



HAL
open science

On the condition number of the finite element method for the Laplace-Beltrami operator

Marcial Nguemfouo, Michael Ndjinga

► **To cite this version:**

Marcial Nguemfouo, Michael Ndjinga. On the condition number of the finite element method for the Laplace-Beltrami operator. *Journal of Elliptic and Parabolic Equations*, 2023, pp.1-28. 10.1007/s41808-023-00251-7. cea-04431293

HAL Id: cea-04431293

<https://cea.hal.science/cea-04431293v1>

Submitted on 1 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the condition number of the finite element method for the Laplace-Beltrami operator

Michael Ndjinga¹, Marcial Nguemfou²

¹University of Paris-Saclay, CEA Saclay, ISAS, DM2S, STMF,
91191 Gif-sur-Yvette, France,
michael.ndjinga@cea.fr

²University of Yaounde I, Department of Mathematics,
P.O. Box 812 Yaoundé, Cameroon,
marcial.nguemfou@facsciences-uy1.cm

Abstract

We give an upper bound for the condition number of the finite element operator for the Laplace-Beltrami operator on closed surfaces immersed in \mathbb{R}^3 . The expression is similar to the condition number of the Laplace operator in the Euclidean case, with the curvature affecting the condition number through the Poincaré constant. However in the case of closed surfaces the finite element matrix is singular and the linear system is solved for a unique solution with zero mean.

Keywords– Condition number, Laplace-Beltrami operator, Finite element method, Elliptic equation, closed surfaces

1 Introduction

The discretisation of surface partial differential equations using finite element methods is motivated by important applications related to physical and biological phenomena [1]. It is also used to answer theoretical questions in geometry [2] and to model complex systems on computers [3]. As is the case in the Euclidean context, the use of the finite element method on curved surfaces requires in practice the resolution of potentially large linear systems [4]. The precision and convergence of this resolution depends highly on the condition number of the finite element matrix [5].

The condition number of the finite element matrix has been extensively studied in the Euclidean case [6, 7]. However few works exist in the case of the Laplace-Beltrami operator on curved surfaces, particularly when they are closed [8].

Moreover, classical preconditioners are based on Incomplete LU factorisations (see section 10.3.1 in [5]), which fails if the matrix is not an M-matrix (see theorem 10.2 in [5]), in particular if the matrix is singular (see definition 1.4 in [5]). Yet, the discretisation of PDEs on closed surfaces yields singular matrices hence, the resolution of the associated linear systems are restricted to meshes with a small number of elements. There is therefore an important need for *ad hoc* preconditioners for the numerical simulation of PDEs on closed surfaces. A first step in this project is to give an estimate of the condition number of the finite element matrix.

We are interested in the finite element approximation of the Poisson problem

$$-\Delta_{\Gamma} u = f \text{ on } \Gamma, \quad (1)$$

where $\Gamma \subset \mathbb{R}^3$ is a closed C^2 manifold and the weak solution u is sought for in $H^1(\Gamma)$. Γ being a closed manifold, u and f must have zero mean for the problem to admit a unique solution ($\int_{\Gamma} u = \int_{\Gamma} f = 0$).

[4] adapted the classical Euclidean finite element approach in order to deal with curved surfaces. Existence theorems and asymptotic error estimates are proven in [4, 9]. An important feature of the method is the avoidance of charts both in the problem formulation and the numerical method. The surface finite element method is based simply on triangulated surfaces and requires the geometry solely through knowledge of the vertices and normal vector of each triangle.

Following [4], we first approximate the domain Γ by the surface of a polyhedron Γ_h with triangular faces. The finite elements (i.e. the 3D triangles) are planar as in the Euclidean case, but they are no longer share the same normal vector. The right hand side f is approximated by a function $f_h \in L^2(\Gamma_h)$ with zero mean.

The solution u of (1) on Γ is then approximated by the solution u_h of the following Poisson problem on Γ_h :

$$-\Delta_{\Gamma_h} u_h = f_h \text{ on } \Gamma_h, \quad (2)$$

where $\Gamma_h \subset \mathbb{R}^3$ is a closed piecewise triangular surface and the weak solution u_h is sought for in $H^1(\Gamma_h)$. Γ_h being a closed manifold, u_h and f_h must have zero mean for (2) to admit a unique solution ($\int_{\Gamma_h} u_h = \int_{\Gamma_h} f_h = 0$).

The finite element method proposed in [4] then consists in approximating $u_h \in H^1(\Gamma_h)$ by its projection \tilde{u}_h on $PL(\Gamma_h) \subset H^1(\Gamma_h)$, the subspace of continuous piecewise linear functions. Doing so, the partial differential operator Δ_{Γ_h} is approximated by a finite dimension linear operator with matrix $A_{\Delta_{\Gamma_h}}$. This approximation uses the variational formulation of (2) and reduces the Poisson problem (1) to the resolution of a linear system

$$A_{\Delta_{\Gamma_h}} X = b, \quad (3)$$

where X and b are vectors with zero mean.

The linear systems obtained using this technique are sparse and can be very large. The most practical way to solve such very large linear systems is to resort to an iterative method. Since the convergence rate of such methods is strongly affected by the condition number $\mathcal{K}_h(A_{\Delta_{\Gamma_h}})$ of the finite element matrix $A_{\Delta_{\Gamma_h}}$, studying the condition number $\mathcal{K}_h(A_{\Delta_{\Gamma_h}})$ has both a theoretical and practical importance in the study of the finite element method.

One technical difficulty is that in the case of closed surfaces, the finite element matrix is not invertible on \mathbb{R}^d since the domain has no boundary. The finite element operator is however invertible on the space of zero mean vectors. The other technical difficulty is the handling of different normal vectors arising from the non zero curvature. We are still able to adapt the Euclidean approach following [6] to derive an upper bound of the condition number.

The article is organised as follows. In section 2 we recall the definition and properties of the Laplace-Beltrami operator. In section 3 we recall the existence and uniqueness of solutions to the Poisson problem for the Laplace-Beltrami operator on closed surfaces. Section 4 is devoted to the finite element discretisation. Section 5 is concerned with the upper bound of the condition number of the finite element operator and Section 6 is devoted to numerical simulation of the Poisson problem on a sphere in order to illustrate the numerical method presented in this paper, and motivate the search for efficient preconditioners.

2 Calculus on hypersurfaces

In standard calculus, the gradient, divergence and Laplace operators are classically defined on the Euclidean space \mathbb{R}^d . They can however also be defined on a smooth manifold Γ in an intrinsic way (no immersion into \mathbb{R}^d) using a Riemannian metric (see for example [11], section 1.2) or the Hodge operator (see [2] section 11.2). We chose instead for simplicity and pedagogy to define the differential operators on embedded manifolds $\Gamma \subset \mathbb{R}^3$ following [4] (Definition 2.1). This approach is also easier to handle from a practical point of view in numerical methods. For instance Lemma 2.1 is used in the determination of the finite element coefficient from each triangle (A).

The definition of differential operators on hypersurfaces can be performed without local charts, using rather the so-called Fermi coordinates [9]. The Fermi coordinates are also useful in the proof of the Poincaré inequality (see theorems 2.8 and 2.12 in [9]).

We define embedded surfaces in subsection 2.1, the tangential gradient $\vec{\nabla}_\Gamma$ in subsection 2.2, the tangential divergence $\vec{\nabla}_\Gamma \cdot$ in subsection 2.3, and then the Laplace-Beltrami operator Δ_Γ as the composition of divergence and gradient in subsection 2.4.

2.1 Embedded surfaces and Fermi coordinates

Following [9], we start with the definition of embedded surfaces and their normals.

Definition 2.1 (Embedded C^k -hypersurface)

$\Gamma \subset \mathbb{R}^3$ is called a C^k -hypersurface if for each point $x_0 \in \Gamma$, there exists an open set $U_{x_0} \subset \mathbb{R}^3$ containing x_0 and a function $\phi_{x_0} \in C^k(U_{x_0})$ with the following properties

$$\begin{aligned} \vec{\nabla} \phi_{x_0} &\neq 0 \text{ on } \Gamma \cap U_{x_0} \\ \Gamma \cap U_{x_0} &= \{x \in U_{x_0} \mid \phi_{x_0}(x) = 0\}. \end{aligned} \quad (4)$$

For any C^1 -hypersurface $\Gamma \subset \mathbb{R}^3$, one can define the **unit normal** at any point $x \in U_{x_0}$ as

$$\vec{n}(x) = \frac{\vec{\nabla} \phi_{x_0}(x)}{\|\vec{\nabla} \phi_{x_0}(x)\|}, \quad (5)$$

where $\vec{\nabla}$ denotes the classical \mathbb{R}^3 gradient.

We define the **δ -strip around Γ** as

$$U_{\delta,\Gamma} = \{x \in \mathbb{R}^3, \text{dist}(x, \Gamma) < \delta\}. \quad (6)$$

For δ small enough it is possible to define a projection operator $a_\Gamma : U_\delta \rightarrow \Gamma$ onto Γ and a signed distance function $d_\Gamma : U_\delta \rightarrow \mathbb{R}$. $a_\Gamma(x)$ and $d_\Gamma(x)$ are called the Fermi coordinates of x and their existence is given by the following theorem.

Theorem 2.1 (Fermi coordinates)

Let Γ be an embedded C^2 hypersurface. There exists $\delta_{\text{Fermi}} > 0$ such that for every point $x \in U_{\delta_{\text{Fermi}}}$, there exists a unique point $a_\Gamma(x) \in \Gamma$ such that

$$\forall x \in U_{\delta_{\text{Fermi}}}, \quad x = a_\Gamma(x) + d_\Gamma(x)\vec{n}(x). \quad (7)$$

where $d_\Gamma \in C^2(U_{\delta_{\text{Fermi}}})$ is the signed distance function.

Proof: see Lemma 2.8 in [9].

□

In the following, the Fermi-coordinates will allow us to extend a function defined on Γ to a neighbourhood of Γ , which will prove useful in the definitions of the tangential gradient and the tangential divergence (sections 2.2 and 2.3).

2.2 The tangential gradient

The notions of hypersurface (Definition 2.1) and associated normal vectors (5) defined in the previous section are useful in the following definition of the projected gradient.

Definition 2.2 (Projected \mathbb{R}^3 gradient)

Let Γ be an embedded C^1 hypersurface, U_Γ a neighbourhood of Γ in \mathbb{R}^3 , and $\bar{u} \in C^1(U_\Gamma)$. The projected gradient of \bar{u} on Γ is

$$\vec{\nabla}_{P\Gamma}\bar{u} = \vec{\nabla}\bar{u} - (\vec{n} \cdot \vec{\nabla}\bar{u})\vec{n}. \quad (8)$$

From the projected gradient of a function $\bar{u} \in C^1(U_\Gamma)$ on a C^1 hypersurface Γ , one can define the tangential gradient of a function $u \in C^1(\Gamma)$ on a C^2 hypersurface Γ as follows.

Definition 2.3 (Tangential gradient)

Let Γ be an embedded C^2 hypersurface, $u \in C^1(\Gamma)$. Let $\bar{u} \in C^1(U_{\delta_{Fermi},\Gamma})$ such that $\bar{u}|_\Gamma = u$. The tangential gradient of u is

$$\vec{\nabla}_\Gamma u = \vec{\nabla}_{P\Gamma}\bar{u}. \quad (9)$$

The assumption that Γ be C^2 in definition 2.5 comes from the fact that the existence of an extension \bar{u} of u on a neighbourhood of Γ requires Fermi coordinates since $\bar{u}(x) = u(a_\Gamma(x))$.

The value of the tangential gradient of $u \in C^1(U_{\delta_{Fermi},\Gamma})$ on an embedded hypersurface Γ does not depend on the choice of \bar{u} . It depends only on the values taken by u on Γ as expressed in the following lemma (see [9] Lemma 2.4).

One key property of the tangential gradient is stated in Lemma 2.1. It will be useful in section A to calculate the coefficients of the finite element matrix.

Lemma 2.1 *The tangential gradient belongs to the tangent plane and is orthogonal to the normal vector*

$$(\vec{\nabla}_\Gamma u) \cdot \vec{n} = 0.$$

As for the classical gradient, there is a Poincaré's inequality involving the tangential gradient.

Theorem 2.2 (Poincaré's inequality)

Assume that Γ is an embedded C^3 hypersurface. There exists a constant c such that, for every function $f \in H^1(\Gamma)$ with $\int_\Gamma f = 0$, we have the inequality

$$\|f\|_{L^2(\Gamma)} \leq c \|\nabla_\Gamma f\|_{L^2(\Gamma)}. \quad (10)$$

Proof: see theorem 2.12 in [9].

□

2.3 The tangential divergence

The tangential divergence operator on Γ is defined as the contribution to the full divergence arising from the tangent space to Γ . We first define the projected divergence on Γ as follows.

Definition 2.4 (Projected \mathbb{R}^3 divergence)

Let Γ be an embedded C^1 hypersurface, U_Γ a neighbourhood of Γ in \mathbb{R}^3 , and $\bar{\mathbf{V}} \in C^1(U_\Gamma)^3$ a vector field. The projected divergence of $\bar{\mathbf{V}}$ is

$$\text{div}_{P\Gamma}\bar{\mathbf{V}} = \vec{\nabla} \cdot \bar{\mathbf{V}} - {}^t\vec{n}(\vec{\nabla} \cdot \bar{\mathbf{V}})\vec{n}$$

From the projected divergence of a function $\bar{u} \in C^1(U_\Gamma)$, one can define the tangential divergence of a function $u \in C^1(\Gamma)$ as follows.

Definition 2.5 (Tangential divergence)

Let Γ be an embedded C^2 hypersurface, $\mathbf{V} \in C^1(\Gamma)^3$ a vector field. Let $\bar{\mathbf{V}} \in C^1(U_{\delta_{\text{Fermi}}\Gamma})^3$ such that $\bar{\mathbf{V}}|_\Gamma = \mathbf{V}$. The tangential divergence of \mathbf{V} is

$$\operatorname{div}_\Gamma \mathbf{V} = \operatorname{div}_{P_\Gamma} \bar{\mathbf{V}}. \quad (11)$$

The assumption that Γ be C^2 in Definition 2.5 comes from the fact the existence of an extension $\bar{\mathbf{V}}$ of \mathbf{V} on a neighbourhood of Γ requires Fermi coordinates since $\bar{\mathbf{V}}(x) = \bar{\mathbf{V}}(a_\Gamma(x))$.

2.4 The Laplace-Beltrami operator

Now that we have defined the tangential gradient and divergence in sections 2.2 and 2.3, the Laplace-Beltrami-operator on Γ , can be defined as the composition of the tangential divergence and gradient of a function $u \in C^2(\Gamma)$.

Definition 2.6 (Laplace-Beltrami operator)

Let Γ be an embedded C^2 hypersurface. The Laplace-Beltrami operator applied to $u \in C^2(\Gamma)$ is

$$\Delta_\Gamma u = \operatorname{div}_\Gamma \vec{\nabla}_\Gamma u \quad (12)$$

For any smooth embedded hypersurface Γ , the following Green's formula is a consequence of the Stokes' theorem (see theorem 6.25 in [10]):

$$\forall v \in C^1(\Gamma), u \in C^2(\Gamma), \quad \int_\Gamma v \Delta_\Gamma u = - \int_\Gamma \vec{\nabla}_\Gamma u \cdot \vec{\nabla}_\Gamma v \quad (13)$$

A consequence of the Green's formula (13) is that the operator $-\Delta_\Gamma$ is symmetric and positive.

3 The Poisson problem on a closed surface

Now that we have defined the Laplace-Beltrami operator in section 2.4, we can study the Poisson problem on closed hypersurfaces. We start by setting the problem and the relevant functional spaces in section 3.1. We then give the weak formulation of the problem in section 3.2. We end up by summarising the existence result in section 3.3.

3.1 Definition and functional spaces

Let Γ be a closed C^2 hypersurface in \mathbb{R}^3 . Since Γ is closed ($\partial\Gamma = \emptyset$), all the constant functions are in the kernel of Δ_Γ . Δ_Γ is therefore not invertible on the space of functions $u \in C^2(\Gamma)$.

We therefore have to impose the global condition $\int_\Gamma u = 0$ to guarantee the uniqueness of solutions. We define $L_0^2(\Gamma)$ (resp. $H_0^1(\Gamma)$) the space of measurable functions that are square integrable (resp. weakly differentiable with square integrable weak derivative) with zero mean on Γ .

$$L_0^2(\Gamma) = \left\{ f \in L^2(\Gamma), \int_\Gamma f = 0 \right\}, \quad H_0^1(\Gamma) = \left\{ f \in H^1(\Gamma), \int_\Gamma f = 0 \right\}. \quad (14)$$

Let $f \in L_0^2(\Gamma)$. $u \in C^2(\Gamma)$ is a **classical solution** of the Poisson problem provided that

$$-\Delta_\Gamma u = f \text{ on } \Gamma, \quad (15)$$

$$\int_\Gamma u = 0. \quad (16)$$

3.2 Weak form of the Poisson problem

The classical Laplace-Beltrami operator (Definition 2.6) acts on C^2 functions. A classical solution of (15) is therefore a function $u \in C^2(\Gamma)$. Unfortunately such a strong solution doesn't always exist even if f is assumed continuous (see [12] section 3.1.2 in the Euclidean case).

The variational formulation for (15) is the following:

$$\text{Find } u \in H_0^1(\Gamma) \text{ such that } \forall v \in H_0^1(\Gamma), \quad \int_\Gamma \vec{\nabla}_\Gamma u \cdot \vec{\nabla}_\Gamma v = \int_\Gamma f v. \quad (17)$$

We obtain (17) from (15) by applying the Green's formula (13). (17) implies (15) only if $u \in C^2(\Gamma)$. A solution $u \in H_0^1(\Gamma)$ of (17) is called **weak solution** for the Poisson problem. Indeed, it is only once weakly differentiable whereas a strong (classical) solution is twice differentiable.

3.3 Existence result

The existence and uniqueness of weak solutions for the variational formulation (17) of problem (15) is a classical result for smooth manifolds Γ (see for instance [2] chapter 4 section 1.2). The following statement is taken from [4] Theorem 1 b).

Theorem 3.1 (Existence and uniqueness of weak solutions on a C^3 manifold)

Let Γ be a closed embedded C^3 hypersurface in \mathbb{R}^3 . For every $f \in L^2(\Gamma)$ with $\int_\Gamma f = 0$, there exists a weak solution $u \in H_0^1(\Gamma)$ of $-\Delta u = f$ on Γ . Furthermore u is unique up to a constant.

Proof: see theorem 1 b) in [4].

□

4 The finite element method

The finite element method (FEM), is a numerical method for solving problems of engineering and mathematical physics ([1], [12]). We first approximate the closed hypersurface Γ by a closed polyhedral surface Γ_h with triangular faces. We approximate the right hand side function $f \in L_0^2(\Gamma)$ by a function $f_h \in L_0^2(\Gamma_h)$.

In section 4.1, we look for $\tilde{u}_h \in PL_0(\Gamma_h)$ the projection of the solution $u_h \in H_0^1(\Gamma_h)$ of the Poisson problem $-\Delta_{\Gamma_h} \tilde{u}_h = f_h$, on the space of continuous piecewise linear functions with zero mean $PL_0(\Gamma_h)$.

The finite element matrix is symmetric and positive but not invertible (section 4.2). However the finite element linear system admits a unique solution provided the right hand side has zero mean.

4.1 The finite element matrix

We consider a closed triangulated surface Γ_h having $n \in \mathbb{N}$ nodes. We define $PL_0(\Gamma_h) \subset H_0^1(\Gamma_h)$, the subspace of piecewise linear functions on Γ_h that have zero mean. Functions ϕ of $PL_0(\Gamma_h)$ are linear in the sense that they take the form $\phi(\vec{x}) = \alpha^T x + \beta^T y + \gamma^T z + \zeta^T$ on each of the triangles \mathcal{T} composing Γ_h .

The discrete form of the variational formulation of the Poisson equation (15) is the following.

$$\text{Find } \tilde{u}_h \in PL_0(\Gamma_h) \text{ such that } \forall \tilde{v}_h \in PL_0(\Gamma_h), \int_{\Gamma_h} \vec{\nabla}_{\Gamma_h} \tilde{u}_h \cdot \vec{\nabla}_{\Gamma_h} \tilde{v}_h = \int_{\Gamma_h} f_h \tilde{v}_h. \quad (18)$$

$PL_0(\Gamma_h)$ is generated by the n nodal functions $\phi_i : \Gamma_h \rightarrow \mathbb{R}$, $i = 1, \dots, n$ such that $\phi_i(x_j) = \delta_{ij}$.

The solution \tilde{u}_h of the discrete Poisson problem (18) must therefore satisfy the following system of equations

$$\forall i \in \{1, \dots, n\}, \quad \int_{\Gamma_h} \vec{\nabla}_{\Gamma_h} \tilde{u}_h \cdot \vec{\nabla}_{\Gamma_h} \phi_i = \int_{\Gamma_h} f_h \phi_i, \quad (19)$$

which takes the algebraic form

$$A_{\Delta_{\Gamma_h}} X = b_h, \quad (20)$$

where the unknown vector $X = {}^t(u_1, \dots, u_n)$ is the vector of components of \tilde{u}_h on the nodal basis :

$$\tilde{u}_h = \sum_{i=1}^n u_i \phi_i. \quad (21)$$

The coefficients of the system matrix $A_{\Delta_{\Gamma_h}} = (a_{ij})_{i,j=1,\dots,n}$, and of the right hand side vector $b_h = {}^t(b_1, \dots, b_n)$ are given by

$$a_{ij} = \int_{\Gamma_h} \vec{\nabla}_{\Gamma_h} \phi_i \cdot \vec{\nabla}_{\Gamma_h} \phi_j, \quad b_j = \int_{\Gamma_h} f \phi_j. \quad (22)$$

Theorem 4.1 (Properties of $A_{\Delta_{\Gamma_h}}$)

Consider a polyhedral surface Γ_h . The finite element matrix $A_{\Delta_{\Gamma_h}}$ (equation 22) satisfies

- $A_{\Delta_{\Gamma_h}}$ is symmetric and positive
- $\ker A_{\Delta_{\Gamma_h}}$ is the set of constant functions
- $\text{Im} A_{\Delta_{\Gamma_h}}$ is the set of functions with zero mean

$A_{\Delta_{\Gamma_h}}$ is not invertible since constants are in its kernel, hence the linear system (20) is singular. However it admits a unique solution with zero mean provided the right hand side has zero mean (see theorem 4.2 in the next section).

4.2 Existence of a finite element approximation

Due to the absence of boundary, a technical difficulty in the numerical solution of linear systems arising from PDEs on closed surfaces is that the solution should be sought for in spaces of function with nil average.

First we note that since $A_{\Delta_{\Gamma_h}}$ is singular, we need to impose a condition on the right-hand side b_h for the solvability of the discrete system (20). Following the continuous setting (Theorem 3.1), we impose that the discrete right hand side function f_h must have zero mean :

$$\int_{\Gamma_h} f_h = \sum_{i=1}^n b_i \int_{\Gamma_h} \phi_i = 0. \quad (23)$$

With the latter condition on the right hand side, since Δ_{Γ_h} is an endomorphism of $PL_0(\Gamma_h)$, we can state the following existence theorem.

Theorem 4.2 (Existence theorem for the linear system)

Let $PL_0(\Gamma_h)$ be the space of piecewise linear finite elements on the discrete surface Γ_h . Let $f_h \in PL_0(\Gamma_h)$ with $\int_{\Gamma_h} f_h = 0$. Then there exists a unique discrete solution $\tilde{u}_h \in PL_0(\Gamma_h)$ to the discrete Poisson problem (18) with the property that $\int_{\Gamma_h} \tilde{u}_h = 0$.

5 Estimate of the condition number

The condition number of an invertible matrix A relative to the norm $\|\cdot\|$ is : $cond(A) = \|A\| \times \|A^{-1}\|$. The Euclidean norm is quite often used in applications and the expression of the L^2 -condition number for a symmetric matrix is

$$cond_2(A) = \frac{\lambda_{\max}(|A|)}{\lambda_{\min}(|A|)}, \quad (24)$$

where $\lambda_{\max}(|A|)$ (resp. $\lambda_{\min}(|A|)$) is the largest (resp. smallest) eigenvalue of A in absolute value.

In the case of the finite element matrix $A_{\Delta_{\Gamma_h}}$, obtained from the discretisation of the Laplace-Beltrami operator on a closed piecewise triangular surface Γ_h (equation 22), one has to deal firstly with the fact that $A_{\Delta_{\Gamma_h}}$ is not invertible, secondly with the presence of a curvature field.

In the sequel, each triangle of Γ_h is denoted \mathcal{T} , its area is denoted $|\mathcal{T}|$ and its diameter is denoted

$$h_{\mathcal{T}} = diam(\mathcal{T}) = \max_{x,y \in \mathcal{T}} \|x - y\|. \quad (25)$$

Theorem 5.1 (Condition number of the finite element matrix)

There exists a constant c such that for any closed piecewise triangular surface $\Gamma_h \subset \mathbb{R}^3$, the condition number of $A_{\Delta_{\Gamma_h}}$ (equation 22) satisfies

$$\mathcal{K}_h(A_{\Delta_{\Gamma_h}}) \leq \frac{n \max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{m(\Gamma_h) c \min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2},$$

where $m(\Gamma_h)$ is the Poincaré's constant of Γ_h .

As is the case in the Euclidean context, the condition number of the finite element matrix for the Laplace-Beltrami operator is in $\mathcal{O}(h^{-2})$. The curvature of the surface affects the condition number through the Poincaré constant $m(\Gamma_h)$.

The proof of theorem 5.1 consists in five lemma adapting to the curved surfaces the steps followed in the Euclidean case (section 9.1.4 in [6]). We start with a lemma regarding the affine geometry of 3D triangles.

Lemma 5.1

Let $A_0, B_0, C_0, A, B, C \in \mathbb{R}^3$ such that A_0, B_0, C_0 as well as A, B, C are not aligned. Let \mathcal{T}_0 (resp. \mathcal{T}) be the triangle formed by A_0, B_0 and C_0 (resp A, B and C). There exists an affine operator

$$\begin{aligned} T_{\mathcal{T}} : \mathcal{T} &\rightarrow \mathcal{T}_0 \\ x &\rightarrow J_{\mathcal{T}}x + b_{\mathcal{T}} \end{aligned}$$

where $J_{\mathcal{T}}$ is a 3×3 matrix and $b_{\mathcal{T}} \in \mathbb{R}^3$.

Proof:

Since A, B, C are not aligned, they define a plane (\mathcal{P}) with unit normal \vec{n} and A, B, C is an affine frame of reference of (\mathcal{P}) .

Since A_0, B_0, C_0 are not aligned, they define a plane (\mathcal{P}_0) with unit normal \vec{n}_0 and A_0, B_0, C_0 is an affine frame of reference of (\mathcal{P}_0) .

Since $(\overrightarrow{AB}, \overrightarrow{AC}, \vec{n})$ and $(\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0)$ form a basis of \mathbb{R}^3 , $J_{\mathcal{T}}$ is defined as the transition matrix from the basis $(\overrightarrow{AB}, \overrightarrow{AC}, \vec{n})$ to the basis $(\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0)$, and we have

$$\begin{aligned} J_{\mathcal{T}}\overrightarrow{AB} &= \overrightarrow{A_0B_0} \\ J_{\mathcal{T}}\overrightarrow{AC} &= \overrightarrow{A_0C_0} \\ J_{\mathcal{T}}\vec{n} &= \vec{n}_0. \end{aligned}$$

Defining

$$b_{\mathcal{T}} = A_0 - J_{\mathcal{T}}A,$$

the affine transformation $T_{\mathcal{T}}(x) = J_{\mathcal{T}}x + b_{\mathcal{T}}$ satisfies

$$\begin{aligned} T_{\mathcal{T}}(A) &= A_0 \\ T_{\mathcal{T}}(B) &= B_0 \\ T_{\mathcal{T}}(C) &= C_0. \end{aligned}$$

Let $P \in \mathcal{T}$, since $(\overrightarrow{AB}, \overrightarrow{AC})$ is a basis of (\mathcal{P}) there are coefficients $\alpha, \beta \in \mathbb{R}$ such that

$$P = A + \alpha\overrightarrow{AB} + \beta\overrightarrow{AC},$$

hence

$$\begin{aligned} T_{\mathcal{T}}(P) &= T_{\mathcal{T}}(A) + \alpha J_{\mathcal{T}}\overrightarrow{AB} + \beta J_{\mathcal{T}}\overrightarrow{AC} \\ &= A_0 + \alpha\overrightarrow{A_0B_0} + \beta\overrightarrow{A_0C_0} \\ &= P_0 \in \mathcal{T}_0. \end{aligned}$$

Hence $T_{\mathcal{T}} : \mathcal{T} \rightarrow \mathcal{T}_0$, and the theorem is proved. □

Lemma 5.2

Under the hypotheses of Lemma 5.1, letting $M_{\mathcal{T}} = [\overrightarrow{AB}, \overrightarrow{AC}, \vec{n}]$ be the transition matrix from the canonical basis to the basis $(\overrightarrow{AB}, \overrightarrow{AC}, \vec{n})$, and $M_{\mathcal{T}_0} = [\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0]$ be the transition matrix from the canonical basis to the basis $(\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0)$, $J_{\mathcal{T}}$ takes the form

$$J_{\mathcal{T}} = M_{\mathcal{T}_0}M_{\mathcal{T}}^{-1}$$

and furthermore

$$\det(J_{\mathcal{T}}) = \frac{|\mathcal{T}_0|}{|\mathcal{T}|}.$$

Proof:

Since $J_{\mathcal{T}}$ is the transition matrix from the basis $(\overrightarrow{AB}, \overrightarrow{AC}, \vec{n})$ to the basis $(\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0)$, we have

$$J_{\mathcal{T}}M_{\mathcal{T}} = M_{\mathcal{T}_0}.$$

Since $M_{\mathcal{T}}$ is invertible, multiplying the previous relation by $M_{\mathcal{T}}^{-1}$ we obtain $J_{\mathcal{T}} = M_{\mathcal{T}_0} M_{\mathcal{T}}^{-1}$. Furthermore, taking the determinant yields

$$\det(J_{\mathcal{T}}) = \frac{|\mathcal{T}_0|}{|\mathcal{T}|},$$

since

$$\begin{aligned} \det(M_{\mathcal{T}}) &= \det(\overrightarrow{AB}, \overrightarrow{AC}, \vec{n}) = (\overrightarrow{AB} \wedge \overrightarrow{AC}) \cdot \vec{n} = |\mathcal{T}| \\ \det(M_{\mathcal{T}_0}) &= \det(\overrightarrow{A_0B_0}, \overrightarrow{A_0C_0}, \vec{n}_0) = (\overrightarrow{A_0B_0} \wedge \overrightarrow{A_0C_0}) \cdot \vec{n}_0 = |\mathcal{T}_0| \end{aligned}$$

□

Let \mathcal{T}_0 be a reference triangle in \mathbb{R}^3 with three non aligned vertices $\vec{x}_1, \vec{x}_2, \vec{x}_3$. In the proof of the main theorem, \mathcal{T}_0 will be chosen such that $|\mathcal{T}_0|^2 = \max_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2$. The constant c in the main theorem is given by the following lemma.

Lemma 5.3

Denote $\phi_i^0, i \in \{1, 2, 3\}$ the shape functions associated to each node of \mathcal{T}_0 , and define the function

$$\begin{aligned} f_0 : \mathbb{R}^n &\rightarrow \mathbb{R} \\ (u_1, u_2, \dots, u_n) &\rightarrow \int_{\mathcal{T}_0} \left| \sum_{i=1}^n u_i \phi_i^0 \right|^2. \end{aligned}$$

f_0 is a strictly positive quadratic form on \mathbb{R}^n .

Furthermore, there exist $c > 0, C > 0$ such that

$$\forall (u_1, u_2, \dots, u_n) \in \mathbb{R}^n, \quad c \sum_{i=1}^n u_i^2 \leq \int_{\mathcal{T}_0} \left| \sum_{i=1}^n u_i \phi_i^0 \right|^2 \leq C \sum_{i=1}^n u_i^2.$$

Proof:

f_0 is a positive quadratic form since

$$\forall (u_1, u_2, \dots, u_n) \in \mathbb{R}^n, \quad \int_{\mathcal{T}_0} \left| \sum_{i=1}^n u_i \phi_i^0 \right|^2 \geq 0.$$

First remark that:

$$\begin{aligned} f_0(U) = 0 &\iff \int_{\mathcal{T}_0} \left| \sum_{i=1}^n u_i \phi_i^0 \right|^2 = 0 \\ &\iff \left| \sum_{i=1}^n u_i \phi_i^0 \right|^2 = 0 \\ &\iff \sum_{i=1}^n u_i \phi_i^0 = 0 \\ &\iff \forall x \in \mathcal{T}_0, \sum_{i=1}^n u_i \phi_i^0(x) = 0. \end{aligned}$$

Since $\phi_i^0(x)$ are basis functions, we deduce that

$$f_0(U) = 0 \iff U = 0,$$

that is f_0 is a strictly positive quadratic form.

Secondly, we seek $c > 0$ and $C > 0$ such that

$$\forall U = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n, \quad c \sum_{i=1}^n u_i^2 \leq f_0(U) \leq C \sum_{i=1}^n u_i^2.$$

Since f_0 is a strictly positive quadratic form on \mathbb{R}^3 , its matrix M is symmetric and positive definite. M thus admits a minimum eigenvalue $\lambda_{\min} > 0$ and a maximum eigenvalue $\lambda_{\max} > 0$ (spectral theorem for symmetric matrices).

Since $f_0(U) = {}^t U M U$ we deduce

$$\forall U \in \mathbb{R}^n, \quad \lambda_{\min} \sum_{i=1}^n u_i^2 \leq f_0(U) \leq \lambda_{\max} \sum_{i=1}^n u_i^2.$$

Taking $c = \lambda_{\min} > 0$ and $C = \lambda_{\max} > 0$ we obtain the desired result. \square

In the following two lemma (lemma 5.4 and Lemma 5.5) the lower bound and the upper bound of the following bilinear form a_h

$$a_h(\tilde{u}_h, \tilde{v}_h) = \int_{\Gamma_h} \vec{\nabla}_{\Gamma_h} \tilde{u}_h \cdot \vec{\nabla}_{\Gamma_h} \tilde{v}_h, \quad (26)$$

on $PL_0(\Gamma_h)$ are given.

Lemma 5.4 (Lower bound of a_h)

We have

$$\forall \tilde{u}_h = \sum_{i=1}^n u_i \phi_i \in PL_0(\Gamma_h), \quad a_h(\tilde{u}_h, \tilde{u}_h) \geq m(\Gamma_h) c n_c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2} \sum_{i=1}^n u_i^2$$

Where $m(\Gamma_h)$ is a Poincaré's constant, $n(i)$ is the number of cells surrounding the node i and $n_c = \min_i n(i)$.

Proof:

Using Poincaré inequality we have

$$\begin{aligned} a_h(\tilde{u}_h, \tilde{u}_h) &\geq m(\Gamma_h) \int_{\Gamma_h} |\tilde{u}_h|^2 \\ &\geq m(\Gamma_h) \sum_{\mathcal{T} \in \mathcal{T}_h} \int_{\mathcal{T}} |\tilde{u}_h|^2. \end{aligned}$$

Let $N_{\mathcal{T}}$ denote the set of nodes surrounding the cell \mathcal{T} . Since $\phi_i(x) = 0$ when $x \in \mathcal{T}$ and $i \notin N_{\mathcal{T}}$, we have

$$a_h(\tilde{u}_h, \tilde{u}_h) \geq m(\Gamma_h) \sum_{\mathcal{T} \in \mathcal{T}_h} \int_{\mathcal{T}} \left| \sum_{i \in N_{\mathcal{T}}} u_i \phi_i \right|^2.$$

Using a change of variable we obtain

$$a_h(\tilde{u}_h, \tilde{u}_h) \geq m(\Gamma_h) \sum_{\mathcal{T} \in \mathcal{T}_h} |\det(J_{\mathcal{T}})|^{-2} \int_{\mathcal{T}_0} \left| \sum_{i \in N_{\mathcal{T}}} u_i \phi_{T_{\mathcal{T}}(i)}^0 \right|^2,$$

where $T_{\mathcal{T}}(i)$ represents the node of \mathcal{T}_0 that is the image of the node i by $T_{\mathcal{T}}$. Using Lemma 5.3 we have

$$a_h(\tilde{u}_h, \tilde{u}_h) \geq m(\Gamma_h) c \sum_{\mathcal{T} \in \mathcal{T}_h} |\det(J_{\mathcal{T}})|^{-2} \sum_{i \in N_{\mathcal{T}}} u_i^2.$$

Using Lemma 5.2 we have

$$\begin{aligned}
a_h(\tilde{u}_h, \tilde{u}_h) &\geq m(\Gamma_h)c \sum_{\mathcal{T} \in \mathcal{T}_h} \frac{|\mathcal{T}|^2}{|\mathcal{T}_0|^2} \sum_{i \in N_{\mathcal{T}}} u_i^2 \\
&\geq m(\Gamma_h)c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2} \sum_{\mathcal{T} \in \mathcal{T}_h} \sum_{i \in N_{\mathcal{T}}} u_i^2 \\
&\geq m(\Gamma_h)c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2} \sum_i n(i) u_i^2.
\end{aligned}$$

Hence we finally have

$$a_h(\tilde{u}_h, \tilde{u}_h) \geq m(\Gamma_h)cn_c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2} \sum_i u_i^2,$$

where $n(i)$ is the number of cells surrounding the node i and

$$n_c = \min_i n(i).$$

□

Lemma 5.5 (Upper bound of a_h)

Under the hypotheses of Lemma 5.4, we have

$$\forall \tilde{u}_h = \sum_{i=1}^n u_i \phi_i \in PL_0(\Gamma_h), \quad a_h(\tilde{u}_h, \tilde{u}_h) \leq 3n_c \frac{h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|} \sum_i u_i^2.$$

Proof:

We decompose a_h over the mesh cells \mathcal{T}_h .

$$\begin{aligned}
a_h(\tilde{u}_h, \tilde{u}_h) &= \sum_{\mathcal{T} \in \mathcal{T}_h} \int_{\mathcal{T}} |\vec{\nabla}_{\Gamma_h} \tilde{u}_h|^2 \\
&= \sum_{\mathcal{T} \in \mathcal{T}_h} \int_{\mathcal{T}} \left| \sum_{i \in N_{\mathcal{T}}} u_i \vec{\nabla}_{\Gamma_h} \phi_i(\vec{x}) \right|^2 \\
&\leq 3 \sum_{\mathcal{T} \in \mathcal{T}_h} \sum_{i \in N_{\mathcal{T}}} u_i^2 \int_{\mathcal{T}} |\vec{\nabla}_{\Gamma_h} \phi_i(\vec{x})|^2. \tag{27}
\end{aligned}$$

The next step consists in the explicit calculation of $\vec{\nabla}_{\Gamma_h} \phi_i(\vec{x})$ in each triangle \mathcal{T} . For more details we refer to A.

Given a triangle \mathcal{T} having nodes $s_1^{\mathcal{T}}, s_2^{\mathcal{T}}$ and $s_3^{\mathcal{T}}$, the gradient of the nodal function associated to $s_1^{\mathcal{T}}$ is

$$\forall \vec{x} \in \mathcal{T}, \quad \vec{\nabla}_{\Gamma_h} \phi_{s_1^{\mathcal{T}}}(\vec{x}) = \frac{1}{-2|\mathcal{T}|} \begin{pmatrix} a_1 \\ b_1 \\ c_1 \end{pmatrix},$$

where

$$\begin{aligned}
a_1 &= ((\vec{x}_{s_3^{\mathcal{T}}})_y - (\vec{x}_{s_2^{\mathcal{T}}})_y)n_z^{\mathcal{T}} + ((\vec{x}_{s_2^{\mathcal{T}}})_z - (\vec{x}_{s_3^{\mathcal{T}}})_z)n_y^{\mathcal{T}} \\
b_1 &= ((\vec{x}_{s_2^{\mathcal{T}}})_x - (\vec{x}_{s_3^{\mathcal{T}}})_x)n_z^{\mathcal{T}} + ((\vec{x}_{s_3^{\mathcal{T}}})_z - (\vec{x}_{s_2^{\mathcal{T}}})_z)n_x^{\mathcal{T}} \\
c_1 &= ((\vec{x}_{s_2^{\mathcal{T}}})_y - (\vec{x}_{s_3^{\mathcal{T}}})_y)n_x^{\mathcal{T}} + ((\vec{x}_{s_3^{\mathcal{T}}})_x - (\vec{x}_{s_2^{\mathcal{T}}})_x)n_y^{\mathcal{T}}.
\end{aligned}$$

Using the inequality $\forall r, s \in \mathbb{R}, (r + s)^2 \leq 2(r^2 + s^2)$, we have :

$$\begin{aligned}
\|\vec{\nabla}_{\Gamma_h} \phi_{s_1^T}(\vec{x})\|_2 &\leq \frac{1}{2|\mathcal{T}|} \sqrt{a_1^2 + b_1^2 + c_1^2} \\
&\leq \frac{1}{2|\mathcal{T}|} \sqrt{(h_{\mathcal{T}} n_z^T + h_{\mathcal{T}} n_y^T)^2 + (h_{\mathcal{T}} n_x^T + h_{\mathcal{T}} n_y^T)^2 + (h_{\mathcal{T}} n_z^T + h_{\mathcal{T}} n_x^T)^2} \\
&\leq \frac{1}{2|\mathcal{T}|} \sqrt{2((h_{\mathcal{T}} n_z^T)^2 + (h_{\mathcal{T}} n_y^T)^2) + 2((h_{\mathcal{T}} n_x^T)^2 + (h_{\mathcal{T}} n_y^T)^2) + 2((h_{\mathcal{T}} n_z^T)^2 + (h_{\mathcal{T}} n_x^T)^2)} \\
&\leq \frac{1}{2|\mathcal{T}|} \sqrt{4((h_{\mathcal{T}} n_z^T)^2 + (h_{\mathcal{T}} n_y^T)^2 + (h_{\mathcal{T}} n_x^T)^2)} \\
&\leq \frac{1}{2|\mathcal{T}|} 2h_{\mathcal{T}} \sqrt{(n_z^T)^2 + (n_y^T)^2 + (n_x^T)^2} \\
&\leq \frac{1}{2|\mathcal{T}|} 2h_{\mathcal{T}}.
\end{aligned}$$

Hence

$$\|\vec{\nabla}_{\Gamma_h} \phi_{s_1^T}(\vec{x})\|_2 \leq \frac{h_{\mathcal{T}}}{|\mathcal{T}|}.$$

Using the same approach we obtain an estimate of the gradient of the two remaining nodal functions

$$\|\vec{\nabla}_{\Gamma_h} \phi_{s_2^T}(\vec{x})\|_2 \leq \frac{h_{\mathcal{T}}}{|\mathcal{T}|}, \quad \|\vec{\nabla}_{\Gamma_h} \phi_{s_3^T}(\vec{x})\|_2 \leq \frac{h_{\mathcal{T}}}{|\mathcal{T}|}.$$

Finally, for any shape function ϕ_i , we have

$$\forall i \in \{s_1^T, s_2^T, s_3^T\}, \|\vec{\nabla}_{\Gamma_h} \phi_i(\vec{x})\| \leq \frac{h_{\mathcal{T}}}{|\mathcal{T}|}.$$

This allows us to deduce the final result from (27):

$$a_h(\tilde{u}_h, \tilde{u}_h) \leq 3n_c \frac{\max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2} \sum_i u_i^2. \quad (28)$$

□

Proof of the main theorem

Using Lemma 5.5, we have

$$\forall \tilde{u}_h \in PL_0(\Gamma_h), \quad \frac{a_h(\tilde{u}_h, \tilde{u}_h)}{\sum_{i=1}^n u_i^2} = \frac{{}^t U_h A_{\Delta_{\Gamma_h}} U_h}{\|U_h\|^2} \leq nn_c \frac{\max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2},$$

and we deduce an upper bound of the spectrum of the finite element operator on $PL_0(\Gamma_h)$

$$\lambda_{\max} \leq nn_c \frac{\max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}. \quad (29)$$

Using Lemma 5.4, we have

$$\forall \tilde{u}_h \in PL_0(\Gamma_h), \quad \frac{a_h(\tilde{u}_h, \tilde{u}_h)}{\sum_{i=1}^n u_i^2} = \frac{{}^t U_h A_{\Delta_{\Gamma_h}} U_h}{\|U_h\|^2} \geq m(\Gamma_h) cn_c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2},$$

and we deduce a lower bound of the spectrum of the finite element operator on $PL_0(\Gamma_h)$

$$\lambda_{\min} \geq m(\Gamma_h) cn_c \frac{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{|\mathcal{T}_0|^2}. \quad (30)$$

Using the definition of $\mathcal{K}_h(A_{\Delta_{\Gamma_h}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$, (30) and (29) yield the following inequality

$$\begin{aligned} \forall \mathcal{T}_0, \quad \mathcal{K}_h(A_{\Delta_{\Gamma_h}}) &\leq \frac{\max_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{m(\Gamma_h)cn_c |\mathcal{T}_0|^2} \times \frac{3n_c \max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|} \\ &\leq \frac{\max_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}{m(\Gamma_h)c |\mathcal{T}_0|^2} \times \frac{3 \max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{\min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|}. \end{aligned}$$

Choosing \mathcal{T}_0 such that $|\mathcal{T}_0|^2 = \max_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2$, we finally have

$$\mathcal{K}_h(A_{\Delta_{\Gamma_h}}) \leq \frac{3 \max_{\mathcal{T} \in \mathcal{T}_h} h_{\mathcal{T}}^2}{m(\Gamma_h)c \min_{\mathcal{T} \in \mathcal{T}_h} |\mathcal{T}|^2}.$$

□

6 Some numerical results

In order to illustrate the numerical method presented in this paper, and motivate the search for efficient preconditioners, we present the numerical simulation of the Poisson problem on a sphere. The continuous problem is presented in section 6.1. Then in section 6.2 we present a sequence of refined meshes of the sphere. The results obtained from the finite element simulation of the Poisson problem on the mesh sequence are given in section 6.3. We analyse these results in section 6.4 and give the convergence curve (picture 4), iteration curve (picture 5), the residual curve (picture 6), and the condition number curve (picture 7).

For the design and meshing of the domain we use GEOMETRY and MESH modules of the software SALOME (see [14, 20]).

For the coding of the script, we use Python with the open-source Linux based library SOLVERLAB [21] which is very practical for the manipulation of large matrices, vectors, meshes and fields. It (SOLVERLAB) can handle finite element and finite volume discretizations, read general 1D, 2D and 3D geometries and meshes generated by SALOME.

For the numerical resolution of our discrete problem, we use an iterative solver because the stiffness matrix $A_{\Delta_{\Gamma_h}}$ is large, sparse (see [5]) and singular. The library PETSc [15], encapsulated in SOLVERLAB, provides linear solvers for singular systems.

For the visualization of the result, we use the PARAVIS module included in SALOME (see [20]).

6.1 The Poisson problem on the unit sphere

We consider the unit sphere defined as follows

$$\Gamma_{sphere} = \{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\}.$$

The sphere is a C^∞ manifold of dimension 2 embedded in \mathbb{R}^3 , hence the Laplace-Beltrami operator $\Delta_{\Gamma_{sphere}}$ is well defined on Γ_{sphere} .

Using the spherical coordinates (θ, ϕ) where θ is the longitude and ϕ the latitude, the Laplace-Beltrami operator takes the following form for any function $u \in C^2(\Gamma_{sphere})$:

$$\Delta_{\Gamma_{sphere}} u = \frac{1}{\sin(\phi)} \frac{\partial}{\partial \phi} \left(\sin \phi \frac{\partial u}{\partial \phi} \right) + \frac{1}{\sin(\phi)^2} \frac{\partial^2 u}{\partial \theta^2}.$$

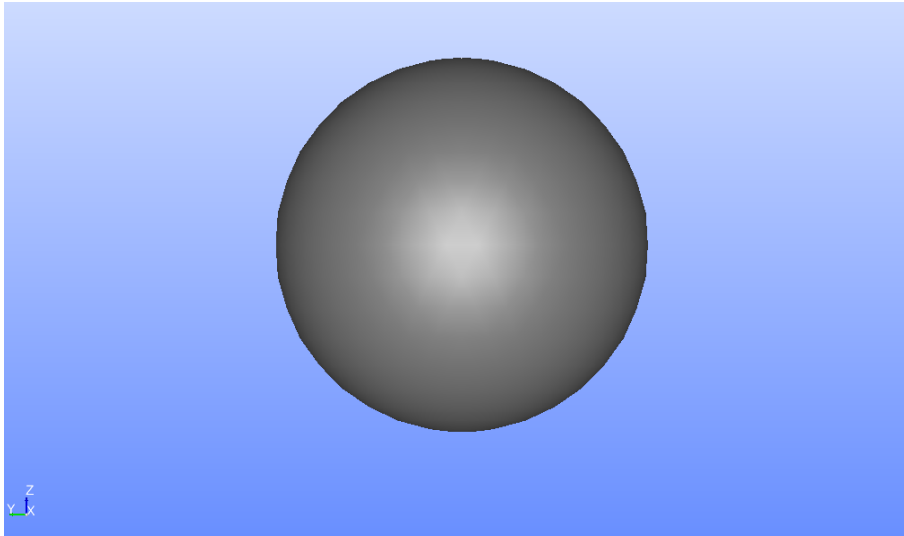


Figure 1: The unit sphere

We consider the following **Poisson problem** on the sphere Γ_{sphere}

$$\begin{cases} -\Delta_{\Gamma_{sphere}} u = f_{sphere} \text{ on } \Gamma_{sphere} \\ \int_{\Gamma_{sphere}} u = 0 \end{cases}, \quad (31)$$

With the following choice for f_{sphere} :

$$f_{sphere}(x, y, z) = \frac{12}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} (3x^2y - y^3).$$

The exact solution u of (31) is given by (see [13]):

$$u_{sphere}(x, y, z) = \frac{1}{(x^2 + y^2 + z^2)^{\frac{3}{2}}} (3x^2y - y^3) = \frac{1}{12} f_{sphere}.$$

One can check that f_{sphere} and u_{sphere} are **zero mean functions**.

Our objective is to solve numerically the Poisson problem (31) using the finite element method described in section 4. We first build a sequence of refined meshes in section 6.2. Then we run the simulation on each mesh. Some results are displayed in section 6.3, and an analysis of the results is performed in section 6.4.

6.2 Meshing of the sphere

In order to assess the Finite Element discretisation of the Poisson problem on the unit sphere, we build a sequence of refined meshes that will enable us to measure the convergence and computational time of the numerical method. The CAD model of the sphere (picture 1) was done with in the GEOMETRY module of the platform SALOME. For the design meshing of the sphere we use the MESH module of the platform SALOME (see [14, 22, 20]).

Below are screenshots of the meshes used in our convergence and computational time analysis. The meshes are generated by a Delaunay type triangulation of the surface with all the nodes belonging to the sphere.

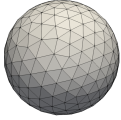
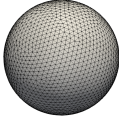
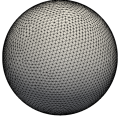
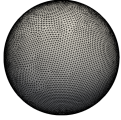
meshSphere 1	meshSphere 2	meshSphere 3	meshSphere 4
			
288 cells	2638 cells	4512 cells	10773 cells

Figure 2: Mesh of domain

Using the SOLVERLAB python module, the right hand side of the Poisson problem (31) is interpolated on the meshes, and the rigidity matrices are filled with a sparse matrix structured encapsulated from PETSc [15]. The linear systems are then solved using a conjugate gradient algorithm [5, 16] after the setting of a non zero nullspace [19].

6.3 Visualization of the results

For the numerical resolution of our discrete problem, we use an iterative solver because the stiffness matrix $A_{\Delta\Gamma_h}$ is large and sparse (see [5]).

For the visualization of the result, we use the PARAVIS module of the platform SALOME (see [20]).

Below are visualizations of the numerical results obtained on the different meshes of picture 2.

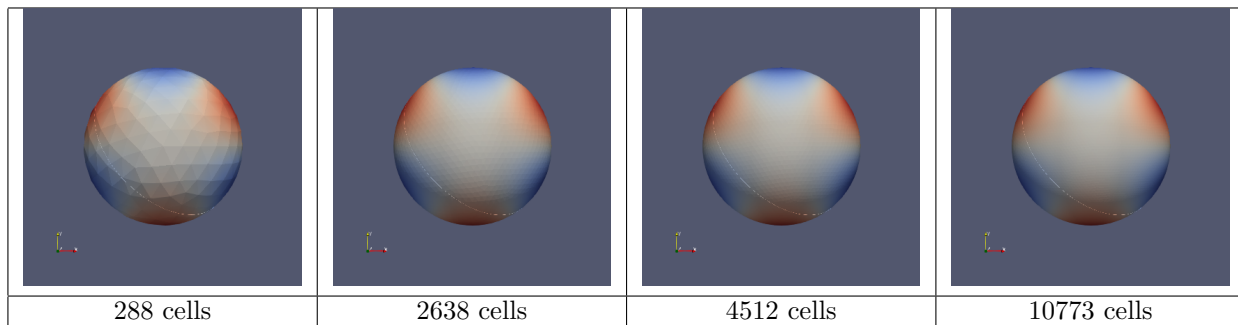


Figure 3: Numerical results of the finite elements on the unit sphere

6.4 Discussion of the numerical results

Numerical convergence of the finite element method The picture 4 displays the evolution of the numerical error $\|u_h - u_{sphere}\|$ with the cell minimal diameter h , in logarithmic scale. The numerical error $\|u_h - u_{sphere}\|$ is taken as the supremum of $|u_h(x_i) - u_{sphere}(x_i)|$ over all the nodes x_i .

The theoretical error is composed of two contributions. The first is the interpolation error of u_{sphere} from the sphere Γ_{sphere} to a polyhedron Γ_h and is $\mathcal{O}(h)$. The second contribution is the approximation error coming from the finite element discreti-

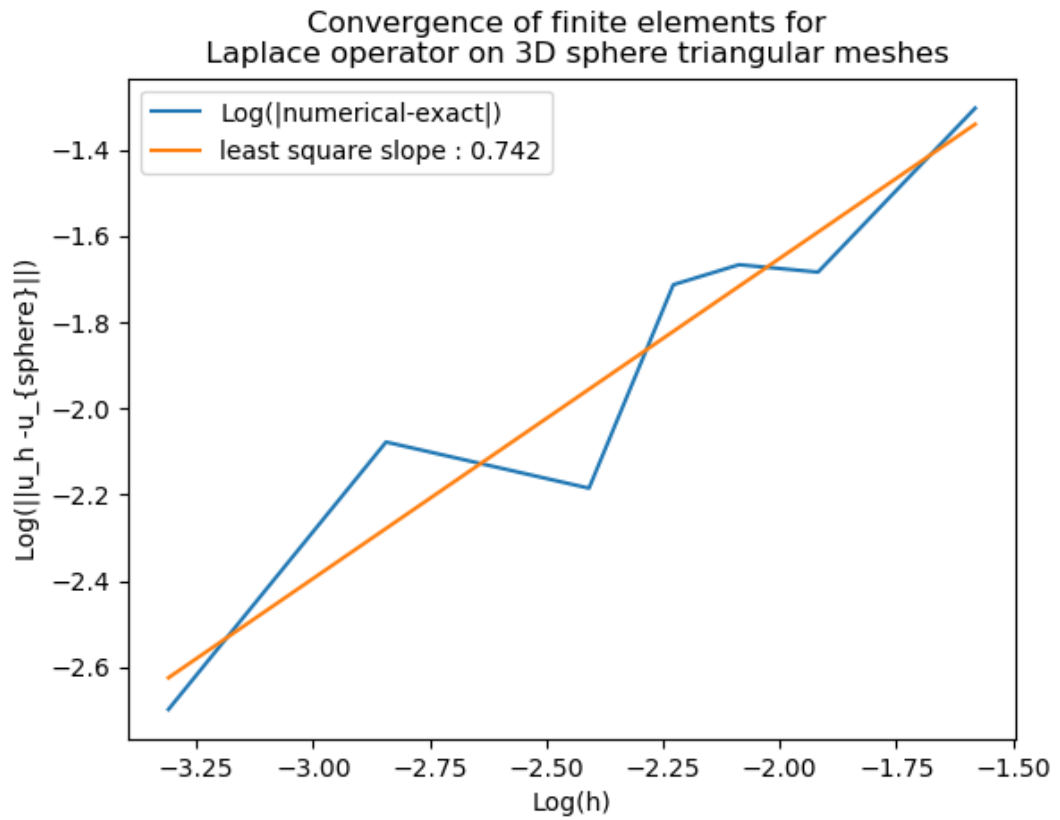


Figure 4: Convergence of the finite element method on the sphere

sation of the Poisson equation. This error is $\mathcal{O}(h^2)$ [4, Theorem 8]. The total error is therefore $\mathcal{O}(h)$.

We observe on picture 4 that the method converges with a numerical order of approximately 0.742. We expect that using meshes larger than 1 million cells we will come closer to the theoretical value of 1.

Number of CG iterations for the finite element method on the sphere

It is not possible to use a direct solvers for the numerical resolution of the linear system $A_{\Delta\Gamma_h}X = b_h$, since they apply only to invertible matrices. We used instead the Conjugate Gradient (CG) [16] method with Incomplete LU factorisation [17] as preconditioner to solve our singular linear system thanks to PETSc algorithm [18, 19]. The picture 5 displays the evolution of the number of CG iterations with the number of nodes. The number of iterations increases linearly with the number of nodes of the mesh.

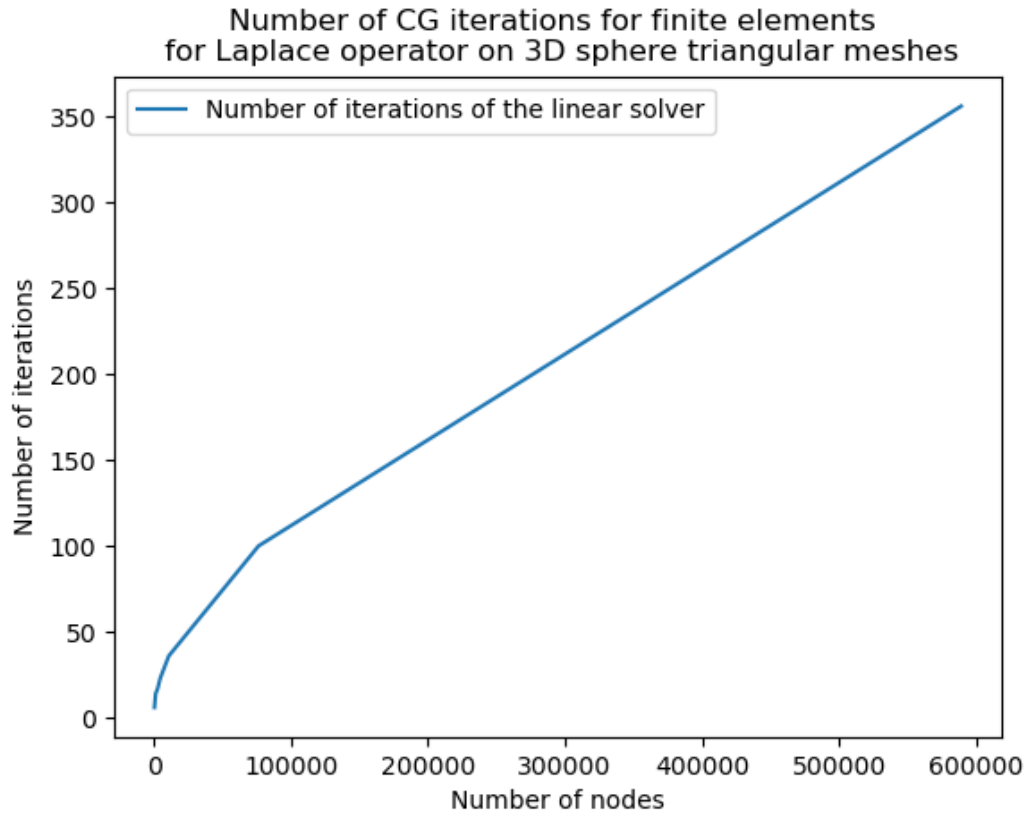


Figure 5: Number of CG iterations for the finite element method on the sphere

CG Residual for the finite elements methods on the sphere The picture 6 displays the evolution in Logarithmic scale of the residual $\epsilon_h = \|A_{\Delta\Gamma_h} X - b_h\|$ with the number of nodes in the mesh \mathcal{T}_h . This residual ϵ_h evolves from about 10^{-2} to about 10^{-5} as the number of nodes evolves from 288 to 600000. This is because we had to adapt the precision of the linear solver to the number of nodes. Indeed the matrix of the linear system $A_{\Delta\Gamma_h} X = b_h$ is singular and thus b_h should belong to the range of $A_{\Delta\Gamma_h}$ up to machine precision. The theoretical range of $A_{\Delta\Gamma_h}$ consists in vectors of zero mean, but at the computer level zero becomes machine precision and thus b_h should have mean lower than machine precision. If the machine precision is too small (say 10^{-10}) then the linear solver will fail when the integral of b_h is larger than 10^{-10} .

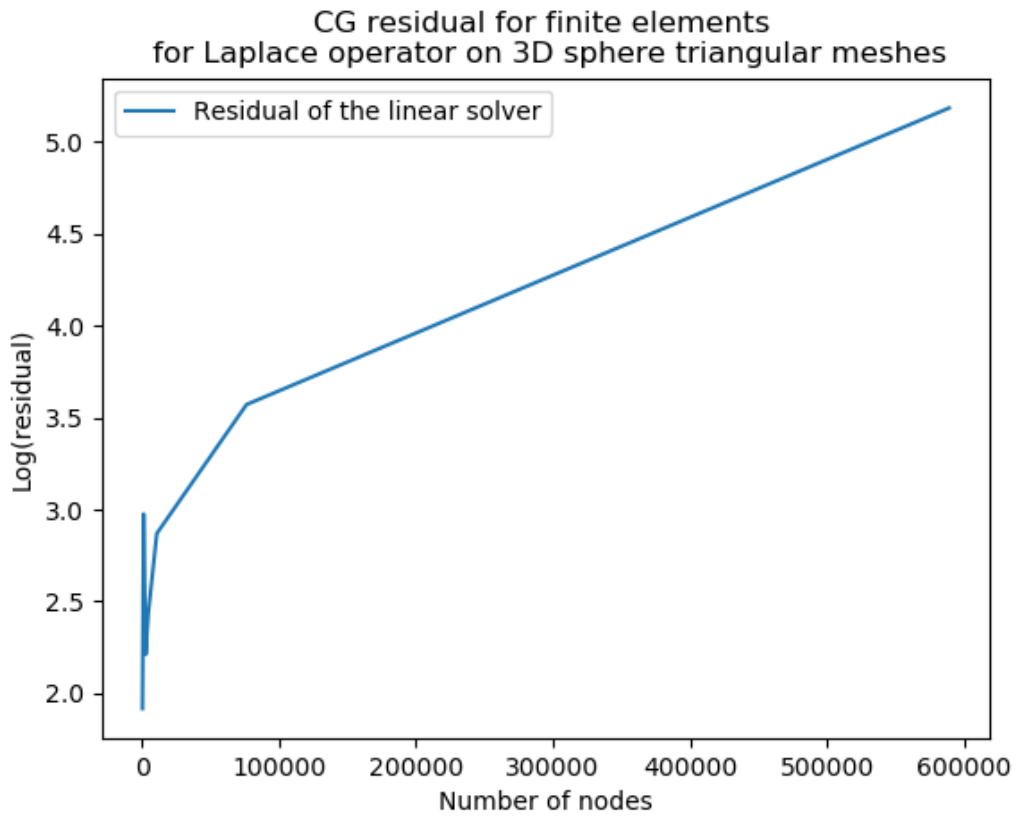


Figure 6: CG residual for the finite element method on the sphere

Condition number of the finite element matrix on the sphere The picture 7 displays the evolution of the condition number with the cell minimal diameter h , in logarithmic scale. We observe first that the condition number increase as $h \rightarrow \infty$.

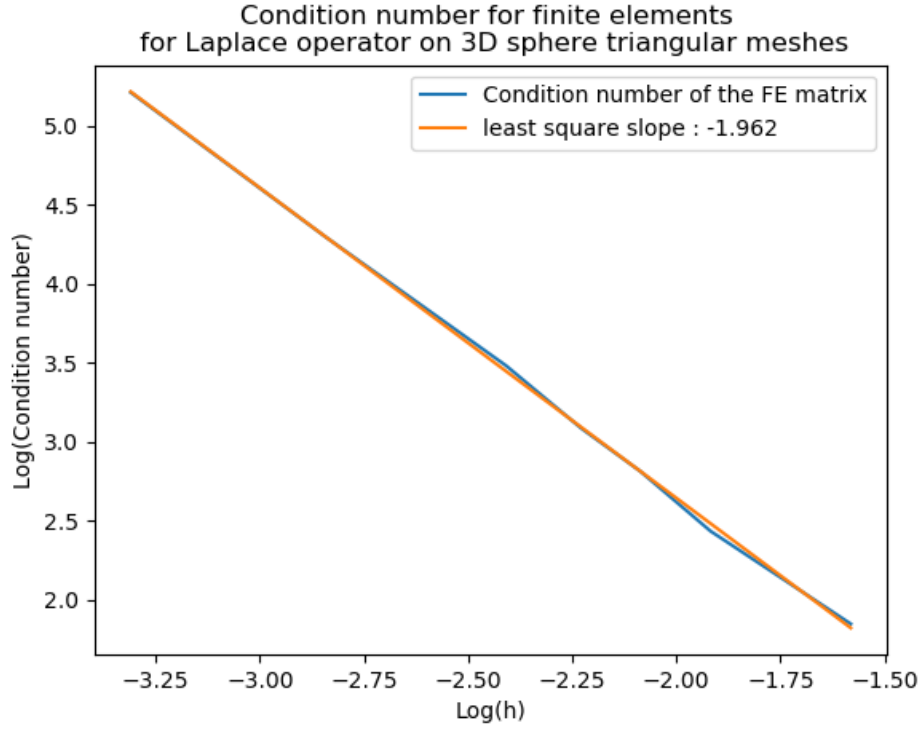


Figure 7: Condition number for the finite element method on the sphere

Furthermore, the condition number grows as h to the power -1.962 , which is very close to the theoretical value of 2 given in our main result in theorem 5.1.

7 Conclusion and perspectives

We have seen that the properties of the finite element matrix on 3D surfaces are very similar to those in Euclidean spaces. The main difference when the surface is closed is that the matrix is not invertible, which makes the proofs more technical. However, adapting the technique used in the Euclidean context, we gave an upper bound for the condition number in $\mathcal{O}(h^{-2})$. Using a similar techniques, it is possible to derive a lower bound of the condition number following for instance [6] in the Euclidean case. In section 6, we have given some numerical results of the simulation of the Poisson problem on a sphere. The results showed that the discretisation converges and that the condition number increases nearly as $\mathcal{O}(h^{-2})$.

The numerical simulations performed in section 6 showed the number of iterations of the linear solver increases linearly with the mesh size. This yields a sharp increase of the computational time as the number of nodes increases. Our simulation used the Incomplete LU factorisations as preconditioner but more advanced techniques such as multigrid perform better in the Euclidean context. Their adaptation to the curved would be based on a fine analysis of the finite element matrix taking into account its singularity. This work can therefore be seen as a first step in the design and analysis of advanced preconditioners for singular systems, particularly those arising from the discretisation of PDEs on closed surfaces.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

A Coefficients of the finite element matrix

We compute the coefficients of the finite element matrix (22) on a closed piecewise triangular surface Γ_h . The mesh \mathcal{T}_h of the domain Γ_h is composed of triangular elements $(\mathcal{T}_k)_{k \geq 1}$ having non zero area. The n vertices of Γ_h are denoted $\vec{x}_1, \vec{x}_2, \vec{x}_3, \dots, \vec{x}_n$. To each vertex \vec{x}_i , we associate a nodal function, $\phi_i : \Gamma_h \rightarrow \mathbb{R} \in PL_0(\Gamma_h)$ such that $\phi_i(x_j) = \delta_{ij}$.

We observe that in 3D, functions $\phi \in PL_0(\Gamma_h)$ take the following form on each triangle $\mathcal{T} \in \mathcal{T}_h$

$$\phi(\vec{x}) = \alpha^T x + \beta^T y + \gamma^T z + \zeta^T. \quad (32)$$

The gradient of any function $\phi \in PL_0(\Gamma_h)$ is thus constant on each triangle $\mathcal{T} \in \mathcal{T}_h$ and its components can be deduced from (32) and (9) :

$$\forall \vec{x} \in \mathcal{T}, \quad \vec{\nabla}_{\Gamma_h} \phi(\vec{x}) = \left(\alpha^T, \beta^T, \gamma^T \right). \quad (33)$$

To determine the gradient of a function $\phi \in PL_0(\Gamma_h)$ on a triangle \mathcal{T} , we first remark that according to Lemma 2.1, we have:

$$n_x^T \alpha^T + n_y^T \beta^T + n_z^T \gamma^T = 0. \quad (34)$$

$$\vec{\nabla}_{\Gamma_h} \phi_{s_2^{\mathcal{T}}}(\vec{x}) = \frac{1}{-2|\mathcal{T}|} \left(\begin{array}{c|ccc} 0 & (\vec{x}_{s_1^{\mathcal{T}}})_y & (\vec{x}_{s_1^{\mathcal{T}}})_z & 1 \\ 1 & (\vec{x}_{s_2^{\mathcal{T}}})_y & (\vec{x}_{s_2^{\mathcal{T}}})_z & 1 \\ 0 & (\vec{x}_{s_3^{\mathcal{T}}})_y & (\vec{x}_{s_3^{\mathcal{T}}})_z & 1 \\ 0 & n_y & n_z & 0 \\ (\vec{x}_{s_1^{\mathcal{T}}})_x & 0 & (\vec{x}_{s_1^{\mathcal{T}}})_z & 1 \\ (\vec{x}_{s_2^{\mathcal{T}}})_x & 1 & (\vec{x}_{s_2^{\mathcal{T}}})_z & 1 \\ (\vec{x}_{s_3^{\mathcal{T}}})_x & 0 & (\vec{x}_{s_3^{\mathcal{T}}})_z & 1 \\ n_x & 0 & n_z & 0 \\ (\vec{x}_{s_1^{\mathcal{T}}})_x & (\vec{x}_{s_1^{\mathcal{T}}})_y & 0 & 1 \\ (\vec{x}_{s_2^{\mathcal{T}}})_x & (\vec{x}_{s_2^{\mathcal{T}}})_y & 1 & 1 \\ (\vec{x}_{s_3^{\mathcal{T}}})_x & (\vec{x}_{s_3^{\mathcal{T}}})_y & 0 & 1 \\ n_x & n_y & 0 & 0 \end{array} \right)$$

$$\vec{\nabla}_{\Gamma_h} \phi_{s_3^{\mathcal{T}}}(\vec{x}) = \frac{1}{-2|\mathcal{T}|} \left(\begin{array}{c|ccc} 0 & (\vec{x}_{s_1^{\mathcal{T}}})_y & (\vec{x}_{s_1^{\mathcal{T}}})_z & 1 \\ 0 & (\vec{x}_{s_2^{\mathcal{T}}})_y & (\vec{x}_{s_2^{\mathcal{T}}})_z & 1 \\ 1 & (\vec{x}_{s_3^{\mathcal{T}}})_y & (\vec{x}_{s_3^{\mathcal{T}}})_z & 1 \\ 0 & n_y & n_z & 0 \\ (\vec{x}_{s_1^{\mathcal{T}}})_x & 0 & (\vec{x}_{s_1^{\mathcal{T}}})_z & 1 \\ (\vec{x}_{s_2^{\mathcal{T}}})_x & 0 & (\vec{x}_{s_2^{\mathcal{T}}})_z & 1 \\ (\vec{x}_{s_3^{\mathcal{T}}})_x & 1 & (\vec{x}_{s_3^{\mathcal{T}}})_z & 1 \\ n_x & 0 & n_z & 0 \\ (\vec{x}_{s_1^{\mathcal{T}}})_x & (\vec{x}_{s_1^{\mathcal{T}}})_y & 0 & 1 \\ (\vec{x}_{s_2^{\mathcal{T}}})_x & (\vec{x}_{s_2^{\mathcal{T}}})_y & 0 & 1 \\ (\vec{x}_{s_3^{\mathcal{T}}})_x & (\vec{x}_{s_3^{\mathcal{T}}})_y & 1 & 1 \\ n_x & n_y & 0 & 0 \end{array} \right).$$

References

- [1] C. M. Elliott, B. Stinner, Computation of two-phase biomembranes with phase dependent material parameters using surface finite elements, *Commun. Comput. Phys.* **13**(2010), 325-360.
- [2] T. Aubin, Some nonlinear problems in Riemannian geometry, Springer, 1998.
- [3] K. Hildebrandt, K. Polthier, On approximation of the Laplace-Beltrami operator and the Willmore energy of surfaces, *Eurographics Symposium on Geometry Processing* **30** (2011).
- [4] G. Dziuk, Finite elements for the Beltrami operator on arbitrary surfaces. in *Partial differential equations and calculus of variations*, S. Hildebrandt and R. Leis, eds., vol. 1357 of *Lecture Notes in Mathematics*, Springer, 1988, pp. 142-155.
- [5] Y. Saad. Iterative Methods for Sparse Linear Systems. *PWS Publishing Company, Boston*, 1996.
- [6] A. Ern, and J-L. Guermond, Theory and practice of finite elements. Vol. **159**. Springer Science & Business Media, 2013.
- [7] A. Ern, and J-L. Guermond, Evaluation for the condition number in linear system arising in finite element approximations, *ESAIM: Mathematical Modelling and Numerical Analysis*, Vol. **40**, issue 1, 2006
- [8] E. Burman, P. Hansbo, M. G. Larson, A. Massing, Cut Finite Element Methods for Partial Differential Equations on Embedded Manifolds of Arbitrary Codimensions <https://doi.org/10.48550/arXiv.1610.01660>

- [9] G. Dziuk and C. M. Elliott, Finite element methods for surface PDEs, *Acta Numerica* 2013, pp. 289-396
- [10] J. Lafontaine, An introduction to differentiable manifolds, Springer, 2015
- [11] E. Hebey, Nonlinear analysis on manifolds ; Sobolev spaces and inequalities, *Courant Lecture Notes* Vol.5 2000.
- [12] G. Allaire, Numerical analysis and optimization - an introduction to mathematical modeling and numerical simulation, *Oxford University Press*, 2007
- [13] M. A. Olshanskii, A. Reusken and J. Grande, A finite element method for elliptic equations on surfaces, *SIAM J. Numer. Anal.* **47**(2009) 3339-3358.
- [14] Ribes, Andre, and Christian Caremoli. *Salome platform component model for numerical simulation*. Computer Software and Applications Conference, 2007. COMPSAC 2007. 31st Annual International. Vol. 2. IEEE,2007.
- [15] S. Balay, S. Abhyankar, M. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, A. Dener, V. Eijkhout, W. Gropp, D. Karpeyev, D. Kaushik, M. Knepley, D. May, L. Curfman McInnes, R. Mills, T. Munson, K. Rupp, P. Sanan, B. Smith, S. Zampini, H. Zhang, and H. Zhang, PETSc/TAO Users Manual, Argonne National Laboratory, ANL-21/39 - Revision 3.17, 2021, <https://petsc.org/release/docs/manual/manual.pdf>.
- [16] PETSc Conjugate Gradient method, <https://petsc.org/release/docs/manualpages/KSP/KSPCG.html>
- [17] PETSc ILU preconditioner, <https://petsc.org/release/docs/manualpages/PC/PCILU.html>
- [18] PETSc linear solver, <https://petsc.org/release/docs/manualpages/KSP/KSPSolve.html>
- [19] PETSc matrix nullspace, <https://petsc.org/release/docs/manualpages/Mat/MatSetNullSpace.html>
- [20] <http://www.salome-platform.org/>
- [21] <https://github.com/ndjinga/SOLVERLAB>
- [22] A. Ribes, A. Bruneton, A. Geay, SALOME: an Open-Source simulation platform integrating ParaView, 10.13140/RG.2.2.12107.08485, 2017