Discrete maximum principles for nonlinear elliptic finite element problems on surfaces with boundary

JÁNOS KARÁTSON†,
Department of Applied Analysis & MTA-ELTE Numerical Analysis and Large Networks Research Group, ELTE University;
Department of Analysis, Technical University; Budapest, Hungary;
AND

BALÁZS KOVÁCS‡
Mathematisches Institut, Universität Tübingen,
Auf der Morgenstelle 10, 72076 Tübingen, Germany
AND

SERGEY KOROTOV§
Department of Computing, Mathematics and Physics,
Western Norway University of Applied Sciences, Bergen, Norway

[Received on October 25, 2018; revised on ]

The maximum principle forms an important qualitative property of second order elliptic equations, therefore its discrete analogues, the so-called discrete maximum principles (DMPs) have drawn much attention owing to their role in reinforcing the qualitative reliability of the given numerical scheme. In this paper DMPs are established for nonlinear finite element problems on surfaces with boundary, corresponding to the classical pointwise maximum principles on Riemannian manifolds in the spirit of Pucci et al. Various real-life examples illustrate the scope of the results.

Keywords: nonlinear elliptic problems on surfaces; maximum principles; discrete maximum principles; Riemannian manifolds; surface finite element method; simplicial mesh.

1. Introduction

The maximum principle forms an important qualitative property of second order elliptic equations Gilbarg & Trudinger (1983); Protter & Weinberger (1984); Pucci & Serrin (2007). We will refer to such results as continuous maximum principles (CMPs). Typical maximum principles often arise in a form stating that a solution attains a nonnegative maximum on the boundary, i.e.

\[ \max_S u \leq \max \{0, \max_{\partial S} u\}. \] (1.1)

CMPs have also been studied on Riemannian manifolds. Thereby the main subject is often an Omori–Yau type maximum principle, but in this paper we are interested in classical pointwise versions, such as the ones studied by Antonini, Mugnai and Pucci in Antonini et al. (2007), which give analogous results to (1.1).

The discrete analogues of CMPs, the so-called discrete maximum principles (DMPs) have drawn much attention. The DMP is in fact an important measure of the qualitative reliability of the numerical scheme, otherwise one could get unphysical numerical solutions like negative concentrations, etc.

We are interested for DMPs on a surface with boundary, in the context of the surface finite element method (SFEM). Originally, the surface FEM was developed by Dziuk (1988) for elliptic problems on closed surfaces (and further developed for evolving surfaces by Dziuk & Elliott (2007, 2013)). The theory was extended to elliptic problems on surfaces with boundaries in the preprint Burman et al. (2015). In such a case a DMP reproduces the above relation for the surface finite element solution \( u_h \) instead of \( u \). Various DMPs in Euclidean domains, including geometric conditions on the computational meshes for FEM solutions, have been given, e.g. by Ciarlet (1970); Ciarlet & Raviart (1973); Drăgănescu et al. (2005); Xu & Zikatanov (1999). The authors’ previous work,
e.g. Karátson & Korotov (2005); Korotov & Křížek (2001); Korotov et al. (2001), involves various types of linear and nonlinear equations and systems, and typical geometric conditions are nonobtuseness or acuteness in the case of simplicial meshes.

To our knowledge, no DMP has been established for (even linear) elliptic problems on surfaces yet. In this paper we provide such results in the spirit of Antonini et al. (2007) for nonlinear finite element problems on manifolds with boundary. We suitably adapt a treatment of the finite element matrix from our earlier paper on the flat domain case Karátson & Korotov (2005). We work in the mentioned surface finite element setting, cf. Burman et al. (2015); Dziuk (1988), and use the properties of the lift operator to achieve the arising stability requirements, thus various discrete maximum-minimum principles are established. We note that discrete maximum principles for parabolic reaction–diffusion equations on closed surfaces have been treated by Frittelli et al. (2017a, 2016, 2017b).

In our paper the issue of generating meshes with suitable angle properties is also addressed. The theoretical results of simplicial meshes.

2. Formulation and properties of the problem

Our formulation of the problem is based on Antonini et al. (2007), in which nonlinear elliptic problems are studied on manifolds. We follow Antonini et al. (2007) for the description of a general class of elliptic problems on the manifold, and show that a proper CMP can be derived. Then, in order to study numerical solutions of real-life problems of this class, we will consider surfaces embedded in Euclidean spaces and employ proper SFEM discretizations, following Burman et al. (2015); Dziuk (1988) for the description of the SFEM, respectively. These preliminaries will help us to reach our goal in the next section, i.e. to adapt our approach from Karátson & Korotov (2005) (developed for Euclidean domains) to verify DMPs for proper SFEM discretizations.

2.1 Nonlinear elliptic problems on Riemannian manifolds

Following the setting and notations of Antonini et al. (2007), let \( \mathcal{M} \) be a smooth complete \( d \)-dimensional Riemannian manifold, and let \( S \subset \mathcal{M} \) be a closed bounded and regular subdomain of \( \mathcal{M} \), so that \( \overline{S} \) is a smooth manifold with boundary. In the sequel we shall write \( TS \times_\mathcal{M} \mathbb{R} \) in place of \( TS \times S (S \times \mathbb{R}) \) to denote the fibered product bundle. In analogy with the Euclidean case, points of \( TS \times_\mathcal{M} \mathbb{R} \) will be denoted with \((x, z, \xi)\), where \((x, \xi) \in TS \) and \((x, z) \in S \times \mathbb{R} \). Integrals will be taken with respect to the natural Riemannian measure, denoted by \( d\mathcal{M} \).

The surface gradient, i.e. the usual Riemannian gradient, of a function \( u \) on \( S \) will be denoted by \( \nabla_S u \), that is, \( \nabla_S u = \sum_{i,j=1}^{d} g^{ij} \frac{\partial u}{\partial x^j} e_i \) where \( g^{ij} \) is the inverse of the first fundamental form, cf. Dziuk & Elliott (2013). Further, \( \Delta_S \) and \( \text{div}_S \) denote the Laplace–Beltrami operator and the surface divergence, respectively.

For \( p \geq 1 \), the Lebesgue and Sobolev norms are defined as

\[
\|u\|_{p,\mathcal{M}} := \left( \int_{\mathcal{M}} |u|^p \, d\mathcal{M} \right)^{1/p} \quad \text{and} \quad \|u\|_{1,p,\mathcal{M}} := \|u\|_{p,\mathcal{M}} + \|\nabla_S u\|_{p,\mathcal{M}},
\]

respectively, where \(|u| := g(u,u)^{1/2}\) using the tensor metric \(g\), further, \(W^{1,p}(\mathcal{M})\) and \(W_0^{1,p}(S)\) are defined as the closure of \(C^0(\mathcal{M})\) and \(C_0^0(S)\), respectively, in the Sobolev norm \(\| \cdot \|_{1,p,\mathcal{M}}\). We define \(p' := p/(p-1)\) if \(p > 1\) and \(p' := \infty\) if \(p = 1\) (see e.g. Dziuk (1988); Dziuk & Elliott (2013)).

There are similar Sobolev embedding estimates of \(W^{1,p}(S)\) available as in the Euclidean case. Let \(p^* := \frac{dp}{d-p}\) if \(d > p\) and \(p^* := \infty\) if \(d \leq p\). Here \(S\) is a closed bounded subdomain of the complete Riemannian \(d\)-manifold \(\mathcal{M}\), hence it is a compact Riemannian \(d\)-manifold itself. Therefore, by Aubin (1998), for all \(1 \leq p_1 < p^*\) we have \(W^{1,p}(S) \subset L^{p_1}(S)\) and

\[
\|v\|_{p_1,S} \leq k_1 \|v\|_{1,p,S} \quad (\forall v \in W^{1,p}(S))
\]

for some constant \(k_1 > 0\) depending on \(p, p_1\) and \(S\), but independent of \(v\).

Now let us consider a nonlinear boundary value problem of the following type, involving a scalar nonlinear coefficient in the principal part:

\[
\begin{cases}
-\text{div}_S \left(h(x,u,\nabla_S u)\nabla_S u\right) + q(x,u) = f(x) & \text{on} S, \\
u = g(x) & \text{on} \partial S,
\end{cases}
\]

with the bounded surface \(S \subset \mathcal{M}\), with boundary \(\partial S\), under the following conditions:

**Assumptions 2.1.**
(A1) Let \( p \geq 2 \) be a given parameter.

(A2) The scalar functions \( b : S \times \mathbb{R} \to \mathbb{R}^d, \quad q : S \times \mathbb{R} \to \mathbb{R} \) are continuous, further, \( g = g^1 |_{\partial S} \) for some \( g^* \in W^{1,\gamma}(S) \), where \( \gamma \geq p \) if \( p > d \) and \( \gamma > d \) if \( p \leq d \).

(A3) (Uniform ellipticity.) The function \( b \) satisfies
\[
\mu_0 + \mu_1 |\xi|^p - 2 \leq b(x, z, \xi) \leq M_0 + M_1 |\xi|^p - 2
\]
with constants \( \mu_0, \mu_1, M_0, M_1 > 0 \) independent of \((x, z, \xi) \in S \times \mathbb{R} \times \mathbb{R}^d \).

(A4) The function \( z \mapsto q(x, z) \) is nondecreasing for any fixed \( x \in S \). Further, let \( 2 \leq p_1 < +\infty \) if \( d \leq p \), or \( 2 \leq p_1 < \beta := \frac{2d}{p} \) if \( d > p \). Then there exist constants \( \alpha, \beta \geq 0 \) such that for any \( x \in S \) and \( z \in \mathbb{R} \) the following growth conditions hold:
\[
0 \leq q(x, 0) \equiv q(x, z) \quad (\forall z \leq 0), \quad q(x, z) - q(x, 0) \leq \alpha z + \beta |z|^p - 1 \quad (\forall z \geq 0).
\]

Further, there holds \( f \in L^{p_1}(S) \).

The weak solution \( u \in W^{1,p}(S) \) of problem (2.2) will be defined as follows:
\[
\int_S [b(x, u, \nabla_S u) \cdot \nabla_S v + q(x, u)v] \, d\mathcal{M} = \int_S f v \, d\mathcal{M} \quad (\forall v \in W^{1,p}_0(S)) \tag{2.3}
\]
and
\[
u - g^* \in W^{1,p}_0(S). \tag{2.4}
\]

**Remark 2.1** As mentioned before, related problems have been studied in the context of the DMP on usual Euclidean domains in our earlier paper Karátson & Korotov (2005), i.e. the above problem (2.2) is a surface analogue of the problems considered therein. Now in (2.2) we only consider Dirichlet boundary conditions, on the other hand, compared to Karátson & Korotov (2005), we impose more general growth conditions allowing \( p \neq 2 \) as well, which involves a more general setting using the Banach space \( W^{1,p}_0(S) \).

In the sequel we omit the sign \( d\mathcal{M} \) in the integrals for brevity, i.e. \( \int_S f := \int_S f \, d\mathcal{M} \).

### 2.2 Continuous maximum principles on the manifold

#### 2.2.1 Preliminaries

We will rely on Antonini et al. (2007), where CMPs are proved for a general class of elliptic inequalities on manifolds. We summarize the required background in a rewritten form involving elliptic equations with right-hand side of given sign.

Let us consider an elliptic equation
\[
- \text{div}_S K(x, u, \nabla_S u) + F(x, u, \nabla_S u) = k(x) \tag{2.5}
\]
on the domain \( S \subset \mathcal{M} \) under the following conditions:

**Assumptions 2.2.1.**

(i) The functions \( K : T S \times S \to T S \) and \( F : T S \times S \to \mathbb{R} \) are continuous, further, \( K(x, z, \xi) \in T_x \mathcal{M} \) for all \((x, z) \in S \times \mathbb{R}\) and \( \xi \in T_x \mathcal{M} \).

(ii) There exist constants \( a_1 > 0 \) and \( b_1, a_2 \geq 0 \) such that for all \((x, z, \xi) \in T S \times S \times \mathbb{R} \) there holds
\[
K(x, z, \xi) \cdot \xi \geq a_1 |\xi|^p - a_2 |z|^p, \quad F(x, z, \xi) \geq -b_1 |\xi|^p - 1.
\]

The treatment by Antonini et al. (2007) involves \( p \)-regular weak solutions of (2.5), defined as functions \( u \in L^{1,\infty}_0(S) \) satisfying \( K(\cdot, u, \nabla_S u) \in L^{p_1}_0(S, T S), \quad F(\cdot, u, \nabla_S u) \in L^{p_1}_0(S) \) and demanding (2.3) only for compactly supported \( v \in W^{1,p}_0(S) \). The following result holds:

**Theorem 2.1** Let Assumptions 2.2.1 hold with \( a_2 = 0 \) in item (ii). Let \( u \) be a \( p \)-regular weak solution of (2.5) such that \( u \in W^{1,p}_0(S) \), and let the right-hand side \( k \leq 0 \) on \( S \). If \( u \leq M \) on \( \partial S \) for some constant \( M \geq 0 \), then \( u \leq M \) in \( S \).

In fact, this result is the reformulation of (Antonini et al., 2007, Theorem 3.3.) for equations instead of inequalities and with rearranged signs. Furthermore, the \( p \)-regularity condition on \( F \) is only used to have \( F(\cdot, u, \nabla_S u) \in L^{1,\infty}_0(S) \) for all compactly supported \( v \in W^{1,p}_0(S) \). Owing to property (2.1), this can be also achieved by the modified assumption \( F(\cdot, u, \nabla_S u) \in L^{p_1'}_0(S) \) on \( F \), where \( p_1' := p_1/(p_1 - 1) \).
2.2.2 **CMP for the studied class.** We may now derive the CMP \( \tilde{u} \) in the vein of inequality \((1.1)\).

**THEOREM 2.2** Let Assumptions 2.1 hold. Assume that \( u \) is a weak solution of problem \((2.2)\), defined as in \((2.3)-(2.4)\), which satisfies \( u \in C(\bar{S}) \). Let
\[
f(x) - q(x,0) \leq 0 \quad (\text{a.e. } \forall x \in S).
\]

1. Then
\[
\max_{\bar{S}} u \leq \max \{0, \max_{\bar{S}} g\}.
\]
2. In particular, if \( g \geq 0 \) then
\[
\max_{\bar{S}} u = \max_{\bar{S}} g,
\]
and, if \( g \leq 0 \) then we have the nonpositivity property
\[
\max_{\bar{S}} u \leq 0.
\]
3. If we do not assume \( u \in C(\bar{S}) \) but only \( g \in L^\infty(S) \), then the above statements hold with replacing each \( \max \) by \( \esssup \), respectively.

**Proof.** It can be verified with direct calculations that problem \((2.2)\) under Assumptions 2.1 is a special case of problem \((2.5)\) under Assumptions 2.2.1. In particular, the estimates
\[
|K(\cdot, u, \nabla_S u)|^{p'} \leq M_3 |\nabla_S u|^{p'} + M_4 |\nabla_S u|^{(p-1)p'} \leq M_5 |\nabla_S u|^p,
\]
\[
|F(\cdot, u, \nabla_S u)|^{p'} \leq \alpha |u|^{p'} + \beta |u|^{(p-1)p'} \leq \gamma |u|^p,
\]
together with properties \( p' \geq 2 \leq p \) and \((2.1)\), show that \( u \) is \( p \)-regular. The remaining parts are left to the reader. \( \Box \)

In the special case \( q \equiv 0 \), the equality \((2.8)\) holds without assuming \( g \geq 0 \):

**THEOREM 2.3** Consider problem \((2.2)\) with \( q \equiv 0 \), under the assumptions of Theorem 2.2. That is, let \((A1)-(A3)\) hold, \( u \in C(\bar{S}) \), and assumption \((2.6)\) now takes the form \( f(x) \leq 0 \) \( (\forall x \in S) \). Then
\[
\max_{\bar{S}} u = \max_{\bar{S}} g.
\]

**Proof.** If \( \max_{\bar{S}} g \geq 0 \) then \((2.7)\) implies \((2.9)\). Let \( \max_{\bar{S}} g < 0 \), say, \( \max_{\bar{S}} g = -K \) with some \( K > 0 \). Then the function \( w := u + K \) satisfies the same type of problem with right-hand side \( f \) and b.c. \( g + K \), respectively, hence Theorem 2.2 is valid for this problem as well, and \((2.7)\) for \( w \) yields \( \max_{\bar{S}} w \leq \max \{0, \max_{\bar{S}} (g + K)\} = 0 \). Then
\[
\max_{\bar{S}} u \leq -K = \max_{\bar{S}} g.
\]

**REMARK 2.2** The corresponding minimum principles and nonnegativity property hold if the sign conditions on \( f \) are reversed.

2.3 **Computational scheme: the surface finite element method (SFEM)**

In order to study numerical solutions of real-life problems of the type \((2.2)\), we consider surfaces in Euclidean spaces. That is, from now on, we assume that the manifold \( \mathcal{M} \) of the previous subsections can be embedded in \( \mathbb{R}^{d+1} \), moreover, it admits an associated signed distance function \( \delta \), satisfying \( \bar{S} = \{ x \in \mathbb{R}^{d+1} \mid \delta(x) = 0 \} \) and \( \nabla \delta = v \). This enables us to use the SFEM setting of Dziuk (1988); Dziuk & Elliott (2013). The surface gradient from Section 2.1 coincides with the tangential gradient, see Dziuk & Elliott (2013). We note that this embedding assumption is not a very strong restriction in the practically most relevant case \( d = 2 \): as described e.g. by Han & Hong (2006), a wide class of two-dimensional manifolds can be embedded in \( \mathbb{R}^3 \) (globally and isometrically) under some proper conditions on their smoothness and Gauss curvature.

The smooth surface \( S \) with boundary \( \partial S \) is approximated by a triangulated one, denoted by \( S_h \), whose vertices \( B_1, B_2, \ldots, B_n \) lie on the surface, including boundary nodes of \( S_h \) lying on \( \partial S \), and it is given as
\[
S_h = \bigcup_{T \in T_h} T.
\]

We always assume that the simplices \( T \) form an admissible triangulation (i.e. partition) \( T_h \), with \( h \) denoting the maximum diameter. Admissible triangulations were introduced by (Dziuk & Elliott, 2007, Section 5.1): \( S_h \) is a
uniform triangulation, i.e. every \( T \in \mathcal{T}_h \) satisfies that the inner radius \( \sigma_h \) is bounded from below by \( ch \) with \( c > 0 \), and \( S_h \) is not a global double covering of \( S \). The discrete tangential gradient on the discrete surface \( S_h \) is given by

\[
\nabla_{S_h} \phi = \nabla \hat{\phi} - (\nabla \hat{\phi} \cdot v_h) v_h,
\]

understood in a piecewise sense, with \( v_h \) denoting the normal to \( S_h \), cf. Dziuk (1988). The boundary of the approximation surface is simply the collection of edges that connect two boundary nodes, and is denoted by \( \partial S_h \), which is an approximation of \( \partial S \).

We define the surface finite element discretisation of our problem, following Dziuk (1988) and Burman et al. (2015). We use simplicial elements and continuous piecewise linear functions. The finite element subspace \( V_h \) is spanned by the continuous, piecewise linear basis functions \( \chi_j \), satisfying \( \chi_j(B_i) = \delta_{ij} \) for all \( i, j = 1, 2, \ldots, \hat{n} \), therefore

\[
V_h = \text{span} \{ \chi_1, \chi_2, \ldots, \chi_{\hat{n}} \}.
\]

Now, let \( n < \hat{n} \) be such that

\[
B_1, B_2, \ldots, B_n \quad \text{and} \quad B_{n+1}, \ldots, B_{\hat{n}}
\]

are the vertices that lie on \( S_h \) but not on \( \partial S_h \), and that lie only on \( \partial S_h \), respectively. Then the basis functions \( \chi_1, \chi_2, \ldots, \chi_{\hat{n}} \) satisfy the homogeneous Dirichlet boundary condition on \( \partial S_h \). Hence, we can define

\[
V^0_h := \text{span} \{ \chi_1, \chi_2, \ldots, \chi_n \} \subset W^{1,p}_0(S_h).
\]

Further, let

\[
g_h = \sum_{j=n+1}^{\hat{n}} g_j \chi_j \in V_h
\]

(with \( g_j \in \mathbb{R} \)) be the piecewise linear approximation of the function \( g \) on \( \partial S_h \) (and on the neighbouring elements).

**2.3.1 The lift operator.** In the following we recall the so-called lift operator, which was introduced by Dziuk (1988) and further investigated by Dziuk & Elliott (2007). The lift operator projects a finite element function on the discrete surface onto a function on the smooth surface.

Using the oriented distance function \( \delta \) (cf. (Dziuk & Elliott, 2007, Section 2.1)), for a continuous function \( \eta_h : S_h \rightarrow \mathbb{R} \) its lift is defined as

\[
\eta^l_h(P) := \eta_h(x), \quad x \in S_h,
\]

where for every \( x \in S_h \) the value \( P = P(x) \in S \) is uniquely defined via

\[
x = P + \nu(P)\delta(x).
\]

By \( \eta^{-l} \) we mean the function whose lift is \( \eta \). Further, we have the lifted finite element spaces

\[
V^l_h := \{ \phi_h = \phi^l_h \mid \phi_h \in V_h \}, \quad (V^l_h)^f := \{ \phi_h = \phi^l_h \mid \phi_h \in V^0_h \}.
\]

The following equivalence between discrete and continuous norms and semi-norms holds, shown by Dziuk (1988); Demlow (2009), for functions \( \eta_h : S_h \rightarrow \mathbb{R} \) with lift \( \eta^l_h : S \rightarrow \mathbb{R} \). This will help to ensure stability of the discrete problem using the surface approximation.

**Lemma 2.1** (equivalence of norms, Dziuk (1988); Demlow (2009) ) The original and lifted norms are equivalent in both the \( L^p \) and \( W^{1,p} \) case for \( 1 \leq p \leq \infty \), independently of sufficiently small mesh sizes \( h \). That is, there exist constants \( c, h_0 > 0 \) such that for all \( 0 < h \leq h_0 \), for all \( \eta_h \in L^p(S_h) \) with \( \nabla S_h \eta_h \in L^p(S_h) \),

\[
\begin{align*}
c^{-1} \| \eta_h \|_{L^p(S_h)} & \leq \| \eta^l_h \|_{L^p(S)} \leq c \| \eta_h \|_{L^p(S_h)}, \\
c^{-1} \| \nabla S_h \eta_h \|_{L^p(S_h)} & \leq \| \nabla \eta^l_h \|_{L^p(S)} \leq c \| \nabla S_h \eta_h \|_{L^p(S_h)}, \\
c^{-1} \| \eta_h \|_{W^{1,p}(S_h)} & \leq \| \eta^l_h \|_{W^{1,p}(S)} \leq c \| \eta_h \|_{W^{1,p}(S_h)}.
\end{align*}
\]

From the Sobolev embedding (2.1) (for surfaces with boundary) its discrete analogue directly follows, via the norm equivalence from above.

**Lemma 2.2** (Sobolev embedding) Let \( 1 \leq p_1 < p^* \). Then there exists constants \( \hat{k}_1 > 0 \) and \( h_0 > 0 \), such that for all \( 0 < h \leq h_0 \),

\[
\| \nu \|_{p_1, S_h} \leq \hat{k}_1 \| \nu \|_{1,p,S_h} \quad (\forall \nu \in W^{1,p}(S_h)).
\]
2.3.2 The discrete problem. In order to find the approximate solution of (2.2), we have to solve the counterpart of (2.3)–(2.4) in $V_h$. We note here that in order to transfer the assumptions on the coefficient functions (A1)–(A4) to the discrete level we have to use negative lifts in the spatial variable. For any $x \in S_h$ with corresponding lift $P \in S$, cf. (2.11), we set

$$b^{-1}(x, \cdot, \cdot) = b(P, \cdot, \cdot),$$

and similarly using the negative lift in the first argument of all coefficient functions ($f$, $q$, $r$, etc.), as well. No lift is used in the other arguments. By this construction all the properties of (A1)–(A4) are clearly transferred to these lifted functions, with $S_h$ instead of $S$. Now first subtract $q^{-1}(x, 0)$ from both sides of the original problem, that is, we have

$$\begin{cases}
- \text{div}_S \left( b(x, u, \nabla_S u) \nabla_S u \right) + \hat{q}(x, u) = \hat{f}(x) & \text{on } S, \\
u = g(x) & \text{on } \partial S
\end{cases}
$$

(2.13)

where $\hat{q}(x, z) := q(x, z) - q(x, 0)$, $\hat{f}(x) := f(x) - q(x, 0)$ for all $x \in S$, $z \in \mathbb{R}$. We consider the corresponding discrete problem: find $u_h \in V_h$ such that

$$\int_{S_h} \left( b^{-1}(x, u_h, \nabla_S u_h) \nabla_S u_h \cdot \nabla_S v_h + q^{-1}(x, u_h)v_h \right) = \int_{S_h} \hat{f}^{-1}v_h \quad (\forall v_h \in V^0_h),
$$

(2.14)

and

$$u_h - g_h \in V^0_h
$$

(2.15)

where

$$\hat{q}^{-1}(x, z) := q^{-1}(x, z) - q^{-1}(x, 0), \quad \hat{f}^{-1}(x) := f^{-1}(x) - q^{-1}(x, 0)
$$

(2.16)

for all $x \in S_h$, $z \in \mathbb{R}$. Further, denote

$$r^{-1}(x, z) := \begin{cases}
\frac{q^{-1}(x, z) - q^{-1}(x, 0)}{z}, & \text{if } z > 0, \\
0, & \text{if } z \leq 0.
\end{cases}
$$

Then, using also assumption (A4),

$$\hat{q}^{-1}(x, z) = r^{-1}(x, z)z, \quad 0 \leq r^{-1}(x, z) \leq \alpha + \beta |z|^{p-2} \quad (\forall x \in S_h, z \in \mathbb{R}).
$$

(2.17)

In what follows, we will rewrite problem (2.14) as

$$\int_{S_h} \left( b^{-1}(x, u_h, \nabla_S u_h) \nabla_S u_h \cdot \nabla_S v_h + r^{-1}(x, u_h)u_h v_h \right) = \int_{S_h} \hat{f}^{-1}v_h \quad (\forall v_h \in V^0_h).
$$

(2.18)

We set

$$u_h = \sum_{j=1}^n c_j \chi_j,
$$

(2.19)

and look for the coefficients $c_1, c_2, \ldots, c_n$.

**Remark 2.3** (Existence and uniqueness.) The goal of this paper is to establish a qualitative property of any FEM solution that the discrete problem (2.14) may have, hence we only give some comments on the existence and uniqueness issue. This can be ensured under some additional assumptions, which are, however, not very restrictive from practical aspect, and cover, for instance, p-Laplacian type operators or nonlinear heat conduction problems.

(i) Existence and uniqueness holds for a general class of problems that can be reduced to monotone operators, see Faragó & Karátson (2002), Zeidler (1986). For our problem (2.2), assume that the principal part is uniformly monotone and locally Lipschitz continuous w.r.t. $\xi$, i.e.,

$$b(x, z_1, \xi_1)\xi_1 - b(x, z_2, \xi_2)\xi_2 \leq (c_1 + c_2 \max \{ \xi_1, |\xi_2| \})^p \xi_1 - \xi_2 |
$$

(2.20)

$$|b(x, z_1, \xi_1)\xi_1 - b(x, z_2, \xi_2)\xi_2| \leq (c_1 + c_2 \max \{ \xi_1, |\xi_2| \})^{p-2} \xi_1 - \xi_2 |
$$

(2.21)

with some constants $c_0, c_1, c_2 > 0$ independent of the arbitrary arguments. Let us first consider homogeneous boundary conditions. The problem (2.14) can be written as an equation

$$\langle F(u_h), v_h \rangle = \langle b, v_h \rangle \quad (\forall v_h \in V^0_h)$$
for some operator $F : V^0_h \to (V^0_h)'$. Using Hölder inequalities, it follows in a standard way that $F$ inherits uniform monotonicity and local Lipschitz continuity, which implies existence and uniqueness. Further, the inhomogeneous problem can be reduced to the above using a standard translation with the Dirichlet lift.

(ii) For non-monotone problems on Euclidean domains, one can prove existence if the coefficient $b$ is properly bounded, and also uniqueness for heat conduction type coefficients, see Douglas et al. (1971), Krejčí & Neittaanmäki (1996), Pospíšek (1994) for details. These ideas might be also adapted in a straightforward way to surfaces using the analogous Sobolev space theory.

2.3.3 The nonlinear algebraic system. Now we turn to the nonlinear algebraic system corresponding to (2.14). For any $\tilde{c} = (c_1, \ldots, c_{\tilde{n}}) \in \mathbb{R}^{\tilde{n}}$, $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, \tilde{n}$, we set

$$b_{ij}(\tilde{c}) = \int_{S_h} b^{-1} \left( x, \sum_{k=1}^{\tilde{n}} c_k \nabla h \cdot \nabla \chi_{k} \right) \nabla h \cdot \nabla \chi_{j}, \quad r_{ij}(\tilde{c}) = \int_{S_h} r^{-1} \left( x, \sum_{k=1}^{\tilde{n}} c_k \nabla h \right) \chi_{j},$$

$$a_{ij}(\tilde{c}) = b_{ij}(\tilde{c}) + r_{ij}(\tilde{c}), \quad d_i = \int_{S_h} \hat{f}^{-1} \chi_{i}.$$

Putting (2.19) and $v_h = \chi_i$ into (2.18), we obtain the $n \times \tilde{n}$ system of algebraic equations

$$\sum_{j=1}^{\tilde{n}} a_{ij}(\tilde{c}) c_j = d_i, \quad i = 1, 2, \ldots, n. \tag{2.22}$$

Using the notations

$$\mathbf{A}(\tilde{c}) = \{a_{ij}(\tilde{c})\}, \quad \text{and} \quad \mathbf{d} = \{d_j\}, \quad \mathbf{c} = \{c_j\}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, \tilde{n},$$

$$\tilde{\mathbf{A}}(\tilde{c}) = \{a_{ij}(\tilde{c})\}, \quad \text{and} \quad \tilde{\mathbf{c}} = \{c_j\}, \quad \tilde{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix}, \quad i = 1, 2, \ldots, n, \quad j = n + 1, \ldots, \tilde{n},$$

the system (2.22) turns into

$$\mathbf{A}(\tilde{c}) \mathbf{c} + \tilde{\mathbf{A}}(\tilde{c}) \tilde{\mathbf{c}} = \mathbf{d}. \tag{2.23}$$

Defining further

$$\tilde{\mathbf{A}}(\tilde{c}) = \begin{bmatrix} \mathbf{A}(\tilde{c}) & \tilde{\mathbf{A}}(\tilde{c}) \end{bmatrix}, \quad \tilde{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix}, \quad \tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{d}} \end{bmatrix}, \tag{2.24}$$

we rewrite (2.23) as follows

$$\tilde{\mathbf{A}}(\tilde{c}) \tilde{\mathbf{c}} = \tilde{\mathbf{d}}.$$

In order to obtain a system with a square matrix, we enlarge our system to an $\tilde{n} \times \tilde{n}$ one. Namely, since $u_h = g_h$ on $\partial S_h$, the coordinates $c_i$ with $n + 1 \leq i \leq \tilde{n}$ satisfy automatically $c_i = g_i$, i.e.

$$\tilde{\mathbf{c}} = \tilde{\mathbf{g}}, \quad \text{where} \quad \tilde{\mathbf{g}} = \{g_j\}, \quad j = n + 1, \ldots, \tilde{n}.$$

That is, we can replace (2.23) by the equivalent system

$$\begin{bmatrix} \mathbf{A}(\tilde{c}) & \tilde{\mathbf{A}}(\tilde{c}) \\ \mathbf{0} & \mathbf{1} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{d}} \end{bmatrix}.$$

REMARK 2.4 The solution of the arising nonlinear finite element problems can rely on various efficient methods, see e.g. Faragó & Karátson (2002).

3. Discrete maximum principles

3.1 Classical matrix maximum principles

Let us consider a linear algebraic system of equations of order $(n+m) \times (n+m)$:

$$\tilde{\mathbf{A}} \tilde{\mathbf{c}} = \tilde{\mathbf{b}},$$

where the matrix $\tilde{\mathbf{A}}$ has the following structure:

$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \tilde{\mathbf{A}} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}. \tag{3.1}$$
In the above, $I$ is an $m \times m$ identity matrix, $0$ is a $m \times n$ zero matrix. In a (surface) finite element setting, such a partitioning arises corresponding to interior and boundary points.

We first recall some classical definitions and results, see, e.g., Ciarlet (1970); Varga (1962). We follow the terminology of Faragó (2010). Throughout, inequalities for matrices or vectors are understood elementwise, and $c$, $\bar{c}$ and $\tilde{c}$ denote vectors consisting of $n$, $m$ or $n + m$ numbers, respectively.

**Definition 3.1** The matrix $\bar{A}$ in (3.1) satisfies

(a) the discrete weak maximum principle (DwMP) if for an arbitrary vector $\bar{c} = (c_1, \ldots, c_{n+m})^T \in \mathbb{R}^{n+m}$ satisfying $(\bar{A}\bar{c})_i \leq 0$, $i = 1, 2, \ldots, n$, one has

$$\max_{i=1,2,\ldots,n+m} c_i \leq \max_{i=n+1,\ldots,n+m} c_i;$$

(b) the discrete strict weak maximum principle (DWMP) if for any vector $\bar{c} = (c_1, \ldots, c_{n+m})^T \in \mathbb{R}^{n+m}$ satisfying $(\bar{A}\bar{c})_i \leq 0$, $i = 1, 2, \ldots, n$, one has

$$\max_{i=1,2,\ldots,n+m} c_i = \max_{i=n+1,\ldots,n+m} c_i.$$

(DMPs without the term 'weak' also assert that only constant vectors may attain a maximum for 'interior' indices, but we do not address this property here.)

Based on Ciarlet & Raviart (1973); Faragó (2010) we have the following result, formulated in this way by Karátson & Korotov (2015):

**Theorem 3.2** If the matrix $\bar{A}$ in (3.1) satisfies the following conditions:

(i) $a_{ij} \leq 0$ \hspace{1em} ($\forall i = 1, \ldots, n; \hspace{0.5em} j = 1, \ldots, n + m; \hspace{0.5em} i \neq j)$,

(ii) $\sum_{j=1}^{n+m} a_{ij} \geq 0$ \hspace{1em} ($\forall i = 1, \ldots, n$),

(iii) $A$ is positive definite,

then $\bar{A}$ possesses the DwMP.

If the inequality in condition (ii) is replaced by equality, then $\bar{A}$ possesses the DWMP.

### 3.2 The discrete maximum principle for the nonlinear elliptic problem

#### 3.2.1 The main results

The derivation of our main results needs the following lemmas on the stability of the boundary data and then on the norm-boundedness of the SFEM solutions. This means that we can achieve stability of the discrete problem, thanks to exploiting the lift properties of subsection 2.3.1 to control the geometric error due to the surface approximation. For the seminorms we will use the shorthand notation

$$|v_h|_{1,p,S_h} := \|\nabla S_h v_h\|_{p,S_h}.$$

**Lemma 3.1** Let $g^*$ and $\gamma$ be as given in Assumption 2.1 (A2). Then there exists a constant $c > 0$ such that for all $h \leq h_0$ with a $h_0 > 0$ sufficiently small,

(a) the discrete function $g_h$, from (2.10), satisfies

$$\|g_h\|_{1,p,S_h} \leq c\|g^*\|_{1,\gamma,S};$$

(b) the FEM solution $u_h$ satisfies

$$\|u_h\|_{1,p,S_h} \leq c\left(\|u_h\|_{1,p,S_h} + \|g^*\|_{1,\gamma,S}\right).$$

**Proof.** (a) The assumption on $\gamma$ in (A2) implies that in each case we have $\gamma \geq p$ and $\gamma > d$. Since the compact manifold $S_h$ has bounded measure, the relation $\gamma \geq p$ implies that

$$\|g_h\|_{1,p,S_h} \leq c_1\|g_h\|_{1,\gamma,S_h}$$

for some constant $c_1 > 0$ independent of $h$. Further, the relation $\gamma > d$ implies that $W^{1,\gamma}(S_h) \subset C(S_h)$, see, e.g., Adams (1975). Hence, by (Ciarlet, 1978, Thm. 3.1.6), the stability of the Lagrange interpolation holds for $(g^*)^{-1}$ in $W^{1,\gamma}$-norm on each triangle and hence also on all $S_h$, therefore

$$\|g_h\|_{1,\gamma,S_h} \leq c_2\|(g^*)^{-1}\|_{1,\gamma,S_h}.$$
for some constant $c_2 > 0$ independent of $h$. Finally, Lemma 2.1 implies that

$$\| (g^*)^{-l} \|_{1, \gamma, S} \leq c_3 \| g^* \|_{1, \gamma, S}$$

for some constant $c_3 > 0$ independent of $h$. These three estimates together yield the desired result.

(b) The $W^{1, p}$-norms and seminorms are equivalent on $W_0^{1, p}(S_h)$, with constants independent of $h$, via Poincaré's inequality on the smooth surface $S$ (cf. Dziuk (1988)) and by the equivalence of norms (Lemma 2.1). In particular,

$$\| v_h \|_{1, p, S_h} \leq c_1 \| v_h \|_{1, p, S_h} \quad (\forall v_h \in W_0^{1, p}(S_h)).$$

(3.4)

Now denote $v_h := u_h - g_h \in V_h^0$. Together with (3.2), we have

$$\| u_h \|_{1, p, S_h} \leq \| v_h \|_{1, p, S_h} + \| g_h \|_{1, p, S_h} \leq c_1 \| v_h \|_{1, p, S_h} + \| g_h \|_{1, p, S_h} \leq c_1 \| u_h \|_{1, p, S_h} + c_2 \| g_h \|_{1, p, S_h} \leq c_3 \| u_h \|_{1, p, S_h} + \| g^* \|_{1, \gamma, S}.$$  

□

**Lemma 3.2** The norms $\| u_h \|_{1, p, S_h}$ are bounded independently of $h$.

**Proof.** We have $v_h := u_h - g_h \in V_h^0$ by (2.15), hence we can use it as a test function in the discrete weak problem (2.14). Rearranging the terms, we obtain

$$\int_{S_h} b^{*-l}(x, u_h, \nabla_S u_h) \nabla_S u_h \, dx = \int_{S_h} b^{*-l}(x, u_h, \nabla_S u_h) (\nabla_S g_h + \hat{q}^{-l}(u_h - g_h)) \, dx.$$  

(3.5)

We wish to estimate this equality appropriately from both above and below. First, let us subtract $\hat{q}^{-l}(x, g_h)(u_h - g_h)$ from the integrands on both sides. On the left, the second term becomes

$$(\hat{q}^{-l}(u_h) - \hat{q}^{-l}(x, g_h))(u_h - g_h) \geq 0,$$

since $\hat{q}$ and thus $\hat{q}^{-l}$ is nondecreasing w.r.t. the last variable. (Recall that $\hat{q}$ was defined in (2.16).) Hence the left-hand side can be estimated below by

$$\int_{S_h} b^{*-l}(x, u_h, \nabla_S u_h) \nabla_S u_h \, dx \geq \mu_1 \int_{S_h} |\nabla_S u_h|^2 = \mu_1 \| u_h \|^p_{1, p, S_h},$$  

(3.6)

where we have used that assumption (A3) yields $b^{*-l}(x, \cdot, \xi) |\xi|^2 \geq \mu_1 |\xi|^p (x \in S_h)$. For the first term on the right-hand side, we have from (A3) that

$$\int_{S_h} b^{*-l}(x, u_h, \nabla_S u_h) \nabla_S u_h \cdot \nabla_S g_h \leq \int_{S_h} (M_0 |\nabla_S u_h| + M_1 \| u_h \|^{p-1}_{1, p, S_h} |\nabla_S g_h|) \leq M_0 \| u_h \|_{1, p', S_h} \| g_h \|_{1, p, S_h} + M_1 \| u_h \|^{p-1}_{1, p, S_h} \| g_h \|_{1, p, S_h} \leq c(M_0 \| u_h \|_{1, p, S_h} + M_1 \| u_h \|^{p-1}_{1, p, S_h}) \| g_h \|_{1, p, S_h},$$

using Hölder's inequality with $\frac{1}{p} + \frac{1}{p'} = 1$, where $p' \leq 2 \leq p$, and owing to the boundedness of the surface. The second term of the integrand on the right-hand side becomes $(f^{*-l} - \hat{q}^{-l}(x, g_h))(u_h - g_h)$ after the mentioned subtraction. Here assumption (A4) implies that

$$f^{*-l} \in L^{p'_1}(S_h), \quad |\hat{q}^{-l}(x, g_h)| \leq \alpha |g_h| + \beta |g_h|^{p-1}.$$  

Using proper Hölder's inequalities, we obtain that

$$\int_{S_h} (f^{*-l} - \hat{q}^{-l}(x, g_h))(u_h - g_h) \leq \left( \| f^{*-l} \|_{p'_1, S_h} + \alpha \| g_h \|_{p'_1, S_h} + \beta \| g_h \|^{p-1}_{p'_1, S_h} \right) \left( \| u_h \|_{p, S_h} + \| g_h \|_{p, S_h} \right).$$  

(3.7)

Here we can use the lift property $\| f^{*-l} \|_{p'_1, S_h} \leq c \| f \|_{p'_1, S}$, further, the relation $p'_1 \leq 2 \leq p_1$ implies $\| g_h \|_{p'_1, S_h} \leq c \| g_h \|_{p_1, S_h}$. In addition, using Lemma 2.2, $\| g_h \|_{p_1, S_h}$ and $\| u_h \|_{p_1, S_h}$ can be further replaced by $c \| g_h \|_{1, p, S_h}$ and $c \| u_h \|_{1, p, S_h}$, respectively. Let us substitute the obtained upper estimates in (3.7), and also replace $\| g_h \|_{1, p, S_h}$ by $c \| g^* \|_{1, \gamma, S}$ due to Lemma 3.1 (a): thus the r.h.s. of (3.7) is bounded above by

$$c \left( \| f \|_{p'_1, S} + \| g^* \|_{1, \gamma, S} + \| g^* \|^{p-1}_{1, \gamma, S} \right) \left( \| u_h \|_{1, p, S_h} + \| g^* \|_{1, \gamma, S} \right).$$
Here, owing to (3.3), \( \|u_h\|_{1,p,S_h} \) can be replaced by the seminorm \( |u_h|_{1,p,S_h} \) up to a constant multiple in the bracket. Altogether, combining the obtained lower and upper bounds on the l.h.s. and r.h.s. of (3.5), respectively, we have

\[
\mu_1 |u_h|_{1,p,S_h} \leq C \left( |u_h|_{1,p,S_h} + |u_h|_{1,p,S_h}^{-1} \right) + C \left( \left\| f \right\|_{1,S} + \left\| g^s \right\|_{1,S} + \left\| g^s \right\|_{1,S}^{1-1} \right) \left( \left| u_h \right|_{1,p,S_h} + \left| g^s \right|_{1,S} \right)
\]

for some constant \( C > 0 \). Thus we obtain a "polynomial" of degree \( p - 1 \) of \( |u_h|_{1,p,S_h} \) on the right-hand side, and a power \( p \) of \( |u_h|_{1,p,S_h} \) on the left-hand side. Since the latter is bounded by the former, this implies that \( |u_h|_{1,p,S_h} \) is bounded. Finally, owing to Lemma 3.1 (b), we obtain that \( \|u_h\|_{1,p,S_h} \) is also bounded independently of \( h \).

Now we can verify the underlying matrix maximum principle for the stiffness matrix of our elliptic problem.

**Theorem 3.3** Let (A1)–(A4) hold and let us consider a family of simplicial triangulations \( \mathcal{T}_h \) \( (h > 0) \) satisfying the following property: for any \( i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, \bar{n} \quad (i \neq j) \)

\[
\nabla S_h \chi_i \cdot \nabla S_h \chi_j \leq -\frac{\sigma_0}{h^2} < 0
\]

on \( \text{supp} \chi_i \cap \text{supp} \chi_j \subset S_h \) with \( \sigma_0 > 0 \) independent of \( i, j \) and \( h \). Let the simplicial triangulations \( \mathcal{T}_h \) be regular, i.e. there exist constants \( m_1, m_2 > 0 \) such that for any \( h > 0 \) and any simplex \( T_h \in \mathcal{T}_h \)

\[
m_1 h^d \leq \text{meas}(T_h) \leq m_2 h^d
\]

(where \( \text{meas}(T_h) \) denotes the \( d \)-dimensional measure of \( T_h \)).

Then for sufficiently small \( h \), the matrix \( \tilde{A}(\tilde{\varepsilon}) \) defined in (2.24) has the following properties:

(i) \( a_{ij}(\tilde{\varepsilon}) \leq 0, \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, \bar{n} \) such that \( i \neq j \).

(ii) \( \sum_{j=1}^{\bar{n}} a_{ij}(\tilde{\varepsilon}) \geq 0, \quad i = 1, 2, \ldots, n \).

(iii) \( \tilde{A}(\tilde{\varepsilon}) \) is positive definite.

**Proof.** Let us recall that for any \( i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, \bar{n} \),

\[
a_{ij}(\tilde{\varepsilon}) = \int_{S_h} \left( b^{-1}(x,u_h) \nabla S_h \chi_i \cdot \nabla S_h \chi_j + r^{-1}(x,u_h) \chi_i \chi_j \right).
\]

Now we prove properties (i)–(iii).

(i) Let \( i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, \bar{n} \) with \( i \neq j \) and let \( S_{ij}^j \) denote the interior of \( \text{supp} \chi_i \cap \text{supp} \chi_j \subset S_h \). If \( S_{ij}^j = \emptyset \) then

\[
a_{ij}(\tilde{\varepsilon}) = 0.
\]

If \( S_{ij}^j \neq \emptyset \) then we use property (A3), namely that \( b(x,u_h,\nabla S_h u_h) \geq \mu_0 \), and (3.8), further, \( r(x,u_h) \geq 0 \) and the fact \( 0 \leq \chi \leq 1 \quad (i = 1, 2, \ldots, \bar{n}) \), which imply

\[
a_{ij}(\tilde{\varepsilon}) \leq -\frac{\sigma_0}{h^2} u_0 \text{meas}(S_{ij}^j) + \int_{S_{ij}^j} r^{-1}(x,u_h).
\]

(3.10)

Here, from (2.17),

\[
\int_{S_{ij}^j} r^{-1}(x,u_h) \leq \int_{S_{ij}^j} \left( \alpha + \beta |u_h|^{p_{ij} - 2} \right).
\]

We now further estimate the integral from above. From assumption (A4) we have \( 2 \leq p_1 < \hat{p} \), where \( \hat{p} = \frac{2d}{2d - p} \) (if \( d > p \)) and \( \hat{p} := +\infty \) (if \( d \leq p \)). Let us also consider the Sobolev embedding estimate (2.1), where \( p^* := \frac{pd}{d-p} \) (if \( d > p \)) and \( p^* := +\infty \) (if \( d \leq p \)) were defined. Since \( 2 \leq p \) from (A1), we clearly have

\[
\hat{p} \leq p^*.
\]

(3.11)

Hence \( p_1 \) also satisfies \( p_1 < p^* \), which implies that the Sobolev embedding estimate (2.1) holds with this \( p_1 \). Now, assume first that \( p_1 > 2 \) and let us fix a real number \( r \) satisfying

\[
\frac{d}{2} < r \leq \frac{p^*}{p_1 - 2}.
\]

(3.12)
Such a number \( r \) always exists, since for \( d \leq p \) there is no upper bound \( (p^* := \infty) \), and for \( d > p \) the assumption \( p_1 < \hat{p} \) implies

\[
p_1 - 2 < \hat{p} - 2 = \frac{2p}{d-p} = \frac{2p^*}{d}, \quad \text{which yields} \quad \frac{d}{2} < \frac{p^*}{p_1-2}.
\]

Further, the assumption \( d \geq 2 \) on the dimension of the manifold implies \( r > 1 \). (In theory the trivial case \( d = 1 \) is also allowed, then we must assume that \( r > 1 \) is chosen, which is allowed since \( \frac{p^*}{p_1-2} > 1 \).)

Now let \( s > 1 \) be chosen such that \( \frac{1}{r} + \frac{1}{s} = 1 \). Then Hölder’s inequality implies

\[
\int_{S_h^i} |u_h|^{p_1-2} \leq \left[ \|u_h\|_{L^1(S_h^i)} \right]^{p_1-2} \left[ \|u_h\|_{L^r(S_h^i)} \right] = \text{meas}(S_h^i)^{1/s} \|u_h\|^{|p_1-2|} \left[ \|u_h\|_{L^r(S_h^i)} \right].
\]

(3.13)

Here \((p_1 - 2)r \leq p^* \) and (2.1) imply \( \|u_h\|_{L^r(S_h^i)} \leq k_1 \|u_h\|_{L^r(S_h^i)} \). Here, by Lemma 3.2, \( \|u_h\|_{1,p,S_h} \) is bounded independently of \( h \). Hence, (3.13) is bounded as

\[
\int_{S_h^i} |u_h|^{p_1-2} \leq K_1 \text{meas}(S_h^i)^{1/s}.
\]

(3.14)

with some constant \( K_1 > 0 \) independent of \( h \). Finally, if \( p_1 = 2 \) then the corresponding equality (3.14) holds trivially with \( s = 1 \). In addition, the integral of the positive constant \( \alpha \) is simply \( \int_{S_h^i} \alpha = \alpha \text{meas}(S_h^i) \). Substituting the two estimates into (3.10), we obtain

\[
a_{ij}(\bar{e}) \leq \left( \frac{C_0}{h^2} + \alpha \right) \text{meas}(S_h^i) + \beta K_1 \text{meas}(S_h^i)^{1/s}.
\]

For sufficiently small \( h \), we can write

\[
a_{ij}(\bar{e}) \leq A^{ij}(h) := -\frac{C_0}{h^2} \text{meas}(S_h^i) + C_1 \text{meas}(S_h^i)^{1/s},
\]

more precisely, there exist positive constants \( h_0, C_0 \) and \( C_1 \) independently of \( h \) and \( i, j \) such that the above inequality holds for \( h < h_0 \). Then, using that \( \frac{1}{r} + \frac{1}{s} = 1 \) and the regularity (3.9) of the mesh, we have

\[
A^{ij}(h) = \text{meas}(S_h^i)^{1/s} \left( -\frac{C_0}{h^2} \text{meas}(S_h^i)^{1/r} + C_1 \right) \leq \text{meas}(S_h^i)^{1/s} \left( -C_2 h^{-2r(d/r)} + C_1 \right).
\]

Since (3.12) implies \( \frac{d}{r} < 2 \), the term in brackets tends to \( -\infty \) as \( h \to 0 \) and hence \( A^{ij}(h) < 0 \) for \( h < h_0 \).

Altogether, we obtain that there exists \( h_0 > 0 \) such that for \( h \leq h_0 \) and all \( i, j \), \( a_{ij}(\bar{e}) < 0 \).

(ii) For any \( i = 1, 2, \ldots, n \),

\[
\sum_{j=1}^n a_{ij}(\bar{e}) = \int_{S_h} \left( b^{-l}(x, u_h, \nabla S_h, u_h) \nabla S_h \chi_i \cdot \nabla S_h \left( \sum_{j=1}^n \chi_j \right) + r^{-l}(x, u_h) \chi_i \left( \sum_{j=1}^n \chi_j \right) \right) \geq 0,
\]

(3.15)

using the fact that \( \sum_{j=1}^n \chi_j \equiv 1 \) and that \( r, \chi_i \) are nonnegative.

(iii) For any vector \( \mathbf{d} \in \mathbb{R}^n \), \( \mathbf{d} \neq 0 \) and corresponding finite element function \( v_h = \sum_{j=1}^n d_j \chi_j \neq 0 \), using that \( b \geq \mu_0 \) and \( r \geq 0 \), we have

\[
A(\bar{e}) \mathbf{d} \cdot \mathbf{d} - \sum_{i,j=1}^n a_{ij}(\bar{e}) d_id_j = \int_{S_h} \left( b^{-l}(x, u_h, \nabla S_h, u_h) \left| \nabla S_h v_h \right|^2 + r^{-l}(x, u_h) \left| u_h \right|^2 \right) \geq \mu_0 \int_{S_h} \left| \nabla S_h v_h \right|^2 > 0
\]

hence \( A(\bar{e}) \) is positive definite. \( \square \)

Now it is straightforward to derive the analogue of Theorem 2.2 for system (2.23), i.e. the discrete maximum principle.

**Theorem 3.4 (Discrete maximum principle)** Let problem (2.2) satisfy

\[
f(x) - q(x, 0) \leq 0 \quad (x \in S)
\]

(3.16)
and consider the discretization given in Section 2.3. Under the conditions of Theorem 3.3 (in particular, for sufficiently small \( h \)), we have
\[
\max_{S_h} u_h \leq \max_{\partial S_h} \{0, \max_{\partial S_h} g_h\}. \tag{3.17}
\]
In particular, if \( g \geq 0 \) then
\[
\max_{S_h} u_h = \max_{\partial S_h} g_h, \tag{3.18}
\]
and if \( g \leq 0 \) then we have the discrete nonpositivity property
\[
\max_{S_h} u_h \leq 0. \tag{3.19}
\]

**Proof.** It follows from Theorem 3.3 and the properties of piecewise linear functions. Namely, Theorem 3.3 states that the matrix \( \bar{A} (\bar{c}) \) satisfies the conditions of Theorem 3.2, further, (3.16) implies that \( \bar{A} (\bar{c}) \bar{c} = \bar{d} = \int_{S_h} \hat{f} \chi' = 0 \). Hence, for our system \( \bar{A} (\bar{c}) \bar{c} = \bar{d} \), Theorem 3.2 provides the algebraic matrix maximum principle DwMP, which is equivalent to
\[
\max_{B_i \in S_h} u_h (B_i) \leq \max_{B_i \cap \partial S_h} \{0, \max_{B_i \cap \partial S_h} g_h (B_i)\} \tag{3.20}
\]
for the node points. Since \( u_h \) is a piecewise linear function, hence (3.20) is also true in the elements between the node points, i.e. on the whole domain \( S_h \): \( \max_{S_h} u_h \leq \max_{\partial S_h} \{0, \max_{\partial S_h} g_h\} \). Finally, (3.18) and (3.19) are trivial consequences of (3.17).

### 3.2.2 Related results

One can verify in the same way the discrete minimum principle for system (2.23), in analogy with Theorem 3.4:

**THEOREM 3.5** Let problem (2.2) satisfy
\[
f(x) - q(x, 0) \geq 0 \quad (x \in S)
\]
and consider the discretisation given in subsection 2.3. Under the conditions of Theorem 3.3, we have
\[
\min_{S_h} u_h \geq \min_{\partial S_h} \{0, \max_{\partial S_h} g_h\}.
\]
In particular, if \( g \leq 0 \) then
\[
\min_{S_h} u_h = \min_{\partial S_h} g_h,
\]
and if \( g \geq 0 \) then we have the discrete nonnegativity property
\[
\min_{S_h} u_h \geq 0.
\]

In the special case \( q \equiv 0 \), the counterpart of Theorem 2.3 is valid, i.e. equality holds without assuming any sign condition on \( g \). We formulate this for both the maximum and minimum principles. Moreover, the strict negativity in (3.8) can be replaced by a weaker nonnegativity condition, and no regularity condition on the mesh like (3.9) needs to be assumed.

**THEOREM 3.6** Let us consider the following special case of problem (2.2):
\[
\begin{align*}
- \text{div}_{S_h} \left( b(x, u, \nabla_{S_h} u) \nabla_{S_h} u \right) &= f(x) & \text{on } S, \\
\text{u} &= \text{g(x)} & \text{on } \partial S.
\end{align*}
\tag{3.21}
\]
under assumptions (A1)–(A3). Let the triangulation \( \mathcal{T}_h \) satisfy the following property: for any \( i = 1, 2, \ldots, n, j = 1, 2, \ldots, n \ (i \neq j) \)
\[
\nabla_{S_i} \chi_i \cdot \nabla_{S_j} \chi_j \leq 0.
\tag{3.22}
\]
Then the following results hold:
1. If \( f \leq 0 \) then \( \max_{S_h} u_h = \max_{\partial S_h} g_h \).
2. If \( f \geq 0 \) then \( \min_{S_h} u_h = \min_{\partial S_h} g_h \).
3. If \( f = 0 \) then the ranges of \( u_h \) and \( g_h \) coincide, i.e. we have \( [\min_{S_h} u_h, \max_{S_h} u_h] = [\min_{\partial S_h} g_h, \max_{\partial S_h} g_h] \) for the corresponding intervals.
Proof. To verify (1), we rely on Theorem 3.2 again, whose conditions follow similarly as in Theorem 3.3. The difference in the proof arises in proving property (i), i.e.

$$a_{ij}(\vec{e}) \leq 0,$$

which now follows trivially: since the assumption $q \equiv 0$ implies $r \equiv 0$, we simply have

$$a_{ij}(\vec{e}) = \int_{S_h} b^{-1}(x, u_h, \nabla S_h \chi_i) \cdot \nabla S_h \chi_j$$

and thus (3.22) and the condition $b \geq \mu_0 > 0$ readily yield (3.23). Further, in property (ii), (3.15) is simply replaced by

$$\sum_{j=1}^\hat{a} a_{ij}(\vec{e}) = \int_{S_h} b^{-1}(x, u_h, \nabla S_h \chi_h) \cdot \nabla S_h \left( \sum_{j=1}^\hat{a} \chi_j \right) = 0,$$

hence we can apply the final statement of Theorem 3.2 to obtain that DWMP holds. This implies, similarly to the argument of Theorem (3.5) on piecewise linear functions, that $\max_{S_h} u_h = \max_{\delta S_h} g_h$.

Statement (2) follows from (1) by replacing $u$ by $-u$, and (3) is a direct consequence of (1) and (2). \hfill \Box

3.2.3 On geometric mesh properties. Conditions (3.8) and (3.22) are easy to check, since the values $\nabla S_h \chi_i$, $\nabla S_h \chi_j$ are constant on each element. These conditions have a straightforward geometric interpretation: the angles must be uniformly acute for (3.8) and nonobtuse for (3.22).

The currently available constructions of acute triangulations on surfaces of various types and close issues have been recently surveyed in the paper (Zamfirescu, 2013, Section 3), see also the literature therein. We may notice as well that the acuteness property is stable with respect to small perturbations of the vertices involved, therefore e.g. constructions of acute triangulations on planes (see Zamfirescu (2013); Brandts et al. (2009)) can be easily adapted for producing acute triangulations for certain (not very curved) surfaces in 3D.

We underline that many real-life problems appearing in biology, biophysics and biochemistry are posed on sphere-like surfaces. In typical cases the surface at hand there is a membrane, a surface of a cell, or of a nucleous, a surface of a crystal, or some other spherical object, see for instance Dziuk & Elliott (2007, 2013) and the references therein. Acute triangulations and refinement on a spherical surface will be discussed in Section 4.

For general curved surfaces, the issue of refinements preserving acuteness (and the construction of families of acute triangulations) is a more involved task than the case of nonobtuse triangulations (similarly to the case of Euclidean domains). For the latter, for example, a constructive proof of existence of a family of nonobtuse triangulations of surfaces of cylindric-type 3D domains is given by Korotov (2012). It is based on the construction of conforming nonobtuse tetrahedral meshes and the fact that faces of nonobtuse tetrahedra are nonobtuse triangles.

In general, triangulations for implicitly defined surfaces (i.e. given by a distance function as $S = \{ \delta(x) = 0 \}$, cf. Dziuk (1988); Dziuk & Elliott (2013)) can be obtained efficiently by many techniques when the surface is closed, for instance, see the references given by Persson & Strang (2004). For our numerical experiments we used (a modification of) DistMesh developed by Persson & Strang (2004). Triangulations of surfaces with boundary are then easily obtained by a modification of DistMesh, however such triangulations do not satisfy the necessary angle conditions in general.

A way of remedy for mesh generation with angle conditions is decreasing possibly large angles using arbitrary Lagrangian Eulerian maps. Such an approach has been developed by Kovács (2016) to generate meshes with acute angles (or to generate meshes of good quality) for general closed surfaces. The algorithm there is based on a constraint system and on arbitrary Lagrangian Eulerian (ALE) maps, see in particular Section 5.3 therein for meshes with angle conditions. This technique can be adapted to surfaces with boundary, see Section 4 below where we have done this for our numerical tests.

3.3 Examples

We may mention some important real-life examples from the paper Antonini et al. (2007) where the corresponding CMP has been proved. The arising equations include the following models:

(i) gas dynamics:

$$-\text{div}_S \left( \rho (|\nabla_S u|^2) \nabla_S u \right) = 0$$

where the function $\rho$, which describes the relation of the velocity and the density, is determined by Bernoulli’s law:
(ii) surface \( p \)-Laplacian:
\[
-\text{div}_S\left(|\nabla_S u|^{p-2} \nabla_S u\right) = 0
\]
which minimizes the \( p \)-Dirichlet norm on \( S \);
(iii) radiative cooling:
\[
-\text{div}_S\left(\kappa(x,u)|\nabla_S u|^{p-2} \nabla_S u\right) + \sigma u^4 = 0
\]
where \( \kappa \) is the coefficient of heat conduction and \( \sigma \) is the radiation, assumed to be constant. (Here \( u \geq 0 \) is of physical interest Keller (1969), hence the nonlinearity is defined as \( q(x,z) := z^4 \) for \( z \geq 0 \) only and as \( q(x,z) \equiv 0 \) for \( z \leq 0 \).)

According to our results, if the finite element discretization of the corresponding boundary value problem satisfies the angle conditions described in subsection 3.2.3, then the numerical solution satisfies the DMP. The typical situation is that the boundary function \( g \) describes a nonnegative physical quantity: \( g \geq 0 \). Then Theorem 3.5 ensures the discrete nonnegativity property, i.e. that the numerical solution satisfies (3.19):
\[
\min_S u_h \geq 0
\]
in accordance with the physical reality.

4. Numerical experiments
In this section we present illustrative results of numerical tests performed for radiative cooling and \( p \)-Laplacian models from Section 3.3 posed on different surfaces with boundary. As a conclusion, we may observe the validity of the discrete maximum-minimum principle in each numerical example.

4.1 Generating acute meshes
As described in Subsection 3.2.3, the generation of acute or nonobtuse surface triangulations need special care. We first show examples where the particular surface properties can be exploited to generate such meshes directly, and then we also demonstrate how to apply the recently developed generation approach Kovács (2016). The studied surfaces are thus a hemisphere, a semi-torus, and an elaborate surface with four holes.

For the case of the hemisphere, the surface was approximated by subsequent refinements of the initial grid, taken to be the half of an icosahedron. In the refinement step each triangle was first divided into four similar triangles and then newly generated vertices were projected back to the surface.

For the case semi-torus, the meshes were generated via an initial subdivision into round strips, such that the vertices on one side of each strip lie on equal distances and on the other side similarly but in "chess-order", and finally linking the vertices as in Figure 2 (middle).

For the elaborate surface with four holes (a variant with boundary of an example by Elliott & Venkataraman (2015)), we adapted the techniques of Kovács (2016) to surfaces with boundary, also handling the extra difficulty of pairs of triangles with a joint edge (almost) perpendicular to the boundary, which are not accounted by the original algorithm. Using this method we generated triangular meshes with acute angles. The surface with four holes is associated with the distance function
\[
\delta(x) = G(s_1^2) + G(s_2^2) + \frac{x_3^2}{0.12} - 1,
\]
with \( G(s) = 31.25s(s - 0.36)/(s - 0.95) \), then the surface is defined by \( S = \{ x \in \mathbb{R}^3 | \delta(x) = 0, x_3 \geq 0 \} \).

The quality of all three meshes, in terms of minimal and maximal angles, is shown in Figure 1.

4.2 Radiative cooling on a hemisphere
First, we carry out some numerical experiments for the radiative cooling model from (iii) of subsection 3.3 with \( \kappa = 1 \) and \( p = 2 \):
\[
-\Delta_S u + \sigma u^4 = 0 \quad \text{in} \ S, \quad \begin{align*}
u &= g &\text{on } \partial S,
\end{align*}
\]
with the nonlinearity defined as \( q(x,z) := z^4 \) for \( z \geq 0 \) and \( q(x,z) \equiv 0 \) for \( z \leq 0 \). We choose \( \sigma = 5 \), and select the boundary function as \( g(x,y,z) = 1 + xy \). We assume that \( S \) is a hemisphere of the radius 1, therefore
Ran(g) = [0.5, 1.5], in particular, g is a nonnegative function over the boundary. This problem, as well as the ones below, was solved with a damped Newton iteration.

The minimum and maximum values of the numerical solutions for the radiative cooling problem on different meshes are presented in Table 1 (along the minimum and maximum angles in the mesh).

<table>
<thead>
<tr>
<th>dof</th>
<th>min{α}</th>
<th>max{α}</th>
<th>min{uₜ₀}</th>
<th>max{uₜ₀}</th>
</tr>
</thead>
<tbody>
<tr>
<td>91</td>
<td>54.39</td>
<td>71.20</td>
<td>0.5041</td>
<td>1.4755</td>
</tr>
<tr>
<td>341</td>
<td>54.09</td>
<td>71.80</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>1321</td>
<td>54.02</td>
<td>71.95</td>
<td>0.5</td>
<td>1.5</td>
</tr>
<tr>
<td>5201</td>
<td>54.00</td>
<td>71.98</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Table 1. Minimum and maximum values of the numerical solutions for the radiative cooling problem on the hemisphere

4.3 \( p \)-Laplacian on a semi-torus

On the semi-torus we performed the experiment for the \( p \)-Laplacian with \( p = 4 \), i.e.

\[
- \text{div}_S (|\nabla_S u|^2 \nabla_S u) = 0 \quad \text{in} \ S, \\
\quad u = g \quad \text{on} \ \partial S.
\]

Homogeneous \( p \)-Laplacian equations with \( p = 4 \) arise, e.g. in rheology, see Busuioc & Cioranescu (2000). We prescribe the following boundary data: \( g(x,y,z) = 10 + x \), hence Ran(g) = [3, 17]. The discrete maximum principle for the \( p \)-Laplace problem is illustrated by Table 2, similarly as in the previous table before.

<table>
<thead>
<tr>
<th>dof</th>
<th>min{α}</th>
<th>max{α}</th>
<th>min{uₜ₀}</th>
<th>max{uₜ₀}</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>29.39</td>
<td>89.82</td>
<td>3.13397</td>
<td>17</td>
</tr>
<tr>
<td>136</td>
<td>31.24</td>
<td>80.16</td>
<td>3.00629</td>
<td>17</td>
</tr>
<tr>
<td>528</td>
<td>31.72</td>
<td>76.37</td>
<td>3.00034</td>
<td>17</td>
</tr>
<tr>
<td>2080</td>
<td>31.85</td>
<td>74.92</td>
<td>3.00002</td>
<td>17</td>
</tr>
<tr>
<td>8256</td>
<td>31.88</td>
<td>74.39</td>
<td>3.00000</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2. Minimum and maximum values of the numerical solutions for the \( p \)-Laplace problem on the semi-torus
4.4 Radiative cooling on a surface with four holes

Let us consider again the radiative cooling problem (4.1) on the surface

\[ S = \{ x \in \mathbb{R}^3 \mid \delta(x) = 0, \ x_3 \geq 0 \}, \] (4.2)

with the distance function \( \delta(x) = G(x_1^2) + G(x_2^2) + \frac{x_3^2}{0.17} - 1 \), and where \( G(s) = 31.25(s - 0.36)(s - 0.95) \). We prescribe the following boundary data on \( \partial S \):

\[ g(x,y,z) = 1 + xy, \] hence \( \text{Ran}(g) = [0.02345468, 1.97653755] \).

Here the minimum and maximum values over the boundary were obtained by numerically computing the extrema of the constrained problem.

Similarly to the previous tables before, the minimum and maximum values of the numerical solutions for the radiative cooling problem on different meshes approximating the surface with four holes are presented in Table 3.

<table>
<thead>
<tr>
<th>dof</th>
<th>min({\alpha})</th>
<th>max({\alpha})</th>
<th>min({u_h})</th>
<th>max({u_h})</th>
</tr>
</thead>
<tbody>
<tr>
<td>375</td>
<td>32.68</td>
<td>88.34</td>
<td>0.05137</td>
<td>1.94184</td>
</tr>
<tr>
<td>464</td>
<td>33.66</td>
<td>87.26</td>
<td>0.04997</td>
<td>1.94978</td>
</tr>
<tr>
<td>672</td>
<td>37.45</td>
<td>87.99</td>
<td>0.02776</td>
<td>1.97317</td>
</tr>
<tr>
<td>1025</td>
<td>34.30</td>
<td>88.10</td>
<td>0.02886</td>
<td>1.97553</td>
</tr>
</tbody>
</table>

Table 3. Minimum and maximum values of the numerical solutions for the radiative cooling problem on the surface with four holes

Finally, Figure 2 reports on the surface meshes and on the numerical solutions for all three examples. The plotted results correspond to the last rows in each table, respectively.

Acknowledgements

We thank two anonymous referees for their constructive and helpful comments.

The work of J. Karátson is supported by the Hungarian Scientific Research Fund OTKA, No. K112157 and SNN125119. The work of B. Kovács is funded by Deutsche Forschungsgemeinschaft, SFB 1173. A grant from the Eötvös Loránd University let B.K. spend a week in Budapest, allowing personal discussions with J.K.

REFERENCES


FIG. 2. Surface meshes and the numerical solutions for the three nonlinear problems (from top to bottom).


