Helsinki University of Technology Institute of Mathematics Research Reports

Espoo 2009 A570

DISCRETE MAXIMUM PRINCIPLES FOR FEM SOLUTIONS OF NONLINEAR ELLIPTIC SYSTEMS

János Karátson Sergey Korotov

TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Helsinki University of Technology Institute of Mathematics Research Reports
Espoo 2009 A570

DISCRETE MAXIMUM PRINCIPLES FOR FEM SOLUTIONS OF NONLINEAR ELLIPTIC SYSTEMS

János Karátson Sergey Korotov

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Mathematics and Systems Analysis

János Karátson, Sergey Korotov: *Discrete maximum principles for FEM solutions of nonlinear elliptic systems*; Helsinki University of Technology Institute of Mathematics Research Reports A570 (2009).

Abstract: *The discrete maximum principle (DMP) is an important qualitative property of various discretized elliptic equations. Conditions that ensure the DMP have drawn much attention, including geometric properties for FEM discretizations. This chapter starts with a brief summary of some background on the DMP, including the algebraic case, and nonobtuseness or acuteness type conditions for FEM. When lower order terms are included in the operator, the DMP can be ensured for sufficiently fine mesh, under uniform acuteness or strict non-narrowness in the case of simplicial or rectangular FEM meshes, respectively. (Similar conditions also appear for prismatic FEM.) Our main interest is formed by nonlinear elliptic systems under standard linear or bilinear FEM discretizations. We first present our previous results on systems with second and zeroth order terms, then extend them to the case involving first order terms. The presentation includes a detailed exposition of the required theory, which needs a generalization of the usual underlying algebraic DMP and some Hilbert space background. The geometric properties of the FEM mesh are also discussed. In many applications the DMP implies (or reduces to) a natural requirement of nonnegativity for the approximations of the corresponding nonnegative physical quantities. Such applications are given to reaction-diffusion processes and diffusion-dominated transport systems, respectively.*

AMS subject classifications: 65N50, 65M50

Keywords: elliptic system, finite element method, discrete maximum principle

Correspondence

Department of Applied Analysis and Computational Mathematics
Eötvös Loránd University, H-1518, Budapest, Pf. 120, Hungary

Institute of Mathematics, Helsinki University of Technology
P.O. Box 1100, FI-02015 TKK, Finland

karatson@cs.elte.hu, sergey.korotov@hut.fi

ISBN 978-951-22-9923-2 (print)
ISBN 978-951-22-9924-9 (PDF)
ISSN 0784-3143 (print)
ISSN 1797-5867 (PDF)

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Mathematics and Systems Analysis
P.O. Box 1100, FI-02015 TKK, Finland
email: math@tkk.fi http://math.tkk.fi/

1 Introduction

The maximum principle forms an important qualitative property of second order elliptic equations [40, 46]. Therefore its discrete analogues, the so-called discrete maximum principles (DMPs) have drawn much attention. Various DMPs, including geometric conditions on the computational meshes for FEM solutions, have been given e.g. in [4, 7, 8, 10, 11, 12, 19, 24, 33, 37, 38, 41, 47, 51, 53] for linear and [6, 25, 26, 27, 35] for nonlinear problems. For elliptic operators with only principal part, if the discretized operator L_h and the FEM solution u_h satisfy $L_h u_h \leq 0$, then the DMP has the simple form $\max\limits_{\overline{\Omega}} u_h = \max\limits_{\partial\Omega} u_h$. On the other hand, for operators with lower order terms as well, one has the weaker statement

$$\max_{\overline{\Omega}} u_h \leq \max\{0, \max_{\partial\Omega} u_h\}. \tag{1}$$

Also, in the latter case one can only provide the DMP for sufficiently fine mesh and needs stronger acuteness type conditions in the case of standard simplicial FEM meshes. Similar conditions on the shape and size of meshes also appear in the case when bilinear or prismatic finite elements are used [19].

The DMP has been extended to systems for the first time in [26]. The class of systems considered there has a coupling which is cooperative and weakly diagonally dominant, these conditions on the coupling also appear in the underlying continuous maximum principle [9, 16, 42, 43]. In the case of mixed boundary conditions and nonpositive right-hand sides, we have

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \leq \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} u_k^h\} \tag{2}$$

where Γ_D is the Dirichlet boundary and k is the number of equations. The acuteness type conditions for simplicial FE meshes have also been suitably weakened in [26].

This chapter is devoted to the DMP for elliptic systems of general type. Its goal is twofold. First, after giving a proper background including algebraic properties and a suitable Hilbert space theory, we summarize our previous results on elliptic systems with second and zeroth order terms. Then, based on these, we develop various new results on systems which are regularly perturbed by first order terms, i.e. contain non-dominating convection type terms. Our general goal is to ensure (2). Further, in many applications the DMP reduces to the natural requirement of nonnegativity for the appropriate discrete quantities, hence this will be also addressed. Some applications are mentioned briefly to reaction-diffusion processes and transport systems, respectively.

We note that the DMP for a single nonsymmetric equation has been studied extensively in the last three decades, see e.g. the early papers [28, 44, 45], the monograph [23] and the references therein. Here a major issue is to provide a DMP for singularly perturbed (i.e. convection-dominated) problems, along with the construction of suitably stabilized Galerkin methods, see e.g. [13, 14, 29, 39, 48, 50]. The present chapter is devoted to regularly perturbed problems, with focus on the generalization of the standard DMP for Galerkin methods to systems; extension of these results to stabilized Galerkin methods may be the subject of further research.

2 Discrete maximum principles in different settings

2.1 Algebraic background and the 'matrix maximum principle'

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

$$\bar{\mathbf{A}}\bar{\mathbf{c}} = \bar{\mathbf{d}}, \tag{3}$$

where the matrix $\bar{\mathbf{A}}$ and the vectors $\bar{\mathbf{d}}$, $\bar{\mathbf{c}}$ have the following structure:

$$\bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & \tilde{\mathbf{A}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \qquad \bar{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{d}} \end{bmatrix}, \qquad \bar{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix} \tag{4}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix and $\mathbf{0}$ is the $m \times k$ zero matrix. Then (3) becomes

$$\begin{bmatrix} \mathbf{A} & \tilde{\mathbf{A}} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{d}} \end{bmatrix}. \tag{5}$$

First we recall a basic definition in the study of DMP (cf. [52, p. 23]):

Definition 2.1 A square $k \times k$ matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{k}$ is called *irreducibly diagonally dominant* if it satisfies the following conditions:

(i) $\mathbf{A}$ is irreducible, i.e., for any $i \neq j$ there exists a sequence of nonzero entries $\{a_{i,i_1}, a_{i_1,i_2}, \dots, a_{i_s,j}\}$ of A, where $i, i_1, i_2, \dots, i_s, j$ are distinct indices,

(ii) $\mathbf{A}$ is diagonally dominant, i.e., $|a_{ii}| \geq \sum\limits_{\substack{j=1\\ j\neq i}}^{k} |a_{ij}|,\ i = 1, ..., k,$

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

$$|a_{i_0,i_0}| > \sum_{\substack{j=1\\ j\neq i_0}}^{k} |a_{i_0,j}|.$$

Following [11], we introduce

Definition 2.2 A $(k+m) \times (k+m)$ matrix $\bar{\mathbf{A}}$ with the structure (4) is said to be of *generalized nonnegative type* if the following properties hold:

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

(ii) $a_{ij} \leq 0, \quad i = 1, ..., k, \ j = 1, ..., k+m \quad (i \neq j),$

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

(iv) There exists an index $i_0 \in \{1, \dots, k\}$ for which

$$\sum_{j=1}^{k} a_{i_0,j} > 0. \tag{6}$$

Remark 2.1 In the original definition in [11, p. 343], it is assumed instead of the above property (iv) that the principal block $\mathbf{A}$ is irreducibly diagonally dominant. However, if we assume that $\mathbf{A}$ is also irreducible, as will be done in Theorem 2.1, then its irreducibly diagonal dominance follows directly from Definition 2.2 under the given sign conditions on a_{ij}. We also note that a well-known theorem [52, p. 85] implies in this case that $\mathbf{A^{-1}} > 0$, i.e., the entries of the matrix $\mathbf{A^{-1}}$ are positive.

Many known results on various discrete maximum principles are based on the following theorem, considered as 'matrix maximum principle' (for a proof, see e.g. [11, Th. 3]).

Theorem 2.1 *Let $\bar{\mathbf{A}}$ be a $(k+m) \times (k+m)$ matrix with the structure (4), and assume that $\bar{\mathbf{A}}$ is of generalized nonnegative type in the sense of Definition 2.2, further, that $\mathbf{A}$ is irreducible.*

If the vector $\bar{\mathbf{c}} = (c_1, ..., c_{k+m})^T \in \mathbf{R}^{k+m}$ (where $(\, . \,)^T$ denotes the transposed) is such that $(\bar{\mathbf{A}}\bar{\mathbf{c}})_i \leq 0, \ i = 1, ..., k,$ then

$$\max_{i=1,...,k+m} c_i \leq \max\{0, \max_{i=k+1,...,k+m} c_i\}. \tag{7}$$

The irreducibility of $\mathbf{A}$ is a technical condition which is sometimes difficult to check in applications, see e.g. [15, 20]. As shown in [26], it can be omitted from the assumptions if (iv) is suitably strengthened. This requires two definitions.

Definition 2.3 Let $\mathbf{A}$ be an arbitrary $k \times k$ matrix. The *irreducible blocks* of $\mathbf{A}$ are the matrices $\mathbf{A}^{(l)} \quad (l = 1, \dots, q)$ defined as follows.

Let us call the indices $i, j \in \{1, \dots, k\}$ *connectible* if there exists a sequence of nonzero entries $\{a_{i,i_1}, a_{i_1,i_2}, \dots, a_{i_s,j}\}$ of $\mathbf{A}$, where $i, i_1, i_2, \dots, i_s, j \in \{1, \dots, k\}$ are distinct indices. Further, let us call the indices i, j mutually connectible if both i, j and j, i are connectible in the above sense. (Clearly, mutual connectibility is an equivalence relation.) Let $N_1, \dots, N_q$ be the equivalence classes, i.e. the maximal sets of mutually connectible indices. (Clearly, $\mathbf{A}$ is irreducible iff $q = 1$.) Letting $N_l = \{s_1^{(l)}, \dots, s_{k_l}^{(l)}\}$ for $l = 1, \dots, q$, we have $k_1 + \dots + k_q = k$. Then we define for all $l = 1, \dots, q$ the $k_l \times k_l$ matrix $\mathbf{A}^{(l)}$ by $\mathbf{A}_{pq}^{(l)} := a_{s_p^{(l)}, s_q^{(l)}} \quad (p, q = 1, \dots, k_l).$

Remark 2.2 One may prove (cf. [1, Th. 4.2]) that by a proper permutation of indices, $\mathbf{A}$ becomes a block lower triangular matrix with the irreducible diagonal blocks $\mathbf{A}^{(l)}$.

Definition 2.4 A $(k+m) \times (k+m)$ matrix $\bar{\mathbf{A}}$ with the structure (4) is said to be of *generalized nonnegative type with irreducible blocks* if properties (i)-(iii) of Definition 2.2 hold, further, property (iv) therein is replaced by the following stronger one:

(iv') For each irreducible component of $\mathbf{A}$ there exists an index $i_0 = i_0(l) \in N_l = \{s_1^{(l)}, \ldots, s_{k_l}^{(l)}\}$ for which $\sum\limits_{j=1}^{k} a_{i_0,j} > 0$.

Remark 2.3 Let assumptions (i)-(iii) hold in Definitions 2.2 or 2.4. Then for a given index $i_0 \in \{1, \ldots, k\}$, a sufficient condition for (6) to hold is that:

there exists an index $j_0 \in \{k+1, \ldots, k+m\}$ for which $a_{i_0,j_0} < 0$.

Namely, using also assumptions (ii) and (iii), respectively, we then have

$$\sum_{j=1}^{k} a_{i_0,j} > \sum_{j=1}^{k} a_{i_0,j} + a_{i_0,j_0} \geq \sum_{j=1}^{k} a_{i_0,j} + a_{i_0,j_0} + \sum_{\substack{j=k+1 \\ j \neq j_0}}^{k+m} a_{i_0,j} = \sum_{j=1}^{k+m} a_{i_0,j} \geq 0.$$

Theorem 2.2 [26]. *Let $\bar{\mathbf{A}}$ be a $(k+m) \times (k+m)$ matrix with the structure (4), and assume that $\bar{\mathbf{A}}$ is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.*

If the vector $\bar{\mathbf{c}} = (c_1, ..., c_{k+m})^T \in \mathbf{R}^{k+m}$ is such that $d_i \equiv (\bar{\mathbf{A}}\bar{\mathbf{c}})_i \leq 0,\ i = 1, ..., k$, then (7) holds.

Consequently, in what follows, our main goal is to provide the stiffness matrix of the problems considered to be of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

2.2 Some motivation for the DMP

2.2.1 Linear equations and continuous maximum principles

First we recall the *(continuous) maximum principle* (CMP) as it usually stands for linear second order elliptic problems. Let L denote the following linear operator, acting on smooth functions defined in a bounded domain Ω:

$$Lu \equiv -\operatorname{div}\Big(a(x)\,\nabla u\Big) + h(x)u, \tag{8}$$

where the coefficients $a \in C^1(\Omega)$ and $h \in C(\Omega)$ are such that $0 < \mu_0 \leq a(x) \leq \mu_1$ and $0 \leq h(x) \leq \mu_1$ with positive constants μ_0 and μ_1 independent of $x \in \Omega$. Further, we assume that $\Omega \subset \mathbf{R}^d, d = 2, 3, ...$, has a piecewise smooth and Lipschitz continuous boundary $\partial\Omega$. The following basic result is found e.g. in [18, 40].

Theorem 2.3 *Let $u \in C^2(\Omega) \cap C(\overline{\Omega})$ be such that $Lu \le 0$ in Ω, then*

$$\max_{\overline{\Omega}} u \le \max\{0, \max_{\partial\Omega} u\}. \tag{9}$$

If, in addition, $h \equiv 0$, then

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u. \tag{10}$$

Theorem 2.3 is also valid for more general differential operators, but we shall only use the operators in the form (8) in what follows. In the context of boundary value problems, we immediately obtain from Theorem 2.3 the following result:

Corollary 2.1 *Let $u \in C^2(\Omega) \cap C(\overline{\Omega})$ be a solution of the problem*

$$\begin{cases} Lu = f & \text{in } \Omega, \\ u = g & \text{on } \partial\Omega \end{cases} \tag{11}$$

where $g \in C(\partial\Omega)$. If $f \le 0$ in Ω, then

$$\max_{\overline{\Omega}} u \le \max\{0, \max_{\partial\Omega} g\}. \tag{12}$$

If, in addition, $h \equiv 0$, then

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} g. \tag{13}$$

The analogous *(continuous) minimum principles* can be immediately formulated by changing the sign condition (i.e., replacing u by $-u$). In this contribution we will be interested in the weak form (12) of the CMP.

Let us now consider a discretization of problem (11). The most widespread methods in this context are the finite element method (FEM) and finite difference method (FDM). Both of these discretizations normally lead to linear algebraic systems of the form (5), where the block decomposition corresponds to interior and boundary mesh points, respectively. For such discretizations, the goal is to ensure that the 'matrix maximum principle' (7) holds, i.e. to apply Theorem 2.1.

In this contribution we are mostly interested in FEM discretizations on simplices. Simplicial elements are most popular and present a basic special case of the FEM, because we can treat many complicated geometries with simplices. We emphasize here an important property related to FEM. Under standard assumptions, see later (66)-(67) (which hold e.g. for usual linear, bilinear or prismatic finite elements), statement (7) directly means that

$$\max_{\overline{\Omega}} u^h \le \max\{0, \max_{\partial\Omega} g^h\} \tag{14}$$

for the discrete solution u^h of the boundary value problem (11). (Here g^h is the linear interpolant of g.) That is, the exact analogue of (12) is valid.

The main conditions that arise in this context are nonobtuseness (for problems with only principal part, i.e. $h \equiv 0$) or uniform acuteness conditions (for problems with lower order term) on the mesh. These conditions originate from the early papers [11, 12], see also [35]. Such geometric conditions will be discussed briefly in subsection 2.3.

When formulating discrete maximum principles, we will be interested in families of meshes and not in a given (single) mesh. Hence the following notion will be crucial for our study:

Definition 2.5 A set of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ is said to be a *family of FEM subspaces* if for any $\varepsilon > 0$ there exists $V_h \in \mathcal{V}$ with $h < \varepsilon$.

When the results are formulated in terms of simplicial FE meshes, one similarly defines the families of meshes $\mathcal{T} = \{\mathcal{T}_h\}_{h\to 0}$.

2.2.2 The DMP for a single nonlinear elliptic equation

The DMP for mixed nonlinear boundary value problems was first proved in [25]. Let us consider the problem

$$\left\{ \begin{aligned} -\operatorname{div}\Big(b(x,\nabla u)\,\nabla u\Big) + q(x,u) &= f(x) \quad \text{in } \Omega, \\ b(x,\nabla u)\tfrac{\partial u}{\partial \nu} + s(x,u) &= \gamma(x) \quad \text{on } \Gamma_N, \\ u &= g(x) \quad \text{on } \Gamma_D, \end{aligned} \right. \tag{15}$$

where Ω is a bounded domain in $\mathbf{R}^d$, under the following

Assumptions 2.2.2.

(A1) Ω has a piecewise smooth and Lipschitz continuous boundary $\partial\Omega$; $\Gamma_N, \Gamma_D \subset \partial\Omega$ are measurable open sets, such that $\Gamma_N \cap \Gamma_D = \emptyset$ and $\overline{\Gamma}_N \cup \overline{\Gamma}_D = \partial\Omega$.

(A2) The scalar functions $b : \overline{\Omega} \times \mathbf{R}^d \to \mathbf{R}$, $q : \overline{\Omega} \times \mathbf{R} \to \mathbf{R}$ and $s : \overline{\Gamma}_N \times \mathbf{R} \to \mathbf{R}$ are continuously differentiable in their domains of definition. Further, $f \in L^2(\Omega)$, $\gamma \in L^2(\Gamma_N)$ and $g = {g^*}_{|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

(A3) The function b satisfies

$$0 < \mu_0 \le b(x,\eta) \le \mu_1 \tag{16}$$

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 any 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 $\alpha_1 \in L^{d/2}(\Omega)$,

$\alpha_2 \in L^{d-1}(\Gamma_N)$ and a constant $\beta \geq 0$ such that for any $x \in \Omega$ (or $x \in \Gamma_N$, resp.) and $\xi \in \mathbf{R}$

$$0 \leq \frac{\partial q(x,\xi)}{\partial \xi} \leq \alpha_1(x) + \beta|\xi|^{p_1-2}, \qquad 0 \leq \frac{\partial s(x,\xi)}{\partial \xi} \leq \alpha_2(x) + \beta|\xi|^{p_2-2}. \tag{17}$$

(A5) Either $\Gamma_D \neq \emptyset$, or q increases strictly and at least linearly at ∞ in the sense that

$$q(x,\xi) \geq c_1|\xi| - c_2(x) \tag{18}$$

(with a constant $c_1 > 0$ and a function $c_2 \in L^1(\Omega)$) $\forall (x,\xi) \in \Omega \times \mathbf{R}$, or s increases strictly and at least linearly at ∞ in the same sense.

Theorem 2.4 [25]. *Let (A1)–(A5) hold and let us consider a family of simplicial FEM meshes $\mathcal{T} = \{\mathcal{T}_h\}_{h\to 0}$ satisfying the following property: for any $i = 1, ..., n,\ j = 1, ..., \bar{n}$ $(i \neq j)$, the basis functions satisfy*

$$\nabla\phi_i \cdot \nabla\phi_j \leq -\frac{\sigma_0}{h^2} < 0 \tag{19}$$

on supp $\phi_i \cap$ supp ϕ_j with $\sigma_0 > 0$ independent of i, j and h.

If the simplicial meshes $\mathcal{T}_h$ are regular, i.e., there exist constants $m_1, m_2 > 0$ such that for any $h > 0$ and any simplex $T_h \in \mathcal{T}_h$

$$m_1 h^d \leq meas(T_h) \leq m_2 h^d \tag{20}$$

(where $meas(T_h)$ denotes the d-dimensional measure of T_h), then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (46) is of generalized nonnegative type in the sense of Definition 2.2, further, $\mathbf{A}$ is irreducible.

Consequently, by Theorem 2.1, under the conditions of Theorem 2.4 and standard assumptions for the FEM mesh (see later (66)-(67)), if

$$f(x) - q(x,0) \leq 0,\ x \in \Omega, \qquad \text{and} \qquad \gamma(x) - s(x,0) \leq 0,\ x \in \Gamma_N. \tag{21}$$

then we have the DMP

$$\max_{\overline{\Omega}} u^h \leq \max\{0, \max_{\Gamma_D} g_h\}. \tag{22}$$

We note that for problems with only principal part, i.e. $q \equiv 0$ and $s \equiv 0$, it suffices to assume the weaker condition

$$\nabla\phi_i \cdot \nabla\phi_j \leq 0 \tag{23}$$

instead of (19), and we obtain the stronger DMP $\max\limits_{\overline{\Omega}} u^h = \max\limits_{\Gamma_D} g^h$.

Remark 2.4 It was also proved in [25] that, more generally, the above properties of the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ are also valid if the simplicial FE meshes $\mathcal{T}_h$ are only quasi-regular in the following sense: the left-hand side of (20) is replaced by

$$c_1 h^\gamma \leq meas(T_h)\,, \tag{24}$$

where $\gamma \geq d$ satisfies

$$\begin{aligned}
&2 \leq \gamma < 3 \quad \text{if } d = 2,\\
&3 \leq \gamma < \min\{\tfrac{12}{p_1-2},\ 5 - \tfrac{p_2}{2}\} \quad \text{if } d = 3,\\
&d \leq \gamma < \min\{\tfrac{4d}{(p_1-2)(d-2)},\ 3 + \tfrac{(4-p_2)(d-2)}{2}\} \quad \text{if } d > 3
\end{aligned} \tag{25}$$

with p_1, p_2 from assumption (A4) for problem (15).

2.3 Geometric properties to ensure the DMP

The values $\nabla\phi_i \cdot \nabla\phi_j$ are constant on each element, hence conditions (19) and (23) are not difficult to check, moreover, these conditions have a nice geometric interpretation. We briefly discuss some of these famous geometric properties, without the goal to give a detailed discussion. Some less strong assumptions will be discussed in subsection 3.4.

Conditions (19) and (23) have the following geometric meaning in view of well-known results. In order to satisfy condition (19) in the case of a simplicial mesh, it is sufficient if the employed mesh is uniformly acute, and similarly, condition (23) is satisfied if the employed mesh is nonobtuse [12, 35]. In the case of bilinear elements, condition (19) is equivalent to the so-called condition of non-narrow mesh, see [10]. The same issue for prismatic finite elements was recently treated in [19], where a convenient notion of (strictly) well-shaped prismatic partition is introduced.

We note that conditions (19) and (23) are sufficient but not necessary. For simplicial FEM, the DMP may still hold if some obtuse interior angles occur in the simplices of the meshes, i.e. if $\nabla\phi_i \cdot \nabla\phi_j$ is positive on each element. Namely, (19) was imposed to ensure the validity of the estimate

$$b_{ij}(\bar{\mathbf{c}}) = \int\limits_{\Omega_{ij}} b(x, \nabla u^h)\ \nabla\phi_i \cdot \nabla\phi_j \, dx \leq -\frac{\sigma_1}{h^2} < 0 \tag{26}$$

with $\sigma_1 > 0$ independent of i, j and h, where $\Omega_{ij} = supp\ \phi_i \cap supp\ \phi_j$. However, using (16), we have in general

$$b_{ij}(\bar{\mathbf{c}}) \leq \mu_0 \sum_{K_l \in \mathcal{K}^-} \text{meas}\,(K_l)\ \nabla\phi_i \cdot \nabla\phi_j\ + \mu_1 \sum_{K_l \in \mathcal{K}^+} \text{meas}\,(K_l)\ \nabla\phi_i \cdot \nabla\phi_j\,, \tag{27}$$

with notations

$$\mathcal{K}^- = \{K \in \mathcal{T}_h :\ \nabla\phi_i \cdot \nabla\phi_j < 0 \text{ on } K\}, \quad \mathcal{K}^+ = \{K \in \mathcal{T}_h :\ \nabla\phi_i \cdot \nabla\phi_j \geq 0 \text{ on } K\}.$$

Then, it suffices to require that the expression in (27) is estimated above by $-\sigma_1/h^2$, which may allow the set $\mathcal{K}^+$ to be nonempty in certain situations. For linear problems, such weakened acute type conditions are given in e.g. in [33, 47].

One often wishes to solve the problem on finer meshes in order to obtain a more accurate approximation. These geometric conditions then need special attention. Namely, if we propose a global refinement of the initial mesh using some refinement technique (see e.g. [31]), then we must take care that the refined mesh preserves the desired acuteness (or nonobtuseness) property. Obviously, this is an easy task in the two-dimensional case, since, using the standard "2D red refinement" [31], we obtain a mesh consisting only of acute or nonobtuse triangular elements if the initial mesh had only acute or nonobtuse triangles, respectively. If we consider a tetrahedral mesh, the task is far from being trivial since in general it is not possible to refine any tetrahedron into eight subtetrahedra similar to it using "3D red refinement" (cf. [31]). A new technique, the so-called "3D yellow refinement" was developed in [30], which allows a global refinement of a nonobtuse tetrahedral mesh so that the resulting (conforming) mesh preserves the property of nonobtuseness. For local nonobtuse refinements (also in higher dimensions), see [2] and [32]. A construction of regular meshes, using a technique different from the red-refinement by midlines, is proposed in [34].

2.4 An algebraic DMP in Hilbert space

When dealing with elliptic systems, it is useful to state an algebraic (matrix) DMP in a Hilbert space setting in order to provide a clean line of thoughts. Namely, this setting will help an organized derivation of the corresponding results under the considered different conditions. The discussion below is based on [26], where it was applied to systems with second and zeroth order terms.

2.4.1 Formulation of the operator equation

Let H be a real Hilbert space and $H_0 \subset H$ a given subspace. We consider the following operator equation: for given vectors $\psi, g^* \in H$, find $u \in H$ such that

$$\langle A(u), v\rangle = \langle \psi, v\rangle \qquad (v \in H_0) \tag{28}$$

$$\text{and} \quad u - g^* \in H_0 \tag{29}$$

with an operator $A : H \to H$ satisfying the following conditions:

Assumptions 2.4.1.

(i) The operator $A : H \to H$ has the form

$$A(u) = B(u)u + N(u)u + R(u)u \tag{30}$$

where B, N and R are given operators mapping from H to $\mathcal{B}(H)$. (Here $\mathcal{B}(H)$ denotes the set of bounded linear operators in H.)

(ii) There exists a constant $m > 0$ such that

$$\Big\langle \Big(B(u) + N(u)\Big)v, v\Big\rangle \geq m \, \|v\|^2 \qquad (u \in H, \ v \in H_0). \tag{31}$$

(iii) There exist subsets of 'positive vectors' $D, P \subset H$ such that for any $u \in H$ and $v \in D$, we have

$$\langle R(u)w, v\rangle \geq 0 \tag{32}$$

provided that either $w \in P$ or $w = v \in D$.

(iv) There exists a continuous function $M_{NR} : \mathbf{R}^+ \to \mathbf{R}^+$ and another norm $|||.|||$ on H such that

$$\Big\langle \Big(N(u) + R(u)\Big)z, v\Big\rangle \leq M_{NR}(\|u\|) \, |||z||| \, |||v||| \qquad (u, z, v \in H). \tag{33}$$

In practice for PDE problems, g^* plays the role of boundary condition and H_0 will be the subspace corresponding to homogeneous boundary conditions, further, $B(u)$ is the principal part of A.

Assumptions 2.4.1 are not in general known to imply existence and uniqueness for (28)-(29). The following extra conditions already ensure well-posedness:

Assumptions 2.4.2.

(i) The operator A is Gateaux differentiable, further, A' is bihemicontinuous (i.e. mappings $(s, t) \mapsto A'(u + sk + tw)h$ are continuous from $\mathbf{R}^2$ to H).

(ii) There exists a continuous function $M_A : \mathbf{R}^+ \to \mathbf{R}^+$ such that

$$\langle A'(u)w, v\rangle \leq M_A(\|u\|) \, \|w\| \, \|v\| \qquad (u \in H, \ w, v \in H_0). \tag{34}$$

(iii) There exists a constant $m > 0$ such that

$$\langle A'(u)v, v\rangle \geq m \, \|v\|^2 \qquad (u \in H, \ v \in H_0). \tag{35}$$

Proposition 2.1 *If Assumptions 2.4.1–2.4.2 hold, then problem (28)-(29) is well-posed.*

The proof is based on uniform monotonicity and local Lipschitz continuity, see e.g. [17]. The proof of an equivalent formulation of Proposition 2.1 is given in [26].

2.4.2 Galerkin type discretization

Let $n_0 \leq n$ be positive integers and $\phi_1, ..., \phi_n \in H$ be given linearly independent vectors such that $\phi_1, ..., \phi_{n_0} \in H_0$. We consider the finite dimensional subspaces

$$V_h = \text{span}\{\phi_1, ..., \phi_n\} \ \subset H, \qquad V_h^0 = \text{span}\{\phi_1, ..., \phi_{n_0}\} \ \subset H_0 \tag{36}$$

with a real positive parameter $h > 0$. In practice, as is usual for FEM, h is inversely proportional to n, and one will consider a family of such subspaces in the sense of Definition 2.5.

We formulate here some connectivity type properties for these subspaces that we will need later. For this, certain pairs $\{\phi_i, \phi_j\} \in V_h \times V_h$ are called '*neighbouring basis vectors*', and then i, j are called '*neighbouring indices*'. The only requirement for the set of these pairs is that they satisfy Assumptions 2.4.3 below, given in terms of the *graph of neighbouring indices*, by which we mean the following. The corresponding indices $\{1, \ldots, n_0\}$ or $\{1, \ldots, n\}$, respectively, are represented as vertices of the graph, and the ith and jth vertices are connected by an edge iff i, j are neighbouring indices.

Assumptions 2.4.3. The set $\{1, \ldots, n\}$ can be partitioned into disjoint sets $S_1, \ldots, S_r$ such that for each $k = 1, \ldots, r$,

(i) both $S_k^0 := S_k \cap \{1, \ldots, n_0\}$ and $\tilde{S}_k := S_k \cap \{n_0+1, \ldots, n\}$ are nonempty;

(ii) the graph of all neighbouring indices in S_k^0 is connected;

(iii) the graph of all neighbouring indices in S_k is connected.

(In later PDE applications, these properties are meant to express that the supports of basis functions cover the domain, both its interior and the boundary.)

Now let $g^h = \sum\limits_{j=n_0+1}^{n} g_j \phi_j \ \in V_h$ be a given approximation of the component of g^* in $H \setminus H_0$. To find the Galerkin solution of (28)-(29) in V_h, we solve the following problem: find $u^h \in V_h$ such that

$$\langle A(u^h), v^h \rangle = \langle \psi, v^h \rangle \qquad (v^h \in V_h^0) \tag{37}$$

$$\text{and} \quad u^h - g^h \in V_h^0. \tag{38}$$

Using (30), we can rewrite (37) as

$$\langle B(u^h)u^h, v^h \rangle + \langle N(u^h)u^h, v^h \rangle + \langle R(u^h)u^h, v^h \rangle = \langle \psi, v^h \rangle \qquad (v^h \in V_h^0). \tag{39}$$

Let us now formulate the nonlinear algebraic system corresponding to (39). We set

$$u^h = \sum_{j=1}^{n} c_j \phi_j, \tag{40}$$

and look for the coefficients $c_1, \ldots, c_n$. For any $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$, $i = 1, ..., n_0$ and $j = 1, ..., n$, we set

$$b_{ij}(\bar{\mathbf{c}}) := \langle B(u^h)\phi_j, \phi_i\rangle, \qquad n_{ij}(\bar{\mathbf{c}}) := \langle N(u^h)\phi_j, \phi_i\rangle, \qquad r_{ij}(\bar{\mathbf{c}}) := \langle R(u^h)\phi_j, \phi_i\rangle,$$

$$a_{ij}(\bar{\mathbf{c}}) := b_{ij}(\bar{\mathbf{c}}) + n_{ij}(\bar{\mathbf{c}}) + r_{ij}(\bar{\mathbf{c}}), \qquad d_i := \langle \psi, \phi_i\rangle. \tag{41}$$

Putting (40) and $v = \phi_i$ into (39), we obtain the $n_0 \times n$ system of algebraic equations

$$\sum_{j=1}^{n} a_{ij}(\bar{\mathbf{c}})\, c_j = d_i \qquad (i = 1, ..., n_0). \tag{42}$$

Using the notations

$$\mathbf{A}(\bar{\mathbf{c}}) := \{a_{ij}(\bar{\mathbf{c}})\},\ i, j = 1, ..., n_0, \quad \tilde{\mathbf{A}}(\bar{\mathbf{c}}) := \{a_{ij}(\mathbf{c})\},\ i = 1, ..., n_0;\ j = n_0{+}1, ..., n,$$

$$\mathbf{d} := \{d_j\},\ \mathbf{c} := \{c_j\}, \quad j = 1, ..., n_0, \quad \text{and} \quad \tilde{\mathbf{c}} := \{c_j\}, \quad j = n_0 + 1, ..., n, \tag{43}$$

system (42) turns into

$$\mathbf{A}(\bar{\mathbf{c}})\mathbf{c} + \tilde{\mathbf{A}}(\bar{\mathbf{c}})\tilde{\mathbf{c}} = \mathbf{d}. \tag{44}$$

In order to obtain a system with a square matrix, we enlarge our system to an $n \times n$ one. Since $u^h - g^h \in V_h^0$, the coordinates c_i with $n_0 + 1 \le i \le n$ satisfy automatically $c_i = g_i$, i.e.,

$$\tilde{\mathbf{c}} = \tilde{\mathbf{g}} := \{g_j\}, \quad j = n_0 + 1, ..., n,$$

hence we can replace (44) by the equivalent system

$$\begin{bmatrix} \mathbf{A}(\bar{\mathbf{c}}) & \tilde{\mathbf{A}}(\bar{\mathbf{c}}) \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \tilde{\mathbf{g}} \end{bmatrix}. \tag{45}$$

Defining further

$$\bar{\mathbf{A}}(\bar{\mathbf{c}}) := \begin{bmatrix} \mathbf{A}(\bar{\mathbf{c}}) & \tilde{\mathbf{A}}(\bar{\mathbf{c}}) \\ 0 & \mathbf{I} \end{bmatrix}, \qquad \bar{\mathbf{c}} := \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix}, \tag{46}$$

we rewrite (44) as follows:

$$\bar{\mathbf{A}}(\bar{\mathbf{c}})\bar{\mathbf{c}} = \mathbf{d}. \tag{47}$$

2.4.3 Maximum principle for the abstract discretized problem

When formulating a discrete maximum principle for system (47), the notion of family of subspaces will be used in analogy of Definition 2.5. First we give sufficient conditions for the generalized nonnegativity of the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$.

Theorem 2.5 *Let Assumptions 2.4.1 and 2.4.3 hold. Let us consider the discretization of operator equation (28)-(29) in a family of subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ with bases as in (36). Let $u^h \in V_h$ be the solution of (39) and let the following properties hold:*

(a) For all $\phi_i \in V_h^0$ and $\phi_j \in V_h$, one of the following holds: either

$$\langle B(u^h)\phi_j, \phi_i\rangle = \langle N(u^h)\phi_j, \phi_i\rangle = 0 \quad \textit{and} \quad \langle R(u^h)\phi_j, \phi_i\rangle \leq 0, \tag{48}$$

or

$$\langle B(u^h)\phi_j, \phi_i\rangle \leq -M_B(h) \tag{49}$$

with a proper function $M_B : \mathbf{R}^+ \to \mathbf{R}^+$ (independent of h, ϕ_i, ϕ_j) such that, defining

$$T(h) := \sup\{|||\phi_i||| : \ \phi_i \in V_h)\}\,, \tag{50}$$

we have

$$\lim_{h\to 0} \frac{M_B(h)}{T(h)^2} = +\infty. \tag{51}$$

(b) If, in particular, $\phi_i \in V_h^0$ and $\phi_j \in V_h$ are neighbouring basis vectors (as defined for Assumptions 2.4.3), then (49)-(51) hold.

(c) $M_{NR}(\|u^h\|)$ is bounded as $h \to 0$, where M_{NR} is the function in Assumption 2.4.1 (iv).

(d) For all $u \in H$ and $h > 0$, $\sum\limits_{j=1}^{n} \phi_j \in ker B(u) \cap ker N(u)$.

(e) For all $h > 0$, $i = 1, ..., n$, we have $\phi_i \in D$ and $\sum\limits_{j=1}^{n} \phi_j \in P$ for the sets D, P introduced in Assumption 2.4.1 (iii).

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (46) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

The proof is given in [26, Theorem 3.1] for the case $N(u) \equiv 0$. It is easy to see that the inclusion of $N(u)$ gives no difference in the proof, since $N(u)$ has been joined either to $B(u)$ or to $R(u)$ both in Assumptions 2.4.1 and in the appropriate conditions of Theorem 2.5. Hence in each step of the proof one has the same condition now with $N(u)$ as it was in [26, Theorem 3.1] without $N(u)$. (When applying Theorem 2.5 later, we will need $N(u)$ for the first order terms. Since it could not be joined either only to $B(u)$ or only to $R(u)$ above, we could not use [26, Theorem 3.1] formally in the original way.)

By Theorem 2.2, we immediately obtain the corresponding algebraic *discrete maximum principle:*

Corollary 2.2 *Let the assumptions of Theorem 2.5 hold. For sufficiently small h, if $d_i \leq 0$ $(i = 1, ..., n_0)$ in (43) and $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ is the solution of (47), then*

$$\max_{i=1,\ldots,n} c_i \leq \max\{0, \max_{i=n_0+1,\ldots,n} c_i\}. \tag{52}$$

Remark 2.5 Assumption (c) of Theorem 2.5 follows in particular if Assumptions 2.4.2 are added to Assumptions 2.4.1 as done in Proposition 2.1, provided that the functions $g^h \in V_h$ in (38) are bounded in H-norm as $h \to 0$. (In practice, the usual choices for g^h even produce $g^h \to g^*$ in H-norm.) In fact, in this case $\|u^h\|$ is bounded as $h \to 0$; then the continuity of M_{NR} yields that $M_{NR}(\|u^h\|)$ is bounded too. For more details see [26].

Remark 2.6 Assumptions 2.4.1. (iv) can be weakened such that one may allow different norms both for N and R and in the factors, i.e. (33) is replaced by

$$\langle N(u)z, v\rangle \le M_N(\|u\|)\, |||z|||_{N_1}\, |||v|||_{N_2}\,, \tag{53}$$

$$\big\langle R(u)w, v\big\rangle \le M_R(\|u\|)\, |||w|||_{R_1}\, |||v|||_{R_2} \tag{54}$$

(for all $u, w, v \in H$). Then Theorem 2.5 remains true if we appropriately replace (50) by

$$T(h) := \sup\Big\{\max\Big\{|||\phi_j|||_{N_1}|||\phi_i|||_{N_2}\,,\ \ |||\phi_j|||_{R_1}|||\phi_i|||_{R_2}\Big\} :\ \phi_i, \phi_j \in V_h\Big\}^{1/2}\,, \tag{55}$$

and require in assumption (c) that both $M_N(\|u^h\|)$ and $M_R(\|u^h\|)$ are bounded as $h \to 0$.

3 Discrete maximum principles for elliptic reaction-diffusion type systems

We first study various types of nonlinear elliptic systems with second and zeroth order terms, quoting our results from [26]. The considered domain Ω and the diffusion coefficient functions b_k ($k = 1, \dots, s$) will satisfy common properties, formulated below:

Assumptions 3.0.

(i) $\Omega \subset \mathbf{R}^d$ is a bounded piecewise C^1 domain; Γ_D, Γ_N are disjoint open measurable subsets of $\partial\Omega$ such that $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$ and $\Gamma_D \neq \emptyset$.

(ii) (Ellipticity.) There exists $m > 0$ such that $b_k \ge m$ holds pointwise for all $k = 1, \dots, s$.

3.1 Systems with nonlinear coefficients

3.1.1 Formulation of the problem

First we consider nonlinear elliptic systems of the form

$$\left.\begin{aligned} -\mathrm{div}\,\Big(b_k(x,u,\nabla u)\,\nabla u_k\Big) + \sum_{l=1}^{s} V_{kl}(x,u,\nabla u)\,u_l &= f_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\}\ (k = 1, \dots, s) \tag{56}$$

with unknown function $u = (u_1, \dots, u_s)^T$, under the following assumptions. Here ∇u denotes the $s \times d$ tensor with rows ∇u_k $(k = 1, \dots, s)$, further, 'a.e.' means Lebesgue almost everywhere and inequalities for functions are understood a.e. pointwise for all possible arguments.

Assumptions 3.1.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and boundedness.) For all $k, l = 1, \dots, s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$ and $V_{kl} \in L^\infty(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$.

(iii) (Cooperativity.) We have

$$V_{kl} \le 0 \qquad (k, l = 1, \dots, s,\ k \ne l). \tag{57}$$

(iv) (Weak diagonal dominance.) We have

$$\sum_{l=1}^{s} V_{kl} \ge 0 \qquad (k = 1, \dots, s). \tag{58}$$

(v) For all $k = 1, \dots, s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^*_k \in H^1(\Omega)$.

Remark 3.1 (i) Assumptions (57)-(58) imply

$$V_{kk} \ge 0 \qquad (k = 1, \dots, s). \tag{59}$$

(ii) One may consider additional terms on the Neumann boundary, see Remark 3.4 later.

For the weak formulation of such problems, we define the Sobolev space

$$H^1_D(\Omega) := \{z \in H^1(\Omega) :\ z_{|\Gamma_D} = 0\}. \tag{60}$$

The weak formulation of problem (56) then reads as follows: find $u \in H^1(\Omega)^s$ such that

$$\langle A(u), v\rangle = \langle \psi, v\rangle \qquad (\forall v \in H^1_D(\Omega)^s) \tag{61}$$

$$\text{and} \quad u - g^* \in H^1_D(\Omega)^s, \tag{62}$$

where

$$\langle A(u), v\rangle = \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, u, \nabla u)\, \nabla u_k \cdot \nabla v_k + \sum_{k,l=1}^{s} V_{kl}(x, u, \nabla u)\, u_l\, v_k \Big) \tag{63}$$

for given $u = (u_1, \dots, u_s) \in H^1(\Omega)^s$ and $v = (v_1, \dots, v_s) \in H^1_D(\Omega)^s$, further,

$$\langle \psi, v\rangle = \int_\Omega \sum_{k=1}^{s} f_k v_k + \int_{\Gamma_N} \sum_{k=1}^{s} \gamma_k v_k \tag{64}$$

for given $v = (v_1, \dots, v_s) \in H^1_D(\Omega)^s$, and $g^* := (g^*_1, \dots, g^*_s)$.

3.1.2 Finite element discretization

We define the finite element discretization of problem (56) in the following way. First, let $\bar{n}_0 \leq \bar{n}$ be positive integers and let us choose basis functions

$$\varphi_1, \ldots, \varphi_{\bar{n}_0} \in H^1_D(\Omega), \qquad \varphi_{\bar{n}_0+1}, \ldots, \varphi_{\bar{n}} \in H^1(\Omega) \setminus H^1_D(\Omega), \tag{65}$$

which correspond to homogeneous and inhomogeneous boundary conditions on Γ_D, respectively. (For simplicity, we will refer to them as 'interior basis functions' and 'boundary basis functions', respectively, thus adopting the terminology of Dirichlet problems even in the general case.) These basis functions are assumed to be continuous and to satisfy

$$\varphi_p \geq 0 \quad (p = 1, \ldots, \bar{n}), \qquad \sum_{p=1}^{\bar{n}} \varphi_p \equiv 1, \tag{66}$$

further, that there exist node points $B_p \in \Omega$ $(p = 1, \ldots, \bar{n}_0)$ and $B_p \in \Gamma_D$ $(p = \bar{n}_0 + 1, \ldots, \bar{n})$ such that

$$\varphi_p(B_q) = \delta_{pq} \tag{67}$$

where δ_{pq} is the Kronecker symbol; and finally, there exists a constant $c > 0$ (independent of the basis functions) such that

$$\max |\nabla \varphi_t| \leq \frac{c}{diam(\mathrm{supp}\, \varphi_t)} \tag{68}$$

where supp denotes the support, i.e. the closure of the set where the function does not vanish. These conditions hold e.g. for standard linear, bilinear or prismatic finite elements. Finally, we assume that any two interior basis functions can be connected with a chain of interior basis functions with overlapping support. By its geometric meaning, this assumption obviously holds for any reasonable FE mesh.

We in fact need a basis in the corresponding product spaces, which we define by repeating the above functions in each of the s coordinates and setting zero in the other coordinates. That is, let $n_0 := s\bar{n}_0$ and $n := s\bar{n}$. First, for any $1 \leq i \leq n_0$,

$$\text{if } \ i = (k-1)\bar{n}_0 + p \quad \text{for some } 1 \leq k \leq s \text{ and } 1 \leq p \leq \bar{n}_0, \ \text{ then}$$

$$\phi_i := (0, \ldots, 0, \varphi_p, 0, \ldots, 0) \qquad \text{where } \ \varphi_p \ \text{ stands at the } k\text{-th entry}, \tag{69}$$

that is, $(\phi_i)_m = \varphi_p$ if $m = k$ and $(\phi_i)_m = 0$ if $m \neq k$. From these, we let

$$V_h^0 := \mathrm{span}\{\phi_1, ..., \phi_{n_0}\} \ \subset H^1_D(\Omega)^s. \tag{70}$$

Similarly, for any $n_0 + 1 \leq i \leq n$, if

$$i = n_0 + (k-1)(\bar{n} - \bar{n}_0) + p - \bar{n}_0 \quad \text{for some } 1 \leq k \leq s \text{ and } \bar{n}_0 + 1 \leq p \leq \bar{n}, \text{ then}$$

$$\phi_i := (0, \ldots, 0, \varphi_p, 0, \ldots, 0)^T \qquad \text{where } \ \varphi_p \ \text{ stands at the } k\text{-th entry}, \tag{71}$$

that is, $(\phi_i)_m = \varphi_p$ if $m = k$ and $(\phi_i)_m = 0$ if $m \neq k$. From (70) and these, we let

$$V_h := \text{span}\{\phi_1, ..., \phi_n\} \ \subset H^1(\Omega)^s. \tag{72}$$

Using the above FEM subspaces, the finite element discretization of problem (56) leads to the task of finding $u^h \in V_h$ such that

$$\langle A(u^h), v^h\rangle = \langle \psi, v^h\rangle \qquad (\forall v^h \in V_h^0) \tag{73}$$

$$\text{and} \quad u^h - g^h \in V_h^0, \quad \text{i.e.,} \ \ u^h = g^h \text{ on } \Gamma_D \tag{74}$$

(where $g^h = \sum\limits_{j=n_0+1}^{n} g_j\phi_j \ \in V_h$ is the approximation of g^* on Γ_D). Then, setting $u^h = \sum\limits_{j=1}^{n} c_j\phi_j$ and $v = \phi_i$ $(i = 1, \ldots, n_0)$ in (61) (just as in (40)-(42)), we obtain the $n_0 \times n$ system of algebraic equations

$$\sum_{j=1}^{n} a_{ij}(\bar{\mathbf{c}})\, c_j = d_i \qquad (i = 1, ..., n_0), \tag{75}$$

where for any $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ and $i = 1, ..., n_0$, $j = 1, ..., n$,

$$a_{ij}(\bar{\mathbf{c}}) := \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, u^h, \nabla u^h)\,(\nabla\phi_j)_k\cdot(\nabla\phi_i)_k + \sum_{k,l=1}^{s} V_{kl}(x, u^h, \nabla u^h)\,(\phi_j)_l\,(\phi_i)_k\Big) \tag{76}$$

$$\text{and} \qquad d_i := \int_\Omega \sum_{k=1}^{s} f_k(\phi_i)_k + \int_{\Gamma_N} \sum_{k=1}^{s} \gamma_k(\phi_i)_k\,. \tag{77}$$

In the same way as for (47), we enlarge system (75) to a square one by adding an identity block, and write it briefly as

$$\bar{\mathbf{A}}(\bar{\mathbf{c}})\bar{\mathbf{c}} = \mathbf{d}\,. \tag{78}$$

That is, for $i = 1, ..., n_0$ and $j = 1, ..., n$, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ has the entry $a_{ij}(\bar{\mathbf{c}})$ from (76).

In what follows, we will need notions of (patch-)regularity of the considered FE meshes, cf. [3].

Definition 3.1 Let $\Omega \subset \mathbf{R}^d$ and let us consider a family of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ constructed as above. Here $h > 0$ is the mesh parameter, proportional to the maximal diameter of the supports of the basis functions $\phi_1, ..., \phi_n$. The corresponding family of meshes will be called

(a) *regular from above* if there exists a constant $c_0 > 0$ such that for any $V_h \in \mathcal{V}$ and basis function $\varphi_p \in V_h$,

$$meas(\text{supp}\,\varphi_p) \leq c_0 h^d \tag{79}$$

(where *meas* denotes d-dimensional measure and supp denotes the support, i.e. the closure of the set where the function does not vanish);

(b) *regular* if there exist constants $c_1, c_2 > 0$ such that for any $V_h \in \mathcal{V}$ and basis function $\varphi_p \in V_h$,

$$c_1 h^d \leq meas(\operatorname{supp} \varphi_p) \leq c_2 h^d \,; \tag{80}$$

(c) *quasi-regular* if (80) is replaced by

$$c_1 h^\gamma \leq meas(\operatorname{supp} \varphi_p) \leq c_2 h^d \tag{81}$$

for some fixed constant

$$d \leq \gamma < d + 2. \tag{82}$$

3.1.3 Discrete maximum principle for systems with nonlinear coefficients

The theory of subsection 2.4 can be applied to derive a DMP for problem (56). The underlying operators have the following properties:

Lemma 3.1 [26, Lemma 4.1]. *For any $u \in H^1(\Omega)^s$, let us define the operators $B(u)$ and $R(u)$ via*

$$\begin{aligned} \langle B(u)z, v\rangle &= \int_\Omega \sum_{k=1}^s b_k(x, u, \nabla u)\, \nabla z_k \cdot \nabla v_k \\ \langle R(u)z, v\rangle &= \int_\Omega \sum_{k,l=1}^s V_{kl}(x, u, \nabla u)\, z_l\, v_k \end{aligned} \tag{83}$$

($z \in H^1(\Omega)^s$, $v \in H^1_D(\Omega)^s$). Together with the operator A, defined in (63), the operators $B(u)$ and $R(u)$, together with $N(u) \equiv 0$, satisfy Assumptions 2.4.1 in the spaces $H = H^1(\Omega)^s$ and $H_0 = H^1_D(\Omega)^s$, and with the new norm

$$|||v|||^2 := \|v\|^2_{L^2(\Omega)^s} = \int_\Omega \sum_{k=1}^s v_k^2 \,. \tag{84}$$

Now let us consider the finite element discretization for problem (56), developed in the previous subsection. One can then derive from Theorem 2.5 the following nonnegativity result for the stiffness matrix:

Theorem 3.1 [26, Theorem 4.1]. *Let problem (56) satisfy Assumptions 3.1. Let us consider a family of finite element subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ satisfying the following property: there exists a real number γ satisfying*

$$d \leq \gamma < d + 2$$

(where d is the space dimension) such that for any $p = 1, ..., \bar{n}_0$, $t = 1, ..., \bar{n}$ $(p \neq t)$, if $meas(\operatorname{supp} \varphi_p \cap \operatorname{supp} \varphi_t) > 0$ then

$$\nabla\varphi_t \cdot \nabla\varphi_p \leq 0 \quad on\ \Omega \quad and \quad \int_\Omega \nabla\varphi_t \cdot \nabla\varphi_p \leq -K_0\, h^{\gamma-2} \tag{85}$$

with some constant $K_0 > 0$ independent of p, t and h. Further, let the family of associated meshes be regular from above, according to Definition 3.1.

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (76) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

By (77) we have $d_i \le 0 \ \ (i = 1, ..., n_0)$, hence Corollary 2.2 immediately yields

Corollary 3.1 *Let the assumptions of Theorem 3.1 hold and let $f_k \le 0$, $\gamma_k \le 0$ $(k = 1, \dots, s)$. For sufficiently small h, if $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ is the solution of (75) with matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (76), then*

$$\max_{i=1,\dots,n} c_i \le \max\{0, \max_{i=n_0+1,\dots,n} c_i\}. \tag{86}$$

The meaning of (86) is as follows. Let us split the vector $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ as in (46), i.e. $\bar{\mathbf{c}} = \begin{bmatrix} \mathbf{c} \\ \tilde{\mathbf{c}} \end{bmatrix}$, where $\mathbf{c} = (c_1, ..., c_{n_0})^T$ and $\tilde{\mathbf{c}} = (c_{n_0+1}, ..., c_n)^T$. Following the notions introduced after (65), the vectors $\mathbf{c}$ and $\tilde{\mathbf{c}}$ contain the coefficients of the 'interior basis functions' and 'boundary basis functions', respectively. Then (86) states that the maximal coordinate is nonpositive or arises for a boundary basis function.

Our main interest is the meaning of Corollary 3.1 for the FEM solution $u^h = (u_1^h, \dots, u_s^h)^T$ itself.

Theorem 3.2 [26, Theorem 4.2]. *Let the basis functions satisfy (66)-(67). If (86) holds for the FEM solution $u^h = (u_1^h, \dots, u_s^h)^T$, then u^h satisfies*

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \le \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g_k^h\}. \tag{87}$$

Thus we obtain the *discrete maximum principle* for system (56):

Corollary 3.2 *Let the assumptions of Theorem 3.1 hold and let*

$$f_k \le 0, \qquad \gamma_k \le 0 \qquad (k = 1, \dots, s).$$

Let the basis functions satisfy (66)-(67). Then for sufficiently small h, if $u^h = (u_1^h, \dots, u_s^h)^T$ is the FEM solution of system (56), then

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \le \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g_k^h\}. \tag{88}$$

Remark 3.2 (i) Let $f_k \le 0$, $\gamma_k \le 0$ for all k. The result (88) can be divided in two cases, both of which are remarkable: if at least one of the functions g_k^h has positive values on Γ_D then

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h = \max_{k=1,\dots,s} \max_{\Gamma_D} g_k^h \tag{89}$$

(which can be called more directly a discrete maximum principle than (88)), and if $g_k \le 0$ on Γ_D for all k, then we obtain the nonpositivity property

$$u_k^h \le 0 \quad \text{on } \Omega \text{ for all } k\,. \tag{90}$$

(ii) Analogously, if $f_k \ge 0$, $\gamma_k \ge 0$ for all k, then (by reversing signs) we can derive the corresponding discrete minimum principles instead of (88) and (89), or the corresponding nonnegativity property instead of (90).

Remark 3.3 The key assumption for the meshes in the above results is property (85). A simple but stronger sufficient condition to satisfy (85) is (19), provided that the family of meshes is quasi-regular according to Definition 3.1. For simplicial FEM, assumption (19) corresponds to acute triangulations. Less strong assumptions to satisfy (85) will be discussed in subsection 3.4.

Remark 3.4 The results of this section may hold as well if there are additional terms $\sum\limits_{l=1}^{s} \omega_{kl}(x,u,\nabla u)\, u_l$ on the Neumann boundary Γ_N, which we did not include for technical simplicity. Then ω_{kl} must satisfy similar properties as assumed for V_{kl} in (57)-(58).

3.2 Systems with general reaction terms of sublinear growth

In (56) both the principal and lower-order parts of the equations were given as containing products of coefficients with ∇u_k and u_l, respectively. Whereas this is widespread in real models for the principal part, the lower order terms are usually not given in such a coefficient form. Now we consider problems where the dependence on the lower order terms is given as general functions of x and u. In this section these functions are allowed to grow at most linearly, in which case one can reduce the problem to the previous one (56) directly. (Superlinear growth of q_k will be dealt with in the next section.)

Accordingly, let us now consider the system

$$\left.\begin{aligned} -\mathrm{div}\,\Big(b_k(x,u,\nabla u)\,\nabla u_k\Big) + q_k(x,u_1,\dots,u_s) &= f_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \ (k=1,\dots,s) \tag{91}$$

under the following assumptions:

Assumptions 3.2.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and boundedness.) For all $k,l = 1,\dots,s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$ and $q_k \in W^{1,\infty}(\Omega \times \mathbf{R}^s)$.

(iii) (Cooperativity.) We have

$$\frac{\partial q_k}{\partial \xi_l}(x,\xi) \le 0 \qquad (k,l = 1,\dots,s,\ k \ne l;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{92}$$

(iv) (Weak diagonal dominance for the Jacobians.) We have

$$\sum_{l=1}^{s} \frac{\partial q_k}{\partial \xi_l}(x,\xi) \ge 0 \qquad (k = 1,\dots,s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{93}$$

(v) For all $k = 1, \dots, s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

Remark 3.5 Similarly to (59), assumptions (92)-(93) now imply

$$\frac{\partial q_k}{\partial \xi_k}(x, \xi) \geq 0 \qquad (k = 1, \dots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{94}$$

The basic idea to deal with problem (91) is to reduce it to (56) via suitably defined functions $V_{kl} : \Omega \times \mathbf{R}^s \to \mathbf{R}$. Namely, let

$$V_{kl}(x, \xi) := \int_0^1 \frac{\partial q_k}{\partial \xi_l}(x, t\xi)\, dt \qquad (k, l = 1, \dots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{95}$$

Then the Newton-Leibniz formula yields

$$q_k(x, \xi) = q_k(x, 0) + \sum_{l=1}^{s} V_{kl}(x, \xi)\, \xi_l \qquad (k = 1, \dots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{96}$$

Defining

$$\hat{f}_k(x) := f_k(x) - q_k(x, 0) \qquad (k = 1, \dots, s), \tag{97}$$

problem (91) then becomes

$$\left.\begin{aligned} -\mathrm{div}\, \Big(b_k(x, u, \nabla u)\, \nabla u_k\Big) + \sum_{l=1}^{s} V_{kl}(x, u)\, u_l &= \hat{f}_k(x) \quad \text{a.e. in } \Omega, \\ b_k(x, u, \nabla u) \frac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \ (k = 1, \dots, s), \tag{98}$$

which is a special case of (56). Here the assumption $q_k \in W^{1,\infty}(\Omega \times \mathbf{R}^s)$ yields that $V_{kl} \in L^\infty(\Omega \times \mathbf{R}^s)$ $(k, l = 1, \dots, s)$. Clearly, assumptions (92) and (93) imply that the functions V_{kl} defined in (95) satisfy (57) and (58), respectively. The remaining items of Assumptions 3.1 and 3.2 coincide, therefore system (98) satisfies Assumptions 3.2.

Consequently, for a finite element discretization developed as in subsection 3.1.2, Theorem 3.2 yields the discrete maximum principle (87) for suitable discretizations of (98), provided $\hat{f}_k \leq 0$ and $\gamma_k \leq 0$ $(k = 1, \dots, s)$. For the original system (91), we thus obtain

Corollary 3.3 *Let problem (91) satisfy Assumptions 3.2, and let its FEM discretization satisfy the corresponding conditions of Theorem 3.1. If*

$$f_k \leq q_k(x, 0), \qquad \gamma_k \leq 0 \qquad (k = 1, \dots, s)$$

and $u^h = (u^h_1, \dots, u^h_s)^T$ is the FEM solution of system (91), then for sufficiently small h,

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u^h_k \leq \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g^h_k\}. \tag{99}$$

3.3 Systems with general reaction terms of superlinear growth

In the previous section we have required the functions q_k to grow at most linearly via the condition $q_k \in W^{1,\infty}(\Omega \times \mathbf{R}^s)$. However, this is a strong restriction and is not satisfied even by (nonlinear) polynomials of u_k that often arise in reaction-diffusion problems. In this section we extend the previous results to problems where the functions q_k may grow polynomially. This generalization, however, needs stronger assumptions in other parts of the problem, because we now need the monotonicity of the corresponding operator in the proof of the DMP. For this to hold, the row-diagonal dominance for the Jacobians in Assumption 3.2 (iv) must be strengthened to diagonal dominance w.r.t. both rows and columns. (In addition, the principal part must be more specific too, but this is not so much restrictive since in practice it is even linear.)

Accordingly, let us now consider the system

$$\left.\begin{aligned} -\mathrm{div}\,\Big(b_k(x,\nabla u_k)\,\nabla u_k\Big) + q_k(x,u_1,\dots,u_s) &= f_k(x) \quad \text{a.e. in } \Omega, \\ b_k(x,\nabla u_k)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \quad (k=1,\dots,s) \tag{100}$$

under the following assumptions:

Assumptions 3.3.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and growth.) For all $k,l = 1,\dots,s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^d)$ and $q_k \in C^1(\Omega \times \mathbf{R}^s)$. Further, let

$$2 \le p < p^*, \quad \text{where } p^* := \tfrac{2d}{d-2} \ \text{ if } d \ge 3 \ \text{ and } \ p^* := +\infty \text{ if } d=2; \tag{101}$$

then there exist constants $\beta_1, \beta_2 \ge 0$ such that

$$\left|\frac{\partial q_k}{\partial \xi_l}(x,\xi)\right| \le \beta_1 + \beta_2|\xi|^{p-2} \qquad (k,l=1,\dots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{102}$$

(iii) (Ellipticity.) Defining $a_k(x,\eta) := b_k(x,\eta)\eta$ for all k, the Jacobian matrices $\frac{\partial}{\partial\eta}\, a_k(x,\eta)$ are uniformly spectrally bounded from both below and above.

(iv) (Cooperativity.) We have

$$\frac{\partial q_k}{\partial \xi_l}(x,\xi) \le 0 \qquad (k,l=1,\dots,s,\ k\ne l;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{103}$$

(v) (Weak diagonal dominance for the Jacobians w.r.t. rows and columns.) We have

$$\sum_{l=1}^{s} \frac{\partial q_k}{\partial \xi_l}(x,\xi) \geq 0, \qquad \sum_{l=1}^{s} \frac{\partial q_l}{\partial \xi_k}(x,\xi) \geq 0 \quad (k=1,\ldots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{104}$$

(vi) For all $k=1,\ldots,s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

Remark 3.6 (i) Similarly to (59), assumptions (103)-(104) now imply

$$\frac{\partial q_k}{\partial \xi_k}(x,\xi) \geq 0 \qquad (k=1,\ldots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{105}$$

(ii) Similarly to Remark 3.4, one may include additional terms $s_k(x,u_1,\ldots,u_s)$ on the Neumann boundary Γ_N, which we omit here for technical simplicity; then s_k must satisfy similar properties as assumed for q_k.

To handle system (100), we start as in the previous subsection by reducing it to a system with nonlinear coefficients: if the functions V_{kl} and $\hat{f}_k$ $(k,l = 1,\ldots,s)$ are defined as in (95) and (97), respectively, then (100) takes a form similar to (98):

$$\left.\begin{aligned} -\mathrm{div}\Big(b_k(x,\nabla u)\,\nabla u_k\Big) + \sum_{l=1}^{s} V_{kl}(x,u)\,u_l &= \hat{f}_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \ (k=1,\ldots,s). \tag{106}$$

The difference compared to the previous subsection is the superlinear growth allowed in (102), which does not let us apply Theorem 3.2 directly as we did for system (91). Instead, we must reprove Theorem 3.1 under Assumptions 3.3.

First, when considering a finite element discretization developed as in subsection 3.1.2, we need a strengthened assumption for the quasi-regularity of the mesh.

Definition 3.2 Let $\Omega \subset \mathbf{R}^d$ and let us consider a family of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ constructed as in subsection 3.1.2. Here $h > 0$ is the mesh parameter, proportional to the maximal diameter of the supports of the basis functions $\phi_1, ..., \phi_n$. The corresponding mesh will be called *quasi-regular* w.r.t. problem (100) if

$$c_1 h^\gamma \leq meas(\mathrm{supp}\,\varphi_p) \leq c_2 h^d\,, \tag{107}$$

where the positive real number γ satisfies

$$d \leq \gamma < \gamma_d^*(p) := 2d - \frac{(d-2)p}{2} \tag{108}$$

with p from Assumption 3.3 (ii).

Remark 3.7 Assumption (108) makes sense for γ since by (101),

$$d < d + d(1 - \tfrac{p}{p^*}) = \gamma_d^*(p)\,. \tag{109}$$

Note on the other hand that $\gamma_d^*(p) \leq \gamma_d^*(2) = d+2$, which is in accordance with (82). Further, we have, in particular, in 2D: $\gamma_2^*(p) \equiv 4$ for all $2 \leq p < \infty$, and in 3D: $\gamma_3^*(p) = 6 - (p/2)$ (where $2 \leq p \leq 6$, and accordingly $3 \leq \gamma_3^*(p) \leq 5$).

Next, as an analogue of Lemma 3.1, the following technical result holds for problem (100):

Lemma 3.2 [26]. *Let Assumptions 3.3 hold. Analogously to (83), for any $u \in H^1(\Omega)^s$ let us define the operators $B(u)$ and $R(u)$ via*

$$\langle B(u)w, v\rangle = \int_\Omega \sum_{k=1}^{s} b_k(x, \nabla u)\, \nabla w_k \cdot \nabla v_k, \quad \langle R(u)w, v\rangle = \int_\Omega \sum_{k,l=1}^{s} V_{kl}(x,u)\, w_l\, v_k$$

($w \in H^1(\Omega)^s$, $v \in H^1_D(\Omega)^s$). Together with $A(u) := B(u)u + R(u)u$, the operators $B(u)$ and $R(u)$ satisfy Assumptions 2.4.1-2.4.2.

Then one can derive the desired nonnegativity result for the stiffness matrix, i.e. the analogue of Theorem 3.1 for system (100). Here the entries of $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ are

$$a_{ij}(\bar{\mathbf{c}}) = \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, \nabla u^h)\, (\nabla \phi_j)_k \cdot (\nabla \phi_i)_k + \sum_{k,l=1}^{s} V_{kl}(x, u^h)\, (\phi_j)_l\, (\phi_i)_k \Big), \tag{110}$$

where by (95),

$$V_{kl}(x, u^h(x)) = \int_0^1 \frac{\partial q_k}{\partial \xi_l}(x, t u^h(x))\, dt \qquad (k, l = 1, \dots, s;\ x \in \Omega). \tag{111}$$

Theorem 3.3 [26]. *Let problem (100) satisfy Assumptions 3.3. Let us consider a family of finite element subspaces V_h ($h \to 0$) satisfying the following property: there exists a real number γ satisfying (108) such that for any indices $p = 1, ..., \bar{n}_0$, $t = 1, ..., \bar{n}$ ($p \neq t$), if meas$(\operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t) > 0$ then*

$$\nabla\varphi_t \cdot \nabla\varphi_p \leq 0 \quad \textit{on } \Omega \quad \textit{and} \quad \int_\Omega \nabla\varphi_t \cdot \nabla\varphi_p \leq -K_0\, h^{\gamma-2} \tag{112}$$

with some constant $K_0 > 0$ independent of p, t and h. Further, let the family of meshes be regular from above, according to Definition 3.1.

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (110) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

Similarly as in Corollary 3.3, using Theorem 3.3, Corollary 2.2 and Theorem 3.2, respectively, we obtain the *discrete maximum principle* for system (100):

Corollary 3.4 *Let problem (100) satisfy Assumptions 3.3, and let its FEM discretization satisfy the conditions of Theorem 3.3. If*

$$f_k \le q_k(x,0), \qquad \gamma_k \le 0 \qquad (k=1,\dots,s)$$

then for sufficiently small h, the FEM solution $u^h = (u_1^h,\dots,u_s^h)$ of system (100) satisfies

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \le \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g_k^h\}. \tag{113}$$

Remark 3.8 As pointed out in Remark 3.2, the result (113) can be divided in two cases: a 'more direct' DMP (89) or the nonpositivity property (90). Further, if $f_k \ge q_k(x,0)$, $\gamma_k \ge 0$ for all k, then (by reversing signs) one can derive the corresponding discrete minimum principle or nonnegativity property. We formulate the latter below for its practical importance.

Corollary 3.5 *Let problem (100) satisfy Assumptions 3.3, and let its FEM discretization satisfy the conditions of Theorem 3.3. If*

$$f_k \ge q_k(x,0), \quad \gamma_k \ge 0, \quad g_k \ge 0 \qquad (k=1,\dots,s)$$

then for sufficiently small h, the FEM solution $u^h = (u_1^h,\dots,u_s^h)^T$ of system (100) satisfies

$$u_k^h \ge 0 \quad \text{on } \Omega \qquad (k=1,\dots,s). \tag{114}$$

3.4 Sufficient conditions and their geometric meaning

The key assumption for the FEM subspaces V_h and the associated meshes in the above results has been the following property, see (85) in Theorem 3.1 and (112) in Theorem 3.3. There exists a real number γ satisfying (82) or (108), respectively, such that for any indices $p = 1,...,\bar{n}_0$, $t = 1,...,\bar{n}$ $(p \ne t)$, if $meas(\operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t) > 0$ then

$$\nabla\varphi_t \cdot \nabla\varphi_p \le 0 \text{ on } \Omega \quad \text{and} \tag{115}$$

$$\int_\Omega \nabla\varphi_t \cdot \nabla\varphi_p \le -K_0\, h^{\gamma-2} \tag{116}$$

with some constant $K_0 > 0$ independent of p, t and h. (The family of meshes must also be regular from above as in (79), but that requirement obviously holds for the usual definition of the mesh parameter h as the maximal diameter of elements.)

A classical way to satisfy such conditions is a pointwise inequality like (19) together with suitable mesh regularity, see Remark 3.3. However, one can ensure (115)-(116) with less strong conditions as well. We summarize some possibilities below.

Proposition 3.1 *Let the family of FEM discretizations $\mathcal{V} = \{V_h\}_{h\to 0}$ satisfy either of the following conditions, where φ_t, φ_p are arbitrary basis functions such that $p = 1, ..., \bar{n}_0,\ t = 1, ..., \bar{n},\ p \neq t$, we let*

$$\Omega_{pt} := \operatorname{supp} \varphi_p \cap \operatorname{supp} \varphi_t \,,$$

further, let

$$\sigma > 0 \quad \textit{and} \quad c_1, c_2, c_3 > 0$$

denote constants independent of the indices p, t and the mesh parameter h, and finally, d is the space dimension and γ satisfies (108).

(i) Let the basis functions satisfy

$$\nabla\varphi_t \cdot \nabla\varphi_p \leq -\frac{\sigma}{h^2} < 0 \quad \textit{on } \Omega_{pt}, \tag{117}$$

and the family of meshes be quasi-regular as in (107).

(ii) Let there exist $0 < \varepsilon \leq \gamma - d$ such that the basis functions satisfy

$$\nabla\varphi_t \cdot \nabla\varphi_p \leq -\frac{\sigma}{h^{2-\varepsilon}} < 0 \quad \textit{on } \Omega_{pt}, \tag{118}$$

but let the quasi-regularity (107) of the family of meshes be now strengthened to

$$c_1 h^{\gamma-\varepsilon} \leq \textit{meas}(\operatorname{supp} \varphi_p) \leq c_2 h^d \,. \tag{119}$$

(iii) Let there exist subsets $\Omega_{pt}^+ \subset \Omega_{pt}$ for all p, t such that the basis functions satisfy

$$\nabla\varphi_t \cdot \nabla\varphi_p \leq -\frac{\sigma}{h^2} < 0 \quad \textit{on } \ \Omega_{pt}^+ \quad \textit{and} \quad \nabla\varphi_t \cdot \nabla\varphi_p \leq 0 \ \ \textit{on } \ \Omega_{pt} \setminus \Omega_{pt}^+ \tag{120}$$

and we have

$$\frac{\textit{meas}(\Omega_{pt}^+)}{\textit{meas}(\Omega_{pt})} \geq c_3 > 0 \,, \tag{121}$$

further, let the family of meshes be quasi-regular as in (107).

Then (115)-(116) holds.

Proof is obvious.

As discussed in subsection 2.3, conditions (115) and (117) have nice geometric interpretations for simplicial, bilinear and for prismatic finite elements, but these conditions are often restrictive. The weaker conditions (118) and (120) allow in theory easier refinement procedures as the property of (strict) acuteness is often hard to preserve in refinement procedures, e.g. by bisection algorithms [5, 34]. First, (118) may allow the acute mesh angles to deteriorate (i.e. tend to 90°) as $h \to 0$. Namely, if a family of simplicial meshes is regular then $|\nabla\varphi_t| = O(h^{-1})$ for all linear basis functions: hence, considering two basis functions φ_p, φ_t and letting α denote the angle of their gradients on a given simplex, the sufficient condition

$$\cos\alpha \le -\sigma h^{\varepsilon} \tag{122}$$

(with some constant $\sigma > 0$ independent of h) implies

$$\nabla\varphi_t \cdot \nabla\varphi_p = |\nabla\varphi_t|\, |\nabla\varphi_p| \cos\alpha \le -\frac{\sigma\, h^{\varepsilon}}{h^2}\,,$$

i.e. (118) holds. Clearly, if $h \to 0$ then (122) allows $\cos\alpha \to 0$, i.e. $\alpha \to 90^\circ$, for the angle of gradients, in which case the corresponding mesh angle also tends to 90°. (In particular, for problem (56), when (108) coincides with $d \le \gamma < d+2$ as in (82), then $\gamma - d$ can be chosen arbitrarily close to 2. Hence the exponent $2-\varepsilon$ in (118) can be arbitrarily close to 0, i.e. the decay of mesh angles to 90° may be fast as $h \to 0$.)

Second, (120) means that one can allow some right mesh angles, but each Ω_{pt}, which consists of a finite number of elements, must contain some elements with acute mesh angles and the measure of these must not asymptotically vanish.

4 Discrete maximum principles for elliptic systems including first order terms

In this section we give various new results for elliptic systems including first order terms. We consider four types of systems, in which the diffusion and reaction terms are nonlinear: first the reaction terms are given as the unknown functions multiplied by nonlinear coefficients, later the reactions are general functions of x and the u_k. On the other hand, the first three types of problems involve linear convection terms and general Dirichlet data, whereas the last type contains nonlinear convection terms and homogeneous Dirichlet data. The allowed growth of the reaction terms is sublinear in two cases and superlinear in the other two cases. These differences require suitable modifications in the assumptions and the treatment.

4.1 Nonsymmetric systems with nonlinear reaction coefficients

First we consider nonlinear elliptic systems of the form

$$\left.\begin{aligned} -\mathrm{div}\left(b_k(x,u,\nabla u)\,\nabla u_k\right) + \mathbf{w}_k(x)\cdot\nabla u_k + \sum_{l=1}^{s} V_{kl}(x,u,\nabla u)\, u_l &= f_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \tag{123}$$

$(k = 1, \dots, s)$. The notations follow those of section 3.

Assumptions 4.1.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and boundedness.) For all $k,l = 1,\dots,s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$, $\mathbf{w}_k \in W^{1,\infty}(\Omega)$ and $V_{kl} \in L^\infty(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$.

(iii) (Coercivity.) We have $\operatorname{div} \mathbf{w}_k \le 0$ on Ω and $\mathbf{w}_k \cdot \nu \ge 0$ on Γ_N $(k = 1,\dots,s)$.

(iv) (Cooperativity.) We have

$$V_{kl} \le 0 \qquad (k,l = 1,\dots,s,\ k \ne l). \tag{124}$$

(v) (Weak diagonal dominance.) We have

$$\sum_{l=1}^{s} V_{kl} \ge 0 \qquad (k = 1,\dots,s). \tag{125}$$

(vi) For all $k = 1,\dots,s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^*_k \in H^1(\Omega)$.

Remark 4.1 (i) Assumptions (124)-(125) imply (59).

(ii) One may consider additional terms on the Neumann boundary as in Remark 3.4.

For the weak formulation of problem (123), we suitable modify (61)–(62): find $u \in H^1(\Omega)^s$ such that

$$\langle A(u), v\rangle = \langle \psi, v\rangle \qquad (\forall v \in H^1_D(\Omega)^s) \tag{126}$$

$$\text{and} \quad u - g^* \in H^1_D(\Omega)^s, \tag{127}$$

where

$$\begin{aligned}\langle A(u), v\rangle = \int_\Omega \Big(\sum_{k=1}^{s} b_k(x,u,\nabla u)\, \nabla u_k \cdot \nabla v_k \\ + \sum_{k=1}^{s} (\mathbf{w}_k(x) \cdot \nabla u_k)\, v_k + \sum_{k,l=1}^{s} V_{kl}(x,u,\nabla u)\, u_l\, v_k \Big)\end{aligned} \tag{128}$$

for given $u = (u_1,\dots,u_s) \in H^1(\Omega)^s$ and $v = (v_1,\dots,v_s) \in H^1_D(\Omega)^s$, further,

$$\langle \psi, v\rangle = \int_\Omega \sum_{k=1}^{s} f_k v_k + \int_{\Gamma_N} \sum_{k=1}^{s} \gamma_k v_k \tag{129}$$

for given $v = (v_1,\dots,v_s) \in H^1_D(\Omega)^s$, and finally $g^* := (g^*_1,\dots,g^*_s)$.

The finite element discretization of problem (123) is defined in the same way as in subsection 3.1.2. Using the FEM subspaces (70) and (72), one seeks $u^h \in V_h$ such that

$$\langle A(u^h), v^h\rangle = \langle \psi, v^h\rangle \qquad (\forall v^h \in V^0_h) \tag{130}$$

$$\text{and} \quad u^h - g^h \in V^0_h, \quad \text{i.e.,} \ \ u^h = g^h \text{ on } \Gamma_D \tag{131}$$

(where $g^h = \sum_{j=n_0+1}^{n} g_j \phi_j \ \in V_h$ is the approximation of g^* on Γ_D). The only difference is that the entries of the stiffness matrix become, for any $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ and $i = 1, ..., n_0$, $j = 1, ..., n$,

$$a_{ij}(\bar{\mathbf{c}}) := \int_\Omega \Bigg(\sum_{k=1}^{s} b_k(x, u^h, \nabla u^h) \, (\nabla \phi_j)_k \cdot (\nabla \phi_i)_k + \sum_{k=1}^{s} \Big(\mathbf{w}_k(x) \cdot (\nabla \phi_j)_k \Big) \, (\phi_i)_k \\ + \sum_{k,l=1}^{s} V_{kl}(x, u^h, \nabla u^h) \, (\phi_j)_l \, (\phi_i)_k \Bigg) \tag{132}$$

instead of (76). With these, similarly to (75), we have the $n_0 \times n$ system of algebraic equations

$$\sum_{j=1}^{n} a_{ij}(\bar{\mathbf{c}}) \, c_j = d_i \qquad (i = 1, ..., n_0). \tag{133}$$

Now our goal is to derive a DMP for problem (123) using Theorem 2.5. For this, we first define the underlying operators as in (83), for which Assumptions 2.4.1 must hold.

Lemma 4.1 *Let Assumptions 4.1 hold. For any $u \in H^1(\Omega)^s$, let us define the operators $B(u)$, $N(u) \equiv N$ and $R(u)$ via*

$$\langle B(u)z, v \rangle = \int_\Omega \sum_{k=1}^{s} b_k(x, u, \nabla u) \, \nabla z_k \cdot \nabla v_k, \quad \langle Nz, v \rangle = \int_\Omega \sum_{k=1}^{s} (\mathbf{w}_k(x) \cdot \nabla z_k) \, v_k, \\ \langle R(u)z, v \rangle = \int_\Omega \sum_{k,l=1}^{s} V_{kl}(x, u, \nabla u) \, z_l \, v_k \tag{134}$$

($z \in H^1(\Omega)^s$, $v \in H^1_D(\Omega)^s$). Together with the operator A, defined in (128), the operators $B(u)$, $N(u)$ and $R(u)$ satisfy Assumptions 2.4.1, modified according to Remark 2.6, in the spaces $H = H^1(\Omega)^s$ and $H_0 = H^1_D(\Omega)^s$.

Proof. By Lemma 3.1, we only need to prove those statements that do not concern only $B(u)$ or $R(u)$. We define

$$\|v\|^2 := \sum_{k=1}^{s} (\int_\Omega |\nabla v_k|^2 + \int_{\Gamma_D} |v_k|^2) \tag{135}$$

on $H^1(\Omega)^s$, which is a norm since $\Gamma_D \neq \emptyset$. Then for $v \in H^1_D(\Omega)^s$ we have $\|v\|^2 = \sum_{k=1}^{s} \int_\Omega |\nabla v_k|^2$.

(i) It is obvious from (128) and (134) that $A(u) = B(u)u + N(u)u + R(u)u$.

(ii) By Lemma 3.1, we have $\langle B(u)v, v\rangle \geq m\,\|v\|^2 \quad (u \in H^1(\Omega)^s,\ v \in H^1_D(\Omega)^s)$. Further, the assumptions and the divergence theorem imply for all $v_k \in H^1_D(\Omega)$ that

$$2\int_\Omega (\mathbf{w}_k \cdot \nabla v_k)v_k \ = -\int_\Omega (\operatorname{div}\mathbf{w}_k)\, v_k^2\, dx + \int_{\partial\Omega} (\mathbf{w}_k \cdot \nu)\, v_k^2\, d\sigma \geq 0. \quad (136)$$

Summing up for k and dividing by 2, we obtain that $\langle Nv, v\rangle \geq 0$. Hence (31) is valid.

(iii) This follows from Lemma 3.1, where P and D were defined as follows. Let $D \subset H^1(\Omega)^s$ consist of the functions that have only one nonzero coordinate that is nonnegative, i.e. $v \in D$ iff $v = (0, \dots, 0, g, 0, \dots, 0)^T$ with g at the k-th entry for some $1 \leq k \leq s$ and $g \in H^1(\Omega)$, $g \geq 0$. Further, let $P \subset H^1(\Omega)^s$ consist of the functions that have identical nonnegative coordinates, i.e. $v \in P$ iff $v = (y, \dots, y)$ for some $y \in H^1(\Omega)$, $y \geq 0$.

(iv) By Lemma 3.1, we have for all $u, w, v \in H^1(\Omega)^s$

$$\langle R(u)z, v\rangle \leq M_R(\|u\|)\, |||z|||\, |||v|||,$$

for the new norm $|||v|||^2 = \|v\|^2_{L^2(\Omega)^s}$, i.e. (54) holds. In fact [26, Lemma 4.1], one has the constant function $M_R(r) \equiv s\tilde{V}$, where $\tilde{V} := \max\limits_{k,l} \|V_{kl}\|_{L^\infty}$. For N, we have

$$\langle Nz, v\rangle \leq \|\mathbf{w}\|_{L^\infty(\Omega)^s} \|\nabla z\|_{L^2(\Omega)^s}\, \|v\|_{L^2(\Omega)^s} \leq \|\mathbf{w}\|_{L^\infty(\Omega)^s} \|z\|\, \|v\|_{L^2(\Omega)^s},$$

where $\|\mathbf{w}\|_{L^\infty(\Omega)^s} := \sup_{k,x} |\mathbf{w}_k(x)|$, i.e. (53) holds for the constant function $M_N(r) \equiv \|\mathbf{w}\|_{L^\infty(\Omega)^s}$ and the norms $|||z|||_{N_1} := \|z\|$, $|||v|||_{N_2} := \|v\|_{L^2(\Omega)^s}$. ■

Now let us consider the finite element discretization for problem (56), developed as in subsection 3.1.2. First we need a strengthened assumption for the regularity of the mesh.

Definition 4.1 Let $\Omega \subset \mathbf{R}^d$ and let us consider a family of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ constructed as in subsection 3.1.2. Here $h > 0$ is the mesh parameter, proportional to the maximal diameter of the supports of the basis functions $\phi_1, ..., \phi_n$. The corresponding mesh will be called *quasi-regular* w.r.t. problem (123) if

$$c_1 h^\gamma \leq meas(\operatorname{supp}\varphi_p) \leq c_2 h^d\,, \quad (137)$$

where the positive real number γ satisfies

$$d \leq \gamma < \frac{d(d+2)}{d+1}\,. \quad (138)$$

One can then prove the desired nonnegativity result for the stiffness matrix:

Theorem 4.1 *Let problem (123) satisfy Assumptions 4.1. Let us consider a family of finite element subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$, such that the corresponding family of meshes is quasi-regular according to Definition 4.1, further, for any $p = 1, ..., \bar{n}_0$, $t = 1, ..., \bar{n}$ $(p \neq t)$, if $meas(\operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t) > 0$ then*

$$\nabla\varphi_t \cdot \nabla\varphi_p \le 0 \quad \text{on } \Omega \quad \text{and} \quad \int_\Omega \nabla\varphi_t \cdot \nabla\varphi_p \le -K_0\, h^{\gamma-2} \tag{139}$$

where γ is from (138) and $K_0 > 0$ is a constant independent of p, t and h.

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (132) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

Proof. We wish to apply Theorem 2.5, modified according to Remark 2.6, in the spaces $H = H^1(\Omega)^s$ and $H_0 = H^1_D(\Omega)^s$. We use that the operators $B(u)$ and $R(u)$ satisfy the corresponding assumptions, since this was used in Theorem 3.1.

With the operator A defined in (128), our problem (126)-(127) coincides with (28)-(29). The FEM subspaces (70) and (72) fall into the class (36). Using the operators $B(u)$, $N(u) \equiv N$ and $R(u)$ in (134), the discrete problem (130)-(131) turns into the form (39) such that by Lemma 4.1, these operators satisfy Assumptions 2.4.1, modified according to Remark 2.6.

Now we follow the proof of Theorem 3.1, see [26, Theorem 4.1]. First we define neighbouring basis functions satisfying Assumptions 2.4.3. Let $\phi_i, \phi_j \in V_h$. Using definitions (69) and (71), assume that ϕ_i has φ_p at its k-th entry and ϕ_j has φ_t at its l-th entry. Then we call ϕ_i and ϕ_j neighbouring basis functions if $k = l$ and $meas(\operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t) > 0$. Let $N := \{1, \dots, n\}$ as before. For any $k = 1, \dots, s$ let

$$S^0_k := \{i \in N : \ i = (k-1)\bar{n}_0 + p \text{ for some } 1 \le p \le \bar{n}_0\},$$

$$\tilde{S}_k := \{i \in N : \ i = n_0 + (k-1)(\bar{n} - \bar{n}_0) + p - \bar{n}_0 \text{ for some } \bar{n}_0 + 1 \le p \le \bar{n}\},$$

$$S_k := S^0_k \cup \tilde{S}_k\,,$$

i.e. by (69) and (71), the basis functions ϕ_i with index $i \in S_k$ have a nonzero coordinate φ_p for some p at the k-th entry, and in particular, $i \in S^0_k$ if this φ_p is an 'interior' basis function (i.e. $1 \le p \le \bar{n}_0$) and $i \in \tilde{S}_k$ if this φ_p is a 'boundary' basis function (i.e. $\bar{n}_0 + 1 \le p \le \bar{n}$). By [26, Theorem 4.1], these neighbouring basis functions satisfy Assumptions 2.4.3.

Our remaining task is to check assumptions (a)-(e) of Theorem 2.5.

(a) Let $\phi_i \in V^0_h$, $\phi_j \in V_h$, and let ϕ_i have φ_p at its k-th entry and ϕ_j have φ_t at its l-th entry. We must prove that either (48) or (49)-(51) holds.

If $k \neq l$, then from [26, Theorem 4.1] we have $\langle B(u^h)\phi_j, \phi_i\rangle = 0$ and $\langle R(u^h)\phi_j, \phi_i\rangle \le 0$. Here ϕ_i and ϕ_j have no common nonzero coordinates, hence $\langle N\phi_j, \phi_i\rangle = 0$, i.e. (48) holds.

If $k = l$, then Assumption 3.0 (ii) and (85) yield

$$\langle B(u^h)\phi_j, \phi_i\rangle = \int_\Omega b_k(x, u^h, \nabla u^h)\, \nabla\varphi_t \cdot \nabla\varphi_p \le m \int_{\Omega_{pt}} \nabla\varphi_t \cdot \nabla\varphi_p \quad (140)$$

where $\Omega_{pt} := \operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t$. If $meas(\Omega_{pt}) = 0$ then $\langle B(u^h)\phi_j, \phi_i\rangle = \langle N\phi_j, \phi_i\rangle = 0$ and we have $\langle R(u^h)\phi_j, \phi_i\rangle \le 0$ similarly as before, hence (48) holds again. If $meas(\Omega_{pt}) > 0$ then (85) and (140) imply

$$\langle B(u^h)\phi_j, \phi_i\rangle \le -mK_0 h^{\gamma-2} \equiv -\hat{c}_1\, h^{\gamma-2} =: -M_B(h) \quad (141)$$

and we must check (51). Let us estimate $T(h)^2$ from (55). As seen in the proof of Lemma 4.1, $|||z|||_{N_1} = \|z\|$ and $|||v|||_{N_2} = |||v|||_{R_1} = |||v|||_{R_2} = \|v\|_{L^2(\Omega)^s}$. Here $\|z\|$ denotes the $H^1(\Omega)$-norm, which we replace here by the equivalent norm $\|v\|^2_{H^1(\Omega)^s} = \sum_{k=1}^{s} (\int_\Omega |\nabla v_k|^2 + \int_\Omega |v_k|^2)$. Hence we must estimate

$$T(h)^2 = \sup\Big\{ \max\Big\{ \|\phi_i\|_{H^1(\Omega)^s}\|\phi_j\|_{L^2(\Omega)^s}\, ,\ \|\phi_i\|_{L^2(\Omega)^s}\|\phi_j\|_{L^2(\Omega)^s} \Big\} : \\ \phi_i, \phi_j \in V_h \Big\}. \quad (142)$$

The L_2-norm of the basis functions satisfies the following estimate, where ϕ_j has φ_t at its l-th entry as before, and we use (137) and that (66) implies $\varphi_t \le 1$:

$$\|\phi_j\|^2_{L^2(\Omega)^s} = \|\varphi_t\|^2_{L^2(\Omega)} \le \int\limits_{\operatorname{supp}\varphi_t} 1 \ = meas(\operatorname{supp}\varphi_t) \le c_2 h^d \quad (143)$$

for all j. For the H^1-norm, let us first estimate the gradient term. Using the previous argument, (137) and (68), respectively,

$$\|\nabla\phi_j\|^2_{L^2(\Omega)^s} = \|\nabla\varphi_t\|^2_{L^2(\Omega)} = \int\limits_{\operatorname{supp}\varphi_t} |\nabla\varphi_t|^2 \ \le \frac{meas(\operatorname{supp}\varphi_t)}{diam^2(\operatorname{supp}\varphi_t)}\,. \quad (144)$$

Here $diam(\operatorname{supp}\varphi_t) \ge c_1 h^{\gamma/d}$ for some $c_1 > 0$ independent of h, since otherwise the l.h.s. of (137) would fail. With this and the r.h.s. of (137), we obtain

$$\|\nabla\phi_j\|^2_{L^2(\Omega)^s} \le c_3 h^{d-\frac{2\gamma}{d}} \quad (145)$$

for some $c_3 > 0$ independent of h. Since this is larger (as $h \to 0$) than the L^2-norm estimate in (143), we also have

$$\|\phi_j\|^2_{H^1(\Omega)^s} \le c_3 h^{d-\frac{2\gamma}{d}}\,, \quad (146)$$

and from (143) and (146) we obtain $\|\phi_i\|^2_{H^1(\Omega)^s}\|\phi_j\|^2_{L^2(\Omega)^s} \le const. \cdot h^{2d-\frac{2\gamma}{d}}$ for all i, j. Also, in (142) the first term in the max is the greater than the second one, hence

$$T(h)^2 \le \sup\Big\{ \|\phi_i\|_{H^1(\Omega)^s}\|\phi_j\|_{L^2(\Omega)^s} :\ \phi_i, \phi_j \in V_h \Big\} \le c_4 h^{d-\frac{\gamma}{d}}$$

for some $c_4 > 0$ independent of h. From this, using (141) and that (138) implies $\gamma + \frac{\gamma}{d} < d + 2$, we obtain

$$\lim_{h \to 0} \frac{M_B(h)}{T(h)^2} \geq c_5 \lim_{h \to 0} h^{\gamma - 2 - d + \frac{\gamma}{d}} = +\infty. \tag{147}$$

(b) Let $\phi_i \in V_h^0$ and $\phi_j \in V_h$ be neighbouring basis vectors, i.e, as defined before in the proof, $k = l$ and $meas(\operatorname{supp} \varphi_p \cap \operatorname{supp} \varphi_t) > 0$. Then, as seen just above, we obtain (141) and (147), which coincide with (49)-(51).

(c) According to Remark 2.6, it is required here that $M_N(\|u^h\|)$ and $M_R(\|u^h\|)$ are bounded as $h \to 0$. Since we have the constant bounds $M_R(r) \equiv s\tilde{V}$ and $M_N(r) \equiv \|\mathbf{w}\|_{L^\infty(\Omega)^s}$, see part (iv) of the proof of Lemma 4.1, these are trivially bounded.

(d) For all $u \in H^1(\Omega)^s$ and $h > 0$, the definition of the functions ϕ_j and assumption (66) imply

$$\sum_{j=1}^{n} \phi_j = \begin{pmatrix} \sum\limits_{p=1}^{\bar{n}} \varphi_p \\ \sum\limits_{p=1}^{\bar{n}} \varphi_p \\ \dots \\ \sum\limits_{p=1}^{\bar{n}} \varphi_p \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} =: \mathbf{1}\,. \tag{148}$$

Then by (134)

$$\langle B(u)(\sum_{j=1}^{n} \phi_j), v\rangle = \langle B(u)\mathbf{1}, v\rangle = \int_\Omega \sum_{k=1}^{s} b_k(x, u, \nabla u)\, \nabla 1 \cdot \nabla v_k = 0 \quad \text{and}$$

$$\langle N(\sum_{j=1}^{n} \phi_j), v\rangle = \langle N\mathbf{1}, v\rangle = \int_\Omega \sum_{k=1}^{s} (\mathbf{w}_k(x) \cdot \nabla 1)\, v_k = 0$$

for all $v \in H_D^1(\Omega)^s$, i.e. $\sum\limits_{j=1}^{n} \phi_j$ belongs to both $\ker B(u)$ and $\ker N$.

(e) This was proved in Theorem 3.1, see [26, Theorem 4.1]. ■

Similarly to Corollary 3.2 before, we thus obtain

Corollary 4.1 *Let the assumptions of Theorem 4.1 hold and let*

$$f_k \leq 0, \qquad \gamma_k \leq 0 \qquad (k = 1, \dots, s).$$

Let the basis functions satisfy (66)-(68). Then for sufficiently small h, if $u^h = (u_1^h, \dots, u_s^h)^T$ is the FEM solution of system (123), then

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \leq \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g_k^h\}. \tag{149}$$

4.2 Nonsymmetric systems with sublinear reaction terms

Now we consider problems where the dependence on the lower order terms is given as general functions of x and u, growing at most linearly, thus following subsection 3.2. One can then reduce the problem to the previous one (123) directly. Accordingly, let us consider the system

$$\left.\begin{aligned} -\mathrm{div}\left(b_k(x,u,\nabla u)\,\nabla u_k\right) + \mathbf{w}_k(x)\cdot\nabla u_k + q_k(x,u_1,...,u_s) &= f_k(x) \quad \text{a.e. in } \Omega, \\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \tag{150}$$

$(k = 1,\ldots,s)$.

Assumptions 4.2.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and boundedness.) For all $k,l = 1,\ldots,s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^s \times \mathbf{R}^{s\times d})$, $\mathbf{w}_k \in W^{1,\infty}(\Omega)$ and $q_k \in W^{1,\infty}(\Omega \times \mathbf{R}^s)$.

(iii) (Coercivity.) We have $\mathrm{div}\,\mathbf{w}_k \le 0$ on Ω and $\mathbf{w}_k \cdot \nu \ge 0$ on Γ_N $(k = 1,\ldots,s)$.

(iv) (Cooperativity.) We have

$$\frac{\partial q_k}{\partial \xi_l}(x,\xi) \le 0 \qquad (k,l = 1,\ldots,s,\ k \ne l;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{151}$$

(v) (Weak diagonal dominance for the Jacobians.) We have

$$\sum_{l=1}^{s} \frac{\partial q_k}{\partial \xi_l}(x,\xi) \ge 0 \qquad (k = 1,\ldots,s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{152}$$

(vi) For all $k = 1,\ldots,s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

The basic idea to deal with problem (150) is to reduce it to (123), similarly as in Subsection 3.2. Defining the functions $V_{kl} : \Omega \times \mathbf{R}^s \to \mathbf{R}$ and $\hat{f}_k$ as in (95) and (97), respectively, problem (150) becomes

$$\left.\begin{aligned} -\mathrm{div}\left(b_k(x,u,\nabla u)\,\nabla u_k\right) + \mathbf{w}_k(x)\cdot\nabla u_k + \sum_{l=1}^{s} V_{kl}(x,u,\nabla u)\,u_l &= \hat{f}_k(x) \quad \text{a.e. in } \Omega, \\ b_k(x,u,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \tag{153}$$

$(k = 1,\ldots,s)$, which is a special case of (150).

Consequently, for a finite element discretization developed as in subsection 3.1.2, Corollary 4.1 discrete maximum principle (149) for suitable discretizations of (153), provided $\hat{f}_k \le 0$ and $\gamma_k \le 0$ $(k = 1,\ldots,s)$. For the original system (91), we thus obtain

Corollary 4.2 *Let problem (150) satisfy Assumptions 4.2, and let its FEM discretization satisfy the corresponding conditions of Theorem 4.1. If*

$$f_k \le q_k(x,0), \qquad \gamma_k \le 0 \qquad (k=1,\dots,s)$$

and $u^h = (u_1^h,\dots,u_s^h)^T$ *is the FEM solution of system (150), then for sufficiently small* h*,*

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h \le \max_{k=1,\dots,s} \max\{0, \max_{\Gamma_D} g_k^h\}. \tag{154}$$

4.3 Nonsymmetric systems with superlinear reaction terms

$$\left.\begin{aligned} -\mathrm{div}\Big(b_k(x,\nabla u)\,\nabla u_k\Big) + \mathbf{w}_k(x)\cdot\nabla u_k + q_k(x,u_1,...,u_s) &= f_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,\nabla u)\tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \tag{155}$$

$(k=1,\dots,s)$.

Assumptions 4.3.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and growth.) For all $k,l=1,\dots,s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^d)$, $\mathbf{w}_k \in W^{1,\infty}(\Omega)$ and $q_k \in C^1(\Omega \times \mathbf{R}^s)$. Further, let

$$2 \le p < p^*, \quad \text{where } p^* := \tfrac{2d}{d-2} \text{ if } d \ge 3 \text{ and } p^* := +\infty \text{ if } d=2; \tag{156}$$

then there exist constants $\beta_1, \beta_2 \ge 0$ such that

$$\left|\frac{\partial q_k}{\partial \xi_l}(x,\xi)\right| \le \beta_1 + \beta_2|\xi|^{p-2} \qquad (k,l=1,\dots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{157}$$

(iii) (Ellipticity.) Defining $a_k(x,\eta) := b_k(x,\eta)\eta$ for all k, the Jacobian matrices $\frac{\partial}{\partial\eta}\,a_k(x,\eta)$ are uniformly spectrally bounded from both below and above.

(iv) (Coercivity.) We have $\mathrm{div}\,\mathbf{w}_k \le 0$ on Ω and $\mathbf{w}_k\cdot\nu \ge 0$ on Γ_N $(k=1,\dots,s)$.

(v) (Cooperativity.) We have

$$\frac{\partial q_k}{\partial \xi_l}(x,\xi) \le 0 \qquad (k,l=1,\dots,s,\ k\ne l;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{158}$$

(vi) (Weak diagonal dominance for the Jacobians w.r.t. rows and columns.) We have

$$\sum_{l=1}^{s}\frac{\partial q_k}{\partial \xi_l}(x,\xi) \ge 0, \qquad \sum_{l=1}^{s}\frac{\partial q_l}{\partial \xi_k}(x,\xi) \ge 0 \qquad (k=1,\dots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{159}$$

(vii) For all $k = 1, \dots, s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

We note that Remark 3.6 is valid for the above system as well.

Now we proceed as in subsection 3.3. System (155) is first reduced to a system with nonlinear coefficients like (153). Owing to the superlinear growth allowed in (157), we must reprove Theorem 4.1 under Assumptions 4.3. For this, we first define the operators

$$\langle B(u)z, v\rangle = \int_\Omega \sum_{k=1}^s b_k(x, \nabla u)\, \nabla z_k \cdot \nabla v_k, \qquad \langle Nz, v\rangle = \int_\Omega \sum_{k=1}^s (\mathbf{w}_k(x) \cdot \nabla z_k)\, v_k,$$
$$\langle R(u)z, v\rangle = \int_\Omega \sum_{k,l=1}^s V_{kl}(x, u)\, z_l\, v_k \tag{160}$$

($z \in H^1(\Omega)^s$, $v \in H^1_D(\Omega)^s$).

Lemma 4.2 *Let Assumptions 4.3 hold. For any $u \in H^1(\Omega)^s$, the above operators $B(u)$, $N(u) \equiv N$ and $R(u)$, together with the operator $A(u) = B(u)u + Nu + R(u)u$, satisfy Assumptions 2.4.1, modified according to Remark 2.6, and Assumptions 2.4.2, in the spaces $H = H^1(\Omega)^s$ and $H_0 = H^1_D(\Omega)^s$.*

Proof. This follows from Lemma 4.1 and Lemma 3.2, using the arguments of the former for $B(u)$ and N, and (under the polynomial growth) the arguments of the latter for $R(u)$. ■

We recall the new norms for (53)–(54): we have

$$|||z|||_{N_1} = \|z\|_{H^1(\Omega)^s} \quad \text{and} \quad |||v|||_{N_2} = \|v\|_{L^2(\Omega)^s} \tag{161}$$

from Lemma 4.1, and

$$|||v|||^2_{R_1} = |||v|||^2_{R_2} = \|v\|^2_{L^{2q}(\Omega)^s} := \Big\| \sum_{k=1}^s v_k^2 \Big\|_{L^q(\Omega)} \qquad (v \in H^1(\Omega)^s) \tag{162}$$

from Lemma 3.2, see [26], where r is any fixed real number satisfying

$$\frac{d}{2+d-\gamma} < r \le \frac{p^*}{p-2}\,. \tag{163}$$

and then q in (162) is chosen as

$$\frac{1}{r} + \frac{1}{q} = 1. \tag{164}$$

We also need to strengthen Definition 4.1 on the quasi-regularity of the mesh.

Definition 4.2 Let $\Omega \subset \mathbf{R}^d$ and let us consider a family of FEM subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$ constructed as in subsection 3.1.2. The corresponding mesh will be called *quasi-regular* w.r.t. problem (155) if

$$c_1 h^\gamma \le meas(\operatorname{supp}\varphi_p) \le c_2 h^d\,, \tag{165}$$

where the positive real number γ satisfies

$$d \le \gamma < \min\Big\{\gamma_d^*(p), \quad \frac{d(d+2)}{d+1}\Big\} \tag{166}$$

with p from (157) and $\gamma_d^*(p)$ from (108).

Now we can derive the desired nonnegativity result for the stiffness matrix. Here the entries of $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ are, for any $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ and $i = 1, ..., n_0$, $j = 1, ..., n$,

$$\begin{aligned} a_{ij}(\bar{\mathbf{c}}) := \int_\Omega \Bigg(&\sum_{k=1}^s b_k(x, \nabla u^h)\,(\nabla\phi_j)_k \cdot (\nabla\phi_i)_k \\ &+ \sum_{k=1}^s \Big(\mathbf{w}_k(x)\cdot(\nabla\phi_j)_k\Big)\,(\phi_i)_k + \sum_{k,l=1}^s V_{kl}(x,u^h)\,(\phi_j)_l\,(\phi_i)_k\Bigg) \end{aligned} \tag{167}$$

where, as in (111),

$$V_{kl}(x, u^h(x)) = \int_0^1 \frac{\partial q_k}{\partial \xi_l}(x, tu^h(x))\,dt \qquad (k,l = 1,\ldots,s;\ x\in\Omega). \tag{168}$$

Theorem 4.2 *Let problem (155) satisfy Assumptions 4.3. Let us consider a family of finite element subspaces $\mathcal{V} = \{V_h\}_{h\to 0}$, such that the corresponding family of meshes is quasi-regular according to Definition 4.2, further, for any $p = 1, ..., \bar{n}_0$, $t = 1, ..., \bar{n}$ $(p \ne t)$, if $meas(\operatorname{supp}\varphi_p \cap \operatorname{supp}\varphi_t) > 0$ then (139) holds, where γ is from (166) and $K_0 > 0$ is a constant independent of p, t and h.*

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (167) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

Proof. The proof is similar to that of Theorem 4.1, combining it with the proof of Theorem 3.3 [26, Theorem 4.3]. We will only point out the differences.

First, owing to (161)–(162), instead of (142) we must estimate

$$\begin{aligned} T(h)^2 = \sup\Big\{ \max\Big\{&\|\phi_i\|_{H^1(\Omega)^s}\|\phi_j\|_{L^2(\Omega)^s}\,,\ \|\phi_i\|_{L^{2q}(\Omega)^s}\|\phi_j\|_{L^{2q}(\Omega)^s}\Big\}: \\ &\phi_i, \phi_j \in V_h\Big\}. \end{aligned} \tag{169}$$

We have for all i, j

$$\|\phi_i\|_{H^1(\Omega)^s}\|\phi_j\|_{L^2(\Omega)^s} \le c_4 h^{d-\frac{\gamma}{d}}$$

from Theorem 4.1 and

$$\|\phi_i\|_{L^{2q}(\Omega)^s}\|\phi_j\|_{L^{2q}(\Omega)^s} \le c_5\, h^{d/q}$$

from Theorem 3.3, hence

$$T(h)^2 \le c_6 \max(h^{d/q}, h^{d-\frac{\gamma}{d}}).$$

Here in (166), γ has been chosen such that both $\gamma - 2 - \frac{d}{q} < 0$, see (143) in [26], and $\gamma - 2 - d + \frac{\gamma}{d} < 0$, see (147). Therefore

$$\lim_{h\to 0} \frac{M_B(h)}{T(h)^2} \ge c_7 \lim_{h\to 0} \min\{h^{\gamma-2-\frac{d}{q}},\ h^{\gamma-2-d+\frac{\gamma}{d}}\} = +\infty. \qquad (170)$$

(Here c_4, c_5 etc. denote positive constants.)

It is left to verify assumption (c), i.e. that $M_N(\|u^h\|)$ and $M_R(\|u^h\|)$ are bounded as $h \to 0$. For the former, we have the constant bound $M_N(\|u^h\|) \equiv \|\mathbf{w}\|_{L^\infty(\Omega)^s}$, see part (iv) of the proof of Lemma 4.1. For the latter, the boundedness of $M_R(\|u^h\|)$ was proved in [26, Theorem 4.3].

The other parts of the proof coincide with that of Theorem 4.1. ■

As before, we can derive the corresponding DMP:

Corollary 4.3 *Let problem (155) satisfy Assumptions 4.3, and let its FEM discretization satisfy the corresponding conditions of Theorem 4.2. If* $f_k \le q_k(x,0)$ *and* $\gamma_k \le 0$ $(k = 1,\dots,s)$*, then (154) holds.*

4.4 Nonsymmetric systems with nonlinear convection coefficients

Finally we study a system containing nonlinear convection terms. The required strengthening in the other assumptions is the strong uniform diagonal dominance (172)–(173) and the homogenity of the Dirichlet data. The applicability of these conditions will be illustrated in the example in subsection 5.3.

Let us consider the system

$$\left.\begin{aligned} -\mathrm{div}\left(b_k(x,\nabla u)\,\nabla u_k\right) + \mathbf{w}_k(x,u)\cdot\nabla u_k + q_k(x,u_1,...,u_s) &= f_k(x) \quad \text{a.e. in } \Omega,\\ b_k(x,\nabla u)\tfrac{\partial u_k}{\partial\nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N,\\ u_k &= 0 \qquad \text{a.e. on } \Gamma_D \end{aligned}\right\} \qquad (171)$$

$(k = 1,\dots,s)$.

Assumptions 4.4.

(i) The domain Ω and the diffusion coefficients b_k satisfy Assumptions 3.0.

(ii) (Smoothness and growth.) For all $k, l = 1, \ldots, s$ we have $b_k \in (C^1 \cap L^\infty)(\Omega \times \mathbf{R}^d)$ and $q_k \in C^1(\Omega \times \mathbf{R}^s)$, further, (156) and (157) hold.

(iii) (Ellipticity.) Defining $a_k(x, \eta) := b_k(x, \eta)\eta$ for all k, the Jacobian matrices $\frac{\partial}{\partial \eta}\, a_k(x, \eta)$ are uniformly spectrally bounded from both below and above.

(iv) (Bounded convection term.) We have $\mathbf{w}_k \in L^\infty(\Omega \times \mathbf{R})$ $(k = 1, \ldots, s)$.

(v) (Cooperativity.) We have $\dfrac{\partial q_k}{\partial \xi_l}(x, \xi) \le 0 \quad (k, l = 1, \ldots, s,\ k \ne l;\ x \in \Omega,\ \xi \in \mathbf{R}^s)$.

(vi) (Uniform diagonal dominance for the Jacobians w.r.t. rows and columns.) There exists $\mu > 0$ such that

$$\sum_{l=1}^{s} \frac{\partial q_k}{\partial \xi_l}(x, \xi) \ge \mu, \qquad \sum_{l=1}^{s} \frac{\partial q_l}{\partial \xi_k}(x, \xi) \ge \mu \qquad (k = 1, \ldots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{172}$$

Moreover,

$$\mu > \frac{\|\mathbf{w}\|^2_{L^\infty(\Omega)^s}}{4m} \tag{173}$$

where $\|\mathbf{w}\|_{L^\infty(\Omega)^s} := \sup\limits_{\substack{k=1,\ldots,s \\ (x,\xi) \in \Omega \times \mathbf{R}^s}} |\mathbf{w}_k(x, \xi)|$ and m is from Assumption 3.0 (ii).

(vii) For all $k = 1, \ldots, s$ we have $f_k \in L^2(\Omega)$ and $\gamma_k \in L^2(\Gamma_N)$.

We proceed similarly as in the previous subsection, system (171) is reduced to a system with nonlinear coefficients as before via the functions $V_{kl} : \Omega \times \mathbf{R}^s \to \mathbf{R}$ and $\hat{f}_k$ from (95) and (97), respectively. The difference is is the nonlinear convection term. Taking this into account, we must reprove Theorem 4.2 under Assumptions 4.4, but only those parts are addressed where the convection term is involved.

The operator corresponding to our problem is

$$\begin{aligned} \langle A(u), v \rangle = \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, \nabla u)\, \nabla u_k \cdot \nabla v_k + \sum_{k=1}^{s} (\mathbf{w}_k(x, u) \cdot \nabla u_k)\, v_k \\ + \sum_{k,l=1}^{s} V_{kl}(x, u)\, u_l\, v_k \Big) \end{aligned} \tag{174}$$

$(u \in H^1(\Omega)^s,\ v \in H^1_D(\Omega)^s)$. First we properly modify (160), where the main point is to compensate for the presence of the convection term in the positivity of the operator without a coercivity condition on $\mathbf{w}_k$. We define the operators

$$\begin{aligned}\langle B(u)z,v\rangle &= \int_\Omega \sum_{k=1}^{s}\Big(b_k(x,\nabla u)\,\nabla z_k\cdot\nabla v_k + \,\mu z_k v_k\Big)\\ \langle N(u)z,v\rangle &= \int_\Omega \sum_{k=1}^{s}(\mathbf{w}_k(x,u)\cdot\nabla z_k)\,v_k \\ \langle R(u)z,v\rangle &= \int_\Omega \Big(\sum_{k,l=1}^{s} V_{kl}(x,u)\,z_l\,v_k - \mu\sum_{k=1}^{s} z_k v_k\Big)\end{aligned} \tag{175}$$

($z \in H^1(\Omega)^s$, $v \in H^1_D(\Omega)^s$). We note that (95) and (172) yield

$$\sum_{l=1}^{s} V_{kl}(x,\xi) \geq \mu \qquad (k=1,\dots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s), \tag{176}$$

and hence, since $V_{kl}(x,\xi)\leq 0$ for $k\neq l$ by Assumption 4.4 (v), we also have

$$V_{kk}(x,\xi) \geq \mu \qquad (k=1,\dots,s;\ x\in\Omega,\ \xi\in\mathbf{R}^s). \tag{177}$$

Lemma 4.3 *Let Assumptions 4.4 hold. For any $u\in H^1(\Omega)^s$, the operators $B(u)$, $N(u)$ and $R(u)$, together with the operator $A(u)$ in (174), satisfy Assumptions 2.4.1, modified according to Remark 2.6, in the spaces $H = H^1(\Omega)^s$ and $H_0 = H^1_D(\Omega)^s$.*

Proof. We must reprove those parts of Lemma 4.2 that involve the convection term or the modifications of $B(u)$ and $R(u)$ with the term containing μ.

(i) It is obvious from (174) and (175) that $A(u) = B(u)u+N(u)u+R(u)u$.

(ii) We must prove (31). Here for all $u\in H^1(\Omega)^s$ and $v\in H^1_D(\Omega)^s$,

$$\begin{aligned}&\Big\langle\Big(B(u)+N(u)\Big)v,v\Big\rangle\\ &\quad= \int_\Omega\sum_{k=1}^{s}\Big(b_k(x,\nabla u)\,|\nabla v_k|^2 + \,\mu v_k^2\Big) + \int_\Omega\sum_{k=1}^{s}(\mathbf{w}_k(x,u)\cdot\nabla v_k)\,v_k\\ &\quad\geq m\|\nabla v\|^2_{L^2(\Omega)^s} + \mu\|v\|^2_{L^2(\Omega)^s} - \omega\|\nabla v\|_{L^2(\Omega)^s}\,\|v\|_{L^2(\Omega)^s}\end{aligned} \tag{178}$$

where $\omega := \|\mathbf{w}\|_{L^\infty(\Omega)^s}$. Using the basic inequality $xy \leq \frac{1}{2}\Big(\varepsilon x^2 + \frac{1}{\varepsilon}y^2\Big)$ ($\varepsilon>0,\ \ x,y\in\mathbf{R}$) for the last two factors, we obtain

$$\Big\langle\Big(B(u)+N(u)\Big)v,v\Big\rangle \geq \Big(m-\frac{\omega\varepsilon}{2}\Big)\|\nabla v\|^2_{L^2(\Omega)^s} + \Big(\mu - \frac{\omega}{2\varepsilon}\Big)\|v\|^2_{L^2(\Omega)^s}\,.$$

Choosing $\varepsilon := \frac{\omega}{2\mu}$, we have

$$\Big\langle\Big(B(u)+N(u)\Big)v,v\Big\rangle \geq \hat{m}\,\|\nabla v\|^2_{L^2(\Omega)^s} \equiv \hat{m}\,\|v\|^2$$

where $\hat{m} := m - \frac{\omega^2}{4\mu} > 0$ by (173).

(iii) Let us consider the sets P and D, defined in paragraph (iii) of the proof of Lemma 4.1. That is, $v \in D$ iff $v = (0, \dots, 0, g, 0, \dots, 0)^T$ with g at the k-th entry for some $1 \le k \le s$ and $g \in H^1(\Omega)$, $g \ge 0$. Further, $v \in P$ iff $v = (y, \dots, y)$ for some $y \in H^1(\Omega)$, $y \ge 0$. We must prove that for any $u \in H^1(\Omega)^s$ and $v \in D$, we have

$$\langle R(u)z, v\rangle \ge 0 \tag{179}$$

provided that either $z \in P$ or $z = v \in D$.

If $z \in P$, then

$$\langle R(u)z, v\rangle = \int_\Omega \Big(\sum_{l=1}^s V_{kl}(x,u) - \mu \Big)\, yg \ge 0$$

by (176) and that $y, g \ge 0$. If $z = v \in D$, then by (177)

$$\langle R(u)v, v\rangle = \int_\Omega \Big(V_{kk}(x,u) - \mu \Big)\, g^2 \ge 0.$$

(iv) This follows in the same way as in Lemma 4.2. For $N(u)$, we can similarly factor out $\|\mathbf{w}\|_{L^\infty(\Omega)^s}$. For $R(u)$, the new norms can remain $|||v|||_{R_1}^2 = |||v|||_{R_2}^2 = \|v\|_{L^{2q}(\Omega)^s}^2$ as in (162), since the additional term in (175) can be bounded by the product L^2-norm $\|.\|_{L^2(\Omega)^s}$, which is (up to a constant factor) not larger than the norm $\|.\|_{L^{2q}(\Omega)^s}^2$ owing to the Sobolev inequality. ∎

Now we can derive the nonnegativity of the stiffness matrix. Here the entries of $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ are, for any $\bar{\mathbf{c}} = (c_1, ..., c_n)^T \in \mathbf{R}^n$ and $i = 1, ..., n_0$, $j = 1, ..., n$,

$$\begin{aligned} a_{ij}(\bar{\mathbf{c}}) := \int_\Omega \Bigg(& \sum_{k=1}^s b_k(x, \nabla u^h)\, (\nabla \phi_j)_k \cdot (\nabla \phi_i)_k \\ & + \sum_{k=1}^s \Big(\mathbf{w}_k(x, u^h) \cdot (\nabla \phi_j)_k \Big) (\phi_i)_k + \sum_{k,l=1}^s V_{kl}(x, u^h)\, (\phi_j)_l\, (\phi_i)_k \Bigg) \end{aligned} \tag{180}$$

where $V_{kl}(x, u^h)$ is as in (168).

Theorem 4.3 *Let problem (171) satisfy Assumptions 4.4. Let us consider a family of finite element subspaces $\mathcal{V} = \{V_h\}_{h \to 0}$, such that the corresponding family of meshes is quasi-regular according to Definition 4.2, further, for any $p = 1, ..., \bar{n}_0$, $t = 1, ..., \bar{n}$ $(p \ne t)$, if $meas(\mathrm{supp}\, \varphi_p \cap \mathrm{supp}\, \varphi_t) > 0$ then (139) holds, where γ is from (166) and $K_0 > 0$ is a constant independent of p, t and h.*

Then for sufficiently small h, the matrix $\bar{\mathbf{A}}(\bar{\mathbf{c}})$ defined in (180) is of generalized nonnegative type with irreducible blocks in the sense of Definition 2.4.

Proof. The proof is similar to that of Theorem 4.2, with a few differences. First, the proof for assumption (a) relies on (141), where by (175), now $\langle B(u^h)\phi_j, \phi_i\rangle$ contains the additional term $\int_\Omega \mu \sum_{k=1}^{s} (\phi_j)_k (\phi_i)_k$. However, this term is bounded by μs, hence altogether (141) is preserved with another constant instead of $\hat{c}_1$ and still tends to $-\infty$. In the other parts of the proof we only need the sum of $B(u)$ and $R(u)$, in which the additional terms vanish by definition.

Finally, Theorem 4.2 contains the boundedness of $M_R(\|u^h\|)$, see the end of its proof, which fact is quoted from [26, Theorem 4.3]. This part of the proof uses Assumptions 2.4.2, which have not yet been proved now. Assumptions 2.4.2 are used in [26, Theorem 4.3] to have uniform monotonicity of A in order to prove that

$$\langle A(u^h) - A(g^h), u^h - g^h\rangle \geq m\,\|u^h - g^h\|^2\,, \tag{181}$$

since this implies the boundedness of $\|u^h\|$ if we assume the boundedness of $\|g^h\|$ (as $h \to 0$). These properties are derived in [26, Remark 3.1]. Now we have $g^h = 0$ by the homogeneous Dirichlet data in (171), hence we only need (181) for the special case $g^h = 0$. Therefore, to prove our theorem, it suffices instead of Assumptions 2.4.2 to verify

$$\langle A(u^h), u^h\rangle \geq \tilde{m}\,\|u^h\|^2 \qquad (h > 0) \tag{182}$$

for some constant $\tilde{m} > 0$, independent of the FEM solution u^h of our problem.

Since $u^h = 0$ on Γ_D, we can substitute $u = v = u^h$ in (174):

$$\begin{aligned}\langle A(u^h), u^h\rangle = \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, \nabla u^h)\,|\nabla u_k^h|^2 + \sum_{k=1}^{s} (\mathbf{w}_k(x, u^h)\cdot \nabla u_k^h)\,u_k^h\Big) \\ + \int_\Omega \sum_{k,l=1}^{s} V_{kl}(x, u^h)\,u_l^h\,u_k^h\end{aligned} \tag{183}$$

$$= \int_\Omega \Big(\sum_{k=1}^{s} b_k(x, \nabla u^h)\,|\nabla u_k^h|^2 + \mu|u_k^h|^2 + \sum_{k=1}^{s} (\mathbf{w}_k(x, u^h)\cdot \nabla u_k^h)\,u_k^h\Big) \tag{184}$$

$$+ \int_\Omega \sum_{k,l=1}^{s} \Big(V_{kl}(x, u^h)\,u_l^h\,u_k^h - \mu|u_k^h|^2\Big). \tag{185}$$

We can estimate (184) in the same way as in (178), and obtain the lower bound $\hat{m}\,\|u^h\|^2$ where $\hat{m} := m - \frac{\omega^2}{4\mu} > 0$. For (185), note that (172) and (95) imply that μ is a lower uniform spectral bound for the matrices $V(x, \xi)$, i.e.

$$V(x, \xi)\,\zeta \cdot \zeta \equiv \sum_{k,l=1}^{s} V_{kl}(x, \xi)\,\zeta_l\,\zeta_k \geq \mu|\zeta|^2 \tag{186}$$

(for all $(x, \xi) \in \Omega \times \mathbf{R}^s$ and $\zeta \in \mathbf{R}^s$), which yields that the expression in (185) is nonnegative. Altogether, (182) holds with $\tilde{m} := \hat{m}$. ■

As before, we can derive the corresponding DMP (154) under the conditions of Theorem 4.3. Since now $g = 0$, this becomes the discrete nonpositivity property $u_k^h \leq 0$. One can similarly obtain the discrete nonnegativity property, which is more noteworthy to formulate here:

Corollary 4.4 *Let problem (171) satisfy Assumptions 4.4, and let its FEM discretization satisfy the corresponding conditions of Theorem 4.3. If $f_k \geq q_k(x,0)$ and $\gamma_k \geq 0 \quad (k = 1,\dots,s)$, then for sufficiently small h, the FEM solution $u^h = (u_1^h, \dots, u_s^h)^T$ of system (171) satisfies*

$$u_k^h \geq 0 \quad \textit{on } \Omega \qquad (k = 1,\dots,s). \tag{187}$$

5 Some real-life examples

5.1 Reaction-diffusion systems in chemistry

The following result is quoted from [26]. The steady states of certain reaction-diffusion processes in chemistry are described by systems of the following form:

$$\left.\begin{aligned} -b_k\Delta u_k + \ P_k(x,u_1,\dots,u_s) &= f_k(x) \quad \text{in } \Omega, \\ b_k \tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{on } \Gamma_N, \\ u_k &= g_k(x) \quad \text{on } \Gamma_D \end{aligned}\right\} \quad (k = 1,\dots,s). \tag{188}$$

Here, for all k, the quantity u_k describes the concentration of the kth species, and P_k is a polynomial which characterizes the rate of the reactions involving the k-th species. A common way to describe such reactions is the so-called mass action type kinetics [21, 22], which implies that P_k has no constant term for any k, in other words, $P_k(x,0) \equiv 0$ on Ω for all k. Further, the reaction between different species is often proportional to the product of their concentration, in which case $P_k(x,u_1,\dots,u_s) = a_{kk}(x)u_k^\alpha + \sum_{k \neq l} a_{kl}(x)u_k u_l$. The function $f_k \geq 0$ describes a source independent of concentrations.

We consider system (188) under the following conditions, such that it becomes a special case of system (100). As pointed out later, such chemical models describe processes with cross-catalysis and strong autoinhibiton.

Assumptions 5.1.

(i) $\Omega \subset \mathbf{R}^d$ is a bounded piecewise C^1 domain, where $d = 2$ or 3, and Γ_D, Γ_N are disjoint open measurable subsets of $\partial\Omega$ such that $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$.

(ii) (Smoothness and growth.) For all $k, l = 1,\dots,s$, the functions P_k are polynomials of arbitrary degree if $d = 2$ and of degree at most 4 if $d = 3$, further, $P_k(x,0) \equiv 0$ on Ω.

(iii) (Ellipticity.) $b_k > 0$ $(k = 1, \ldots, s)$ are given numbers.

(iv) (Cooperativity.) We have

$$\frac{\partial P_k}{\partial \xi_l}(x, \xi) \leq 0 \qquad (k, l = 1, \ldots, s,\ k \neq l;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{189}$$

(v) (Weak diagonal dominance for the Jacobians w.r.t. rows and columns.) We have

$$\sum_{l=1}^{s} \frac{\partial P_k}{\partial \xi_l}(x, \xi) \geq 0, \qquad \sum_{l=1}^{s} \frac{\partial P_l}{\partial \xi_k}(x, \xi) \geq 0 \quad (k = 1, \ldots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{190}$$

(vi) For all $k = 1, \ldots, s$ we have $f_k \in L^2(\Omega)$, $\gamma_k \in L^2(\Gamma_N)$, $g_k = g^*_{k|\Gamma_D}$ with $g^* \in H^1(\Omega)$.

Similarly to (94), assumptions (189)-(190) now imply

$$\frac{\partial P_k}{\partial \xi_k}(x, \xi) \geq 0 \qquad (k = 1, \ldots, s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{191}$$

Returning to the model described by system (188), the chemical meaning of the cooperativity (189) is cross-catalysis, whereas (191) means autoinhibiton. Cross-catalysis arises e.g. in gradient systems [49]. Condition (190) means that autoinhibition is strong enough to ensure both weak diagonal dominances.

By definition, the concentrations u_k are nonnegative, therefore a proper numerical model must produce such numerical solutions. We can use Corollary 3.5 to obtain the required property:

Corollary 5.1 *Let problem (188) satisfy Assumptions 5.1, and let its FEM discretization satisfy the conditions of Theorem 3.3. If*

$$f_k \geq 0, \quad \gamma_k \geq 0, \quad g_k \geq 0 \qquad (k = 1, \ldots, s)$$

then for sufficiently small h, the FEM solution $u^h = (u_1^h, \ldots, u_s^h)^T$ of system (188) satisfies

$$u_k^h \geq 0 \quad \text{on } \Omega \qquad (k = 1, \ldots, s). \tag{192}$$

5.2 Linear elliptic systems

Maximum principles or nonnegativity preservation for linear elliptic systems have attracted great interest, as mentioned in the introduction. Hence it is worthwile to derive the corresponding DMPs from the previous results. Following [26], let us therefore consider linear elliptic systems of the form

$$\left.\begin{aligned} -\mathrm{div}\,(b_k(x)\,\nabla u_k) + \sum_{l=1}^{s} V_{kl}(x)\,u_l \;&=\; f_k(x) \quad \text{a.e. in } \Omega, \\ b_k(x)\tfrac{\partial u_k}{\partial \nu} \;&=\; \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k \;&=\; g_k(x) \quad \text{a.e. on } \Gamma_D \end{aligned}\right\} \qquad (k = 1, \dots, s) \tag{193}$$

where for all $k, l = 1, \dots, s$ we have $b_k \in W^{1,\infty}(\Omega)$ and $V_{kl} \in L^\infty(\Omega)$.

Let Assumptions 3.1 hold (where in fact we do not need assumption (ii)). Then (193) is a special case of (56), hence Corollary 3.2 holds, as well as the analogous results mentioned in Remark 3.2. Here we formulate two of these that follow the most studied CMP results:

Corollary 5.2 *Let problem (193) satisfy Assumptions 3.1, let its FEM discretization satisfy the conditions of Theorem 3.1 and let h be sufficiently small. If $u^h = (u_1^h, \dots, u_s^h)^T$ is the FEM solution of system (193), then the following properties hold.*

(1) If $f_k \le 0$, $\gamma_k \le 0$ $(k = 1, \dots, s)$ and $\max\limits_{k=1,\dots,s} \max\limits_{\Gamma_D} g_k^h > 0$, then

$$\max_{k=1,\dots,s} \max_{\overline{\Omega}} u_k^h = \max_{k=1,\dots,s} \max_{\Gamma_D} g_k^h\,. \tag{194}$$

(2) If $f_k \ge 0$, $\gamma_k \ge 0$ and $g_k \ge 0$ $(k = 1, \dots, s)$, then

$$u_k^h \ge 0 \quad \textit{on } \Omega \qquad (k = 1, \dots, s). \tag{195}$$

5.3 Nonsymmetric transport systems

The description of nonlinear transport processes for certain agents (pollutants), involving diffusion, convection and reaction, often leads to systems of the form

$$\left.\begin{aligned} -b_k\Delta u_k + \mathbf{w}_k(x,u)\cdot\nabla u_k + P_k(x, u_1, ..., u_s) &= f_k(x) \quad \text{a.e. in } \Omega, \\ b_k \tfrac{\partial u_k}{\partial \nu} &= \gamma_k(x) \quad \text{a.e. on } \Gamma_N, \\ u_k &= \;0 \quad\quad\;\; \text{a.e. on } \Gamma_D \end{aligned}\right\} \tag{196}$$

$(k = 1, \dots, s)$. We consider diffusion-dominated processes, i.e. when the fixed numbers $b_k > 0$ are comparable to the magnitude of the coefficients $\mathbf{w}_k$. Here $u_k \ge 0$ are the concentrations of the agents. One expects any numerical solution method to reproduce the nonnegativity of the solution.

Assumptions 5.3.

(i) The numbers b_k and functions P_k, f_k and γ_k satisfy Assumptions 5.1.

(ii) We have $\mathbf{w}_k \in L^\infty(\Omega \times \mathbf{R})$ $(k = 1, \dots, s)$.

(iii) There exists $\mu > 0$ such that

$$\sum_{l=1}^{s} \frac{\partial P_k}{\partial \xi_l}(x,\xi) \geq \mu, \quad \sum_{l=1}^{s} \frac{\partial P_l}{\partial \xi_k}(x,\xi) \geq \mu \quad (k = 1,\ldots,s;\ x \in \Omega,\ \xi \in \mathbf{R}^s). \tag{197}$$

Moreover,

$$\mu > \frac{\|\mathbf{w}\|^2_{L^\infty(\Omega)^s}}{4m} \tag{198}$$

where $\|\mathbf{w}\|_{L^\infty(\Omega)^s} := \sup_{\substack{k=1,\ldots,s \\ (x,\xi)\in\Omega\times\mathbf{R}^s}} |\mathbf{w}_k(x,\xi)|$ and $m := \min_k b_k > 0$.

Systems of the form (196) typically arise from the time discretization of the time-dependent transport system

$$\frac{\partial u_k}{\partial t} - b_k \Delta u_k + \mathbf{w}_k(x,u) \cdot \nabla u_k + R_k(x,u_1,...,u_s) \;=\; g_k(x,t) \tag{199}$$

with the boundary conditions of (196) and an initial condition $u_k(x,0) = u_0(x)$ $(x \in \Omega)$. Here $\mathbf{w}_k(x,u)$ is the convective term, e.g. wind, and R_k is a polynomial which characterizes the rate of the reactions involving the k-th species, as in subsection 5.1. Here the R_k do not satisfy a condition like (197), this will come instead from the numerical process below.

The standard numerical solution first uses a time discretization, resulting in the following equations, where u_k^i denotes the solution on the ith time level t_i:

$$\frac{u_k^i - u_k^{i-1}}{\tau} - b_k \Delta u_k^i + \mathbf{w}_k(x,u^i) \cdot \nabla u_k^i + R_k(x,u_1^i,...,u_s^i) \;=\; g_k^i(x)\,.$$

Rearranging this as

$$-b_k \Delta u_k^i + \mathbf{w}_k(x,u^i) \cdot \nabla u_k^i + \Big(R_k(x,u_1^i,...,u_s^i) \,+ \frac{1}{\tau}\, u_k^i\Big) \;=\; g_k^i(x) + \frac{1}{\tau}\, u_k^{i-1},$$

we obtain a system for the unknown function u_k^i in the form (196) with coefficients

$$P_k(x,\xi_1,...,\xi_s) := R_k(x,\xi_1,...,\xi_s) \,+ \frac{1}{\tau}\,\xi_k \tag{200}$$

and $f_k(x) := g_k^i(x) + \frac{1}{\tau}\, u_k^{i-1}(x)$. Then the strong uniform diagonal dominance (197)–(198) can be ensured as follows. Assume that we have an estimate

$$\inf_{\substack{k=1,\ldots,s \\ (x,\xi)\in\Omega\times\mathbf{R}^s}} \sum_{l=1}^{s} \frac{\partial R_k}{\partial \xi_l}(x,\xi) \geq -\mu_0, \qquad \inf_{\substack{k=1,\ldots,s \\ (x,\xi)\in\Omega\times\mathbf{R}^s}} \sum_{l=1}^{s} \frac{\partial R_l}{\partial \xi_k}(x,\xi) \geq -\mu_0$$

for some $\mu_0 \geq 0$, and let μ be a number satisfying (198). Then we can choose the time-step τ to be small enough, namely, $\tau \leq \frac{1}{\mu_0+\mu}$. In this case, using (200), we obtain

$$\sum_{l=1}^{s} \frac{\partial P_k}{\partial \xi_l}(x,\xi) \geq -\mu_0 + \frac{1}{\tau} \geq -\mu_0 + (\mu_0 + \mu) = \mu,$$

and similarly for the other sum in (197).

Under the above conditions, system (196) is a special case of system (171), hence we can apply Corollary 4.4. Here, as mentioned in subsection 5.1, $P_k(x,0) \equiv 0$ on Ω for all k, further, we have homogeneous Dirichlet boundary conditions. Hence the result has the following form:

Corollary 5.3 *Let problem (196) satisfy Assumptions 5.3, and let its FEM discretization satisfy the corresponding conditions of Theorem 4.3. If $f_k \geq 0$ and $\gamma_k \geq 0 \quad (k = 1, \dots, s)$, then for sufficiently small h, the FEM solution $u^h = (u_1^h, \dots, u_s^h)^T$ of system (196) satisfies*

$$u_k^h \geq 0 \quad \text{on } \Omega \qquad (k = 1, \dots, s). \tag{201}$$

References

[1] Axelsson, O., *Iterative Solution Methods,* Cambridge University Press, 1994.

[2] Brandts, J., Korotov, S., Křížek, M., Dissection of the path-simplex in $\mathbf{R^n}$ into n path-subsimplices, *Linear Algebra Appl.* 421 (2007), no. 2-3, 382–393.

[3] Brandts, J., Korotov, S., Křížek, M., On the equivalence of regularity criteria for triangular and tetrahedral finite element partitions, *Comput. Math. Appl.* 55 (2008), 2227–2233.

[4] Brandts, J., Korotov, S., Křížek, M., The discrete maximum principle for linear simplicial finite element approximations of a reaction-diffusion problem, *Linear Algebra Appl.* 429 (2008), 2344–2357.

[5] Brandts, J., Korotov, S., Křížek, M., Šolc, J., On nonobtuse simplicial partitions, *SIAM Rev.*, to appear.

[6] Burman, E., Ern, A., Nonlinear diffusion and the discrete maximum principle for stabilized Galerkin approximations of the convection-diffusion-reaction equation, *Comput. Methods Appl. Mech. Engrg.* 191 (2002), 3833–3855.

[7] Burman, E., Ern, A., Discrete maximum principle for Galerkin approximations of the Laplace operator on arbitrary meshes, *C. R. Acad. Paris, Ser I* 338 (2004), 641–646.

[8] Burman, E., Ern, A., Stabilized Galerkin approximation of convection-diffusion-reaction equations: Discrete maximum principle and convergence, *Math. Comput.* 74 (2005), 1637-1652.

[9] Caristi, G., Mitidieri, E., Further results on maximum principles for noncooperative elliptic systems, *Nonlinear Anal.* 17 (1991), no. 6, 547–558.

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

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

[12] 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.

[13] Codina, R., *A finite element formulation for the numerical solution of the convection-diffusion equation*, Monograph, 14. Centro Internacional de Métodos Numéricos en Ingeniería, Barcelona, 1993.

[14] Donea, J., Huerta, A., *Finite Element Methods for Flow Problems*, John Wiley and Sons, 2003.

[15] Draganescu, A., Dupont, T. F., Scott, L. R., Failure of the discrete maximum principle for an elliptic finite element problem, *Math. Comp.* 74 (2005), no. 249, 1–23.

[16] de Figueiredo, D. G., Mitidieri, E., Maximum principles for cooperative elliptic systems, *C. R. Acad. Sci. Paris Sér. I Math.* 310 (1990), no. 2, 49–52.

[17] 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.

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

[19] Hannukainen, A., Korotov, S., Vejchodský, T., Discrete maximum principles for FE solutions of the diffusion-reaction problem on prismatic meshes, *J. Comput. Appl. Math.* 226 (2009), 275–287.

[20] Hannukainen, A., Korotov, S., Vejchodský, T., On Weakening Conditions for Discrete Maximum Principles for Linear Finite Element Schemes, in: *Numerical Analysis and Applications*, eds. S. Margenov, L.G. Vulkov and J.Wasniewski, Lecture Notes Comp. Sci. No. 5434, pp. 297-304, Springer, 2009.

[21] Hárs, V., Tóth, J., On the inverse problem of reaction kinetics, In: *Qualitative Theory of Differential Equations* (Szeged, Hungary, 1979), Coll. Math. Soc. János Bolyai 30, ed. M. Farkas, North-Holland - János Bolyai Mathematical Society, Budapest, 1981, pp. 363-379.

[22] Horn, F., Jackson, R., General mass action kinetics, *Arch. Rat. Mech. Anal.* 47 (1972), 81–116.

[23] Ikeda, T., *Maximum principle in finite element models for convection-diffusion phenomena,* Lecture Notes in Numerical and Applied Analysis, 4. North-Holland Mathematics Studies, 76. Kinokuniya Book Store Co., Ltd., Tokyo, 1983.

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

[25] 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.

[26] Karátson, J., Korotov, S., A discrete maximum principle in Hilbert space with applications to nonlinear cooperative elliptic systems, to appear in *SIAM J. Numer. Anal.*

[27] Karátson J., Korotov, S., Křížek, M., On discrete maximum principles for nonlinear elliptic problems, *Math. Comput. Simul.* 76 (2007) pp. 99–108.

[28] Kikuchi, F., Discrete maximum principle and artificial viscosity in finite element approximations to convective diffusion equations, Institute of Space and Aeronautical Science, University of Tokyo, Report no. 550, vol. 42, Sept. 1977, p. 153-166.

[29] Knobloch, P., Numerical solution of convection-diffusion equations using upwinding techniques satisfying the discrete maximum principle, *Proceedings of Czech-Japanese Seminar in Applied Mathematics* 2005, 69–76, COE Lect. Note, 3, Kyushu Univ. The 21 Century COE Program, Fukuoka, 2006.

[30] Korotov, S., Křížek, M., Acute type refinements of tetrahedral partitions of polyhedral domains, *SIAM J. Numer. Anal.* 39 (2001), 724–733.

[31] Korotov, S., Křížek, M., Tetrahedral partitions and their refinements, In: *Proc. Conf. Finite Element Methods: Three-dimensional Problems, Univ. of Jyväskylä, GAKUTO Internat. Ser. Math. Sci. Appl.,* vol. 15, Gakkotosho, Tokyo, 2001, 118–134.

[32] Korotov, S., Křížek, M., Global and local refinement techniques yielding nonobtuse tetrahedral partitions, *Comput. Math. Appl.* 50 (2005), 1105–1113.

[33] Korotov, S., Křížek, M., Neittaanmäki, P., Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle, *Math. Comp.* 70 (2001), 107–119.

[34] Korotov, S., Křížek, M., Kropáč, A., Strong regularity of a family of face-to-face partitions generated by the longest-edge bisection algorithm, *Comp. Math. Math. Phys.* 9 (2008), 1687–1698.

[35] Křížek, M., Lin Qun, On diagonal dominance of stiffness matrices in 3D, *East-West J. Numer. Math.* 3 (1995), 59–69.

[36] Křížek, M., Neittaanmäki, P., *Mathematical and numerical modelling in electrical engineering: theory and applications*, Kluwer Academic Publishers, 1996.

[37] Kuzmin, D., On the design of algebraic flux correction schemes for quadratic finite elements, *J. Comput. Appl. Math.* 218 (2008), no. 1, 79–87.

[38] Kuzmin, D., Shashkov, M.J., Svyatskiy, D., A constrained finite element method satisfying the discrete maximum principle for anisotropic diffusion problems, *J. Comput. Phys.* 228 (2009), 3448–3463.

[39] Mizukami, A., Hughes, T. J. R., A Petrov-Galerkin finite element method for convection-dominated flows: an accurate upwinding technique for satisfying the maximum principle, *Comput. Methods Appl. Mech. Engrg.* 50 (1985), no. 2, 181–193.

[40] Ladyzhenskaya, O. A., Ural'tseva, N. N., *Linear and quasilinear elliptic equations*, Leon Ehrenpreis Academic Press, New York-London, 1968.

[41] Liska, R., Shashkov, M., Enforcing the discrete maximum principle for linear finite element solutions of second order elliptic problems, *Commun. Comput. Phys.* 3 (2008), no. 4, 852–877.

[42] López-Gómez, J., Molina-Meyer, M., The maximum principle for cooperative weakly coupled elliptic systems and some applications, *Diff. Int. Equations* 7 (1994), no. 2, 383–398.

[43] Mitidieri, E., Sweers, G., Weakly coupled elliptic systems and positivity, *Math. Nachr.* 173 (1995), 259–286.

[44] Ohmori, K., The discrete maximum principle for nonconforming finite element approximations to stationary convective diffusion equations, *Math. Rep. Toyama Univ.* 2 (1979), 33–52.

[45] Ohmori, K., Correction to: "The discrete maximum principle for nonconforming finite element approximations to stationary convective diffusion equations" [Math. Rep. Toyama Univ. 2 (1979), 33–52], *Math. Rep. Toyama Univ.* 4 (1981), 179–182.

[46] Protter, M. H., Weinberger, H. F., *Maximum principles in differential equations*, Springer-Verlag, New York, 1984.

[47] 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.

[48] Stynes, M., Steady-state convection-diffusion problems, *Acta Numer.* 14 (2005), 445–508.

[49] Tóth, J., Gradient systems are cross-catalytic, *Reaction Kinetics and Catalysis Letters* 12 (3) (1979), 253–257.

[50] Vabishchevich, P. N., Samarskii, A. A., Monotone difference schemes for convection-diffusion problems on triangular grids, , *Comput. Math. Math. Phys.* 42 (2002), no. 9, 1317–1330

[51] Vejchodský, T., Šolín, P., Discrete maximum principle for higher-order finite elements in 1D, *Math. Comp.* 76 (2007), no. 260, 1833–1846.

[52] Varga, R., *Matrix iterative analysis*, Prentice Hall, New Jersey, 1962.

[53] Xu, J., Zikatanov, L., A monotone finite element scheme for convection-diffusion equations, *Math. Comp.* 68 (1999), 1429–1446.

(continued from the back cover)

A564 Antti H. Niemi
Best bilinear shell element: flat, twisted or curved?
February 2009

A563 Dmitri Kuzmin, Sergey Korotov
Goal-oriented a posteriori error estimates for transport problems
February 2009

A562 Antti H. Niemi
A bilinear shell element based on a refined shallow shell model
December 2008

A561 Antti Hannukainen, Sergey Korotov, Michal Krizek
On nodal superconvergence in 3D by averaging piecewise linear, bilinear, and trilinear FE approximations
December 2008

A560 Sampsa Pursiainen
Computational methods in electromagnetic biomedical inverse problems
November 2008

A559 Sergey Korotov, Michal Krizek, Jakub Solc
On a discrete maximum principle for linear FE solutions of elliptic problems with a nondiagonal coefficient matrix
November 2008

A558 José Igor Morlanes, Antti Rasila, Tommi Sottinen
Empirical evidence on arbitrage by changing the stock exchange
December 2008

A557 Mika Juntunen, Rolf Stenberg
Analysis of finite element methods for the Brinkman problem
April 2009

A556 Lourenço Beirão da Veiga, Jarkko Niiranen, Rolf Stenberg
A posteriori error analysis for the Morley plate element with general boundary conditions
December 2008

HELSINKI UNIVERSITY OF TECHNOLOGY INSTITUTE OF MATHEMATICS
RESEARCH REPORTS

The reports are available at *http://math.tkk.fi/reports/* .
The list of reports is continued inside the back cover.

A569 Antti Hannukainen, Mika Juntunen, Rolf Stenberg
Computations with finite element methods for the Brinkman problem
April 2009

A568 Olavi Nevanlinna
Computing the spectrum and representing the resolvent
April 2009

A567 Antti Hannukainen, Sergey Korotov, Michal Krizek
On a bisection algorithm that produces conforming locally refined simplicial meshes
April 2009

A566 Mika Juntunen, Rolf Stenberg
A residual based a posteriori estimator for the reaction–diffusion problem
February 2009

A565 Ehsan Azmoodeh, Yulia Mishura, Esko Valkeila
On hedging European options in geometric fractional Brownian motion market model
February 2009

ISBN 978-951-22-9923-2 (print)
ISBN 978-951-22-9924-9 (PDF)
ISSN 0784-3143 (print)
ISSN 1797-5867 (PDF)