LA-UR-14-27620
Approved for public release; distribution is unlimited.

Title: A non-linear constrained optimization technique for the mimetic finite difference method

Author(s): Manzini, Gianmarco
Svyatskiy, Daniil
Bertolazzi, Enrico
Frego, Marco

Intended for: Report

Issued: 2014-09-30

A non-linear constrained optimization technique for the mimetic finite difference method

E. Bertolazzi, [a] M. Frego, [a] G. Manzini, [b,c,d] D. Svyatskiy [b,c,d]

[a] *Dipartimento di Ingegneria Industriale, Università di Trento, Via Mesiano 77, I – 38050 Trento, Italy*

[b] *Istituto di Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche (IMATI-CNR), via Ferrata 1, I – 27100 Pavia, Italy,*

[c] *Centro di Simulazione Numerica Avanzata (CeSNA) –IUSS Pavia, v.le Lungo Ticino Sforza 56, I – 27100 Pavia, Italy*

[d] *Los Alamos National Laboratory, Theoretical Division, Group T-5, MS B284, Los Alamos, NM-87545, USA*

Abstract

We propose a strategy for the construction of monotone schemes in the framework of the mimetic finite difference method for the approximation of diffusion problems on unstructured polygonal and polyhedral meshes.

Key words: Mimetic finite difference method, polygonal mesh, polyhedral mesh, Karesh-Kuhn-Tucker optimization, monotone method, maximum principle

1. Introduction

A major property of the solutions of elliptic problems is the existence of Maximum and Minimum Principles [17, 18]. The possibility of reproducing at the discrete level this fundamental property of the exact solutions has been extensively investigated in the literature concerning finite volumes and finite elements for linear and nonlinear parabolic and elliptic partial differential equations [7, 10, 16, 20, 21, 28, 35]. In fact, it turns out that proving the existence of a *Discrete Maximum Principle* (DMP) for the approximate solution is a crucial step towards the development of robust and accurate numerical methods. This property may be crucial, for example, in the numerical modeling of multiphase flow problems in heterogeneous porous media with anisotropic diffusivity [28].

The DMP mimics the maximum principle of the analytical problem in some sense that we need to specify. In fact, since we may find different formulations of the maximum principle in the continuum [19], it is not surprising that there may exist a number of different formulations of the DMP. For example, a numerical method is said to satisfy a maximum principle if the solution p on Ω is less than or equal to p on $\partial\Omega$ when the source term is zero [34]. Another common formulation of a DMP is based on the requirement that the inverse of the stiffness matrix arising from the discretization be a nonnegative matrix, i.e., a matrix with nonnegative coefficients. To formalize this statement, let the system of discrete equations $Ap_h = f_h$ be given by the discretization of the elliptic problem $L(p) = f$ with the additional assumption of a homogeneous Dirichlet condition on the boundary (f_h being a suitable approximation of f and p_h the numerical solution that approximates the exact solution p). If $A^{-1} \geq 0$, i.e. if each element of A^{-1} is nonnegative, then $f_h \geq 0$ implies that $p_h \geq 0$. Since this property is formally similar to the property describing the maximum principle of the continuum case, we say that the discrete formulation satisfies a DMP.

According to [33], we refer to a nonsingular matrix A whose inverse has the sign property defined above as a *monotone* matrix and we say that a numerical method is monotone if it leads to a monotone matrix [28]. Such a monotonicity property holds when matrix A is an M-matrix [1] and building a numerical method that leads to an M-matrix is sufficient to ensure the monotonicity of the approximate solution. This fact explains how M-matrices have been so successful in the spatial finite difference discretizations of second-order elliptic problems [3–6, 12–14], while only a few papers deal with discretizations not yielding an M-matrix [4, 6, 28, 34].

The property of providing an M-matrix is strictly correlated to the mesh and to the diffusion tensor. Therefore, classical finite volume and finite element schemes may fail to satisfy a DMP for strong anisotropic diffusion tensors and/or distorted meshes [15, 28]. In the case of a scalar diffusion tensor, the monotonicity of the numerical scheme may be achieved for a specific range of the diffusion value and under some restrictive conditions on the shape of the parallelograms [15, 28]. A different approach that also leads schemes preserving the maximum principle is based on a nonlinear discretization in finite volumes [2, 25, 29] and in finite elements [22]. The positivity condition can also be recovered from a numerical solution that does not satisfy a maximum principle using a post-processing technique based on a "repair" concept [26]. For higher-order finite element and finite volume methods the analysis of the DMP mainly cope with 1-D problems [30, 34].

In [23], a set of local monotonicity criteria is proposed that can be imposed in the construction of the mimetic inner product to obtain *positivity* and *sparsity* at the elemental level. These conditions hold for quite general (also anisotropic) diffusion tensors and several families of meshes that are widely in use in the scientific community. For instance, parallelograms in 2-D and oblique parallelepipeds in 3-D are used to represent tilted layers, while meshes from Adaptive Mesh Refinement (AMR) techniques [32] are used to increase local accuracy of the numerical solution. Moreover, using such conditions implies that the resulting matrix for the Lagrange multipliers of the mixed-hybrid form is a nonsingular M-matrix, and its inverse matrix has only nonnegative elements. Concerning accuracy and stability, using these criteria does not affect the convergence behavior of the numerical flux and the superconvergence of the pressure variable [8]).

The outline of this paper is as follows. In Section 2 we discuss how the mixed MFD scheme for diffusion problems can be reformulated in the *mixed-hybrid* form by introducing an additional unknown on each mesh edge (in 2-D problems) and on each mesh face (in 3-D problems) to approximate the average of the trace of the scalar variable. An estimate of the approximation error is derived in a suitable mesh-dependent norm. The use of such additional degrees of freedom makes it possible a different algebraic implementation of the family of mixed mimetic schemes, which leads to a well-conditioned linear system for the Lagrangian multiplier unknowns through static condensation. In Section 3, we recall the Hopf's Lemma, which is the fundamental theoretical result on Maximum/Minimum Principles, and study the algebraic monotonicity conditions for mixed-hybrid discretizations. Then, we formulate sufficient monotonicity conditions for the mixed-hybrid MFD method. We study separately the case of simplicial meshes, quadrilateral meshes, hexahedral meshes and orthogonal AMR meshes. In Section 4, we illustrate how these condition perform on a set of representative test cases.

2. The mixed-hybrid mimetic finite difference method

Let Ω be a bounded, simply connected, open subset of $\mathbb{R}^d$ for $d = 2, 3$ with boundary Γ. For simplicity, we assume that Ω is either a polyhedral domain for $d = 3$ or a polygonal domain for $d = 2$. We consider the diffusion of a scalar quantity p in an anisotropic heterogeneous medium filling Ω, which is governed by the second-order elliptic partial differential equation

$$-\mathrm{div}(\mathsf{K}\nabla p) = f \qquad \text{in } \Omega, \tag{1}$$

$$p = g^D \qquad \text{on } \Gamma^D, \tag{2}$$

$$\mathbf{n} \cdot \mathsf{K}\nabla p = g^N \qquad \text{on } \Gamma^N, \tag{3}$$

where K is the diffusion tensor coefficient, f is the volumetric source, g^D is the Dirichlet condition defined on Γ^D, $\mathbf{n}$ is the unit normal vector to Γ pointing out of Ω, and g^N is the Neumann condition defined on

$\Gamma^N \subset \Gamma$. We require that $\Gamma = \bar{\Gamma}^D \cup \bar{\Gamma}^N$. To guarantee the uniqueness of the solution to (1)-(3) and to the discrete problem that will be defined in the next section, we also require a positive measure of Γ^D, i.e. $|\Gamma^D| > 0$. Moreover, we assume that f, g^D and g^N are given functions that belong to $L^2(\Omega)$, $H^{1/2}(\Gamma^D)$ and $H^{-1/2}(\Gamma^N)$, respectively. We also assume that K is a *strongly elliptic* tensor on Ω, i.e., $\mathsf{K}(\mathbf{x})$ is a symmetric and positive definite (SPD) matrix for every point $\mathbf{x} \in \Omega$. We will refer to this regularity assumption on K as "Assumption (K)". From this assumption it follows immediately that $\mathsf{K}(\mathbf{x})$ is a non-singular matrix for every $\mathbf{x} \in \Omega$, and that $\mathsf{K}^{-1}(\mathbf{x})$ is also a symmetric and positive definite matrix for every $\mathbf{x} \in \Omega$.

We reformulate problem (1)-(3) in a mixed form by introducing the vector variable $\mathbf{u}$, the flux of p, through the equation

$$\mathbf{u} + \mathsf{K}\nabla p = 0, \tag{4}$$

and rewriting equation (1) as follows

$$\operatorname{div} \mathbf{u} = f. \tag{5}$$

Under the assumptions introduced so far both differential formulations (1) and (4)-(5) with boundary conditions (2)-(3) are mathematically well-posed [17].

Let Ω_h be a partition of Ω into control volumes P. We assume that Ω_h belongs to a sequence of mesh partitions of Ω that satisfy the shape-regularity condition stated in [8]. Accordingly, the control volumes P are general polyhedrons (polygons in 2-D) that may have non-convex shapes, as, for example, the ones encountered in finite volume methods. Moreover, we assume that Ω_h is conformal, i.e., the intersection of any two distinct cells P' and P'' is either empty, or a few mesh points, or a few mesh edges or a few mesh faces (in 3-D), since the adjacent cells may share more than one edge or face. Furthermore, we assume that Ω_h is *face-connected*, i.e. any two control volumes can be connected by a continuous line going through centers of interior mesh faces. This last assumption is not restrictive since the majority of computational meshes are face-connected.

The degrees of freedom of the scalar variable p are associated with both mesh control volumes P and mesh faces f. In the first case, the degree of freedom is denoted by p_{P} and approximates the average of p over P. In the second case, the degree of freedom is denoted by p_{f} and approximates the average of p over f.

The degrees of freedom of the vector variable $\mathbf{u}$ are associated with the mesh faces and approximate the normal components of $\mathbf{u}$. We denote these degrees of freedom by $u_{\mathsf{P},\mathsf{f}}$, $u_{\mathsf{P},\mathsf{f}} \approx \mathbf{u} \cdot \mathbf{n}_{\mathsf{P},\mathsf{f}}$, where $\mathbf{n}_{\mathsf{P},\mathsf{f}}$ is the normal vector to f pointing out of P. Consequently, any internal face, which is shared by two cells P' and P'', is characterized by two distinct flux unknowns $u_{\mathsf{P}',\mathsf{f}}$ and $u_{\mathsf{P}'',\mathsf{f}}$ that must satisfy the condition of flux conservation:

$$u_{\mathsf{P}',\mathsf{f}} + u_{\mathsf{P}'',\mathsf{f}} = 0 \quad \text{where} \quad \mathsf{f} \subseteq \partial\mathsf{P}' \cap \partial\mathsf{P}''. \tag{6}$$

Let the boundary of cell P be formed by the m faces f_i, $i = 1, \ldots, m$, with measure $|\mathsf{f}_i|$ (length in 2-D, surface area in 3-D). We now consider the numerical discretization of (4) that reads as

$$\begin{pmatrix} u_{\mathsf{P},\mathsf{f}_1} \\ u_{\mathsf{P},\mathsf{f}_2} \\ \vdots \\ u_{\mathsf{P},\mathsf{f}_m} \end{pmatrix} = -\mathsf{W}_{\mathsf{P}} \begin{pmatrix} |\mathsf{f}_1|(p_{\mathsf{f}_1} - p_{\mathsf{P}}) \\ |\mathsf{f}_2|(p_{\mathsf{f}_2} - p_{\mathsf{P}}) \\ \vdots \\ |\mathsf{f}_m|(p_{\mathsf{f}_m} - p_{\mathsf{P}}) \end{pmatrix}, \tag{7}$$

where $\mathsf{W}_{\mathsf{P}} = (w_{ij})$ is a symmetric and positive definite (SPD) with size $m \times m$. The construction of W_{P} will be detailed in the next subsection.

Let $\mathbf{u}_{\mathsf{P}} = (u_{\mathsf{P},\mathsf{f}_1}, u_{\mathsf{P},\mathsf{f}_2}, \ldots, u_{\mathsf{P},\mathsf{f}_m})^T$ be the m-sized vector of numerical fluxes across faces f_i of P. We approximate (5) using vector $\mathbf{u}_{\mathsf{P}}$ as:

$$\operatorname{div}_{\mathsf{P}} \mathbf{u}_{\mathsf{P}} = \overline{f}_{\mathsf{P}}, \tag{8}$$

where $\overline{f}_{\mathsf{P}}$ is the average of f over P, and

$$\operatorname{div}_{\mathsf{P}} \mathbf{u}_{\mathsf{P}} = \frac{1}{|\mathsf{P}|} \sum_{i=1}^{m} |\mathsf{f}_i| u_{\mathsf{P},\mathsf{f}_i}.$$

The Dirichlet boundary condition (2) is taken into account by setting each p_{f} corresponding to a face $\mathsf{f} \in \Gamma^D$ as follows

$$p_{\mathsf{f}} = \overline{g}_{\mathsf{f}}^D := \frac{1}{|\mathsf{f}|} \int_{\mathsf{f}} g^D \, \mathrm{d}S \quad \forall \mathsf{f} \in \partial \mathsf{P} \cap \Gamma^D. \tag{9}$$

The Neumann boundary condition (3) is taken into account by setting the numerical fluxes corresponding to $\mathsf{f} \in \Gamma^N$ as follows

$$u_{\mathsf{f}} = \overline{g}_{\mathsf{f}}^N := \frac{1}{|\mathsf{f}|} \int_{\mathsf{f}} g^N \, \mathrm{d}S \quad \forall \mathsf{f} \in \Gamma^N. \tag{10}$$

The hybrid mimetic scheme is defined by formulas (6), (7), (8) with bounday conditions specified by (9), and (10).

2.1. *Construction of the matrix* W_{P}

In the mixed-hybrid MFD method, matrix W_{P} is built in accordance with a *stability* and a *consistency* conditions [9]. The stability condition states that

$$\frac{\sigma_*}{|\mathsf{P}|} \, \mathbf{u}_{\mathsf{P}}^T \mathbf{u}_{\mathsf{P}} \leq \mathbf{u}_{\mathsf{P}}^T \mathsf{W}_{\mathsf{P}} \mathbf{u}_{\mathsf{P}} \leq \frac{\sigma^*}{|\mathsf{P}|} \, \mathbf{u}_{\mathsf{P}}^T \mathbf{u}_{\mathsf{P}} \tag{11}$$

for any $N^{\mathcal{F}}$-sized flux vector $\mathbf{u}_{\mathsf{P}}$ where σ_* and σ^* are two constant factors independent of P and of the mesh Ω_h. This condition states that matrix W_{P} is spectrally equivalent to the scalar matrix $|\mathsf{P}|^{-1} \, \mathsf{I}_{\mathsf{P}}$, where I_{P} is the $m \times m$-sized identity matrix.

Let $\mathbf{x}_{\mathsf{P}}$ and $\mathbf{x}_{\mathsf{f}}$ be the centers of gravity of cell P and face f, respectively. Furthermore, let $\mathbf{n}_{\mathsf{f}_i}$ be the outward unit normal vector to the i-th face f_i. Now, let us introduce the following two $m \times d$-sized matrices

$$\mathsf{R}_{\mathsf{P}} = \begin{pmatrix} |\mathsf{f}_1|(\mathbf{x}_{\mathsf{f}_1} - \mathbf{x}_{\mathsf{P}})^T \\ |\mathsf{f}_2|(\mathbf{x}_{\mathsf{f}_2} - \mathbf{x}_{\mathsf{P}})^T \\ \vdots \\ |\mathsf{f}_m|(\mathbf{x}_{\mathsf{f}_m} - \mathbf{x}_{\mathsf{P}})^T \end{pmatrix} \quad \text{and} \quad \mathsf{N}_{\mathsf{P}} = \begin{pmatrix} \mathbf{n}_{\mathsf{f}_1}^T \\ \mathbf{n}_{\mathsf{f}_2}^T \\ \vdots \\ \mathbf{n}_{\mathsf{f}_m}^T \end{pmatrix} \mathsf{K}. \tag{12}$$

A straightforward calculation shows that

$$\mathsf{N}_{\mathsf{P}}^T \, \mathsf{R}_{\mathsf{P}} = |\mathsf{P}| \, \mathsf{K},$$

from which we deduce that $\mathsf{N}_{\mathsf{P}}^T \mathsf{R}_{\mathsf{P}}$ is a $d \times d$-sized SPD matrix The consistency condition takes the form

$$\mathsf{W}_{\mathsf{P}} \mathsf{R}_{\mathsf{P}} = \mathsf{N}_{\mathsf{P}},$$

and, in according to this formula, matrix W_{P} is given by

$$\mathsf{W}_{\mathsf{P}} = \mathsf{W}_{\mathsf{P}}^0 + \mathsf{W}_{\mathsf{P}}^1 = \mathsf{N}_{\mathsf{P}} (\mathsf{N}_{\mathsf{P}}^T \mathsf{R}_{\mathsf{P}})^{-1} \mathsf{N}_{\mathsf{P}}^T + \mathsf{D}_{\mathsf{P}} \mathsf{U}_{\mathsf{P}} \mathsf{D}_{\mathsf{P}}^T, \tag{13}$$

where D_{P} is a maximum rank $d \times (m-d)$-sized matrix such that $\mathsf{R}_{\mathsf{P}}^T \mathsf{D}_{\mathsf{P}} = 0$, and U_{P} is a $(m-d) \times (m-d)$-sized matrix of parameters.

Two possible choices for the matrix of parameter are given by taking $\widetilde{\mathsf{U}}_{\mathsf{P}} = \gamma_{\mathsf{P}} \mathsf{I}_{\mathsf{P}}$ or $\widetilde{\mathsf{U}}_{\mathsf{P}} = \gamma_{\mathsf{P}} (\mathsf{D}_{\mathsf{P}}^T \mathsf{D}_{\mathsf{P}})^{-1}$.

Remark 2.1 *On a rectangular mesh there exists a choice of* U_{P} *such that* W_{P} *is diagonal. When* W_{P} *is diagonal, the MFD method reduces to the cell-centered discretization. However, the method remains second-order accurate only when the mesh is* K*-orthogonal. Extension of the mimetic construction to multiple fluxes per mesh edge (face in 3D) allows us to reproduce a family of MPFA O-methods [24]. This connection has been used to analyze convergence and stability of the MPFA methods.On simplicial meshes the MFD method is strongly connected with the lowest order Raviart-Thomas mixed finite element method. For a specific choice of the parameters on triangular and tetrahedral meshes we recover such method [11].*

3. Maximum and minimum principles, monotonicity conditions and the Karush-Kuhn-Tucker optimization strategy

3.1. *Hopf's lemma, strong and weak maximum principles*

Let $L(\cdot)$ be a second-order elliptic operator with regular coefficients and f a square integrable function, both defined in the open domain Ω. We consider the elliptic problem $L(p) = f$ for the solution field p. The maximum principle states that p cannot have any minimum in Ω whenever the source term is nonnegative. More precisely, if $f \geq 0$ and there exists a point $\mathbf{x}_0 \in \Omega$ such that $p(\mathbf{x}_0) \geq p(\mathbf{x})$ for all other $\mathbf{x} \in \Omega$, then the solution field p is constant in Ω. This result, also known as *Hopf's lemma* [18], is proved under the condition that the diffusion tensor is continuously differentiable. Other versions of the maximum principle also exist [31]. Let us recall the fundamental results from the theory of elliptic partial differential equations in accordance with the *strong* and the *weak* formulations.

Theorem 3.1 (Strong Maximum Principle) *Let us suppose that p satisfies*

$$-\mathrm{div}(\mathsf{K}\nabla p) \leq 0 \quad \textit{in } \Omega$$

under assumption (K) on K and for Ω enough regular. If p attains a nonnegative maximum $\widehat{p}$ at an interior point of Ω, then:

$$p = \widehat{p} \quad \textit{in } \overline{\Omega}.$$

Theorem 3.2 (Weak Maximum Principle) *Let us suppose that p satisfies*

$$-\mathrm{div}(\mathsf{K}\nabla p) \leq 0 \quad \textit{in } \Omega$$

under assumption (K) on K and for Ω enough regular. Then,

$$\max_{\mathbf{x}\in\overline{\Omega}} p(\mathbf{x}) \leq \max\Big(0, \max_{\mathbf{x}\in\Gamma} p(\mathbf{x})\Big).$$

From the Weak Maximum Principle it is immediate to derive a *monotonicity property* for the Dirichlet problem. In case of mixed boundary conditions, the monotonicity property is as follows [20].

Corollary 3.1 (Monotonicity Property) *Let us suppose that p satisfies*

$$\begin{aligned} -\mathrm{div}(\mathsf{K}\nabla p) &\geq 0 \quad \textit{in } \Omega, \\ \mathbf{n}\cdot\mathsf{K}\nabla p &\geq 0 \quad \textit{on } \Gamma^N, \\ p &\geq 0 \quad \textit{on } \Gamma^D, \end{aligned}$$

under assumption (K) on K and for Ω enough regular. Then,

$$p \geq 0 \qquad \textit{in } \Omega.$$

3.2. *Algebraic monotonicity conditions*

Now, let us review the derivation of the algebraic conditions that allows us to select the monotone schemes within the family of the mixed MFD schemes. To obtain sufficient criteria for a monotone mimetic-hybrid scheme, we introduce these two assumptions on the pattern of the local matrix $\mathsf{W_P}$.

Assumption 3.3 *Let $\mathsf{W_P} = (w_{ij})_{i,j=1}^m$. Then,*
(A1), matrix $\mathsf{W_P}$ satisfies the geometric constraint:

$$w_{ii}|\mathsf{f}_i| + \sum_{j\neq i} w_{ij}|\mathsf{f}_j| \geq 0 \quad \forall i,$$

and the inequality is strict for at least one matrix row;

(A2), matrix $\mathsf{W_P}$ *is a Z-matrix, i.e.,* $w_{ij} \leq 0$ *for* $i \neq j$.

Using assumptions (A1) and (A2), we recall the fundamental monotonicity result that are proved in [23]. Let $\mathcal{F}$ denote the set of all mesh faces except the Dirichlet boundary faces.

Theorem 3.4 (Discrete Maximum Principle) *Let* $p_h = (p_\mathsf{P})_{\mathsf{P}\in\Omega_h}$ *and* $\lambda_h = (p_\mathsf{f})_{\mathsf{f}\in\mathcal{F}}$ *be the solution of the hybrid mimetic method under assumptions (A1) and (A2) applied to problem* (1)-(3)*. If* f *is a nonnegative function in* Ω *and* g^D *and* g^N *are nonnegative functions on* Γ^D *and* Γ^N*, respectively, then* $p_\mathsf{P} \geq 0$ *for any* $\mathsf{P} \in \Omega_h$.

A *Discrete Minimum Principle* can also be obtained if f is a nonpositive function.

Theorem 3.5 (Discrete Minimum Principle) *Let* $p_h = (p_\mathsf{P})_{\mathsf{P}\in\Omega_h}$ *and* $\lambda_h = (p_\mathsf{f})_{\mathsf{f}\in\mathcal{F}}$ *be the solution of the hybrid mimetic method under assumptions (A1) and (A2) applied to the problem* (1)-(3)*. If* f *is a nonpositive function in* Ω *and* g^D *and* g^N *are nonpositive functions on* Γ^D *and* Γ^N*, respectively, then* $p_\mathsf{P} \leq 0$ *for any* $\mathsf{P} \in \Omega_h$.

Using both Theorems 3.4 and 3.5 allows us to obtain a discrete version of Theorem 3.2.

Theorem 3.6 *Let* $p_h = (p_\mathsf{P})_{\mathsf{P}\in\Omega_h}$ *and* $\lambda_h = (p_\mathsf{f})_{\mathsf{f}\in\mathcal{F}}$ *be the solution of the hybrid mimetic method under assumptions (A1) and (A2) applied to the problem* (1)-(3)*. Furthermore, let* $f = 0$ *and* $\Gamma^N = \emptyset$*. Then, the values of* p_P *for any* $\mathsf{P} \in \Omega_h$ *are bounded by the maximum and minimum components of the Dirichlet conditions set in* (9).

3.3. *Karush-Kuhn-Tucker optimization technique*

The local sufficient conditions of assumptions (A1) and (A2) can be rewritten as a Nonlinear Programming Problem (NLP) for the SPD matrix $\mathsf{U_P}$. To take care of the fact that $\mathsf{U_P}$ is an SPD matrix, we consider its Cholesky decomposition $\mathsf{U_P} = \mathsf{L_P}\mathsf{L_P}^T$, where $\mathsf{L_P}$ is a lower triangular matrix. Matrix $\mathsf{W_P}(\mathsf{L_P})$ defined in (13) becomes

$$\mathsf{W_P}(\mathsf{L_P}) = \mathsf{W_P^0} + (\mathsf{D_P}\mathsf{L_P})(\mathsf{D_P}\mathsf{L_P})^T. \tag{14}$$

We rescale $\mathsf{W_P}(\mathsf{L_P})$ as follows

$$\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P}) = \widetilde{\mathsf{W}}_\mathsf{P}^0 + (\widetilde{\mathsf{D}}_\mathsf{P}\widetilde{\mathsf{L}}_\mathsf{P})(\widetilde{\mathsf{D}}_\mathsf{P}\widetilde{\mathsf{L}}_\mathsf{P})^T, \tag{15}$$

where the scaled matrices are defined as

$$\alpha^2\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P}) = \mathsf{C_P}\mathsf{W_P}(\mathsf{L_P})\mathsf{C_P}, \qquad \alpha^2\widetilde{\mathsf{W}}_\mathsf{P}^0 = \mathsf{C_P}\mathsf{W_P^0}\mathsf{C_P}, \qquad \beta\widetilde{\mathsf{D}}_\mathsf{P} = \mathsf{C_P}\mathsf{D_P}, \qquad (\alpha/\beta)\,\widetilde{\mathsf{L}}_\mathsf{P} = \mathsf{L_P}.$$

through the scaling factors α and β and the diagonal matrix collecting the measures of the cell faces $\mathsf{C_P} = \mathsf{diag}(|\mathsf{f}_1|, |\mathsf{f}_2|, \ldots, |\mathsf{f}_m|)$. We determine the scaling factors α and β by respectively requiring that $\|\widetilde{\mathsf{W}}_\mathsf{P}^0\|_2 = 1$ and $\|\widetilde{\mathsf{D}}_\mathsf{P}\widetilde{\mathsf{D}}_\mathsf{P}^T\|_2 = 1$, where $\|\cdot\|_2$ denote the matrix norm induced by the euclidean norm. We obtain:

$$\alpha = \sqrt{\|\mathsf{C_P}\mathsf{W_P^0}\mathsf{C_P}\|_2}, \qquad \beta = \sqrt{\|\mathsf{C_P}\mathsf{D_P}\mathsf{D_P}^T\mathsf{C_P}\|_2}.$$

According to Assumptions (A1)-(A2), we have to maximize the diagonal dominance of the scaled matrix $\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P})$ with the constraint that the off-diagonal elements of this matrix are zero or negative. This is equivalent to maximize a quantity F by changing the matrix of parameter $\mathsf{L_P}$ so that the following inequalities are satisfied:

$$\widetilde{w}_{ii}(\widetilde{\mathsf{L}}_\mathsf{P}) + \sum_{j\neq i} \widetilde{w}_{ij}(\widetilde{\mathsf{L}}_\mathsf{P}) \geq F \qquad \forall i, \tag{16a}$$

$$\widetilde{w}_{ij}(\widetilde{\mathsf{L}}_\mathsf{P}) \leq 0 \qquad \forall i \neq j. \tag{16b}$$

If such a constrained optimization problem admits a solution $\mathsf{L_P}$ that is associated with a non-negative quantity F, then matrix $\mathsf{W_P}$ satisfies both conditions (A1) and (A2), and discrete Maximum and Minimum principles for the mimetic scheme can be proved. However, a solution to this problem does not exist in

general because the sign condition in (16b) on the off-diagonal elements in $\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P})$ can be too restrictive. We tackle this issue by substituting (16b) with the weaker condition

$$\widetilde{w}_{ij}(\widetilde{\mathsf{L}}_\mathsf{P}) \leq \max\{0, E\} \qquad \forall i \neq j. \tag{17}$$

Now, conditions (16a)-(17) define a non-empty solution set for some pair (E, F). If we can determine a parameter matrix L_P such that $E \leq 0$ and $F \geq 0$, assumptions (A1)-(A2) are satisfied.

To determine the best possible choice of L_P, we apply a *divide-and-conquer* strategy implemented in a two-step algorithm. In the first step, we solve an NLP problem to determine a parameter matrix $\widetilde{\mathsf{L}}_\mathsf{P}$ such that, if possible, $E \leq 0$ and $\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P})$ is a Z-matrix. Matrix $\widetilde{\mathsf{L}}_\mathsf{P}$ can be optionally required to be close to a matrix $\widehat{\mathsf{L}}_\mathsf{P}$ such that $\widehat{\mathsf{L}}_\mathsf{P}\widehat{\mathsf{L}}_\mathsf{P}^T = \gamma_\mathsf{P}\mathsf{I}_\mathsf{P}$ or $\widehat{\mathsf{L}}_\mathsf{P}\widehat{\mathsf{L}}_\mathsf{P}^T = \gamma_\mathsf{P}(\widetilde{\mathsf{D}}_\mathsf{P}^T\widetilde{\mathsf{D}}_\mathsf{P})^{-1}$. In practice, matrix $\widehat{\mathsf{L}}_\mathsf{P}$ is returned by the Cholesky decomposition of one of the (scaled) matrices $\widetilde{\mathsf{U}}_\mathsf{P}$ listed at the end of subsection 2.1. This NLP problem always admits a solution $\widetilde{\mathsf{L}}_\mathsf{P}$ but the resulting matrix $\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P})$ is not guaranteed to be a Z-matrix. However, if the first step is successful, i.e., $E \leq 0$, and $\widetilde{\mathsf{W}}_\mathsf{P}(\widetilde{\mathsf{L}}_\mathsf{P})$ is a Z-matrix, we perform a second step by solving an NLP problem that maximizes the diagonal dominance of $\mathsf{W}_\mathsf{P}(\mathsf{L}_\mathsf{P})$.

The first NLP problem reads as:

$$\text{minimize the objective function:} \qquad f(E, \widetilde{\mathsf{L}}_\mathsf{P}) = E + \omega_L \sum_{i \leq j} (\widetilde{l}_{ij} - \widehat{l}_{ij})^2 \tag{18a}$$

$$\text{subject to the unilateral constraints:} \qquad f_{pq}(E, \widetilde{\mathsf{L}}_\mathsf{P}) = E - \widetilde{w}_{pq}(\widetilde{\mathsf{L}}_\mathsf{P}) \geq 0, \qquad \forall p < q \tag{18b}$$

where ω_L is a tuning factor that controls how $\widetilde{\mathsf{L}}_\mathsf{P}$ differs from $\widehat{\mathsf{L}}_\mathsf{P}$. The quantity E controls the off-diagonal elements of matrix W_P and if the solution of problem (18) satisfies $E \leq 0$, then condition (A2) is satisfied. On the other hand, if E is strictly positive, the optimization algorithm cannot produce a Z-matrix.

The second NLP problem read as:

$$\text{minimize the objective function:} \qquad g(F, \widetilde{\mathsf{L}}_\mathsf{P}) = -F + \omega_L \sum_{i \leq j} (\widetilde{l}_{ij} - \hat{l}_{ij})^2 \tag{19a}$$

$$\text{subject to unilateral constraints:} \qquad g_{pq}(F, \widetilde{\mathsf{L}}_\mathsf{P}) = E^+ - \widetilde{w}_{pq}(\widetilde{\mathsf{L}}_\mathsf{P}) \geq 0, \qquad \forall p < q \tag{19b}$$

$$\text{and:} \qquad g_{pp}(F, \widetilde{\mathsf{L}}_\mathsf{P}) = \widetilde{w}_{pp}(\widetilde{\mathsf{L}}_\mathsf{P}) + \sum_{j \neq p} \widetilde{w}_{pj}(\widetilde{\mathsf{L}}_\mathsf{P}) - F \geq 0, \qquad \forall p, \tag{19c}$$

where ω_L is the same tuning factor of step 1 that controls how $\widetilde{\mathsf{L}}_\mathsf{P}$ differs from $\widehat{\mathsf{L}}_\mathsf{P}$. The quantity F controls the property of W_P of being a diagonally dominant matrix.

Thus, the first step is used only for the computation of $E^+ = \max\{0, E\}$, which defines a feasible set of constraints for the second step. If $E^+ = 0$ and the solution of NLP (19) satisfies $F > 0$, then solution respects assumptions (A1) and (A2).

The solutions of NLPs (18) and (19) are characterized by the Karush-Kuhn-Tucker (KKT) first order necessary conditions (see [27] chapter 12). The first order KKT necessary condition uses the two Lagrangian functionals defined by:

$$\mathcal{L}(E, \widetilde{\mathsf{L}}_\mathsf{P}, \boldsymbol{\mu}) = f(E, \widetilde{\mathsf{L}}_\mathsf{P}) - \sum_{p<q} \mu_{pq} f_{pq}(E, \mathsf{L}_\mathsf{P}), \qquad \text{for NLP (18)}$$

$$\mathcal{L}(F, \widetilde{\mathsf{L}}_\mathsf{P}, \boldsymbol{\mu}) = g(F, \widetilde{\mathsf{L}}_\mathsf{P}) - \sum_{p \leq q} \mu_{pq} g_{pq}(F, \mathsf{L}_\mathsf{P}), \qquad \text{for NLP (19)}$$

where we introduced the Lagrangian multipliers μ_{pq} to take into account the constraints f_{pq} and g_{pq}. The first order KKT condition becomes

$$\text{for NLP (18):} \begin{cases} \dfrac{\partial \mathcal{L}}{\partial E} = \dfrac{\partial \mathcal{L}}{\partial l_{ij}} = 0 \\ \mu_{pq} f_{pq}(E, \widetilde{\mathsf{L_P}}) = 0 \\ f_{pq}(E, \widetilde{\mathsf{L_P}}) \geq 0 \\ \mu_{pq} \geq 0 \end{cases} \qquad \text{for NLP (19):} \begin{cases} \dfrac{\partial \mathcal{L}}{\partial F} = \dfrac{\partial \mathcal{L}}{\partial l_{ij}} = 0 \\ \mu_{pq} g_{pq}(F, \widetilde{\mathsf{L_P}}) = 0 \\ g_{pq}(F, \widetilde{\mathsf{L_P}}) \geq 0 \\ \mu_{pq} \geq 0 \end{cases} \tag{20}$$

3.4. *Implementation of the two-step algorithm*

Our implementation of the two-step algorithm is based the open source software IPOPT [36]. IPOPT requires the gradient and Hessian of both the objective function and the unilateral constraints. The gradient of the objective function $f(E, \widetilde{\mathsf{L_P}})$ and of the constraints $f_{pq}(E, \widetilde{\mathsf{L_P}})$ has the following components:

$$\frac{\partial f}{\partial E} = 1, \qquad \frac{\partial f}{\partial \widetilde{l}_{ij}} = 2\omega_L(\widetilde{l}_{ij} - \widetilde{l}^0_{ij}), \qquad \frac{\partial f_{pq}}{\partial E} = 1, \qquad \frac{\partial f_{pq}}{\partial \widetilde{l}_{ij}} = -\frac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}} \tag{21}$$

The gradient of the objective function $g(F, \widetilde{\mathsf{L_P}})$ and of the constraints $g_{pq}(F, \widetilde{\mathsf{L_P}})$ has the following components:

$$\frac{\partial g}{\partial F} = -1, \qquad \frac{\partial g}{\partial \widetilde{l}_{ij}} = 2\omega_L(\widetilde{l}_{ij} - \widetilde{l}^0_{ij}), \qquad \frac{\partial g_{pq}}{\partial F} = -1, \qquad \frac{\partial g_{pq}}{\partial \widetilde{l}_{ij}} = \begin{cases} -\dfrac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}} & p < q, \\ \displaystyle\sum_k \frac{\partial \widetilde{w}_{pk}}{\partial \widetilde{l}_{ij}} & p = q. \end{cases} \tag{22}$$

The Hessian matrices of the targets $f(E, \widetilde{\mathsf{L_P}})$ and $g(F, \widetilde{\mathsf{L_P}})$ satisfy

$$\nabla^2 f(E, \widetilde{\mathsf{L_P}}) = \begin{pmatrix} 0 & 0 \\ 0 & 2\omega_L \mathsf{I_P} \end{pmatrix}, \qquad \nabla^2 g(F, \widetilde{\mathsf{L_P}}) = \begin{pmatrix} 0 & 0 \\ 0 & 2\omega_L \mathsf{I_P} \end{pmatrix},$$

The Hessian matrices of the constraints $f_{pq}(E, \widetilde{\mathsf{L_P}})$ and $g_{pq}(F, \widetilde{\mathsf{L_P}})$ satisfy

$$\nabla^2 f_{pq}(E, \widetilde{\mathsf{L_P}}) = \begin{pmatrix} 0 & 0 \\ 0 & G_{pq} \end{pmatrix}, \qquad \nabla^2 g_{pq}(F, \widetilde{\mathsf{L_P}}) = \begin{pmatrix} 0 & 0 \\ 0 & G_{pq} \end{pmatrix},$$

where

$$G_{pq} = \begin{cases} -\left(\dfrac{\partial^2 \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij} \partial \widetilde{l}_{i'j'}} \right) & p < q, \\ \left(\displaystyle\sum_k \frac{\partial^2 \widetilde{w}_{pk}}{\partial \widetilde{l}_{ij} \partial \widetilde{l}_{i'j'}} \right) & p = q. \end{cases}$$

The partial derivative of $\widetilde{w}_{pq}$ can be computed efficiently by the formula provided by the following lemma.

Lemma 3.1 *Let* $\widetilde{\mathsf{D_P}} = (\widetilde{d}_{ij})$ *then*

$$\frac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}} = \sum_{k \geq j} \widetilde{l}_{kj} (\widetilde{d}_{qk} \widetilde{d}_{pi} + \widetilde{d}_{pk} \widetilde{d}_{qi}). \tag{23}$$

$$\frac{\partial^2 \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij} \partial \widetilde{l}_{i'j'}} = \begin{cases} \widetilde{d}_{qi'} \widetilde{d}_{pi} + \widetilde{d}_{pi'} \widetilde{d}_{qi} & j' = j \text{ and } i' \geq j; \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Proof 3.1 *Let* $\mathbf{e}_p$ *be the* p*-th vector of the canonical basis of* $\mathbb{R}^p$*, i.e., the vector that has* 1 *in the* p*-th position and zero elsewhere. Since* $\widetilde{w}_{pq}(\widetilde{\mathsf{L_P}}) = \widetilde{w}^0_{pq} + ((\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})(\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})^T)_{pq}$ *and* $\widetilde{w}^0_{pq}$ *is constant, we have:*

$$\frac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}}(\widetilde{\mathsf{L_P}}) = \frac{\partial}{\partial \widetilde{l}_{ij}} \Big((\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})(\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})^T \Big)_{pq} = \frac{\partial}{\partial \widetilde{l}_{ij}} \Big(\mathbf{e}_p^T (\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})(\widetilde{\mathsf{D_P}}\widetilde{\mathsf{L_P}})^T \mathbf{e}_q \Big)_{pq}.$$

Using the chain rule, the derivative becomes:

$$\frac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}}(\widetilde{\mathsf{L}}_\mathsf{P}) = \mathbf{e}_p^T \left(\widetilde{\mathsf{D}}_\mathsf{P} \frac{\partial \widetilde{\mathsf{L}}_\mathsf{P}}{\partial \widetilde{l}_{ij}} \right) (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})^T \mathbf{e}_q + \mathbf{e}_p^T \left(\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P} \right) \left(\widetilde{\mathsf{D}}_\mathsf{P} \frac{\partial \widetilde{\mathsf{L}}_\mathsf{P}}{\partial \widetilde{l}_{ij}} \right)^T \mathbf{e}_q,$$

and since $\partial \widetilde{\mathsf{L}}_\mathsf{P} / \partial \widetilde{l}_{ij} = \mathbf{e}_i \mathbf{e}_j^T$ *a straightforward calculation gives:*

$$\begin{aligned}
\frac{\partial \widetilde{w}_{pq}}{\partial \widetilde{l}_{ij}}(\widetilde{\mathsf{L}}_\mathsf{P}) &= \mathbf{e}_p^T (\widetilde{\mathsf{D}}_\mathsf{P} \mathbf{e}_i \mathbf{e}_j^T)(\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})^T \mathbf{e}_q + \mathbf{e}_p^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})(\widetilde{\mathsf{D}}_\mathsf{P} \mathbf{e}_i \mathbf{e}_j^T)^T \mathbf{e}_q \\
&= (\mathbf{e}_p^T \widetilde{\mathsf{D}}_\mathsf{P} \mathbf{e}_i) \mathbf{e}_j^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})^T \mathbf{e}_q + \mathbf{e}_p^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P}) \mathbf{e}_j (\mathbf{e}_i^T \widetilde{\mathsf{D}}_\mathsf{P}^T \mathbf{e}_q) \\
&= \widetilde{d}_{pi} \mathbf{e}_j^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})^T \mathbf{e}_q + \mathbf{e}_p^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P}) \mathbf{e}_j \widetilde{d}_{qi} \\
&= \widetilde{d}_{pi} \mathbf{e}_q^T \widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P} \mathbf{e}_j + \mathbf{e}_p^T (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P}) \mathbf{e}_j \widetilde{d}_{qi} \\
&= \widetilde{d}_{pi} (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})_{qj} + (\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})_{pj} \widetilde{d}_{qi}.
\end{aligned}$$

Equation (23) *follows by noticing that*

$$(\widetilde{\mathsf{D}}_\mathsf{P} \widetilde{\mathsf{L}}_\mathsf{P})_{sj} = \sum_{k \geq j} \widetilde{d}_{sk} \widetilde{l}_{kj} \quad \textit{for } s = p, q$$

since $\widetilde{\mathsf{L}}_\mathsf{P}$ *is a lower triangular matrix. Equation* (24) *follows immediately.*

An initial guess for NLP (18) is any $\widetilde{\mathsf{L}}_\mathsf{P}$, for example $\widetilde{\mathsf{L}}_\mathsf{P} = \widehat{\mathsf{L}}_\mathsf{P}$ while a guess for E is $E = \max\{\widetilde{w}_{ij}(\widetilde{\mathsf{L}}_\mathsf{P})\}$. An initial guess for NLP (19) is the computed $\widetilde{\mathsf{L}}_\mathsf{P}$ from NLP (18) with F

$$F = \min_p \left(\widetilde{w}_{pp}(\widetilde{\mathsf{L}}_\mathsf{P}) + \sum_{j \neq p} \widetilde{w}_{pj}(\widetilde{\mathsf{L}}_\mathsf{P}) \right).$$

Remark 3.7 *When* $E \leq 0$ *matrix* $\mathsf{W}_\mathsf{P}(\mathsf{L}_\mathsf{P})$ *is a Z-matrix, thus minimization of* E *should be done only for* $E > 0$*. Hence, the optimization problem can be written as*

$$\textit{minimize:} \qquad -\omega_F F + E^+, \qquad \textit{subject to } (16a) \textit{ and } (17) \tag{25}$$

where ω_F *is a weight for the diagonal dominance and* $E^+ = \max\{0, E\}$ *is the positive part of* E*. This minimization is not smooth and does not guarantee that* $\widetilde{\mathsf{L}}_\mathsf{P}$ *be nonsingular.*

A smoothed approximation of NLP (25) *uses the following regularized functions:*

$$f(E, F, \widetilde{\mathsf{L}}_\mathsf{P}) = E_\varepsilon^+ - \omega_F F + \omega_L \sum_{i \leq j} (\widetilde{l}_{ij} - \widetilde{l}_{ij}^0)^2, \tag{26}$$

$$g_{pq}(E, F, \widetilde{\mathsf{L}}_\mathsf{P}) = \begin{cases} \widetilde{w}_{pp}(\widetilde{\mathsf{L}}_\mathsf{P}) + \sum_{j \neq p} \widetilde{w}_{pj}(\widetilde{\mathsf{L}}_\mathsf{P}) - F & p = q \\ E - \widetilde{w}_{pq}(\widetilde{\mathsf{L}}_\mathsf{P}) & p \neq q \end{cases} \tag{27}$$

where E_ε^+ *is a smoothed positive part of* E *defined as*

$$E_\varepsilon^+ = \frac{E - \varepsilon + \sqrt{\varepsilon^2 + E^2}}{2}. \tag{28}$$

Moreover, ω_L *weights a convex term which avoids that* $\widetilde{\mathsf{L}}_\mathsf{P}$ *becomes too large and speeds up convergence. Matrix* L_P^0 *is the scaled version of* $\mathsf{L}_\mathsf{P} \mathsf{L}_\mathsf{P}^T = \gamma_\mathsf{P} \mathsf{I}_\mathsf{P}$ *or* $\mathsf{L}_\mathsf{P} \mathsf{L}_\mathsf{P}^T = \gamma_\mathsf{P} (\mathsf{D}_\mathsf{P}^T \mathsf{D}_\mathsf{P})^{-1}$*. Finally,* $\varepsilon > 0$*,* $\omega_F > 0$ *and* $\omega_L \geq 0$ *are tuning parameters. The smoothed NLP problem is*

$$\textit{minimize the objective function:} \qquad f(E, F, \widetilde{\mathsf{L}}_\mathsf{P}) \tag{29a}$$

$$\textit{subject to unilateral constraints:} \qquad g_{pq}(E, F, \widetilde{\mathsf{L}}_\mathsf{P}) \geq 0, \qquad \forall p \leq q \tag{29b}$$

4. Numerical experiments

4.1. *Effects related to the mesh and permeability*

In this subsection, we investigate how mesh distortion impacts on the property of the stiffness matrix of being an M-matrix. In particular, we consider three different test cases that are representative of realistic meshes.

Mesh tilting We review how the optimization algorithm discussed in this paper works on the sequence of tilted meshes considered in [23]. The theoretical analysis of the M-matrix properties of the stiffness matrices on these meshes has been presented in [23]. According to [23], the monotonicity property is satisfied for angles $\theta \geq \overline{\theta}$, where the value of this latter depends on the aspect ratio and the eigenvalues of the diffusion tensor. For K equal to the identity matrix the critical angle θ for some aspect ratios satisfying the monotonicity condition (A1)-(A2) are reported in Table 1. Fig. 1 shows the value of the parameter E on the mesh cells for $\theta = 60°$ and aspect ratio equal to 2 (left plot) and $\theta = 84.5$ and aspect ratio equal to 10 (right plot). For such aspect ratios, these are the critical values of θ for which the monotonicity condition (A1) is first violated. In both cases, as shown in Fig. 2 (top panels), the parameters E and F are, respectively, negative and positive, thus ensuring the monotonicity of the discretization. The bottom panels of Fig. 2 also shows the dependence of the condition number of the global stiffness matrix and the number of negative values of the formal inverse (computed using the routine `dgetri` of LaPack). The behavior shown by these curves is in perfect agreement with the theoretical results of [23]. In other words, the proposed optimization algorithm applied on this type of meshes is always able to find the stiffness matrix with M-matrix properties when such solution exists. The four curves reported in the plots of this figure are for aspect ratios equal to 1, 2, 5, and 10. The condition number versus θ is clearly deteriorating and its value increases significantly when the θ becomes bigger that $\overline{\theta}$. The number of negative values of the inverse of the stiffness matrix, which must be zero if the stiffness matrix is an M-matrix, also increases significantly for $\theta > \overline{\theta}$.

Anisotropic permeability

We solve the Poisson problem with the 2D anisotropic permeability:

$$\mathsf{K} = \begin{pmatrix} \epsilon\, x^2 + y^2 + \epsilon & -(1-\epsilon)xy \\ -(1-\epsilon)xy & x^2 + \epsilon\, y^2 + \epsilon \end{pmatrix}. \tag{30}$$

The loading term is given by

$$f(x,y) = \begin{cases} 1 & \text{if } (x,y) \in [3/8, 5/8], \\ 0 & \text{otherwise,} \end{cases} \tag{31}$$

and we impose boundary conditions everywhere. The exact solution $u(x,y)$ is unknown but the minimum principle states that it is non-negative. Numerical oscillations were seen in the solution computed by MPFA.

Fig. 3 shows the numerical solution for different values of the parameter ϵ. Figs 4 shows the maximum value of E (top-right plot), the minimum value of F (bottom-left plot), the condition number (bottom-left plot), and the number of negative entries ($bottom - rightplot$) versus ϵ.

Mesh deformation Here, we investigate how this algorithm performs on a sequence of progressively distorted meshes. Given a regular partition of the domain $\Omega =]0,1[\times]0,1[$, the position of the internal nodes with coordinates (ξ_{ij}, η_{ij}) is remapped according to

$$x_{ij} = \xi_{ij} + w\big(1 - e^{\sin(\pi\xi_{ij})\sin(\pi\eta_{ij})}\big),$$

$$y_{ij} = \eta_{ij} + w\big(1 - e^{\sin(\pi\xi_{ij})\sin(\pi\eta_{ij})}\big),$$

where w is the distortion parameter. Fig.5 shows the cell values of the parameters E and F for $w = 0.11$, when the monotonicity (A1) is first violated. Fig. 6 characterizes the behavior of the method for values of w ranging from 0 to 0.16. In particular, we show the maximum value of E (top-right plot), the minimum value of F (bottom-left plot), the condition number (bottom-left plot), and the number of negative entries

($bottom - rightplot$). The same calculation is repeated for a mesh of hexagons. The results are shown in Figs. 7 and 7.

Mesh pinching Here, we apply the algorithm to a sequence of meshes that are pinched along the vertical direction on one side. In this test case, the node on the top-right corner is collapsed along the vertical direction towards the bottom-right corner. Fig. 9 shows the cell values of the parameters E and F on the most distorted mesh such that the monotonicity condition is first violated. Fig. 10 show the maximum value of E (top-right plot), the minimum value of F (bottom-left plot), the condition number (bottom-left plot), and the number of negative entries ($bottom - rightplot$) versus the parameter that controls the position of the vertically collapsing node.

5. Conclusions

In this work, we proposed a new strategy for the construction of monotone schemes in the framework of the mimetic finite difference method. These schemes are suitable to the approximation of diffusion problems on unstructured polygonal and polyhedral meshes. The monotonicity of the resulting mimetic method is obtained by the construction of the local stabilization term of the stiffness matrix that satisfies two sufficient conditions in order to have an M-matrix. The presented numerical tests show that whenever the local mimetic family of stiffness matrices contains at least one an M-matrix as a member, a non-linear algorithm based on KKT optimization allows us to determine this matrix. In other situation the imposed requirement are too restrictive for the local mimetic family.

A set of numerical experiments illustrates the performance of the method.

A.R.	1	2	5	10
θ_{min}	$--$	60	78.4	84.3

Table 1
Tilted quadrilateral cells; K is the identity matrix; minimum angle $\overline{\theta}$ for some aspect ratios satisfying the monotonicity condition (A1)-(A2)

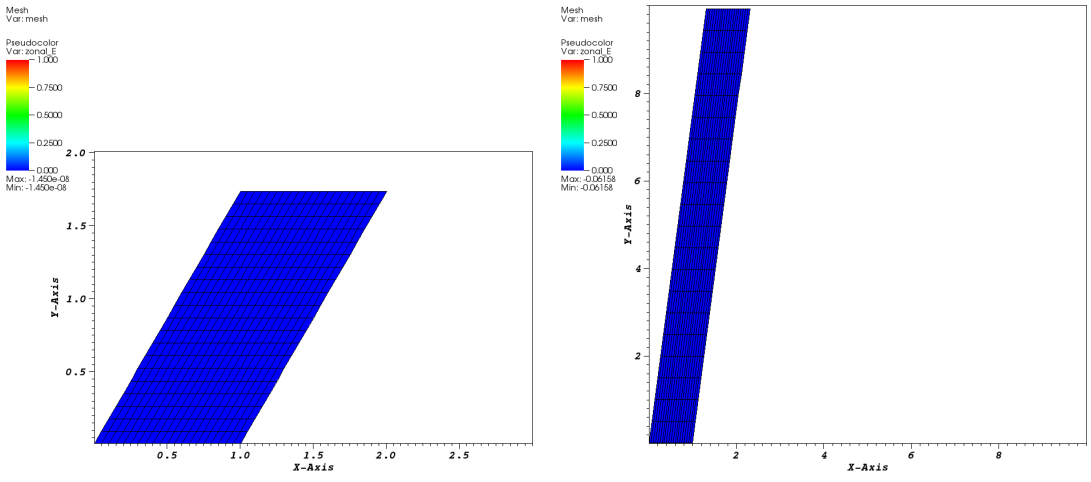

Fig. 1. Tilted quadrilateral cells; map of F when the monotonicity condition (A1) is first violated, i.e., for (left panel) $\overline{\theta} = 60$ when aspect ratio is 2 and (right panel) $\overline{\theta} = 84.3$ when aspect ratio is 10. In both cases, E is strictly negative.

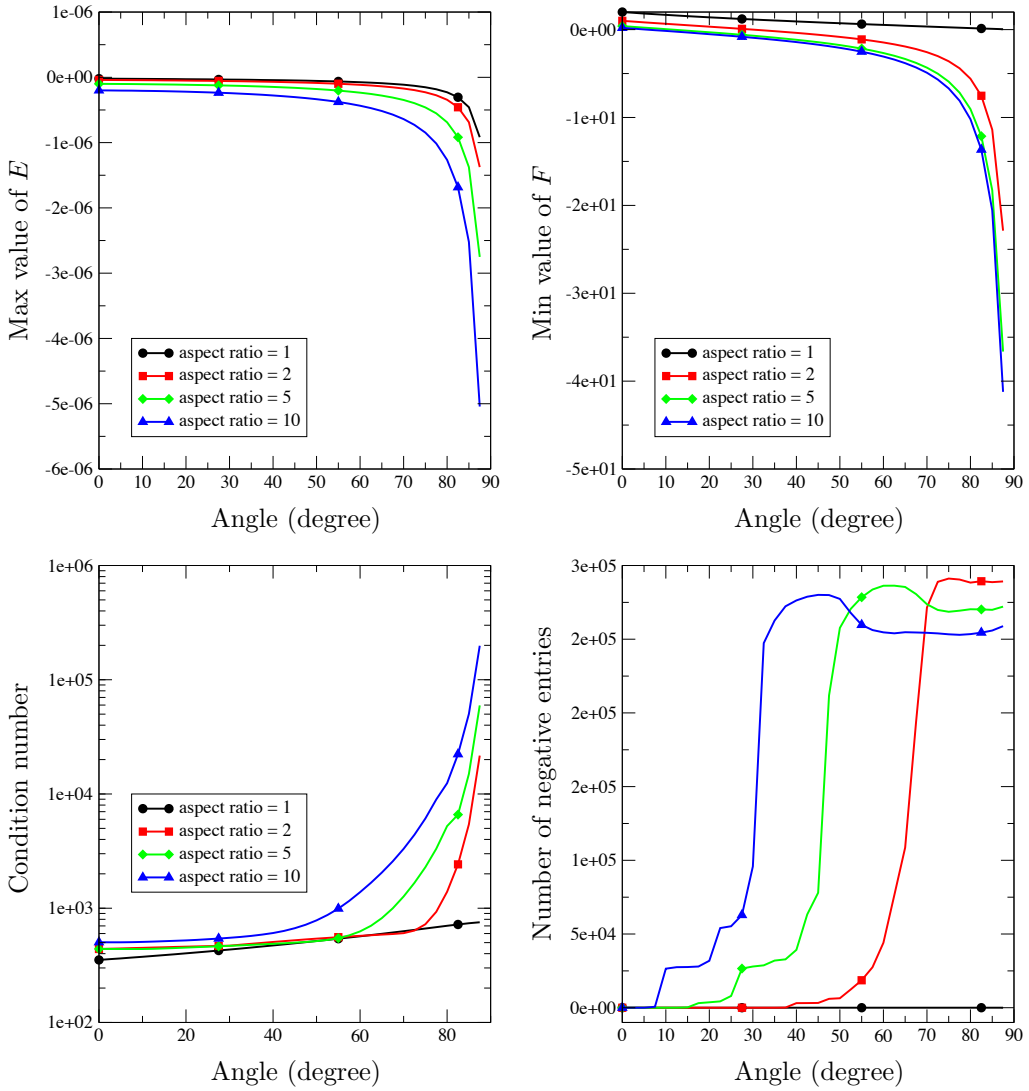

Fig. 2. Tilted quadrilateral cells; K is the identity matrix; the four plots corresponds to the aspect ratios 1, 2, 5 and 10. These curves shows the behavior of the parameters E (top-left plot) and F (top-right plot) as well as the condition number (bottom-left) and the number of negative entries of the formal inverse of the stiffness matrix with respect to the value of the titling angle θ. The critical values of θ that were found theoretically in [23] are perfectly reflected by these experimental results.

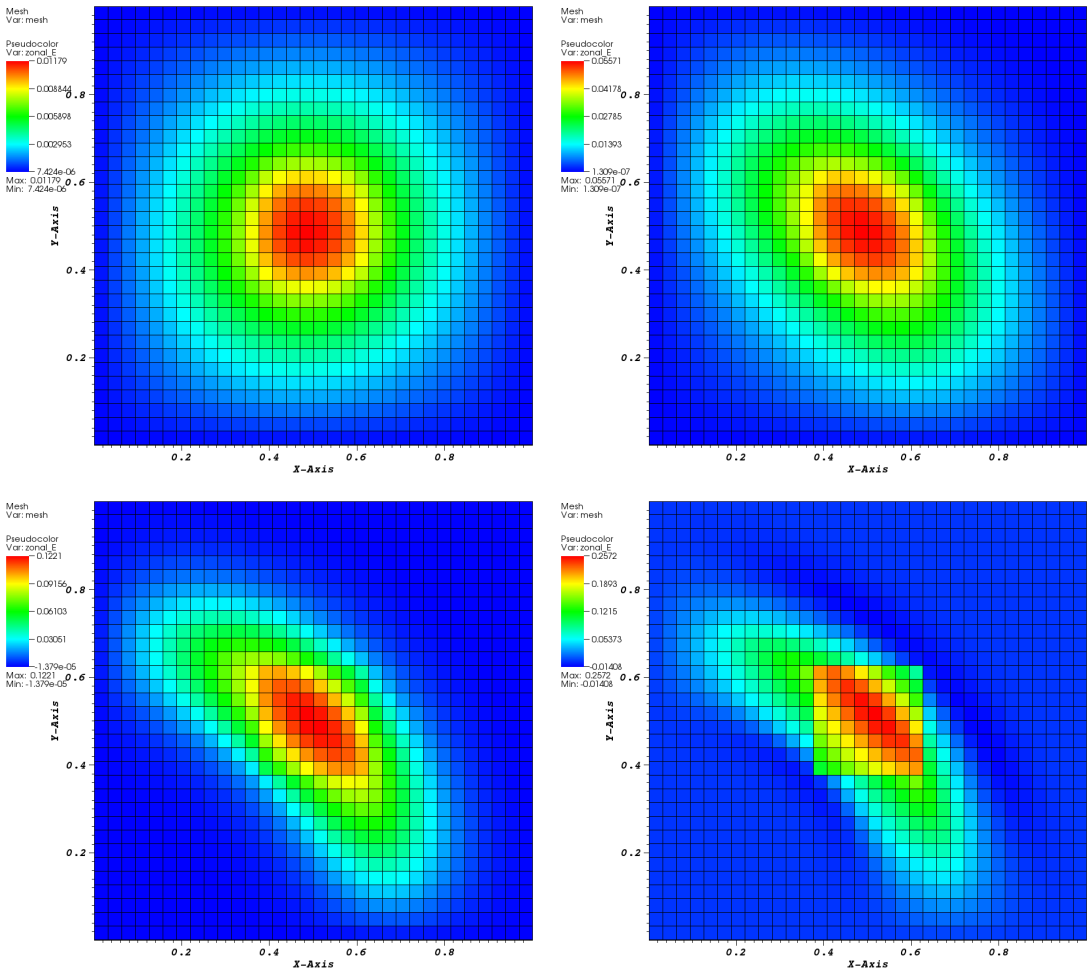

Fig. 3. Anisotropic permeability on regular quadrilateral cells; K is given by (30) numerical solution for $\epsilon = 1$ (top-left), $\epsilon = 10^{-1}$ (top-right), $\epsilon = 10^{-2}$ (bottom-left), $\epsilon = 10^{-3}$ (bottom-right).

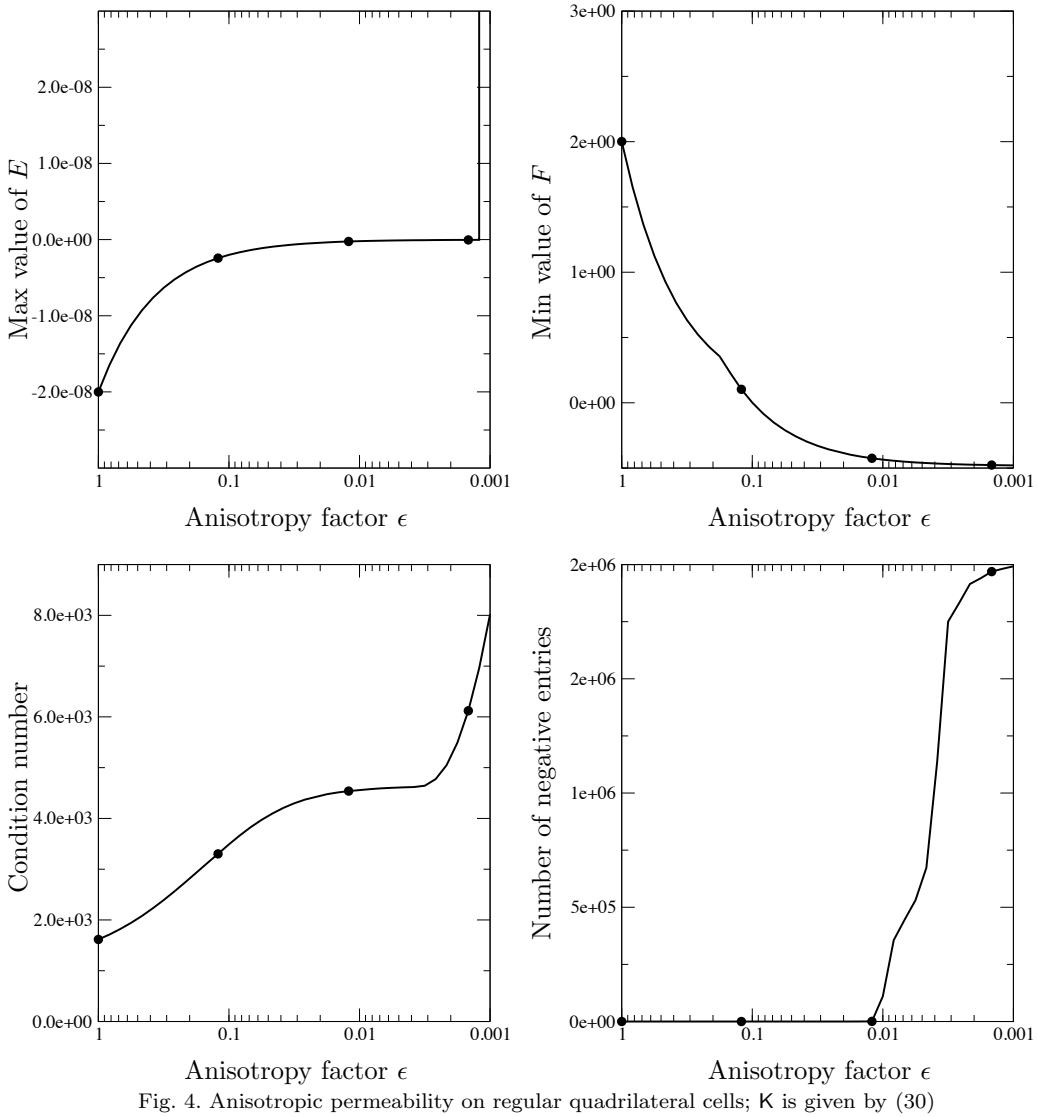

Fig. 4. Anisotropic permeability on regular quadrilateral cells; K is given by (30)

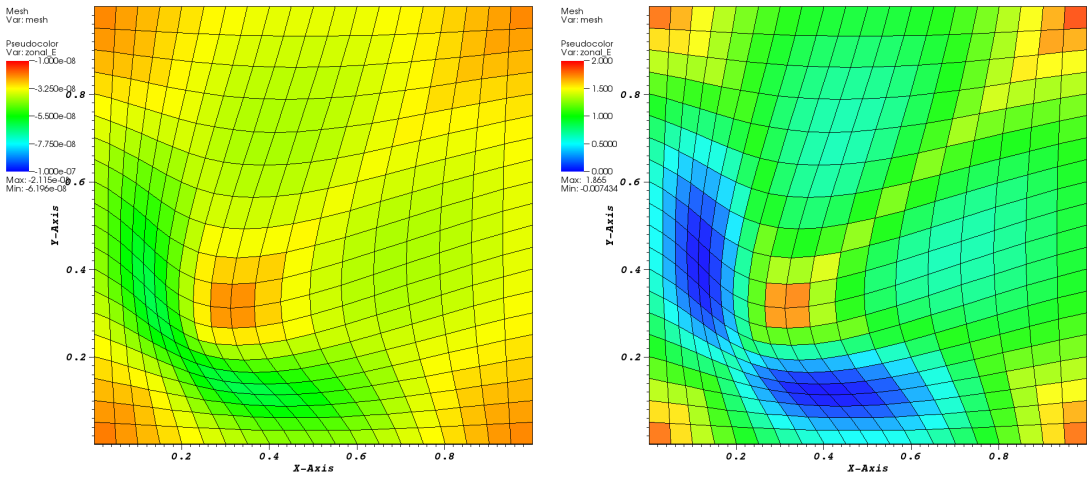

Fig. 5. Distorted quadrilateral cells; map of E (left) and F (right) when the monotonicity condition (A1) is first violated

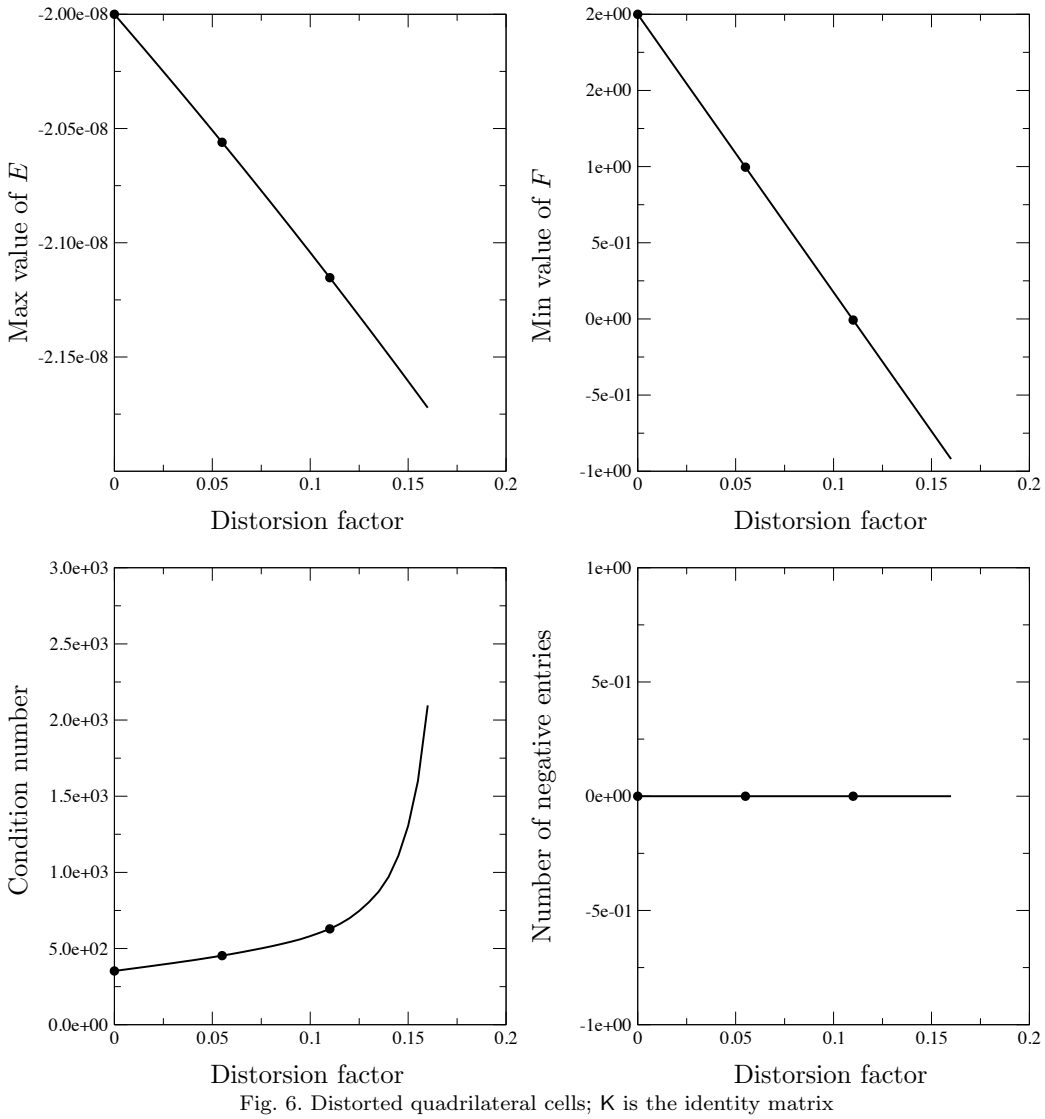

Fig. 6. Distorted quadrilateral cells; K is the identity matrix

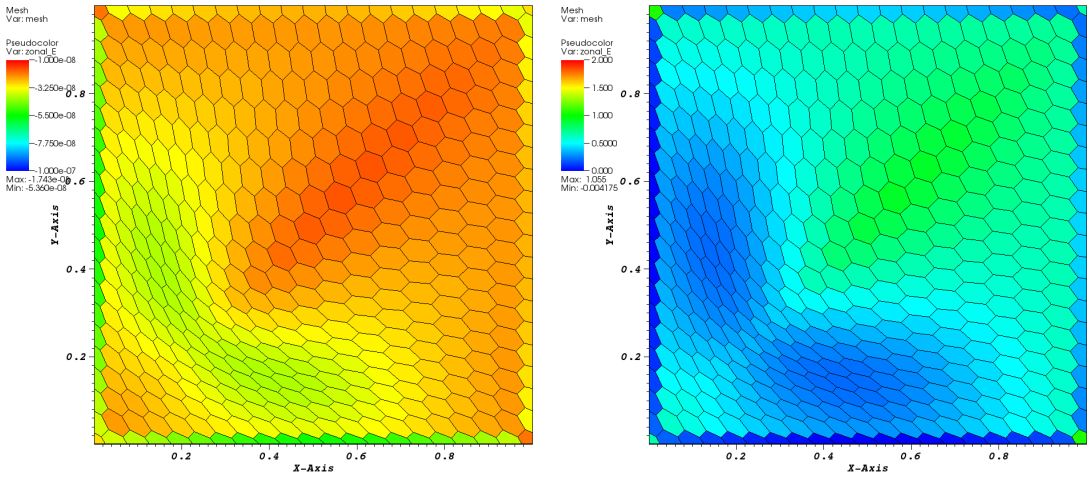

Fig. 7. Distorted hexagonal cells; map of E (left) and F (right) when the monotonicity condition (A1) is first violated. Note that the violation occurs only for the F quantity.

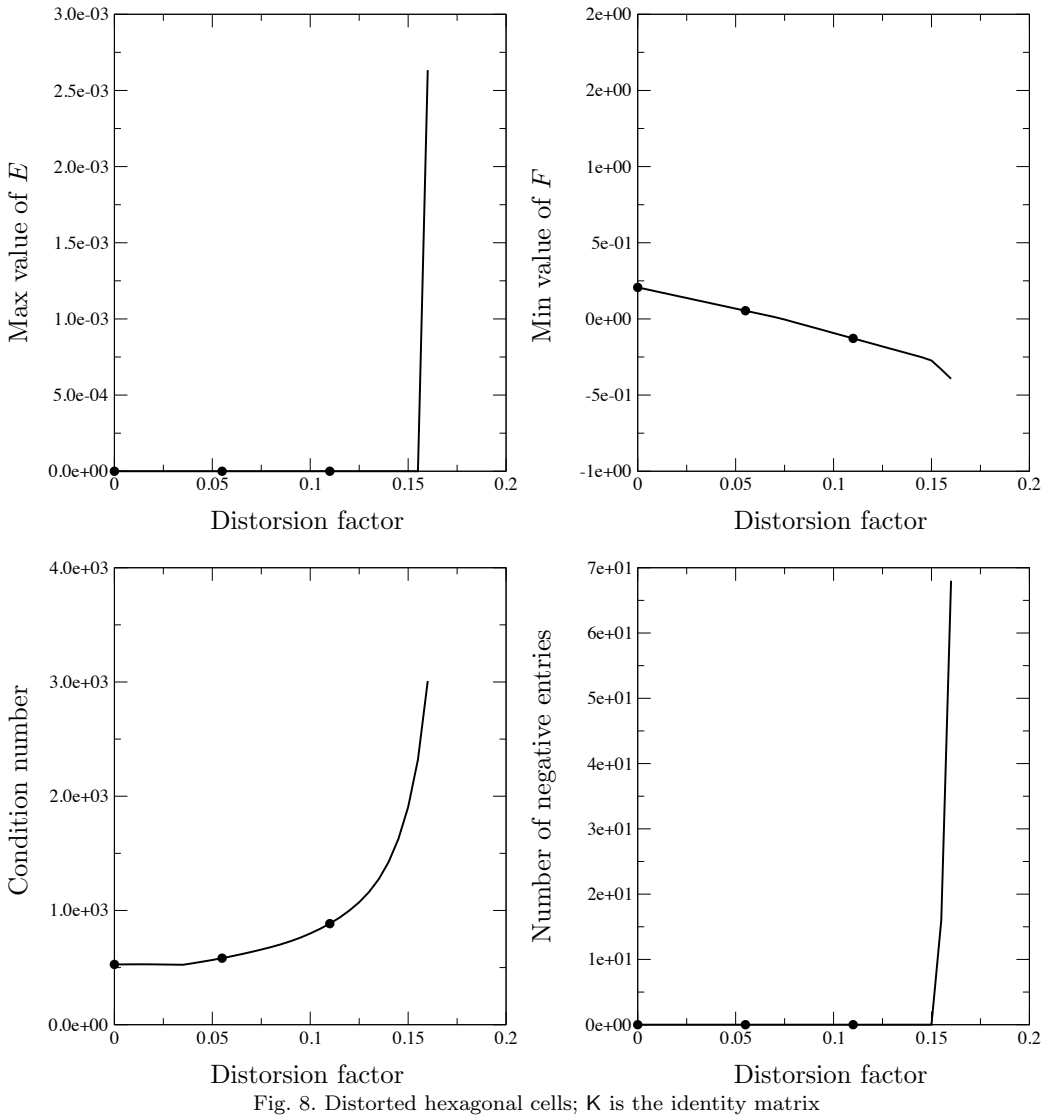

Fig. 8. Distorted hexagonal cells; K is the identity matrix

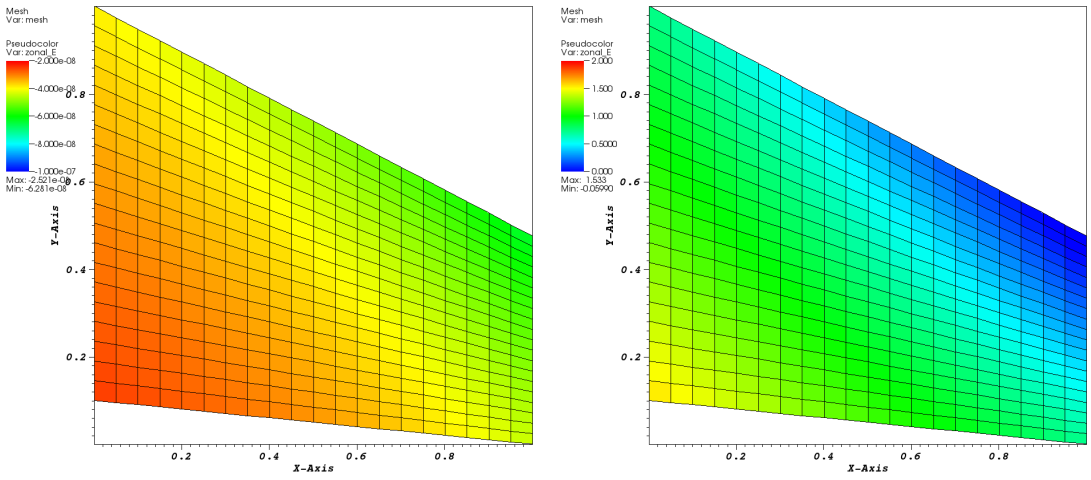

Fig. 9. Slope, quadrilateral cells; map of E (left) and F (right) when the monotonicity condition (A1) is first violated

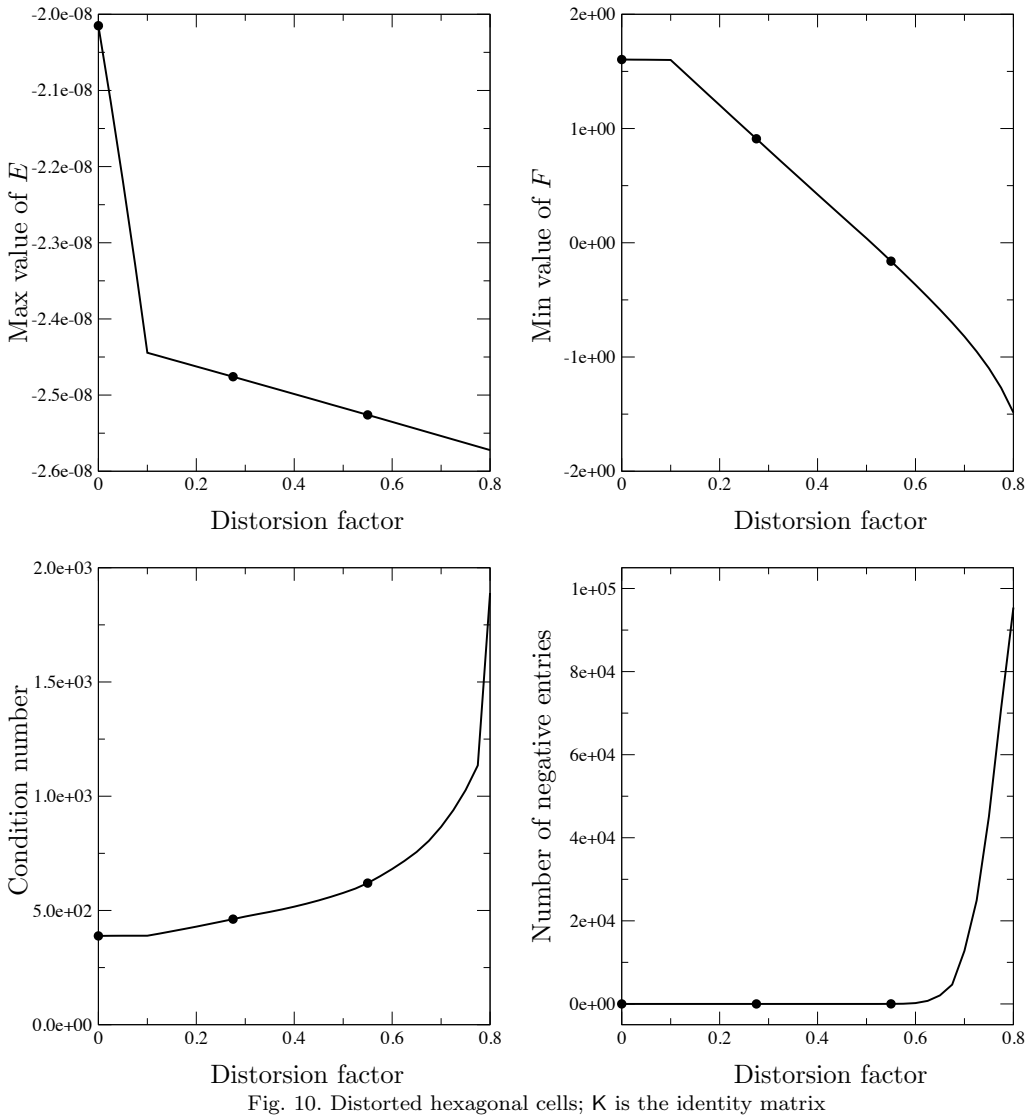

Fig. 10. Distorted hexagonal cells; K is the identity matrix

References

[1] A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Classics in Applied Mathematics. SIAM, New York, 1994.

[2] E. Bertolazzi and G. Manzini. A second-order maximum principle preserving finite volume method for steady convection-diffusion problems. *SIAM, J. Numer. Anal.*, 43(5):2172–2199, 2005.

[3] J. H. Bramble. Fourth-order finite difference analogues of the dirichlet problem for poisson's equation in three and four dimensions. *Math. Comp.*, 17:217–222, 1963.

[4] J. H.; Bramble and B. E. Hubbard. A priori bounds on the discretization error in the numerical solution of the dirichlet problem. *Contributions to Differential Equations*, 2:229–252, 1963.

[5] J. H. Bramble and B. E. Hubbard. New monotone type approximations for elliptic problems. *Mathematics of Computation*, 18(87):349–367, 1964.

[6] J. H. Bramble and B. E. Hubbard. On a finite difference analogue of an elliptic boundary problem which is neither diagonally dominant nor of non-negative type. *J. Math. and Phys.*, 43:117–132, 1964.

[7] J. H. Brandts, S. Korotov, and M. Krizek. The discrete maximum principle for linear simplicial finite element approximations of a reaction-diffusion problem. *Linear Algebra and its Applications*, 429(10):2344–2357, 2008. Special Issue in honor of Richard S. Varga.

[8] F. Brezzi, K. Lipnikov, and M. Shashkov. Convergence of the mimetic finite difference method for diffusion problems on polyhedral meshes. *SIAM J. Numer. Anal.*, 43(5):1872–1896, 2005.

[9] F. Brezzi, K. Lipnikov, and V. Simoncini. A family of mimetic finite difference methods on polygonal and polyhedral meshes. *Math. Models Methods Appl. Sci.*, 15(10):1533–1551, 2005.

[10] E. Burman and A. Ern. Discrete maximum principle for galerkin approximations of the laplace operator on arbitrary meshes. *Comptes Rendus Mathematique*, 338(8):641–646, 2004.

[11] A. Cangiani and G. Manzini. Flux reconstruction and pressure post-processing in mimetic finite difference methods. *Comput. Methods Appl. Mech. Engrg.*, 197(9-12):933–945, 2008.

[12] P. G. Ciarlet. Some results in the theory of nonnegative matrices. *Linear Algebra and its Applications*, 1(1):139–152, 1968.

[13] P. G. Ciarlet. Discrete maximum principle for finite-difference operators. *Aequationes Mathematicae*, 4(3):338–352, 1970.

[14] P. G. Ciarlet and P.-A. Raviart. Maximum principle and uniform convergence for the finite element method. *Computer Methods in Applied Mechanics and Engineering*, 2(1):17–31, 1973.

[15] A. Draganescu, T. F. Dupont, and L. R. Scott. Failure of the discrete maximum principle for an elliptic finite element problem. *Math. Comp.*, 74(249):1–23, 2004.

[16] M. A. T. Elshebli. Discrete maximum principle for the finite element solution of linear non-stationary diffusion-reaction problems. *Applied Mathematical Modelling*, 32(8):1530–1541, 2008. Special issue on numerical and computational issues related to applied mathematical modelling.

[17] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monograph and Studies in Mathematics*. Pitman, Boston, 1985.

[18] E. Hopf. Elementare Bemerkungen uber die L osungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus. *Sitzungsber. Preuss. Akad. Wiss.*, 19, 1927.

[19] K. Ishihara. Strong and weak discrete maximum principles for matrices associated with elliptic problems. *Linear Algebra and its Applications*, 88-89:431–448, 1987.

[20] J. Karátson, S. Korotov, and M. Krizek. On discrete maximum principles for nonlinear elliptic problems. *Mathematics and Computers in Simulation*, 76(1-3):99–108, 2007.

[21] S. Korotov, M. Krízek, and P. Neittaanm aki. Weakened acute type condition for tetrahedral triangulations and the discrete maximum principle,. *Math. Comp.*, 223:107–119, 2001.

[22] D. Kuzmin, M. J. Shashkov, and D. Svyatskiy. A constrained finite element method satisfying the discrete maximum principle for anisotropic diffusion problems. *J. Comput. Phys.*, 228(9):3448–63, 2009.

[23] K. Lipnikov, G. Manzini, and D. Svyatskiy. Analysis of the monotonicity conditions in the mimetic finite difference method for elliptic problems. *J. Comput. Phys.*, 230(7):2620–2642, 2011.

[24] K. Lipnikov, M. Shashkov, and I. Yotov. Local flux mimetic finite difference methods. *Numer. Math.*, 112(1):115–152, 2009.

[25] K. Lipnikov, D. Svyatskiy, and Y. Vassilevski. A monotone finite volume method for advection-diffusion equations on unstructured polygonal meshes. *J. Comput. Phys.*, 229(11):4017–4032, 2010.

[26] R. Liska and M. Shaskov. Enforcing the discrete maximum principle for linear finite element solutions of second-order elliptic problems. *Commun. Comput. Phys.*, 3:852–877, 2008.

[27] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.

[28] J. M. Nordbotten, I. Aavatsmark, and G. T. Eigestad. Monotonicity of control volume methods. *Numer. Math.*, 106(2):255–288, 2007.

[29] C. Le Potier. Finite volume scheme satisfying maximum and minimum principles for anisotropic diffusion operators. In R. Eymard and J.-M. Herard, editors, *Finite Volumes for Complex Applications V*, pages 103–118, 2008.

[30] P. Solin and T. Vejchodsky. A weak discrete maximum principle for hp-fem. *J. Comput. Appl. Math.*, 209(1):54–65, 2007.

[31] G. Stampacchia. Le problème de dirichlet pour les èquations elliptiques du second ordre à coefficients discontinus. *Ann. Inst. Fourier*, 15:189–258, 1965.

[32] T. J. Linde T. Plewa and V. G. Weirs, editors. *Adaptive mesh refinement, theory and applications*, volume 41 of *Lecture notes in computational science and engineering*. Springer, 2005.
[33] R. Varga. *Matrix Iterative Methods*. Prentice-Hall, Inc., 1962.
[34] R. Varga. On a discrete maximum principle. *SIAM J. Numer. Anal.*, 3(2):355–359, 1966.
[35] T. Vejchodsky, S. Korotov, and A. Hannukainen. Discrete maximum principle for parabolic problems solved by prismatic finite elements. *Mathematics and Computers in Simulation*, 2009. In Press, Corrected Proof.
[36] Andreas Wächter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.