A finite element method for surface PDEs: matrix properties

Maxim A. Olshanskii · Arnold Reusken

Received: 3 October 2008 / Revised: 19 June 2009 / Published online: 11 September 2009 © Springer-Verlag 2009

Abstract We consider a recently introduced new finite element approach for the discretization of elliptic partial differential equations on surfaces. The main idea of this method is to use finite element spaces that are induced by triangulations of an "outer" domain to discretize the partial differential equation on the surface. The method is particularly suitable for problems in which there is a coupling with a problem in an outer domain that contains the surface, for example, two-phase flow problems. It has been proved that the method has optimal order of convergence both in the H^1 and in the L^2 -norm. In this paper, we address linear algebra aspects of this new finite element method. In particular the conditioning of the mass and stiffness matrix is investigated. For the two-dimensional case we present an analysis which proves that the (effective) spectral condition number of the diagonally scaled mass matrix and the diagonally scaled stiffness matrix behaves like $h^{-3} |\ln h|$ and $h^{-2} |\ln h|$, respectively, where *h* is the mesh size of the outer triangulation.

Mathematics Subject Classification (2000) 58J32 · 65N15 · 65N30 · 76D45 · 76T99

M. A. Olshanskii

Department of Mechanics and Mathematics,

Moscow State M.V. Lomonosov University, 119899 Moscow, Russia e-mail: Maxim.Olshanskii@mtu-net.ru

A. Reusken (⊠) Institut für Geometrie und Praktische Mathematik, RWTH-Aachen University, 52056 Aachen, Germany e-mail: reusken@igpm.rwth-aachen.de

Partially supported by the the Russian Foundation for Basic Research through the projects 08-01-00159 and 09-01-00115. This work was supported by the German Research Foundation through SFB 540.

1 Introduction

Certain mathematical models involve elliptic partial differential equations posed on surfaces. This occurs, for example, in multiphase fluids if one takes so-called surface active agents (surfactants) into account. These surfactants induce tangential surface tension forces and thus cause Marangoni phenomena [6,7]. In mathematical models surface equations are often coupled with other equations that are formulated in a (fixed) domain which contains the surface. In such a setting a common approach is to use a splitting scheme that allows to solve at each time step a sequence of simpler (decoupled) equations. Doing so one has to solve numerically at each time step an elliptic type of equation on a surface. The surface may vary from one time step to another and usually only some discrete approximation of the surface is (implicitly) available. A well-known finite element method for solving elliptic equations on surfaces, initiated by the paper [5], consists of approximating the surface by a piecewise polygonal surface and using a finite element space on a triangulation of this discrete surface, cf. [3,6]. If the surface is changing in time, then this approach leads to time-dependent triangulations and time-dependent finite element spaces. Implementing this requires substantial data handling and programming effort. Another approach has recently been introduced in [2]. The method in that paper applies to cases in which the surface is given implicitly by some level set function and the key idea is to solve the partial differential equation on a narrow band around the surface. Unfitted finite element spaces on this narrow band are used for discretization.

In the recent paper [9] we introduced a new technique for the numerical solution of an elliptic equation posed on a hypersurface. The main idea is to use time-*in*dependent finite element spaces that are induced by triangulations of an "outer" domain to discretize the partial differential equation on the surface. This method is particularly suitable for problems in which the surface is given implicitly by a level set or VOF function and in which there is a coupling with a flow problem in a fixed outer domain. If in such problems one uses finite element techniques for the discretization of the flow equations in the outer domain, this immediately results in an easy to implement discretization method for the surface equation. If the surface varies in time, one has to recompute the surface mass and stiffness matrix using the same data structures each time. Moreover, quadrature routines that are needed for these computations are often available already, since they are needed in other surface related calculations, for example the computation of surface tension forces. Opposite to the method in [2], in the paper [9] we do not use an extension of the surface partial differential equation but instead use a restriction of the outer finite element spaces.

In [9] it is shown that this new method has optimal order of convergence in H^1 and L^2 norms. The analysis requires shape regularity of the outer triangulation, but *does not* require any type of shape regularity for discrete surface elements.

In the present paper, we address linear algebra aspects of this new finite element method. In particular the conditioning of the mass and stiffness matrix is investigated. Numerical experiments in two- and three-dimensional examples (treated in Sect. 2.2) indicate that in the 3D case *both for the diagonally scaled mass and stiffness matrix* (effective) spectral condition numbers behave as $O(h^{-2})$ and in the 2D case the behaviour of these condition numbers is $O(h^{-2})$ and $O(h^{-3})$, respectively.

493

Here *h* denotes the mesh size of the *outer* triangulation, which is assumed to be quasiuniform in a small neighbourhood of the surface. For the two-dimensional case we present an analysis which proves these conditioning properties (up to an additional logarithmic term $|\ln h|$) under certain assumptions on distribution of the nodes near the surface. The plausibility of these assumptions is discussed. We believe that this analysis can be extended to the three-dimensional case, but would require a lot of additional technical manipulations, see Sect. 3.5.

The remainder of the paper is organized as follows. In Sect. 2.1, we describe the finite element method that is introduced in [9]. In Sect. 2.2, we give results of some numerical experiments. These results illustrate the optimal order of convergence of the method and conditioning properties. In Sect. 3, we present an analysis of conditioning properties for the two-dimensional case. We introduce necessary notation in Sect. 3.1. In Sect. 3.2, we collect some preliminaries and assumptions for the analysis. Eigenvalue bounds for the diagonally scaled mass matrix are derived in Sect. 3.3. The stiffness matrix is treated in Sect. 3.4. The plausibility of the assumptions and further possible extensions of the analysis are discussed in Sect. 3.5.

2 Surface finite element method

2.1 Description of the method

In this section, we describe the finite element method from [9] for the three-dimensional case. The modifications needed for the two-dimensional case are obvious.

We assume that Ω is an open subset in \mathbb{R}^3 and Γ a connected C^2 compact hypersurface contained in Ω . For a sufficiently smooth function $g : \Omega \to \mathbb{R}$ the tangential derivative (along Γ) is defined by

$$\nabla_{\Gamma}g=\nabla g-\nabla g\cdot\mathbf{n}_{\Gamma}\,\mathbf{n}_{\Gamma}.$$

The *Laplace–Beltrami operator* on Γ is defined by

$$\Delta_{\Gamma}g := \nabla_{\Gamma} \cdot \nabla_{\Gamma}g.$$

We consider the Laplace–Beltrami problem in weak form: For given $f \in L^2(\Gamma)$ with $\int_{\Gamma} f \, d\mathbf{s} = 0$, determine $u \in H^1(\Gamma)$ with $\int_{\Gamma} u \, d\mathbf{s} = 0$ such that

$$\int_{\Gamma} \nabla_{\Gamma} u \nabla_{\Gamma} v \, \mathrm{d}\mathbf{s} = \int_{\Gamma} f v \, \mathrm{d}\mathbf{s} \quad \text{for all } v \in H^{1}(\Gamma).$$
(2.1)

The solution *u* is unique and satisfies $u \in H^2(\Gamma)$ with $||u||_{H^2(\Gamma)} \leq c ||f||_{L^2(\Gamma)}$ and a constant *c* independent of *f*, cf. [5].

For the discretization of this problem one needs an approximation Γ_h of Γ . We assume that this approximate manifold is constructed as follows. Let $\{\mathcal{T}_h\}_{h>0}$ be a family of tetrahedral triangulations of a *fixed* domain $\Omega \subset \mathbb{R}^3$ that contains Γ . These triangulations are assumed to be regular, consistent and stable [1]. Take $\mathcal{T}_h \in \{\mathcal{T}_h\}_{h>0}$.

We assume that Γ_h is a $C^{0,1}$ surface without a boundary and Γ_h can be partitioned in planar segments, triangles or quadrilaterals, consistent with the outer triangulation \mathcal{T}_h . This can be formally defined as follows. For any tetrahedron $S_T \in \mathcal{T}_h$ such that $\text{meas}_2(S_T \cap \Gamma_h) > 0$ define $T = S_T \cap \Gamma_h$. We assume that each T is *planar*, i.e., either a triangle or a quadrilateral. Thus Γ_h can be decomposed as

$$\Gamma_h = \cup_{T \in \mathcal{F}_h} T,$$

where \mathcal{F}_h is the set of all triangles or quadrilaterals T such that $T = S_T \cap \Gamma_h$ for some tetrahedron $S_T \in \mathcal{T}_h$. If T coincides with a face of an element in \mathcal{T}_h than the corresponding S_T is not unique. In this case, we chose one arbitrary but fixed tetrahedron S_T which has T as a face. We emphasize that although the family $\{\mathcal{T}_h\}_{h>0}$ is *shape-regular* the family $\{\mathcal{F}_h\}_{h>0}$ in general is *not shape-regular*. In our examples \mathcal{F}_h contains strongly deteriorated triangles that have very small angles and neighboring triangles can have very different areas, cf. Fig. 1.

The main idea of the method from [9] is that for discretization of the problem (2.1) we use a finite element space induced by the continuous linear finite elements on T_h . This is done as follows. We define a subdomain that contains Γ_h :

$$\omega_h := \bigcup_{T \in \mathcal{F}_h} S_T. \tag{2.2}$$

This subdomain in \mathbb{R}^3 is partitioned in tetrahedra that form a subset of \mathcal{T}_h . We introduce the finite element space

$$V_h := \{ v_h \in C(\omega_h) \mid v_{|S_T} \in P_1 \text{ for all } T \in \mathcal{F}_h \}.$$

$$(2.3)$$

This space induces the following space on Γ_h :

$$V_h^{\Gamma} := \{\psi_h \in H^1(\Gamma_h) \mid \exists v_h \in V_h : \psi_h = v_h|_{\Gamma_h}\}.$$
(2.4)

This space is used for a Galerkin discretization of (2.1): determine $u_h \in V_h^{\Gamma}$ with $\int_{\Gamma_h} u_h d\mathbf{s}_h = 0$ such that

$$\int_{\Gamma_h} \nabla_{\Gamma_h} u_h \nabla_{\Gamma_h} \psi_h \, \mathrm{d}\mathbf{s}_h = \int_{\Gamma_h} f_h \psi_h \, \mathrm{d}\mathbf{s}_h \quad \text{for all } \psi_h \in V_h^{\Gamma},$$

with f_h an extension of f such that $\int_{\Gamma_h} f_h d\mathbf{s}_h = 0$ (cf. [9] for details). Due to the Lax-Milgram lemma this problem has a unique solution u_h . In [9] we analyze the discretization quality of this method. In this analysis we assume Γ_h to be sufficiently close to Γ in the following sense. Let $U \subset \mathbb{R}^3$ be a neighborhood of Γ and $d : U \to \mathbb{R}$ the signed distance function: $|d(x)| = \operatorname{dist}(x, \Gamma)$. We assume that

ess
$$\sup_{x \in \Gamma_h} |d(x)| \le c_0 h^2$$
,
ess $\sup_{x \in \Gamma_h} ||\nabla d(x) - \mathbf{n}_h(x)|| \le \tilde{c}_0 h$,

 $\underline{ \mathcal{D}}$ Springer

hold, with $\mathbf{n}_h(x)$ a suitably oriented unit normal to Γ_h at $x \in \Gamma_h$. Under these assumptions the following optimal discretization error bounds are proven:

$$\|\nabla_{\Gamma_h}(u^e - u_h)\|_{L^2(\Gamma_h)} \le C \, h \|f\|_{L^2(\Gamma)},\tag{2.5}$$

$$\|u^{e} - u_{h}\|_{L^{2}(\Gamma_{h})} \le C h^{2} \|f\|_{L^{2}(\Gamma)}, \qquad (2.6)$$

with u^e a suitable extension of u and with a constant C independent of f and h.

2.2 Results of numerical experiments

In this section, we present results of a few numerical experiments. We distinguish between 3D and 2D cases. The numerical experiments suggest among other things, that in the 2D case the conditioning of matrices is more sensitive to the distribution of nodes of the outer triangulation near the surface than in the 3D case. The analysis given in Sect. 3 supports this observation.

2.2.1 3D example

As a first test problem we consider the Laplace-Beltrami equation

$$-\Delta_{\Gamma}u + u = f$$
 on Γ ,

with $\Gamma = {\mathbf{x} \in \mathbb{R}^3 \mid ||\mathbf{x}||_2 = 1}$ and $\Omega = (-2, 2)^3$. This example is taken from [2]. The zero order term is added to guarantee a unique solution. The source term *f* is taken such that the solution is given by

$$u(\mathbf{x}) = a \frac{\|\mathbf{x}\|^2}{12 + \|\mathbf{x}\|^2} \left(3x_1^2 x_2 - x_2^3 \right), \quad \mathbf{x} = (x_1, x_2, x_3) \in \Omega,$$

with $a = -\frac{13}{8}\sqrt{\frac{35}{\pi}}$. A family $\{\mathcal{T}_l\}_{l\geq 0}$ of tetrahedral triangulations of Ω is constructed as follows. We triangulate Ω by starting with a uniform subdivision into 48 tetrahedra with mesh size $h_0 = \sqrt{3}$. Then we apply an adaptive red-green refinement-algorithm (implemented in the software package DROPS [4]) in which in each refinement step the tetrahedra that contain Γ are refined such that on level l = 1, 2, ... we have

$$h_T \leq \sqrt{3} \ 2^{-l} =: h_l \text{ for all } T \in \mathcal{T}_l \text{ with } T \cap \Gamma \neq \emptyset.$$

The family $\{\mathcal{T}_l\}_{l\geq 0}$ is consistent and shape-regular. The interface Γ is the zero-level of $\varphi(\mathbf{x}) := \|\mathbf{x}\|^2 - 1$. Let *I* be the standard nodal interpolation operator on \mathcal{T}_l . The discrete interface is given by $\Gamma_{h_l} := \{\mathbf{x} \in \Omega \mid I(\varphi)(\mathbf{x}) = 0\}$. Let $\{\phi_i\}_{1\leq i\leq m}$ be the nodal basis functions corresponding to the vertices of the tetrahedra in ω_h , cf. (2.2). The entries $\int_{\Gamma_h} \nabla_{\Gamma_h} \phi_i \cdot \nabla_{\Gamma_h} \phi_j + \phi_i \phi_j d\mathbf{s}$ of the stiffness matrix are computed within machine accuracy. For the right-handside we use a quadrature-rule that is exact up to order five. The discrete problem is solved using a standard CG method with symmetric

Discretization errors reduction	Level l	$\ u-u_h\ _{L^2(\Gamma_h)}$	Factor					
	1	0.1124	_					
	2	0.03244	3.47					
	3	0.008843	3.67					
	4	0.002186	4.05					
	5	0.0005483	3.99					
	6	0.0001365	4.02					
	7	3.411e-05	4.00					



Fig. 1 Detail of the induced triangulation of Γ_h (*left*) and level lines of the discrete solution u_h

Gauss-Seidel preconditioner to a relative tolerance of 10^{-6} . The number of iterations needed on level l = 1, 2, ..., 7, is 14, 26, 53, 104, 201, 435, 849, respectively.

In [9] a discretization error analysis of this method is presented, which shows that it has optimal order of convergence, both in the H^1 - and L^2 -norm. The discretization errors in the $L^2(\Gamma_h)$ -norm are given in Table 1, cf. [9].

These results clearly show the h_l^2 behaviour as predicted by the analysis given in [9], cf. (2.6). To illustrate the fact that in this approach the triangulation of the approximate manifold Γ_h is strongly shape-irregular we show a part of this triangulation in Fig. 1. The discrete solution is visualized in Fig. 1.

The mass matrix M and stiffness matrix A have entries

$$M_{i,j} = \int_{\Gamma_h} \phi_i \phi_j \, \mathrm{d}\mathbf{s}_h, \ A_{i,j} = \int_{\Gamma_h} \nabla_{\Gamma_h} \phi_i \cdot \nabla_{\Gamma_h} \phi_j \, \mathrm{d}\mathbf{s}_h, \ 1 \leq i, j \leq m.$$

Define $\mathbf{D}_M := \operatorname{diag}(\mathbf{M}), \ \mathbf{D}_A := \operatorname{diag}(\mathbf{A})$ and the scaled matrices

$$\tilde{\mathbf{M}} := \mathbf{D}_M^{-\frac{1}{2}} \mathbf{M} \mathbf{D}_M^{-\frac{1}{2}}, \ \tilde{\mathbf{A}} := \mathbf{D}_A^{-\frac{1}{2}} \mathbf{A} \mathbf{D}_A^{-\frac{1}{2}}.$$

for different refinement levels we computed the largest and smallest eigenvalues of $\tilde{\mathbf{M}}$ and $\tilde{\mathbf{A}}$. The results are given in Tables 2 and 3.

Table 1 and error

k

Table 2Eigenvalues of scaledmass matrix \tilde{M}	Level l	т	Factor	λ1	λ2	λ_m	λ_m/λ_2	Factor
	1	112	_	3.8 e−17	0.0261	2.86	109	_
	2	472	4.2	4.0 e−17	0.0058	2.83	488	4.5
	3	1922	4.1	0	0.0012	2.83	2358	4.8
	4	7646	4.0	0	0.00029	2.83	9759	4.1
Table 3 Eigenvalues of scaled								
stiffness matrix $\tilde{\mathbf{A}}$	Level <i>l</i>	т	Factor	$\lambda_1 \lambda_2$	λ3	λ_m	λ_m/λ_3	Factor
	1	112	_	0 0	0.055	2.17	39.5	_
	2	472	4.2	0 0	0.013	2.26	174	4.4
	3	1922	4.1	0 0	0.0028	2.47	882	5.0
	4	7646	4.0	0 0	0.00069	2.61	3783	4.3
(5)	for the 3D ex	xel=1 vel=2 vel=3 of. O(k)	10 10 (W) ^y 10 10 10	-1 Correction Correcti	matrix eigenv	alues fo	r the 3D ex	ample el=1 el=2 el=3 O(k)

Fig. 2 Eigenvalue distributions for scaled mass matrix $\mathbf{\tilde{N}}$ (*left*) and for scaled stiffness matrix $\mathbf{\tilde{S}}$ (*right*) for the 3D example

k

These results show that for the scaled mass matrix there is one eigenvalue very close to or equal to zero and for the effective condition number we have $\frac{\lambda_m}{\lambda_2} \sim m \sim h_l^{-2}$. For the scaled stiffness matrix we observe that there are two eigenvalues close to or equal to zero and an effective condition number $\frac{\lambda_m}{\lambda_3} \sim m \sim h_l^{-2}$. In Fig. 2 for both matrices the eigenvalues λ_j , with $j \geq 2$ (mass matrix), $j \geq 3$ (stiffness matrix) are shown.

2.2.2 Structured 2D example

We also performed a numerical experiment with a very structured two-dimensional triangulation and a simple "surface" as illustrated in Fig. 3. The number of vertices is denoted by n_V ($n_V = 11$ in Fig. 3).

The surface is given by $\Gamma = [0, 1] = [m_1, m_{n_V-1}]$. The mesh size of the triangulation is $h = \frac{2}{n_V-3}$. The vertices $v_1, v_3, \ldots, v_{n_V-2}$ and $v_0, v_2, \ldots, v_{n_V-1}$ are on lines parallel to Γ and the distances of the upper and lower lines to Γ are given by $\frac{\delta}{2}h$



Fig. 3 Example with a uniform triangulation

Table 4Eigenvalues of scaledmass matrix $\tilde{\mathbf{M}}$	δ	n_V	λ_1	λ_2	λ_{n_V}	λ_{n_V}/λ_2	Factor
	0.3	17	0	1.01e-2	2.42	239	_
		33	0	2.20e-3	2.42	1.10e+3	4.60
		65	0	5.14e-4	2.42	4.70e+3	4.27
		129	0	1.24e-4	2.42	1.95e+4	4.13
		257	0	3.06e-5	2.42	7.89e+4	4.06
	0.5	65	0	5.14e-4	2.40	4.72e+3	-
	0.1		0	5.14e-4	2.46	4.79e+3	
	0.01		0	5.14e-4	2.50	4.86e+3	
	0.001		0	5.14e-4	2.50	4.86e+3	

δ	n_V	λ_1	λ_2	λ3	λ_{n_V}	λ_{n_V}/λ_3	Factor
0.3	17	0	0	5.25e-2	2.0	38.1	_
	33	0	0	1.54e-2	2.0	130	3.41
	65	0	0	4.27e-3	2.0	468	3.60
	129	0	0	1.13e-3	2.0	1.77e+3	3.77
	257	0	0	2.92e-4	2.0	6.85e+3	3.88
0.5	65	0	0	4.27e-3	2.0	468	_
0.1		0	0	4.27e-3	2.0	468	
0.01		0	0	4.27e-3	2.0	468	
0.001		0	0	4.27e-3	2.0	468	

Table 5	Eigenvalues of scaled
stiffness	matrix Ã

and $\frac{1-\delta}{2}h$, respectively, with a parameter $\delta \in (0, 1)$ ($\delta = \frac{1}{2}$ in Fig. 3). In this case a dimension argument immediately yields that both the mass and stiffness matrix are singular. For different values of n_V and of δ we computed the eigenvalues of the scaled mass and stiffness matrix. The results are given in Tables 4 and 5.

These results clearly suggest that the condition numbers of both the diagonally scaled mass and the diagonally scaled stiffness matrix behave like h^{-2} for $h \rightarrow 0$.





Moreover, one observes for this particular example that the conditioning is insensitive to the distance of the surface Γ to the nodes of the outer triangulation.

2.2.3 Less structured 2D examples

We consider two 1D surfaces in $\Omega := (0, 1)^2$. One is the ellipse given by

$$\frac{\left(x-\frac{1}{2}\right)^2}{a^2} + \frac{\left(y-\frac{1}{2}\right)^2}{b^2} = 1, \quad a = \frac{2}{5}, \ b = \frac{9}{40},$$

another one is the pedal curve given by

$$\tilde{x}(t) = \frac{ab^2 \cos(t)}{a^2 \sin^2 t + b^2 \cos^2 t}, \quad \tilde{y}(t) = -\frac{ba^2 \cos(t)}{a^2 \sin^2 t + b^2 \cos^2 t}, \quad x = \frac{\tilde{x} + 2}{4}, \quad y = \frac{\tilde{y} + 2}{4},$$

with $a = \frac{20}{11}$, $b = \frac{6}{11}$. In both cases we use a uniform triangulation of Ω as in Fig. 4. The pedal curve and the coarsest triangulation $(h = \sqrt{2} 2^{-3})$ are illustrated in Fig. 4. Finer triangulations on meshes with $h_l = \sqrt{2} 2^{-l}$ are obtained using regular refinement. We use surface finite element spaces as introduced above. Eigenvalue distributions of the resulting scaled mass matrix $\tilde{\mathbf{M}} = \tilde{\mathbf{M}}_l$ and scaled stiffness matrix $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_l$ for several refinement levels *l* are shown in Fig. 5.

The situation appears to be more delicate now: (i) There are few very small eigenvalues (we will call them 'outliers') which do not obey any clear asymptotic; this irregular behavior of the few smallest eigenvalues is especially well seen for the case of the ellipse. (ii) Apart from these outliers, we observe for the $m_l \times m_l$ scaled stiffness matrix $\tilde{\mathbf{A}}_l$ an eigenvalue distribution $\lambda_k(\tilde{\mathbf{A}}_l) \sim (\frac{k}{m_l})^2$, which due to $m_l \sim h_l^{-1}$ results in an $O(h_l^{-2})$ effective condition number. (iii) For the scaled mass matrix $\tilde{\mathbf{M}}_l$ we observe a different behaviour: for the major part of the spectrum we have $\lambda_k(\tilde{\mathbf{M}}_l) \sim (\frac{k}{m_l})^2$ again; for eigenvalues with $k \leq \sqrt{m_l}$, however, the behaviour appears to be of the



Fig. 5 Eigenvalue distributions for scaled mass matrix $\hat{\mathbf{M}}$ and for scaled stiffness matrix $\hat{\mathbf{A}}$ for ellipse (*left*) and pedal (*right*)

form $\lambda_k(\tilde{\mathbf{M}}_l) \sim (\frac{k}{m_l})^3$ (again few outliers may fail to obey any clear asymptotic). This leads to an $O(h_l^{-3})$ effective condition number, which is worse compared to the 3D and the regular 2D case discussed above.

Remark 1 In the analysis in Sect. 3, we only study effective condition numbers. It is well-known that for many methods the rate of convergence depends not only on the condition number but also on the eigenvalue distribution. In Fig. 5, we see that the worse asymptotics in case of the scaled mass matrix occurs only in a relatively small part of the spectrum. Furthermore in all cases we see relatively large distances between eigenvalues in the lower part of the spectrum. Besides the condition number, these properties will also affect the convergence behaviour of Krylov subspace methods applied to this type of problem. We performed an experiment where we applied the CG method to systems with $\tilde{\mathbf{M}}_l$ and $\tilde{\mathbf{A}}_l$. We take $\mathbf{b}_l = \tilde{\mathbf{M}}_l(1, 1, \ldots, 1)^T$ and starting vector equal to zero; the same for $\tilde{\mathbf{A}}_l$. We perform m_l iterations of the CG method. The convergence is measured by computing the energy norm of the error \mathbf{e}^k , i.e. $\|\mathbf{e}^k\|_{\tilde{\mathbf{M}}_l}$ and $\|\mathbf{e}^k\|_{\tilde{\mathbf{A}}_l}$, respectively. For l = 6, 7, 8 the convergence behaviour is shown in Fig. 6 (left). As a measure for the rate of convergence we computed $r(\tilde{\mathbf{M}}_l) = \frac{1}{m_l} \log(\|\mathbf{e}^{m_l}\|_{\tilde{\mathbf{M}}_l}\|\mathbf{e}^0\|_{\tilde{\mathbf{M}}_l}^{-1})$; similar for $r(\tilde{\mathbf{A}}_l)$. The results are given in the table in Fig. 6. We observe that the systems with matrix $\tilde{\mathbf{M}}_l$ are more difficult to solve than



Fig. 6 Convergence behaviour of CG for levels l = 6, 7, 8 in the first m_l iterations (*left*) and rate of convergence (*right*)

the ones with $\tilde{\mathbf{A}}_l$. For increasing *l*, the deterioration in $r(\tilde{\mathbf{M}}_l)$ is somewhat stronger, but comparable with the one for $r(\tilde{\mathbf{A}}_l)$. Thus, for $\tilde{\mathbf{M}}_l$ the deterioration is milder as predicted by the $O(h_l^{-3})$ effective condition number.

3 Analysis

This section gives the analysis supporting the $O(h^{-2})$ and $O(h^{-3})$ condition number estimates for the scaled stiffness matrix and mass matrix, respectively. The section is organized as follows. In Sect. 3.1, we define the surface mass and stiffness matrices and give an introductory example. Sect. 3.2 introduces some further definitions and notations and collects assumptions we need for our analysis. Some of these assumptions are introduced exclusively for the sake of analysis and are not expected to hold for most practical problems, while other assumptions turn out to be quite realistic. The plausibility of the assumptions is discussed in Sect. 3.5, which goes right after Sects. 3.3 and 3.4 containing the main theoretical results of the paper. Even for a simple 2D academic case the analysis appears to be rather technical. Possible extensions of theoretical results are discussed in Sect. 3.5.

3.1 Mass and stiffness matrices and notation

We take $\Gamma = [0, 1]$ and consider a family of quasi-uniform triangulations $\{\mathcal{T}_h\}_{h>0}$ as illustrated in Fig. 7, i.e., for each $T \in \mathcal{T}_h$ we have meas₁($\Gamma \cap T$) > 0 and the endpoints x = 0 and x = 1 of Γ lie on an edge of some $T \in \mathcal{T}_h$. The numbering of vertices v_i and intersection points m_i is as indicated in Fig. 7. We distinguish between the set of leafs *L* with corresponding index set ℓ and the set of nodes *N* (= vertices that are not leafs) with corresponding index set $\{1, 2, ..., n\}$. In the example in Fig. 7 we have $L = \{v_{1,1}, v_{6,1}, v_{9,1}, v_{9,2}, v_{13,1}\}, \ell = \{(1, 1), (6, 1), (9, 1), (9, 2), (13, 1)\},$



Fig. 8 Directed graph corresponding to the 1D example

 $N = \{v_1, v_2, \ldots, v_{13}\}$. Note that for $i = (i_1, i_2) \in \ell$ we have $1 \le i_1 \le n$. The set of all vertices is denoted by $V = L \cup N$, and $|V| = n_V$. The corresponding index set is denoted by $\mathcal{I} = \{1, 2, \ldots, n\} \cup \ell$. This distinction between leafs and nodes is more clear, if in the triangulation we delete all edges between vertices that are on the same side of Γ . For the example in Fig. 7 this results in a directed graph shown in Fig. 8. For each node $v_i \in N$ the number of leafs attached to v_i is denoted by l_i (in our example: $l_1 = l_6 = l_{13} = 1$, $l_9 = 2$, $l_j = 0$ for all other j). The intersection points m_j are numbered as indicated in Fig. 7. In the analysis it is convenient to use the following notation: if $v_i, v_{i+1} \in N$ we define $m_{i,0} := m_i, m_{i,l_i+1} := m_{i+1}$, and $m_{1,0} := m_{1,1}, m_{n,l_n+1} := m_{n,l_n}$. Using this, the subdivision of Γ into the intersections with the triangles $T \in \mathcal{T}_h$ can be written as

$$\Gamma = \bigcup_{1 \le i \le n} \bigcup_{1 \le j \le l_i + 1} [m_{i, j-1}, m_{i, j}].$$
(3.1)

We define $h := \sup\{\operatorname{diam}(T) \mid T \in \mathcal{T}_h\}$, $\omega_h := \bigcup\{T \mid T \in \mathcal{T}_h\}$, the linear finite element space $V_h = \{v \in C(\Omega_h) \mid v_{|T} \in \mathcal{P}_1 \text{ for all } T \in \mathcal{T}_h\}$ of dimension n_V , and the induced finite element space $V_h^{\Gamma} = \{w \in C(\Gamma) \mid w = v_{|\Gamma} \text{ for some } v \in V_h\}$ as in (2.3) and (2.4), respectively. These spaces V_h and V_h^{Γ} are called *outer* and *surface* finite element spaces, respectively.

For the implementation it is very convenient to use the nodal basis functions of the outer finite element space for representing functions in the surface finite element space. Let $\{\phi_i \mid i \in \mathcal{I}\}$ be the set of standard nodal basis functions in V_h , i.e., ϕ_i has value one at node v_i and zero values at all other $v \in V$, $v \neq v_i$. Clearly

$$V_h^{\Gamma} = \operatorname{span}\{(\phi_i)|_{\Gamma} \mid i \in \mathcal{I}\}$$

holds. A dimension argument shows that these functions are not independent and thus do not form a basis V_h^{Γ} . This set of generating functions is used for the implementation

of a finite element discretization of scalar elliptic partial differential equations on Γ , using the surface space V_h^{Γ} . The corresponding mass and stiffness matrices are given by

$$\langle \mathbf{M}\mathbf{u}, \mathbf{u} \rangle = \int_{0}^{1} u_{h}(x)^{2} dx, \quad \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle = \int_{0}^{1} u_{h}'(x)^{2} dx,$$

with $u_{h} = \sum_{i \in \mathcal{I}} u_{i}(\phi_{i})|_{\Gamma}, \quad \mathbf{u} := (u_{i})_{i \in \mathcal{I}} \in \mathbb{R}^{n_{V}}.$ (3.2)

Both matrices are singular. The effective condition number of \mathbf{M} (or \mathbf{A}) is defined as the ratio of the largest and smallest *nonzero* eigenvalue of \mathbf{M} (or \mathbf{A}). Below we derive bounds for the effective condition of diagonally scaled mass and stiffness matrices.

3.2 Preliminaries and assumptions

In this section, we derive some results that will be used in the analysis of the massand stiffness matrix in the following sections.

The following identities hold for $u \in V_h$:

$$u(m_i) = \phi_{i-1}(m_i)u(v_{i-1}) + \phi_i(m_i)u(v_i) \quad \text{for } 1 \le i \le n,$$
(3.3)

$$u(m_i) = \phi_{i_1}(m_i)u(v_{i_1}) + \phi_i(m_i)u(v_i) \quad \text{for } i = (i_1, i_2) \in \ell.$$
(3.4)

We introduce the notation

$$\widetilde{u}_{i} := \phi_{i}(m_{i})u(v_{i}) \text{ for } i \in \mathcal{I},
\psi_{i} := u(m_{i}) \text{ for } i \in \mathcal{I},
\xi_{i} := \begin{cases} \frac{\phi_{i}(m_{i+1})}{\phi_{i}(m_{i})} & \text{for } 1 \leq i \leq n-1, \\ \frac{\phi_{i_{1}}(m_{i})}{\phi_{i_{1}}(m_{i_{1}})} & \text{for } i = (i_{1}, i_{2}) \in \ell, \end{cases}$$
(3.5)

and obtain the relations

$$\psi_i = \xi_{i-1} \tilde{u}_{i-1} + \tilde{u}_i \quad \text{for } 2 \le i \le n,$$
(3.6)

$$\psi_i = \xi_i \tilde{u}_{i_1} + \tilde{u}_i \quad \text{for } i = (i_1, i_2) \in \ell.$$
 (3.7)

For $v_i = (x_i, y_i) \in V$ we denote the distance of v_i to the *x*-axis by $|y_i| =: d(v_i)$. We introduce the following assumption on the triangulations $\{\mathcal{T}_h\}_{h>0}$.

Assumption 1 For $v_i \in N$ let $v_j, v_r \in V$ be such that $v_i v_j$ and $v_i v_r$ intersect Γ . We assume:

$$\frac{d(v_j)}{d(v_r)} \le c_1, \quad \text{with } c_1 \text{ independent of } i, j, r \text{ and } h.$$
(3.8)

For the derivation of lower bounds for mass and stiffness matrices we will need a further assumption on the triangulations $\{T_h\}_{h>0}$:

Assumption 2 For j = 1, ..., n denote $g_j := \sum_{k=1}^{\ell_j + 1} (\xi_{j,k} - \xi_{j,k-1})^2$. Define, for $\alpha \in [0, 1]$:

$$N(\alpha) := \{ v_j \in N \mid d(v_j) \} \le h^{\alpha} \max\left\{ hg_j, \max_{i=j, j+2, \dots} d(v_i) \right\}.$$
 (3.9)

Assume that there is an *h*-independent constant c_1 such that $|N(\alpha)| \le c_1 h^{\alpha-1}$ for all $\alpha \in [0, 1]$.

Assumption 2 can be interpreted as a quantitative description on how the set of nodes having a certain (maximal) distance to Γ (as specified in (3.9)) becomes smaller if this distance gets smaller. The plausibility of both assumptions is discussed in Sect. 3.5. In particular, it will be shown that Assumption 2 can be replaced by a simple (although stronger) assumption on the distribution of the nodes of the outer triangulation near the surface.

In the remainder of the paper, to simplify the notation, we use $f \sim g$ iff there are generic constants $c_1 > 0$ and c_2 independent of h, such that $c_1g \leq f \leq c_2g$.

Lemma 3.1 For ξ_i as in (3.5) we have

$$\Pi_{k=j}^{i}\xi_{k} = \left(\frac{1}{d(v_{j-1})} + \frac{1}{d(v_{j})}\right)\frac{1}{\frac{1}{d(v_{i})} + \frac{1}{d(v_{i+1})}} \quad for \ 1 \le j \le i \le n-1.$$
(3.10)

Furthermore, if Assumption 1 is satisfied we have

$$\xi_i \sim 1 \ \text{for} \ 1 \le i \le n-1, \ i \in \ell.$$
 (3.11)

Proof From geometric properties we get

$$\phi_i(m_i) = \frac{d(v_{i-1})}{d(v_i) + d(v_{i-1})} \quad \text{for } 1 \le i \le n,$$
(3.12)

$$\phi_{i_1}(m_i) = \frac{d(v_i)}{d(v_{i_1}) + d(v_i)} \quad \text{for } i = (i_1, i_2) \in \ell.$$
(3.13)

Using this in the definition of ξ_i we obtain

$$\xi_{i} = \begin{cases} \frac{d(v_{i+1})}{d(v_{i-1})} \frac{d(v_{i-1}) + d(v_{i})}{d(v_{i}) + d(v_{i+1})} & \text{for } 1 \le i \le n-1, \\ \frac{d(v_{i})}{d(v_{i-1})} \frac{d(v_{i-1}) + d(v_{i-1})}{d(v_{i-1}) + d(v_{i})} & \text{for } i = (i_{1}, i_{2}) \in \ell. \end{cases}$$
(3.14)

In both cases ξ_i is of the form

$$\xi_i = a\left(\frac{\frac{1}{a}+z}{1+z}\right),$$

namely with $a = \frac{d(v_{i+1})}{d(v_{i-1})}$, $z = \frac{d(v_i)}{d(v_{i+1})}$ if $1 \le i \le n-1$, and $a = \frac{d(v_i)}{d(v_{i-1})}$, $z = \frac{d(v_i)}{d(v_i)}$ if $i \in \ell$. Note that z > 0 and from Assumption 1 it follows that $a \sim 1$. Furthermore:

$$\frac{1}{a} \le \frac{\frac{1}{a} + z}{1 + z} \le 1 \text{ for } z \ge 0, \ a \ge 1,$$

$$1 \le \frac{\frac{1}{a} + z}{1 + z} \le \frac{1}{a} \text{ for } z \ge 0, \ 0 < a \le 1$$

This yields min{a, 1} $\leq \xi_i \leq \max\{1, a\}$ and thus the result in (3.11) is proved.

For $1 \le i \le n - 1$ the representation of ξ_i in (3.14) can be rewritten as

$$\xi_i = \left(\frac{1}{d(v_{i-1})} + \frac{1}{d(v_i)}\right) \frac{1}{\frac{1}{d(v_i)} + \frac{1}{d(v_{i+1})}}.$$

Using this the result in (3.10) immediately follows.

We introduce the notation: $\Delta_i := m_{i+1} - m_i (= m_{i,l_i+1} - m_{i,0})$ for i = 1, ..., n, and $\Delta_0 := \Delta_1, \ \Delta_{n+1} := \Delta_n$. Due to quasi-uniformity of $\{\mathcal{T}_h\}_{h>0}$ and (3.8) the following holds:

$$|\operatorname{supp}(\phi_i) \cap \Gamma| \sim \Delta_{i_1} \quad \text{for } i = (i_1, i_2) \in \ell,$$

$$|\operatorname{supp}(\phi_i) \cap \Gamma| \sim \Delta_{i-1} + \Delta_i + \Delta_{i+1} \sim h \quad \text{for } 1 \le i \le n.$$

Lemma 3.2 Assume that Assumption 1 holds. Then we have

$$\|\phi_i\|_{\Gamma}^2 := \int_0^1 \phi_i(x)^2 \, dx \sim \Delta_{i_1} \phi_i(m_i)^2 \quad \text{for all } i = (i_1, i_2) \in \ell, \tag{3.15}$$

$$\|\phi_i\|_{\Gamma}^2 \sim h \,\phi_i(m_i)^2 \quad for \ 1 \le i \le n,$$
 (3.16)

$$\|(\phi_i)_x\|_{\Gamma}^2 := \int_0^1 \phi_i'(x)^2 \, dx \sim \frac{1}{\Delta_{i_1}} \phi_i(m_i)^2 \quad \text{for all } i = (i_1, i_2) \in \ell, \tag{3.17}$$

$$\|(\phi_i)_x\|_{\Gamma}^2 \sim \left(\frac{1}{\Delta_{i-1}} + \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} \left(\xi_{i,j} - \xi_{i,j-1}\right)^2 + \frac{1}{\Delta_{i+1}}\right) \phi_i(m_i)^2 \quad for \ 1 \le i \le n.$$
(3.18)

Proof First we consider $i = (i_1, i_2) =: (p, q) \in \ell$. Note that $\operatorname{supp}(\phi_i) \cap \Gamma = [m_{p,q-1}, m_{p,q+1}]$ and that $\phi_i(m_{p,q-1}) = \phi_i(m_{p,q+1}) = 0$. For a *linear* function g we have $\int_a^b g(x)^2 dx \sim (b-a)(g(a)^2 + g(b)^2)$. Thus we get

$$\int_{0}^{1} \phi_{i}(x)^{2} dx = \int_{m_{p,q-1}}^{m_{p,q}} \phi_{i}(x)^{2} dx + \int_{m_{p,q}}^{m_{p,q+1}} \phi_{i}(x)^{2} dx$$
$$\sim \phi_{i}(m_{p,q})^{2} (m_{p,q} - m_{p,q-1}) + \phi_{i}(m_{p,q})^{2} (m_{p,q+1} - m_{p,q})$$
$$= \phi_{i}(m_{i})^{2} (m_{p,q+1} - m_{p,q-1}) = \phi_{i}(m_{i})^{2} |\operatorname{supp}(\phi_{i}) \cap \Gamma| \sim \Delta_{i_{1}} \phi_{i}(m_{i})^{2}.$$

This proves the result in (3.15). Furthermore:

$$\int_{0}^{1} \phi_{i}'(x)^{2} dx = \int_{m_{p,q-1}}^{m_{p,q}} \phi_{i}'(x)^{2} dx + \int_{m_{p,q}}^{m_{p,q+1}} \phi_{i}'(x)^{2} dx$$
$$\sim \phi_{i}(m_{p,q})^{2} \left(\frac{1}{m_{p,q} - m_{p,q-1}} + \frac{1}{m_{p,q+1} - m_{p,q}}\right) \sim \frac{1}{\Delta_{i_{1}}} \phi_{i}(m_{p,q})^{2},$$

which proves the result in (3.17). Here we used the relation $\Delta_p \sim m_{p,q} - m_{p,q-1}$, which holds thanks to (3.8) and the angle condition for the outer triangulation.

We now consider $1 \le i \le n$. We use the notation $m_{0,j} = 0$ for all j and $m_{n+1,j} = 1$ for all j. The support supp $(\phi_i) \cap \Gamma = [m_{i-1,l_{i-1}}, m_{i+1,1}]$ is split into subintervals (cf. (3.1)) as:

$$[m_{i-1,l_{i-1}}, m_{i-1,l_{i-1}+1}] \cup (\bigcup_{1 \le j \le l_i+1} [m_{i,j-1}, m_{i,j}]) \cup [m_{i+1,0}, m_{i+1,1}]$$

Note that $\phi_i(m_{i-1,l_{i-1}}) = \phi_i(m_{i+1,1}) = 0$ and $m_{i-1,l_{i-1}+1} = m_i$, $m_{i+1,0} = m_{i+1}$. We obtain

$$\int_{0}^{1} \phi_{i}(x)^{2} dx = \int_{m_{i-1,l_{i-1}}}^{m_{i-1,l_{i-1}}+1} \phi_{i}(x)^{2} dx + \sum_{j=1}^{l_{i}+1} \int_{m_{i,j-1}}^{m_{i,j}} \phi_{i}(x)^{2} dx + \int_{m_{i+1,0}}^{m_{i+1,1}} \phi_{i}(x)^{2} dx$$

$$\sim \left(m_{i-1,l_{i-1}+1} - m_{i-1,l_{i-1}}\right) \phi_{i}(m_{i})^{2}$$

$$+ \sum_{j=1}^{l_{i}+1} \left(m_{i,j} - m_{i,j-1}\right) \left(\phi_{i}(m_{i,j})^{2} + \phi_{i}(m_{i,j-1})^{2}\right)$$

$$+ \left(m_{i+1,1} - m_{i+1,0}\right) \phi_{i}(m_{i+1})^{2}$$

$$= \phi_i(m_i)^2 \left[m_{i-1,l_{i-1}+1} - m_{i-1,l_{i-1}} + \sum_{j=1}^{l_i+1} (m_{i,j} - m_{i,j-1}) \left(\xi_{i,j}^2 + \xi_{i,j-1}^2 \right) + (m_{i+1,1} - m_{i+1,0}) \xi_i^2 \right],$$

with $\xi_{i,j}$, ξ_i as in (3.5), $\xi_{i,0} = \frac{\phi_i(m_{i,0})}{\phi_i(m_i)} = 1$, and for i < n, $\xi_{i,l_i+1} = \frac{\phi_i(m_{i,l_i+1})}{\phi_i(m_i)} = \frac{\phi_i(m_{i,l_i+1})}{\phi_i(m_i)} = \xi_i$. Using (3.11) we get

$$\int_{0}^{1} \phi_{i}(x)^{2} dx \sim \phi_{i}(m_{i})^{2} \left[m_{i-1,l_{i-1}+1} - m_{i-1,l_{i-1}} + \sum_{j=1}^{l_{i}+1} \left(m_{i,j} - m_{i,j-1} \right) + \left(m_{i+1,1} - m_{i+1,0} \right) \right]$$
$$= \phi_{i}(m_{i})^{2} |\operatorname{supp}(\phi_{i}) \cap \Gamma| \sim h\phi_{i}(m_{i})^{2}.$$

Hence the result in (3.16) holds. We also have:

$$\int_{0}^{1} \phi_{i}'(x)^{2} dx = \int_{m_{i-1,l_{i-1}}}^{m_{i-1,l_{i-1}}+1} \phi_{i}'(x)^{2} dx + \sum_{j=1}^{l_{i+1}} \int_{m_{i,j-1}}^{m_{i,j}} \phi_{i}'(x)^{2} dx + \int_{m_{i+1,0}}^{m_{i+1,1}} \phi_{i}'(x)^{2} dx$$
$$\sim \frac{\phi_{i}(m_{i})^{2}}{\Delta_{i-1}} + \sum_{j=1}^{l_{i+1}} \frac{(\phi_{i}(m_{i,j}) - \phi_{i}(m_{i,j-1}))^{2}}{\Delta_{i}} + \frac{\phi_{i}(m_{i+1})^{2}}{\Delta_{i+1}}$$
$$= \phi_{i}(m_{i})^{2} \left(\frac{1}{\Delta_{i-1}} + \frac{1}{\Delta_{i}} \sum_{j=1}^{l_{i+1}} (\xi_{i,j} - \xi_{i,j-1})^{2} + \frac{\xi_{i}^{2}}{\Delta_{i+1}}\right).$$

Using $\xi_i \sim 1$ this proves the result in (3.18).

3.3 Analysis for the mass matrix

In this section, we derive bounds for the (effective) condition number of the mass matrix **M** defined in (3.2). We define $\mathbf{D}_M := \operatorname{diag}(\mathbf{M}) = \operatorname{diag}(\|\phi_i\|_{\Gamma}^2)_{i \in \mathcal{I}}$. By $\langle \cdot, \cdot \rangle$ we denote the Euclidean inner product.

Lemma 3.3 Assume that Assumption 1 is satisfied. For all $\mathbf{u} = (u_i)_{i \in \mathcal{I}} \in \mathbb{R}^{n_V}$, $\mathbf{u} \neq 0$, we have

$$\frac{\langle \mathbf{M}\mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_{M}\mathbf{u}, \mathbf{u} \rangle} \sim \frac{h \sum_{i=2}^{n} \psi_{i}^{2} + \sum_{i=(i_{1},i_{2}) \in \ell} \Delta_{i_{1}} \psi_{i}^{2}}{h \sum_{i=1}^{n} \tilde{u}_{i}^{2} + \sum_{i=(i_{1},i_{2}) \in \ell} \Delta_{i_{1}} \tilde{u}_{i}^{2}}$$
(3.19)

with $\psi_i = u(m_i), u := \sum_{i \in \mathcal{I}} u_i \phi_i, \tilde{u}_i = \phi_i(m_i) u_i.$

Proof The identity $\langle \mathbf{D}_M \mathbf{u}, \mathbf{u} \rangle = \sum_{i \in \mathcal{I}} \|\phi_i\|_{\Gamma}^2 u_i^2$ follows directly from the definition of \mathbf{D}_M . Furthermore, using Lemma 3.2, we obtain:

$$\sum_{i \in \mathcal{I}} \|\phi_i\|_{\Gamma}^2 u_i^2 = \sum_{i=1}^n \|\phi_i\|_{\Gamma}^2 u_i^2 + \sum_{i \in \ell} \|\phi_i\|_{\Gamma}^2 u_i^2 \sim h \sum_{i=1}^n \phi_i(m_i)^2 u_i^2 + \sum_{i \in \ell} \Delta_{i_1} \phi_i(m_i)^2 u_i^2$$
$$= h \sum_{i=1}^n \tilde{u}_i^2 + \sum_{i \in \ell} \Delta_{i_1} \tilde{u}_i^2.$$

We now consider the nominator. For two neighboring point m_p and m_q we introduce the mesh sizes $h_p^- := m_p - m_q$ if $m_q < m_p$, $h_p^+ := m_q - m_p$ if $m_q > m_p$ and $h_p := h_p^- + h_p^+$. Furthermore, $h_1 := h_1^+$, $h_{n,1} := h_{n,1}^-$. Using this we get

$$\langle \mathbf{M}\mathbf{u}, \mathbf{u} \rangle = \int_{0}^{1} u(x)^{2} dx = \sum_{i=1}^{n} \sum_{j=1}^{l_{i}+1} \int_{m_{i,j-1}}^{m_{i,j}} u(x)^{2} dx \sim \sum_{i=1}^{n} \sum_{j=1}^{l_{i}+1} (m_{i,j} - m_{i,j-1}) \left(u(m_{i,j})^{2} + u(m_{i,j-1})^{2} \right) = \sum_{i=1}^{n} \sum_{j=1}^{l_{i}+1} h_{i,j}^{-} \left(\psi_{i,j}^{2} + \psi_{i,j-1}^{2} \right) \sim \sum_{i=1}^{n} \sum_{j=0}^{l_{i}} h_{i,j} \psi_{i,j}^{2} = \sum_{i=2}^{n} h_{i} \psi_{i}^{2} + \sum_{i \in \ell} h_{i} \psi_{i}^{2} \sim h \sum_{i=2}^{n} \psi_{i}^{2} + \sum_{i \in \ell} h_{i} \psi_{i}^{2}.$$

From this and $h_i \sim \Delta_{i_1}$ for $i = (i_1, i_2) \in \ell$ the result in (3.19) follows.

Theorem 3.4 Assume that Assumption 1 is satisfied. There exists a constant C independent of h such that

$$\frac{\langle \mathbf{M}\mathbf{u},\mathbf{u}\rangle}{\langle \mathbf{D}_M\mathbf{u},\mathbf{u}\rangle} \leq C \quad for \ all \ \mathbf{u} \in \mathbb{R}^{n_V}, \ \mathbf{u} \neq 0.$$

Proof Using (3.6) and (3.11) we obtain, for $2 \le i \le n$,

$$\psi_i^2 \le c(\tilde{u}_{i-1}^2 + \tilde{u}_i^2).$$

Deringer

Hence,

$$h\sum_{i=2}^{n}\psi_{i}^{2} \le c h\sum_{i=1}^{n}\tilde{u}_{i}^{2}.$$
(3.20)

For $i = (i_1, i_2) \in \ell$ we have, using (3.7) and (3.11),

$$\Delta_{i_1} \psi_i^2 \le c \Delta_{i_1} (\tilde{u}_{i_1}^2 + \tilde{u}_i^2) \le c (h \tilde{u}_{i_1}^2 + \Delta_{i_1} \tilde{u}_i^2).$$

This yields

$$\sum_{i\in\ell}\Delta_{i_1}\psi_i^2 \le c\left(h\sum_{i=1}^n \tilde{u}_i^2 + \sum_{i\in\ell}\Delta_{i_1}\tilde{u}_i^2\right).$$
(3.21)

Combination of (3.20), (3.21) and the result in Lemma 3.3 proves the result.

Theorem 3.5 Assume that Assumptions 1 and 2 are satisfied. There exists a constant C > 0 independent of h such that

$$\frac{\langle \mathbf{M}\mathbf{u},\mathbf{u}\rangle}{\langle \mathbf{D}_M\mathbf{u},\mathbf{u}\rangle} \geq Ch^3 |\ln h|^{-1}$$

for all $\mathbf{u} = (u_i)_{i \in \mathcal{I}} \in \mathbb{R}^{n_V}$, $\mathbf{u} \neq 0$, with $u_1 = 0$ and $u_{k+1} = 0$ if $v_k \in N(1)$ (k < n). *Proof* First consider the case $N(1) = \emptyset$. For $2 \le i \le n$ we have, using (3.6) and $u_1 = 0$:

$$|\tilde{u}_i| \le |\psi_i| + \xi_{i-1} |\tilde{u}_{i-1}| \le \sum_{j=2}^i \left(\prod_{k=j}^{i-1} \xi_k \right) |\psi_j|.$$

From this we get

$$\sum_{i=2}^{n} \tilde{u}_{i}^{2} \leq \left(\sum_{i=2}^{n} \sum_{j=2}^{i} \left(\Pi_{k=j}^{i-1} \xi_{k}\right)^{2}\right) \sum_{j=2}^{n} \psi_{j}^{2}.$$
(3.22)

Using Assumption 2 the factor $\sum_{i=2}^{n} \sum_{j=2}^{i} (\prod_{k=j}^{i-1} \xi_k)^2$ can be estimated as follows. To shorten notation we write $d_i := d(v_i)$. Using the result in (3.10) we obtain

$$\left(\Pi_{k=j}^{i-1}\xi_k\right)^2 \le \min\{d_{i-1}, d_i\}^2 \left(\frac{1}{d_{j-1}} + \frac{1}{d_j}\right)^2,$$

hence,

$$\sum_{j=2}^{i} \left(\prod_{k=j}^{i-1} \xi_k \right)^2 \le 4 \sum_{j=1}^{i} \frac{\min\{d_{i-1}, d_i\}^2}{d_j^2},$$

and

$$\sum_{i=2}^{n} \sum_{j=2}^{i} \left(\prod_{k=j}^{i-1} \xi_k \right)^2 \le 4 \sum_{i=2}^{n} \sum_{j=1}^{i} \frac{\min\{d_{i-1}, d_i\}^2}{d_j^2} \le 4 \sum_{j=1}^{n} \sum_{i=j}^{n} \frac{\min\{d_{i-1}, d_i\}^2}{d_j^2} \le 8 \sum_{j=1}^{n} \sum_{i=j, j+2, j+4, \dots}^{n} \frac{d_i^2}{d_j^2} \le 8n \sum_{j=1}^{n} \left(\frac{\max_{i=j, j+2, \dots} d_i}{d_j} \right)^2 =: 8n \sum_{j=1}^{n} \beta_j^2.$$

Define $\hat{N}(\alpha) := \{ v_j \in N \mid d(v_j) \leq h^{\alpha} \max_{i=j,j+2,\dots} d(v_i) \} = \{ v_j \in N \mid \beta_j \geq h^{-\alpha} \} \subset N(\alpha)$. Note that $\hat{N}(0) = N$, $\hat{N}(1) = \emptyset$ and $|\hat{N}(\alpha)| \leq c_1 h^{\alpha-1}$. Furthermore, for $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ we have $\#\{\beta_j \mid \beta_j \in [h^{-\alpha_1}, h^{-\alpha_2})\} = |\hat{N}(\alpha_1)| - |\hat{N}(\alpha_2)|$. Using this and Assumption 2 we obtain:

$$8n\sum_{j=1}^{n}\beta_{j}^{2} \leq -ch^{-1}\int_{0}^{1}h^{-2\alpha}\,\mathrm{d}|\hat{N}(\alpha)| \leq -c\,h^{-2}\int_{0}^{1}h^{-2\alpha}\,\mathrm{d}h^{\alpha}$$
$$\leq ch^{-2}|\ln h|\int_{0}^{1}h^{-\alpha}\,\mathrm{d}\alpha \leq ch^{-3}|\ln h|.$$

Using this bound in (3.22), in combination with $\tilde{u}_1 = 0$, $|\tilde{u}_0| = |\psi_1|$ we obtain

$$\sum_{i=0}^{n} \tilde{u}_{i}^{2} \le ch^{-3} |\ln h| \sum_{i=1}^{n} \psi_{i}^{2}.$$
(3.23)

We now consider the case $N(1) \neq \emptyset$. We take |N(1)| = 1, say $N(1) = \{v_k\}$, hence $u_{k+1} = 0$. Using the above arguments both on the triangulation starting with v_0 and ending at v_k and on the one starting at v_k and ending at v_n we obtain results as in (3.23) with $\sum_{i=0,1}^{n}$ replaced by $\sum_{i=0,1}^{k-1}$ and with $\sum_{i=0,1}^{n}$ replaced by $\sum_{i=k,k+1}^{n}$, respectively. Adding these two results we see that (3.23) holds in this case, too. The case $|N(1)| \ge 2$ can be treated by repetition of this splitting argument.

We now treat the second term in the denominator in (3.19) for the general case $|N(1)| \ge 0$. For $i = (i_1, i_2) \in \ell$ we get, using (3.7) and (3.11):

$$|\tilde{u}_i| \le c |\tilde{u}_{i_1}| + |\psi_i|,$$

hence,

$$\Delta_{i_1}\tilde{u}_i^2 \leq c\left(h\tilde{u}_{i_1}^2 + \Delta_{i_1}\psi_i^2\right),\,$$

🖉 Springer

which yields, using (3.23),

$$\sum_{i \in \ell} \Delta_{i_1} \tilde{u}_i^2 \le c \left(h \sum_{i=2}^n \tilde{u}_i^2 + \sum_{i \in \ell} \Delta_{i_1} \psi_i^2 \right) \le c h^{-3} |\ln h| \left(h \sum_{i=2}^n \psi_i^2 + \sum_{i \in \ell} \Delta_{i_1} \psi_i^2 \right).$$
(3.24)

Combination of (3.23) and (3.24) with the result in Lemma 3.3 completes the proof.

We now present a main result of this paper on the conditioning of the scaled mass matrix.

Theorem 3.6 Let $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n_V}$ be the eigenvalues of $\mathbf{D}_M^{-1}\mathbf{M}$. Assume that Assumptions 1 and 2 are satisfied. Then

$$\lambda_1 = 0, \quad \lambda_2 > 0, \quad and \quad \frac{\lambda_{n_V}}{\lambda_m} \le Ch^{-3} |\ln h|$$

$$(3.25)$$

holds with a constant C independent of h, and m = |N(1)| + 2.

Proof The matrix **M** has dimension $n_V \times n_V$. The number of intersection points m_j is $n_V - 1$ and thus $\dim(V_h^{\Gamma}) \leq n_V - 1$ holds. This implies $\dim(\operatorname{range}(\mathbf{M})) = \dim(V_h^{\Gamma}) \leq n_V - 1$, and thus $\dim(\ker(\mathbf{M})) \geq 1$, which implies $\lambda_1 = 0$. For any $\mathbf{u} \neq 0$ with $u_1 = 0$ we have $\langle \mathbf{Mu}, \mathbf{u} \rangle > 0$ and thus from the Courant-Fischer representation of eigenvalues it follows that $\lambda_2 > 0$ holds. From the Courant-Fischer representation and Theorem 3.5 we obtain, with W_m the family of m - 1-dimensional subspaces of \mathbb{R}^{n_V} ,

$$\lambda_m = \sup_{S \in W_m} \inf_{\mathbf{u} \in S^\perp} \frac{\langle \mathbf{M} \mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_M \mathbf{u}, \mathbf{u} \rangle} \geq \inf_{\substack{\mathbf{u} \in \mathbb{R}^n \vee, u_1 = 0 \\ u_{i+1} = 0 \text{ if } v_i \in N(1)}} \frac{\langle \mathbf{M} \mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_M \mathbf{u}, \mathbf{u} \rangle} \geq Ch^3 |\ln h|^{-1}.$$

In combination with the result in Theorem 3.4 this yields $\frac{\lambda_{n_V}}{\lambda_m} \le Ch^{-3} |\ln h|$. \Box

Note that due to Assumption 2 we have $|N(1)| \le c_1$ with some *h*-independent constant $c_1 = O(1)$. Thus in (3.25) we have a bound on an "effective" condition number in the sense that a fixed (independent of *h*) number of m - 1 smallest eigenvalues are not taken into account.

3.4 Analysis for the stiffness matrix

In this section, we derive bounds for the (effective) condition number of the stiffness matrix \mathbf{A} defined in (3.2).

Let $\mathbf{D}_A = \text{diag}(\mathbf{A})$ be the diagonal of the stiffness matrix.

Lemma 3.7 Assume that Assumption 1 holds. For all $\mathbf{u} = (u_i)_{i \in \mathcal{I}} \in \mathbb{R}^{n_V}$, $\mathbf{u} \neq 0$, we have

$$\frac{\langle \mathbf{A}\mathbf{u},\mathbf{u}\rangle}{\langle \mathbf{D}_{A}\mathbf{u},\mathbf{u}\rangle} \sim \frac{\sum_{i=1}^{n} \frac{1}{\Delta_{i}} \sum_{j=1}^{l_{i}+1} (\psi_{i,j} - \psi_{i,j-1})^{2}}{\sum_{i=1}^{n} \left(\frac{1}{\Delta_{i-1}} + \frac{1}{\Delta_{i}} \sum_{j=1}^{l_{i}+1} \left(\xi_{i,j} - \xi_{i,j-1}\right)^{2} + \frac{1}{\Delta_{i+1}}\right) \tilde{u}_{i}^{2} + \sum_{i \in \ell} \frac{1}{\Delta_{i_{1}}} \tilde{u}_{i}^{2}},$$

with $\psi_i = u(m_i)$, $u := \sum_{i \in \mathcal{I}} u_i \phi_i$, $\tilde{u}_i = \phi_i(m_i) u_i$.

Proof The identity $\langle \mathbf{D}_A \mathbf{u}, \mathbf{u} \rangle = \sum_{i \in \mathcal{I}} \|(\phi_i)_x\|_{\Gamma}^2 u_i^2$ follows directly from the definition of \mathbf{D}_A . Furthermore, using Lemma 3.2 we obtain, with $g_i := \sum_{j=1}^{l_i+1} (\xi_{i,j} - \xi_{i,j-1})^2$:

$$\begin{split} \sum_{i \in \mathcal{I}} \|(\phi_i)_x\|_{\Gamma}^2 u_i^2 &= \sum_{i=1}^n \|(\phi_i)_x\|_{\Gamma}^2 u_i^2 + \sum_{i \in \ell} \|(\phi_i)_x\|_{\Gamma}^2 u_i^2 \\ &\sim \sum_{i=1}^n \left(\frac{1}{\Delta_{i-1}} + \frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}}\right) \phi_i(m_i)^2 u_i^2 + \sum_{i \in \ell} \frac{1}{\Delta_{i_1}} \phi_i(m_i)^2 u_i^2 \\ &= \sum_{i=1}^n \left(\frac{1}{\Delta_{i-1}} + \frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}}\right) \tilde{u}_i^2 + \sum_{i \in \ell} \frac{1}{\Delta_{i_1}} \tilde{u}_i^2. \end{split}$$

For the nominator we have:

$$\langle \mathbf{Au}, \mathbf{u} \rangle = \int_{0}^{1} u'(x)^{2} dx = \sum_{i=1}^{n} \sum_{j=1}^{l_{i}+1} \int_{m_{i,j-1}}^{m_{i,j}} u'(x)^{2} dx$$

= $\sum_{i=1}^{n} \sum_{j=1}^{l_{i}+1} \frac{\left(u(m_{i,j}) - u(m_{i,j-1})\right)^{2}}{m_{i,j} - m_{i,j-1}} \sim \sum_{i=1}^{n} \frac{1}{\Delta_{i}} \sum_{j=1}^{l_{i}+1} (\psi_{i,j} - \psi_{i,j-1})^{2}.$

This completes the proof.

Theorem 3.8 Assume that Assumption 1 holds. There exists a constant C independent of h such that

$$\frac{\langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_A \mathbf{u}, \mathbf{u} \rangle} \le C \quad for \ all \ \mathbf{u} \in \mathbb{R}^{n_V}, \ \mathbf{u} \neq 0.$$

Proof We use Lemma 3.7. Using (3.6) and (3.7) we obtain

$$\psi_{i,1} - \psi_{i,0} = \psi_{i,1} - \psi_i = \tilde{u}_{i,1} - \xi_{i-1}\tilde{u}_{i-1} + (\xi_{i,1} - 1)\tilde{u}_i$$
$$= \tilde{u}_{i,1} - \xi_{i-1}\tilde{u}_{i-1} + (\xi_{i,1} - \xi_{i,0})\tilde{u}_i$$

and for $2 \le j \le l_i + 1$

$$\psi_{i,j} - \psi_{i,j-1} = \tilde{u}_{i,j} - \tilde{u}_{i,j-1} + (\xi_{i,j} - \xi_{i,j-1})\tilde{u}_i.$$

D Springer

Using $\xi_i \sim 1$ this yields, with $\tilde{u}_{i,0} := \tilde{u}_{i-1}$,

$$(\psi_{i,j} - \psi_{i,j-1})^2 \le c \left(\tilde{u}_{i,j}^2 + \tilde{u}_{i,j-1}^2 + (\xi_{i,j} - \xi_{i,j-1})^2 \tilde{u}_i^2 \right) \quad \text{for } 1 \le j \le l_i + 1.$$

Hence, with $g_i := \sum_{j=1}^{l_i+1} (\xi_{i,j} - \xi_{i,j-1})^2$ we obtain

$$\sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2 \le c \left(\tilde{u}_{i-1}^2 + \tilde{u}_{i+1}^2 + g_i \tilde{u}_i^2 + \sum_{j=1}^{l_i} \tilde{u}_{i,j}^2 \right)$$

and thus

$$\sum_{i=1}^{n} \frac{1}{\Delta_{i}} \sum_{j=1}^{l_{i}+1} (\psi_{i,j} - \psi_{i,j-1})^{2} \leq c \sum_{i=1}^{n} \frac{1}{\Delta_{i}} \left(\tilde{u}_{i-1}^{2} + \tilde{u}_{i+1}^{2} + g_{i} \tilde{u}_{i}^{2} \right) + c \sum_{i=(i_{1},i_{2})\in\ell} \frac{1}{\Delta_{i_{1}}} \tilde{u}_{i}^{2}$$
$$\leq c \sum_{i=1}^{n} \left(\frac{1}{\Delta_{i-1}} + \frac{g_{i}}{\Delta_{i}} + \frac{1}{\Delta_{i+1}} \right) \tilde{u}_{i}^{2} + c \sum_{i=(i_{1},i_{2})\in\ell} \frac{1}{\Delta_{i_{1}}} \tilde{u}_{i}^{2},$$

which completes the proof.

We now derive a lower bound for the smallest nonzero eigenvalue of $\mathbf{D}_A^{-1}\mathbf{A}$. For this it turns out to be more convenient to consider $u_i := u(v_i) = \phi_i(m_i)^{-1}\tilde{u}_i$ instead of \tilde{u}_i .

Lemma 3.9 For $u_i = u(v_i)$ we have the recursion

$$u_i = (1 - \alpha_i)u_{i-1} + \alpha_i u_{i-2} + \frac{1}{\phi_i(m_i)}(\psi_i - \psi_{i-1}), \quad i = 2, \dots, n,$$

with

$$\alpha_i := \frac{d(v_{i-1}) + d(v_i)}{d(v_{i-2}) + d(v_{i-1})}.$$

For $u_0 = u_1 := 0$ the solution of this recursion is given by

$$u_i = \sum_{j=1}^{i-1} \left(d(v_j) + (-1)^{i-j-1} d(v_i) \right) \frac{1}{d(v_j)} (\psi_{j+1} - \psi_j), \quad i = 2, \dots, n.$$
(3.26)

Proof From (3.3) we get

$$\psi_i = \phi_{i-1}(m_i)u_{i-1} + \phi_i(m_i)u_i$$

$$\psi_{i-1} = \phi_{i-2}(m_{i-1})u_{i-2} + \phi_{i-1}(m_{i-1})u_{i-1}.$$

and thus, using $\phi_{j-1}(m_j) = 1 - \phi_j(m_j)$, we have

$$u_{i} = \left(1 + \frac{\phi_{i-1}(m_{i-1}) - 1}{\phi_{i}(m_{i})}\right)u_{i-1} + \frac{1 - \phi_{i-1}(m_{i-1})}{\phi_{i}(m_{i})}u_{i-2} + \frac{1}{\phi_{i}(m_{i})}(\psi_{i} - \psi_{i-1})$$
$$= (1 - \alpha_{i})u_{i-1} + \alpha_{i}u_{i-2} + \frac{1}{\phi_{i}(m_{i})}(\psi_{i} - \psi_{i-1})$$

with $\alpha_i := \frac{1-\phi_{i-1}(m_{i-1})}{\phi_i(m_i)}$. Using the formula in (3.12) we get

$$\alpha_i = \frac{d(v_{i-1}) + d(v_i)}{d(v_{i-2}) + d(v_{i-1})}.$$

The representation

$$u_{i} = \sum_{k=2}^{i} \sum_{j=1}^{k-1} (-1)^{k+1-j} \left(d(v_{k-1}) + d(v_{k}) \right) \frac{1}{d(v_{j})} (\psi_{j+1} - \psi_{j})$$
(3.27)

can be shown by induction as follows. For i = 2 we get (using (3.12)),

$$u_2 = (d(v_1) + d(v_2)) \frac{1}{d(v_1)} (\psi_2 - \psi_1) = \frac{1}{\phi_2(m_2)} (\psi_2 - \psi_1),$$

which also follows from the recursion formula if we take $u_0 = u_1 = 0$. Assume that the representation formula (3.27) is correct for indices less than or equal to i - 1. We then obtain

$$(1 - \alpha_i)u_{i-1} + \alpha_i u_{i-2} + \frac{1}{\phi_i(m_i)}(\psi_i - \psi_{i-1})$$

$$= -\alpha_i(u_{i-1} - u_{i-2}) + u_{i-1} + \frac{1}{\phi_i(m_i)}(\psi_i - \psi_{i-1})$$

$$= -\alpha_i \sum_{j=1}^{i-2} (-1)^{i-j} (d(v_{i-2}) + d(v_{i-1})) \frac{1}{d(v_j)}(\psi_{j+1} - \psi_j)$$

$$+ \sum_{k=2}^{i-1} \sum_{j=1}^{k-1} (-1)^{k+1-j} (d(v_{k-1}) + d(v_k)) \frac{1}{d(v_j)}(\psi_{j+1} - \psi_j)$$

$$+ \frac{d(v_{i-1}) + d(v_i)}{d(v_{i-1})}(\psi_i - \psi_{i-1})$$

$$= \sum_{j=1}^{i-1} (-1)^{i+1-j} (d(v_{i-1}) + d(v_i)) \frac{1}{d(v_j)}(\psi_{j+1} - \psi_j)$$

$$+ \sum_{k=2}^{i-1} \sum_{j=1}^{k-1} (-1)^{k+1-j} (d(v_{k-1}) + d(v_k)) \frac{1}{d(v_j)}(\psi_{j+1} - \psi_j)$$

$$=\sum_{k=2}^{i}\sum_{j=1}^{k-1}(-1)^{k+1-j}\left(d(v_{k-1})+d(v_{k})\right)\frac{1}{d(v_{j})}(\psi_{j+1}-\psi_{j})$$

and thus the representation for u_i in (3.27). From this we obtain, by changing the order of summation:

$$u_i = \sum_{j=1}^{i-1} \left(\sum_{k=j+1}^{i} (-1)^{k+1-j} \left(d(v_{k-1}) + d(v_k) \right) \right) \frac{1}{d(v_j)} (\psi_{j+1} - \psi_j).$$

The representation in (3.26) immediately follows from this one.

Theorem 3.10 Assume the Assumptions 1 and 2 hold. There exists a constant C > 0 independent of h such that

$$\frac{\langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_{A}\mathbf{u}, \mathbf{u} \rangle} \ge Ch^{2} |\ln h|^{-1}$$

for all $\mathbf{u} = (u_i)_{i \in \mathcal{I}} \in \mathbb{R}^{n_V}$, $\mathbf{u} \neq 0$, with $u_0 = u_1 = 0$ and $u_k = u_{k+1} = 0$ if $v_k \in N(1)$ (k < n).

Proof We continue to use the notation $d_i := d(v_i), u_i := u(v_i)$. First assume $N(1) = \emptyset$. We use the representation in Lemma 3.7 and first consider the term $\sum_{i=1}^{n} (\frac{1}{\Delta_{i-1}} + \frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}})\tilde{u}_i^2$ in the denominator. The Assumption 1 and the regularity of the outer triangulation imply that the angles between Γ and all sides of the triangles intersecting Γ are uniformly (w.r.t. *h*) bounded away from zero. Hence we have $d_i \sim \Delta_i$ $(1 \le i \le n)$, and $\tilde{u}_i = \phi_i(m_i)u_i \sim \frac{d_{i-1}}{h}u_i$ $(1 \le i \le n)$. Using this and the result in (3.26) we get

$$\tilde{u}_i^2 \le c \frac{d_{i-1}^2}{h^2} u_i^2 \le c \frac{d_{i-1}^2}{h^2} \left(\sum_{j=1}^{i-1} (d_j^2 + d_i^2) \frac{1}{d_j} \right) \sum_{j=1}^n \frac{1}{d_j} (\psi_{j+1} - \psi_j)^2.$$

For the last term we have

$$\sum_{i=1}^{n} \frac{1}{d_i} (\psi_{i+1} - \psi_i)^2 \le c \sum_{i=1}^{n} \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2,$$

Condition (3.8) and $d_j \sim \Delta_j$ yield $d_{i-1} \sim \Delta_{i-1} \sim \Delta_{i+1}$. Using this and $u_1 = 0$, we get

$$\sum_{i=1}^{n} \left(\frac{1}{\Delta_{i-1}} + \frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}} \right) \tilde{u}_i^2 \le c \sum_{i=1}^{n} \left(\frac{1}{d_{i-1}} + \frac{g_i}{d_i} \right) \tilde{u}_i^2$$
$$\le \left[\frac{c}{h^2} \sum_{i=2}^{n} \left(d_{i-1} + \frac{g_i d_{i-1}^2}{d_i} \right) \sum_{j=1}^{i-1} \left(d_j + \frac{d_i^2}{d_j} \right) \right] \sum_{i=1}^{n} \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2.$$

🖉 Springer

We estimate the factor in the square brackets as follows. Using $d_{i-1}d_i \leq c \min \{d_{i-1}, d_i\}h$ we get:

$$\frac{c}{h^2} \sum_{i=2}^n \left(d_{i-1} + \frac{g_i d_{i-1}^2}{d_i} \right) \sum_{j=1}^{i-1} \left(d_j + \frac{d_i^2}{d_j} \right)$$

$$\leq \frac{c}{h^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \frac{d_{i-1} d_i^2 + d_{i-1}^2 d_i}{d_j} + \frac{c}{h^2} \sum_{i=2}^n \sum_{j=1}^{i-1} d_{i-1} d_j + \frac{c}{h^2} \sum_{i=2}^n \frac{g_i d_{i-1}^2}{d_i} \sum_{j=1}^{i-1} d_j$$

$$\leq c \sum_{i=2}^n \sum_{j=1}^{i-1} \frac{\min\{d_{i-1}, d_i\}}{d_j} + ch^{-2} + \frac{c}{h} \sum_{i=2}^n \frac{hg_i}{d_i}.$$

The first term on the righthand side can be bounded by $ch^{-2} |\ln h|$ using the same arguments as in the proof of Theorem 3.5. The third term can be treated in a similar way as follows. Define $\beta_j := \frac{hg_j}{d_j}$, $\tilde{N}(\alpha) := \{v_j \in N \mid d(v_j) \leq h^{\alpha+1}g_j\} = \{v_j \in N \mid \beta_j \geq h^{-\alpha}\} \subset N(\alpha)$. Note that $\tilde{N}(0) = N$, $\tilde{N}(1) = \emptyset$ and $|\tilde{N}(\alpha)| \leq c_1 h^{\alpha-1}$. For $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ we have $\#\{\beta_j \mid \beta_j \in (h^{-\alpha_1}, h^{-\alpha_2}]\} = |\tilde{N}(\alpha_1)| - |\tilde{N}(\alpha_2)|$. Using this and Assumption 2 we obtain:

$$\frac{c}{h}\sum_{i=2}^{n}\beta_{i} \leq -ch^{-1}\int_{0}^{1}h^{-\alpha}d|\tilde{N}(\alpha)| \leq -ch^{-2}\int_{0}^{1}h^{-\alpha}dh^{\alpha} \leq ch^{-2}|\ln h|.$$

Collecting these results (and using $\tilde{u}_0 = \tilde{u}_1 = 0$) we get

$$\sum_{i=0}^{n} \left(\frac{1}{\Delta_{i-1}} + \frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}} \right) \tilde{u}_i^2 \le ch^{-2} |\ln h| \sum_{i=1}^{n} \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2.$$
(3.28)

We now consider the case $N(1) \neq \emptyset$. We take |N(1)| = 1, say $N(1) = \{v_k\}$, hence $u_k = u_{k+1} = 0$. Using the above arguments both on the triangulation starting with v_0 and ending at v_k and on the one starting at v_k and ending at v_n we obtain results as in (3.28) with $\sum_{i=0,1}^{n}$ replaced by $\sum_{i=0,1}^{k-1}$ and with $\sum_{i=0,1}^{n}$ replaced by $\sum_{i=k,k+1}^{n}$, respectively. Adding these two results we see that (3.28) holds in this case, too. The case $|N(1)| \ge 2$ can be treated by repetition of this splitting argument.

We now treat the term $\sum_{i \in \ell} \frac{1}{\Delta_{i_1}} \tilde{u}_i^2$ in the denominator in Lemma 3.7 for the general case $|N(1)| \ge 0$. Note that

$$\sum_{i \in \ell} \frac{1}{\Delta_{i_1}} \tilde{u}_i^2 = \sum_{1 \le i \le n, \ l_i > 0} \sum_{j=1}^{l_i} \frac{1}{\Delta_i} \tilde{u}_{i,j}^2.$$

Springer

Using (3.7) we get, for an *i* with $l_i \ge 2$:

$$\tilde{u}_{i,j} - \tilde{u}_{i,j-1} = \psi_{i,j} - \psi_{i,j-1} - (\xi_{i,j} - \xi_{i,j-1})\tilde{u}_i$$

and with (3.6) and $\psi_{i,0} := \psi_i, \xi_{i,0} := 1$:

$$\tilde{u}_{i,1} - \xi_{i-1}\tilde{u}_{i-1} = \psi_{i,1} - \psi_{i,0} - (\xi_{i,1} - \xi_{i,0})\tilde{u}_i.$$

This yields, for $1 \le j \le l_i$:

$$\begin{split} \tilde{u}_{i,j}^2 &\leq c \left(\tilde{u}_{i-1}^2 + \sum_{j=1}^{l_i} (\psi_{i,j} - \psi_{i,j-1})^2 + \sum_{j=1}^{l_i} (\xi_{i,j} - \xi_{i,j-1})^2 \tilde{u}_i^2 \right) \\ &\leq c \left(\tilde{u}_{i-1}^2 + \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2 + g_i \tilde{u}_i^2 \right). \end{split}$$

Thus we get

$$\sum_{1 \le i \le n, \ l_i > 0} \sum_{j=1}^{l_i} \frac{1}{\Delta_i} \tilde{u}_{i,j}^2 \le c \sum_{i=1}^n \frac{1}{\Delta_i} \tilde{u}_{i-1}^2 + c \sum_{i=1}^n \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2 + c \sum_{i=1}^n \frac{g_i}{\Delta_i} \tilde{u}_i^2$$
$$\le c \sum_{i=1}^n \left(\frac{g_i}{\Delta_i} + \frac{1}{\Delta_{i+1}} \right) \tilde{u}_i^2 + c \sum_{i=1}^n \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2$$

Using the bound in (3.28) we obtain

$$\sum_{1 \le i \le n, \ l_i > 0} \sum_{j=1}^{l_i} \frac{1}{\Delta_i} \tilde{u}_{i,j}^2 \le ch^{-2} |\ln h| \sum_{i=1}^n \frac{1}{\Delta_i} \sum_{j=1}^{l_i+1} (\psi_{i,j} - \psi_{i,j-1})^2,$$

and combination of this with the result in (3.28) completes the proof.

Theorem 3.11 Let $0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n_V}$ be the eigenvalues of $\mathbf{D}_A^{-1}\mathbf{A}$. Assume that Assumptions 1 and 2 are satisfied. Then

$$\lambda_1 = 0, \quad \lambda_3 > 0, \quad and \quad \frac{\lambda_{n_V}}{\lambda_m} \le Ch^{-2} |\ln h|$$

holds, with a constant C independent of h and m = 2|N(1)| + 3.

Proof A dimension argument as in the proof of Theorem 3.6 yields $\lambda_1 = 0$. From the Courant-Fischer representation and Theorem 3.10 we obtain, with W_m the family of (m-1)-dimensional subspaces of \mathbb{R}^{n_V} ,

$$\lambda_m = \sup_{S \in W_m} \inf_{\mathbf{u} \in S^{\perp}} \frac{\langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_A \mathbf{u}, \mathbf{u} \rangle} \ge \inf_{\substack{\mathbf{u} \in \mathbb{R}^{n_V, u_0} = u_1 = 0\\ u_{i+1} = u_i = 0 \text{ if } v_i \in N(1)}} \frac{\langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle}{\langle \mathbf{D}_A \mathbf{u}, \mathbf{u} \rangle} \ge Ch^2 |\ln h|^{-1}.$$

In combination with the result in Theorem 3.8 this yields $\frac{\lambda_{n_V}}{\lambda_m} \le Ch^{-2} |\ln h|$. \Box

3.5 Further discussions and extensions

3.5.1 Discussion of Assumptions 1 and 2

Assumption 1 poses a restriction on how the surface Γ divides any triangle $T \in T_h$. Since T_h satisfies a minimal angle condition one easily finds that Assumption 1 implies that the angles between Γ and all sides of the triangles that intersect Γ are uniformly (w.r.t. *h*) bounded away from zero. This condition may be satisfied in certain structured cases, as in Sect. 2.2.2, in a general, however, there appears to be no reason why it should hold. We introduced Assumption 1 to make our analysis work and currently we do not see how to avoid it. Assumption 1, however is weak enough to allow a subdivision of Γ which is *not* quasi-uniform, cf. the experiment with varying δ in Sect. 2.2.2. In our applications (where Γ is an approximation of the zero level of a level set function, cf. Sect. 2.2) it is not very realistic to assume quasi-uniformity of the induced triangulation of Γ , cf. Fig. 1.

Assumption 2 gives a condition on the distribution of nodes near the surface Γ in terms of their distances to Γ . In general, the condition $|N(\alpha)| \leq c_1 h^{\alpha-1}$ means that the set of nodes having a certain (maximal) distance to Γ (as specified in (3.9)) becomes smaller if this distance gets smaller. However, in the structured 2D experiment in Sect. 2.2.2, we can have many nodes (namely $\sim \frac{1}{2}n$) with arbitrarily small distances to Γ . In that experiment, however, we have $d(v_j) = \max_{i=j, j+2,...} d(v_i)$ and $g_j = 0$ for all *j* (the triangulation is "parallel" to Γ). Thus we have N(0) = N, $N(\alpha) = \emptyset$ for all $\alpha \in (0, 1]$ and Assumption 2 is fulfilled. In more practical unstructured cases it looks natural to use instead of Assumption 2 the following *stronger*, but simpler assumption:

Define, for $\alpha \in [0, 1]$:

$$N(\alpha) := \{ v_j \in N \mid d(v_j) \le h^{1+\alpha} \},$$
(3.29)

and assume that there is an *h*-independent constant c_1 such that $|N(\alpha)| \le c_1 h^{\alpha-1}$ for all $\alpha \in [0, 1]$. To validate the plausibility of this assumption we computed $N(\alpha)$ as in (3.29) for the ellipse and pedal curves, see Sect. 2.2.3. Figure 9 shows $|N(\alpha)|$ vs. α for different refinement levels *l*. The plot in the logarithmic scale for $|N(\alpha)|$ shows that the assumption on $h^{\alpha-1}$ asymptotic seems very plausible. Furthermore, Table 6 shows the value of |N(1)| for different refinement levels (all values are multiples of 4 due to the symmetries of both curves). We recall that |N(1)| appears in the statement of eigenvalues low bounds in Theorems 3.6 and 3.11 and it is related to the presence of few outliers in the spectrum of mass and stiffness matrices. Finally, we note that for the structured 2D example from Sect. 2.2.2 a $O(h^{-2}|\ln h|)$ condition number bound



Fig. 9 Distribution of node distances to the curve Γ for ellipse and pedal curve

Table 6 $ N(1) $ for different refinement levels; $h_l = \sqrt{2} 2^{-l}$	l	4	5	6	7	8	4	5	6	7	8
		Ellij	ose				Peda	al curv	ve .		
	m_l	70	138	278	554	1110	82	170	350	690	1378
	N(1)	0	8	16	12	4	0	12	8	4	16

for the mass matrix can be proved using a stronger condition than the one formulated in Assumption 2, cf. [8].

3.5.2 On analysis for the 3D case

The numerical results from Sect. 2.2.1, Fig. 2, Tables 2 and 3, strongly suggest that in the (unstructured) 3D case for both the mass and stiffness matrices we have effective condition number that behave like $O(h^{-2})$ and furthermore that no outliers occur in the spectrum. This in contrast to the 2D case where for the effective condition number of the mass matrix in general only an $O(h^{-3})$ (up to a possible log-term) estimate holds and outliers do occur in the spectrum. A generalization of our analysis to the 3D case requires a lot of further technical manipulations and will be presented elsewhere. Here we give a brief explanation, why the 3D case may be more advantageous regarding the above-mentioned spectral properties.

The proof of the lower bound for both mass and stiffness matrices is based on bounds for the scaled value of a finite element function in the *i*th outer node $(|\tilde{u}_i|)$ in terms of the values of the same function (or differences of values) in a sequence of surface nodes, cf. (3.3) and (3.4). All sequences start from one basis outer node v_1 (two nodes v_0 , v_1 for (3.4)), where the function vanishes. We were able to prove bounds for the coefficients of these sequences, using Assumption 1 and 2 on the outer nodes distances to the surface. In the 2D case *all* nodes v_j , $1 \le j \le i$, cf. fig. 8, between the basis node v_1 and the node v_i , where the function is estimated, are involved in the sequences. In the 3D case, however, thanks to the additional space dimension one can consider many different sequences of outer nodes 'connecting' a basis node with a given node v_i , where the function is estimated. Therefore it may be possible that a sequence is found that contains no "bad" nodes, e.g. nodes having very small distance to Γ . This would lead to better estimates for the small eigenvalues of the mass and stiffness matrices.

3.5.3 On extension to smooth surfaces

As discussed above the key points of our analysis are the estimates (3.3) and (3.4) together with suitable bounds for the coefficients occurring in these bounds. Similar bounds hold if Γ is a smooth curve. Estimating the coefficients, however, is then even more technical (although do-able), and uses smoothness assumptions on the surface, i.e. that Γ is locally an $O(h^2)$ perturbation of a line. To avoid these further technical complications we decided to restrict the analysis in this paper to the case $\Gamma = [0, 1]$.

Acknowledgments We thank the referees for their valuable comments which led to substantial improvements of the original version of this paper.

References

- 1. Braess, D.: Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics, 3rd edn. Cambridge University Press, London (2007)
- Deckelnick, K., Dziuk, G., Elliott, C.M., Heine, C.-J.: An *h*-narrow band finite element method for elliptic equations on implicit surfaces. IMA J. Numer. Anal., doi:10.1093/imanum/drn049 (2009)
- Demlow, A., Dziuk, G.: An adaptive finite element method for the Laplace–Beltrami operator on implicitly defined surfaces. SIAM J. Numer. Anal. 45, 421–442 (2007)
- DROPS package. http://www.igpm.rwth-aachen.de/DROPS/
- Dziuk, G.: Finite elements for the beltrami operator on arbitrary surfaces. In: Hildebrandt, S., Leis, R. (eds.) Partial Differential Equations and Calculus of Variations. Lecture Notes in Mathematics, vol. 1357, pp. 142–155. Springer, Heidelberg (1988)
- James, A., Lowengrub, J.: A surfactant-conserving volume-of-fluid method for interfacial flows with insoluble surfactant. J. Comp. Phys. 201, 685–722 (2004)
- Muradoglu, M., Tryggvason, G.: A front-tracking method for computation of interfacial flows with soluble surfactant. J. Comput. Phys. 227, 2238–2262 (2008)
- 8. Olshanskii, M.A., Reusken, A.: A finite element method for surface PDEs: Matrix properties, Preprint 287, IGPM, RWTH Aachen, earlier version of this paper (2008)
- 9. Olshanskii, M.A., Reusken, A., Grande, J.: An eulerian finite element method for elliptic equations on moving surfaces. Accepted to SIAM J. Numer. Anal. (2009)