INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING
Volume 6, Number 1, Pages 1–16

DISCRETE MAXIMUM PRINCIPLES FOR FEM SOLUTIONS OF SOME NONLINEAR ELLIPTIC INTERFACE PROBLEMS

JÁNOS KARÁTSON AND SERGEY KOROTOV

Abstract. Discrete maximum principles are proved for finite element discretizations of nonlinear elliptic interface problems with jumps of the normal derivatives. The geometric conditions in the case of simplicial meshes are suitable acuteness or nonobtuseness properties.

Key Words. Nonlinear elliptic problem, interface problem, maximum principle, discrete maximum principle, finite element method, simplicial mesh.

1. Introduction

The maximum principle forms an important qualitative property of second order elliptic boundary value problems [12, 25, 29]. Consequently, the discrete analogues of the maximum principle (so-called discrete maximum principles, DMPs) have drawn much attention. Various DMPs have been formulated and proved including the case of finite difference, finite volume and finite element approximations, and corresponding geometric conditions on the computational meshes have been given, see, e.g., [3, 5, 6, 7, 9, 13, 21, 30, 31, 33] for linear and [16, 17, 22] for nonlinear problems with standard (i.e., Dirichlet, and in [16, 17] mixed) boundary conditions.

In this paper we address interface problems, which arise in various branches of material science, biochemistry, multiphase flow etc., often when two or more distinct materials are involved with different conductivities or densities. Another (for our work, motivating) example is from localized reaction-diffusion problems [14, 15], see at the end of this paper. Many special numerical methods have been designed for interface problems, see, e.g., [14, 27, 28, 26], but maximum principles have received less attention than for the case of standard boundary value problems. A continuous minimum principle for a related problem is given in [11]. The discrete maximum principle for suitable finite difference discretizations of linear interface problems has been proved in [27].

Our goal is to derive maximum principles for nonlinear elliptic interface problems when finite element discretization is involved. The present paper is the extension of our paper [16] to a class of such problems, and relies on a similar technique using weak formulation and positivity conditions that ensure well-posedness. Our result is based on the observation that we can recast the considered interface problem to a weak formulation, which is similar to that of the mixed problem studied in [16]. We consider matching conditions for the solution on the interface, i.e., the jump is allowed for the normal derivatives but not for the solution itself. Problems

Received by the editors May 15 2007 and, in revised form, February 12 2008.

2000 *Mathematics Subject Classification.* 35B50, 35J65, 65N30, 65N50.

The first author was supported by the Hungarian Research Fund OTKA under grant NO. K67819 and by HAS under the Bolyai János Scholarship. The second author was supported by the Academy Research Fellowship NO. 208628 from the Academy of Finland.

with jump of the solution or without well-posedness may be the subject of further research.

The paper is organized as follows. The formulation of the problem is presented in Section 2 with focus on the suitable weak form of the problem, and a corresponding continuous maximum principle is enclosed. The finite element discretization is described in Section 3. Discrete maximum principles are derived and examples are given in Section 4.

2. Nonlinear elliptic interface problems

2.1. Formulation of the problem. We investigate nonlinear interface problems of the following form:

$$\left\{ \begin{aligned} -\operatorname{div}\Big(b(x,\nabla u)\,\nabla u\Big) + q(x,u) \;&=\; f(x) \quad \text{in } \Omega \setminus \Gamma, \\ [\,u]_\Gamma \;&=\; 0 \qquad\quad \text{on } \Gamma, \\ \big[\,b(x,\nabla u)\tfrac{\partial u}{\partial \nu}\big]_\Gamma + s(x,u) \;&=\; \gamma(x) \quad \text{on } \Gamma, \\ u \;&=\; g(x) \quad \text{on } \partial\Omega, \end{aligned} \right. \tag{1}$$

where $\partial\Omega$ denotes the boundary of the domain Ω and the interface Γ is a surface lying in Ω, further, ν denotes the outward normal unit vector, $[\,u]_\Gamma$ and $\big[\,b(x,\nabla u)\frac{\partial u}{\partial \nu}\big]_\Gamma$ denote the jump (i.e., the difference of the limits from the two sides of the interface Γ) of the solution u and the flux $b(x,\nabla u)\frac{\partial u}{\partial \nu}$, respectively. We impose the following

Assumptions 2.1:

(A1) Ω is a bounded open domain in $\mathbf{R}^d$, $d \in \{1,2,\dots\}$ the interface $\Gamma \subset \Omega$ and the boundary $\partial\Omega$ are piecewise smooth and Lipschitz continuous $(d-1)$-dimensional surfaces.

(A2) The scalar functions $b : \Omega \times \mathbf{R}^d \to \mathbf{R}$, $\;q : \Omega \times \mathbf{R} \to \mathbf{R}$ and $s : \Gamma \times \mathbf{R} \to \mathbf{R}$ are measurable and bounded w.r.t. their first variable $x \in \Omega$ (resp. $x \in \Gamma$) and continuously differentiable w.r.t. their second variable $\eta \in \mathbf{R}^d$ (resp. $\xi \in \mathbf{R}$). Further, $f \in L^2(\Omega)$, $\gamma \in L^2(\Gamma)$ and $g \in H^1(\Omega)$.

(A3) The function b satisfies

$$0 < \mu_0 \le b(x,\eta) \le \mu_1 \tag{2}$$

with positive constants μ_0 and μ_1 independent of (x,η), further, the diadic product matrix $\eta \cdot \frac{\partial b(x,\eta)}{\partial \eta}$ is symmetric positive semidefinite and bounded in matrix norm by some positive constant μ_2 independent of (x,η).

(A4) Let $2 \le p_1$ if $d = 2$, or $2 \le p_1 \le \frac{2d}{d-2}$ if $d > 2$, further, let $2 \le p_2$ if $d = 2$, or $2 \le p_2 \le \frac{2d-2}{d-2}$ if $d > 2$. There exist functions $\alpha_1 \in L^{d/2}(\Omega)$, $\alpha_2 \in L^{d-1}(\Gamma)$ and a constant $\beta \ge 0$ such that for any $x \in \Omega$ (or $x \in \Gamma$, resp.) and $\xi \in \mathbf{R}$

$$0 \le \frac{\partial q(x,\xi)}{\partial \xi} \le \alpha_1(x) + \beta|\xi|^{p_1-2}, \qquad 0 \le \frac{\partial s(x,\xi)}{\partial \xi} \le \alpha_2(x) + \beta|\xi|^{p_2-2}. \tag{3}$$

Remark 2.1. Problem (1) contains some widespread interface models as special cases, see, e.g., [15, 28] and also the models addressed in subsection 4.4.

Remark 2.2. (i) The role of assumption (A3) is to ensure that the Jacobian matrices $J(x,\eta) := \frac{\partial}{\partial \eta}\Big(b(x,\eta)\,\eta\Big)$ are symmetric and satisfy the uniform ellipticity property $\mu_0|\zeta|^2 \le \zeta^T J(x,\eta)\;\zeta \le \mu_3|\zeta|^2$, $\;\;\zeta \in \mathbf{R}^d$ (with $\mu_3 = \mu_1 + \mu_2$), which will be required for well-posedness. For instance, assumption (A3) holds for coefficients of the form $b(x,\eta) = a(x,|\eta|)$ (see [10, 23] for such nonlinearities), where the C^1 function $a : \Omega \times \mathbf{R}^+ \to \mathbf{R}$ satisfies $0 < \mu_0 \le a(x,r) \le \frac{\partial}{\partial r}(a(x,r)\,r) \le \mu_3$ $(r > 0)$. More specially, one may have $b(x,\eta) = a(x)$ (i.e., linear principal part) with a measurable function a satisfying $0 < \mu_0 \le a(x) \le \mu_3$.

(ii) The conditions on the exponents p_1 and p_2 in assumption (A4) ensure the embeddings of $H^1(\Omega)$ into $L^{p_1}(\Omega)$ and of the space of traces of $H^1(\Omega)$ on Γ into $L^{p_2}(\Gamma)$, respectively, see [1].

For the study of such problems, one needs a precise definition of solution. First we define the classical solution as a smooth function for which (1) can be properly understood pointwise. Here we assume in addition that the interface Γ is a closed surface, or more generally, it is any compact subset of an (also piecewise smooth and Lipschitz continuous) closed surface $\hat{\Gamma} \subset \Omega$ as illustrated in Figure 1. Let us denote by Ω_0 the domain enclosed by the surface $\hat{\Gamma}$, i.e., $\partial\Omega_0 = \hat{\Gamma}$.

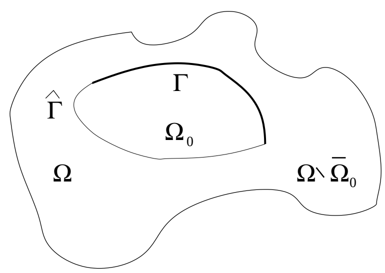

FIGURE 1. Interface in a domain

Definition 2.1. We call $u : \overline{\Omega} \to \mathbf{R}$ a *classical solution of problem (1)* if $u \in C^2(\Omega \setminus \Gamma)$, $u\big|_{\overline{\Omega}_0} \in C^1(\overline{\Omega}_0)$, $u\big|_{\overline{\Omega\setminus\Omega_0}} \in C^1(\overline{\Omega \setminus \Omega_0})$ and u satisfies (1) pointwise.

We note that if there exists a classical solution u, and if b, q, s are also C^1 w.r.t. x (which often holds in practice when b, q, s are just independent of x), then the equalities in (1) imply additional properties for the input data, namely, they satisfy $f \in C(\Omega \setminus \Gamma)$, $\gamma \in C(\Gamma)$ and $g \in C(\overline{\Omega})$. That is, these properties are necessary for the existence of a classical solution u. Sufficent conditions for classical solvability may be much harder to formulate and are, however, not required in this paper. Instead, a suitable weak form will turn to be relevant in our context.

In what follows, we pass on to this suitable weak form. In the next subsection we define the weak solution and justify its relevance by its relation to the classical solution. Afterwards in this paper, this weak formulation will be used to define finite element discretization and to state the corresponding discrete maximum principle.

2.2. Weak formulation and its coherence with the strong form.

Definition 2.2. A *weak solution* of problem (1) is a function $u^* \in H^1(\Omega)$ satisfying

$$(4) \qquad \int_\Omega \Big(b(x,\nabla u^*)\ \nabla u^* \cdot \nabla v + q(x,u^*)v\Big)\,dx + \int_\Gamma s(x,u^*)v\,d\sigma = \int_\Omega fv\,dx + \int_\Gamma \gamma v\,d\sigma \qquad \forall v \in H_0^1(\Omega)$$

$$\text{and} \quad u^* = g \quad \text{on } \partial\Omega. \tag{5}$$

Proposition 2.1. *(i) A classical solution of problem (1) is also a weak solution.*

(ii) Conversely, if a weak solution satisfies the smoothness requirements of Definition 2.1, then it is a classical solution.

PROOF. (i) Let $x \in \hat{\Gamma}$, let ν denote the normal unit vector pointing out of Ω_0, and let $\hat{\nu} := -\nu$ (the normal unit vector pointing out of $\Omega \setminus \Omega_0$). The jump of $b(x, \nabla u)\frac{\partial u}{\partial \nu}$ at x is the difference of the limits of $b(\cdot, \nabla u)\frac{\partial u}{\partial \nu}$ at x from Ω_0 and from $\Omega \setminus \Omega_0$. Using the definition $\frac{\partial u}{\partial \nu}(x) := \lim\limits_{t \to 0^+} \frac{1}{t} \Big(u(x) - u(x - t\nu) \Big)$, we thus have

$$\left[b(x, \nabla u)\frac{\partial u}{\partial \nu} \right]_{x \in \Gamma} :=$$

$$b\Big(x, \nabla(u|_{\overline{\Omega}_0})(x)\Big) \lim_{t \to 0^+} \frac{1}{t} \Big(u(x) - u(x - t\nu) \Big) - b\Big(x, \nabla(u|_{\overline{\Omega \setminus \Omega_0}})(x)\Big) \lim_{t \to 0^-} \frac{1}{t} \Big(u(x) - u(x - t\nu) \Big) =$$

$$b\Big(x, \nabla(u|_{\overline{\Omega}_0})(x)\Big) \lim_{t \to 0^+} \frac{1}{t} \Big(u(x) - u(x - t\nu) \Big) + b\Big(x, \nabla(u|_{\overline{\Omega \setminus \Omega_0}})(x)\Big) \lim_{s \to 0^+} \frac{1}{s} \Big(u(x) - u(x - s\hat{\nu}) \Big) =$$

$$\left(b\Big(x, \nabla(u|_{\overline{\Omega}_0})(x)\Big) \frac{\partial u}{\partial \nu} + b\Big(x, \nabla(u|_{\overline{\Omega \setminus \Omega_0}})(x)\Big) \frac{\partial u}{\partial \hat{\nu}} \right)_{x \in \Gamma}. \tag{6}$$

Now let u be a classical solution. The assumptions imply that $u \in H^1(\Omega)$, and (5) holds trivially. For any $v \in H_0^1(\Omega)$, Green's formula for equation (1) on Ω_0 and $\Omega \setminus \Omega_0$, respectively, yields

$$\int_{\Omega_0} f v \, dx = \int_{\Omega_0} \Big(b(x, \nabla u) \, \nabla u \cdot \nabla v + q(x, u) v \Big) \, dx - \int_{\hat{\Gamma}} b\Big(x, \nabla(u|_{\overline{\Omega}_0})\Big) \frac{\partial u}{\partial \nu} \, v \, d\sigma$$

and

$$\int_{\Omega \setminus \Omega_0} f v \, dx = \int_{\Omega \setminus \Omega_0} \Big(b(x, \nabla u) \, \nabla u \cdot \nabla v + q(x, u) v \Big) \, dx - \int_{\hat{\Gamma}} b\Big(x, \nabla(u|_{\overline{\Omega \setminus \Omega_0}})\Big) \frac{\partial u}{\partial \hat{\nu}} \, v \, d\sigma.$$

Summing up, the integrand on $\hat{\Gamma}$ becomes the jump on Γ (using (6)) and zero on $\hat{\Gamma} \setminus \Gamma$ (since ∇u is continuous there and $\hat{\nu} = -\nu$). In virtue of the jump condition in (1), we altogether obtain

$$\int_{\Omega} f v \, dx = \int_{\Omega} \Big(b(x, \nabla u) \, \nabla u \cdot \nabla v + q(x, u) v \Big) \, dx - \int_{\Gamma} \left[b(x, \nabla u) \frac{\partial u}{\partial \nu} \right]_{\Gamma} v \, d\sigma \tag{7}$$

$$= \int_{\Omega} \Big(b(x, \nabla u) \, \nabla u \cdot \nabla v + q(x, u) v \Big) \, dx + \int_{\Gamma} \Big((s(x, u) - \gamma \Big) \, v \, d\sigma\,. \tag{8}$$

(ii) This direction follows from the above in the standard way, using the argument in the opposite direction. Let u be a weak solution. First, for any $v \in H_0^1(\Omega)$ satisfying $v_{|\Gamma} = 0$, (8) and Green's formula imply that the first equation in (1) holds. The latter in turn yields that the integrals on Γ in (7) and (8) coincide for all $v \in H_0^1(\Omega)$, hence the third equation in (1) (the normal jump condition) also holds. The second equation in (1) (the zero jump condition on u) follows even from $u \in H^1(\Omega)$, and the last equation in (1) is explicitly required in (5). ∎

2.3. Existence and uniqueness.

Theorem 2.1. *Under Assumptions 2.1, problem (1) has a unique weak solution $u^* \in H^1(\Omega)$, i.e. that satisfies (4)-(5).*

Proof. (i) We first prove the theorem for homogeneous boundary condition, i.e. when $g = 0$. In this case the weak solution u^* can be obtained using monotone operators, in a similar way as in [10, Chap. 6], therefore we only indicate the main steps of the proof. First, we define

(9)
$$\langle F(u), v\rangle \;=\; \int\limits_\Omega \Big(b(x,\nabla u)\,\nabla u\cdot\nabla v + q(x,u)v - fv\Big)\,dx + \int\limits_\Gamma \Big(s(x,u)v - \gamma v\Big)\,d\sigma,\; v \in H_0^1(\Omega),$$

where the growth conditions in (A1)–(A4) ensure that the arising integrals are finite. Let $J(x,\eta) := \frac{\partial}{\partial\eta}\Big(b(x,\eta)\,\eta\Big)$ as in Remark 2.2. Then, from (A3)-(A4), we obtain that the Gateaux derivative $F'(u)$ exists, is self-adjoint for all u and satisfies

(10)
$$\langle F'(u)v, v\rangle \;=\; \int\limits_\Omega \Big(J(x,\nabla u)\,\nabla v\cdot\nabla v + q_u'(x,u)v^2\Big)\,dx + \int\limits_\Gamma s_u'(x,u)v^2\,d\sigma \geq \mu_0 \int\limits_\Omega |\nabla v|^2\,dx$$

(for all $u, v \in H_0^1(\Omega)$), where q_u', s_u' denote derivatives w.r.t. u. Using the standard Sobolev norm defined via

$$\|v\|_1^2 = \int_\Omega |\nabla v|^2\,dx \qquad (v \in H_0^1(\Omega)), \tag{11}$$

the uniform ellipticity (10) implies that the operator equation $F(u) = 0$ has a unique solution $u^* \in H_0^1(\Omega)$. Here $F(u^*) = 0$ is equivalent to (4), i.e., u^* is the weak solution.

(ii) For non-homogeneous boundary conditions the problem can be reduced to the homogeneous case using a usual translation. Let $g \in H^1(\Omega)$ be arbitrary and let us require (5) on the boundary. Then we look for u^* in the form $u^* = u + g$, in which case $u = 0$ on $\partial\Omega$. Substituting this sum into (4), we observe that u must satisfy the same problem with homogeneous boundary conditions and with coefficients

$$\hat b(x,\eta) = b(x,\eta + \nabla g(x)), \quad \hat q(x,\xi) = q(x,\xi + g(x)), \quad \hat s(x,\xi) = s(x,\xi + g(x)).$$

Here $g(x)$ is independent of ξ, η, hence these coefficients remain C^1 in their second variable and satisfy the same growth conditions as b, q, s. This implies existence for u, and then the same for u^* owing to the relation $u^* = u + g$.

To prove uniqueness, note first that the operator F in (9) makes sense for all $u \in H_0^1(\Omega)$, as well as estimate (10). (In the latter, the homogeneous boundary condition was only necessary for v to ensure that the r.h.s. of (10) defines a norm for v.) With this, the requirements (4)-(5) of the weak solution can be written as

$$\langle F(u^*), v\rangle \;=\; 0 \quad \forall v \in H_0^1(\Omega), \qquad u^* \;=\; g \quad \text{on } \partial\Omega. \tag{12}$$

Now let $\hat u$ be another weak solution, i.e. that also satisfies (12). Then subtracting (12) for $\hat u$ from (12) for u^*, and setting $v := u^* - \hat u$ (which is in $H_0^1(\Omega)$), we obtain

$$\langle F(u^*) - F(\hat u), u^* - \hat u\rangle \;=\; 0. \tag{13}$$

On the other hand, (10) implies

$$\langle F(u^*) - F(\hat u), u^* - \hat u\rangle = \langle F'\big(\hat u + \theta\big(u^* - \hat u\big)\big)(u^* - \hat u), u^* - \hat u\rangle \geq \mu_0 \int\limits_\Omega |\nabla(u^* - \hat u)|^2\,dx,$$

and these two mean that the norm (11) of the function $u^* - \hat{u} \in H_0^1(\Omega)$ equals zero, i.e. $u^* = \hat{u}$. ∎

2.4. Continuous maximum principles. In Theorems 2.2 and 2.3 we formulate and prove two continuous maximum principles for our PDE problem (1). These statements provide the properties whose discrete analogues can be expected for suitable FEM solutions.

Theorem 2.2. *Let Assumptions 2.1 hold and*

$$f(x) - q(x,0) \le 0, \ x \in \Omega, \quad \textit{and} \quad \gamma(x) - s(x,0) \le 0, \ x \in \Gamma. \tag{14}$$

If the weak solution u of problem (1) belongs to $C^1(\Omega \setminus \Gamma) \cap C(\overline{\Omega})$, then

$$\max_{\overline{\Omega}} u \le \max\{0, \max_{\partial\Omega} g\}. \tag{15}$$

In particular, if $g \ge 0$, then $\max_{\overline{\Omega}} u = \max_{\partial\Omega} g$, and if $g \le 0$, then we have the nonpositivity property $\max_{\overline{\Omega}} u \le 0$.

In general, if $u \in H^1(\Omega)$ only (without the above regularity assumption) and g is a.e. bounded on $\partial\Omega$, then the same statements hold if $\max u$ and $\max g$ are replaced by $\operatorname{ess\,sup} u$ and $\operatorname{ess\,sup} g$, respectively.

Proof. We only prove the regular case, following [16]. The general case is similar, if $\max u$ and $\max g$ are replaced by $\operatorname{ess\,sup} u$ and $\operatorname{ess\,sup} g$, respectively. Let

$$r(x,\xi) := \begin{cases} \frac{q(x,\xi)-q(x,0)}{\xi}, & \text{if } \xi \neq 0, \\ \frac{\partial q}{\partial \xi}(x,0), & \text{if } \xi = 0, \end{cases} \qquad z(x,\xi) := \begin{cases} \frac{s(x,\xi)-s(x,0)}{\xi}, & \text{if } \xi \neq 0, \\ \frac{\partial s}{\partial \xi}(x,0), & \text{if } \xi = 0. \end{cases} \tag{16}$$

Here, using (A2), the functions r and z are continuous in ξ. Further, in view of (A4), we have $r(x,\xi) \ge 0$, $z(x,\xi) \ge 0$. We define $\tilde{a}(x) := b(x, \nabla u(x))$ $(x \in \Omega \setminus \Gamma)$, $\tilde{h}(x) := r(x, u(x))$ $(x \in \Omega)$, $\tilde{k}(x) := z(x, u(x))$ $(x \in \Gamma)$. Using also the notations

$$\hat{f}(x) := f(x) - q(x,0) \qquad \text{and} \quad \hat{\gamma}(x) := \gamma(x) - s(x,0), \tag{17}$$

the weak formulation of problem (1) is rewritten as

$$\int_\Omega \big(\tilde{a}\ \nabla u \cdot \nabla v + \tilde{h} uv\big)\, dx + \int_\Gamma \tilde{k} uv \, d\sigma = \int_\Omega \hat{f} v \, dx + \int_\Gamma \hat{\gamma} v \, d\sigma \qquad \forall v \in H_0^1(\Omega). \tag{18}$$

Let $M := \max\{0, \max_{\partial\Omega} g\}$, and we introduce the continuous and piecewise C^1 function $v := \max\{u - M, 0\}$. Thus $v \in H^1(\Omega)$, we have $v \ge 0$ and $v_{|\partial\Omega} = 0$, further, $u(x) = v(x) + M$ for any $x \in \overline{\Omega}$ unless $v(x) = 0$. Hence, for this v the left-hand side of (18) satisfies

$$\int_\Omega \big(\tilde{a}\ \nabla u \cdot \nabla v + \tilde{h} uv\big)\, dx + \int_\Gamma \tilde{k} uv \, d\sigma = \int_\Omega \big(\tilde{a}\, |\nabla v|^2 + \tilde{h} \cdot (v+M) v\big)\, dx + \int_\Gamma \tilde{k} \cdot (v+M) v \, d\sigma \ge 0$$

since the functions $\tilde{a}, \tilde{h}, \tilde{k}, v$ and the constant M are nonnegative. On the other hand, the assumptions $\hat{f} \le 0$, $\hat{\gamma} \le 0$ imply that for this v the right-hand side of (18) is nonpositive, which together imply the relation

$$\int_\Omega \big(\tilde{a}\, |\nabla v|^2 + \tilde{h} \cdot (v+M) v\big)\, dx + \int_\Gamma \tilde{k} \cdot (v+M) v \, d\sigma = 0.$$

By assumption (A3), here $\tilde{a}$ has a positive minimum, hence $|\nabla v| = 0$, i.e., v is constant. We have seen that $v_{|\partial\Omega} = 0$, hence we obtain that $v \equiv 0$, which just means that (15) holds. ∎

The following special case provides equality of maxima on $\partial\Omega$ without assuming $g \geq 0$:

Theorem 2.3. *Let $q \equiv 0$ and $s \equiv 0$ in problem (1). Let us impose the assumptions of Theorem 2.2, which now means that (A1)–(A3) are satisfied, $u \in C^1(\Omega \setminus \Gamma) \cap C(\overline{\Omega})$, and (14) takes the form*

$$f(x) \leq 0, \; x \in \Omega \quad \text{and} \quad \gamma(x) \leq 0, \; x \in \Gamma. \tag{19}$$

Then

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} g. \tag{20}$$

(If $u \in H^1(\Omega)$ only and g is a.e. bounded on $\partial\Omega$, then $\operatorname{ess\,sup} u = \operatorname{ess\,sup} g$ on $\partial\Omega$.)

Proof. We only prove the regular case again. If $\max_{\partial\Omega} g \geq 0$ then (15) implies (20). Let $\max_{\partial\Omega} g < 0$, say, $\max_{\partial\Omega} g = -K$ with some $K > 0$. Then the function $w := u + K$ satisfies the same mixed problem with right-hand sides f, γ and $g + K$, respectively, hence Theorem 2.2 is valid for this problem as well, and (15) for w yields $\max_{\overline{\Omega}} w \leq \max\{0, \max_{\partial\Omega} (g + K)\} = 0$. Then $\max_{\overline{\Omega}} u \leq -K = \max_{\partial\Omega} g$. ∎

Remark 2.3. Analogously to Theorems 2.2 and 2.3, corresponding minimum principles and nonnegativity property hold if the sign conditions in (14) and (19) are reversed.

3. Finite element discretization of the problem

In the sequel we impose a basic assumption that Ω is a polytopic domain and the interface Γ is also polytopic. We note that if $\partial\Omega$ or Γ are curved then the convergence of the discrete solution to the exact one is a much more difficult problem, out of the scope of this paper. Even for the simpler case of Dirichlet problems in 3D without interface, such an analysis has been given only recently in [18].

We introduce a finite element discretization of our problem with simplicial elements and continuous piecewise linear basis functions. Thus, let $\mathcal{T}_h$ be a conforming triangulation of $\overline{\Omega}$ into simplices (denoted by symbol T later on, possibly with some subindices), whose nodes are $B_1, ..., B_{\bar{n}}$, and where $h := \max_{T \in \mathcal{T}_h} \operatorname{diam} T$. Denote by $\phi_1, ..., \phi_{\bar{n}}$ the piecewise linear continuous basis functions defined in a standard way, i.e., $\phi_i(B_j) = \delta_{ij}$ for $i, j = 1, ..., \bar{n}$, where δ_{ij} is the Kronecker symbol. Let V_h denote the finite element subspace spanned by the above basis functions:

$$V_h = \operatorname{span}\{\phi_1, ..., \phi_{\bar{n}}\} \subset H^1(\Omega).$$

Let $n < \bar{n}$ be such that

$$B_1, ..., B_n \tag{21}$$

are the nodes that lie in Ω and let

$$B_{n+1}, ..., B_{\bar{n}} \tag{22}$$

be the nodes that lie on $\partial\Omega$. Then the basis functions $\phi_1, ..., \phi_n$ satisfy homogeneous boundary condition on $\partial\Omega$, i.e., $\phi_i \in H_0^1(\Omega)$. We define

$$V_h^0 = \operatorname{span}\{\phi_1, ..., \phi_n\} \subset H_0^1(\Omega).$$

Further, let

$$g_h = \sum_{j=n+1}^{\bar{n}} g_j \phi_j \ \in V_h \tag{23}$$

(with $g_j \in \mathbf{R}$) be the piecewise linear approximation of the function g on $\partial\Omega$ (and on the neighbouring elements). To find the FEM solution of (4)-(5) in V_h, we solve the following problem: find $u_h \in V_h$ such that

$$\begin{aligned} &\int_\Omega \Big(b(x, \nabla u_h)\ \nabla u_h \cdot \nabla v_h + \ q(x,u_h) v_h \Big)\, dx \ + \int_\Gamma s(x,u_h) v_h \, d\sigma \\ &= \int_\Omega f v_h \, dx \ + \int_\Gamma \gamma v_h \, d\sigma \quad \forall v_h \in V_h^0, \quad \text{and} \qquad u_h \ = \ g_h \quad \text{on } \partial\Omega. \end{aligned} \tag{24}$$

Theorem 3.1. *Under Assumptions 2.1, problem (24) has a unique solution $u_h \in V_h$, and $\|u^* - u_h\|_1 \to 0$ as $h \to 0$.*

Proof. The proof of Theorem 2.1 can be repeated to obtain u_h, just replacing $H^1(\Omega)$ by V_h. The convergence of u_h to u^* in H^1-norm follows in the standard way from the ellipticity of the equation and the fact that the finite-dimesional subspaces V_h satisfy the condition $\lim_{h\to 0} \operatorname{dist}(u, V_h) = 0$ for all $u \in H^1(\Omega)$, where $\operatorname{dist}(u, V_h) \ = \inf_{v_h \in V_h} \|u - v_h\|_1$ (see [8]). ■

Let us now formulate the nonlinear algebraic system corresponding to (24). First we rewrite problem (24) with the notations (16) and (17):

$$\begin{aligned} &\int_\Omega \Big(b(x, \nabla u_h)\ \nabla u_h \cdot \nabla v_h + \ r(x,u_h) u_h v_h \Big)\, dx \ + \int_\Gamma z(x,u_h) u_h v_h \, d\sigma \\ &= \int_\Omega \hat{f} v_h \, dx \ + \int_\Gamma \hat{\gamma} v_h \, d\sigma, \forall v_h \in V_h^0 \end{aligned} \tag{25}$$

We set

$$u_h = \sum_{j=1}^{\bar{n}} c_j \phi_j, \tag{26}$$

and look for the coefficients $c_1, \ldots, c_{\bar{n}}$. For any $\bar{\mathbf{c}} = (c_1, ..., c_{\bar{n}}) \in \mathbf{R}^{\bar{n}}$, $i = 1, ..., n$ and $j = 1, ..., \bar{n}$, we set

$$b_{ij}(\bar{\mathbf{c}}) := \int_\Omega b(x, \sum_{k=1}^{\bar{n}} c_k \nabla \phi_k)\ \nabla\phi_j \cdot \nabla \phi_i \, dx, \qquad r_{ij}(\bar{\mathbf{c}}) := \int_\Omega r(x, \sum_{k=1}^{\bar{n}} c_k \phi_k)\ \phi_j \phi_i \, dx,$$

$$z_{ij}(\bar{\mathbf{c}}) := \int_\Gamma z(x, \sum_{k=1}^{\bar{n}} c_k \phi_k)\ \phi_j \phi_i \, d\sigma, \qquad d_i(\bar{\mathbf{c}}) := \int_\Omega \hat{f} \phi_i \, dx \ + \int_\Gamma \hat{\gamma} \phi_i \, d\sigma\,,$$

$$a_{ij}(\bar{\mathbf{c}}) := b_{ij}(\bar{\mathbf{c}}) + r_{ij}(\bar{\mathbf{c}}) + z_{ij}(\bar{\mathbf{c}}). \tag{27}$$

Putting (26) and $v_h = \phi_i$ into (25), we obtain the $n \times \bar{n}$ system of algebraic equations

$$\sum_{j=1}^{\bar{n}} a_{ij}(\bar{\mathbf{c}})\, c_j = d_i, \quad i = 1, ..., n. \tag{28}$$

Using the notations

(29)

$$\mathbf{A}(\bar{\mathbf{c}}) := \{a_{ij}(\bar{\mathbf{c}})\},\ i,j=1,...,n, \qquad \tilde{\mathbf{A}}(\bar{\mathbf{c}}) := \{a_{ij}(\bar{\mathbf{c}})\},\ i=1,...,n;\ j=n+1,...,\bar{n},$$

$$\mathbf{d} := \{d_j\},\ \mathbf{c} := \{c_j\}, \quad j=1,...,n, \quad \text{and} \quad \tilde{\mathbf{c}} := \{c_j\}, \quad j=n+1,...,\bar{n},$$

system (28) turns into

$$\mathbf{A}(\bar{\mathbf{c}})\mathbf{c} + \tilde{\mathbf{A}}(\bar{\mathbf{c}})\tilde{\mathbf{c}} = \mathbf{d}. \tag{30}$$

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

$$\tilde{\mathbf{c}} = \tilde{\mathbf{g}} := \{g_j\}, \quad j=n+1,...,\bar{n},$$

hence we can replace (30) by the equivalent system

$$\begin{bmatrix} \mathbf{A}(\bar{\mathbf{c}}) & \tilde{\mathbf{A}}(\bar{\mathbf{c}}) \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{g}} \end{bmatrix}. \tag{31}$$

Defining further

$$\bar{\mathbf{A}}(\bar{\mathbf{c}}) := \begin{bmatrix} \mathbf{A}(\bar{\mathbf{c}}) & \tilde{\mathbf{A}}(\bar{\mathbf{c}}) \\ 0 & \mathbf{I} \end{bmatrix}, \qquad \bar{\mathbf{c}} := \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix}, \qquad \bar{\mathbf{d}} := \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{g}} \end{bmatrix}, \tag{32}$$

we rewrite (31) as follows:

$$\bar{\mathbf{A}}(\bar{\mathbf{c}})\bar{\mathbf{c}} = \bar{\mathbf{d}}. \tag{33}$$

Concerning the solvability of system (33), note that it is the nonlinear algebraic system corresponding to (24). This means that (33) is equivalent to (24), where this equivalence is realized by the one-to-one correspondence (26) between the vectors $\bar{\mathbf{c}} \in \mathbf{R}^{\bar{n}}$ and the functions $u_h \in V_h$. Therefore, Theorem 3.1 implies that system (33) has a unique solution.

In practice, to find the solution of system (33) one applies some (Newton-like or other) iterative method. Construction and convergence, the latter in particular due to the involved monotone operator framework, is summarized for such methods e.g. in [10]. An inexact Newton iteration designed especially for interface problems is given in [2]. In what follows, we are instead interested in the discrete maximum principle for system (33).

4. Maximum principle for the discretized problem

4.1. Background. First we recall a basic definition in the study of DMP (cf. [32, p. 23]):

Definition 4.1. A square $n \times n$ matrix $\mathbf{M} = (m_{ij})_{i,j=1}^n$ is called *irreducibly diagonally dominant* if it satisfies the following conditions:

(i) $\mathbf{M}$ is irreducible, i.e., for any $i \ne j$ there exists a sequence of nonzero entries $\{m_{i,i_1}, m_{i_1,i_2}, \ldots, m_{i_s,j}\}$ of $\mathbf{M}$, where $i, i_1, i_2, \ldots, i_s, j$ are distinct indices,

(ii) $\mathbf{M}$ is diagonally dominant, i.e., $|m_{ii}| \ge \sum\limits_{\substack{j=1 \\ j\ne i}}^n |m_{ij}|,\ i=1,...,n,$

(iii) for at least one index $i_0 \in \{1,...,n\}$ the above inequality is strict, i.e.,

$$|m_{i_0,i_0}| > \sum_{\substack{j=1 \\ j\ne i_0}}^n |m_{i_0,j}|.$$

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

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

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

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \tilde{\mathbf{A}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} . \tag{34}$$

Here $\mathbf{I}$ is the $m \times m$ identity matrix and $\mathbf{0}$ is the $m \times n$ zero matrix. Following [7], we introduce

Definition 4.2. An $(n+m)\times(n+m)$ matrix $\bar{\mathbf{A}}$ with the structure (34) is said to be of *generalized nonnegative type* if the following properties hold:

(i) $a_{ii} > 0, \quad i = 1, ..., n,$

(ii) $a_{ij} \le 0, \quad i = 1, ..., n, \ j = 1, ..., n+m \quad (i \ne j),$

(iii) $\sum\limits_{j=1}^{n+m} a_{ij} \ge 0, \quad i = 1, ..., n,$

(iv) There exists an index $i_0 \in \{1, \dots, n\}$ for which $\sum\limits_{j=1}^{n} a_{i_0,j} > 0.$

(v) $\mathbf{A}$ is irreducible.

Remark 4.1. In the original definition in [7, p. 343], it is assumed instead of the above properties (iv)–(v) that the principal block $\mathbf{A}$ is irreducibly diagonally dominant. However, the latter follows directly from Definition 4.2 under the given sign conditions on a_{ij}.

We also note that a well-known theorem [32, p. 85] implies in this case that $\mathbf{A}^{-1} > 0$, i.e., the entries of the matrix $\mathbf{A}^{-1}$ are positive.

The known results on various discrete maximum principles (e.g., [7, 9, 16, 22]) are essentially based on the following theorem:

Theorem 4.1. *Let $\bar{\mathbf{A}}$ be a $(n+m)\times(n+m)$ matrix with the structure (34), and assume that $\bar{\mathbf{A}}$ is of generalized nonnegative type in the sense of Definition 4.2.*

If the vector $\bar{\mathbf{c}} = (c_1, ..., c_{n+m}) \in \mathbf{R}^{n+m}$ is such that $(\bar{\mathbf{A}}\bar{\mathbf{c}})_i \le 0, \ i = 1, ..., n,$ then

$$\max_{i=1,\dots,n+m} c_i \ \le \ \max\{0, \max_{i=n+1,\dots,n+m} c_i\}. \tag{35}$$

If, in addition,

$$\sum_{j=1}^{n+m} a_{ij} = 0, \qquad i = 1, ..., n, \tag{36}$$

then

$$\max_{i=1,\dots,n+m} c_i \ = \ \max_{i=n+1,\dots,n+m} c_i. \tag{37}$$

Proof. As stated in Remark 4.1, $\mathbf{A}$ is irreducibly diagonally dominant. This, together with (i)–(iii), implies both statements (35) and (37), see [7, Th. 3] and [16, Th. 3], respectively. ■

Concerning DMPs for standard boundary value problems, up to our knowledge, the most general case has been considered in our paper [16]. Thereby the following general nonlinear problem with mixed boundary conditions has been considered, given in weak form:

$$u^* \ = \ g \quad \text{on } \Gamma_D \qquad \text{and} \tag{38}$$

(39)
$$\int_\Omega \left[b(x,\nabla u^*)\,\nabla u^*\cdot\nabla v + q(x,u^*)v\right]dx + \int_{\Gamma_N} s(x,u^*)v\,d\sigma = \int_\Omega f v\,dx + \int_{\Gamma_N} \gamma v\,d\sigma, \forall v \in H^1_D(\Omega),$$

with similar growth conditions as given for (1) in the present paper. For the finite element discretization of (38)–(39), defined similarly to Section 3 above, it has been proved that the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ satisfies five properties, equivalent to the generalized nonnegative type in Definition 4.2. Based on Theorem 4.1, the corresponding DMP has been derived in [16]. (See also [17] for problems with suitably modified coefficients.)

4.2. Algebraic conditions for the discrete maximum principle. The following theorem gives a general result, which will allow us to derive various forms of the discrete maximum principle. The sign condition (40) is similar to the one given in [9, 16].

Theorem 4.2. *Let Assumptions 2.1 hold. Let us consider a family of simplicial triangulations $\mathcal{F} = \{\mathcal{T}_h\}_{h\to 0}$ satisfying the following property: for any $\mathcal{T}_h \in \mathcal{F}$ and any $i = 1,...,n,\ j = 1,...,\bar{n}\ (i \neq j)$, we have*

$$\nabla\phi_i \cdot \nabla\phi_j \le -\frac{\sigma_0}{h^2} < 0 \tag{40}$$

on supp $\phi_i \cap$ supp ϕ_j, where $\sigma_0 > 0$ is independent of $\mathcal{T}_h$, and i, j.

(A) *Let the family of triangulations $\mathcal{F}$ be strongly regular, i.e., there exist constants $c_1, c_2 > 0$ such that for any $\mathcal{T}_h \in \mathcal{F}$ and any simplex $T \in \mathcal{T}_h$ we have*

$$c_1 h^d \le \mathrm{meas}_d(T) \le c_2 h^d\,, \tag{41}$$

where meas_d *denotes d-dimensional measure. Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (32) is of generalized nonnegative type in the sense of Definition 4.2.*

(B) *More generally, for statement* (A) *to hold, it suffices to assume instead of (41) that the family $\mathcal{F}$ is only quasi-regular in the following sense: the left-hand side of (41) is replaced by*

$$c_1 h^\gamma \le \mathrm{meas}_d(T) \tag{42}$$

with some $\gamma \ge d$ satisfying

$$2 \le \gamma < 3 \quad \text{if} \quad d = 2, \qquad 3 \le \gamma < \min\{\tfrac{12}{p_1 - 2},\ 5 - \tfrac{p_2}{2}\} \quad \text{if} \quad d = 3 \tag{43}$$

(or in general, $d \le \gamma < \min\{\frac{4d}{(p_1-2)(d-2)},\ 3 + \frac{(4-p_2)(d-2)}{2}\}$ if $d \ge 3$) where p_1 and p_2 are defined in Assumptions 2.1, (A4).

Proof. By Proposition 2.1, problem (1) can be brought to the weak form (4)–(5). Therefore the proof in [16, Theorem 8] for problem (38)–(39) can be adapted if the Neumann boundary Γ_N in [16] is replaced by the interface Γ in (1). ∎

Theorem 4.2 enables us to derive the discrete maximum principle for system (30):

Theorem 4.3. *Let the conditions of Theorem 4.2 hold and let*

$$f(x) - q(x,0) \le 0, \ \ x \in \Omega, \qquad \text{and} \qquad \gamma(x) - s(x,0) \le 0, \ \ x \in \Gamma. \tag{44}$$

Then

$$\max_{\overline{\Omega}} u_h \le \max\{0, \max_{\partial\Omega} g_h\}. \tag{45}$$

In particular, if $g \ge 0$, then $\max_{\overline{\Omega}} u_h = \max_{\partial\Omega} g_h$, and if $g \le 0$, then we have the nonpositivity property $\max_{\overline{\Omega}} u_h \le 0$.

Proof. Theorem 4.2 states that the condition of Theorem 4.1 is satisfied with $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ and $\bar{n}$ substituted for $\bar{\mathbf{A}}$ and $n+m$, respectively. Further, (44) yields that $(\bar{\mathbf{A}}(\bar{\mathbf{c}})\bar{\mathbf{c}})_i \le 0$ for all i. Hence (35) yields

$$\max_{i=1,\dots,\bar{n}} c_i \le \max\{0, \max_{i=n+1,\dots,\bar{n}} c_i\}. \tag{46}$$

Since $c_i = g_i$ for all $i = n+1, \dots, \bar{n}$, estimate (46) is equivalent to (45). ∎

In analogy the following minimum principle for system (30) can be verified in the same way.

Theorem 4.4. *Let the conditions of Theorem 4.2 hold and let*

$$f(x) - q(x,0) \ge 0 \quad (x \in \Omega) \qquad \text{and} \quad \gamma(x) - s(x,0) \ge 0 \quad (x \in \Gamma). \tag{47}$$

Then we have

$$\min_{\overline{\Omega}} u_h \ge \min\{0, \min_{\partial\Omega} g_h\}. \tag{48}$$

In particular, if $g \le 0$, then $\min_{\overline{\Omega}} u_h = \min_{\partial\Omega} g_h$, and if $g \ge 0$, then we have the nonnegativity property $\min_{\overline{\Omega}} u_h \ge 0$.

Let us now consider the special case $q \equiv 0$ and $s \equiv 0$. Then the counterpart of Theorem 2.3 is valid, which we now formulate for both the maximum and minimum principles. Moreover, the strict negativity in (40) can be replaced by the weaker nonnegativity property, regularity conditions on the mesh like (42)–(43) are not required, and the result for a proper mesh holds for all parameters h instead of only sufficiently small h.

Theorem 4.5. *Let us consider the following special case of problem (1):*

$$\left\{ \begin{aligned} -\operatorname{div}\Big(b(x,\nabla u)\,\nabla u\Big) &= f(x) \quad \text{in } \Omega \setminus \Gamma, \\ [u]_\Gamma &= 0 \qquad \text{on } \Gamma, \\ \big[\, b(x,\nabla u)\tfrac{\partial u}{\partial \nu}\big]_\Gamma &= \gamma(x) \quad \text{on } \Gamma, \\ u &= g(x) \quad \text{on } \partial\Omega, \end{aligned} \right. \tag{49}$$

Let (A1)–(A3) of Assumptions 2.1 hold and let the triangulation $\mathcal{T}_h$ satisfy the following property: for any $i = 1, \dots, n$, $j = 1, \dots, \bar{n}$ $(i \ne j)$

$$\nabla\phi_i \cdot \nabla\phi_j \le 0. \tag{50}$$

Then the following results hold:

(A) *If $f \le 0$ and $\gamma \le 0$, then $\max_{\overline{\Omega}} u_h = \max_{\partial\Omega} g_h$.*

(B) *If $f \ge 0$ and $\gamma \ge 0$, then $\min_{\overline{\Omega}} u_h = \min_{\partial\Omega} g_h$.*

(C) *If $f=0$ and $\gamma=0$, then the ranges of u_h and g_h coincide, i.e., we have* $[\min_{\overline{\Omega}} u_h, \max_{\overline{\Omega}} u_h] = [\min_{\partial\Omega} g_h, \max_{\partial\Omega} g_h]$ *for the corresponding intervals.*

Proof. (A) The conditions of Theorem 4.1 follow similarly as in Theorem 4.2. The difference arises in proving property (ii), i.e., $a_{ij}(\bar{\mathbf{c}}) \le 0$, where only (50) is sufficient, since the assumptions $q \equiv 0$ and $s \equiv 0$ imply $r \equiv 0$ and $z \equiv 0$, see (27). In order to apply statement (37) of Theorem 4.1, it remains to verify that $\sum_{j=1}^{\bar{n}} a_{ij}(\bar{\mathbf{c}}) = 0,\ i = 1, ..., n$. Since $r \equiv 0$ and $z \equiv 0$, this follows indeed from the definition of $a_{ij}(\bar{\mathbf{c}})$. Statement (B) follows from (A) by replacing u by $-u$, and (C) is a direct consequence of (A) and (B). ∎

Remark 4.2. Conditions (40) and (50) can be in fact relaxed such that $\nabla\phi_i \cdot \nabla\phi_j$ need not be negative resp. nonpositive on each element, see [16, Remark 6] for details. A sufficient and necessary condition for $\nabla\phi_i \cdot \nabla\phi_j \le 0$ is given in [35].

4.3. Geometric conditions on the mesh. The conditions in the preceding subsection that guarantee the DMP have apparent geometric interpretations for simplicial meshes. This relies on the fact that the values $\nabla\phi_i \cdot \nabla\phi_j$ are constant on each simplicial element, hence conditions (40) and (50) are, in general, not very difficult to check. Indeed, it is shown in [4, 35] that

$$\nabla\phi_i \cdot \nabla\phi_j|_T = -\frac{\mathrm{meas}_{d-1}(S_i)\cdot \mathrm{meas}_{d-1}(S_j)}{d^2(\mathrm{meas}_d(T))^2}\cos(S_i,S_j) \quad \text{for } i \ne j, \tag{51}$$

where T is a d-dimensional simplex with vertices $P_1, \dots, P_{d+1}$, S_i is the face of T opposite to P_i, and $\cos(S_i, S_j)$ is the cosine of the interior angle between faces S_i and S_j.

Thus, in order to satisfy condition (40) or (50), it is sufficient if the employed simplicial triangulations (meshes) are uniformly acute (that is, any angles between adjacent $(d-1)$-dimensional simplicial faces of any triangulations are bounded away from $\frac{\pi}{2}$ by a positive constant) or nonobtuse, respectively. (See [5, 19, 20, 24] for discussion of these questions from a practical point of view, where also mesh refinement procedures preserving the above-mentioned geometrical properties are discussed). When the growth of nonlinearities are bounded, i.e. $p_1 = p_2 = 2$ in (3), then b, z, r are bounded in (25), i.e. the equation behaves like a linear one, and hence one may derive a similar explicit connection between the acceptable angles of the adjacent faces and the mesh widths as in [5]. We note that the conditions of acuteness or nonobtuseness are sufficient but not necessary: as referred to in Remark 4.2, the DMP may still hold if some obtuse interior angles occur in the simplices of the meshes. This is analogous to the case of linear problems [21, 30]. We note that most probably our results cannot be easily extended to the case of meshes consisting of block elements, especially in higher dimensions, see [17, Sect. 5.2] for a relevant discussion.

4.4. Some applications to model problems. We quote three examples of problems where suitable discrete maximum and minimum principles or, in particular, discrete nonnegativity or nonpositivity properties are valid.

4.4.1. Semilinear equations: reaction-diffusion problems with localized autocatalytic chemical reactions. The problem

$$\left\{\begin{aligned} -\Delta u &= f(x) && \text{in } \Omega \setminus \Gamma, \\ [u]_\Gamma &= 0 && \text{on } \Gamma, \\ \left[\tfrac{\partial u}{\partial \nu}\right]_\Gamma + s(x,u) &= 0 && \text{on } \Gamma, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned}\right. \tag{52}$$

in a planar domain $\Omega \subset \mathbf{R}^2$ describes a chemical reaction-diffusion process where the reaction is localized at the curve Γ, further, the reaction is autocatalytic, i.e., the growth of the concentration $u \geq 0$ speeds up the rate of the reaction, that is $\frac{\partial s(x,u)}{\partial u} \geq 0$ (see, e.g., [14, 15]). The reaction function s grows polynomially in u. The fact that there is no reaction without material is expressed by $s(x,0) = 0$, further, we may assume that the source term f is nonnegative. These conditions imply that the requirement $u \geq 0$ is satisfied, see subsection 2.4, moreover, the boundary conditions yield $\min\limits_{\overline{\Omega}} u = 0$. As a special case of Theorem 4.4, we obtain the corresponding discrete minimum principle:

Corollary 4.1. *Let u_h be the FEM solution to problem (52) under a FEM discretization with the acuteness property (40). If h is sufficiently small then*

$$\min_{\overline{\Omega}} u_h = 0.$$

4.4.2. Stefann-Boltzmann nonlinearity. The following problem arises in the study of ion charge distribution in electrolyte media, see [34] and the references therein:

$$\left\{\begin{aligned} -\operatorname{div}\Big(\varepsilon(x)\,\nabla u\Big) + C_1(x)\sinh(ku) + C_2(x) &= 0 && \text{in } \Omega \setminus \Gamma, \\ [u]_\Gamma &= 0 && \text{on } \Gamma, \\ \left[\varepsilon(x)\tfrac{\partial u}{\partial \nu}\right]_\Gamma &= 0 && \text{on } \Gamma, \\ u &= \varphi(x) && \text{on } \partial\Omega, \end{aligned}\right. \tag{53}$$

where ε is a piecewise constant eletric permittivity, $k > 0$ is constant and the coefficients C_1 and C_2 vanish in the interior/exterior of the interface, respectively. In [34] the one-dimensional case is analyzed, in particular, an analogous CMP is given. For problem (53), Theorems 4.3-4.4 imply the following discrete minimum and maximum principles, where φ_h is the corresponding approximation of φ in the used FEM subspace:

Corollary 4.2. *Let u_h be the FEM solution to problem (53) under a FEM discretization with the acuteness property (40). If h is sufficiently small then*

$$\min\{0, \min_{\partial\Omega} \varphi_h\} \leq \min_{\overline{\Omega}} u_h, \quad \max_{\overline{\Omega}} u_h \leq \max\{0, \max_{\partial\Omega} \varphi_h\}.$$

Consequently, if $\varphi_h \geq 0$ then $u_h \geq 0$, and if $\varphi_h \leq 0$ then $u_h \leq 0$.

4.4.3. Linear equations. The following linear interface model arises in many applications such as biochemistry or multiphase flow, see, e.g., [28]:

$$
\tag{54}
\begin{cases}
-\operatorname{div}\Big(k(x)\,\nabla u\Big) \;=\; f(x) & \text{in } \Omega\setminus\Gamma,\\
[u]_\Gamma = 0 & \text{on } \Gamma,\\
\big[\,k(x)\frac{\partial u}{\partial \nu}\big]_\Gamma = \gamma(x) & \text{on } \Gamma,\\
u = 0 & \text{on } \partial\Omega,
\end{cases}
$$

where the bounded measurable function k is discontinuous on Γ. In addition, it suffices to assume that k has a positive lower bound and $f \in L^2(\Omega)$, $\gamma \in L^2(\Gamma)$. Then, as a special case of Theorem 4.5, we obtain the corresponding discrete maximum and minimum principles:

Corollary 4.3. *Let u_h be the FEM solution to problem (54) under a FEM discretization with the nonobtuseness property (50).*

If $f \le 0$ *and* $\gamma \le 0$ *then* $\max_{\overline{\Omega}} u_h = 0$, *and if* $f \ge 0$ *and* $\gamma \ge 0$ *then* $\min_{\overline{\Omega}} u_h = 0$.

References

[1] Adams, R.A., *Sobolev spaces*, Academic Press, New York-London, 1975.

[2] Antal, I., Karátson, J., Mesh independent superlinear convergence of an inner-outer iterative method for semilinear elliptic interface problems, "J. Comp. Appl. Math., in press, DOI 10.1016/j.cam.2008.08.001"

[3] Bertolazzi, E., Manzini, G., A second-order maximum principle preserving finite volume method for steady convection-diffusion problems, *SIAM J. Numer. Anal.* 43 (2005), no. 5, 2172–2199.

[4] Brandts, J., Korotov, S., Křížek, M., Dissection of the path-simplex in R^n into n path-subsimplices, *Linear Algebra Appl.* 421 (2007), 382–393.

[5] Brandts, J., Korotov, S., Křížek, M., The discrete maximum principle for linear simplicial finite element approximations of a reaction-diffusion problem, *Linear Algebra Appl.* (in press).

[6] Christie, I., Hall, C., The maximum principle for bilinear elements, *Internat. J. Numer. Methods Engrg.* 20 (1984), 549–553.

[7] Ciarlet, P. G., Discrete maximum principle for finite-difference operators, *Aequationes Math.* 4 (1970), 338–352.

[8] Ciarlet, P. G., *The finite element method for elliptic problems,* North-Holland, Amsterdam, 1978

[9] Ciarlet, P. G., Raviart, P.-A., Maximum principle and uniform convergence for the finite element method, *Comput. Methods Appl. Mech. Engrg.* 2 (1973), 17–31.

[10] Faragó, I., Karátson, J., *Numerical solution of nonlinear elliptic problems via preconditioning operators. Theory and applications.* Advances in Computation, Volume 11, NOVA Science Publishers, New York, 2002.

[11] Fleishman B. A., Mahar T. J., A minimum principle for superharmonic functions subject to interface conditions, *J. Math. Anal. Appl.*, 80 (1981), 46–56.

[12] Gilbarg, D., Trudinger, N. S., *Elliptic partial differential equations of second order* (2nd edition), Grundlehren der Mathematischen Wissenschaften 224, Springer, 1983.

[13] Ishihara, K., Strong and weak discrete maximum principles for matrices associated with elliptic problems, *Linear Algebra Appl.* 88/89 (1987), 431–448.

[14] Kandilarov, J. D., A monotone iterative method for numerical solution of diffusion equations with nonlinear localized chemical reactions, *Lecture Notes in Comput. Sci.*, vol. 4310 (2007), 615–622.

[15] Kandilarov, J. D., Vulkov, L. G.., Analysis of immersed interface difference schemes for reaction-diffusion problems with singular own sources, *Comput. Methods Appl. Math.* 3 (2003), no. 2, 253–273.

[16] Karátson, J., Korotov, S., Discrete maximum principles for finite element solutions of nonlinear elliptic problems with mixed boundary conditions, *Numer. Math.* 99 (2005), 669–698.

[17] Karátson, J., Korotov, S., Křížek, M., On discrete maximum principles for nonlinear elliptic problems, *Math. Comput. Simul.* 76 (2007), 99–108.
[18] Korotov, S., Křížek, M., Finite element analysis of variational crimes for a quasilinear elliptic problem in 3D, *Numer. Math.* 84 (2000), 549–576.
[19] Korotov, S., Křížek, M., Acute type refinements of tetrahedral partitions of polyhedral domains, *SIAM J. Numer. Anal.* 39 (2001), 724–733.
[20] Korotov, S., Křížek, M., Global and local refinement techniques yielding nonobtuse tetrahedral partitions, *Comput. Math. Appl.* 50 (2005), no. 7, 1105–1113.
[21] Korotov, S., Křížek, M., Neittaanmäki, P., Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle, *Math. Comp.* 70 (2001), 107–119.
[22] Křížek, M., Lin Qun, On diagonal dominance of stiffness matrices in 3D, *East-West J. Numer. Math.* 3 (1995), 59–69.
[23] Křížek, M., Neittaanmäki, P., *Mathematical and numerical modelling in electrical engineering: theory and applications*, Kluwer Academic Publishers, 1996.
[24] Křížek, M., Šolc, J., Acute versus nonobtuse tetrahedralizations, in: *Conjugate gradient algorithms and finite element methods*, 161–170, Sci. Comput., Springer, Berlin, 2004.
[25] Ladyzhenskaya, O. A., Ural'tseva, N. N., *Linear and quasilinear elliptic equations*, Leon Ehrenpreis Academic Press, New York-London, 1968.
[26] LeVeque, R. J., Li, Zh., The immersed interface method for elliptic equations with discontinuous coefficients and singular sources, *SIAM J. Numer. Anal.* 31 (1994), no. 4, 1019–1044.
[27] Li, Zh., A fast iterative algorithm for elliptic interface problems, *SIAM J. Numer. Anal.* 35 (1998), no. 1, 230–254.
[28] Li, Zh., Ito, K., Maximum principle preserving schemes for interface problems with discontinuous coefficients, *SIAM J. Sci. Comput.* 23 (2001), no. 1, 339–361.
[29] Protter, M. H., Weinberger, H. F., *Maximum principles in differential equations*, Springer-Verlag, New York, 1984.
[30] Ruas Santos, V., On the strong maximum principle for some piecewise linear finite element approximate problems of non-positive type, *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 29 (1982), 473–491.
[31] Šolín, P., Vejchodský, T., A weak discrete maximum principle for *hp*-FEM, *J. Comp. Appl. Math.* 209 (2007), 54–65.
[32] Varga, R., *Matrix iterative analysis*, Prentice Hall, New Jersey, 1962.
[33] Vejchodský, T., Šolín, P., Discrete maximum principle for higher-order finite elements in 1D, *Math. Comp.* 76 (2007), no. 260, 1833–1846.
[34] Vulkov, L., Well-posedness and a monotone iterative method for a nonlinear interface problem on disjoint intervals, *Amer. Inst. Physics Conf. Proceedings* 946 (2007).
[35] Xu, J., Zikatanov, L., A monotone finite element scheme for convection-diffusion equations, *Math. Comp.* 68 (1999), 1429–1446.

ELTE University, Department of Applied Analysis, H-1518 Budapest, Pf. 120, Hungary, Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, FI–02015 TKK, Finland

E-mail: `karatson@cs.elte.hu and sergey.korotov@hut.fi`