mathbf{B}_{\rm fem}}_{8 \text{ times}}] . \tag{9.163}$$

Due to the above Lemma and the generalized conditions of the shape regularity, we have the spectral inequalities

$$\underline{\gamma}_B h^{(r)} \varrho_r \mathbf{B}_I \le \mathbf{K}_{I_r} \le \overline{\gamma}_B h^{(r)} \varrho_r \mathbf{B}_I$$

and

$$\underline{\gamma}_B h^{(r)} \varrho_r \mathbf{B}_{I,\mathrm{fem}} \le \mathbf{K}_{I_r} \le \overline{\gamma}_B h^{(r)} \varrho_r \mathbf{B}_{I,\mathrm{fem}}.$$

This means that we can set $\boldsymbol{\mathcal{A}}_I = \mathbf{B}_I$ or $\boldsymbol{\mathcal{A}}_I = \mathbf{B}_{I,\mathrm{fem}}$ in (9.28), and, respectively, $\boldsymbol{\mathcal{K}}_{I_r} = h^{(r)} \varrho_r \mathbf{B}_I$ or $\boldsymbol{\mathcal{K}}_{I_r} = h^{(r)} \varrho_r \mathbf{B}_{I,\mathrm{fem}}$ in (9.20).

9.3.3.2 *Main Components of DD_{loc} Algorithms*

Let us suppose that, in the DD preconditioner, the matrix $\boldsymbol{\mathcal{A}}_I = \mathbf{B}_I$ is used for preconditioning internal Dirichlet problems on finite elements. In view of (9.163), at each iterative step of the leading PCGM with the DD preconditioner, the solution of systems with the matrix $\mathbf{B}_I$ is reduced to solving systems of the form: given $\mathbf{d} \in U^e$, find $\mathbf{v} \in U^e$ such that

$$\mathbf{B}\mathbf{v} = \mathbf{d}, \tag{9.164}$$

where $U^e = U_{I_e}$ is the subspace of vectors, corresponding to the coordinate polynomials $L_{\boldsymbol{\alpha}}$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$, with all α_k, $k = 1, 2, 3$, even, see Subsection 9.2.1. For solving these systems we will implement the secondary domain decomposition method designated by DD_{loc}. In spite of the differences between the bilinear forms (8.1) and (9.161) related to the DD_{loc} preconditioners in 2d and 3d cases, we can attempt to design the DD_{loc} solver in 3d, as an expansion to 3d of the one considered in Subsection 8.1.2. It is clear though that theoretical substantiations of the steps at the derivation of the Schur complement preconditioners would require considerably harder efforts. We do not produce them here and, basically, present a version of the substructuring algorithm only.

Alongside with the multiindex $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ ordering of the d. o. f., which is "global" for the problem (9.164), we will use other indices reflecting the structure of the decomposition in the basis of DD_{loc} algorithm. The vector space $U^e = R^{N^3}$ can be decomposed in the direct sum $U^e = U_{\mathrm{i}} \oplus U_{\mathrm{f}} \oplus U_{\mathrm{w}}$ of three subspaces U_{i}, U_{f} and U_{w}, where

- $U_{\mathrm{i}} = \bigcup_{s,k_s=1}^{3,\,\ell_0} U_{\mathrm{i},\Bbbk}$ is the union of the subsubspaces $U_{\mathrm{i},\Bbbk}$ of the d. o. f. which are internal for each subdomain $\delta_{\Bbbk}$ of the decomposition,
- U_{f} is the subspace of the internal d. o. f. for faces of the decomposition,
- U_{w} is the subspace of the wire basket d. o. f., *i.e.,* d. o. f. living on the edges and at the vertices of the decomposition.

In the correspondence to this structure of the space U^e, the matrix and the right-hand side vector of the system (9.164) can be written in the block

form

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\rm i} & \mathbf{B}_{\rm if} & \mathbf{B}_{\rm iw} \\ \mathbf{B}_{\rm fi} & \mathbf{B}_{\rm f} & \mathbf{B}_{\rm fw} \\ \mathbf{B}_{\rm wi} & \mathbf{B}_{\rm wf} & \mathbf{B}_{\rm w} \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} \mathbf{v}_{\rm i} \\ \mathbf{v}_{\rm f} \\ \mathbf{v}_{\rm w} \end{pmatrix}, \quad \mathbf{d} = \begin{pmatrix} \mathbf{d}_{\rm i} \\ \mathbf{d}_{\rm f} \\ \mathbf{d}_{\rm w} \end{pmatrix}, \tag{9.165}$$

where obviously

$$\mathbf{B}_{\rm i} = \operatorname{diag} [\mathbf{B}_{{\rm i},\Bbbk}]_{k_1,k_2,k_3=1}^{\ell_0}$$

are the independent blocks $\mathbf{B}_{{\rm i},\Bbbk}$ corresponding to subdomains $\delta_\Bbbk$. Below we list the most important facts concerning the *DD* loc solver for the system (9.164)–(9.165):

1) Matrices of local discrete Dirichlet problems have the form

$$\begin{aligned} \mathbf{B}_{{\rm i},\Bbbk} = {} & \psi_{k_1} \mathbf{I}_{n_1} \otimes \mathbf{\Delta}_{n_2} \otimes \mathbf{\Delta}_{n_2} + \psi_{k_2} \mathbf{\Delta}_{n_1} \otimes \mathbf{I}_{n_2} \otimes \mathbf{\Delta}_{n_3} + \\ & \psi_{k_3} \mathbf{\Delta}_{n_1} \otimes \mathbf{\Delta}_{n_2} \otimes \mathbf{I}_{n_3}, \end{aligned} \tag{9.166}$$

where $\mathbf{\Delta}_{n_s} = \operatorname{tridiag} [-1, 2, -1]$ and $\mathbf{I}_{n_s} = \operatorname{diag} [1]$ are $(n_s - 1) \times n_s - 1)$ matrices, and $n_s = n_s(k_s)$ is the number of fine mesh intervals on the edge of $\delta_\Bbbk$ parallel to the axis ξ_s. The corresponding blocks $\mathbf{B}_{{\rm i},\Bbbk,{\rm fem}}$ of the matrix $\mathbf{B}_{\rm fem}$ are induced by the bilinear forms

$$\begin{aligned} a_{\delta_\mathbf{k}}(u, v) = {} & \tfrac{1}{h} \int\limits_{\delta_\Bbbk} [\, \psi_{k_3}(z) u_{xy} v_{xy} + \psi_{k_2}(y) u_{xz} v_{xz} + \\ & \psi_{k_1}(x) u_{yz} v_{yz} \,]\, dx\, dy\, dz\,, \end{aligned} \tag{9.167}$$

and are represented in the same form as (9.166), but with the unity matrices replaced with the tridiagonal ones. Clearly, the systems with system matrices $\mathbf{B}_{{\rm i},\Bbbk}$ can efficiently be solved by means of the 3d FDFT with the total computational cost $\mathcal{O}(N^3 \log N)$ for all $\Bbbk$. In a more efficient way, especially for the subdomains $\delta_\Bbbk$ with high aspect ratios of the edges, we can use the combination of the 2d FDFT in the two shortest directions and the Gauss elimination for the tridiagonal systems (Thomas algorithm also called progonka in the Russian literature) in the longest one. Besides, some optimal, *i.e.,* requiring $\mathcal{O}(N^3)$ arithmetic operations, multilevel methods may also be implemented, *e.g.,* in form of a BPX type preconditioning. The multilevel method described in Subsection 9.3.1 is an example which is clearly applicable here. It becomes much simpler for systems with the system matrices $\mathbf{B}_{{\rm i},\Bbbk}$ because of the constant coefficients in the bilinear form (9.167), corresponding to such matrices. In the literature there are many other suggestions which can be used in optimal solvers for systems with matrices $\mathbf{B}_{{\rm i},\Bbbk}$. As usual, if some secondary iterative solver is used for solving systems with matrices $\mathbf{B}_{{\rm i},\Bbbk}$, then the most efficient way is to use it as an inexact iterative solver.

2) The space U^e can be represented as the sum $U^e = U_\mathrm{i} \oplus U_\mathrm{f,tr} \oplus U_\mathrm{w}$, where the subspace $U_\mathrm{f,tr} = U_\mathrm{f}$ is spanned by the basis which may be called the *piecewise discrete harmonic trigonometric basis*. The elements of this basis are uniquely defined by the following two properties:

i) $U_\mathrm{f,tr} \perp_\mathbf{B} U_\mathrm{i}$, *i.e.*, the subspace $U_\mathrm{f,tr}$ is orthogonal to U_i with respect to the scalar product $(\cdot,\cdot)_\mathbf{B} := (\mathbf{B}\,\cdot,\cdot)$.

ii) The trace of each basis vector on the interface boundary is nonzero only on one face, and, on each face $F_{\Bbbk,s,t}$, these traces are the vectors which are denoted as $\boldsymbol{\mu}^t_{s,\alpha,\beta}$ and have the components

$$\boldsymbol{\mu}^t_{s,\beta,\gamma}\,\big|_{F_{\Bbbk,s,t}} = \left\{ \sin(\frac{\beta i\pi}{n_{s+1}}) \sin(\frac{\gamma j\pi}{n_{s+2}}) \right\}_{i,j=0}^{n_{k_{s+1}},n_{k_{s+2}}} .$$

Here $s = 1, 2, 3$ stands for the axis ξ_s, to which the face

$$\xi_s \equiv \xi_{s,l}, \quad l = k_s - 1 + t, \quad t = 0, 1,$$

is orthogonal, and $s+1$ as well as $s+2$ are understood modulo 3. Further, $\beta = 1, \ldots, n_{s+1} - 1$ and $\gamma = 1, \ldots, n_{s+2} - 1$ are local indices of the face basis vectors and the face internal nodes, whereas $i = 0, 1, \ldots, n_{s+1}$ and $j = 0, 1, \ldots, n_{s+2}$ are local indices of all the face nodes, including its boundary nodes.

It can be shown that a consequence of the first two properties are the next two ones:

iii) The basis vectors of $U_\mathrm{f,tr}$ corresponding to any face are orthogonal in the sense of the scalar product $(\cdot,\cdot)_\mathbf{B}$. However, basis vectors for adjacent faces are not orthogonal. Additionally, each face basis vector is coupled with a single basis vector of the opposite face, and the former basis vector has the same local indices for the face, see also ii). Namely, for any fixed s and any β, γ in the ranges pointed out in ii), the vector $\boldsymbol{\mu}^t_{s,\beta,\gamma}$ is only coupled with $\boldsymbol{\mu}^{1-t}_{s,\beta,\gamma}$.

iv) After the transformation to the piecewise discrete-harmonic trigonometric basis, system (9.165) gets the form

$$\widetilde{\mathbf{B}}\,\widetilde{\mathbf{v}} = \widetilde{\mathbf{d}}\,, \tag{9.168}$$

with

$$\widetilde{\mathbf{B}} = \begin{pmatrix} \mathbf{B}_\mathrm{i} & \mathbf{O} & \mathbf{B}_\mathrm{iw} \\ \mathbf{O} & \mathbf{S}_\mathrm{f,tr} & \mathbf{B}_\mathrm{fw,tr} \\ \mathbf{B}_\mathrm{wi} & \mathbf{B}_\mathrm{wf,tr} & \mathbf{B}_\mathrm{w} \end{pmatrix}, \quad \widetilde{\mathbf{v}} = \begin{pmatrix} \mathbf{v}_\mathrm{i} \\ \mathbf{v}_\mathrm{f,tr} \\ \mathbf{v}_\mathrm{w} \end{pmatrix}, \quad \widetilde{\mathbf{d}} = \begin{pmatrix} \mathbf{d}_\mathrm{i} \\ \mathbf{d}_\mathrm{f,tr} \\ \mathbf{d}_\mathrm{i} \end{pmatrix}. \tag{9.169}$$

Let us turn to the block

$$\widetilde{\mathbf{B}}_\mathrm{b,tr} = \begin{pmatrix} \mathbf{S}_\mathrm{f,tr} & \mathbf{B}_\mathrm{fw,tr} \\ \mathbf{B}_\mathrm{wf,tr} & \mathbf{B}_\mathrm{w} \end{pmatrix},$$

which is the restriction of the matrix $\widetilde{\mathbf{B}}$ to the subspace $U_{\rm b} := U_{\rm f,trig} \oplus U_{\rm w}$. It is important to see that the entries of $\mathbf{S}_{\rm f,tr}$, as in the 2d case, can be directly calculated with the use of explicit expressions. It can also be expected that the computation of these nonzero entries by means of the explicit expressions costs not more than $\mathcal{O}(N^3)$ arithmetic operations. The blocks $\mathbf{B}_{\rm fw}$ and $\mathbf{B}_{\rm fw,tr}$ have the dimension $\mathcal{O}(N) \times \mathcal{O}(N \log^2 N)$. Therefore, $\mathbf{B}_{\rm fw,tr}$ can be obtained from $\mathbf{B}_{\rm fw}$ by FDFT for $\mathcal{O}(N^3 \log^3 N)$ a.o. However, in the version of *DD* $_{\rm loc}$ algorithm, which we here consider, the operation of the calculation of the matrix $\mathbf{B}_{\rm fw,tr}$ can be avoided. We only need the matrix-vector mutiplication by $\mathbf{B}_{\rm fw}$, and then the application of the FDFT to the obtained vector. These operations are cheaper by one order of N.

3) Two versions of the *DD* $_{\rm loc}$ preconditioner-solver can be considered. In the first version, we neglect coupling between the internal subdomain d. o. f. and the wire basket d. o. f., and set

$$\boldsymbol{\mathcal{B}} = \begin{pmatrix} \mathbf{B}_{\rm i} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\mathcal{S}}_{\rm tr} \end{pmatrix},$$

where $\boldsymbol{\mathcal{S}}_{\rm tr}$ is defined by means of the factorized representation of its inverse

$$\boldsymbol{\mathcal{S}}_{\rm tr}^{-1} = \begin{pmatrix} \mathbf{I}_{\rm f} & -\mathbb{S}_{\rm f,it}^{-1}\mathbf{B}_{\rm fw,tr} \\ \mathbf{O} & \mathbf{I}_{\rm w} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{S}}_{\rm f,tr}^{-1} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{B}}_{\rm w}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\rm f} & \mathbf{O} \\ -\mathbf{B}_{\rm wf,tr}\mathbb{S}_{\rm f,it}^{-1} & \mathbf{I}_{\rm w} \end{pmatrix}. \qquad (9.170)$$

Here $\mathbb{S}_{\rm f,it}$ and $\boldsymbol{\mathcal{S}}_{\rm f,tr}$ are some preconditioners for $\mathbf{S}_{\rm f,tr}$, whereas $\boldsymbol{\mathcal{B}}_{\rm w}$ is a preconditioner for $\mathbf{B}_{\rm w}$. We can safely take $\boldsymbol{\mathcal{B}}_{\rm w} = \mathbf{B}_{\rm w}$. Indeed, the dimension of $\mathbf{B}_{\rm w}$ is $\mathcal{O}(N(\log N)^2)$. Therefore, even under the assumption that it is completely filled in and the classical Gausiian elimination procedure is used, the solution of each system with the system matrix $\mathbf{B}_{\rm w}$ will cost not more than $\mathcal{O}(N^3 \log^6 N)$. Preconditioners for $\mathbf{S}_{\rm f,tr}$ will be discussed below.

For outlining another version, we look at the factorization

$$\widetilde{\mathbf{B}}^{-1} = \begin{pmatrix} \mathbf{I}_{\rm i} & -\mathbf{B}_{\rm i}^{-1}\mathbf{B}_{\rm ib} \\ \mathbf{O} & \mathbf{I}_{\rm b} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{\rm i}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{S}_{b,\rm tr}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\rm i} & \mathbf{O} \\ -\mathbf{B}_{\rm bi}\mathbf{B}_{\rm i}^{-1} & \mathbf{I}_{\rm b} \end{pmatrix},$$

with $\mathbf{B}_{\rm b,i}^T = (\mathbf{O}^T, \mathbf{B}_{\rm wi}^T)$, and consider the matrix-vector multiplication by $\widetilde{\mathbf{B}}^{-1}$. We conclude that all operations, except for solving the system with the system matrix $\mathbf{S}_{\rm b,tr}$, may be performed by means of $\mathcal{O}(N^3)$ or $\mathcal{O}(N^3 \log N)$ a.o. This is because of the reasons pointed out in i) and of the sparsity of the matrix $\mathbf{B}_{\rm i,w}$. Therefore, deriving the fast *DD* $_{\rm loc}$ solver is reduced to designing a fast preconditioner-solver for the matrix

$$\mathbf{S}_{\rm tr} = \begin{pmatrix} \mathbf{S}_{\rm f,tr} & \mathbf{B}_{\rm fw,tr}^{(1)} \\ \mathbf{B}_{\rm wf,tr}^{(1)} & \mathbf{B}_{\rm w}^{(1)} \end{pmatrix},$$

with $\mathbf{B}_{\mathrm{w}}^{(1)} = \mathbf{B}_{\mathrm{w}} - \mathbf{B}_{\mathrm{wi}}\mathbf{B}_{\mathrm{i}}^{-1}\mathbf{B}_{\mathrm{iw}}$. The main components of the algorithm are represented by the expressions

$$\boldsymbol{\mathcal{B}}_{\star}^{-1} = \begin{pmatrix} \mathbf{I}_{\mathrm{i}} & -\mathbf{B}_{\mathrm{i}}^{-1}\mathbf{B}_{\mathrm{ib}} \\ \mathbf{O} & \mathbf{I}_{\mathrm{b}} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{\mathrm{i}}^{-1} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{S}}_{\mathrm{tr}\star}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\mathrm{i}} & \mathbf{O} \\ -\mathbf{B}_{\mathrm{bi}}\mathbf{B}_{\mathrm{i}}^{-1} & \mathbf{I}_{\mathrm{b}} \end{pmatrix}, \qquad (9.171)$$

with $\mathbf{B}_{\mathrm{b,i}}^{T} = (\mathbf{O}^{T}, \mathbf{B}_{\mathrm{wi}}^{T})$, and

$$\boldsymbol{\mathcal{S}}_{\mathrm{tr}\star}^{-1} = \begin{pmatrix} \mathbf{I}_{\mathrm{f}} & -\mathbb{S}_{\mathrm{f,it}}^{-1}\mathbf{B}_{\mathrm{fw,tr}}^{(1)} \\ \mathbf{O} & \mathbf{I}_{\mathrm{w}} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}^{-1} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{B}}_{\mathrm{w}\star}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\mathrm{f}} & \mathbf{O} \\ -\mathbf{B}_{\mathrm{wf,tr}}^{(1)}\mathbb{S}_{\mathrm{f,it}}^{-1} & \mathbf{I}_{\mathrm{w}} \end{pmatrix}, \qquad (9.172)$$

where $\boldsymbol{\mathcal{B}}_{\mathrm{w}\star}$ is a preconditioner for

$$\mathbf{B}_{\mathrm{w}}^{(2)} = \mathbf{B}_{\mathrm{w}}^{(1)} - \mathbf{B}_{\mathrm{wf,tr}}^{(1)}\mathbf{S}_{\mathrm{f}}^{-1}\mathbf{B}_{\mathrm{fw,tr}}^{(1)}\,.$$

General ways of calculation of $\mathbf{B}_{\mathrm{w}}^{(2)}$ are not cheap. Thus, the use of it for $\boldsymbol{\mathcal{B}}_{\mathrm{w}\star}$ is possible only after designing sufficiently cheap procedure for calculating $\mathbf{B}_{\mathrm{w}}^{(2)}$.

4) The aspect ratios of edges of subdomains $\delta_{\mathbf{k}}$ as well as the aspect ratios of the coefficients in front of the fourth-order mixed finite-differences in (9.162), can be very small. For this reason, we obtain the preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}$ for $\mathbf{S}_{\mathrm{f,tr}}$ by zeroing only all off-diagonal blocks coupling the d. o. f. for adjacent faces in $\mathbf{S}_{\mathrm{f,tr}}$. The system with such a preconditioner as system matrix, though more complex than one obtained after decoupling all faces, can be solved by a direct method by means of $\mathcal{O}(N^2 \log N)$ arithmetic operations. Indeed, as follows from iii), the preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}$ is, up to the perturbation, nothing than the *block diagonal matrix with* $3[N^2 - 2N(l_0 - 1) + (l_0 - 1)^2]$ *independent tridiagonal blocks* of the size $l_0 - 1$. Numerical experiment in the range of $p \leq 500$ showed that the relative condition number $\mathrm{cond}\,[\boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}^{-1}\mathbf{S}_{\mathrm{f,tr}}]$ is sufficiently good and, namely, is poly-logarithmic, *i.e.,* $\mathcal{O}((\log N)^{\varkappa})$ with some fixed $\varkappa$ that is not too large. Indeed, for the pointed out range of p, Figure 9.1 allows us to accept some $\varkappa \leq 4$. This also approves the choice

$$\mathbb{S}_{\mathrm{f,it}} = \boldsymbol{\mathcal{I}}\,[\mathbf{S}_{\mathrm{f,tr}}, \boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}, \nu]\,,$$

as preconditioner for (9.170) and (9.172), based on inexact iterative solver for the systems with the matrix $\mathbf{S}_{\mathrm{f,tr}}$, and, besides, allows us to replace $\boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}$ by $\mathbb{S}_{\mathrm{f,it}}$ in the expressions above.

In $DD_{\,\mathrm{loc}}$ algorithm, multiplications by $\mathbf{S}_{\,\mathrm{tr}}$ can be circumvented by a series of operations in the nodal basis corresponding to matrix-vector multiplications by $\mathbf{S} = \mathbf{B}_b - \mathbf{B}_{\mathrm{b,i}}\mathbf{B}_{\mathrm{i}}^{-1}\mathbf{B}_{\mathrm{i,b}}$ and FDFT. In the same way, multiplications by $\mathbf{S}_{\,\mathrm{f,tr}}$ can be performed.

9.3.3.3 *A Version of the $DD_{\rm loc}$ Algorithm*

$DD_{\rm loc}$ algorithms, in which Schur complements are not calculated explicitly, have a similar computational complexity. Which of the two versions is advantageous may be decided on the basis of numerical experimentation. If the Schur complement $\mathbf{S}_{\rm b}$ is not calculated explicitly, then the number of multiplications by $\mathbf{B}_{\rm i}^{-1}$ increases, but in a fixed number of times. We note that if the lower index "tr" is absent in the notations of matrices below, then the initial basis, which is "nodal" for the finite-difference interpretation of the $DD_{\rm loc}$ preconditioner, is assumed. One possible $DD_{\rm loc}$ preconditioner $\boldsymbol{\mathcal{B}}$ of this type for the matrices $\mathbf{A}_e, \boldsymbol{\Lambda}$ and $\mathbf{B}$ is defined by the factorization

$$\boldsymbol{\mathcal{B}}^{-1} = \begin{pmatrix} \mathbf{I}_{\rm i} & -\mathbf{B}_{\rm i}^{-1}\mathbf{B}_{\rm ib} \\ \mathbf{O} & \mathbf{I}_{\rm b} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{\rm i}^{-1} & \mathbf{O} \\ \mathbf{O} & \boldsymbol{\mathcal{S}}_{\rm b}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\rm i} & \mathbf{O} \\ -\mathbf{B}_{\rm bi}\mathbf{B}_{\rm i}^{-1} & \mathbf{I}_{\rm b} \end{pmatrix},$$

where

$$\boldsymbol{\mathcal{S}}_{\rm b}^{-1} = \begin{pmatrix} \mathbf{I}_{\rm f} & -\mathbb{S}_{\rm f}^{-1}\mathbf{S}_{\rm fw} \\ \mathbf{O} & \mathbf{I}_{\rm w} \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathcal{S}}_{\rm f}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbb{S}_{\rm w}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{\rm f} & \mathbf{O} \\ -\mathbf{S}_{\rm w,f}\mathbb{S}_{\rm f}^{-1} & \mathbf{I}_{\rm w} \end{pmatrix},$$

$\mathbf{S}_{\rm f} = \mathbf{B}_{\rm f} - \mathbf{B}_{\rm fi}\mathbf{B}_{\rm i}^{-1}\mathbf{B}_{\rm if}$, $\mathbf{S}_{\rm w} = \mathbf{B}_{\rm w} - \mathbf{B}_{\rm wi}\mathbf{B}_{\rm i}^{-1}\mathbf{B}_{\rm iw}$, $\boldsymbol{\mathcal{S}}_{\rm f}^{-1} = \boldsymbol{\mathcal{F}}_{\rm f}^{-1}\mathbf{S}_{\rm f,it}^{-1}\boldsymbol{\mathcal{F}}_{\rm f}$, $\mathbb{S}_{\rm f}^{-1} = \boldsymbol{\mathcal{F}}_{\rm f}^{-1}\mathbb{S}_{\rm f,it}^{-1}\boldsymbol{\mathcal{F}}_{\rm f}$, $\mathbf{S}_{\rm fw} = \mathbf{B}_{\rm fw} - \mathbf{B}_{\rm fi}\mathbf{B}_{\rm i}^{-1}\mathbf{B}_{\rm iw}$ and $\mathbb{S}_{\rm w} = \mathbf{S}_{\rm w} - \mathbf{S}_{\rm wf}\mathbf{S}_{\rm f}^{-1}\mathbf{S}_{\rm fw}$. Here $\boldsymbol{\mathcal{F}}_{\rm f}$ is the matrix of the 2d DFFT to the subdomainwise discrete-harmonic trigonometric basis on the faces of the decomposition, whereas

$$\mathbb{S}_{\rm f,it} = \boldsymbol{\mathcal{I}}\left[\mathbf{S}_{\rm f,tr}, \boldsymbol{\mathcal{S}}_{\rm f,tr}, k_1\right] \quad \text{and} \quad \mathbf{S}_{\rm f,it} = \boldsymbol{\mathcal{I}}\left[\mathbf{S}_{\rm f,tr}, \boldsymbol{\mathcal{S}}_{\rm f,tr}, k_2\right].$$

The matrices $\mathbf{S}_{\rm f}, \mathbf{S}_{\rm w}, \mathbf{S}_{\rm fw}, \mathbf{S}_{\rm wf}, \mathbb{S}_{\rm w}, \boldsymbol{\mathcal{S}}_{\rm f}^{-1}$ and $\mathbb{S}_{\rm f}^{-1}$ are not calculated explicitly and they are not stored in the memory. For instance, the product $\mathbf{S}_{\rm fw}\mathbf{v}_{\rm w}$ is calculated as $\mathbf{S}_{\rm fw}\mathbf{v}_{\rm w} = \mathbf{B}_{\rm fw}\mathbf{v}_{\rm w} - \mathbf{B}_{\rm fi}(\mathbf{B}_{\rm i}^{-1}(\mathbf{B}_{\rm iw}\mathbf{v}_{\rm w}))$. The matrix $\mathbb{S}_{\rm w}$ can be replaced by the preconditioner $\boldsymbol{\mathcal{S}}_{\rm w} = \mathbf{S}_{\rm w} - \mathbf{S}_{\rm wf}\mathbf{S}_{\rm f,it}^{-1}\mathbf{S}_{\rm fw}$. If the condition number of $\boldsymbol{\mathcal{S}}_{\rm f}^{-1}\boldsymbol{\mathcal{S}}_{\rm f}$ is almost optimal, then the described version of $\mathrm{DD}_{\rm loc}$ algorithm is fast provided that the computations are arranged appropriately.

Further improvements of efficiency can be achieved by additional shifts from the substructuring algorithm to the DD algorithm. For instance, in the expressions for $\mathbb{S}_{\rm f,it}$ and $\mathbf{S}_{\rm f,it}$, the Schur complement $\mathbf{S}_{\rm f,tr}$ can be replaced by some preconditioner-multiplicator $\boldsymbol{\mathcal{S}}_{\rm f,it}$, which is close to $\mathbf{S}_{\rm f,it}$ in the spectrum and cheaper for matrix-vector multiplications. Simultaneously, the interface Schur complement preconditioner $\boldsymbol{\mathcal{S}}_{\rm b}$ can be subjected to the same type adjustments. It can be also implicitly defined by the inexact solver as $\mathbf{S}_{\rm b,it} = \boldsymbol{\mathcal{I}}_{\circ}\left[\boldsymbol{\mathcal{S}}_{\rm b}^{(1)}, \boldsymbol{\mathcal{S}}_{\rm b}^{(2)}\right]$ with $\boldsymbol{\mathcal{S}}_{\rm b}^{(2)}$ being obtained by simplification of $\boldsymbol{\mathcal{S}}_{\rm b}$ given above. The preconditioner-multiplicator $\boldsymbol{\mathcal{S}}_{\rm b}^{(1)}$ can be constructed with the use of the techniques, which are similar to those for deriving $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$

in Subsubsection 8.1.2.3 and developed in Chapter 6 for deriving the more general preconditioner-multiplicator $\boldsymbol{\mathcal{S}}_1$. However, the preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{b}}^{(1)}$ for 3d is considerably more complex than its counterpart $\boldsymbol{\mathcal{S}}_{\mathbb{III},1}$ for 2d.

9.3.3.4 *Numerical Experiment*

Positive features of the DD_{loc} algorithm are efficient and simple preconditioner-solvers for two basic DD components: the internal subproblems on the subdomains of the decomposition and the internal subproblems on the faces of the subdomains of the decomposition. These preconditioner-solvers deal with the major bulk of unknowns, whereas the number of the remaining unknowns does not exceed $3N(\log N)^2$. The definition of the latter preconditioner differs from the common way of designing face preconditioners for requlars discretizations of regular elliptic problems, *e.g.,* in the wire basket algorithms discussed in Subsection 9.1.4. In order to take into account properties of highly orthotropic discrete problems with the dominating fourth-order mixed finite-differences more accurately, we retain coupling of the opposite faces of subdomains. This increases the computational complexity of the solvers for the Schur complement preconditioner only by a constant, but prevents some losses in the relative condition number. Numerical experiments approve effectiveness of the preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{f}}$ for $\mathbf{S}_{\mathrm{f}}$ (and equivalently of the preconditioner $\boldsymbol{\mathcal{S}}_{\mathrm{f,tr}}$ for $\mathbf{S}_{\mathrm{f,tr}}$) and the way of their use in inexact solvers. The construction of $\boldsymbol{\mathcal{S}}_{\mathrm{f}}$ implies that $\lambda_{\max}[\boldsymbol{\mathcal{S}}_{\mathrm{f}}^{-1}\mathbf{S}_{\mathrm{f}}] \le 6$. In order to estimate $\lambda_{\min}[\boldsymbol{\mathcal{S}}_{\mathrm{f}}^{-1}\mathbf{S}_{\mathrm{f}}]$, we note that the matrices $\boldsymbol{\mathcal{S}}_{\mathrm{f}}$ and $\mathbf{S}_{\mathrm{f}}$ are obtained by assembling the matrices $\boldsymbol{\mathcal{S}}_{\mathrm{f},\mathbb{k}}$ and $\mathbf{S}_{\mathrm{f},\mathbb{k}}$, respectively, for the subdomains $\delta_{\mathbb{k}}$ of the decomposition. Hence, $\lambda_{\min}[\boldsymbol{\mathcal{S}}_{\mathrm{f}}^{-1}\mathbf{S}_{\mathrm{f}}] \ge \min_{\mathbb{k}} \lambda_{\min}[\boldsymbol{\mathcal{S}}_{\mathrm{f},\mathbb{k}}^{-1}\mathbf{S}_{\mathrm{f},\mathbb{k}}]$, where $\lambda_{\min}[\boldsymbol{\mathcal{S}}_{\mathrm{f},\mathbb{k}}^{-1}\mathbf{S}_{\mathrm{f},\mathbb{k}}]$ denotes the minimal eigenvalue of the nonzero spectrum of the eigenvalue problem $\mathbf{S}_{\mathrm{f},\mathbb{k}}\mathbf{x} = \lambda \boldsymbol{\mathcal{S}}_{\mathrm{f},\mathbb{k}}\mathbf{x}$.

The dependence of $1/\min_{\mathbb{k}} \lambda_{\min}[\boldsymbol{\mathcal{S}}_{\mathrm{f},\mathbb{k}}^{-1}\mathbf{S}_{\mathrm{f},\mathbb{k}}]$ on N, that was numerically computed in [Anufriev and Korneev (2004)], is shown in the Figure 9.1 by circles. The solid line is the graph of $1 + 0.8(\ln N)^{1.75}$.

9.3.4 *Fast Domain Decomposition Solver for Spectral Elements*

FD/FE preconditioners for the spectral and hierarchical reference p-elements look significantly different, especially in the 3d case. While the former are FE/FD approximations to the Laplace operator on nonuniform

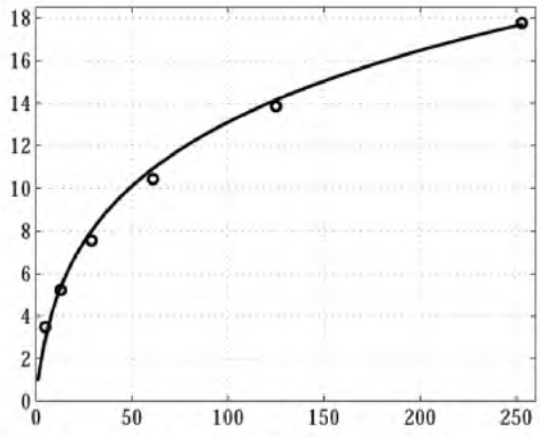

Fig. 9.1 Dependence of $1/\lambda_{\min}\,[\boldsymbol{\mathcal{S}}_{\mathrm{f}}^{-1}\mathbf{S}_{\mathrm{f}}]$ on N.

spectral meshes, the latter are interpreted as FD/FE approximations of the partial differential equation with mixed fourth-order derivatives for the main terms and anysotropic deteriorating coefficients on uniform meshes. In the contrast to these distinctions, the $DD_{\,\mathrm{loc}}$ algorithms, based on the same approach to the arrangement of basic components, may be applied in both cases. For this reason, in this subsection, we practically remind only the decomposition mesh. After its definition, the construction of the $DD_{\,\mathrm{loc}}$ preconditioner-solver could be repeated step by step, if not to look for deeper particularities. This subsection also shows that irregular boundary value problems different from the types considered earlier can efficiently be solved by the domain decomposition method.

9.3.4.1 *Coarse Decomposition Mesh and FD Preconditioner on a Piecewise Uniform Mesh*

For convenience, as in 2d, we assume $p = 2N$. The generalization of the *spectral mesh*, termed also the fine spectral mesh, the *pseudo-spectral mesh*, the *coarse decomposition mesh* and the *shifted piecewise uniform fine mesh* from 2d to 3d is straightforward. We base our considerations on the preconditioner $\boldsymbol{\mathcal{A}}_I := \boldsymbol{\mathcal{A}}_{I,\mathrm{p/s}}$, denoted for brevity $\boldsymbol{\mathcal{A}}_I$, which is nothing but the FE matrix defined by means of the pseudo-spectral mesh in Subsection 9.2.3. However, similar $DD_{\,\mathrm{loc}}$ solvers can be constructed on the basis

of the finite-difference/finite element preconditioners introduced there for GLL and GLC spectral elements. As usual, for designing fast solvers-preconditioners, first of all we modify the source FE preconditioner with the aim of adaptation to the particular type $DD_{\,\mathrm{loc}}$ preconditioner-solver.

Algorithms for finding coordinates $\zeta_0 = -1,\ \zeta_k,\ k = 0, 1, \ldots, 2\ell_0$, $\zeta_{2\ell_0} = 1$ and mesh steps $\overline{h}_k = \zeta_k - \zeta_{k-1}$ of the 1d coarse decomposition mesh are the same as in Subsection 8.3.1. Here we assume that the mesh steps $\overline{h}_k$ satisfy the conditions (8.126), p. 327, see additionally Lemma 8.6 and Corollary 8.6. The nodes of 3d coarse rectangular mesh are the tensor products of the coordinates of this 1d coarse mesh.

The cells of the decomposition mesh are the parallelepipeds

$$\delta_{\Bbbk} = \{\boldsymbol{\xi} : \zeta_{k_s-1} < \xi_s < \zeta_{k_s},\ \ s = 1, 2, 3\}\,, \quad 1 \le k_s \le 2\ell_0\,,$$

the faces of which $F_{\Bbbk,s,t}$ lie in the planes $\xi_s \equiv \zeta_{k_s+t-1}$, $t = 0, 1$, where $\Bbbk = (k_1, k_2, k_3)$. Similarly to the 2d case, the piecewise uniform *shifted fine mesh* $\xi_s = \vartheta_i,\ i = 0, 1, \ldots, p,$ is introduced in such a way that

– it covers the coarse mesh, *i.e.,* $\vartheta_{\widetilde{i}(k)} = \eta_{\widetilde{i}(k)} = \zeta_k$ for $k = 0, 1, \ldots, 2\ell_0$, and

– on each nest $\delta_{\Bbbk}$ of the coarse mesh, it is uniform with the steps

$$h_{k_s} = \overline{h}_{k_s}/n_{k_s}, \quad n_k = \widetilde{i}(k) - \widetilde{i}(k-1)\,.$$

It is clear that the numbers of the mesh planes crossing each nest $\overline{\delta}_{\Bbbk}$ are the same for the fine mesh and the shifted fine mesh. For the steps of the piecewise uniform shifted fine mesh, we also use the notation $h^{(i)} := \vartheta_i - \vartheta_{i-1}$.

Let $\mathcal{H}_D(\tau_0)$ be the space of functions which are continuous on τ_0 and trilinear on each cell of the fine shifted mesh, whereas $\mathring{\mathcal{H}}_D(\tau_0)$ denotes the subspace of FE functions vanishing on $\partial\tau_0$. The modified preconditioner $\overline{\mathbf{B}}$ for $\mathbf{A}$ is defined as the stiffness matrix generated by the Dirichlet integral on the space $\mathcal{H}_D(\tau_0)$ with the nodal basis. This preconditioner can be written in a form that is similar to (9.141). Let $\mathcal{H}_D(-1, 1)$ be the space of functions which are continuous on [-1,1] and linear on each interval of the fine shifted mesh. Let us also introduce the matrices $\mathbb{B}_1$ and $\mathbb{B}_0$ generated by the bilinear forms $(v', w')_{(-1,1)}$ and $(v, w)_{(-1,1)}$ on the space $\mathcal{H}_D(-1, 1)$, respectively. Then we obviously have the representation

$$\overline{\mathbf{B}} = \mathbb{B}_0 \otimes \mathbb{B}_0 \otimes \mathbb{B}_1 + \mathbb{B}_0 \otimes \mathbb{B}_1 \otimes \mathbb{B}_0 + \mathbb{B}_1 \otimes \mathbb{B}_0 \otimes \mathbb{B}_0\,.$$

The $DD_{\,\mathrm{loc}}$ algorithm is also applicable to the simpler, 7-point stencil preconditioner

$$\overline{\mathbf{B}}_+ = \mathbf{D}_0 \otimes \mathbf{D}_0 \otimes \mathbb{B}_1 + \mathbf{D}_0 \otimes \mathbb{B}_1 \otimes \mathbf{D}_0 + \mathbb{B}_1 \otimes \mathbf{D}_0 \otimes \mathbf{D}_0$$

without lossing the efficiency, where $\mathbf{D}_0$ is the diagonal matrix with the main diagonal from $\mathbb{B}_0$. We can also take $\mathbf{D}_0 = \operatorname{diag}[\widetilde{h}^{(i)}]_{i=1}^{2\ell_0}$ with $\widetilde{h}^{(i)} = 0.5(h^{(i-1)} + h^{(i)})$ and $h^{(-1)}, h^{(2\ell_0+1)} = 0$.

Lemma 9.11. *Let* $\mathbf{A}$ *be the GLL spectral reference element stiffness matrix and* $\overline{\mathbf{B}}$ *be the FE matrix generated on the fine shifted mesh corresponding to the spectral or pseudo spectral mesh. Then the inequalities*

$$\underline{\gamma}_B \overline{\mathbf{B}} \le \mathbf{A} \le \overline{\gamma}_B \overline{\mathbf{B}} \quad \textit{and} \quad \underline{\gamma}_B \overline{\mathbf{B}}_+ \le \mathbf{A} \le \overline{\gamma}_B \overline{\mathbf{B}}_+ \,,$$

hold with positive constants independent of p.

Proof. The proof is similar to the proof for 2d, see Lemma 8.7 and Corollary 8.7. The constants depend only on c_0 in (8.126). □

Therefore, matrix $\mathbf{B}$, which designates the block of $\overline{\mathbf{B}}$ related to the internal degrees of freedom, is a good preconditioner for the internal Dirichlet problems on spectral reference elements. Moreover, the *DD* $_{\rm loc}$ type algorithm can be implemented for solving systems with this matrix as the system matrix. This algorithm is analogous to that one considered in Subsection 9.3.3 for the hierarchical finite elements in respects of handling Dirichlet problems on subdomains of decomposition and the inter-subdomain problem with the use of the subdomainwise trigonometric basis. At the same time the involved local matrices for subdomains and faces are significantly different first of all due to the significant difference of the energy integrals defining the stiffness matrix FE/FD preconditioners. Besides, stiffness matrices $\mathbf{A}$ of hierarchical reference elements are 8×8 block diagonal, whereas spectral elements does not possess such a structure.

9.3.4.2 *On Components of* $DD_{\rm loc}$ *Algorithms*

We consider the system: given $\mathbf{d} \in U_I$, find $\mathbf{v} \in U_I$ such that

$$\mathbf{B}\mathbf{v} = \mathbf{d} \,, \tag{9.173}$$

where $U_I = R^{(p-1)^3}$ is the subspace of vectors corresponding to the internal degrees of freedom of the reference element. It can be decomposed into the direct sum $U_I = U_{\rm i} \oplus U_{\rm f} \oplus U_{\rm w}$ of three subspaces, where

- $U_{\rm i} = \bigcup_{s,k_s=1}^{3,\,2\ell_0} U_{{\rm i},\Bbbk}$ is the union of subsubspaces $U_{{\rm i},\Bbbk}$ of internal d. o. f. for each subdomain $\delta_\Bbbk$ of the decomposition,
- $U_{\rm f}$ is the subspace of the internal unknowns for the faces of the decomposition,
- $U_{\rm w}$ is the subspace of the wire-basket unknowns.

With respect to this decomposition, the matrix and the solution and right-hand side vectors of the system (9.164) can be written in the block form

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{\rm i} & \mathbf{B}_{\rm i,b} \\ \mathbf{B}_{\rm b,i} & \mathbf{B}_{\rm b} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{\rm i} & \mathbf{B}_{\rm if} & \mathbf{B}_{\rm iw} \\ \mathbf{B}_{\rm fi} & \mathbf{B}_{\rm f} & \mathbf{B}_{\rm fw} \\ \mathbf{B}_{\rm wi} & \mathbf{B}_{\rm wf} & \mathbf{B}_{\rm w} \end{pmatrix}, \ \mathbf{v} = \begin{pmatrix} \mathbf{v}_{\rm i} \\ \mathbf{v}_{\rm f} \\ \mathbf{v}_{\rm w} \end{pmatrix}, \ \mathbf{d} = \begin{pmatrix} \mathbf{d}_{\rm i} \\ \mathbf{d}_{\rm f} \\ \mathbf{d}_{\rm w} \end{pmatrix}, \tag{9.174}$$

where $\mathbf{B}_{\rm i} = \mathrm{diag}\,[\mathbf{B}_{\rm i,\mathbb{k}}]_{k_1,k_2,k_3=1}^{2\ell_0}$ is the block-diagonal matrix with the diagonal blocks $\mathbf{B}_{\rm i,\mathbb{k}}$ corresponding to the subdomains $\delta_{\mathbb{k}}$. Below we briefly discuss a few facts to be taken into consideration at the implementation of the *DD* loc solver for the system (9.173), (9.174).

The matrices of the local discrete Dirichelet problems have the form

$$\begin{aligned} \mathbf{B}_{\rm i,\mathbb{k}} = {} & \tfrac{h_{k_2}h_{k_3}}{h_{k_1}} \mathbf{\Delta}_{n_{k_1}} \otimes \mathbf{M}_{n_{k_2}} \otimes \mathbf{M}_{n_{k_3}} + \tfrac{h_{k_1}h_{k_3}}{h_{k_2}} \mathbf{M}_{n_{k_1}} \otimes \mathbf{\Delta}_{n_{k_2}} \otimes \mathbf{M}_{n_{k_3}} \\ & + \tfrac{h_{k_1}h_{k_2}}{h_{k_3}} \mathbf{M}_{n_{k_1}} \otimes \mathbf{M}_{n_{k_2}} \otimes \mathbf{\Delta}_{n_{k_3}} \ , \end{aligned} \tag{9.175}$$

where

$$\mathbf{\Delta}_n = \mathrm{tridiag}\,[-1,2,-1] \quad \text{and} \quad \mathbf{M}_n = \frac{1}{6}\,\mathrm{diag}\,[1,4,1]$$

are matrices of the dimension $n \times n$. Systems with these matrices can efficiently be solved, *e.g.,* by means of the FDFT with $\mathcal{O}(N^3 \log N)$ arithmetic operations. The structure of the matrix $\mathbf{B}_{\rm i,\mathbb{k}}$ allows us to use also various optimal multilevel solvers, For instance, solvers presented for 2d problems in Subsubsection 8.1.3.3 are easily expanded to 3d discrete problems governed by $\mathbf{B}_{\rm i,\mathbb{k}}$.

The space U_I can be represented by the sum $U_I = U_{\rm i} \oplus U_{\rm f,tr} \oplus U_{\rm w}$, where the subspace $U_{\rm f,tr}$ is spanned by the basis which may be termed the *piecewise discrete harmonic trigonometric basis.* The properties i)–iv) listed in Subsection 9.3.1 hold for this basis as well. We reformulate only the two most important of them, from which the others follow easily:

i) $U_{\rm i} \perp_{\mathbf{B}} U_{\rm f,tr}$, *i.e.*, the subspaces $U_{\rm i}$ and $U_{\rm f,tr}$ are orthogonal with respect to the scalar product $(\cdot,\cdot)_{\mathbf{B}}$.

ii) The trace of each basis vector on the interface boundary is nonzero only on one face. On each face $F_{\mathbb{k},s,t}$, these traces are vectors, which we denote $\boldsymbol{\mu}^t_{s,\alpha,\beta}$. They have the components

$$\boldsymbol{\mu}^t_{s,\alpha,\beta}\,\big|_{F_{\mathbb{k},s,t}} = \left\{ \sin(\frac{\alpha\pi i}{n_{k_{s+1}}}) \sin(\frac{\beta\pi j}{n_{k_{s+2}}}) \right\}_{i,j=0}^{n_{k_{s+1}},n_{k_{s+2}}} ,$$

where $s+t$ has to be understood modulo 3. Here, n_{k_s} is the number of fine mesh intervals on the edge of $\delta_{\mathbb{k}}$ parallel to the axis ξ_s, *cf.* 9.175).

The numbers α, β and i, j are the local indices of the face basis vectors and the face nodes, respectively, increasing towards the positive directions of the corresponding axes, *i.e.,* $\alpha = 1, \ldots, n_{k_{s+1}} - 1,\ i = 0, 1, \ldots, n_{k_{s+1}}$, and $\beta = 1, \ldots, n_{k_{s+2}} - 1,\ j = 0, 1, \ldots, n_{k_{s+2}}$.

The basis vectors of $U_{\mathrm{f,tr}}$ living on any single face are orthogonal with respect to the scalar product $(\cdot, \cdot)_{\mathbf{B}}$. However, the basis vectors of the adjacent faces are not orthogonal. Apart from that, each face basis vector is coupled with a single basis vector of the opposite face having the same local indices. After the transformation to the piecewise discrete harmonic trigonometric basis, the system (9.174) transforms to the one, which we write as

$$\widetilde{\mathbf{B}}\,\widetilde{\mathbf{v}} = \widetilde{\mathbf{d}}\,. \tag{9.176}$$

Its block form is identical to (9.169) up to the dimensions of the blocks, for which we can use similar notations. The further description of the *DD* $_{\mathrm{loc}}$ preconditioner-solver literally repeats Subsubsection 9.3.3.2.

It is worth noting that the *DD* $_{\mathrm{loc}}$ algorithm for spectral elements is more convenient for analysis than the corresponding algorithm for hierarchical elements. This is for the reason that, in the basis of the former, we have preconditioners which are FD/FE approximations to the Poisson equation.

9.3.5 *Multiresolution Wavelet Solver for Internal Dirichlet Problems for Spectral Elements*

It was established in Subsection 9.2.4 that in order to obtain a fast preconditioner-solver for the internal stiffness matrices $\mathbf{A}_I = \mathbf{A}_{I,\mathrm{sp}}$ or $\mathbf{K}_I^{(r)}$ of spectral elements, it is sufficient to design such a fast solver for the preconditioner $\mathbf{\Lambda}_\diamond$ or $\mathbf{\Lambda}_\dagger$. Construction of the multilevel solver for the matrix $\mathbf{\Lambda}_\diamond$, based on the multiresolution wavelet analysis, closely follows the lines of Subsection 9.3.1, and we intentionally use the same as in this section notations, where it is possible without confusion. However it is necessary to keep in mind that the objects under the same notation are different in these two subsections, *e.g.*, in dimension.

Without loss of generality, it is assumed $p = 2N,\ N = 2^{l_0 - 1}$. For each $l = 1, 2, \ldots, l_0$, one can introduce the uniform mesh $x_i^l,\ i = 0, 1, 2, \ldots, 2N_l,\ N_l = 2^{l-1},\ x_0 = -1,\ x_{2N_l} = 1$ of the size $\hbar_l = 2^{1-l}$ on the interval $(-1, 1)$ and the space $\mathcal{V}_l(-1, 1)$ of the continuous on (-1,1) piecewise linear functions, vanishing at the ends of this interval. The dimension of $\mathcal{V}_l(-1, 1)$ is $\mathcal{N}_l = p_l - 1 = 2^l - 1$ with $p_{\ell_0} = p$. Let $\phi_i^l \in \mathcal{V}_l(-1, 1)$

be the nodal basis function for the node x_i^l, so that $\phi_i^l(x_j^l) = \delta_{i,j}$ and $\mathcal{V}_l(-1,1) = \text{span}\,[\phi_i^l]_{i=1}^{p_l-1}$. This basis induces the Gram matrices

$$\boldsymbol{\Delta}_l = \hbar_l \left\{ \langle (\phi_i^l)', (\phi_j^l)' \rangle_{\chi=1} \right\}_{i,j=1}^{p_l-1}, \quad \boldsymbol{\mathcal{M}}_l = \hbar_l^{-1} \left\{ \langle \phi_i^l, \phi_j^l \rangle_{\chi=\phi} \right\}_{i,j=1}^{p_l-1},$$

where

$$\langle v, u \rangle_\chi := \int_0^1 \chi^2 v\, u\, dx\,,$$

and ϕ is the function introduced in (7.152), see p. 253. These matrices are related to the matrices entering the expression (9.144) for $\boldsymbol{\Lambda}_\diamond$ as follows:

$$\begin{aligned} \boldsymbol{\Delta}_l &= \boldsymbol{\Delta}_{\diamond,l} = \text{diag}\,[-1,2,-1]_1^{p_l-1}\,, \\ \boldsymbol{\Delta}_{l_0} &= \boldsymbol{\Delta}_\diamond\,, \\ \boldsymbol{\mathcal{M}}_l &\asymp \boldsymbol{\mathcal{D}}_{\diamond,l} = \text{diag}\,[1,4,\ldots,(N_l-1),N_l,(N_l-1),\ldots,4,1]\,, \\ \boldsymbol{\mathcal{M}}_{l_0} &\asymp \boldsymbol{\mathcal{D}}_\diamond\,. \end{aligned} \tag{9.177}$$

We now introduce the notations $\mathcal{V} = \mathcal{V}_{l_0}$ and $\mathcal{W}_1 = \mathcal{V}_1$. The space decomposition $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2 \oplus \ldots \oplus \mathcal{W}_{l_0}$ results from the representation of each subspace $\mathcal{V}_k$ by the direct sum $\mathcal{V}_k = \mathcal{V}_{k-1} \oplus \mathcal{W}_k$. Let $\{\psi_i^k\}_{i\in\mathfrak{N}_k}$ be a *single scale "k" basis* in $\mathcal{W}_k$, that is $\mathcal{W}_k = \text{span}\,[\psi_i^k]_{i\in\mathfrak{N}_k}$, where $\mathfrak{N}_k$ is the set of the respective indices i, $1 \le i \le p_l - 1$. Under accepted assumptions

$$\dim\,[\mathfrak{N}_l] = \dim\,[\mathcal{W}_l] = 2^{l-1}\,.$$

By $\Psi_l := \{\psi_i^k\}_{(i,k)\in\omega_l}$ is denoted the *multiscale wavelet basis*, which is composed of the single scale bases for $k = 1,2,\ldots,l$, implying $\omega_l = \{(i,k) : i \in \mathfrak{N}_k,\ k = 1,2,\ldots,l\}$.

The multiscale wavelet basis induces the matrices

$$\begin{aligned} \boldsymbol{\Delta}_{\text{wt}} &= \left\{ \langle (\psi_i^k)', (\psi_j^l)' \rangle_1 \right\}_{(i,k),(j,l)\in\omega_{l_0}}\,, \\ \boldsymbol{\mathcal{M}}_{\text{wt}} &= \left\{ \langle \psi_i^k, \psi_j^l \rangle_x \right) \}_{(i,k),(j,l)\in\omega_{l_0}}\,, \end{aligned} \tag{9.178}$$

and the diagonal matrices

$$\begin{aligned} \mathbb{D}_1 &= \lfloor \boldsymbol{\Delta}_{\text{wt}} \rfloor_{\text{diag}} = \text{diag}\,[\langle (\psi_i^k)', (\psi_i^k)' \rangle_1]_{(i,k)\in\omega_{l_0}}\,, \\ \mathbb{D}_0 &= \lfloor \boldsymbol{\mathcal{M}}_{\text{wt}} \rfloor_{\text{diag}} = \text{diag}\,[\langle \psi_i^k, \psi_i^k \rangle_x]_{(i,k)\in\omega_{l_0}}\,, \end{aligned} \tag{9.179}$$

which have the same look as the matrices (9.148), (9.149), but are different in view of difference of the wavelet spaces.

The matrix $\mathbf{Q}$ denotes the transformation matrix from the multiscale wavelet basis to the basis $\{\phi_i^{l_0}\}_{i=1}^{p-1}$. If $\mathbf{v}$ and $\mathbf{v}_{\text{wavelet}}$ are the vectors of the coefficients of a function from $\mathcal{V}(0,1)$ in the initial single-scale finite element nodal basis and multi-scale wavelet bases, respectively, then $\mathbf{v}_{\text{wavelet}} = \mathbf{Q}\mathbf{v}$.

Theorem 9.9. *There exist multiscale wavelet bases* Ψ_{l_0} *such that*

i) the matrices $\boldsymbol{\Delta}_{\rm wt}$ *and* $\boldsymbol{\mathcal{M}}_{\rm wt}$ *are simultaneously spectrally equivalent to their diagonals* $\mathbb{D}_1$ *and* $\mathbb{D}_0$*, respectively, uniformly in* p*,* i.e.*,*

$$\mathbb{D}_1 \prec \boldsymbol{\Delta}_{\rm wt} \prec \mathbb{D}_1 \quad \textit{and} \quad \mathbb{D}_0 \prec \boldsymbol{\mathcal{M}}_{\rm wt} \prec \mathbb{D}_0\,, \tag{9.180}$$

and

ii) the multiplications $\mathbf{Q}\,\mathbf{v}$ *require not more than* $\mathcal{O}(p)$ *arithmetic operations.*

Proof. The proof requires only minor adjustments of the proof of Theorem 9.6, given in [Beuchler *et al.* (2004)]. One of the main difficulties of this proof is taking into account the deterioration of the weight at $x = 0$. Although it is in part compensated by the Dirichlet boundary condition at that point, it complicates the construction of the wavelets. In the case under consideration, such problem arises at both ends of the interval $(-1, 1)$, but is resolved in the same way as in the cited paper. Practically, even the same wavelets, except for a few new wavelets with the supports in the vicinity of the origin $x = 0$, can be used. Let us introduce two superelements $\mathcal{E}^r, r = 1, 2$, defined on $(-1, 0)$ and $(0, 1)$, respectively, and respective restrictions $\mathcal{V}_l^1(-1, 0)$, $\mathcal{V}_l^2(0, 1)$ of the FE space $\mathcal{V}_l(-1, 1)$. Up to the shifting the interval, the difference between the space $\mathcal{V}_{l+1}^1(-1, 0)$ and the space $\mathcal{V}_l(0, 1)$, used in Subsection 9.3.1, is only in the boundary conditions

$$v\Big|_{x=-1} = 0\,, \quad \forall v \in \mathcal{V}_{l+1}^1(-1, 0)\,, \quad \text{and} \quad v\Big|_{x=0,1} = 0\,, \quad \forall v \in \mathcal{V}_l(0, 1)\,,$$

which are different at the right end. However, for the subspace $\mathcal{V}_{l+1}^1(-1, 0)$ the result similar to Theorem 9.6 remains valid. It approves the existence of the multiscale bases Ψ_{l+1}^1, which in general differ from the bases Ψ^l, see Subsection 9.3.1, only by a few basis wavelet functions with the supports in the vicinity of $x = 0$. Thus, we obtain the matrices $\boldsymbol{\Delta}_{\rm wt}^{(r)}$ and $\boldsymbol{\mathcal{M}}_{\rm wt}^{(r)}$ for $r = 1$. Due to the symmetry of the superelement $\mathcal{E}^1$ to $\mathcal{E}^2$ with respect to $x = 0$, their matrices for differ only by the perturbation. One obtains the matrices for the whole interval by the proper assembling $(-1, 1)$: $\boldsymbol{\Delta}_{\rm wt} = \boldsymbol{\Delta}_{\rm wt}^{(1)} \biguplus \boldsymbol{\Delta}_{\rm wt}^{(2)}$ and $\boldsymbol{\mathcal{M}}_{\rm wt} = \boldsymbol{\mathcal{M}}_{\rm wt}^{(1)} \biguplus \boldsymbol{\mathcal{M}}_{\rm wt}^{(2)}$. For these matrices the properties *i*), *ii*) are fulfilled. □

The theorem below is a consequence of the structure of the preconditioner $\boldsymbol{\Lambda}_\diamond$ and Theorem 9.9.

Theorem 9.10. *Let*

$$\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}}^{-1} = \begin{cases} (\mathbf{Q}^T \otimes \mathbf{Q}^T)[\,\mathbb{D}_0 \otimes \mathbb{D}_1 + \mathbb{D}_1 \otimes \mathbb{D}_0\,]^{-1}(\mathbf{Q} \otimes \mathbf{Q}), & d = 2, \\ (\mathbf{Q}^T \otimes \mathbf{Q}^T \otimes \mathbf{Q}^T)[\mathbb{D}_0 \otimes \mathbb{D}_1 \otimes \mathbb{D}_1 + \mathbb{D}_1 \otimes \mathbb{D}_0 \otimes \mathbb{D}_1 \\ \qquad +\mathbb{D}_1 \otimes \mathbb{D}_1 \otimes \mathbb{D}_0]^{-1}(\mathbf{Q} \otimes \mathbf{Q} \otimes \mathbf{Q}), & d = 3\,, \end{cases} \tag{9.181}$$

then $\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}} \asymp \boldsymbol{\Lambda}_\diamond$ *and therefore*

$$\mathrm{cond}\,[\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}}^{-1}\boldsymbol{\Lambda}_\diamond] \prec 1\,. \tag{9.182}$$

The computational cost of the operation $\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}}^{-1}\mathbf{v}$ *is* $\mathcal{O}(p^d)$ *for any* $\mathbf{v} \in U_I$.

Proof. The proof of the spectral equivalence $\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}} \asymp \boldsymbol{\Lambda}_\diamond$ directly follows from (9.177), Theorem 9.9 and the representations (9.144) and (9.181) of the matrices $\boldsymbol{\Lambda}_\diamond$ and $\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}}$ by sums of the Kronecker products of the matrices induced by 1d problems. The outline of the proof follows the steps of the proof of Theorem 9.7. First we introduce the tentative preconditioner $\widehat{\boldsymbol{\Lambda}}_\diamond \asymp \boldsymbol{\Lambda}_\diamond$ by replacing in the expression (9.144) for $\boldsymbol{\Lambda}_\diamond$ the matrices $\boldsymbol{\mathcal{D}}_\diamond$ by $\boldsymbol{\mathcal{M}}_{l_0}$. Then we represent $\widehat{\boldsymbol{\Lambda}}_\diamond$ in the factorized form $\widehat{\boldsymbol{\Lambda}}_\diamond = \boldsymbol{\mathfrak{Q}}^\top \widehat{\boldsymbol{\Lambda}}_{\diamond,\mathrm{wt}} \boldsymbol{\mathfrak{Q}}$, where in 2d $\boldsymbol{\mathfrak{Q}} = \mathbf{Q} \otimes \mathbf{Q}$ and $\widehat{\boldsymbol{\Lambda}}_{\diamond,\mathrm{wt}} = \boldsymbol{\mathcal{M}}_{\mathrm{wt}} \otimes \boldsymbol{\Delta}_{\mathrm{wt}} + \boldsymbol{\Delta}_{\mathrm{wt}} \otimes \boldsymbol{\mathcal{M}}_{\mathrm{wt}}$. At the last step, we use (9.180) and come to $\boldsymbol{\Lambda}_{\diamond,\mathrm{wt}} \asymp \boldsymbol{\Lambda}_\diamond$. □

Appendix A

Technical Proofs

A.1 Proof of Theorem 8.4

The estimates of the constants $c_\circ$ and c_k in (8.9) and (8.10) are produced in a superelementwise fashion with the dint of which it is reduced to the estimation of some bounds of spectrums for stiffness matrices of superelements or their blocks relative to their preconditioners. At that, square or triangular superelements can be used. In other words, for domains of superelements of level l, one can chose either the square nests of the orthogonal mesh of level $l-1$ or the triangles of the triangulation of level $l-1$. For definiteness, it is assumed that all triangulations of π_1 are obtained by the subdivision of each square nest in two triangles by the diagonal parallel to the straight line joining $(0,0)$ and $(1,1)$. For the estimates of c_1, c_2 we use triangular superelements, which require consideration of matrix eigenvalue problems of smaller dimensions.

Let $\overline{N}_l = N_l + 1$, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, $\gamma_k = 1, 2, \ldots, \overline{N}_{l-1}$, and

$$\delta_{\boldsymbol{\gamma}}^{l-1} = \{x : (\gamma_k - 1)\hbar_{l-1} < x_k < \gamma_k \hbar_{l-1},\ \hbar_{l-1} = 1/\overline{N}_{l-1},\ k = 1, 2\} \quad \text{(A.1)}$$

be the square nests of the $(l-1)$-level orthogonal mesh, see Fig. A.1. They are taken for domains of the square superelements of the level l, for whose stiffness matrices we introduce the notations $\mathbf{B}_{\square}(\boldsymbol{\gamma}) = \hbar^2 \mathbb{A}_{\square}(\boldsymbol{\gamma})$. For the domains of the two upper and lower triangular superelements, defined on $\delta_{\boldsymbol{\gamma}}^{l-1}$, we use notations $\delta_{\triangledown,\boldsymbol{\gamma}}^{l-1}$ and $\delta_{\triangle,\boldsymbol{\gamma}}^{l-1}$, respectively, whereas $\mathbf{B}_{\triangledown}(\boldsymbol{\gamma}) = \hbar^2 \mathbb{A}_{\triangledown}(\boldsymbol{\gamma})$ and $\mathbf{B}_{\triangle}(\boldsymbol{\gamma}) = \hbar^2 \mathbb{A}_{\triangle}(\boldsymbol{\gamma})$ stand for their stiffness matrices. Superelement stiffness matrices can be represented in the block form, *i.e.*,

$$\mathbb{A}(\boldsymbol{\gamma}) = \begin{pmatrix} \mathbb{A}_1(\boldsymbol{\gamma}) & \mathbb{A}_{12}(\boldsymbol{\gamma}) \\ \mathbb{A}_{21}(\boldsymbol{\gamma}) & \mathbb{A}_2(\boldsymbol{\gamma}) \end{pmatrix}, \quad \text{(A.2)}$$

where blocks $\mathbb{A}_1(\boldsymbol{\gamma})$ and $\mathbb{A}_2(\boldsymbol{\gamma})$ correspond to the nodes from the sets $X_{l,w}$ and X_{l-1}, respectively, and the indices characterizing the geometrical type

(square, upper- or lower-triangular) of the superelement are omitted. For the described superelements, we will use the notations $\mathcal{E}_{\square}(\boldsymbol{\gamma}), \mathcal{E}_{\triangledown}(\boldsymbol{\gamma})$ and $\mathcal{E}_{\triangle}(\boldsymbol{\gamma})$. Note that index l, characterizing the level in the notations of superelements, their stiffness matrices etc., is as a rule omitted. One reason for doing this in the case of stiffness matrices is the following: indeed, these matrices do not depend on the level and mesh size $\hbar_l$, but only on $\boldsymbol{\gamma}$, see explicit expressions below.

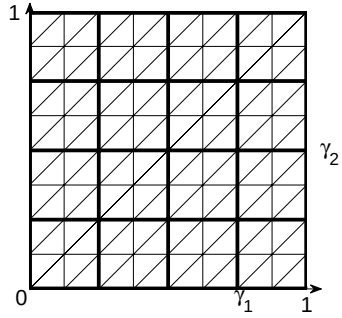

Fig. A.1 Triangulation and domains $\delta_{\boldsymbol{\gamma}}^{l-1}$ of superelements.

If to adopt that $x = \hbar_l(i,j)$, with $i = 2\gamma_1 - 1$, $j = 2\gamma_2 - 1$, is the central l-th level node of the square superelement $\mathcal{E}(\boldsymbol{\gamma})$ on $\overline{\delta}_{\boldsymbol{\gamma}}^{l-1}$, then all l-th level nodes on $\overline{\delta}_{\boldsymbol{\gamma}}^{l-1}$ correspond to the pairs of numbers (i',j') with $i' = i-1, i, i+1$, $j' = j-1, j, j+1$. For writing down entries of matrix $\mathbb{A}(\boldsymbol{\gamma})$, we number these nodes according to the ordering

$$\overset{1}{(i-1,j)}\ \overset{2}{(i,j+1)}\ \overset{3}{(i+1,j)}\ \overset{4}{(i,j-1)}\ \overset{5}{(i,j)}\ \overset{6}{(i-1,j+1)}\ \overset{7}{(i+1,j+1)}\ \overset{8}{(i+1,j-1)}\ \overset{9}{(i-1,j-1)}\,,$$

shown in Fig. A.2. In triangular superelements shown in Fig. A.3 and A.4, the order of the nodes corresponds to their sequence in the above listing, *i.e.*, for $\mathcal{E}_{\triangledown}(\boldsymbol{\gamma})$, we have

$$\begin{array}{cccccc} 1 \Leftarrow 1 & 2 \Leftarrow 2 & 3 \Leftarrow 5 & 4 \Leftarrow 6 & 5 \Leftarrow 7 & 6 \Leftarrow 9 \\ (i-1,j) & (i,j+1) & (i,j) & (i-1,j+1) & (i+1,j+1) & (i-1,j-1)\,. \end{array}$$

We also introduce the numbers

$$\begin{aligned} \alpha(k) &= \frac{1}{2}k^2 + \frac{2}{3}k + \frac{1}{4}, \\ \beta(k) &= \frac{1}{2}k^2 + \frac{1}{3}k + \frac{1}{12}, \\ \upsilon(k) &= \alpha(k-1) + \beta(k) = k^2 + \frac{1}{6}\,. \end{aligned} \tag{A.3}$$

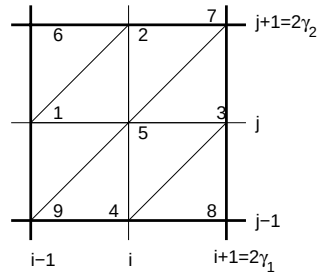

Fig. A.2 Triangulation and domains of square superelements.

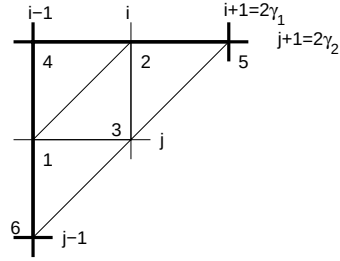

Fig. A.3 Upper triangular superelement.

The entries of the stiffness matrices of superelements are expressed via simple integrals over domains of finite elements of level l. At that, for the blocks of $\mathbb{A}_{\triangledown}(\boldsymbol{\gamma})$, we get

$$\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma}) = \begin{pmatrix} 2\beta(i-1)+\upsilon(j) & 0 & -\upsilon(j) \\ & \upsilon(i)+2\alpha(j) & -\upsilon(i) \\ \text{SYM} & & \upsilon(j)+\upsilon(i) \end{pmatrix},$$

$$\mathbb{A}_{2,\triangledown}(\boldsymbol{\gamma}) = \operatorname{diag}\left[\beta(i-1)+\alpha(j),\, \alpha(j),\, \beta(i-1)\right], \tag{A.4}$$

$$\mathbb{A}_{12,\triangledown}(\boldsymbol{\gamma}) = \begin{pmatrix} -\beta(i-1) & 0 & -\beta(i-1) \\ -\alpha(j) & -\alpha(j) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We also introduce the superelement stiffness matrix $\hat{\mathbb{A}}_{\triangledown}(\boldsymbol{\gamma}) = \mathbf{C}_{\triangledown}^{\top}\mathbb{A}_{\triangledown}(\boldsymbol{\gamma})\,\mathbf{C}_{\triangledown}$, which is obtained by means of the transformation matrix

$$\mathbf{C}_{\triangledown} = \begin{pmatrix} 1 & & \frac{1}{2} & \frac{1}{2} \\ & 1 & \frac{1}{2} & \frac{1}{2} \\ & & 1 & \frac{1}{2} \quad \frac{1}{2} \\ & & & 1 \\ & & & & 1 \\ & & & & & 1 \end{pmatrix},$$

with empty entries implying zeroes. The matrix $\mathbf{C}_{\triangledown}$ transforms the stiffness matrix of the superelement to the basis in which three basis functions, related to the vertices of the superelement, become basis functions of level $l-1$. The upper block on the diagonal of the stiffness matrix is not changed by the transformation, *i.e.*,

$$\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma}) = \hat{\mathbb{A}}_{1,\triangledown}(\boldsymbol{\gamma})\,.$$

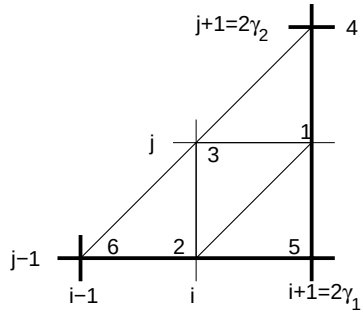

Fig. A.4 Lower triangular superelement.

The preconditioner $\boldsymbol{\mathcal{B}}_{l,\mathrm{w}}$ can be also defined by assembling the preconditioners

$$\boldsymbol{\mathcal{B}}_{1,\triangledown}(\boldsymbol{\gamma}) = \hbar^2\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma}) \quad \text{and} \quad \boldsymbol{\mathcal{B}}_{1,\triangle}(\boldsymbol{\gamma}) = \hbar^2\boldsymbol{\mathcal{A}}_{1,\triangle}(\boldsymbol{\gamma})\,,$$

which result from the matrices $\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})$ and $\mathbb{A}_{1,\triangle}(\boldsymbol{\gamma})$, respectively, by decoupling the d. o. f. of the line $\Im_l^i$, $i = 2\max(\gamma_1, \gamma_2)$, with the d. o. f. of the lines $\Im_l^{i-1}$ and $\Im_l^{i+1}$. If we denote the entries of $\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})$ by $a_{k,l}$, then

$$\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma}) = \begin{pmatrix} a_{1,1} & 0 & a_{1,3} \\ 0 & a_{2,2} & 0 \\ a_{3,1} & 0 & a_{3,3} \end{pmatrix}, \quad \text{for} \quad 1 \le \gamma_1 \le \gamma_2 \le \overline{N}_{l-1}\,, \tag{A.5}$$

and

$$\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma}) = \begin{pmatrix} a_{1,1} & 0 & 0 \\ 0 & a_{2,2} & a_{2,3} \\ 0 & a_{3,2} & a_{3,3} \end{pmatrix}, \quad \text{for} \quad 1 \le \gamma_2 < \gamma_1 \le \overline{N}_{l-1}\,. \tag{A.6}$$

Let the spaces $W_{\triangledown,\gamma}$ and $W_{\triangle,\gamma}$ be the restrictions of the space W_l to the superelements $\mathcal{E}_{\triangledown}(\boldsymbol{\gamma})$ and $\mathcal{E}_{\triangle}(\boldsymbol{\gamma})$, respectively. For positive constants $c_{k,\triangledown}$, $c_{k,\triangle}$, $k = 1, 2$, there hold the inequalities

$$\begin{aligned} c_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{v}^\top \boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{v} \le \mathbf{v}^\top \mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{v} \le c_{2,\triangledown}(\boldsymbol{\gamma})\mathbf{v}^\top \boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{v}\,, \\ c_{1,\triangle}(\boldsymbol{\gamma})\mathbf{v}^\top \boldsymbol{\mathcal{A}}_{1,\triangle}(\boldsymbol{\gamma})\mathbf{v} \le \mathbf{v}^\top \mathbb{A}_{1,\triangle}(\boldsymbol{\gamma})\mathbf{v} \le c_{2,\triangle}(\boldsymbol{\gamma})\mathbf{v}^\top \boldsymbol{\mathcal{A}}_{1,\triangle}(\boldsymbol{\gamma})\mathbf{v}\,, \end{aligned} \tag{A.7}$$

for $\forall\, \mathbf{v} \in W_{\triangledown,\gamma}$ and $\forall\, \mathbf{v} \in W_{\triangle,\gamma}$, respectively, in view that all involved 3×3 matrices are s.p.d.

The matrix $\mathbf{B}_{l,\mathrm{w}}$ and its preconditioner $\boldsymbol{\mathcal{B}}_{l,\mathrm{w}}$ can be defined by identical assembling procedures from the matrices $\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})$, $\mathbb{A}_{1,\triangle}(\boldsymbol{\gamma})$ and $\boldsymbol{\mathcal{B}}_{1,\triangledown}(\boldsymbol{\gamma})$, $\boldsymbol{\mathcal{B}}_{1,\triangle}(\boldsymbol{\gamma})$, respectively. Taking additionally into account the symmetry conditions, one comes to the conclusion that the constants in (8.10) satisfy

$$\begin{aligned} c_1 &\ge \min\nolimits_{1\le\gamma_k\le\overline{N}_l} \min(c_{1,\triangle}(\boldsymbol{\gamma}), c_{1,\triangledown}(\boldsymbol{\gamma})) = \min\nolimits_{1\le\gamma_k\le\overline{N}_l} c_{1,\triangledown}(\boldsymbol{\gamma})\,, \\ c_2 &\le \max\nolimits_{1\le\gamma_k\le\overline{N}_l} \max(c_{2,\triangle}(\boldsymbol{\gamma}), c_{2,\triangledown}(\boldsymbol{\gamma})) = \max\nolimits_{1\le\gamma_k\le\overline{N}_l} c_{2,\triangledown}(\boldsymbol{\gamma})\,. \end{aligned} \tag{A.8}$$

The evaluation of the parameters μ_l and ν of the multigrid iterations assumes sufficiently good estimates for these constants. In order to get them, we consider the eigenvalue problems

$$\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{x} = \lambda \boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{x}\,, \quad \text{for} \quad 1 \le \gamma_2 < \gamma_1 \le \overline{N}_{l-1}\,.$$

With the notations

$$\triangle\,(i,j) := a_{2,2}a_{3,3} - a_{2,1}^2 = \upsilon(j)[\upsilon(i) + 2\alpha(j)] + 2\upsilon(i)\alpha(j)\,,$$

$$\varkappa = \varkappa_1\varkappa_2\,, \quad \varkappa_1 = \frac{\upsilon(j)}{2\beta(i-1) + \upsilon(j)}\,, \quad \varkappa_2 = \frac{\upsilon(j)[\upsilon(i) + 2\alpha(j)]}{\triangle\,(i,j)}\,,$$

we can write

$$(\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma}))^{-1}\,\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma}) = \begin{pmatrix} 1 & 0 & -\frac{\upsilon(j)}{2\beta(i-1)+\upsilon(j)} \\ -\frac{\upsilon(j)\upsilon(i)}{\triangle(i,j)} & 1 & 0 \\ -\frac{\upsilon(j)[\upsilon(i)+2\alpha(j)]}{\triangle(i,j)} & 0 & 1 \end{pmatrix}.$$

Therefore, we have

$$\det[(\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma}))^{-1}\,\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma}) - \lambda\mathbf{I}] = (1-\lambda)\left[(1-\lambda)^2 - \varkappa\right].$$

This implies

$$\lambda_{1,3} = 1 \mp \sqrt{\varkappa}\,, \quad \lambda_2 = 1\,, \tag{A.9}$$

and reduces obtaining (A.8) to the estimation of $\varkappa$.

Obviously, we have

$$\varkappa_1 = \frac{\upsilon(j)}{2\beta(i-1)+\upsilon(j)} = \frac{1}{2+\frac{2(i-1)}{3\upsilon(i-1)}} \le \frac{1}{2}\,, \quad 1 < \gamma_2 < \gamma_1 \le \overline{N}_{l-1}\,, \tag{A.10}$$

now we come to bounding $\varkappa_2$. If a function $\phi(x)$ at $x \ge a$ is positive and monotonically grows (decreases), then for the function $f(x) = \phi(x)/(\phi(x)+c)$, $c = \text{const}$, we have $f' = c\phi'/(\phi(x)+c)^2 > 0$. Therefore, $f(x)$ also monotonically grows (decreases) at $x \ge a$, and $\max_{x\in[a,b]} f(x) < f(d)$, $\forall d \ge b$. Looking at the relation

$$\varkappa_2 = \frac{\upsilon(j)[\upsilon(i)+2\alpha(j)]}{\upsilon(j)[\upsilon(i)+2\alpha(j)]+2\upsilon(i)\alpha(j)} = \frac{\phi(i,j)}{\phi(i,j)+1}\,,$$

where

$$\phi(i,j) = \upsilon(j)\left(\frac{1}{2\alpha(j)} + \frac{1}{\upsilon(i)}\right),$$

we see that for a fixed $i \le \overline{N}$ the function $\phi(i,j)$ is monotonically growing with $j \le i$ and we have the bound $\phi(i,j) \le 2$. Consequently,

$$\varkappa_2 = \frac{1}{1+\frac{1}{\phi(i,j)}} \le \frac{1}{1+\frac{1}{2}} = \frac{2}{3}\,,$$

and combining with (A.10) we get

$$\varkappa \le \frac{1}{3}\,, \quad 1 \le \gamma_2 < \gamma_1 \le \overline{N}_{l-1}\,. \tag{A.11}$$

We turn to the eigenvalues of the matrix eigenvalue problem

$$\mathbb{A}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{x} = \lambda\boldsymbol{\mathcal{A}}_{1,\triangledown}(\boldsymbol{\gamma})\mathbf{x}\,, \quad \text{for} \quad 1 \le \gamma_1 \le \gamma_2 \le \overline{N}_{l-1}\,.$$

From (A.6), it follows that the values of them are defined by (A.9) with $\varkappa$, which now is convenient to represent as

$$\varkappa = \varkappa^{(1)}\varkappa^{(2)}\,, \quad \varkappa^{(1)} = \tfrac{\upsilon(i)}{2\alpha(j)+\upsilon(i)}\,, \quad \varkappa^{(2)} = \tfrac{\upsilon(i)[\upsilon(j)+2\beta(i-1)]}{\triangle(i,j)}\,,$$

$$\triangle\,(i,j) := a_{1,1}a_{3,3} - a_{3,1}^2 = \upsilon(i)[\upsilon(j)+2\beta(i-1)] + 2\upsilon(j)\beta(i-1)\,. \tag{A.12}$$

First of all, we establish that $\varkappa^{(1)}(i,j)$, $\varkappa^{(2)}(i,j)$ and, therefore, $\varkappa(i,j)$ decrease, if $i \le j$ is fixed and j grows. For $\varkappa^{(1)}(i,j)$ this is obvious. Another multiplier can be represented in the form

$$\varkappa^{(2)}(i,j) = \frac{\phi(i,j)}{\phi(i,j)+2\beta(i-1)}\,, \quad \phi(i,j) = \upsilon(i)\left(1+\frac{2\beta(i-1)}{\upsilon(j)}\right)\,.$$

If i is fixed and j grows, then $\phi(i,j)$ decreases for $j \ge i$. As a consequence, by the judgement like the one implemented above at the estimation of $\varkappa_2$, we conclude that $\varkappa^{(2)}(i,j)$ behaves similarly. This proves also a monotonic decrease of $\varkappa(i,j)$ with growing j for a fixed i, $i \le j$. Therefore,

$$\max_{j\ge i}\varkappa(i,j) = \varkappa(i,i)\,, \tag{A.13}$$

and it is left to bound $\max_i \varkappa(i,i)$.

On the one hand, for $\varkappa^{(1)}$, we get

$$\varkappa^{(1)}(i,i) = \frac{1}{2} - \frac{4i+3}{2(6i^2+2i+2)} \le \frac{1}{2}\,. \tag{A.14}$$

On the other hand, in view of (A.12)

$$\varkappa^{(2)}(i,i) = \frac{\upsilon(i)+2\beta(i-1)}{\upsilon(i)+4\beta(i-1)}\,,$$

and, with the notation $k = i-1$, we have

$$\upsilon(i)+2\beta(i-1) = 2k^2+\frac{8}{3}k+\frac{4}{3}\,, \quad \upsilon(i)+4\beta(i-1) = 3k^2+\frac{10}{3}k+\frac{3}{2}\,.$$

Therefore,

$$\varkappa^{(2)}(k+1,k+1) = \frac{2}{3} - \frac{12k+1}{18k^2+20k+9} \le \frac{2}{3}\,, \quad 0 \le k \le i-1\,, \tag{A.15}$$

and, by combining with (A.13) and (A.14), it follows

$$\varkappa(i,j) \le \frac{1}{3}\,, \quad 1 \le i \le j \le \overline{N}_{l-1}\,. \tag{A.16}$$

A direct consequence of (A.16) and (A.9) are the inequalities (A.8) for superelements $\mathcal{E}_\triangledown$ with the constants c_1, c_2 the same as in (8.35). If to

take into account the symmetry properties of the eigenvalue problems, we conclude that the inequalities (A.8) for superelements $\mathcal{E}_{\vartriangle}$ hold as well. We did not paid attention to the Dirichlet boundary condition on $\partial\pi_1$, because it reduces the dimensions of the eigenvalue problems and improves the constants in the spectral inequalities for the superelements having nodes on $\partial\pi_1$. We omit the bounds of these constants. Thus, the statement of Theorem 8.4 concerning the bounds for c_1 and c_2 is proved.

An analytical estimation of c_0 is more complicated. Let the system matrix $\mathbf{B}_l$ be formulated in the nodal basis of the space $\mathring{\mathcal{H}}_l$, and let $\hat{\mathbf{B}}_l = \mathbf{C}_{\pi,l}^{\top}\mathbf{B}_l\mathbf{C}_{\pi,l}$, where $\mathbf{C}_{\pi,l}$ is the transformation matrix to the basis, which is the union of the nodal bases in the subspaces $\mathcal{W}_l$ and $\mathring{\mathcal{H}}_{l-1}$. Further, let

$$\mathbb{B}_{\pi,l} = \mathrm{diag}[\mathbf{B}_{l,w},\, \mathbf{B}_{l-1}]\,.$$

If $\lambda_{\pi,l,\min}$ is the minimal eigenvalue of the eigenvalue problem

$$\hat{\mathbf{B}}_l\mathbf{x} = \lambda_{\pi,l}\mathbb{B}_{\pi,l}\,, \tag{A.17}$$

then taking into account (2.70) of Lemma 2.2, we have

$$c_0 \le \sup_{l\ge 3} c_{0,l}\,, \quad \text{with } \; c_{0,l} = (1-\lambda_{\pi,l,\min})^2\,.$$

We consider $\hbar^2\hat{\mathbb{A}}_{\square}(\boldsymbol{\gamma})$, which is the stiffness matrix of the square superelement $\mathcal{E}_{\square}(\boldsymbol{\gamma})$, transformed to the two level basis, and its 2×2 block diagonal preconditioner $\hbar^2\hat{\mathbb{B}}_{\square}(\boldsymbol{\gamma})$, see (A.19). The quadratic form $\hat{\mathbb{A}}_{\square}(\boldsymbol{\gamma})$ will be estimated from below by the quadratic form $\hat{\mathbb{B}}_{\square}(\boldsymbol{\gamma})$ with the aid of the consistent implementation of the Cauchy inequality with epsilon, the simplest form of which is

$$2v_1v_2 \le \epsilon v_1^2 + \epsilon^{-1}v_2^2\,, \quad \forall\,\epsilon > 0\,. \tag{A.18}$$

The resulting bound will be sufficiently good to approve the numbers $\mu = 3$ and the $\nu = 3$, pointed out in Theorem 8.1.

For convenience, let us introduce the matrix $\mathbb{A}_{\square}(\boldsymbol{\gamma}) = \hbar^{-2}\mathbf{B}_{\square}(\boldsymbol{\gamma})$ and note that, below, we omit box in the lower index. The upper block on diagonal of the matrix $\mathbb{A}(\boldsymbol{\gamma})$ is

$$\mathbb{A}_1(\boldsymbol{\gamma}) =$$

$$\begin{pmatrix} 2\beta(i-1)+ \\ +\alpha(j-1)+\beta(j) & 0 & 0 & 0 & -\alpha(j-1)- \\ & & & & -\beta(j) \\ & \alpha(i-1)+ & & & \\ & +\beta(i)+2\alpha(j) & 0 & 0 & -\alpha(i-1)- \\ & & & & -\beta(i) \\ & & \alpha(i-1)+ & & \\ & & +2\alpha(i)+\beta(j) & 0 & -\alpha(j-1)- \\ & & & & -\beta(j) \\ & \text{SYM} & & \alpha(i-1)+ & -\alpha(i-1)- \\ & & & +\beta(i)+2\beta(j-1) & -\beta(i) \\ & & & & \\ & & & & 2[\alpha(i-1)+ \\ & & & & +\beta(i)+ \\ & & & & +\alpha(j-1)+ \\ & & & & +\beta(j)] \end{pmatrix}.$$

The lower block on the diagonal of $\mathbb{A}(\boldsymbol{\gamma})$ is the diagonal matrix

$$\mathbb{A}_2(\boldsymbol{\gamma}) =$$
$$= \operatorname{diag}\left[\,\beta(i-1)+\alpha(j)\,,\,\alpha(i)+\alpha(j)\,,\,\alpha(i)+\beta(j-1),\,\beta(i-1)+\beta(j-1)\,\right],$$

whereas the off-diagonal block are given by

$$\mathbb{A}_{12}(\boldsymbol{\gamma}) = \begin{pmatrix} -\beta(i-1) & 0 & 0 & -\beta(i-1) \\ -\alpha(j) & -\alpha(j) & 0 & 0 \\ 0 & -\alpha(i) & -\alpha(i) & 0 \\ 0 & 0 & -\beta(j-1) & -\beta(j-1) \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

By definition, $\hat{\mathbb{A}}(\boldsymbol{\gamma}) = \mathbf{C}^\top \mathbb{A}(\boldsymbol{\gamma})\,\mathbf{C}$, where $\mathbf{C}$ is the transformation matrix

$$\mathbf{C} = \begin{pmatrix} 1 & & & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ & 1 & & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ & & 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ & & & 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ & & & & 1 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & & & 1 & & & \\ & & & & & & 1 & & \\ & & & & & & & 1 & \\ & & & & & & & & 1 \end{pmatrix}$$

with empty entries implying zero entries. The preconditioner $\hat{\mathbb{B}}(\boldsymbol{\gamma})$ is obtained by setting $\hat{\mathbb{A}}_{12}(\boldsymbol{\gamma}) = (\hat{\mathbb{A}}_{21}(\boldsymbol{\gamma}))^\top = \mathbf{0}$ in the matrix $\hat{\mathbb{A}}(\boldsymbol{\gamma})$.

If $\lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})$ denotes the lowest nonzero eigenvalue of the matrix eigenvalue problem

$$\hat{\mathbb{A}}(\boldsymbol{\gamma})\mathbf{x} = \lambda_{\mathcal{E}}\hat{\mathbb{B}}(\boldsymbol{\gamma})\mathbf{x}\,, \quad \hat{\mathbb{A}}(\boldsymbol{\gamma})\mathbf{x} \neq \mathbf{0}\,, \tag{A.19}$$

then, according to Lemma 2.2, we have

$$c_0 \leq (1 - \underline{\lambda}_{\mathcal{E},\min})^2\,, \quad \text{where } \underline{\lambda}_{\mathcal{E},\min} = \inf_{l\geq 1} \min_{\boldsymbol{\gamma}} \lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})\,. \tag{A.20}$$

Below we assume that the homogeneous Dirichlet boundary condition on $\partial\pi_1$ is taken into account in the equations (A.19). We use also the notation $\theta_\circ(\boldsymbol{\gamma}) = 1 - \lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})$.

The estimate for c_0 will be attained by the use of the Cauchy inequality with epsilon for the sequential eliminating nonzero entries of the matrix $\hat{\mathbb{A}}_{12}(\boldsymbol{\gamma})$. Such a process will lead to the bound $\hat{\mathbb{A}}(\boldsymbol{\gamma})$ from below by the block diagonal matrix $\hat{\mathbb{B}}(\boldsymbol{\gamma})$. The Cauchy inequality will be implemented in a matrix form for the special symmetric sign-definite 2×2 matrices and the corresponding vectors.

In order to simplify further transformations, we introduce the numbers

$$\begin{aligned} &a = 2\alpha(i), \quad &&b = 2\beta(i-1), \quad &&c = \upsilon(i),\\ &d = 2\alpha(j), \quad &&e = 2\beta(j-1), \quad &&f = \upsilon(j)\,, \end{aligned} \tag{A.21}$$

which allow us to write the block $\hat{\mathbb{A}}_1(\boldsymbol{\gamma})$ as

$$\hat{\mathbb{A}}_1(\boldsymbol{\gamma}) = \begin{pmatrix} b+d & 0 & 0 & 0 & -d \\ 0 & a+f & 0 & 0 & -a \\ 0 & 0 & c+d & 0 & -d \\ 0 & 0 & 0 & a+e & -a \\ -d & -a & -d & -a & 2(a+d) \end{pmatrix}.$$

For the two other blocks, we get

$$\hat{\mathbb{A}}_2(\boldsymbol{\gamma}) = \frac{1}{4}\begin{pmatrix} a+b+d+f & -(d+f) & 0 & -(a+b) \\ -(d+f) & a+c+d+f & -(a+c) & 0 \\ 0 & -(a+c) & a+c+d+e & -(d+e) \\ -(a+b) & 0 & -(d+e) & a+b+d+e \end{pmatrix},$$

and

$$\hat{\mathbb{A}}_{12}(\boldsymbol{\gamma}) = \frac{1}{2}\begin{pmatrix} d & -d & 0 & 0 \\ a & 0 & 0 & -a \\ 0 & 0 & d & -d \\ 0 & -a & a & 0 \\ -(a+d) & a+d & -(a+d) & a+d \end{pmatrix}.$$

Let us now introduce the matrix

$$\mathbb{I} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

and the sub-vectors

$$\mathbf{v}_k = \begin{pmatrix} v_k \\ v_{k+1} \end{pmatrix}, \quad \mathbf{v}_5 = \begin{pmatrix} v_6 \\ v_7 \end{pmatrix}, \quad \mathbf{v}_6 = \begin{pmatrix} v_6 \\ v_9 \end{pmatrix}, \quad k = 1, 2, 3, 4, 7, 8\,,$$

of a vector $\mathbf{v} = (v_1, v_2, \ldots, v_9)^\top$. We notice that the eigenvalue $\lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})$ in (A.20) is the exact constant in the inequality

$$\begin{aligned} 4\mathbf{v}^\top \hat{\mathbb{A}}_\nabla(\boldsymbol{\gamma})\mathbf{v} = 4\,\big[&d(\mathbf{v}_1^\top \mathbb{I}\,\mathbf{v}_1 + \mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_3) + a(\mathbf{v}_2^\top \mathbb{I}\,\mathbf{v}_2 + \mathbf{v}_4^\top \mathbb{I}\,\mathbf{v}_4) \\ &+d(\mathbf{v}_1^\top \mathbb{I}\,\mathbf{v}_5 + \mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_8) + a(\mathbf{v}_2^\top \mathbb{I}\,\mathbf{v}_6 + \mathbf{v}_4^\top \mathbb{I}\,\mathbf{v}_7) \\ &+bv_1^2 + fv_2^2 + cv_3^2 + ev_4^2\big] + (d+f)\mathbf{v}_5^\top \mathbb{I}\,\mathbf{v}_5 \\ &(a+b)\mathbf{v}_6^\top \mathbb{I}\,\mathbf{v}_6 + (a+c)\mathbf{v}_7^\top \mathbb{I}\,\mathbf{v}_7 + (d+e)\mathbf{v}_8^\top \mathbb{I}\,\mathbf{v}_8 \\ \geq\, \lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})\{&4\,\big[d(\mathbf{v}_1^\top \mathbb{I}\,\mathbf{v}_1 + \mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_3) + a(\mathbf{v}_2^\top \mathbb{I}\,\mathbf{v}_2 + \mathbf{v}_4^\top \mathbb{I}\,\mathbf{v}_4) \\ &+bv_1^2 + fv_2^2 + cv_3^2 + ev_4^2\big] + (d+f)\mathbf{v}_5^\top \mathbb{I}\,\mathbf{v}_5 \\ &+(a+b)\mathbf{v}_6^\top \mathbb{I}\,\mathbf{v}_6 + (a+c)\mathbf{v}_7^\top \mathbb{I}\,\mathbf{v}_7 + (d+e)\mathbf{v}_8^\top \mathbb{I}\,\mathbf{v}_8\}\,, \end{aligned} \tag{A.22}$$

which holds for all $\mathbf{v} \in R^9$.

For the first step, we apply the Cauchy inequalities with epsilon

$$\begin{aligned} 2\mathbf{v}_1^\top \mathbb{I}\,\mathbf{v}_5 &\leq \tfrac{1}{\epsilon_d}\mathbf{v}_1^\top \mathbb{I}\,\mathbf{v}_1 + \epsilon_d \mathbf{v}_5^\top \mathbb{I}\,\mathbf{v}_5\,, \\ 2\mathbf{v}_4^\top \mathbb{I}\,\mathbf{v}_7 &\leq \tfrac{1}{\epsilon_a}\mathbf{v}_4^\top \mathbb{I}\,\mathbf{v}_4 + \epsilon_a \mathbf{v}_7^\top \mathbb{I}\,\mathbf{v}_7\,, \end{aligned} \tag{A.23}$$

and define ϵ_d, and ϵ_a by means of a subsidiary value δ according to the equalities

$$\epsilon_d = \frac{\delta(d+f)}{2d} \quad \text{and} \quad \epsilon_a = \frac{\delta(a+c)}{2a}\,. \tag{A.24}$$

The choice of ϵ_d, ϵ_a is motivated by the following considerations. Suppose, $\lambda_{\nabla,\min}(\boldsymbol{\gamma})$ is defined similarly to $\lambda_{\mathcal{E},\min}(\boldsymbol{\gamma})$, but by means of the triangular superelements. It is known that, if the matrix $\mathbf{B}$ is generated by the same finite element space, but at $\varphi(x_k) \equiv 1$ in (8.1), then $(1 - \lambda_{\nabla,\min}(\boldsymbol{\gamma}))^2 = 1/2$ and, therefore, $(1 - \lambda_{\mathcal{E},\min}(\boldsymbol{\gamma}))^2 = 1/2$ for all triangular and square superelements in π_1, see [Maitre and Musy (1982)] and [Jung and Maitre (1999)]. As a consequence, in the case under consideration related to the bilinear form (8.1), we have

$$\lim_{p\to\infty} \lambda_{\mathcal{E},\min}(N_{l-1}, N_{l-1}) = 1 - \frac{1}{\sqrt{2}}\,. \tag{A.25}$$

Additionally we note that, in what follows, δ^2 will play the role of one of the bounds for c_0. This assumes that at the use of the values (A.24) we can consider only $\delta \in (1/\sqrt{2}\,, 1]$, and, for such δ and the corresponding ϵ_d, we get the inequality

$$\Delta_d := (1 - \frac{1}{2\epsilon_d}) - (1-\delta) = \delta - \frac{d}{\delta(d+f)} \geq \delta - \frac{1}{\sqrt{2}} \geq 0\,, \quad \forall\, \delta \geq 1/\sqrt{2}\,.$$

Similarly, we have

$$\Delta_a := (1 - \frac{1}{2\epsilon_a}) - (1-\delta) = \delta - \frac{a}{\delta(a+c)} \geq \delta - \frac{1}{\sqrt{2}} \geq 0\,, \quad \forall\, \delta \geq 1/\sqrt{2}\,.$$

Note, that at the first step, one could use the Cauchy inequalities for the terms $\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6$, $\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_8$ with epsilons, accordingly defined. However, the above relations will not be hold and the resulting bound of c_0 will be ruder.

Combining (A.22)–(A.24) yields the inequality

$$\begin{gathered}
4\mathbf{v}^\top \hat{\mathbb{A}}_\triangledown(\boldsymbol{\gamma})\mathbf{v} \geq 4\left[(1-\delta)\left(d\mathbf{v}_1^\top \mathbb{I}\, \mathbf{v}_1 + a\mathbf{v}_4^\top \mathbb{I}\, \mathbf{v}_4 + bv_1^2 + fv_2^2 + cv_3^2 + ev_4^2\right)\right. \\
+\Delta_d\, d\mathbf{v}_1^\top \mathbb{I}\, \mathbf{v}_1 + \Delta_a\, a\mathbf{v}_4^\top \mathbb{I}\, \mathbf{v}_4 + + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_3 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_8 + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6 \\
\left.+\delta(bv_1^2 + fv_2^2 + cv_3^2 + ev_4^2)\right] + (1-\delta)\left((d+f)\mathbf{v}_5^\top \mathbb{I}\, \mathbf{v}_5 + (a+c)\mathbf{v}_7^\top \mathbb{I}\, \mathbf{v}_7\right) \\
+(a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6 + (d+e)\mathbf{v}_8^\top \mathbb{I}\, \mathbf{v}_8\,, \quad \forall\, \mathbf{v} \in R^9.
\end{gathered} \tag{A.26}$$

and, clearly, it is sufficient to consider only those terms in the right part, which do not have multiplier $(1-\delta)$. The corresponding quadratic form $\Theta(\mathbf{v})$ is given by the formula

$$\begin{aligned}
\Theta(\mathbf{v}) = 4\left[\Delta_d\, d\mathbf{v}_1^\top \mathbb{I}\, \mathbf{v}_1 + \Delta_a\, a\mathbf{v}_4^\top \mathbb{I}\, \mathbf{v}_4 + + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_3 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_8\right. \\
\left.+a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6 + \delta(bv_1^2 + fv_2^2 + cv_3^2 + ev_4^2)\right] \\
+(a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6 + (d+e)\mathbf{v}_8^\top \mathbb{I}\, \mathbf{v}_8\,, \quad \forall\, \mathbf{v} \in R^9.
\end{aligned} \tag{A.27}$$

The choice $\epsilon_b = (\delta b + \Delta_d d)/(\Delta_d d)$ in the inequality $2v_1v_5 \leq \epsilon_b v_1^2 + \epsilon_b^{-1} v_5^2$ allows us to conclude that

$$\Delta_d\, d\mathbf{v}_1^\top \mathbb{I}\, \mathbf{v}_1 + bv_1^2 \geq \Delta_d d(1 - \epsilon_b^{-1})v_5^2 = \frac{\delta b \Delta_d d}{\delta b + \Delta_d d}\, v_5^2\,, \tag{A.28}$$

and, similarly, we get

$$\Delta_a\, a\mathbf{v}_4^\top \mathbb{I}\, \mathbf{v}_4 + e v_4^2 \geq \frac{\delta e \Delta_a a}{\delta e + \Delta_a a}\, v_5^2\,. \tag{A.29}$$

Using $(A.3)$ and $(A.21)$, we can estimate

$$m_d = \frac{\delta b \Delta_d}{\delta b + \Delta_d d} \quad \text{and} \quad m_a = \frac{\delta e \Delta_a}{\delta e + \Delta_a a}$$

as follows:

$$m_d d \leq \delta c \quad \text{and} \quad m_a a \leq \delta f\,. \tag{A.30}$$

Applying (A.28) and (A.29) to the corresponding members of (A.26), then using inequalities (A.30) and $\mathbf{I} \geq 0.5\mathbb{I}$, where $\mathbf{I}$ is the unity matrix, we get

$$\begin{aligned}\Theta(\mathbf{v}) &= 4\left[(m_d d + m_a a)v_5^2 + + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_3 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_8 + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6\right] \\ &\quad + (a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6 + (d+e)\mathbf{v}_8^\top \mathbb{I}\, \mathbf{v}_8 \\ &\geq 4\left[a(1+\frac{m_a}{2})\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + d(1+\frac{m_d}{2})\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_3 + d\mathbf{v}_3^\top \mathbb{I}\, \mathbf{v}_8 + a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6\right] \\ &\quad + (a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6 + (d+e)\mathbf{v}_8^\top \mathbb{I}\, \mathbf{v}_8\,, \quad \forall\, \mathbf{v} \in R^9. \end{aligned} \tag{A.31}$$

Inequality (A.31) implies that it is sufficient to consider

$$\Theta_{1,a}(\mathbf{v}) = 4a\left[(1+\frac{m_a}{2})\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + \mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6\right] + (a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6 \tag{A.32}$$

for $\mathbf{v} \in R^9$, because the remaining terms in the right part can be treated in the same way.

Let us apply the Cauchy inequality with ε_a to the term $a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_6$ with ε_a satisfying the equation

$$\frac{4a(1+m_a/2) - 2a/\varepsilon_a}{4a} = \frac{a+b-2\varepsilon_a a}{a+b}\,.$$

This is a quadratic equation

$$\varepsilon_a^2 + \frac{m_a(a+b)}{4a}\varepsilon_a - \frac{a+b}{4a} = 0$$

with the positive root

$$\varepsilon_a = -\frac{m_a(a+b)}{8a} + \sqrt{\left[\frac{m_a(a+b)}{8a}\right]^2 + \frac{a+b}{4a}}\,.$$

Thus, in the result, we obtain

$$\Theta_{1,a} \geq (1-\delta_{1,a})\left[4a\mathbf{v}_2^\top \mathbb{I}\, \mathbf{v}_2 + (a+b)\mathbf{v}_6^\top \mathbb{I}\, \mathbf{v}_6\right]\,, \qquad \forall\, \mathbf{v} \in R^9\,, \tag{A.33}$$

where

$$\delta_{1,a} = \frac{m_a}{2} - \frac{1}{2\varepsilon_a} = \frac{2a\varepsilon_a}{a+b} = -\frac{m_a}{4} + \sqrt{\frac{m_a^2}{16} + \frac{a}{a+b}}\,. \tag{A.34}$$

In the same way, we arrive at the relations

$$\begin{aligned}\Theta_{1,d} &:= 4d\left[(1+\frac{m_d}{2})\mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_3 + \mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_8\right] + (d+e)\mathbf{v}_8^\top \mathbb{I}\,\mathbf{v}_8 \\ &\geq (1-\delta_{1,d})\left[4d\mathbf{v}_3^\top \mathbb{I}\,\mathbf{v}_3 + (d+e)\mathbf{v}_8^\top \mathbb{I}\,\mathbf{v}_8\right]\,, \quad \forall\,\mathbf{v}\in R^9,\end{aligned} \tag{A.35}$$

with

$$\delta_{1,d} = \frac{m_d}{2} - \frac{1}{2\varepsilon_d} = \frac{2d\varepsilon_d}{d+e} = -\frac{m_d}{4} + \sqrt{\frac{m_d^2}{16} + \frac{d}{d+e}}\,.$$

From (A.26), (A.27), (A.31)-(A.35), it follows that

$$\hat{\mathbb{A}}(\boldsymbol{\gamma}) \geq (1-\delta_0)\hat{\mathbb{B}}(\boldsymbol{\gamma})\,,$$

with $\max[\delta(\boldsymbol{\gamma}), \delta_{1,a}(\boldsymbol{\gamma}), \delta_{1,d}(\boldsymbol{\gamma})]$. Therefore, $1-\delta_0$ is the bound from below for $\lambda^{(l)}_{\mathcal{E},\min}(\boldsymbol{\gamma})$, and

$$c_0 \leq \overline{\delta}_0^2\,, \quad \text{with} \quad \overline{\delta}_0 = \max_{\boldsymbol{\gamma}} \delta(\boldsymbol{\gamma})\,. \tag{A.36}$$

Now we remind that $\delta_{1,a} = \delta_{1,a}(\delta)$ and $\delta_{1,d} = \delta_{1,d}(\delta)$, with $\delta \in (1/\sqrt{2}, 1)$, and, therefore, the quality of the bound depends on the choice of δ. For completing the estimation of c_0, we can solve, *e.g.,* the equation $\delta = \delta_{1,a}(\delta)$, which is

$$\delta + \frac{m_a(\delta)}{4} = \sqrt{\frac{m_a^2(\delta)}{16} + \frac{a}{a+b}}\,, \tag{A.37}$$

as well the equation $\delta = \delta_{1,d}(\delta)$. Squaring (A.37) and inserting the expression for $m_a(\delta)$ turn (A.37) into

$$\delta^2 + \frac{1}{2}\delta m_a(\delta) - \frac{a}{a+b} = 0\,, \quad \text{with} \;\; \delta m_a(\delta) = \frac{\delta^4 e(a+c) - \delta^2 ea}{\delta^2(a+c)(a+e) - a^2}\,.$$

Multiplication by the positive multiplier $2[\delta^2(a+c)(a+e) - a^2]$ results in the biquadratic equation

$$\delta^4(a+c)(2a+3e) - \delta^2 a\left[2a + e + \frac{2}{a+b}(a+c)(a+e)\right] + \frac{2a^3}{a+b} = 0\,,$$

the solution of which is

$$\delta^2 = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}\,, \tag{A.38}$$

with

$$p = -\frac{a}{(a+c)(2a+3e)}\left[2a+e+\frac{2}{a+b}(a+c)(a+e)\right],$$
$$q = \frac{2a^3}{(a+b)(a+c)(2a+3e)}.$$

The direct calculation yields $\delta(2,2) = 0,5694$.

Now it convenient introduce the notations $\delta^{(t)}$ for solutions of the equations $\delta = \delta_{1,t}(\delta)$, $t = a, d$. In view of the symmetry of the discrete eigenvalue problem (A.17) in the respect of the straight line $x_1 = x_2$ and the way of definition of the involved deltas, we have $\max_{\boldsymbol{\gamma}} \delta^{(a)}(\boldsymbol{\gamma}) = \max_{\boldsymbol{\gamma}} \delta^{(d)}(\boldsymbol{\gamma})$. Since it is also known that $\max \theta_\circ^2(\boldsymbol{\gamma}) = \theta_\circ^2(2,2)$, one comes to the estimate

$$c_0 \le \max \theta_\circ^2(\boldsymbol{\gamma}) \le \left(\delta^{(a)}(2,2)\right)^2 = 0.5694\,, \quad \forall\, \gamma_1, \gamma_2 \ge 2\,. \tag{A.39}$$

Instead of the second inequality in (A.39) one can use $\max_{\boldsymbol{\gamma}} \delta^{(a)}(\boldsymbol{\gamma}) \le \sigma$, where σ is the bound for $\sqrt{c_0}$ obtained from (A.38) or in some other way. For approving this inequality it is sufficient for σ to satisfy the two inequalities $p \le 2\sigma^2$ and $\sigma^4 - \sigma^2 p + q \ge 0$. The first is easy to check, the second requires simple arithmetic but relatively lengthy calculations of the coefficients in the expressions of $p' = sp$, $q' = sq$, and $s = (a+b)(a+c)(2a+3e)$ as polynomials, *e.g.*, of the variables $\gamma_1' = \gamma_1 - 2$ and $\gamma_2' = \gamma_2 - 2$.

Suppose that the homogeneous Neumann boundary condition is posed on the boundary $\partial\pi_1$. Then the maximal value of $\theta_\circ(\boldsymbol{\gamma})$ can be expected at the superelement $\mathcal{E}_\square(1,1)$. In view of the monotonic behavior of the coefficients of the bilinear form $a_{\pi_1}^{(1)}(\cdot,\cdot)$, the superelements in the four corners of π_1 can be suspected in relation with this maximum. However, comparing the figures in the right parts of (A.25) and (A.39), we should exclude superelements $\mathcal{E}_\square$ in the right upper corner of π_1 from the start. The minimal aspect ratio of orthotropism on the superelements in the left upper and the right lower corners is infinitely growing with growing p, and as a consequence $\theta_\circ(\boldsymbol{\gamma}) < 1/\sqrt{2}$ for the corresponding $\boldsymbol{\gamma}$. At the Dirichlet boundary condition on $\partial\pi_1$ the stiffness matrix of the superelement $\mathcal{E}_\square(1,1)$ has the dimension 4×4. The symmetry with respect to the diagonal of the superelement, crossing the origin of the coordinates, and Lemma 2.2 allow to reduce finding $\theta_\circ(1,1)$ to linear algebraic equation, the solution of which satisfies $\theta_\circ(1,1) < 1/\sqrt{2}$. The much simpler analysis of the eigenvalue problems for superelements, corresponding $\gamma_k = 0,\ \gamma_{3-k} = 2, 3, \ldots, N,\ k = 1, 2$, gives better bounds of c_0 than in (A.39). This completes the proof of the theorem.

Appendix B

Abbreviations and Notations

B.1 Abbreviations

a.o.	arithmetic operation(s),
a.e.	almost everywhere,
AltSM	Alternating Schwarz Method,
ASM	Additive Schwarz Method,
BETI	Boundary Element Tearing and Interconnecting,
BPS	Bramble-Pasciak-Schatz preconditioner,
BPX	Bramble-Pasciak-Xu preconditioner,
BVP	Boundary Value Problems,
CG	Conjugate Gradient method,
DD	Domain Decomposition,
d.o.f.	degree(s) of freedom,
FD	finite-difference,
FDFT	Fast Discrete Fourier Transform,
FE	Finite Element,
FEM	Finite Element Method,
FETI	Finite Element Tearing and Interconnecting,
GLC	Gauss-Lobatto-Chebyshev,
GLL	Gauss-Lobatto-Legendre,
MDS	Multilevel Diagonal Scaling preconditioner,
MSM	Multiplicative Schwarz Method,
PCG	Preconditioned Conjugate Gradient method,
PDE	Partial Differential Equations,
p. d.	positive definite,
SLAE	System of Linear Algebraic Equations,
s. p. d.	symmetric positive definite.

B.2 Numbers, Vectors and Matrices

R^n	vector space of the dimension n,
$\mathbb{N}$	the set of whole numbers,
$\mathbb{N}^+$	the set of positive whole numbers,
c	a positive constant whose value can change within the same chain of inequalities,
c_{a}	an absolute constant,
$\mathrm{int}\lfloor a \rfloor$	the integer part of the number a,
$\mathrm{int}\lfloor\, a \,\rfloor_+$	the closest to a integer not less than a,
$\mathrm{int}\,\langle a \rangle$	the closest to a integer or one of the two closest integers,
$\prec, \succ, \asymp$	inequalities and two-sided inequalities hold up to **absolute constant** or constants depending on secondary insignificant parameters,
$(\mathbf{w}, \mathbf{v}) = (\mathbf{w}, \mathbf{v})_{R^n} =$ $= \mathbf{w} \cdot \mathbf{v} = \mathbf{w}^\top \mathbf{v}$	Euclidean scalar product in R^n,
$\|\|\mathbf{v}\|\| = (\mathbf{v}, \mathbf{v})^{1/2}$	Euclidean norm of a vector $\mathbf{v} \in R^n$,
$\|\|\mathbf{v}\|\|_{\mathbf{C}} = (\mathbf{C}\mathbf{v}, \mathbf{v})^{1/2}$	the norm of a vector $\mathbf{v} \in R^n$ induced by s.p.d. matrix $\mathbf{C}$ and the seminorm if $\mathbf{C}$ is a symmetric and non-negative matrix with a nontrivial kernel,
$\lambda_{\min}(\mathbf{C})$, $\lambda_{\max}(\mathbf{C})$	minimal and maximal eigenvalues of the non-negative symmetric matrix $\mathbf{C}$,
$\mathbf{C}^+$	the pseudo-inverse matrix to the matrix $\mathbf{C}$,
$\kappa[\mathbf{C}] = \mathrm{cond}\,[\mathbf{C}]$	spectral condition number of the positive definite symmetric matrix $\mathbf{C}$, *i.e.*, $\mathrm{cond}[\mathbf{C}] = \lambda_{max}(\mathbf{C})/\lambda_{min}(\mathbf{C})$,
$\ker\,[\mathbf{C}]$	kernel of a matrix $\mathbf{C}$.,
$\mathrm{diag}[a_i]_{i=1}^n$	diagonal $n \times n$ matrix with the entries $a_1, \ldots, a_N$ on the diagonal,
$\mathrm{tridiag}[b_i, a_i, c_i]_{i=1}^n$	tridiagonal $n \times n$ matrix, where a_i is the entry on the diagonal of the ith row, and b_i and c_i are the entries in the same row nearest from the left and the right, respectively,
$\lfloor \mathbf{C} \rfloor_{\mathrm{diag}}$	is that square diagonal matrix the main diagonal of which coincides with the main diagonal of the matrix $\mathbf{C}$,

$\mathbf{C} \leq \mathbf{B}$	brief record of the inequality $\mathbf{v}^\top \mathbf{C}\mathbf{v} \leq \mathbf{v}^\top \mathbf{B}\mathbf{v}$, $\forall \mathbf{v} \in R^n$, with nonnegative square $n \times n$ matrices $\mathbf{C}$ and $\mathbf{B}$, and, similarly, for $\prec, \succ, \asymp$, see above for the meaning of these signs,
$\mathbf{C} \sim \mathbf{B}$	spectrally equivalent $n \times n$ nonnegative matrices, for which the spectral inequalities $\underline{c}\,\mathbf{v}^\top \mathbf{B}\mathbf{v} \leq \mathbf{v}^\top \mathbf{C}\mathbf{v} \leq \overline{c}\,\mathbf{v}^\top \mathbf{B}\mathbf{v}$ for any vector $\mathbf{v} \in R^n$ and for some positive generic constants $\overline{c}$ and $\underline{c}$,
$\mathbf{C}^{-1}\mathbf{y}$	this usually assumes some procedure for solving the system $\mathbf{C}\mathbf{x} = \mathbf{y}$ with the system matrix $\mathbf{C}$,
$\mathbf{J}[\Theta(\xi)]$	Jacobi matrix of a mapping $\Theta(\xi)$,
$\mathbb{J}[\Theta(\xi)]$	Jacobian, *i.e.*, the determinant of the Jacobi matrix $\mathbf{J}[\Theta(\xi)]$,
ops[*procedure*]	number of arithmetic operations needed for the *procedure* in square brackets.

B.3 Intervals, Domains, Boundaries, Derivatives, Polynomial and Functional Spaces, and Norms

$\mathcal{I} = (-1, 1)$	open interval $(-1, 1)$,
$\mathcal{I}_\pi = (0, \pi)$	open interval $(0, \pi)$,
Ω, $\Gamma = \partial\Omega$	domain and its boundary,
Γ_D and Γ_N	parts of the boundary, where Dirichlet and Natural boundary conditions are specified, respectively,
$u_{,m}(x)$	$u_{,m}(x) = \partial u(x)/\partial x_m$,
$u_{,l,m}(x)$	$u_{,l,m}(x) = \partial^2 u(x)/\partial x_l \partial x_m$,
$\mathcal{P}_p$	space of polynomials of the degree not greater than p, $p \geq 1$, with respect to all variables,
$\mathcal{Q}_p$	space of polynomials of the degree not greater than p, $p \geq 1$, with respect to each variable,
span $[\phi_i]_{i=1}^N$	space spanned by the functions $\phi_1, \ldots, \phi_N$,
$Z = X \oplus Y$	direct sum of the subspaces X and Y of the space Z,
$C(\overline{\Omega})$	space of functions which are continuous on $\overline{\Omega}$,
$C^k(\overline{\Omega})$	space of functions which are k-times continuously differentiable on $\overline{\Omega}$,

$L_2(\Omega)$	space of square integrable functions,
$H^k(\Omega) = W_2^k(\Omega)$	Sobolev space of functions from $L_2(\Omega)$ the distributional derivatives of order up to k of which are also in $L_2(\Omega)$,
$\mathring{H}^k(\Omega) = \mathring{W}_2^k(\Omega)$	subspace of functions in $H^k(\Omega)$ with vanishing traces on the boundary $\Gamma = \partial\Omega$,
$H^{1/2}(\Omega)$	Sobolev-Slobodeckii space,
$H_{00}^{1/2}(\Omega)$	Sobolev-Slobodeckii space,
$\|v\|_\Omega = \|v\|_{0,\Omega}$	norm in the space $L_2(\Omega)$,
$\|v\|_{k,\Omega},\ \|v\|_{k,\Omega}$	norm and seminorm in the space $H^k(\Omega)$,
${}_{00}\|v\|_{1/2,\Omega}$	norm in the space $H_{00}^{1/2}(\Omega)$,
$W_p^k(\Omega)$	Sobolev space of functions from $L_p(\Omega)$ the distributional derivatives of order up to k of which are also in $L_p(\Omega)$,
$L_p^k(\Omega)$	the factor-space of $W_p^k(\Omega)$ by the space of all polynomials of degree at most $k-1$

We refer to [Sobolev (1991)], [Adams and Fournier (2003)], [Maz'ya and Poborchi (1998)] and [Steinbach (2008)] for definitions of the listed functional spaces, norms in these spaces and their properties.

B.4 Finite Element Method

τ_0	domain of a reference element which is the unit cube $(-1,1)^m,\ m \geq 1$, or the unilateral simplex with lengths of edges equal 1 or 2,
π_1	unit cube $(0,1)^m,\ m > 1$,
$\tau = \tau_r$	domain of a finite element,
$\mathcal{R}$	number of finite elements of a finite element assemblage,
$\mathbf{v} \leftrightarrow v$	vector representation $\mathbf{v}$ of the finite element function v also called FE isomorphism,
$x = \mathcal{X}^{(r)}(y)$	mapping $\bar{\tau}_0 \to \bar{\tau}_r$ of the reference element domain on the domain of a finite element, by means of which the finite element on τ_r is defined,
$H_k^{(r)}$	Lame's coefficient of the mapping $x = \mathcal{X}^{(r)}(y)$,
$\mathbf{A}$	reference element stiffness matrix,

$\boldsymbol{\mathcal{A}}$	preconditioner for the reference element stiffness matrix,
$\mathbf{A}_I, \mathbf{A}_B, \mathbf{A}_E, \mathbf{A}_V$	blocks on the diagonal of a reference element stiffness matrix $\mathbf{A}$ for the degrees of freedom living in the interior of reference element, its boundary, edges and vertices, respectively,
$\mathbf{A}_F, \mathbf{A}_W$	blocks on the diagonal of the 3-dimensional reference element stiffness matrix $\mathbf{A}$ for the degrees of freedom living on faces and the wire basket of the reference element,
$\mathcal{H}(\pi_1)$	space of the continuous piecewise linear functions on π_1 induced by the canonical triangulation of π_1 with the nodes in the nodes of the uniform square mesh,
$\mathring{\mathcal{H}}(\pi_1)$	subspace of functions from $\mathcal{H}(\pi_1)$ vanishing on $\partial\pi_1$,
$\mathbf{K}_r$	stiffness matrix of a finite element τ_r,
$\biguplus_r \mathbf{K}_r$	stands for the procedure of assembling, *e.g.*, of a global stiffness matrix $\mathbf{K}$ from the local stiffness matrices $\mathbf{K}_r$ of finite elements such that $\mathbf{K} = \biguplus_r \mathbf{K}_r$; sometimes this procedure is called *topological summation* of the matrices $\mathbf{K}_r$.

B.5 Domain Decomposition

Ω_j	subdomain of decomposition,
$\mathcal{D}_\varkappa$, $\tau_\varkappa^\circ$, $\tau_{0,j}$	pre-images of decomposition subdomains called also reference domains,
$x = \Upsilon_j(y)$	mapping defining subdomain Ω_j by its pre-image reference domain,
Γ_B	inter-subdomain boundary, *i.e.* $\Gamma_B = (\bigcup_{j=1} \partial\Omega_j) \setminus \partial\Omega$,
Q also $\mathcal{Q}$	number of free faces of subdomains of the decomposition,
Q_j	number of free faces of a subdomain of decomposition Ω_j,
F^k	a face,
F_j^k	a face of the subdomain Ω_j.
E_r^l	edge "l" belonging to the finite element $\overline{\tau}_r$, p. 353.

Bibliography

Adams, R. (1975). *Sobolev Spaces* (Academic Press, New York).

Adams, R. and Fournier, J. (2003). *Sobolev Spaces* (Academic Press, New York).

Ainsworth, M. (1996). A preconditioner based on domain decomposition for $h-p$ finite element approximation on quasi-uniform meshes, *SIAM J. Numer. Anal.* **33**, 4, pp. 1358–1376.

Ainsworth, M. and Guo, B. (2000). An additive schwarz preconditioner for p-version boundary element approximation of the hypersingular operator in three dimensions. *Numer. Math.* **85**, 3, pp. 343–366.

Ainsworth, M. and Senior, B. (1997). Aspects of an adaptive hp-finite element method: Adaptive strategy, conforming approximation and efficient solvers, *Comp. Methods Appl. Mech. Engrg.* **150**, 1-4, pp. 65–87.

Andreev, V. (1972). Stability of difference schemes for ellip[tic equations with respect to dirichlet boundary conditions, *Zh. Vichisl. Matem. Matem. Fyz.* **12**, pp. 598–611, (in Russian).

Andreev, V. (1973). Equivalent normalization of mesh functions from $w_2^{1/2}(\gamma)$, in *Issled. Teorii Raznost. Skhem. Ellipt. Parabol.* (Moskow State University Publishin House), pp. 6–39, (in Russian).

Anufriev, I. and Korneev, V. (2004). Fast domain decomposition method for solving 4-th order deteriorating differential equation in 3d domain, in *Mesh Methods for Boundary Value Problems and Applicationa*, pp. 8–14, proceedings of 5-th All-Russian Seminar "Mesh Methods for Boundary Value Problems and Applicationa" held at Kazan, Russia, 17–21 September 2004.

Anufriev, I. and Korneev, V. (2005). Numerical testing domain decomposition method for hp-version fem, *Isvestia VUZov* **512**, 1, pp. 10–24, (in Russian).

Anufriev, I., Korneev, V. and Rytov, A. (2003). Computer testing of fast domain decomposition algorithms for hp discretizations of 2-nd order elliptic equations, in E. Lipitakis (ed.), *Proc. of Hellenic-European Conference on Computer Mathematics and its Applications HERCMA 2003*, Vol. 1 (LEA Publs, Athens).

Argyris, J. and Kelsey, S. (1960). *Energy Theorem and Strucural Analysis* (Plenum Press, London).

Aronszajn, N. (1955). Boundary values of functions with finite Dirichlet integral, in *Conference on Partial Differential Equations, Studies in Eigenvalue Problems* (Univ. of Kansas).

Axellson, O. (1994). *Iterative Solution Methods* (Cambridge University Press, New York).

Axelsson, O. and Vassilevski, P. (1989). Algebraic multilevel preconditioning methods, i. *Num. Math.* **56**, 2–3, pp. 157–177.

Axelsson, O. and Vassilevski, P. (1990). Algebraic multilevel preconditioning methods, ii, *SIAM Journal on Numerical Analysis*.

Babich, V. and Slobodetski, L. (1956). On the boundedness of the dirichlet integral. *Dokl. Akad. Nauk SSSR* **106**, pp. 04–607.

Babuška, I. (1958). On the Schwarz algorithm in the theory of differential equations of mathematical physics, *Tchecosl. Math. J.* **8**, 83, pp. 328–342, (in Russian).

Babuška, I., Craig, A., Mandel, J. and Pitkäranta, J. (1991). Efficient preconditioning for the p-version finite element method in two dimensions, *SIAM J. Numer. Anal.* **28**, 3, pp. 624–661.

Babuška, I. and Strouboulis, T. (2001). *The Finite Element Method and its Reliability* (Clarendon Press, Oxford).

Babuška, I. and Suri, M. (1987). The h-p-version of the finite element method with quasiuniform meshes, *Modélisation mathématique et analyse numérique* **21**, 2, pp. 199–238.

Bahlmann, D. and Korneev, V. (1993). Comparison of high order accuracy finite element methods: nodal points selection and preconditioning, Preprint Nr.237/7, Technische Universität Chemnitz, Faculty of Mathematics, Chemnitz.

Bakhvalov, N. (1966). On the convergence of a relaxation method with natural constraints on the elliptic operator, *USSR Comput. Math. and Math. Phys.* **6**, 5, pp. 101–135.

Bank, R., Jimack, P., Nadeem, S. and Nepomnyaschikh, S. (2002). A weakly overlapping domain decomposition preconditioner for the finite element solution of elliptic partial differential equation, *SIAM J. Sci. Comput.* **23**, 6, pp. 1818–1242.

Bebendorf, M. (2008). *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, *Lecture Notes in Computational Science and Engineering (LNCSE)*, Vol. 63 (Springer-Verlag, Heidelberg).

Belgacem, F. B. (1994). Polynomial extensions of compatible polynomial traces in 3d, *Comput. Methods Appl. Mech. Engrg.* **116**, pp. 235–241.

Bernardi, C., Dauge, M. and Maday, Y. (1993). Polynomials in weighted sobolev spaces: Basics and trace liftings, Publications du Laboratoire d'Analyse Numerique R92039, Universite Pierre et Marie Curie, Paris.

Bernardi, C. and Maday, Y. (1989). Properties of some weighted Sobolev spaces and application to spectral approximations, *SIAM J. Numer. Anal.* **26**, 4, pp. 769–829.

Bernardi, C. and Maday, Y. (1992a). Polynomial interpolation results in Sobolev spaces, *Journal of Computational and Applied Mathematics* **43**, 1–2, pp. 53–80.

Bernardi, C. and Maday, Y. (1992b). *Spectral approximation for elliptic boundary value problems*, *Mathematiques & Applications*, Vol. iv (Springer-Verlag, Paris), (in French).

Bernardi, C. and Maday, Y. (1997). Spectral method, in P. Ciarlet and J.-L. Lions (eds.), *Handbook of numerical analysis, vol. V: Techniques of Scientific computing (Part 2)* (North-Holland), pp. 209–485.

Beuchler, S. (2001). A preconditioner for solving of the inner problem for p-version of the fem, Part II: Algebraic multigrid proof, Preprint SFB393/01-07, Technische Universitat Chemnitz, Chemnitz.

Beuchler, S. (2002). Multigrid solver for the inner problem in domain decomposition methods for p-fem, *SIAM J. Num. Anal.* **40**, 4, pp. 928–944.

Beuchler, S. and Braess, D. (2006). Improvements of some condition numbers estimates for preconditioned system in p-fem. *Numerical Linear Algebra with Applications* **13**, 7, pp. 573–588.

Beuchler, S. and Nepomnyaschikh, S. (2007). Overlapping additive Schwarz preconditioners for elliptic problems with degenerate locally anisotropic coefficients, *SIAM J. Numer. Anal.* **45**, 6, pp. 2321–2344.

Beuchler, S., Schneider, R. and Schwab, C. (2004). Multiresolution weighted norm equivalence and applications, *Numerische Mathematik* **98**, 1, pp. 67–97.

Beuchler, S. and Schöberl, J. (2005). Optimal extensions on tensor product meshes, *Applied Num. Math.* **54**, 3–4, pp. 391–405.

Beuchler, S. and Schöberl, J. (2006). New shape functions for triangular p-fem using integrated jacobi polynomials, *Numerische Mathematik* **103**, 3, pp. 339–366.

Bjørstad, P. and Mandel, J. (1991). On the spectra of sums of orthogonal projections with applications to parallel computing, *BIT* **31**, pp. 76–88.

Bjørstad, P. and Widlund, O. (1986). Iterative methods for the solution of elliptic problems on regions partitioned into substructures, *SIAM J. Numer. Anal.* **23**, 6, pp. 1093–1120.

Börgers, C. (1989). The Neumann-Dirichlet domain decomposition method with inexact solvers on the subdomains, *Numerische Mathematik* **55**, pp. 123–136.

Braess, D. (2007). *Finite Elements: Theory, Fast Solvers and Applications in Solid Mechanics* (Cambridage University Press, Cambridge).

Bramble, J. (1993). *Multigrid methods*, *Pitman Research Notes in Mathematical Sciences*, Vol. 294 (Longman Sci. Tech., Harlow, Essex).

Bramble, J. and Hilbert, S. (1971). Bounds for a class of linear functionals with applications to hermite interpolation, *Numer. Math.* **16**, pp. 362–369.

Bramble, J., Pasciak, J. and Schatz, A. (1986). The constraction of preconditioners for elliptic problems by substructuring, i, *Math. Comp.* **47**, 175, pp. 103–134.

Bramble, J., Pasciak, J. and Schatz, A. (1987). The constraction of preconditioners for elliptic problems by substructuring, ii, *Math. Comp.* **49**, 179, pp. 1–16.

Bramble, J., Pasciak, J. and Schatz, A. (1988). The constraction of preconditioners for elliptic problems by substructuring, iii, *Math. Comp.* **51**, 184, pp. 415–430.

Bramble, J., Pasciak, J. and Schatz, A. (1989). The constraction of preconditioners for elliptic problems by substructuring, iv, *Math. Comp.* **53**, 187, pp. 1–24.

Bramble, J., Pasciak, J. and Xu, J. (1990). Parallel multilevel preconditioners, *Math. Comp.* **55**, pp. 1–22.

Bramble, J. and Xu, J. (1991). Some estimates for a weighted l^2 projections, *Math. Comp.* **50**, pp. 463–476.

Bramble, J. and Zhang, X. (2000). The analysis of multigrid methods, in *Handbook of Numerical Analysis*, Vol. VII (North-Holland, Amsterdam), pp. 173–415.

Brelaz, D. (1979). New method to color the vertices of a graph, *Communications of ASM* **22**, pp. 251–256.

Brenner, S. (1999). The condition number of the Schur complement in domain decomposition, *Numerische Mathematik* **83**, pp. 187–203.

Brenner, S. (2000). Lower bounds for two-level additive Schwarz preconditioners with small overlap, *SIAM J. Sci. Comput.* **21**, 5, pp. 1657–1669.

Cai, X. (1993). An optimal two-level overlapping domain decomposition methods for elliptic problems in two and three dimensions, *SIAM J. Sci. Comput.* **14**, pp. 239–247.

Canuto, C. (1994). Stabilization of spectral methods by finite element bubble functions, *Computer Methods in Applied Mechanics and Engineering* **116**, 1–4, pp. 13–26.

Canuto, C., Hussani, M., Quarteroni, A. and Zang, A. (1988). *Spectral Methods in Fluid Dynamics* (Springer-Verlag, Berlin, Heidelberg).

Canuto, C. and Quarteroni, A. (1982). Approximation results for orthogonal polynomials in sobolev spaces, *Math. Comp.* **38**, pp. 67–86.

Carstensen, C., Kuhn, M. and Langer, U. (1998). Fast parallel solvers for symmetric boundary element domain decomposition equations, *Numerische Mathematik* **79**, pp. 321–347.

Casarin, M. (1997). Quasi-optimal schwarz methods for the conforming spectral element discretization, *SIAM Journal on Numerical Analysis* **34**, 6, pp. 2482–2502.

Chan, T. (1987). Iterative methods for the solution of elliptic problems on regions partitioned into substructures, *SIAM J. Numer. Anal.* **24**, 2, pp. 382–390.

Chan, T. and Mathew, T. (1994). Domain decomposition algorithms, *Acta Numerica* , pp. 61–143.

Chan, T., Smith, B. and Zou, J. (1996). Overlapping Schwarz methods on unstructured meshes using nonmatching coarse grids, *Numer. Math.* **73**, 2, pp. 149–167.

Ciarlet, P. (1978). *The finite element method for elliptic problems* (North-Holland Publishing Company, Amsterdam - New York - Oxford).

Ciarlet, P. and Lions, J. (eds.) (1991). *Handbook of numerical analysis. Volume II: Finite element methods*, Vol. 2 (Elsevier science Ltd.).

Ciarlet, P. and Raviart, P. (1972). Interpolation theory over curved elements with applicatinos to finite element methods, *Computer Methods in Applied Mechanics and Engineering* **1**, pp. 217–249.

Clément, P. (1975). Approximation by finite element functions using local regularization, *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* **9**, R2, pp. 77–84.

Cohen, A. (2000). Wavelet methods in numerical analysis, in P.G.Ciarlet and J.L.Lions (eds.), *Handbook of Numerical Analysis*, Vol. VII (Elsevier, Amsterdam).

Conte, S. and Boor, C. (1972). *Elementary Numerical Analysis* (McGraw-Hill, New York).

Dahmen, W. and Schneider, R. (1999). Composite wavelet bases for operator equations, *Math. Comp.* **68**, pp. 1533–1567.

Demkowicz, L., Gopalakrishnan, J. and Schöberl, J. (2008). Polynomial extension operators. part i, *SIAM J Numerical Analysis* **46**, 6, pp. 3006–3031.

Demkowicz, L., Gopalakrishnan, J. and Schöberl, J. (2009). Polynomial extension operators. part ii, *SIAM J Numerical Analysis* **47**, 5, pp. 3293–3324.

Demkowicz, L., Gopalakrishnan, J. and Schöberl, J. (2012). Polynomial extension operators. part ii, *Mathematics of Computation* **81**, pp. 1289–1326.

Demmel, J., Eisenstat, T., Gilbert, J., Li, X. and JWH, J. L. (1999). A supernodal approach to sparse partial pivoting, *SIAM J. Matrix Anal. Appl.* **20**, pp. 720–755.

Deville, M., Fischer, P. and Mund, E. (2002). *High-Order Methods for Incompressible Fluid Flow* (Cambridge University Press, Cambridge,England).

Deville, M. and Mund, E. (1985). Chebyshev pseudospectral solution of second order elliptic equations with finite element preconditioning, *J. Comput. Phys.* .

Deville, M. and Mund, E. (1990). Finite-element preconditioning for pseudospectral solutions of elliptic problems, *SIAM Journal on Scientific and Statistical Computations* **11**, pp. 311–342.

Dolgov, S., Khoromskij, B., I.Oseledets and Tyrtyshnikov, E. (2011). A reciprocal preconditioner for structured matrices arising from elliptic problems with jumping coefficients, *Linear Algebra Appl.* .

Dolgov, S., Khoromskij, B., I.Oseledets and Tyrtyshnikov, E. (2012). Low-rank tensor structure of solutions to elliptic problems with jumping coefficients, *J. of Comput. Math.* .

Douglas, C., Haase, G. and Langer, U. (2003). *A Tutorial on Elliptic PDE Solvers and Their Parallelization* (SIAM, Philadelphia).

Dryja, M. (1982). A capacitance matrix method for Dirichlet problems on polygonal regions, *Numer. Math.* **39**, 1, pp. 51–64.

Dryja, M. (1984). A finite element-capacitance matrix method for elliptic problems on regions partitioned into substructures, *Numer. Math.* **44**, pp. 153–168.

Dryja, M., M.V.Sarkis and Widlund, O. (1996). Multilevel shwarz methods for elliptic problems with discontinuous coefficients in tree dimensions, *Numer. Math.* **72**, 3, pp. 313–348.

Dryja, M., Smith, B. and Widlund, O. (1994). Schwarz analysis of iterative substructuring algorithms for elliptic problems in three dimensions, *SIAM J. Numer. Anal.* **31**, 6, pp. 1662–1694.

Dryja, M. and Widlund, O. (1987). An additive variant of the schwarz alternating method for the case of many subdomains, Technical report tr-339, also ultracomputer note 131, Department of Computer Science, Courant Institute, New York.

Dryja, M. and Widlund, O. (1994). Domain decomposition algorithms with small overlap, *SIAM J. Sci. Comput.* **15**, 3, pp. 604–620.

Dryja, M. and Widlund, O. (1995). Schwarz methods of neumann-neumann type for three-dimensional elliptic finite element problems, *Communications on Pure and Applied Mathematics* **48**, 2, pp. 121–155.

Dyakonov, E. (1989). *Minimization of computational work. Asymptotically optimal algorithms for elliptic problems* (Nauka, Moscow).

Farhat, C. and Roux, F.-X. (1994). Implicite parallel processing in structural mechanics, in J. Oden (ed.), *Computational Mechanics Advances* (North-Holland, Amsterdam), pp. 1–124.

Fedorenko, R. (1961). A relaxation method for solving elliptic difference equations, *USSR Comput. Math. and Math. Phys.* **1**, 5, pp. 1092–1096.

Fedorenko, R. (1964). The speed of convergence of one iterative process, *USSR Comput. Math. and Math. Phys.* **4**, 3, pp. 227–235.

Gagliardo, E. (1957). Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in più variabili. *Rend. Sem. Mat. Univ. Padova* **27**, pp. 284–305.

George, A. and Liu, J. (1981). *Computer solution of large sparse positive definite systems* (Pretice-Hall Inc., Englewood Cliffs, New Jersey).

Golub, G. and Mayers, D. (1984). The use of preconditioning over irregular regions, in *Computing Methods in Applied Sciences and Engineering* (North-Holland, Amsterdam, New York, Oxford), pp. 3–14.

Golub, G. H. and Ye, Q. (1999). Inexact preconditioned conjugate gradient method with inner-outer iteration, *SIAM J. Sci. Comput.* **21**, 4, pp. 1305–1320.

Gottlieb, D. and Orszag, S. (1977). *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM-CMBS).

Gradshteyn, I. and Ryzhik, I. (1979). *Tables of Integrals, Series and Products*, corrected and enlarged edition edn. (Academic Press, New York).

Grauschopf, T., Griebel, M. and Regler, H. (1997). Additive multilevel preconditioners based on bilinear interpolation, matrix-dependent geometric coarsening and algebraic multigrid coarsening for second-order elliptic PDEs, *Applied Numerical Mathematics* .

Griebel, M. and Oswald, P. (1995a). On the abstract theory of additive and multiplicative schwarz algorithms, *Numerische Mathematik* **70**, 2, pp. 163–180.

Griebel, M. and Oswald, P. (1995b). Tensor product type subspace splittings and multilevel iterative methods for anisothropic problems, *Advances Comput. Math.* **4**, pp. 171–201.

Guo, B. and Cao, W. (1997). Inexact solvers on element interfaces for the p and h-p finite element method, *Computer Methods in Applied Mechanics and Engineering* **150**, 1–4, pp. 173–189.

Guo, B. and Wang, L. (2004). Jacobi approximations in non-uniformly Jacobi-weighted Sobolev spaces, *Journal of Approximation Theory* **128**, 1, pp. 1–41.

Gupta, A. (2002). Recent advances in direct methods for solving unsymmetric sparse systems of linear equations, *ACM Trans. Math. Softw.* **28**, pp. 301–324.

Haase, G., Heise, B., Kuhn, M. and Langer, U. (1997). Adaptive domain decomposition methods for finite and boundary element equations, in W. Wendland (ed.), *Boundary Element Topics* (Springer, Berlin, Heidelberg, New York), pp. 121–147.

Haase, G., Langer, U. and Meyer, A. (1990). A new approach to the dirichlet domain decomposition method, in S. Hengst (ed.), *Fifth Multigrid Seminar, Eberswalde 1990*, pp. 1–59.

Haase, G., Langer, U. and Meyer, A. (1991). The approximate dirichlet domain decomposition method. part i and ii. *Computing* **47**, pp. 137–167.

Haase, G., Langer, U., Meyer, A. and Nepomnyaschikh, S. (1994). Hierarchical extension operators and local multigrid methods in domain decomposition preconditioners, *East-West Journal of Numerical Mathematics* **2**, pp. 173–193.

Haase, G. and Nepomnyaschikh, S. (1997). Extension explicite operators on hierarchical grids, *East-West Journal of Numerical Mathematics* **5**, pp. 231–248.

Hackbusch, W. (1985). *Multi-grid Methods and Applications* (Springer-Verlag, Berlin).

Hackbusch, W. (1994). *Iterative solution of large sparse systems of equations* (Springer-Verlag, New York).

Hackbusch, W. (2003). Direct domain decomposition using the hierarchical matrix technique, in I. Herrera, D. Keyes, O. Widlund and R. Yates (eds.), *Domain Decomposition Methods in Science and Engineering* (Published by the National Autonomous University of Mexico), pp. 39–50, URL `http://www.ddm.org/DD14/`, proceedings of the fourteenth international conference on domain decomposition methods, Cocoyoc, Mexico.

Hackbusch, W. (2004). Multigrid methods for fem and bem applications, in E. Stein, R. de Borst and T. Hughes (eds.), *Encyclopedia of Computational Mechanics. Vol 1 : Fundamentals*, *Chapter 20*, Vol. 1 (John Wiley & Sons), pp. 577–595, mGM-SURVEY.

Hackbusch, W. (2009). *Hierarchische Matrizen : Algorithmen und Analysis* (Springer, Dordrecht).

Hackbusch, W. (2012). *Tensor spaces and numerical tensor calculus*, *Springer series in computational mathematics*, Vol. 42 (Springer, Heidelberg).

Hackbusch, W., Khoromskij, B. N. and Kriemann, R. (2005). Direct schur complement method by domain decomposition based on h-matrix approximation, *Computing and Visualization in the Sciences* **8**, 3–4, pp. 179–188.

Hendrickson, B. and Leland, R. (1995). A multilevel algorithm for partitioning graphs, in *Proceedings of the 1995 ACM/IEEE conference on Supercomputing (CDROM)*, Supercomputing '95 (ACM, New York, NY, USA), ISBN

0-89791-816-9, doi:10.1145/224170.224228, URL `http://doi.acm.org/10.1145/224170.224228`.

Hesthaven, J. (1998). From electrostatics to almost optimal nodal sets for polynomial interpolation in a simplex, *SIAM J. Num. Anal.* **35**, 2, pp. 655–676.

Hsiao, G., Khoromskij, B. and Wendland, W. (2001). Preconditioning for boundary element methods in domain decomposition, *Eng. Analysis with Boundary Elements* **25**, 323–338.

Ivanov, S. and Korneev, V. (1995). Selection of the high order coordinate functions and preconditioning in the frame of the domain decomposition method, *Izvestia Vyshikh Uchebnykh Zavedeniy* **395**, 4, pp. 62–81, (in Russian).

Ivanov, S. and Korneev, V. (1996). Preconditioning in the domain decomposition methods for the p-version with the hierarchical bases, *Matematicheskoie modelirovanie (Mathematical Modeling)* **8**, 9, pp. 63–73.

Jung, M. and Langer, U. (1991). Application of multilevel methods to practical problems, *Surveys on Mathematics for Industry* **1**, pp. 217–257.

Jung, M., Langer, U., Meyer, A., Queck, W. and Schneider, M. (1989). Multigrid preconditioners and their applications, in G. Telschow (ed.), *Third Multigrid Seminar, Biesenthal 1988*, Report R–MATH–03/89 (Karl–Weierstrass–Institut, Berlin), pp. 11–52.

Jung, M. and Maitre, J. (1999). Some remarks on the constant in the strengthened cbs inequality: Estimate for hierarchical finite element discretizations of elasticity problems, *Numerical Methods for Partial Differential Equations*.

Karniadakis, G. and Sherwin, S. (1999). *Spectral hp element methods for CFD* (Oxford University Press).

Katsnelson, V. and Menshikov, V. (1973). On one analog of the alternating Schwarz method, in *Theory of Functions, Functional Analysis and their Applications*, Vol. 17 (Charkov State University Press, Charkow), pp. 206–215.

Keyes, D. and Gropp, W. (1987). A comparison of domain decomposition techniques for elliptic partial differential equations and their parallel implementation, *SIAM Journal on Scientific Computing* **8**, 2, pp. 166–202.

Khoromskij, B. (2011). Introduction to tensor numerical methods in scientific computing, Preprint 06-2011, University of Zuerich, Institute of Mathematics, URL `http://www.math.uzh.ch/fileadmin/math/preprints/0611.pdf`.

Khoromskij, B. and Wittum, G. (1999). Robust schur complement method for strongly anisotropic elliptic equations, *J. Numer. Linear Algebra with Appl.* **6**, pp. 1–33.

Khoromskij, B. and Wittum, G. (2004). *Numerical solution of elliptinterface* (Springer, Berlin - New York).

Kim, S. and Kim, S. D. (2009). Preconditioning on high-order element methods using Chebyshev – Gauss – Lobatto nodes, *Appl. Num. Math.* **59**, pp. 316–333.

Korneev, V. (1970). On the construction of variational difference schemes of high order accuracy, *Vestnik Leningradskogo Universiteta* **19**, pp. 28–40, (in Russian).

Korneev, V. (1977a). *The finite element methods of high order of accuracy* (Leningrad State University, Leningrad), (in Russian).

Korneev, V. (1977b). Iterative methods of solving systems of equations of the finite element method, *USSR Computational Mathematics and Mathematical Physics* **17**, 5, pp. 1213–1233.

Korneev, V. (1979a). Schemes of the finite element method of high orders accuracy for second-order elliptic equations in three-dimensional regions, part i. *USSR Computational Mathematics and Mathematical Physics* **19**, 4, pp. 149–162.

Korneev, V. (1979b). High order schemes of the finite elements method for second-order elliptic equations in three-dimensional domain, part ii, *USSR Computational Mathematics and Mathematical Physics* **19**, 5, pp. 54–61.

Korneev, V. (2001). Almost optimal solver for dirichlet problems on subdomains of decomposition, *Differentsialnye Uravnenia (Differential Equations)* **37**, 7, pp. 958–968, (in Russian).

Korneev, V. (2002a). Fast domain decomposition solvers for *hp* discretizations of 2-nd order elliptic equations, in *Proceedings of the International Conference on Computational Mathematics, 24–28 June, 2002, Novosibirsk* (Novosibirsk), pp. 536–547.

Korneev, V. (2002b). Local dirichlet problems on subdomains of decomposition in *hp* discretizations, and optimal algorithms for their solution, *Matematicheskoie Modelirovanie (Mathematical Modelling)* **14**, 5, pp. 51–74.

Korneev, V. (2012). On domain decomposition preconditioner of bps type for finite element discretizations of 3d elliptic equations, *Computational Mathematics and Mathematical Physics* **52**, 9, pp. 1260–1294.

Korneev, V. (2013a). Fast domain decomposition algorithms for elliptic problems with piecewise variable orthotropism, in T. Apel and O. Steinbach (eds.), *Advanced Finite Element Methods and Applications*, *Lecture Notes in Applied and Computational Mechanics*, Vol. 66 (Springer, Heidelberg - New York - Dordreht - London), pp. 57–89.

Korneev, V. (2013b). Fast domain decomposition type solver for stiffness matrices of reference p-elements, *Computational Methods in Applied Mathematics* .

Korneev, V. and Fish, J. (2000). Two-level methods for 3d problems based on aggregation, *Izvestiya VUZov, Mathematica* **462**, 11, pp. 42–61, (in underlying Russian).

Korneev, V., Flaherty, J., Oden, J. and Fish, J. (2002a). Additive Schwarz algorithms for solving hp-version finite element systems on triangular meshes, *Applied Numerical Mathematics* **43**, 3, pp. 399–421.

Korneev, V., Flaherty, J., Oden, J. and Fish, J. (2002b). Hp-version additive Schwarz algorithms on triangular meshes, *Matematicheskoie modelirovanie (Mathematical Modelling)* **14**, 2, pp. 61–94.

Korneev, V. and Jensen, S. (1997). Preconditioning of the p-version of the finite element method, *Computer Meth. Appl. Mech. Engrg.* **150**, 1–4, pp. 215–238.

Korneev, V. and Jensen, S. (1999). Domain decomposition preconditioning in the hierarchical p-version of the finite element method, *Applied Numerical Mathematics* **29**, 4, pp. 479–518.

Korneev, V. and Langer, U. (2004). Domain decomposition and preconditioning, in E. Stein, R. de Borst and T. Hughes (eds.), *Encyclopedia of Computational Mechanics. Vol 1 : Fundamentals*, *Chapter 22*, Vol. 1 (John Wiley & Sons), pp. 617–647, dD-SURVEY.

Korneev, V., Langer, U. and Xanthis, L. (2003a). Fast adaptive domain decomposition algorithms for *hp*-discretizations of 2d and 3d elliptic equations: recent advances, *Hermis-μπ: An International Journal of Computer Mathematics and its Applications* **4**, pp. 27–44.

Korneev, V., Langer, U. and Xanthis, L. (2003b). On fast domain decomposition solving procedures for *hp*-discretizations of 3d, *Computational Methods in Applied Mathematics* **3**, 4, pp. 536–559.

Korneev, V., Poborchi, S. and Salgado, A. (2007). Interface schur complement preconditioning for piecewise orthotropic discretizations with high aspect ratios, in V. Korneev (ed.), *Fast mesh methods of computational mechanics of continua* (St. Petersburg State University, "Zolotoe sechenie" Publishing House, St. Petersburg), pp. 106–159.

Korneev, V. and Rytov, A. (2005a). On existence of the essential interrelation between spectral and hierarchical elements, in *Mesh methods for boundary value problems and applications*, Proceedings of 6th Allrussian Seminar (Kazan State University, Kazan), pp. 41–150.

Korneev, V. and Rytov, A. (2005b). On existence of the essential interrelation between spectral and hierarchical elements, in *Mesh methods for boundary value problems and applications. Proceedings of 6th Allrussian seminar (Materialy 6-ogo vserossiiskogo seminara)* (Kazan State University, Kazan), pp. 141–150.

Korneev, V. and Rytov, A. (2007). Spectral discretizations of 3d elliptic problems and fast domain decomposition methods, *Computational Mathematics and Mathematical Physics* **10**, pp. 1727–1745.

Korneev, V. and Rytov, A. (2008). Fast domain decomposition algorithm for discretizations of 3d elliptic equations by spectral elements, *Comput. Methods Appl. Mech. Engrg.* **197**, pp. 1433–1446.

Korneev, V., Xanthis, L. and Anufriev, I. (2002c). Hierarchical and Lagrange hp discretizations and fast domain decomposition solvers for them, SFB Report SFB Report 02-18, Johannes Kepler University Linz, SFB F013, Linz.

Korneev, V., Xanthis, L. and Anufriev, I. (2002d). Solving finite element hp discretizations of elliptic problems by fast dd algorithm, *Trudy SPbGPU (Transactions of the St. Petersburg State Polytechnical University)* **485**, pp. 126–153.

Kraus, J. and Margenov, S. (2009). *Robust Algebraic Multilevel Methods and Algorithms*, *Radon Series Comp. Appl. Math.*, Vol. 5 (Walter de Gruyter, Berlin - NewYork).

Kwak, D., Nepomnyaschikh, S. and Pyo, H. (2004). Domain decomposition for model heterogeneous anisotropic problem, *Numerical Linear Algebra with Applications* .

Lallemand, M.-H., Steve, H. and Derviex, A. (1992). Unstructured multigriding by volume agglomeration: Current status, *Computer and Fluids* **21**, pp. 343–397.

Langer, U. and Steinbach, O. (2007). Coupled finite and boundary element domain decomposition methods, in *In "Boundary Element Analysis: Mathematical Aspects and Application", ed. by M. Schanz and O. Steinbach, Lecture Notes in Applied and Computational Mechanic, Volume 29* (Springer, Berlin), pp. 29–59.

Lebedev, V. and Agoshkov, V. (1983). *Poinkaré-Steklov operators and their applications in analysis* (Academy of Sciences USSR, Dept. of Numerical Mathematics, Moscow).

Lions, J. (1988). On the Schwarz alternating methods i. in R. Glowinski, G. Golub, G. Meurant and J. Périaux (eds.), *First International Symposium on Domain Decomposition Methods for Partial Differential Equations, held in Paris, France, January 7–9, 1987* (SIAM, Philadelphia, PA), pp. 1–42.

Lions, J. (1989). On the Schwarz alternating methods ii. in T. Chan, R. Glowinski, J. Périaux and O. Widlund (eds.), *Second International Symposium on Domain Decomposition Methods for Partial Differential Equations, held in Los Angeles, California, January 14–16, 1988* (SIAM, Philadelphia, PA), pp. 47–70.

Lions, J. (1990). On the schwarz alternating methods iii. in T. Chan, R. Glowinski, J. Périaux and O. Widlund (eds.), *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations, held in Houston, Texas, March 20–22, 1989* (SIAM, Philadelphia, PA), pp. 202–223.

Lions, J. and Magenes, E. (1973). *Non-homogeneous boundary value problems and applications* (Springer-Verlag).

Lonsdale, R. (1993). An algebraic multigrid solver for the navier–stokes equations on unstructured meshes, *Int. J. Num. Meth. Heat Fluid Flow* **3**, pp. 3–14.

Maday, Y. (1989). Relèvements de traces polynomiales et interpolations hilbertiennes entre espaces de polynomes, *C.R.Acad.Sci.Paris. Sér. I Math.* **309**, pp. 463–468.

Maitre, J. F. and Musy, F. (1982). The contraction number of a class of two–level methods, an exact evaluation for some finite element subspaces and model problems, in *Multigrid Methods, Proceedings of the Conference held at Köln–Porz, November 23–27, 1981*, *Lecture Notes in Mathematics*, Vol. 960 (Springer–Verlag, Berlin–Heidelberg–New York), pp. 535–544.

Mandel, J. (1994). Hybrid domain decomposition with unstructured subdomains, in A. Quarteroni, J. Periaux, Y. Kuznetsov and O. B. Widlund (eds.), *Domain decomposition methods in science and engineering*, Contemporary Mathematics (AMS, Providence, Rhode Island), pp. 103–112, proceedings ot the "Sixth International Conference on Domain Decomposition Methods in Science and Engineering" held at Como (Italy), June 15–19, 1992.

Mandel, J. and Lett, G. (1991). Domain decomposition preconditioning for p-version finite elelment methods with high aspect ratios, *Applied Numerical Analysis* **8**, pp. 411–425.

Mathew, T. (2008). *Domain Decomposition Methods for the Numerical Solution of Partial Differential Equations*, *Lecture Notes in Computational Science and Engineering*, Vol. 61 (Springer, Berlin - Heidelberg).

Matsokin, A. (1975). Variational-difference method for solving elliptic equations in a circle, Tech. rep., Computer Center of the Sibirian Branch of USSR Academy of Sciences, Novosibirsk.

Matsokin, A. (1980). The ficticious component method and modified difference analog of Schwarz method, in *Vychiclitel'nye methody lineinoi algebry (Numerical methods of linear algebra)* (Computer Center of the Sibirian Branch of USSR Academy of Sciences, Novosibirsk), (in Russian).

Matsokin, A. and Nepomnyaschikh, S. (1981). On the convergence of the alternating subdomain schwarz method without intersections, in Y. A. Kuzntsov (ed.), *Methods of Interpolation and Approximation* (Computer Center of the Sibirian Branch of USSR Academy of Sciences, Novosibirsk), (in Russian).

Matsokin, A. and Nepomnyaschikh, S. (1985). A schwarz alternating method in a subspace, *Soviet Mathematics* **29**, 10, pp. 78–84.

Maz'ya, V. and Poborchi, S. (1998). *Differentiable functions on bad domains* (World Scientific, Singapore - New Jersey - London - Hong Kong).

Melenk, J., Gerdes, K. and Schwab, C. (2001). Fully discrete hp-finite elements: fast quadrature, *Comput. Meth. Appl. Mech. Eng.* **190**, 32–33, pp. 4339–4364.

Meurant, G. (1999). *Computer solution of large linear systems* (Elsevier Science B.V., Amsterdam).

Mikhlin, S. G. (1951). On the Schwarz algorithm, *Dokl.Akad. Nauk S.S.S.R.* **77**, 4, pp. 569–571, (in Russian).

Miller, K. (1965). Numerical analogs to the Schwarz alternating procedure, *Numerische Mathematik* **7**, 2, pp. 91–103.

Morgenstern, D. (1956). Begründung des alternierenden verfahrens durch orthogonalprojektion, *ZAMM* **36**, 7–8, pp. 255–256, (in German).

Munoz-Sola, R. (1997). Polynomial liftings on a tetrahedron and applications to the h-p version of the finite element method in three dimensions, *SIAM J. Numer. Anal.* **34**, 1.

Nepomnyaschikh, S. (1991a). Domain decomposition methods for elliptic problems with discontinuous coefficients, in R. Glowinski, Y. Kuznetsov, G. Meurant, J. Périaux and O. B. Widlund (eds.), *Fourth International Symposium on Domain Decomposition Methods for Partial Differential Equations* (SIAM, Philadelphia), pp. 242–252, moscow, May 21–25, 1990.

Nepomnyaschikh, S. (1991b). Mesh theorems on traces, normalizations of function traces and their inversio, *Russian (cont. Sov.) Journal on Numerical Analysis and Mathematical Modeling* **6**, 3, pp. 223–242.

Nepomnyaschikh, S. (1991c). Method of splitting into subspaces for solving elliptic boundary value problems in complex–form domains, *Russian (cont. Sov.)*

Journal on Numerical Analysis and Mathematical Modeling **6**, 2, pp. 151–168.

Nevanlinna, R. (1939). über des alternierenden Verfahren von Schwarz, *Crelle's J. f. reine u. angew. Math.* **180**, pp. 121–128, (in German).

Nübel, V., Düster, A. and Rank, E. (2001). Adaptive vector integration as an efficient quadrature scheme for p-version finite element matrices, in *European conference on Computational Mechanics – ECCM-2001, June 26–29, 2001, Cracow, Poland*.

Oden, J., Patra, A. and Feng, Y. (1994). Domain decomposition solver for adaptive hp finite elements, in D. Keyes and J. Xu (eds.), *Domain Decomposition Methods in Scientific and Engineering Computing: Proceedings of the Seventh International Conference on Domain Decomposition, October 27-30, 1993, the Pennsylvania State University* (American Mathematical Soc.), pp. 295–301.

Oden, J., Patra, A. and Feng, Y. (1997). Parallel domain decomposition solver for adaptive hp finite element methods, *SIAM J. Num. Anal.* **34**, 6, pp. 2090–2118.

Orszag, A. (1980). Spectral methods for problems in complex geometries, *Journal of Computational Physics* **37**, 1.

Osmolovskii, V. and Rivkind, V. (1981). On the mathod of domain splitting for elliptic equations with jumps in coefficients, *Journal of Computer Mathematics and Mathemqatical physics* **21**, 1, pp. 35–39.

Oswald, P. (1992). On discrete norm estimates related to multilevel preconditioners in the finite element method, in K. Ivanov, P. Petrushev and B. Sendov (eds.), *Proceedings of the International Conference on Constructive Theory of Functions, Varna 1991* (Bulg. Acad. Sci., Sofia), pp. 203–241.

Oswald, P. (1994). *Multilevel Finite Element Approximation: Theory and Application*, Teubner Skripten zur Numerik (B.G. Teubner, Stuttgart).

Oswald, P. (1999a). Interface preconditioners and multilevel extension operators, in C.-H. Lai, P. Bjørstad, M. Cross and O. Widlund (eds.), *Proc. 11th Int. Conf. on Domain Decomposition Methods in Science and Engineering, Greenwich, July 20–24, 1998* (ddm.org), pp. 96–103.

Oswald, P. (1999b). On the robustness of the BPX-preconditioner with respect to jumps in the coeffcients, *Math. Comp.* **68**, pp. 633–650.

Parter, S. and Rothman, E. (1995). Preconditioning legendre spectral collocation methods for elliptic problems, *SIAM Journal on Numerical Analysis* **32**, 2, pp. 333–385.

Pavarino, L. (1994). Additive schwartz methods for the p–version finite element method, *Numerische Mathematik* **66**, 4, pp. 493–515.

Pavarino, L. and Widlund, O. (1996). Polylogarithmic bound for an iterative substructuring method for spectral elements in three dimensions, *SIAM Journal on Numerical Analysis* **33**, 4, pp. 1303–1335.

Pavarino, L. and Widlund, O. (1999a). Iterative substructuring methods for spectral element discretizations of elliptic systems i: Compressible linear elasticity. *SIAM Journal on Numerical Analysis* **37**, 2, pp. 353–374.

Pavarino, L. and Widlund, O. (1999b). Iterative substructuring methods for spectral element discretizations of elliptic systems ii: Mixed methods for linear elasticity and stokes flow. *SIAM Journal on Numerical Analysis* **37**, 2, pp. 375–402.

Pavarino, L. and Widlund, O. (2000a). Iterative substructuring methods for spectral element discretizations of elliptic systems i: Compressible linear elasticity, *SIAM Journal on Numerical Analysis* **37**, 2, pp. 353–374.

Pavarino, L. and Widlund, O. (2000b). Iterative substructuring methods for spectral element discretizations of elliptic systems ii: Compressible linear elasticity, *SIAM Journal on Numerical Analysis* **37**, 2, pp. 375–402.

Peano, A. (1975). *Hierarchies of Conforming Finite Elements*, Ph.D. thesis, Sever Institute of Technology, Washington University, St. Louis, Missouri.

Peano, A. (1976). Hierarchies of conforming finite elements for plane elasticity and plate bending, *Comp. and Maths. with Appl.* **1**, 3–4, pp. 211–224.

Peano, A. (1979). Conforming approximations of Kirchhoff plates and shells, *Int. J. Num. Meth. Engng* **14**, 9, pp. 1273–1291.

Pechstein, C. (2013). *Finite and Boundary Element Tearing and Interconnecting Solvers for Multiscale Problems*, *Lecture Notes in Computational Science and Engineering*, Vol. 90 (Springer, Berlin-Heidelberg).

Pflaum, C. (2000). Robust confergence of multilevel algorithms for convection-difusion equations, *SIAM Journal on Numerical Analysis* **37**, 2, pp. 443–469.

Pothen, A., Simon, H. and Liou, K.-P. (1990). Partitioning sparse matrices with eigenvectors of graphs, *SIAM J. Matrix Anal. Appl.* **11**, 3, pp. 430–452.

Przemieniecki, J. (1963). Matrix structural analysis of substructures, *AIAA Journal* **1**, 1, pp. 138–147.

Quarteroni, A. and Vali, A. (1999). *Domain Decomposition Methods for Partial Differential Equations* (Oxfort Sciences Publications).

Rytov, A. (2006). Numerical testing of the domain decomposition method for deteriorating second-order elliptic equation, *Scientific-Technical bulletin of the St. Petersburg State Polytechnical University (Nauchno-tehnicheskie vedomosti SPbGPU)* **4**, pp. 1–13.

Samarskii, A. and Nikolayev, E. (1989). *Numerical methods for grid equations. Vol. II: Iterative methods* (Birkhäuser Verlag).

Schieweck, N. (1986). A multigrid convergence proof by a strengthened Cauchy inequality for symmetric elliptic boundary value problems, in G. Telschow (ed.), *Second Multigrid Seminar, Garzau 1985*, Report R–MATH–08/86 (Karl–Weierstrass–Institut, Berlin), pp. 289–309.

Schwab, C. (1998). *p and hp Finite Element Methods: Theory and Applications in Solid and Fluid Mechanics* (Clarendon Press).

Schwarz, H. (1869). Über einige Abbildungseigenschaften, *J. Reine Angew. Numer. Math.* **70**, pp. 105–120.

Scott, L. and Zhang, S. (1990). Finite element interpolation of nonsmooth functions satisfying boundary conditions, *Math. Comput.* **54**, pp. 483–493.

Shaidurov, V. (1995). *Multigrid methods for finite elements*, *Mathematics and its Applications*, Vol. 318 (Springer, Berlin, Heidelberg).

Shen, J., Wang, F. and Xu, J. (2000). A finite element multigrid preconditioner for chebyshev-collocation methods, *Applied Numerical Mathematics* **33**, 1-4, pp. 471–477.

Smelov, V. (1980). Justification of iterative process over subdomains in odd p_{2n+1}-approximation for problems of convection theory, Tech. rep., Computer Center of the Sibirian Branch of USSR Academy of Sciences.

Smith, B. (1991). A domain decomposition algorithm for elliptic problems in three dimentions, *Numerische Mathematik* **60**, 4, pp. 219–234.

Smith, B. (1993). A parallel implementation of an iterative substructuring algorithm in three dimensions, *SIAM J. Sci. Comput.* **14**, 2, pp. 406–423.

Smith, B. and Widlund, O. (1990). A domain decomposition algorithm using a hierarvhical basis, *SIAM J. Sci. Stat. Comput.* **11**, 6, pp. 1212–1220.

Smith, B. F., Bjørstad, P. E. and Gropp, W. (1996). *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations* (Cambridge University Press).

Sobolev, S. (1991). *Some Applications of Functional Analysis in Mathematicsl Physics*, *Translations of Mathematical Monographs*, Vol. 90, third edition edn. (American Mathematical Society, Providence, Rhode Island).

Sobolev, S. L. (1936). L'Algorithme de Schwarz dans la Théorie de l'Elasticité, *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* **IV**, (XIII) 6, pp. 243–246.

Starius, G. (1977). Composite mesh difference methods for elliptic boundary value problems, *Numerische Mathematik* **28**, 2, pp. 243–258.

Steinbach, O. (2003). *Stability Estimates for Hybrid Coupled Domain Decompostion Methods* (Springer, Heidelberg, New York).

Steinbach, O. (2008). *Numerical Approximation Methods for Elliptic Boundary Value Problems: Finite and Boundary Elements* (Springer, New York).

Stüben, K. (2001). A review of algebraic multigrid, *Journal of Computational and Applied Mathematics* **128**, pp. 281–309.

Szabo, B. and Babuška, I. (1987). *Finite Element Analysis* (Wiley Interscience, New York).

Szego, G. (1975). *Orthogonal Polynomials*, Amer. Math. Soc. Colloq. (Amer. Math. Soc., Providence, RI).

Tallec, P. L. (1994). Domain decomposition methods in computational mechanics, in J. Oden (ed.), *Computational Mechanics Advances*, Vol. 1 (North-Holland), pp. 121–220.

Tong, C., Chan, T. and Kuo, C. (1991). A decomposition preconditioner based on a change to a multilevel nodal basis, *SIAM J. Sci. Stat. Comput.* **12**, 6, pp. 1486–1495.

Toselli, A. and Widlund, O. (2005). *Domain Decoposition Methods - Algorithms and Theory*, *Springer Series in Computational Mathematics*, Vol. 34 (Springer, Berlin, Heidelberg).

Tsvik, L. (1975). A generalization of the Schwarz algorithm on the case of domains joint without overlap, *Doklady Academy of Sciences USSR* **224**, 2, pp. 309–312.

Vassilevski, P. S. (2008). *Multilevel Block Factorization Preconditioners: Matrix-based Analysis and Algorithms for Solving Finite Element Equations* (Springer, New York).

Verfürth, R. (1996). *A review of a posteriory estimates and adaptive mesh-refining techniques* (Wiley and Teubner, Stuttgart).

Widlund, O. (1992). Some schwarz methods for symmetric and nonsymmetric elliptic problems, in D. Keyes, T. Chan, G. Meurant, J. Scroggs and R. Voig (eds.), *Fifth International Symposium on Domain Decomposition Methods for Partial Differential Equations* (SIAM, Philadelphia), pp. 19–36.

Widlund, O. (1997). Preconditioners for spectral and mortar finite element methods, in R. Glowinski, J. Périaux, Z.-C. Shi and O. Widlund (eds.), *Proceedings of the Eighth International Conference on Domain Decomposition Methods* (John Wiley & Sons, Strasbourg), pp. 19–32.

Xu, J. (1989). *Theory of multilevel methods*, Ph.D. thesis, Cornell Unversity.

Xu, J. (1992). Iterative methods by space decomposition and subspace correction: A unifying approach, *SIAM Review* **34**, pp. 581–613.

Xu, J. and Long, Y. (1988). The error analysis and the improved algorithms for the infinite element method, *Chinese J. Numer. Math. Appl.* **10**, pp. 71–82.

Xu, J. and Zikatanov, L. (2002). The method of alternating projections and the method of subspace corrections in hilbert space, *J. Amer. Math. Soc.* **15**, 3, pp. 573–597.

Xu, J. and Zou, J. (1998). Some nonoverlapping domain decomposition methods, *SIAM Review* **40**, 4, pp. 857–914.

Zhang, X. (1991). *Studies in domain decomposition: Multilevel methods and biharmonic Dirichlet problem*, Ph.D. thesis, Courant Institute of Mathematical Science, New York.

Zhang, X. (1992). Multilevel Schwarz methods, *Numerische Mathematik* **63**, 4, pp. 521–539.

Index