A coupled electro-thermal Discontinuous Galerkin method

L. Homsi[a], C. Geuzaine[b], L. Noels[a,*]

[a] *Computational & Multiscale Mechanics of Materials (CM3),*
Department of Aerospace and Mechanical Engineering,
University of Liège,
Quartier Polytech 1, Allée de la Découverte 9, B-4000 Liège, Belgium
[b] *Department of Electrical Engineering and Computer Science,*
University of Liège,
Quartier Polytech 1, Allée de la Découverte 10, B-4000 Liège, Belgium

Abstract

This paper presents a Discontinuous Galerkin scheme in order to solve the nonlinear elliptic partial differential equations of coupled electro-thermal problems.

In this paper we discuss the fundamental equations for the transport of electricity and heat, in terms of macroscopic variables such as temperature and electric potential. A fully coupled nonlinear weak formulation for electro-thermal problems is developed based on continuum mechanics equations expressed in terms of energetically conjugated pair of fluxes and fields gradients. The weak form can thus be formulated as a Discontinuous Galerkin method.

The existence and uniqueness of the weak form solution are proved. The numerical properties of the nonlinear elliptic problems i.e., consistency and stability, are demonstrated under specific conditions, i.e. use of high enough stabilization parameter and at least quadratic polynomial approximations. Moreover the prior error estimates in the H^1-norm and in the L^2-norm are shown to be optimal in the mesh size with the polynomial approximation degree.

Keywords: Discontinuous Galerkin Method, Electro-thermal coupling, nonlinear elliptic problem, error estimates

1. Introduction

Electro-thermal materials received a significant interest in recent years due to their capability to convert electricity directly into heat and vice versa, which promises a wide range of applications in energy and environment fields. The main interest of this work is to derive a consistent and stable Discontinuous Galerkin (DG) method for two-way electro-thermal coupling analyzes considering electro-thermal effects such as Seebeck and Peltier effects, but also Joule heating. These effects describe the direct conversion of the difference in electric potential into a temperature difference within the system (Peltier effect), and vice versa (Seebeck effect). This is typical of thermo-electric cells which could work in two ways: electric generations [1, 2] and heat pumps which operate in cool or heat modes [3].

[*] Corresponding author, Phone: +32 4 366 48 26, Fax: +32 4 366 95 05
Email addresses: `Lina.Homsi@student.ulg.ac.be` (L. Homsi), `CGeuzaine@ulg.ac.be` (C. Geuzaine), `L.Noels@ulg.ac.be` (L. Noels)

Electro-thermal continuum has extensively been developed in the literature [3, 4, 5, 6]. For example, as a non-exhaustive list, Ebling et al. [1] have implemented thermo-electric elements into the finite element method and have validated it by analytical and experimental results for the figure of merit values. Liu [6] has developed a continuum theory of thermo-electric bodies. He has applied it to predict the effective properties of thermo-electric composites. However he has considered that the temperature and voltage variations are small, which leads to a linear system of partial differential equations. Pérez-Aparicio et al. [5] have proposed an electro-thermal formulation for simple configurations and have provided a comparison between analytical and numerical results.

The constitutive equations that govern electro-thermal coupling can be formulated in term of $(\frac{-\mathrm{V}}{\mathrm{T}}, \frac{1}{\mathrm{T}})$ instead of (V, T), where T is the temperature and V is the electric potential. Such a formulation has been considered in the literature, e.g. by Mahan [4], Yang et al. [7], Liu [6], in order to obtain a conjugated pair of fluxes and fields gradients. Mahan [4] has provided a comparison between the different energy fluxes that have been developed and used by different researchers and concluded that all these different treatments result in the same equation.

In this paper a discontinuous Galerkin weak formulation in terms of energetically conjugated fields gradients and fluxes is proposed. The main strengths of the DG method are the ease in handling elements of different types, the high order accuracy reached for higher polynomial orders and the high scalability in parallel simulations. Indeed the possibility of using irregular and non conforming meshes in algorithm makes it suitable for time dependent transient problems as it allows for easy mesh modification dynamically with time. Above all, since the DG method allows discontinuities of the physical unknowns within the interior of the problem domain, it is a natural approach to capture the jumps across the material interface in electro-thermal coupled problems. However, if not correctly formulated, discontinuous methods can exhibit instabilities, and the numerical results fail to approximate the exact solution. For practitioners, it is important to have methods available which yield reliable results for a wide variety of problems. By using an adequate inter element flux definition combined to stabilization techniques, the shortcomings of non-stabilized DG methods can be overcome, e.g. [8, 9, 10].

Since the seminal work of Reed et al. [11], DG methods have been developed to solve hyperbolic, parabolic, and elliptic problems. The state of the art of DG methods and their developments can be found in [12]. Most of DG methods for elliptic and parabolic problems rely on the Interior Penalty (IP) method. The main principle of IP, as introduced by [13, 14], is to constrain weakly the compatibility instead of building it into the finite element which makes it easier to use discontinuous polynomial spaces of high degree. The interest in the symmetric interior penalty (SIPG) methods, which will be considered in this paper, has been renewed by Wheeler [14] due to demands for optimality of convergence rates with the mesh size $\mathrm{h_s}$ (i.e., the rates of the convergence is k in the H^1-norm and $\mathrm{k}+1$ in the L^2-norm, where k is the polynomial approximation degree). However there exist other possible choices of traces and numerical fluxes as discussed by Arnold et al. [15], who have provided an analysis of a large class of discontinuous methods for second order elliptic problems with different numerical fluxes, and declared that IP, NIPG (Non-Symmetric Interior Penalty), LDG (Local discontinuous Galerkin) and other DG methods are consistent and stables methods. In particular Arnold et al. [15] have proposed a framework for dealing with linear elliptic problems by means of DG methods and demonstrated that DG methods which are consistent, adjoint consistent, and stable achieve optimal error estimates, and that the inconsistent DG methods like the pure penalty methods can still achieve optimal error estimates provided they are super-penalized. Besides, Georgoulis [16] has derived anisotropic hp version error bounds for linear second order elliptic diffusion convection reaction using Discontinuous Galerkin finite element methods (SIPG and NIPG), on shape-regular and anisotropic elements, and for isotropic and anisotropic polynomial degrees for the element bases. He has also observed optimal order of convergence in the L^2-norm for the SIPG formulation when a uniform mesh size refinement for different values of k is employed. Moreover, he has shown that the solution of the adjoint problem suffers from sub-optimal rates of convergence when a NIPG formulation is used. Yadav et al. [17] have extended the DG methods from a linear self-adjoint elliptic problem to a second order nonlinear elliptic problem. The nonlinear system resulting from DG methods is then analyzed based on a fixed point argument. They have shown that the error estimate in the L^2-norm for piece-wise polynomials of degree $\mathrm{k} \geq 1$ is $\mathrm{k}+1$. They have also provided numerical results to illustrate the theoretical results. Gudi et al. [18] have proposed an analysis for the most popular DG schemes such as SIPG, NIPG and LDG methods for nonlinear elliptic problems, and the error estimates have been studied for each of these methods by reformulating the problems in a fixed point form. In addition, according to Gudi et al. [18], optimal errors in the H^1-norm and in L^2-norm are proved for SIPG for polynomial degrees larger or equal to 2, and a loss in the optimality in L^2-norm is observed for NIPG and LDG. In that work a deterioration in the order of convergence in $\mathrm{h_s}$ is noted when linear polynomials are used.

Recently, DG has been used to solve coupled problems. For instance Wheeler and Sun [19] have proposed a primal DG method with interior penalty (IP) terms to solve coupled reactive transport in porous media. In that work, a cut off operator is used in the DG scheme to treat the coupling and achieve convergence. They have declared that optimal convergence rates for both flow and transport terms can be achieved if the same polynomial degree of approximation is used. However if they are different, the behavior for the coupled system is controlled by the part with the lowest degree of approximation, and the error estimate in $\mathrm{L}^2(\mathrm{H}^1)$-norm is nearly optimal in k with a loss of $\frac{1}{2}$ when polynomials with different degrees are used. Furthermore, Zheng et al. [20] have proposed a DG method to solve the thermo-elastic coupling problems due to temperature and pressure dependent thermal contact resistance. In that work the DG method is used to simulate the temperature jump, and the mechanical sub-problem is solved by the DG finite element method with a penalty function.

To the authors knowledge there is no development related to DG methods for electro-thermal coupling, which is the aim of this paper. The main advantage of this work is the aptitude to deal with complex geometry and the capability of the formulation to capture the electro-thermal behavior for composite materials with high contrast: one phase has a high electric conductivity (e.g., carbon fiber) and other is a resistive material (e.g., polymers). The key point in being able to develop a stable DG method for electro-thermal coupling is to formulate the equations in terms of energetically conjugated pairs of fluxes and fields gradients. Indeed, the use of energetically consistent pairs allow writing the strong form in a matrix form suitable to the derivation of a SIPG weak form as it will be demonstrated in this paper.

This paper is organized as follows. Section 2 describes the governing equations of electro-thermal materials. In order to develop the DG formulation, the weak form is formulated in terms of a conjugated pair of fluxes and fields gradients, resulting in a particular choice of the test functions $(\delta(\frac{1}{\mathrm{T}}), \delta(\frac{-\mathrm{V}}{\mathrm{T}}))$ and of the trial functions $(\frac{1}{\mathrm{T}}, \frac{-\mathrm{V}}{\mathrm{T}})$. A complete nonlinear coupled finite element algorithm for electro-thermal materials is then developed in Section 3 using the DG method to derive the weak form. This results into a set of nonlinear equations which is implemented within a three-dimensional finite element code. Section 4 focuses on the demonstration of the numerical properties of the DG method under the assumption of a bi-dimensional problem, based on rewriting the nonlinear formulation in a fixed point form following closely the approach described in [21, 18]. The numerical properties of the nonlinear elliptic problem, i.e. consistency and the uniqueness of the solution, can then be demonstrated, and the prior error estimate is shown to be optimal in the mesh size for polynomial approximation degrees $\mathrm{k} > 1$. In Section 5, several examples of applications in one, two, and three dimensions are provided for single and composite materials, in order to validate the accuracy and effectiveness of the electro-thermal DG formulation and to illustrate the algorithmic properties.

We end by some conclusions, remarks, and perspectives in Section 6.

2. Governing equations

In this section an overview of the basic equations that govern the electro-thermal phenomena is presented for a structure characterized by a domain $\Omega \subset \mathbb{R}^{\mathrm{d}}$, with $\mathrm{d} = 2$, or 3 the space dimension, whose external boundary is $\partial\Omega$. In particular we discuss the choice of the conjugated pair of fluxes and fields gradients that will be used to formulate the strong form in a matrix form.

2.1. Strong form

The first balance equation is the electrical charge conservation equation. When assuming a steady state, the solution of the electrical problem consists in solving the following Poisson type equation for the electrical potential

$$\nabla \cdot \mathbf{j}_{\mathrm{e}} = 0 \qquad \forall\, \mathbf{x} \in \Omega\,, \tag{1}$$

where $\mathbf{j}_{\mathrm{e}}$ [A/m^2] denotes the flow of electrical current density vector, which is defined as the rate of charge carriers per unit area or the current per unit area. At zero temperature gradient, the current density $\mathbf{j}_{\mathrm{e}}$ is described by Ohm's law which is the relationship between the electric potential V [V] gradient and the electric current flux per unit area through the electric conductivity $\mathbf{l}$ [S/m], with

$$\mathbf{j}_{\mathrm{e}} = \mathbf{l} \cdot (-\nabla \mathrm{V}). \tag{2}$$

However when T [K] varies inside the body, an electromotive force $(\nabla \mathrm{V})^{\mathrm{s}}$ per unit length appears, and reads

$$(\nabla \mathrm{V})^{\mathrm{s}} = -\alpha \nabla \mathrm{T}, \tag{3}$$

where α [V/K] is the Seebeck coefficient which is in general temperature dependent and defined as the derivative of the electric potential with respect to the temperature. By taking in consideration the Seebeck effect, Eq. (3), and adding it to Ohm's Law, Eq.(2), for systems in which the particle density is homogeneous [4], the current density is rewritten as

$$\mathbf{j}_{\mathrm{e}} = \mathbf{l} \cdot (-\nabla \mathrm{V}) + \alpha \mathbf{l} \cdot (-\nabla \mathrm{T}). \tag{4}$$

The second balance equation is the conservation of the energy flux, which is a combination of the inter exchanges between the thermal and electric energies:

$$\nabla \cdot \mathbf{j}_{\mathrm{y}} = -\partial_{\mathrm{t}} \mathrm{y} \qquad \forall\, \mathbf{x} \in \Omega. \tag{5}$$

The right hand side of this equilibrium equation is the time derivative of the internal energy density y [J/m^3]

$$\mathrm{y} = \mathrm{y}_0 + \mathrm{c}_{\mathrm{v}}\, \mathrm{T}, \tag{6}$$

which consists of the constant y_0 independent of the temperature and of the electric potential, and of the volumetric heat capacity c_{v} [J/(K$\cdot$m^3)] multiplied by the absolute temperature T. Moreover the energy flux $\mathbf{j}_{\mathrm{y}}$ is defined as

$$\mathbf{j}_{\mathrm{y}} = \mathbf{q} + \mathrm{V}\, \mathbf{j}_{\mathrm{e}}, \tag{7}$$

where $\mathbf{q}$ [W/m^2] is the heat flux. On the one hand, at zero electric current density, $\mathbf{j}_\mathrm{e} = 0$ (open circuit), the heat flux is given by the Fourier 's Law

$$\mathbf{q} = \mathbf{k} \cdot (-\nabla \mathrm{T}), \tag{8}$$

where $\mathbf{k}$ [W/(K · m)] denotes the symmetric matrix of thermal conductivity coefficients, which may depend on the temperature. On the other hand, at zero temperature gradient, the heat flux is given by

$$\mathbf{q} = \beta_\alpha \, \mathbf{j}_\mathrm{e} = \alpha \, \mathrm{T} \, \mathbf{j}_\mathrm{e}, \tag{9}$$

where the coupling between the heat flux $\mathbf{q}$ and the electric current density $\mathbf{j}_\mathrm{e}$ is governed by the Peltier coefficient $\beta_\alpha = \alpha \, \mathrm{T}$. By superimposing the previous terms to the Fourier's Law, Eq. (8), the thermal flux can be rewritten as:

$$\mathbf{q} = \mathbf{k} \cdot (-\nabla \mathrm{T}) + \alpha \mathrm{T} \, \mathbf{j}_\mathrm{e} = (\mathbf{k} + \alpha^2 \, \mathrm{T} \mathbf{l}) \cdot (-\nabla \mathrm{T}) + \alpha \mathrm{T} \mathbf{l} \cdot (-\nabla \mathrm{V}). \tag{10}$$

The first term is due to the conduction and the second term corresponds to the joule heating effect.

Let us now define the Sobolev space $\mathrm{W}^\mathrm{s}_\mathrm{r}(\Omega)$, with s a non-negative integer and $\mathrm{r} \in [1, \infty[$, the subspace of all functions from the norm $\mathrm{L}^\mathrm{r}(\Omega)$ whose generalized derivatives up to order s exist and belong to $\mathrm{L}^\mathrm{r}(\Omega)$, which is defined as

$$\mathrm{W}^\mathrm{s}_\mathrm{r}(\Omega) = \left\{ \mathrm{f} \in \mathrm{L}^\mathrm{r}(\Omega), \;\; \partial^\alpha \mathrm{f} \in \mathrm{L}^\mathrm{r}(\Omega) \;\; \forall \mid \alpha \mid \leq \mathrm{s}, \; \mathrm{s} \geq 1 \right\}. \tag{11}$$

When $\mathrm{r} = 2$, the spaces are Hilbert spaces equipped with the scalar product: $\mathrm{W}^\mathrm{s}_2(\Omega) = \mathrm{H}^\mathrm{s}(\Omega)$. For $\mathrm{s} = 0$, the norm is the L^2 norm.

Therefore the conservation laws are written as finding V, $\mathrm{T} \in \mathrm{H}^2(\Omega) \times \mathrm{H}^{2^+}(\Omega)$ such that

$$\nabla \cdot \mathbf{j}_\mathrm{e} = 0 \qquad \forall \, \mathbf{x} \in \Omega, \tag{12}$$

$$\nabla \cdot \mathbf{j}_\mathrm{y} = \nabla \cdot \mathbf{q} + \mathbf{j}_\mathrm{e} \cdot \nabla \mathrm{V} = -\partial_\mathrm{t} \mathrm{y} \qquad \forall \, \mathbf{x} \in \Omega, \tag{13}$$

where T belongs to the manifold H^{2^+}, such that T is always strictly positive.

These equations are completed by suitable boundary conditions, where the boundary $\partial\Omega$ is decomposed into a Dirichlet boundary $\partial_\mathrm{D}\Omega$ and a Neumann boundary $\partial_\mathrm{N}\Omega$ (i.e., $\partial_\mathrm{D}\Omega \cup \partial_\mathrm{N}\Omega = \partial\Omega$, and $\partial_\mathrm{D}\Omega \cap \partial_\mathrm{N}\Omega = 0$). On the Dirichlet BC, one has

$$\mathrm{T} = \bar{\mathrm{T}} > 0 \, , \quad \mathrm{V} = \bar{\mathrm{V}} \qquad \forall \, \mathbf{x} \in \partial_\mathrm{D}\Omega, \tag{14}$$

where $\bar{\mathrm{T}}$ and $\bar{\mathrm{V}}$ are the prescribed temperature and electric potential respectively. The natural Neumann boundary conditions are constraints on the secondary variables: the electric current for the electric charge equation and the energy flux for the energy equation, i.e.

$$\mathbf{j}_\mathrm{e} \cdot \mathbf{n} = \bar{\mathrm{j}}_\mathrm{e} \, , \quad \mathbf{j}_\mathrm{y} \cdot \mathbf{n} = \bar{\mathrm{j}}_\mathrm{y} \;\; \forall \, \mathbf{x} \in \partial_\mathrm{N}\Omega, \tag{15}$$

where $\mathbf{n}$ is the outward unit normal to the boundary $\partial\Omega$. For simplicity we consider the same boundary division into Neumann and Dirichlet parts for the both fields T and V. However in the general case this could be different.

The set of Eqs. (12, 13) can be rewritten under a matrix form. First we rewrite Eqs. (4, 7, 10) under the form

$$\mathbf{j} = \left(\begin{array}{c} \mathbf{j}_{\mathrm{e}} \\ \mathbf{j}_{\mathrm{y}} \end{array} \right) = \left(\begin{array}{cc} \mathbf{l} & \alpha\mathbf{l} \\ \mathrm{V}\mathbf{l} + \alpha\mathrm{T}\mathbf{l} & \mathbf{k} + \alpha\mathrm{V}\mathbf{l} + \alpha^2\mathrm{T}\mathbf{l} \end{array} \right) \left(\begin{array}{c} -\nabla\mathrm{V} \\ -\nabla\mathrm{T} \end{array} \right). \tag{16}$$

The set of governing Eqs. (12, 13) thus becomes finding V, $\mathrm{T} \in \mathrm{H}^2(\Omega) \times \mathrm{H}^{2^+}(\Omega)$ such that

$$\mathrm{div}\,(\mathbf{j}) = \left(\begin{array}{c} 0 \\ -\partial_{\mathrm{t}}\mathrm{y} \end{array} \right) = \mathbf{i}, \tag{17}$$

where we have introduced $\mathbf{i} = \left(\begin{array}{c} 0 \\ -\partial_{\mathrm{t}}\mathrm{y} \end{array} \right)$ for a future use.

2.2. The conjugated driving forces

First the weak form of the conservation of electric charge carriers, Eq. (1), is obtained by taking the inner product of this equation with a suitable scalar test function $\delta\mathrm{f}_{\mathrm{V}} \in \mathrm{H}^1(\Omega')$ over a sub-domain $\Omega' \subset \Omega$, yielding

$$\int_{\Omega'} \nabla \cdot \mathbf{j}_{\mathrm{e}} \delta\mathrm{f}_{\mathrm{V}} \mathrm{d}\Omega' = 0 \quad \forall \delta\mathrm{f}_{\mathrm{V}} \in \mathrm{H}^1(\Omega'). \tag{18}$$

After a simple formal integration by parts and using the divergence theorem, we obtain

$$-\int_{\Omega'} \mathbf{j}_{\mathrm{e}} \cdot \nabla\delta\mathrm{f}_{\mathrm{V}}\ \mathrm{d}\Omega' + \int_{\partial\Omega'} \mathbf{j}_{\mathrm{e}} \cdot \mathbf{n}\delta\mathrm{f}_{\mathrm{V}}\ \mathrm{dS} = 0 \quad \forall \delta\mathrm{f}_{\mathrm{V}} \in \mathrm{H}^1(\Omega'). \tag{19}$$

Secondly, taking the inner product of the second balance equation, Eq. (13), with the test function $\delta\mathrm{f}_{\mathrm{T}} \in \mathrm{H}^1(\Omega')$, over the sub-domain $\Omega' \subset \Omega$ leads to

$$\int_{\Omega'} \nabla \cdot \mathbf{q}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' + \int_{\Omega'} \mathbf{j}_{\mathrm{e}} \cdot \nabla\mathrm{V}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' = -\int_{\Omega'} \partial_{\mathrm{t}}\mathrm{y}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' \quad \forall \delta\mathrm{f}_{\mathrm{T}} \in \mathrm{H}^1(\Omega'). \tag{20}$$

Moreover by applying the divergence theorem, one obtains

$$\int_{\Omega'} \mathbf{q} \cdot \nabla\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' = \int_{\partial\Omega'} \mathbf{q} \cdot \mathbf{n}\delta\mathrm{f}_{\mathrm{T}} \mathrm{dS} + \int_{\Omega'} \nabla\mathrm{V} \cdot \mathbf{j}_{\mathrm{e}}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' + \int_{\Omega'} \partial_{\mathrm{t}}\mathrm{y}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' \quad \forall \delta\mathrm{f}_{\mathrm{T}} \in \mathrm{H}^1(\Omega'). \tag{21}$$

By substituting the internal energy, Eq. (6), and the thermal flux, Eq. (10), this last equation reads

$$\begin{aligned} \int_{\Omega'} (\mathbf{k} \cdot (-\nabla\mathrm{T}) + \alpha\mathrm{T}\mathbf{j}_{\mathrm{e}}) \cdot \nabla\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' = & \int_{\Omega'} \mathrm{c}_{\mathrm{V}}\, \partial_{\mathrm{t}}\mathrm{T}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' + \int_{\Omega'} \nabla\mathrm{V} \cdot \mathbf{j}_{\mathrm{e}}\delta\mathrm{f}_{\mathrm{T}} \mathrm{d}\Omega' \\ & + \int_{\partial\Omega'} (\mathbf{k} \cdot (-\nabla\mathrm{T}) + \alpha\mathrm{T}\mathbf{j}_{\mathrm{e}}) \cdot \mathbf{n}\delta\mathrm{f}_{\mathrm{T}} \mathrm{dS}. \end{aligned} \tag{22}$$

In order to define the conjugated forces, let us substitute $\delta\mathrm{f}_{\mathrm{V}}$ by $-\frac{\mathrm{V}}{\mathrm{T}}$ in Eq. (19). This results into

$$\int_{\partial\Omega'} \mathbf{j}_{\mathrm{e}} \cdot \mathbf{n}(-\frac{\mathrm{V}}{\mathrm{T}})\mathrm{dS} = \int_{\Omega'} \mathbf{j}_{\mathrm{e}} \cdot (-\frac{\nabla\mathrm{V}}{\mathrm{T}} + \frac{\mathrm{V}}{\mathrm{T}^2}\nabla\mathrm{T})\mathrm{d}\Omega'. \tag{23}$$

Substituting $\delta \mathrm{f_T}$ by $\frac{1}{\mathrm{T}} \in \mathrm{H}^{1^+}(\Omega')$ in Eq. (22) leads to:

$$\begin{aligned}\int_{\Omega'} \left((-\nabla \mathrm{T}) \cdot \mathbf{k} \cdot \frac{(-\nabla \mathrm{T})}{\mathrm{T}^2} - \alpha \frac{\mathbf{j}_\mathrm{e}}{\mathrm{T}} \cdot \nabla \mathrm{T} \right) \mathrm{d}\Omega' = \int_{\Omega'} (\frac{\mathrm{c_V}}{\mathrm{T}}\, \partial_\mathrm{t} \mathrm{T}) \mathrm{d}\Omega' + \int_{\Omega} \nabla \mathrm{V} \cdot \frac{\mathbf{j}_\mathrm{e}}{\mathrm{T}} \mathrm{d}\Omega' \\ + \int_{\partial\Omega'} \left(\mathbf{k} \cdot (\frac{-\nabla \mathrm{T}}{\mathrm{T}}) + \alpha \mathbf{j}_\mathrm{e} \right) \cdot \mathbf{n} \mathrm{dS}.\end{aligned} \tag{24}$$

By subtracting Eq. (23) from Eq. (24), one gets

$$\begin{aligned}&\int_{\Omega'} \frac{\mathrm{c_V}}{\mathrm{T}}\, \partial_\mathrm{t} \mathrm{T} \mathrm{d}\Omega' + \int_{\partial\Omega'} \left(\mathbf{k} \cdot (\frac{-\nabla \mathrm{T}}{\mathrm{T}}) + \alpha \mathbf{j}_\mathrm{e} + \mathbf{j}_\mathrm{e} (\frac{\mathrm{V}}{\mathrm{T}}) \right) \cdot \mathbf{n} \mathrm{dS} \\ &= \int_{\Omega'} \left(-\mathbf{j}_\mathrm{e} \cdot \frac{\nabla \mathrm{V}}{\mathrm{T}} + \mathbf{j}_\mathrm{e} \cdot \frac{\nabla \mathrm{V}}{\mathrm{T}} - \mathbf{j}_\mathrm{e} \frac{\mathrm{V}}{\mathrm{T}^2} \cdot \nabla \mathrm{T} \right) \mathrm{d}\Omega' + \int_{\Omega'} \left((-\nabla \mathrm{T}) \cdot \mathbf{k} \cdot \frac{(-\nabla \mathrm{T})}{\mathrm{T}^2} - \alpha \frac{\mathbf{j}_\mathrm{e}}{\mathrm{T}} \cdot \nabla \mathrm{T} \right) \mathrm{d}\Omega',\end{aligned} \tag{25}$$

or

$$\int_{\Omega'} \frac{1}{\mathrm{T}} (\mathrm{c_V}\, \partial_\mathrm{t} \mathrm{T}) \mathrm{d}\Omega' + \int_{\partial\Omega'} \frac{1}{\mathrm{T}} (\mathbf{q} + \mathbf{j}_\mathrm{e} \mathrm{V}) \cdot \mathbf{n} \mathrm{dS} = \int_{\Omega'} \frac{-\nabla \mathrm{T}}{\mathrm{T}^2} \cdot (\mathbf{j}_\mathrm{e} \mathrm{V} - \mathbf{k} \cdot \nabla \mathrm{T} + \alpha \mathbf{j}_\mathrm{e} \mathrm{T})\, \mathrm{d}\Omega'. \tag{26}$$

Henceforth, as $\mathbf{j}_\mathrm{y} = \mathbf{q} + \mathbf{j}_\mathrm{e} \mathrm{V}$, this last result is rewritten as

$$\int_{\Omega'} \partial_\mathrm{t} \mathrm{y} \delta \mathrm{f_T} \mathrm{d}\Omega' + \int_{\partial\Omega'} \mathbf{j}_\mathrm{y} \cdot \mathbf{n} \delta \mathrm{f_T} \mathrm{dS} = \int_{\Omega'} \mathbf{j}_\mathrm{y} \cdot \nabla \delta \mathrm{f_T} \mathrm{d}\Omega'. \tag{27}$$

By this way we recover the conservation equation of the energy flux, Eq. (5), which shows that $\mathbf{j}_\mathrm{e}, \mathbf{j}_\mathrm{y}$ and $\nabla(-\frac{\mathrm{V}}{\mathrm{T}}), \nabla(\frac{1}{\mathrm{T}})$ is a conjugated pair of fluxes and fields gradients as shown in [6].

2.3. Strong form in terms of the conjugated pairs of fluxes and fields gradients

Let us define a 2×1 vector of the unknown fields $\mathbf{M} = \begin{pmatrix} \mathrm{f_V} \\ \mathrm{f_T} \end{pmatrix}$, with $\mathrm{f_V} = -\frac{\mathrm{V}}{\mathrm{T}}$ and $\mathrm{f_T} = \frac{1}{\mathrm{T}}$, then the gradients of the fields vector $\nabla \mathbf{M}$, a $2\mathrm{d} \times 1$ vector in terms of $(\nabla \mathrm{f_V}, \nabla \mathrm{f_T})$, is defined by

$$\begin{pmatrix} \nabla \mathbf{M} \end{pmatrix} = \begin{pmatrix} \nabla \mathrm{f_V} \\ \nabla \mathrm{f_T} \end{pmatrix} = \begin{pmatrix} \nabla(-\frac{\mathrm{V}}{\mathrm{T}}) \\ \nabla(\frac{1}{\mathrm{T}}) \end{pmatrix} = \begin{pmatrix} -\frac{1}{\mathrm{T}} \mathbf{I} & \frac{\mathrm{V}}{\mathrm{T}^2} \mathbf{I} \\ 0 & -\frac{1}{\mathrm{T}^2} \mathbf{I} \end{pmatrix} \begin{pmatrix} \nabla \mathrm{V} \\ \nabla \mathrm{T} \end{pmatrix}. \tag{28}$$

Hence, the fluxes defined by Eq. (16) can be expressed in terms of $\mathrm{f_V}, \mathrm{f_T}$, yielding

$$\mathbf{j} = \begin{pmatrix} \mathbf{j}_\mathrm{e} \\ \mathbf{j}_\mathrm{y} \end{pmatrix} = \begin{pmatrix} \mathbf{l} \mathrm{T} & \mathrm{VT} \mathbf{l} + \alpha \mathrm{T}^2 \mathbf{l} \\ \mathrm{VT} \mathbf{l} + \alpha \mathrm{T}^2 \mathbf{l} & \mathrm{T}^2 \mathbf{k} + 2\alpha \mathrm{T}^2 \mathrm{V} \mathbf{l} + \alpha^2 \mathrm{T}^3 \mathbf{l} + \mathrm{TV}^2 \mathbf{l} \end{pmatrix} \begin{pmatrix} \nabla \mathrm{f_V} \\ \nabla \mathrm{f_T} \end{pmatrix}. \tag{29}$$

The $2\mathrm{d} \times 1$ fluxes vector $\mathbf{j}$ is the product of the fields gradients vector $\nabla \mathbf{M}$, which derived from the state variables $(\mathrm{f_V}, \mathrm{f_T})$, by a coefficients matrix $\mathbf{Z}(\mathrm{V}, \mathrm{T})$ of size $2\mathrm{d} \times 2\mathrm{d}$, which is temperature and electric potential dependent. This formulation of the conjugated forces leads to a symmetric coefficients matrix $\mathbf{Z}(\mathrm{V}, \mathrm{T})$ such that

$$\mathbf{j} = \ \mathbf{Z} \quad \nabla \mathbf{M}. \tag{30}$$

From Eq. (29), the symmetric coefficients matrix $\mathbf{Z}(\mathrm{V},\mathrm{T})$ is positive definite if $\mathbf{Z}_{00}$ and $\mathbf{Z}_{11}$ - $\mathbf{Z}_{10}^{\mathrm{T}}\mathbf{Z}_{00}^{-1}\mathbf{Z}_{01}$ are positive definite. Since $\mathbf{Z}_{00} = \mathbf{l}\mathrm{T}$ is positive definite, and $\mathbf{Z}_{11}$ - $\mathbf{Z}_{10}^{\mathrm{T}}\mathbf{Z}_{00}^{-1}\mathbf{Z}_{01} = \mathbf{k}\mathrm{T}^2$ is also positive definite, then $\mathbf{Z}(\mathrm{V},\mathrm{T})$ is a positive definite matrix.

The coefficient matrix $\mathbf{Z}(\mathrm{V},\mathrm{T})$ in Eq. (29) could also be rewritten in terms of $(\mathrm{f_V},\mathrm{f_T}) = (-\frac{\mathrm{V}}{\mathrm{T}}, \frac{1}{\mathrm{T}})$, since $\mathrm{T} = \frac{1}{\mathrm{f_T}}, \mathrm{V} = -\frac{\mathrm{f_V}}{\mathrm{f_T}}$, as

$$\mathbf{Z}(\mathrm{f_V},\mathrm{f_T}) = \begin{pmatrix} \frac{1}{\mathrm{f_T}}\mathbf{l} & -\frac{\mathrm{f_V}}{\mathrm{f_T^2}}\mathbf{l} + \alpha\frac{1}{\mathrm{f_T^2}}\mathbf{l} \\ -\frac{\mathrm{f_V}}{\mathrm{f_T^2}}\mathbf{l} + \alpha\frac{1}{\mathrm{f_T^2}}\mathbf{l} & \frac{\mathbf{k}}{\mathrm{f_T^2}} - 2\alpha\frac{\mathrm{f_V}}{\mathrm{f_T^3}}\mathbf{l} + \alpha^2\frac{1}{\mathrm{f_T^3}}\mathbf{l} + \frac{\mathrm{f_V^2}}{\mathrm{f_T^3}}\mathbf{l} \end{pmatrix}. \tag{31}$$

Since the coefficients matrix is positive definite, the energy can be defined by

$$\begin{aligned} \nabla\mathbf{M}^{\mathrm{T}}\mathbf{j} &= \nabla\mathbf{M}^{\mathrm{T}}\mathbf{Z}(\mathrm{f_V},\mathrm{f_T})\nabla\mathbf{M} \\ &= \begin{pmatrix} \nabla\mathrm{f_V} & \nabla\mathrm{f_T} \end{pmatrix} \begin{pmatrix} \frac{1}{\mathrm{f_T}}\mathbf{l} & -\frac{\mathrm{f_V}}{\mathrm{f_T^2}}\mathbf{l} + \alpha\frac{1}{\mathrm{f_T^2}}\mathbf{l} \\ -\frac{\mathrm{f_V}}{\mathrm{f_T^2}}\mathbf{l} + \alpha\frac{1}{\mathrm{f_T^2}}\mathbf{l} & \frac{\mathbf{k}}{\mathrm{f_T^2}} - 2\alpha\frac{\mathrm{f_V}}{\mathrm{f_T^3}}\mathbf{l} + \alpha^2\frac{1}{\mathrm{f_T^3}}\mathbf{l} + \frac{\mathrm{f_V^2}}{\mathrm{f_T^3}}\mathbf{l} \end{pmatrix} \begin{pmatrix} \nabla\mathrm{f_V} \\ \nabla\mathrm{f_T} \end{pmatrix} \geq 0. \end{aligned} \tag{32}$$

Finally, the strong form (16, 17) can be expressed as

$$\begin{cases} \mathrm{div}(\mathbf{j}) & = \mathbf{i} \quad\quad \forall\, \mathbf{x} \in \Omega, \\ \mathbf{M} & = \bar{\mathbf{M}} \quad\quad \forall\, \mathbf{x} \in \partial_{\mathrm{D}}\Omega, \\ \bar{\mathbf{n}}^{\mathrm{T}}\,\mathbf{j} & = \bar{\mathbf{j}} \quad\quad \forall\, \mathbf{x} \in \partial_{\mathrm{N}}\Omega, \end{cases} \tag{33}$$

where $\bar{\mathbf{n}} = \begin{pmatrix} \mathbf{n} & 0 \\ 0 & \mathbf{n} \end{pmatrix}$, $\bar{\mathbf{M}} \in \mathrm{L}^2(\partial_{\mathrm{D}}\Omega) \times \mathrm{L}^{2^+}(\partial_{\mathrm{D}}\Omega)$, and $\bar{\mathbf{j}} = \begin{pmatrix} \bar{\mathrm{j}}_{\mathrm{e}} \\ \bar{\mathrm{j}}_{\mathrm{y}} \end{pmatrix}$.

3. Thermo-electrical analysis with the Discontinuous Galerkin (DG) finite element method

Let the domain $\Omega \subset \mathbb{R}^{\mathrm{d}}$ be approximated by a discretized domain $\Omega_{\mathrm{h}} \subset \mathbb{R}^{\mathrm{d}}$ such that $\Omega \approx \Omega_{\mathrm{h}} = \cup_{\mathrm{e}}\Omega^{\mathrm{e}}$, where a finite element in Ω_{h} is denoted by Ω^{e}. The boundary $\partial\Omega_{\mathrm{h}}$ is decomposed into a region of Dirichlet boundary $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$, and a region of Neumann boundary $\partial_{\mathrm{N}}\Omega_{\mathrm{h}}$. The intersecting boundary of the finite elements is denoted by $\partial_{\mathrm{I}}\Omega_{\mathrm{h}} = \cup_{\mathrm{e}}\partial\Omega^{\mathrm{e}} \setminus \partial\Omega_{\mathrm{h}}$ as shown in the Fig. 1, with $\partial_{\mathrm{N}}\Omega_{\mathrm{h}} = \cup_{\mathrm{e}}\partial_{\mathrm{N}}\Omega^{\mathrm{e}}$, $\partial_{\mathrm{D}}\Omega_{\mathrm{h}} = \cup_{\mathrm{e}}\partial_{\mathrm{D}}\Omega^{\mathrm{e}}$, $\partial\Omega_{\mathrm{h}} \cup \partial_{\mathrm{I}}\Omega_{\mathrm{h}} = \cup_{\mathrm{e}}\partial\Omega^{\mathrm{e}}$, and $\partial_{\mathrm{I}}\Omega^{\mathrm{e}} = \partial\Omega^{\mathrm{e}} \bigcap \partial_{\mathrm{I}}\Omega_{\mathrm{h}}$.

Within this finite element discretization, an interior face $(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}} = \partial\Omega^{\mathrm{e}+} \cap \partial\Omega^{\mathrm{e}-}$ is shared by elements $\Omega^{\mathrm{e}+}$ and $\Omega^{\mathrm{e}-}$, and $\mathbf{n}^-$ is the unit normal vector pointing from element $\Omega^{\mathrm{e}-}$ toward element $\Omega^{\mathrm{e}+}$, see Fig. 1. Similarly, an exterior Dirichlet edge $(\partial_{\mathrm{D}}\Omega)^{\mathrm{s}} = \partial\Omega^{\mathrm{e}} \cap \partial_{\mathrm{D}}\Omega_{\mathrm{h}}$ is the intersection between the boundary of the element Ω^{e} and the Dirichlet boundary, and $\mathbf{n}^- = \mathbf{n}$ is used to represent the outward unit normal vector. Finally $(\partial_{\mathrm{DI}}\Omega)^{\mathrm{s}}$is a face either on $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$ or on $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$, with $\sum_{\mathrm{s}}(\partial_{\mathrm{DI}}\Omega)^{\mathrm{s}} = \partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}$.

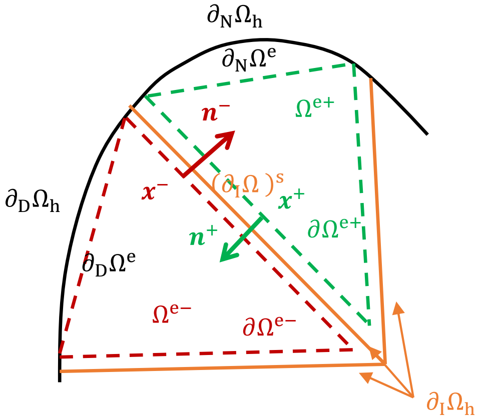


Figure 1: interface between two elements (Ω^{e+}) and (Ω^{e-})

3.1. Weak discontinuous form

The weak form of the matrix form Eq. (33) will be formulated by considering a two-field coupled problem defined in terms of the energetically conjugated pair of fluxes and fields gradients as defined in Section 2.3.

To account for the discontinuity in $\mathbf{M} = \begin{pmatrix} \mathrm{f_V} \\ \mathrm{f_T} \end{pmatrix}$, we can define the associated norm of the standard broken Sobolev space $\mathrm{W^s_r}(\Omega_\mathrm{h})$ of order s and exponent r with $1 \leq \mathrm{r} < \infty$. Starting from the Sobolev space norm and semi norm

$$\begin{cases} \parallel \mathbf{M} \parallel_{\mathrm{W^s_r}(\Omega^\mathrm{e})} & = \left(\sum_{|\alpha|\leq \mathrm{s}} \int_{\Omega_\mathrm{e}} \parallel \partial^\alpha \mathrm{f_T} \parallel^\mathrm{r} \mathrm{dx} + \sum_{|\alpha|\leq \mathrm{s}} \int_{\Omega_\mathrm{e}} \parallel \partial^\alpha \mathrm{f_V} \parallel^\mathrm{r} \mathrm{dx}\right)^{\frac{1}{\mathrm{r}}}, \\ | \mathbf{M} |_{\mathrm{W^s_r}(\Omega^\mathrm{e})} & = \left(\int_{\Omega_\mathrm{e}} \parallel \partial^\mathrm{s} \mathrm{f_T} \parallel^\mathrm{r} \mathrm{dx} + \int_{\Omega_\mathrm{e}} \parallel \partial^\mathrm{s} \mathrm{f_V} \parallel^\mathrm{r} \mathrm{dx}\right)^{\frac{1}{\mathrm{r}}}, \end{cases} \tag{34}$$

the norm and semi norm of the broken Sobolev space read

$$\begin{cases} \parallel \mathbf{M} \parallel_{\mathrm{W^s_r}(\Omega_\mathrm{h})} & = \left(\sum_\mathrm{e} \parallel \mathbf{M} \parallel^\mathrm{r}_{\mathrm{W^s_r}(\Omega^\mathrm{e})}\right)^{\frac{1}{\mathrm{r}}}, \\ | \mathbf{M} |_{\mathrm{W^s_r}(\Omega_\mathrm{h})} & = \left(\sum_\mathrm{e} | \mathbf{M} |^\mathrm{r}_{\mathrm{W^r_s}(\Omega^\mathrm{e})}\right)^{\frac{1}{\mathrm{r}}}. \end{cases} \tag{35}$$

For the case $\mathrm{r} = \infty$, the norm is defined as

$$\parallel \mathbf{M} \parallel_{\mathrm{W^s_\infty}(\Omega_\mathrm{h})} = \max_\mathrm{e} \parallel \mathbf{M} \parallel_{\mathrm{W^s_\infty}(\Omega^\mathrm{e})}, \text{ with } \parallel \mathbf{M} \parallel_{\mathrm{W^s_\infty}(\Omega^\mathrm{e})} = \max_{|\alpha|\leq \mathrm{s}} \{\parallel \partial^\alpha \mathrm{f_V} \parallel_{\mathrm{L}^\infty(\Omega^\mathrm{e})}, \parallel \partial^\alpha \mathrm{f_T} \parallel_{\mathrm{L}^\infty(\Omega^\mathrm{e})}\}. \tag{36}$$

We can define the broken Sobolev spaces for the case $\mathrm{r} = 2$ as

$$\mathrm{X_s^{(+)}} = \left\{\mathbf{M} \in \mathrm{L}^2(\Omega_\mathrm{h}) \times \mathrm{L}^{2^{(+)}}(\Omega_\mathrm{h}) \,|_{\mathbf{M}_{|\Omega^\mathrm{e}} \in \mathrm{H^s}(\Omega^\mathrm{e}) \times \mathrm{H}^{\mathrm{s}^{(+)}}(\Omega^\mathrm{e}) \;\; \forall \Omega^\mathrm{e} \in \Omega_\mathrm{h}}\right\}, \tag{37}$$

and

$$\mathbf{Y} = \left\{\nabla\mathbf{M} \in (\mathrm{L}^2(\Omega_\mathrm{h}))^3 \times (\mathrm{L}^2(\Omega_\mathrm{h}))^3 \,|_{\nabla\mathbf{M}_{|\Omega^\mathrm{e}} \in (\mathrm{H}^1(\Omega^\mathrm{e}))^3 \times (\mathrm{H}^1(\Omega^\mathrm{e}))^3 \;\; \forall \Omega^\mathrm{e} \in \Omega_\mathrm{h}}\right\}, \tag{38}$$

where, we define $\mathrm{X}_{\mathrm{s}}^{+}$ as the manifold such that $\mathrm{f}_{\mathrm{T}} > 0$, while X_{s} is the manifold for which $\mathrm{f}_{\mathrm{T}} \lesseqqgtr 0$, with $\mathrm{X}_{\mathrm{s}}^{+} \subset \mathrm{X}_{\mathrm{s}}$. For conciseness, we have used an abuse of notations in Eq. (37), in which "$\bullet^{(+)}$" holds either for "$\bullet$" or for "$\bullet^{+}$". Still for conciseness, we define $\mathrm{X}^{(+)} = \mathrm{X}_{2}^{(+)}$.

Let us derive the weak form of the governing equation (33) for electro-thermal coupling by multiplying it by a suitable test function $\delta\mathbf{M} \in \mathrm{X}_1$, performing a volume integral, and using the divergence theorem on each element Ω^{e}. This leads to state the problem as finding $\mathbf{M} \in \mathrm{X}_1^{+}$ such that

$$-\sum_{\mathrm{e}}\int_{\Omega^{\mathrm{e}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\ \mathrm{d}\Omega + \sum_{\mathrm{e}}\int_{\partial\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\ \mathrm{dS} = \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\mathbf{i}\mathrm{d}\Omega \quad \forall\delta\mathbf{M} \in \mathrm{X}_1. \tag{39}$$

The surface integral of this last equation is rewritten as

$$\sum_{\mathrm{e}}\int_{\partial\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS} = \sum_{\mathrm{e}}\int_{\partial_{\mathrm{N}}\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS} + \sum_{\mathrm{e}}\int_{\partial_{\mathrm{I}}\Omega^{\mathrm{e}}\cup\partial_{\mathrm{D}}\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\,\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS}. \tag{40}$$

The second term of the right hand side of Eq. (40) can be rewritten using

$$\begin{aligned}
&\sum_{\mathrm{e}}\int_{\partial_{\mathrm{I}}\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS} = \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left(\delta\mathbf{M}^{-^{\mathrm{T}}}\bar{\mathbf{n}}^{-^{\mathrm{T}}}\mathbf{j}^{-}(\mathbf{M},\nabla\mathbf{M}) + \delta\mathbf{M}^{+^{\mathrm{T}}}\bar{\mathbf{n}}^{+^{\mathrm{T}}}\mathbf{j}^{+}(\mathbf{M},\nabla\mathbf{M})\right)\mathrm{dS},\ \bar{\mathbf{n}}^{+} = -\bar{\mathbf{n}}^{-},\\
&\sum_{\mathrm{e}}\int_{\partial_{\mathrm{D}}\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS} = -\int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \left(-\delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{-^{\mathrm{T}}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\right)\mathrm{dS} \text{ and } \bar{\mathbf{n}}^{-} = \bar{\mathbf{n}}.
\end{aligned} \tag{41}$$

In these equations we use the superscript "$-(+)$" to refer to the value of element $\Omega^{\mathrm{e}^{-}}(\Omega^{\mathrm{e}^{+}})$. Moreover we define $\delta\mathbf{M}_{\mathbf{n}} = \begin{pmatrix} \mathbf{n}^{-} & 0 \\ 0 & \mathbf{n}^{-} \end{pmatrix}\delta\mathbf{M}$ and $\mathbf{M}_{\mathbf{n}} = \begin{pmatrix} \mathbf{n}^{-} & 0 \\ 0 & \mathbf{n}^{-} \end{pmatrix}\mathbf{M}$ for future use.

We can introduce trace operators to manipulate the numerical flux and obtain the primal formulation. On $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$, the average $\langle\bullet\rangle$ and the jump $[\![\bullet]\!]$ operators are defined as $\langle\bullet\rangle = \frac{1}{2}(\bullet^{+} + \bullet^{-})$, $[\![\ ]\!] = (\bullet^{+} - \bullet^{-})$. The definition of these two trace operators can be extended on the Dirichlet boundary $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$ as $\langle\bullet\rangle = \bullet$, $[\![\bullet]\!] = (-\bullet)$. Therefore using Eqs. (41), Eq. (39) becomes

$$\begin{aligned}
\sum_{\mathrm{e}}\int_{\Omega^{\mathrm{e}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{d}\Omega + \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\mathbf{i}\mathrm{d}\Omega = &\int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{dS}\\
&-\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}\cup\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} [\![\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})]\!]\ \mathrm{dS}.
\end{aligned} \tag{42}$$

Applying the Neumann boundary conditions specified in Eq. (33) allows this last result to be rewritten as finding $\mathbf{M} \in \mathrm{X}_1^{+}$ such that

$$\int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{j}}\ \mathrm{dS} = \int_{\Omega_{\mathrm{h}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{d}\Omega + \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\mathbf{i}\mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}\cup\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} [\![\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})]\!]\ \mathrm{dS}\ \forall\delta\mathbf{M} \in \mathrm{X}_1. \tag{43}$$

Applying the mathematical identity $[\![\mathrm{ab}]\!] = [\![\mathrm{a}]\!]\langle\mathrm{b}\rangle + [\![\mathrm{b}]\!]\langle\mathrm{a}\rangle$ on $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$, and by neglecting the second term because only consistency of the test functions needs to be enforced, then the consistent flux related to the last term of Eq. (43) reads $[\![\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}]\!]\langle\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\rangle$.

Moreover, on the one hand, due to the discontinuous nature of the trial functions in the DG weak form, the inter-element discontinuity is allowed, so the continuity of unknown variables is enforced weakly by using symmetrization and stabilization terms at the interior elements boundary interface $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$. On the other hand, the Dirichlet boundary condition (33) is also enforced in a weak sense by considering the same symmetrization and stabilization terms at the Dirichlet elements boundary interface $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$. By using the definition of the conjugated force, Eq. (29), the virtual energy flux $\delta\mathbf{j}(\mathbf{M})$ reads

$$\delta\mathbf{j}(\mathbf{M}) = \mathbf{Z}(\mathbf{M})\nabla\delta\mathbf{M}. \tag{44}$$

This last result allows formulating the symmetrization and quadratic stabilization terms so the weak form Eq. (43) becomes finding $\mathbf{M} \in \mathrm{X}_1^+$ such that:

$$\begin{aligned}
&\int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{j}}\mathrm{dS} - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \bar{\mathbf{M}}_{\mathbf{n}}^{\mathrm{T}}\left(\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right)\mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\left(\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right)\bar{\mathbf{M}}_{\mathbf{n}}\mathrm{dS} \\
&= \int_{\Omega_{\mathrm{h}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{d}\Omega + \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\mathbf{i}\mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}\cup\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\rrbracket\left\langle\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\right\rangle\mathrm{dS} \\
&+ \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \llbracket\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\rrbracket\left\langle\mathbf{Z}(\mathbf{M})\nabla\delta\mathbf{M}\right\rangle\mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\rrbracket\left\langle\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right\rangle\mathrm{dS} \\
&+ \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \llbracket\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\rrbracket\left\langle\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\mathbf{M})\right\rangle\llbracket\mathbf{M}_{\mathbf{n}}\rrbracket\mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\rrbracket\left\langle\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right\rangle\llbracket\mathbf{M}_{\mathbf{n}}\rrbracket\mathrm{dS} \quad \forall\delta\mathbf{M} \in \mathrm{X}_1,
\end{aligned} \tag{45}$$

where $\bar{\mathbf{M}}_{\mathbf{n}} = \begin{pmatrix} \mathbf{n} & 0 \\ 0 & \mathbf{n} \end{pmatrix}\bar{\mathbf{M}}$. In this equation $\mathcal{B}$ is the stability parameter which has to be sufficiently high to guarantee stability, as it will be shown in Section 4, and $\mathrm{h_s}$ is the characteristic length of the mesh, which will also be defined in Section 4.

The last two terms of the left hand side of Eq. (45) make sure that the Dirichlet boundary condition (33) is weakly enforced, as it will be shown in Section 4. The last five terms of the right hand side of Eq. (45) are the interface terms, which correspond to a SIPG method:

1. The first term ensures consistency despite the discontinuity of the test function $\delta\mathbf{M}$ between two elements, and involves the consistent numerical flux which is here the traditional average flux.
2. The second and third term achieve symmetry of the weak form and thereby also of the stiffness matrix after FE discretization. It also ensures (weakly) the continuity of solution across element boundaries and the optimal convergence rate in the L^2-norm.
3. The last two terms ensure stability, as it is well known that the discontinuous formulation of elliptic problems requires quadratic terms. The stabilization term depends on a stability parameter, which is independent of mesh size and material properties, as it will be shown in Section 4.

46- ally the weak form (45) is thus summarized as finding $\mathbf{M} \in \mathrm{X}_1^+$ such that:

$$\mathrm{a}(\mathbf{M},\delta\mathbf{M}) = \mathrm{b}(\bar{\mathbf{M}},\delta\mathbf{M}) - \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\mathbf{i}\mathrm{d}\Omega \quad \forall\delta\mathbf{M} \in \mathrm{X}_1, \tag{46}$$

with

$$\begin{aligned} \mathrm{a}(\mathbf{M},\delta\mathbf{M}) = &\int_{\Omega_\mathrm{h}} \nabla\delta\mathbf{M}^\mathrm{T}\mathbf{j}(\mathbf{M},\nabla\mathbf{M})\mathrm{d}\Omega + \int_{\partial_\mathrm{I}\Omega_\mathrm{h}\cup\partial_\mathrm{D}\Omega_\mathrm{h}} [\![\delta\mathbf{M}_\mathbf{n}^\mathrm{T}]\!] \left\langle \mathbf{j}(\mathbf{M},\nabla\mathbf{M})\right\rangle \mathrm{dS} \\ &+ \int_{\partial_\mathrm{I}\Omega_\mathrm{h}} [\![\mathbf{M}_\mathbf{n}^\mathrm{T}]\!] \left\langle \mathbf{Z}(\mathbf{M})\nabla\delta\mathbf{M}\right\rangle \mathrm{dS} + \int_{\partial_\mathrm{D}\Omega_\mathrm{h}} [\![\mathbf{M}_\mathbf{n}^\mathrm{T}]\!] \left\langle \mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right\rangle \mathrm{dS} \\ &+ \int_{\partial_\mathrm{I}\Omega_\mathrm{h}} [\![\delta\mathbf{M}_\mathbf{n}^\mathrm{T}]\!] \left\langle \frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\mathbf{M})\right\rangle [\![\mathbf{M}_\mathbf{n}]\!] \mathrm{dS} + \int_{\partial_\mathrm{D}\Omega_\mathrm{h}} [\![\delta\mathbf{M}_\mathbf{n}^\mathrm{T}]\!] \left\langle \frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right\rangle [\![\mathbf{M}_\mathbf{n}]\!] \mathrm{dS}, \end{aligned} \tag{47}$$

and

$$\mathrm{b}(\bar{\mathbf{M}},\delta\mathbf{M}) = \int_{\partial_\mathrm{N}\Omega_\mathrm{h}} \delta\mathbf{M}^\mathrm{T}\bar{\mathbf{j}}\mathrm{dS} - \int_{\partial_\mathrm{D}\Omega_\mathrm{h}} \bar{\mathbf{M}}_\mathbf{n}^\mathrm{T}\left(\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right)\mathrm{dS} + \int_{\partial_\mathrm{D}\Omega_\mathrm{h}} \delta\mathbf{M}_\mathbf{n}^\mathrm{T}\left(\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right)\bar{\mathbf{M}}_\mathbf{n}\mathrm{dS}. \tag{48}$$

It should be noted that the test functions in the previous equations of the weak formulation belong to X_1, however for the numerical analysis, we will need to be in X_2.

3.2. Finite element discretization

In the finite element method, the trial function $\mathbf{M}$ is approximated by $\mathbf{M}_\mathrm{h}$, which is defined over a finite element Ω^e using the interpolation concepts in terms of the standard shape function $\mathrm{N^a}\in\mathbb{R}$ at node a, see [22], yielding

$$\mathbf{M}_\mathrm{h} = \mathbf{N}^\mathrm{a}\,\mathbf{M}^\mathrm{a}\,, \nabla\mathbf{M}_\mathrm{h} = \nabla\mathbf{N}^\mathrm{a}\,\mathbf{M}^\mathrm{a}, \tag{49}$$

where $\mathbf{M}^\mathrm{a}$ denotes the nodal value of $\mathbf{M}_\mathrm{h}$ at node a, $\mathbf{N}^\mathrm{a} = \begin{pmatrix} \mathrm{N^a} & 0 \\ 0 & \mathrm{N^a}\end{pmatrix}$ is a matrix of the shape functions and $\nabla\mathbf{N}^\mathrm{a} = \begin{pmatrix} \nabla\mathrm{N^a} & 0 \\ 0 & \nabla\mathrm{N^a}\end{pmatrix}$ is a matrix of the shape function gradients. In order to obtain a Galerkin formulation, the test functions are approximated using the same interpolation, i.e.

$$\delta\mathbf{M}_\mathrm{h} = \mathbf{N}^\mathrm{a}\,\delta\mathbf{M}^\mathrm{a}\,, \quad \nabla\delta\mathbf{M}_\mathrm{h} = \nabla\mathbf{N}^\mathrm{a}\,\delta\mathbf{M}^\mathrm{a}. \tag{50}$$

In this work, we assume a constant mesh size on the elements, but the theory can be generalized by considering bounded element sizes such as in [18]. We assume the discretization is shaped with a regular mesh of size $\mathrm{h_s}$ defined as $\frac{|\Omega^\mathrm{e}|}{|\partial\Omega^\mathrm{e}|}$. We also assume shape regularity of Ω_h so that there exist constants c_1, c_2, c_3, and c_4, independent of $\mathrm{h_s}$, such that

$$\begin{aligned} &\mathrm{c}_1\,\mathrm{diam}\left((\partial_\mathrm{DI}\Omega)^\mathrm{s}\right) \le \mathrm{h_s} \le \mathrm{c}_2\,\mathrm{diam}\left((\partial_\mathrm{DI}\Omega)^\mathrm{s}\right), \text{ and} \\ &\mathrm{c}_3\,\mathrm{diam}\left(\Omega^\mathrm{e}\right) \le \mathrm{h_s} \le \mathrm{c}_4\,\mathrm{diam}\left(\Omega^\mathrm{e}\right). \end{aligned} \tag{51}$$

The finite discontinuous polynomial approximation $\mathbf{M}_\mathrm{h} = \begin{pmatrix} \mathrm{f}_{\mathrm{V_h}} \\ \mathrm{f}_{\mathrm{T_h}}\end{pmatrix} \in \mathrm{X}^{\mathrm{k}^+}$ of the solution is thus defined in the space

$$\mathrm{X}^{\mathrm{k}^{(+)}} = \left\{ \mathbf{M}_\mathrm{h} \in \mathrm{L}^2(\Omega_\mathrm{h}) \times \mathrm{L}^{2^{(+)}}(\Omega_\mathrm{h}) \mid_{\mathbf{M}_\mathrm{h}|_{\Omega^\mathrm{e}} \in \mathbb{P}^\mathrm{k}(\Omega^\mathrm{e})\times\mathbb{P}^{\mathrm{k}^{(+)}}(\Omega^\mathrm{e})\ \forall\Omega^\mathrm{e}\in\Omega_\mathrm{h}} \right\}, \tag{52}$$

where $\mathbb{P}^{\mathrm{k}}(\Omega^{\mathrm{e}})$ is the space of polynomial functions of order up to k and $\mathbb{P}^{\mathrm{k}^+}$ means that the polynomial approximation remains positive. In the numerical framework, the positive nature of the polynomial approximation is not directly enforced, but results naturally from the resolution of the well-posed finite element problem.

Using these definitions, the problem becomes finding $\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}^+}$ such that

$$\mathrm{a}(\mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) = \mathrm{b}(\bar{\mathbf{M}}, \delta\mathbf{M}_{\mathrm{h}}) - \int_{\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathrm{h}}^{\mathrm{T}} \mathbf{i} \mathrm{d}\Omega \quad \forall \delta\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}}. \tag{53}$$

The set of Eqs. (53) can be rewritten under the form:

$$\mathbf{F}^{\mathrm{a}}_{\mathrm{ext}}\left(\mathbf{M}^{\mathrm{b}}\right) = \mathbf{F}^{\mathrm{a}}_{\mathrm{int}}\left(\mathbf{M}^{\mathrm{b}}\right) + \mathbf{F}^{\mathrm{a}}_{\mathrm{I}}\left(\mathbf{M}^{\mathrm{b}}\right), \tag{54}$$

where $\mathbf{M}^{\mathrm{b}}$ is the vector of the unknown fields at node b. The nonlinear Eqs. (54) are solved using the Newton Raphson scheme. To this end, the forces are written in a residual form. The predictor, iteration 0, reads $\mathbf{M}^{\mathrm{b}} = \mathbf{M}^{\mathrm{b}0}$, the residual at iteration i reads

$$\mathbf{F}^{\mathrm{a}}_{\mathrm{ext}}\left(\mathbf{M}^{\mathrm{bi}}\right) - \mathbf{F}^{\mathrm{a}}_{\mathrm{int}}\left(\mathbf{M}^{\mathrm{bi}}\right) - \mathbf{F}^{\mathrm{a}}_{\mathrm{I}}\left(\mathbf{M}^{\mathrm{bi}}\right) = \mathbf{R}, \tag{55}$$

and at iteration i, the first order Taylor development yields the system to be solved, i.e.

$$\mathbf{0} = \mathbf{R}^{\mathrm{a}} + \left(\frac{\partial \mathbf{F}^{\mathrm{a}}_{\mathrm{ext}}}{\partial \mathbf{M}^{\mathrm{b}}} - \frac{\partial \mathbf{F}^{\mathrm{a}}_{\mathrm{int}}}{\partial \mathbf{M}^{\mathrm{b}}} - \frac{\partial \mathbf{F}^{\mathrm{a}}_{\mathrm{I}}}{\partial \mathbf{M}^{\mathrm{b}}}\right)\left(\mathbf{M}^{\mathrm{b}} - \mathbf{M}^{\mathrm{bi}}\right). \tag{56}$$

In this last equation

$$\begin{aligned} \mathbf{F}^{\mathrm{a}}_{\mathrm{ext}} = & \sum_{\mathrm{e}} \int_{(\partial_{\mathrm{N}}\Omega)^{\mathrm{s}}} \mathbf{N}^{\mathrm{a}} \bar{\mathbf{j}} \mathrm{dS} - \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{D}}\Omega)^{\mathrm{s}}} \nabla \mathbf{N}^{\mathrm{a}^{\mathrm{T}}} \mathbf{Z}(\bar{\mathbf{M}}) \bar{\mathbf{M}}_{\mathbf{n}} \mathrm{dS} \\ & + \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{D}}\Omega)^{\mathrm{s}}} \mathbf{N}^{\mathrm{a}} \bar{\mathbf{n}}^{\mathrm{T}} \left(\mathbf{Z}(\bar{\mathbf{M}}) \frac{\mathcal{B}}{\mathrm{h_s}}\right) \bar{\mathbf{M}}_{\mathbf{n}} \mathrm{dS}, \end{aligned} \tag{57}$$

$$\mathbf{F}^{\mathrm{a}}_{\mathrm{int}} = \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \nabla \mathbf{N}^{\mathrm{a}^{\mathrm{T}}} \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}}) \mathrm{d}\Omega + \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \mathbf{N}^{\mathrm{a}} \mathbf{i} \mathrm{d}\Omega, \tag{58}$$

$$\mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I}} = \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I1}} + \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I2}} + \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I3}}, \tag{59}$$

with the three contributions to the interface forces on $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$[1]

$$\mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I1}} = \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm \mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-^{\mathrm{T}}} \left\langle \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}}) \right\rangle \mathrm{dS}, \tag{60}$$

$$\mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I2}} = \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\nabla \mathbf{N}^{\mathrm{a}\pm^{\mathrm{T}}} \mathbf{Z}^{\pm}(\mathbf{M}_{\mathrm{h}})\right) [\![\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!] \mathrm{dS}, \tag{61}$$

[1] The contributions on $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$ can be directly deduced by removing the factor (1/2) accordingly to the definition of the average flux on the Dirichlet boundary and by substituting $\mathbf{Z}(\mathbf{M}_{\mathrm{h}})$ by $\mathbf{Z}(\bar{\mathbf{M}})$.

$$\mathbf{F}_{\mathrm{I3}}^{\mathrm{a}\pm} = \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm\mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-\mathrm{T}} \left\langle \mathbf{Z}(\mathbf{M}_{\mathrm{h}}) \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \right\rangle [\![\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!] \,\mathrm{dS}. \tag{62}$$

In these equations the symbol "±" refers to the node $\mathrm{e}^{\pm}$ (with "+" for node a^{+} and "-" for node a^{-}).

This system is solved by means of a Newton-Raphson method with the stiffness matrix computed in Appendix A, where the iterations continue until the convergence to a specified tolerance is achieved.

4. Numerical properties for DG method

In this section, the numerical properties of the weak formulation stated by Eq. (46) are studied in steady state conditions ($\mathbf{i} = 0$), and under the assumption that $d = 2$. It is demonstrated that the framework satisfies two fundamental properties of a numerical method: consistency and stability. Moreover we show that the method possesses the optimal convergence rate with respect to the mesh size.

4.1. Discontinuous space and finite element properties

In this part, we will assume that $\partial_{\mathrm{D}}\Omega_{\mathrm{h}} = \partial\Omega_{\mathrm{h}}$. This assumption is not restrictive but simplifies the demonstrations. Let us also define the norms, which appear in the analysis of the interior penalty, for $\mathbf{M} \in \mathrm{X}_1$

$$||| \mathbf{M} |||_*^2 = \sum_{\mathrm{e}} \|\nabla\mathbf{M}\|^2_{\mathrm{L}^2(\Omega^{\mathrm{e}})} + \sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}}^{-1} \| [\![\mathbf{M}_{\mathbf{n}}]\!] \|^2_{\mathrm{L}^2(\partial\Omega^{\mathrm{e}})}, \tag{63}$$

$$||| \mathbf{M} |||^2 = \sum_{\mathrm{e}} \|\mathbf{M}\|^2_{\mathrm{H}^1(\Omega^{\mathrm{e}})} + \sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}}^{-1} \| [\![\mathbf{M}_{\mathbf{n}}]\!] \|^2_{\mathrm{L}^2(\partial\Omega^{\mathrm{e}})}, \tag{64}$$

and

$$||| \mathbf{M} |||_1^2 = \sum_{\mathrm{e}} \|\mathbf{M}\|^2_{\mathrm{H}^1(\Omega^{\mathrm{e}})} + \sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}} \parallel \mathbf{M} \parallel^2_{\mathrm{H}^1(\partial\Omega^{\mathrm{e}})} + \sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}}^{-1} \| [\![\mathbf{M}_{\mathbf{n}}]\!] \|^2_{\mathrm{L}^2(\partial\Omega^{\mathrm{e}})}, \tag{65}$$

where $\partial\Omega^{\mathrm{e}} = \partial_{\mathrm{I}}\Omega^{\mathrm{e}} \cup \partial_{\mathrm{D}}\Omega^{\mathrm{e}}$. Eqs. (63-65) define norms as $|||\mathbf{M}|||_* = 0$ only when $\mathbf{M} = \mathrm{cst}$ on Ω_{h} and is equal to 0 on $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$.

Lemma 4.1 (Relation between energy norms on the finite element space)**.** *From [14], for $\boldsymbol{M}_h \in X^k$, there exists a positive constant C^k, depending on k, such that*

$$||| \boldsymbol{M}_h |||_1 \leq C^k \, ||| \boldsymbol{M}_h ||| \,. \tag{66}$$

where C^k is a positive constant, independent of the mesh size.

The demonstration directly follows by bounding the extra terms $\sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}} \parallel \mathbf{M} \parallel^2_{\mathrm{L}^2(\partial\Omega^{\mathrm{e}})}$ and $\sum_{\mathrm{e}} \mathrm{h}_{\mathrm{s}} \parallel \nabla\mathbf{M} \parallel^2_{\mathrm{L}^2(\partial\Omega^{\mathrm{e}})}$ of the norm defined by Eq. (65), in comparison to the norm defined by Eq. (64), using successively the trace inequality, Eq. (B.5), and the inverse inequality, Eq. (B.9), for the first term, and the trace inequality on the finite element space, Eq. (B.6), for the second term.

4.2. Consistency

To prove the consistency of the method, the exact solution $\mathbf{M}^{\mathrm{e}} \in \mathrm{H}^2(\Omega) \times \mathrm{H}^{2^+}(\Omega)$ of the problem stated by Eqs. (33) is considered. This implies $[\![\mathbf{M}^{\mathrm{e}}]\!] = 0$, $\langle \mathbf{j} \rangle = \mathbf{j}$ on $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$, and $[\![\mathbf{M}^{\mathrm{e}}]\!] = -\bar{\mathbf{M}} = -\mathbf{M}^{\mathrm{e}}$, $\langle \mathbf{j} \rangle = \mathbf{j} = \mathbf{Z}(\mathbf{M}^{\mathrm{e}})\nabla\mathbf{M}^{\mathrm{e}}$, and $\mathbf{Z}(\mathbf{M}) = \mathbf{Z}(\bar{\mathbf{M}}) = \mathbf{Z}(\mathbf{M}^{\mathrm{e}})$ on $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$. Therefore, Eq. (46) becomes:

$$\begin{aligned}
&\int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{j}}\mathrm{dS} - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \bar{\mathbf{M}}_{\mathbf{n}}^{\mathrm{T}} \left(\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right) \mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}} \left(\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right) \bar{\mathbf{M}}_{\mathbf{n}}\mathrm{dS} \\
&= \int_{\Omega_{\mathrm{h}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} [\![\delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}]\!]\, \mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{dS} - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{dS} \\
&- \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} {\mathbf{M}_{\mathbf{n}}^{\mathrm{e}}}^{\mathrm{T}}\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}} \frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\mathbf{M}_{\mathbf{n}}^{\mathrm{e}}\mathrm{dS} \quad \forall\delta\mathbf{M} \in \mathrm{X}\,.
\end{aligned} \tag{67}$$

Integrating the first term of the right hand side by parts leads to

$$\sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \nabla\delta\mathbf{M}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{d}\Omega = -\sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\nabla\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{d}\Omega + \sum_{\mathrm{e}} \int_{\partial\Omega^{\mathrm{e}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{dS}, \tag{68}$$

and Eq.(67) becomes

$$\begin{aligned}
&\int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}^{\mathrm{T}}\bar{\mathbf{j}}\mathrm{dS} - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \bar{\mathbf{M}}_{\mathbf{n}}^{\mathrm{T}} \left(\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\right) \mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}} \left(\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\right) \bar{\mathbf{M}}_{\mathbf{n}}\mathrm{dS} = \\
&\qquad - \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \delta\mathbf{M}^{\mathrm{T}}\nabla\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{d}\Omega + \int_{\partial_{\mathrm{N}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{dS} \\
&\qquad - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} {\mathbf{M}_{\mathbf{n}}^{\mathrm{e}}}^{\mathrm{T}}\mathbf{Z}(\bar{\mathbf{M}})\nabla\delta\mathbf{M}\mathrm{dS} + \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta\mathbf{M}_{\mathbf{n}}^{\mathrm{T}} \frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\bar{\mathbf{M}})\mathbf{M}_{\mathbf{n}}^{\mathrm{e}}\mathrm{dS} \quad \forall\delta\mathbf{M} \in \mathrm{X}.
\end{aligned} \tag{69}$$

The arbitrary nature of the test functions leads to recover the set of conservation laws and the boundary conditions stated by Eqs. (33).

4.3. Stability of the DG formulation

The demonstration of the stability follows closely the approach developed by [8, 14, 17, 18] for linear and nonlinear elliptic problems. Since the problem is herein coupled, and as the elliptic operator is different, we report the modified main steps of the demonstrations that were initially developed in [17, 18] for $\mathrm{d} = 2$. The main idea to prove the solution uniqueness and to establish the prior error estimates is to linearize the problem by reformulating the nonlinear problem in a fixed point form which is the solution of the linearized problem as proposed in [17, 18, 23].

Starting from the definition of matrix $\mathbf{Z}(\mathbf{M})$, Eq. (31), which is a symmetric and positive definite matrix, as we have proved in Section 2.2, let us define the minimum and maximum eigenvalues of the matrix $\mathbf{Z}(\mathbf{M})$ as $\lambda(\mathbf{M})$ and $\Lambda(\mathbf{M})$, then for all $\xi \in \mathbb{R}_0^{2\mathrm{d}}$

$$0 < \lambda(\mathbf{M})|\xi|^2 \leq \xi_{\mathrm{i}}\mathbf{Z}^{\mathrm{ij}}(\mathbf{M})\xi_{\mathrm{j}} \leq \Lambda(\mathbf{M})|\xi|^2. \tag{70}$$

Also by assuming that $\parallel \mathbf{M} \parallel_{\mathrm{W}^1_\infty} \leq \alpha$, then there is a positive constant C_α such that

$$0 < \mathrm{C}_\alpha < \lambda(\mathbf{M}). \tag{71}$$

In the subsequent analysis, we use the following integral form of the Taylor's expansions of $\mathbf{j}$, defined in Eq. (30), for $(\mathbf{V}, \nabla\mathbf{P}) \in \mathrm{X}^{+} \times \mathbf{Y}$ in terms of $(\mathbf{M}, \nabla\mathbf{M}) \in \mathrm{X}^{+} \times \mathbf{Y}$

$$\begin{aligned} \mathbf{j}(\mathbf{V}, \nabla\mathbf{P}) - \mathbf{j}(\mathbf{M}, \nabla\mathbf{M}) &= -\mathbf{j}_{\mathbf{M}}(\mathbf{M}, \nabla\mathbf{M})(\mathbf{M} - \mathbf{V}) - \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M})(\nabla\mathbf{M} - \nabla\mathbf{P}) + \bar{\mathbf{R}}_{\mathbf{j}}(\mathbf{M} - \mathbf{V}, \nabla\mathbf{M} - \nabla\mathbf{P}) \\ &= -\bar{\mathbf{j}}_{\mathbf{M}}(\mathbf{M}, \nabla\mathbf{M})(\mathbf{M} - \mathbf{V}) - \bar{\mathbf{j}}_{\nabla\mathbf{M}}(\mathbf{M})(\nabla\mathbf{M} - \nabla\mathbf{P}), \end{aligned} \tag{72}$$

where $\mathbf{j}_{\mathbf{M}}$ is the partial derivative of $\mathbf{j}$ with respect to $\mathbf{M}$, $\mathbf{j}_{\nabla\mathbf{M}}$ is the partial derivative of $\mathbf{j}$ with respect to $\nabla\mathbf{M}$ expressed in the matrix form, and where $\bar{\mathbf{j}}_{\mathbf{M}}$, $\bar{\mathbf{j}}_{\nabla\mathbf{M}}$, and $\bar{\mathbf{R}}_{\mathbf{j}}$ are the remainder terms. With $\mathbf{V}^{\mathrm{t}} = \mathbf{M} + \mathrm{t}(\mathbf{V} - \mathbf{M})$, $\nabla\mathbf{P}^{\mathrm{t}} = \nabla\mathbf{M} + \mathrm{t}(\nabla\mathbf{P} - \nabla\mathbf{M})$, we have

$$\bar{\mathbf{j}}_{\mathbf{M}}(\mathbf{M}, \nabla\mathbf{M}) = \int_0^1 \mathbf{j}_{\mathbf{M}}(\mathbf{V}^{\mathrm{t}}, \nabla\mathbf{P}^{\mathrm{t}})\mathrm{dt}, \quad \bar{\mathbf{j}}_{\nabla\mathbf{M}}(\mathbf{M}, \nabla\mathbf{M}) = \int_0^1 \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{V}^{\mathrm{t}}, \nabla\mathbf{P}^{\mathrm{t}})\mathrm{dt}, \tag{73}$$

$$\bar{\mathrm{R}}_{\mathbf{j}}(\mathbf{M} - \mathbf{V}, \nabla\mathbf{M} - \nabla\mathbf{P}) = (\mathbf{M} - \mathbf{V})^{\mathrm{T}}\bar{\mathbf{j}}_{\mathbf{M}\mathbf{M}}(\mathbf{V}, \nabla\mathbf{P})(\mathbf{M} - \mathbf{V}) + 2(\mathbf{M} - \mathbf{V})^{\mathrm{T}}\bar{\mathbf{j}}_{\mathbf{M}\nabla\mathbf{M}}(\mathbf{V}, \nabla\mathbf{P})(\nabla\mathbf{M} - \nabla\mathbf{P}), \tag{74}$$

and

$$\begin{aligned} \bar{\mathbf{j}}_{\mathbf{M}\mathbf{M}}(\mathbf{V}, \nabla\mathbf{P}) &= \int_0^1 (1 - \mathrm{t})\mathbf{j}_{\mathbf{M}\mathbf{M}}(\mathbf{V}^{\mathrm{t}}, \nabla\mathbf{P}^{\mathrm{t}})\mathrm{dt}, \\ \bar{\mathbf{j}}_{\mathbf{M}\nabla\mathbf{M}}(\mathbf{V}, \nabla\mathbf{P}) &= \int_0^1 (1 - \mathrm{t})\mathbf{j}_{\mathbf{M}\nabla\mathbf{M}}(\mathbf{V}^{\mathrm{t}}, \nabla\mathbf{P}^{\mathrm{t}})\mathrm{dt}\,. \end{aligned} \tag{75}$$

Using the definition Eq. (30) of $\mathbf{j}$, we have $\mathbf{j}_{\mathbf{M}} = \frac{\partial\mathbf{Z}}{\partial\mathbf{M}}\nabla\mathbf{M}$, $\mathbf{j}_{\nabla\mathbf{M}} = \mathbf{Z}$, $\mathbf{j}_{\mathbf{M}\mathbf{M}} = \frac{\partial^2\mathbf{Z}}{\partial\mathbf{M}^2}\nabla\mathbf{M}$, $\mathbf{j}_{\mathbf{M}\nabla\mathbf{M}} = \mathbf{j}_{\nabla\mathbf{M}\mathbf{M}} = \frac{\partial\mathbf{Z}}{\partial\mathbf{M}}$, $\mathbf{j}_{\nabla\mathbf{M}\nabla\mathbf{M}} = 0$. If $\mathrm{f_T} \geq \mathrm{f_{T0}} > 0$, then $\bar{\mathbf{j}}_{\mathbf{M}}, \bar{\mathbf{j}}_{\mathbf{M}\mathbf{M}} \in \mathbf{L}^{\infty}\left(\Omega \times \mathbb{R} \times \mathbb{R}_0^{+} \times \mathbb{R}^{\mathrm{d}} \times \mathbb{R}^{\mathrm{d}}\right)$ and $\bar{\mathbf{j}}_{\nabla\mathbf{M}}, \bar{\mathbf{j}}_{\mathbf{M}\nabla\mathbf{M}}, \bar{\mathbf{j}}_{\nabla\mathbf{M}\mathbf{M}} \in \mathbf{L}^{\infty}\left(\Omega \times \mathbb{R} \times \mathbb{R}_0^{+}\right)$. Since $\mathbf{j}$ is a twice continuously differential function with all the derivatives through the second order, which can be shown to be locally bounded in a ball around $\mathbf{M} \in \mathbb{R} \times \mathbb{R}_0^{+}$ following the argumentation of [17, 18] for $d = 2$, we denote by $\mathrm{C_y}$

$$\begin{aligned} \mathrm{C_y} = \max\Big\{ &\parallel \mathbf{j} \parallel_{\mathrm{W}^2_{\infty}(\Omega\times\mathbb{R}\times\mathbb{R}_0^{+}\times\mathbb{R}^{\mathrm{d}}\times\mathbb{R}^{\mathrm{d}})}, \parallel \bar{\mathbf{j}}_{\mathbf{M}}\,,\ \bar{\mathbf{j}}_{\mathbf{M}\mathbf{M}} \parallel_{\mathrm{L}^{\infty}(\Omega\times\mathbb{R}\times\mathbb{R}_0^{+}\times\mathbb{R}^{\mathrm{d}}\times\mathbb{R}^{\mathrm{d}})}, \\ &\parallel \bar{\mathbf{j}}_{\nabla\mathbf{M}}, \ \bar{\mathbf{j}}_{\mathbf{M}\nabla\mathbf{M}}, \ \bar{\mathbf{j}}_{\nabla\mathbf{M}\mathbf{M}} \parallel_{\mathrm{L}^{\infty}(\Omega\times\mathbb{R}\times\mathbb{R}_0^{+})} \Big\}\,. \end{aligned} \tag{76}$$

We can now study the weak form defined by Eq. (46) under the assumptions $\mathbf{i} = 0$ and $\bar{\mathbf{j}}$ independent of $\mathbf{M}$. The problem thus reads as finding $\mathbf{M} \in \mathrm{X}^{+}$ such that

$$\mathrm{a}(\mathbf{M}, \delta\mathbf{M}) = \mathrm{b}(\bar{\mathbf{M}}, \delta\mathbf{M}) \ \ \forall\delta\mathbf{M} \in \mathrm{X}, \tag{77}$$

with $\mathrm{a}(\mathbf{M}, \delta\mathbf{M})$ defined by Eq. (47) and $\mathrm{b}(\bar{\mathbf{M}}, \delta\mathbf{M})$ by Eq. (48).

4.3.1. Derivation of the non-self-adjoint linear elliptic problem

Let us define $\mathbf{M}^{\mathrm{e}} \in \mathrm{H}^2(\Omega) \times \mathrm{H}^{2^{+}}(\Omega)$ the solution of the strong form stated by Eq. (33). Thus as $[\![\mathbf{M}^{\mathrm{e}}]\!] = 0$ on $\partial_{\mathrm{I}}\Omega^{\mathrm{e}}$ and as $[\![\mathbf{M}^{\mathrm{e}}]\!] = -\mathbf{M}^{\mathrm{e}} = -\bar{\mathbf{M}}$ on $\partial_{\mathrm{D}}\Omega^{\mathrm{e}}$, we have

$$\begin{aligned} \mathrm{a}(\mathbf{M}^{\mathrm{e}}, \delta\mathbf{M}^{\mathrm{e}}) = &\int_{\Omega_{\mathrm{h}}} \nabla\delta{\mathbf{M}^{\mathrm{e}}}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left[\!\!\left[\delta{\mathbf{M}^{\mathrm{e}}_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right] \langle\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\rangle\,\mathrm{dS} \\ &- \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta{\mathbf{M}^{\mathrm{e}}_{\mathbf{n}}}^{\mathrm{T}}\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}})\mathrm{dS} - \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \bar{\mathbf{M}}_{\mathbf{n}}^{\mathrm{T}}\mathbf{Z}(\mathbf{M}^{\mathrm{e}})\nabla\delta\mathbf{M}^{\mathrm{e}}\mathrm{dS} \\ &+ \int_{\partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \delta{\mathbf{M}^{\mathrm{e}}_{\mathbf{n}}}^{\mathrm{T}}\frac{\mathcal{B}}{\mathrm{h_s}}\mathbf{Z}(\mathbf{M}^{\mathrm{e}})\bar{\mathbf{M}}_{\mathbf{n}}\mathrm{dS} = \mathrm{b}(\bar{\mathbf{M}}, \delta\mathbf{M}^{\mathrm{e}}) \ \ \forall\delta\mathbf{M}^{\mathrm{e}} \in \mathrm{X}, \end{aligned} \tag{78}$$

as the weak form stated by Eq. (46) is consistent, see Section 4.2. Using the weak formulation (77), we state the Discontinuous Galerkin finite element method for the problem as finding $\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}^+}$, such that

$$\mathrm{a}(\mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) = \mathrm{b}(\bar{\mathbf{M}}, \delta\mathbf{M}_{\mathrm{h}}) \ \ \forall \delta\, \mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}} \subset \mathrm{X}. \tag{79}$$

Therefore, using $\delta\mathbf{M}^{\mathrm{e}} = \delta\mathbf{M}_{\mathrm{h}}$ in Eq. (78), subtracting it from the DG discretization (79), then adding and subtracting successively $\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \langle \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) \nabla\delta\mathbf{M}_{\mathrm{h}} \rangle \, \mathrm{dS}$ and $\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \left\langle \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) \right\rangle \llbracket \delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}} \rrbracket \, \mathrm{dS}$, yields

$$\begin{aligned}
0 = \mathrm{a}(\mathbf{M}^{\mathrm{e}}, \delta\mathbf{M}_{\mathrm{h}}) - \mathrm{a}(\mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) = & \int_{\Omega_{\mathrm{h}}} \nabla\delta\mathbf{M}_{\mathrm{h}}^{\mathrm{T}} \left(\mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}}) - \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla\mathbf{M}_{\mathrm{h}})\right) \mathrm{d}\Omega \\
& + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket \delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \rrbracket \langle \mathbf{j}(\mathbf{M}^{\mathrm{e}}, \nabla\mathbf{M}^{\mathrm{e}}) - \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla\mathbf{M}_{\mathrm{h}}) \rangle \, \mathrm{dS} \\
& + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \langle \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) \nabla\delta\mathbf{M}_{\mathrm{h}} \rangle \, \mathrm{dS} \\
& - \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \langle \left(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) - \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}_{\mathrm{h}})\right) \nabla\delta\mathbf{M}_{\mathrm{h}} \rangle \, \mathrm{dS} \\
& - \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \left\langle \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \left(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) - \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}_{\mathrm{h}})\right) \right\rangle \llbracket \delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}} \rrbracket \, \mathrm{dS} \\
& + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \left\llbracket \mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}} \right\rrbracket \left\langle \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) \right\rangle \llbracket \delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}} \rrbracket \, \mathrm{dS} \ \ \forall \delta\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}}.
\end{aligned} \tag{80}$$

Using the Taylor series defined in Eq. (72) to rewrite the first two terms, the set of Eqs. (80) becomes finding $\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}^+}$ such that:

$$\mathcal{A}(\mathbf{M}^{\mathrm{e}}; \mathbf{M}^{\mathrm{e}} - \mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) + \mathcal{B}(\mathbf{M}^{\mathrm{e}}; \mathbf{M}^{\mathrm{e}} - \mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) = \mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{M}_{\mathrm{h}}; \delta\mathbf{M}_{\mathrm{h}}) \quad \forall \delta\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}}. \tag{81}$$

In this last equation, for given $\boldsymbol{\psi} \in \mathrm{X}^+$, $\boldsymbol{\omega} \in \mathrm{X}$ and $\delta\boldsymbol{\omega} \in \mathrm{X}$, we have defined the following forms:

$$\begin{aligned}
\mathcal{A}(\boldsymbol{\psi}; \boldsymbol{\omega}, \delta\boldsymbol{\omega}) = & \int_{\Omega_{\mathrm{h}}} \nabla\delta\boldsymbol{\omega}^{\mathrm{T}} \mathbf{j}_{\nabla\boldsymbol{\psi}}(\boldsymbol{\psi}) \nabla\boldsymbol{\omega} \mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket \delta\boldsymbol{\omega}_{\mathbf{n}}^{\mathrm{T}} \rrbracket \langle \mathbf{j}_{\nabla\boldsymbol{\psi}}(\boldsymbol{\psi}) \nabla\boldsymbol{\omega} \rangle \, \mathrm{dS} \\
& + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket \boldsymbol{\omega}_{\mathbf{n}}^{\mathrm{T}} \rrbracket \langle \mathbf{j}_{\nabla\boldsymbol{\psi}}(\boldsymbol{\psi}) \nabla\delta\boldsymbol{\omega} \rangle \, \mathrm{dS} + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket \boldsymbol{\omega}_{\mathbf{n}}^{\mathrm{T}} \rrbracket \left\langle \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \mathbf{j}_{\nabla\boldsymbol{\psi}}(\boldsymbol{\psi}) \right\rangle \llbracket \delta\boldsymbol{\omega}_{\mathbf{n}} \rrbracket \, \mathrm{dS},
\end{aligned} \tag{82}$$

$$\mathcal{B}(\boldsymbol{\psi}; \boldsymbol{\omega}, \delta\boldsymbol{\omega}) = \int_{\Omega_{\mathrm{h}}} \nabla\delta\boldsymbol{\omega}^{\mathrm{T}} \left(\mathbf{j}_{\boldsymbol{\psi}}(\boldsymbol{\psi}, \nabla\boldsymbol{\psi})\boldsymbol{\omega}\right) \mathrm{d}\Omega + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} \llbracket \delta\boldsymbol{\omega}_{\mathbf{n}}^{\mathrm{T}} \rrbracket \langle \mathbf{j}_{\boldsymbol{\psi}}(\boldsymbol{\psi}, \nabla\boldsymbol{\psi})\boldsymbol{\omega} \rangle \, \mathrm{dS}. \tag{83}$$

For fixed $\boldsymbol{\psi}$, the form $\mathcal{A}(\boldsymbol{\psi}; ., .)$ and the form $\mathcal{B}(\boldsymbol{\psi}; ., .)$ are bi-linear. The nonlinearilty of Eq. (81) is thus gathered in the term $\mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{M}_{\mathrm{h}}; \delta\mathbf{M}_{\mathrm{h}})$ which reads

$$\begin{aligned}\mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{M}_{\mathrm{h}}; \delta\mathbf{M}_{\mathrm{h}}) = & \int_{\Omega_{\mathrm{h}}} \nabla\delta\mathbf{M}_{\mathrm{h}}^{\mathrm{T}} (\bar{\mathbf{R}}_{\mathbf{j}}(\mathbf{M}^{\mathrm{e}} - \mathbf{M}_{\mathrm{h}}, \nabla\mathbf{M}^{\mathrm{e}} - \nabla\mathbf{M}_{\mathrm{h}}))\mathrm{d}\Omega \\ & + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}} \cup \partial_{\mathrm{D}}\Omega_{\mathrm{h}}} [\![\delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}}]\!] \left\langle \bar{\mathbf{R}}_{\mathbf{j}}(\mathbf{M}^{\mathrm{e}} - \mathbf{M}_{\mathrm{h}}, \nabla\mathbf{M}^{\mathrm{e}} - \nabla\mathbf{M}_{\mathrm{h}}) \right\rangle \mathrm{dS} \\ & + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left[\!\!\left[\mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right] \left\langle (\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) - \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}_{\mathrm{h}})) \nabla\delta\mathbf{M}_{\mathrm{h}} \right\rangle \mathrm{dS} \\ & + \int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}} \left[\!\!\left[\mathbf{M}_{\mathbf{n}}^{\mathrm{e}^{\mathrm{T}}} - \mathbf{M}_{\mathrm{h}_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right] \left\langle \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} (\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}}) - \mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}_{\mathrm{h}})) \right\rangle [\![\delta\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!] \mathrm{dS}.\end{aligned} \tag{84}$$

4.3.2. Solution uniqueness

Let us first define $\boldsymbol{\eta} = \mathrm{I}_{\mathrm{h}}\mathbf{M} - \mathbf{M}^{\mathrm{e}} \in \mathrm{X}$, with $\mathrm{I}_{\mathrm{h}}\mathbf{M} \in \mathrm{X}^{\mathrm{k}^+}$ the interpolant of $\mathbf{M}^{\mathrm{e}}$ in $\mathrm{X}^{\mathrm{k}^+}$. The last relation (81) thus becomes

$$\begin{aligned}\mathcal{A}(\mathbf{M}^{\mathrm{e}}; \mathrm{I}_{\mathrm{h}}\mathbf{M} - \mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) + \mathcal{B}(\mathbf{M}^{\mathrm{e}}; \mathrm{I}_{\mathrm{h}}\mathbf{M} - \mathbf{M}_{\mathrm{h}}, \delta\mathbf{M}_{\mathrm{h}}) = & \mathcal{A}(\mathbf{M}^{\mathrm{e}}; \boldsymbol{\eta}, \delta\mathbf{M}_{\mathrm{h}}) + \mathcal{B}(\mathbf{M}^{\mathrm{e}}; \boldsymbol{\eta}, \delta\mathbf{M}_{\mathrm{h}}) \\ & + \mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{M}_{\mathrm{h}}; \delta\mathbf{M}_{\mathrm{h}}) \quad \forall\delta\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}}.\end{aligned} \tag{85}$$

To prove the existence of a solution $\mathbf{M}_{\mathrm{h}}$ of the problem stated by Eq. (80), which corresponds to the DG finite element discretization (79), the problem is stated in the fixed point formulation and a map $\mathrm{S}_{\mathrm{h}} : \mathrm{X}^{\mathrm{k}^+} \to \mathrm{X}^{\mathrm{k}^+}$ is defined as follows [18]: for a given $\mathbf{y} \in \mathrm{X}^{\mathrm{k}^+}$, find $\mathrm{S}_{\mathrm{h}}(\mathbf{y}) = \mathbf{M}_{\mathbf{y}} \in \mathrm{X}^{\mathrm{k}^+}$, such that

$$\begin{aligned}\mathcal{A}(\mathbf{M}^{\mathrm{e}}; \mathrm{I}_{\mathrm{h}}\mathbf{M} - \mathbf{M}_{\mathbf{y}}, \delta\mathbf{M}_{\mathrm{h}}) + \mathcal{B}(\mathbf{M}^{\mathrm{e}}; \mathrm{I}_{\mathrm{h}}\mathbf{M} - \mathbf{M}_{\mathbf{y}}, \delta\mathbf{M}_{\mathrm{h}}) = \mathcal{A}(\mathbf{M}^{\mathrm{e}}; \boldsymbol{\eta}, \boldsymbol{\delta}\mathbf{M}_{\mathrm{h}}) + \mathcal{B}(\mathbf{M}^{\mathrm{e}}; \boldsymbol{\eta}, \delta\mathbf{M}_{\mathrm{h}}) + \mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{y}; \delta\mathbf{M}_{\mathrm{h}}) \\ \forall\delta\mathbf{M}_{\mathrm{h}} \in \mathrm{X}^{\mathrm{k}}.\end{aligned} \tag{86}$$

The existence of a fixed point of the map S_{h} is equivalent to the existence of a solution $\mathbf{M}_{\mathrm{h}}$ of the discrete problem (79), see [18].

For the following analysis, we denote by C^{k}, a positive generic constant which is independent of the mesh size, but may depend on $\mathrm{C}_{\mathcal{T}}, \mathrm{C}^{\mathrm{k}}_{\mathcal{D}}, \mathrm{C}^{\mathrm{k}}_{\mathcal{I}}, \mathrm{C}^{\mathrm{k}}_{\mathcal{K}}$, and on k which are defined in Appendix B, and can take different values at different places.

To demonstrate the uniqueness, we have recourse to the following Lemmata, following [18].

Lemma 4.2 (Lower bound)**.** *For $\mathcal{B}$ larger than a constant, which depends on the polynomial approximation only, there exist two constants C_1^k and C_2^k, such that*

$$\mathcal{A}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \delta\boldsymbol{M}_h) + \mathcal{B}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \delta\boldsymbol{M}_h) \geq C_1^k \mid\mid\mid \delta\boldsymbol{M}_h \mid\mid\mid_*^2 - C_2^k \parallel \delta\boldsymbol{M}_h \parallel_{L^2(\Omega)}^2 \quad \forall\delta\boldsymbol{M}_h \in X^k, \tag{87}$$

$$\mathcal{A}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \delta\boldsymbol{M}_h) + \mathcal{B}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \delta\boldsymbol{M}_h) \geq C_1^k \mid\mid\mid \delta\boldsymbol{M}_h \mid\mid\mid^2 - C_2^k \parallel \delta\boldsymbol{M}_h \parallel_{L^2(\Omega)}^2 \quad \forall\delta\boldsymbol{M}_h \in X^k. \tag{88}$$

Proceeding by using the bounds (71) and (76), the Cauchy-Schwartz' inequality, the trace inequality on the finite element space (B.6), the trace inequality, Eq. (B.4), and inverse inequality, Eq. (B.9), the ξ-inequality $–\xi > 0 : |\mathrm{ab}| \leq \frac{\xi}{4}\mathrm{a}^2 + \frac{1}{\xi}\mathrm{b}^2$, as in Wheeler et al. [14] and Prudhomme *et al.* [8] analyzes, yields to prove this Lemma 4.2. The two positive constants $\mathrm{C}_1^{\mathrm{k}}, \mathrm{C}_2^{\mathrm{k}}$ are independent of the mesh size, but do depend on k and $\mathcal{B}$. The proof follows the one presented in [24]. In particular, for $\mathrm{C}_1^{\mathrm{k}}$ to be positive, the following constraint on the stabilization parameter should be satisfied: $\mathcal{B} > \frac{\mathrm{C}_{\mathrm{y}}^2}{\mathrm{C}_{\alpha}^2}\mathrm{C}^{\mathrm{k}}$. Therefore for the method to be stable, the stabilization parameter should be large enough depending on the polynomial approximation.

Lemma 4.3 (Upper bound)**.** *There exist $C > 0$ and $C^k > 0$ such that*

$$| \mathcal{A}(\boldsymbol{M}^e; \boldsymbol{u}, \delta\boldsymbol{M}) + \mathcal{B}(\boldsymbol{M}^e; \boldsymbol{u}, \delta\boldsymbol{M}) | \leq C \,|||\, \boldsymbol{u} \,|||_1 \,|||\, \delta\boldsymbol{M} \,|||_1 \quad \forall \boldsymbol{u}\,,\, \delta\boldsymbol{M} \in X, \tag{89}$$

$$| \mathcal{A}(\boldsymbol{M}^e; \boldsymbol{u}, \delta\boldsymbol{M}_h) + \mathcal{B}(\boldsymbol{M}^e; \boldsymbol{u}, \delta\boldsymbol{M}_h) | \leq C^k \,|||\, \boldsymbol{u} \,|||_1 \,|||\, \delta\boldsymbol{M}_h \,||| \quad \forall \boldsymbol{u} \in X,\, \delta\boldsymbol{M}_h \in X^k, \tag{90}$$

$$| \mathcal{A}(\boldsymbol{M}^e; \boldsymbol{u}_h, \delta\boldsymbol{M}_h) + \mathcal{B}(\boldsymbol{M}^e; \boldsymbol{u}_h, \delta\boldsymbol{M}_h) | \leq C^k \,|||\, \boldsymbol{u}_h \,|||\, |||\, \delta\boldsymbol{M}_h \,||| \quad \forall \boldsymbol{u}_h,\, \delta\boldsymbol{M}_h \in X^k. \tag{91}$$

Applying the Hölder's inequality, and the bound (76) on each term of $\mathcal{A}(\mathbf{M}^e; \mathbf{u}, \delta\mathbf{M}) + \mathcal{B}(\mathbf{M}^e; \mathbf{u}, \delta\mathbf{M})$ and then applying the Cauchy-Schwartz' inequality, lead to relation (89). Therefore relations (90) and (91) are easily deduced from the relation between energy norms on the finite element space, Eq. (66). The proof follows the one presented in [24].

Lemma 4.4 (Energy bound)**.** *Let $\boldsymbol{M}^e \in X_s, s \geq 2$, and let $I_h\boldsymbol{M} \in X^k, s \geq 2$, be its interpolant. Therefore, there is a positive constant $C^k > 0$ independent of h_s, such that*

$$|||\, \boldsymbol{M}^e - I_h\boldsymbol{M} \,|||_1 \leq C^k h^{\mu-1} \parallel \boldsymbol{M}^e \parallel_{H^s(\Omega_h)}, \tag{92}$$

with $\mu = min\{s, k+1\}$.

The proof follows from applying the interpolation inequalities, Eq. (B.1) and Eq. (B.3), on the mesh dependent norm (65).

Lemma 4.5 (Auxiliary problem)**.** *We consider the following auxiliary problem, with $\boldsymbol{\phi} \in L^2(\Omega) \times L^2(\Omega)$:*

$$\begin{aligned} -\nabla^T (\boldsymbol{j}_{\nabla\boldsymbol{M}}(\boldsymbol{M}^e)\nabla\boldsymbol{\psi} + \boldsymbol{j}_{\boldsymbol{M}}(\boldsymbol{M}^e, \nabla\boldsymbol{M}^e)\boldsymbol{\psi}) = \boldsymbol{\phi} \quad &on \quad \Omega, \\ \boldsymbol{\psi} = 0 \quad &on \quad \partial\Omega. \end{aligned} \tag{93}$$

Assuming regular ellipticity of the operator, there is a unique solution $\boldsymbol{\psi} \in H^2(\Omega) \times H^2(\Omega)$ to the problem stated by Eq. (93), which satisfies the elliptic property

$$\parallel \boldsymbol{\psi} \parallel_{H^2(\Omega_h)} \leq C \parallel \boldsymbol{\phi} \parallel_{L^2(\Omega_h)} . \tag{94}$$

Moreover, for a given $\boldsymbol{\xi} \in L^2(\Omega_h) \times L^2(\Omega_h)$ there exists a unique $\boldsymbol{\phi}_h \in X^k$ such that

$$\mathcal{A}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \boldsymbol{\phi}_h) + \mathcal{B}(\boldsymbol{M}^e; \delta\boldsymbol{M}_h, \boldsymbol{\phi}_h) = \sum_e \int_{\Omega^e} \boldsymbol{\xi}^T \delta\boldsymbol{M}_h d\Omega \;\; \forall \delta\boldsymbol{M}_h \in X^k, \tag{95}$$

and there is a constant C^k such that :

$$|||\, \boldsymbol{\phi}_h \,||| \leq C^k \parallel \boldsymbol{\xi} \parallel_{L^2(\Omega_h)} . \tag{96}$$

The proof of the first statement is given in [25], by combining [25, Theorem 8.3] to [25, Lemma 9.17]. The proof of the second statement follows follows the methodology described by Gudi *et al.* [18]: The use of Lemma 4.2 and Eq. (95) with $\delta\mathbf{M}_\mathrm{h} = \boldsymbol{\phi}_\mathrm{h}$ allows bounding $|||\, \boldsymbol{\phi}_\mathrm{h} \,|||$ in terms of $\parallel \boldsymbol{\xi} \parallel_{\mathrm{L}^2(\Omega_\mathrm{h})}$ and $\parallel \boldsymbol{\phi}_\mathrm{h} \parallel_{\mathrm{L}^2(\Omega_\mathrm{h})}$; $\parallel \boldsymbol{\phi}_\mathrm{h} \parallel_{\mathrm{L}^2(\Omega_\mathrm{h})}$ is then estimated by considering $\boldsymbol{\phi} = \boldsymbol{\phi}_\mathrm{h} \in \mathrm{X}^\mathrm{k}$ in Eq. (93), multiplying the result by $\boldsymbol{\phi}_\mathrm{h}$ and integrating it by parts on Ω_h yielding $\parallel \boldsymbol{\phi}_\mathrm{h} \parallel^2_{\mathrm{L}^2(\Omega_\mathrm{h})} = \mathcal{A}(\mathbf{M}^e; \boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^e; \boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h})$. Inserting the interpolant

$\mathrm{I}_h\boldsymbol{\psi}$ in these last terms, *i.e.* $\| \boldsymbol{\phi}_\mathrm{h} \|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} = \mathcal{A}(\mathbf{M}^\mathrm{e}; \boldsymbol{\psi} - \mathrm{I}_h\boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \boldsymbol{\psi} - \mathrm{I}_h\boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I}_h\boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{G}^\mathrm{e}; \mathrm{I}_h\boldsymbol{\psi}, \boldsymbol{\phi}_\mathrm{h})$, bounding the last two terms by considering $\delta\mathbf{M}_\mathrm{h} = \mathrm{I}_h\boldsymbol{\psi}$ in Eq. (95), bounding the first two terms by making successive use of Lemmata 4.3 and 4.4, and using the regular ellipticity, Eq. (94), allow deriving the bound $\| \boldsymbol{\phi}_\mathrm{h} \|_{\mathrm{L}^2(\Omega_\mathrm{h})} \leq \mathrm{C}^\mathrm{k} \| \boldsymbol{\xi} \|_{\mathrm{L}^2(\Omega_\mathrm{h})}$, which results into the proof of (96).

The existence of the solution of the discrete problem is demonstrated by proving, using these Lemmata [18], that the map S_h has a fixed point.

Theorem 4.6 (Solution uniqueness to the problem stated by Eq. (86))**.** *The solution $\boldsymbol{M_y}$ to the problem stated by Eq. (86) is unique for a given $\boldsymbol{y} \in X^{k^+}$ with $S_h(\boldsymbol{y}) = \boldsymbol{M_y}$.*

Let us assume that there are two distinct solutions $\mathbf{M}_{\mathbf{y}_1}$, $\mathbf{M}_{\mathbf{y}_2}$, which result into

$$\mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_1}, \delta\mathbf{M}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_1}, \delta\mathbf{M}_\mathrm{h}) = \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_2}, \delta\mathbf{M}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_2}, \delta\mathbf{M}_\mathrm{h}) \quad \forall \ \delta\mathbf{M}_\mathrm{h} \in \mathrm{X}^\mathrm{k}. \tag{97}$$

For fixed $\mathbf{M}^\mathrm{e}$, $\mathcal{A}$ and $\mathcal{B}$ are bi-linear, therefore this last relation becomes

$$\mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2}, \delta\mathbf{M}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2}, \delta\mathbf{M}_\mathrm{h}) = 0 \ \forall \ \delta\mathbf{M}_\mathrm{h} \in \mathrm{X}^\mathrm{k}. \tag{98}$$

Using Lemma 4.5, with $\boldsymbol{\xi} = \delta\mathbf{M}_\mathrm{h} = \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2} \in \mathrm{X}^\mathrm{k}$ results in stating that there is a unique $\boldsymbol{\Phi}_\mathrm{h} \in \mathrm{X}^\mathrm{k}$ solution of the problem Eq. (95), with

$$\mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2}, \boldsymbol{\Phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2}, \boldsymbol{\Phi}_\mathrm{h}) = \| \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2} \|^2_{\mathrm{L}^2(\Omega_\mathrm{h})}, \tag{99}$$

and that $||| \boldsymbol{\Phi}_\mathrm{h} ||| \leq \mathrm{C}^\mathrm{k} \| \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2} \|_{\mathrm{L}^2(\Omega_\mathrm{h})}$. Choosing $\delta\mathbf{M}_\mathrm{h}$ as $\boldsymbol{\Phi}_\mathrm{h}$ in Eq. (98), we have $\| \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2} \|_{\mathrm{L}^2(\Omega_\mathrm{h})} = 0$. Therefore, the solution $\mathrm{S}_\mathrm{h}(\mathbf{y}) = \mathbf{M}_\mathbf{y}$ is unique.

It is now demonstrated that S_h maps a ball $\mathrm{O}_\sigma(\mathrm{I}_\mathrm{h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$ into itself and is continuous in the ball. The ball O_σ is defined with a radius σ and is centered at the interpolant $\mathrm{I}_\mathrm{h}\mathbf{M}$ of $\mathbf{M}^\mathrm{e}$ as

$$\begin{aligned} \mathrm{O}_\sigma(\mathrm{I}_\mathrm{h}\mathbf{M}) &= \left\{ \mathbf{y} \in \mathrm{X}^{\mathrm{k}^+} \text{such that } ||| \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{y} |||_1 \leq \sigma \right\}, \\ \text{with} \quad \sigma &= \frac{||| \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}^\mathrm{e} |||_1}{\mathrm{h}_\mathrm{s}^\varepsilon}, \quad 0 < \varepsilon < \frac{1}{4}. \end{aligned} \tag{100}$$

The idea proposed in [18] is to work on a linearized problem in the ball $\mathrm{O}_\sigma(\mathrm{I}_\mathrm{h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$ around the interpolant $\mathrm{I}_\mathrm{h}\mathbf{M} \in \mathrm{X}^{\mathrm{k}^+}$ of $\mathbf{M}^\mathrm{e} \in \mathrm{X}^+$ so the nonlinear term $\mathbf{j}$ and its derivatives are locally bounded in the ball $\mathrm{O}_\sigma(\mathrm{I}_\mathrm{h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$. We note from Lemma 4.4, Eq. (92) that

$$||| \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}^\mathrm{e} |||_1 \leq \mathrm{C}^\mathrm{k}\mathrm{h}_\mathrm{s}^{\mu-1} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^\mathrm{s}(\Omega_\mathrm{h})} \text{ and } \sigma \leq \mathrm{C}^\mathrm{k}\mathrm{h}_\mathrm{s}^{\mu-1-\varepsilon} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^\mathrm{s}(\Omega_\mathrm{h})}, \tag{101}$$

with $\mu = \min\{\mathrm{s}, \mathrm{k}+1\}$. Assuming $\mathbf{M}^\mathrm{e} \in \mathrm{H}^{\frac{5}{2}}(\Omega) \times \mathrm{H}^{\frac{5}{2}^+}(\Omega)$ and considering $\mathrm{s} = \frac{5}{2}$, $\mathrm{C}_\mathrm{M} = \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^{\frac{5}{2}}(\Omega_\mathrm{h})}$, and $\mu = \frac{5}{2} = \mathrm{s}$, this equation is rewritten

$$||| \mathrm{I}_\mathrm{h}\mathbf{M} - \mathbf{M}^\mathrm{e} |||_1 \leq \mathrm{C}^\mathrm{k}\mathrm{h}_\mathrm{s}^{\frac{3}{2}} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^{\frac{5}{2}}(\Omega_\mathrm{h})} \text{ and } \sigma \leq \mathrm{C}^\mathrm{k}\mathrm{C}_\mathrm{M}\mathrm{h}_\mathrm{s}^{\frac{3}{2}-\varepsilon} \quad \text{if } \mathrm{k} \geq 2. \tag{102}$$

Moreover, it can be shown that $\mathbf{j}(\mathbf{x}; \mathbf{y}, \nabla\mathbf{y})$, $\mathbf{j}_\mathbf{M}(\mathbf{x}; \mathbf{y}, \nabla\mathbf{y})$, $\mathbf{j}_{\mathbf{MM}}(\mathbf{x}; \mathbf{y}, \nabla\mathbf{y})$, $\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{x}; \mathbf{y})$, $\mathbf{j}_{\mathbf{M}\nabla\mathbf{M}}(\mathbf{x}; \mathbf{y})$ are bounded for $\mathbf{x} \in \bar{\Omega}$, $\mathbf{y} \in \mathrm{O}_\sigma(\mathrm{I}_\mathrm{h}\mathbf{M})$, by the same reasoning as in [17] for $\mathrm{d} = 2$, which justifies Eq. (76).

We have now the tools to bound the nonlinear term $\mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}; \delta\mathbf{M}_\mathrm{h})$ of Eq. (86).

Lemma 4.7. *Let $\boldsymbol{y} \in O_\sigma(I_h\boldsymbol{M}) \subset X^{k^+}$, with the ball $O_\sigma(I_h\boldsymbol{M})$ of radius σ defined in Eq. (100) and with $I_h\boldsymbol{M} \in X^{k^+}$ the interpolant of $\boldsymbol{M}^e \in X^+$ in X^{k^+}, and let $\delta\boldsymbol{M}_h \in X^k$, then we have the bound*

$$| \mathcal{N}(\boldsymbol{M}^e, \boldsymbol{y}; \delta\boldsymbol{M}_h) | \le C^k C_y \parallel \boldsymbol{M}^e \parallel_{H^s(\Omega_h)} h_s^{\mu-2-\varepsilon}\sigma \left[| \delta\boldsymbol{M}_h |_{H^1(\Omega_h)} + \left(\sum_e h_s | \delta\boldsymbol{M}_h |^2_{H^1(\partial\Omega^e)}\right)^{\frac{1}{2}} + \left(\sum_e h_s^{-1} \parallel [\![\delta\boldsymbol{M}_{h_{\boldsymbol{n}}}]\!] \parallel^2_{L^2(\partial\Omega^e)}\right)^{\frac{1}{2}}\right]. \tag{103}$$

Moreover, one has

$$| \mathcal{N}(\boldsymbol{M}^e, \boldsymbol{y}; \delta\boldsymbol{M}_h) | \le C^k C_y \parallel \boldsymbol{M}^e \parallel_{H^s(\Omega_h)} {h_s}^{\mu-2-\varepsilon}\sigma \; ||| \, \delta\boldsymbol{M}_h \, |||_1 \le C^k C_y \parallel \boldsymbol{M}^e \parallel_{H^s(\Omega_h)} {h_s}^{\mu-2-\varepsilon}\sigma \; ||| \, \delta\boldsymbol{M}_h \, |||, \tag{104}$$

with $\mu = min\{s,\ k+1\}$, or again

$$| \mathcal{N}(\boldsymbol{M}^e, \boldsymbol{y}; \delta\boldsymbol{M}_h) | \le C^k C_y C_M {h_s}^{\frac{1}{2}-\varepsilon}\sigma \; ||| \, \delta\boldsymbol{M}_h \, ||| \quad \textit{if } k \ge 2. \tag{105}$$

The bound (103) is obtained by defining $\boldsymbol{\zeta} = \mathbf{M}^{\mathrm{e}} - \mathbf{y}$ which can be expanded as $\boldsymbol{\zeta} = \boldsymbol{\eta} + \boldsymbol{\xi}$ with $\boldsymbol{\eta} = \mathbf{M}^{\mathrm{e}} - \mathrm{I_h}\mathbf{M} \in \mathrm{X}$ and $\boldsymbol{\xi} = \mathrm{I_h}\mathbf{M} - \mathbf{y} \in \mathrm{X^k}$. Therefore, every term of Eq. (84) can be bounded separately using Taylor's series (72-74), the generalized Hölder's inequality, the generalized Cauchy-Schwartz' inequality, the definition of $\mathrm{C_y}$ in Eq. (76), the norm definition, Eq. (65), the definition of the ball, Eqs. (100, 101), and some other inequalities which are reported in Appendix B, such as trace inequalities Eqs. (B.4-B.6), inverse inequalities Eqs. (B.7-B.9) in the particular case of $\mathrm{d} = 2$, and the interpolation inequalities Eqs. (B.1-B.3) in the particular case of $\mathrm{d} = 2$. The proof follows from the argumentation reported in [18] and is detailled in [24]. The bound of the nonlinear term $\mathcal{N}(\mathbf{M}^{\mathrm{e}}, \mathbf{y}; \delta\mathbf{M}_{\mathrm{h}})$ is nominated by the term with the largest bound and as a result we get Eq. (103). Moreover, using the definition of the energy norm (65) yields (104). Finally, by using Lemma 4.1, Eq. (104) is rewritten as (105).

We now have the tools to demonstrate that $\mathrm{S_h}$ (i) maps a ball $\mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$ into itself and (ii) is continuous in the ball.

Theorem 4.8 ($\mathrm{S_h}$ maps $\mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M})$ into itself)**.** *Let $0 < h_s < 1$ and σ be defined by Eq. (101), therefore*

$$||| \, I_h\boldsymbol{M} - \boldsymbol{M}_{\boldsymbol{y}} \, ||| \le C^{k'} \sigma h_s^{\varepsilon} \textit{ if } k \ge 2, \tag{106}$$

where, for a mesh size h_s small enough and a given ball size σ, $I_h\boldsymbol{M} - \boldsymbol{M}_{\boldsymbol{y}} \longrightarrow 0$, hence S_h maps $O_\sigma(I_h\boldsymbol{M})$ to itself.

The proof is derived using the bounds obtained in Lemma 4.2, Eq. (88), Lemma 4.3, Eq. (90), Lemma 4.7, Eq. (105), the definition of the ball (100), and the auxiliary problem, Lemma 4.5. More details are reported in Appendix C.

Theorem 4.9 (The continuity of the map $\mathrm{S_h}$ in the ball $\mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M})$)**.** *For $\boldsymbol{y}_1$, $\boldsymbol{y}_2 \in O_\sigma(I_h\boldsymbol{M})$, let $\boldsymbol{M}_{\boldsymbol{y}_1} = S_h(\boldsymbol{y}_1)$, $\boldsymbol{M}_{\boldsymbol{y}_2} = S_h(\boldsymbol{y}_2)$ be solutions of Eq. (86). Then for $0 < h_s < 1$*

$$||| \, \boldsymbol{M}_{\boldsymbol{y}_1} - \boldsymbol{M}_{\boldsymbol{y}_2} \, ||| \le C^k C_y \parallel \boldsymbol{M}^e \parallel_{H^s(\Omega_h)} h_s^{\mu-2-\varepsilon} \; ||| \, \boldsymbol{y}_1 - \boldsymbol{y}_2 \, ||| \, . \tag{107}$$

The proof of this theorem results from using Lemma 4.2, Eq. (88), proceeding similarly as to establish Lemma 4.7, Eq. (105), and then using the auxiliary problem stated in Lemma 4.5. The proof is detailed in Appendix D.

Following the argumentation in [18], using the Theorems 4.8 and 4.9 yields that for small h_s, the maps $\mathrm{S_h}$ has a fixed point: there exists $\mathbf{M}_\mathrm{h} \in \mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M})$ such that $\mathbf{M}_\mathrm{h} = \mathrm{S_h}(\mathbf{M}_\mathrm{h})$. The uniqueness of this point is directly deduced from Theorem 4.9. Moreover, the existence of a unique fixed point of the map $\mathrm{S_h}$ yields the existence of a unique solution $\mathbf{M}_\mathrm{h}$ of the problem stated by Eq. (85), or again the existence of a unique solution $\mathbf{M}_\mathrm{h}$ of the original quasi-linear discrete problem (79), which is the steady state version of the finite element statement (53).

4.3.3. A priori error estimates

As $\mathrm{S_h}$ maps a ball into itself, we can use $\mathbf{M}_\mathrm{h}$ instead of $\mathbf{M}_\mathrm{y}$ in Eq. (106), hence we have

$$||| \mathrm{I_h}\mathbf{M} - \mathbf{M}_\mathrm{h} ||| \leq \mathrm{C}^{\mathrm{k}'}\sigma \mathrm{h}_\mathrm{s}^\varepsilon = \mathrm{C}^{\mathrm{k}'} ||| \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e} |||_1 \,. \tag{108}$$

Now using this last relation, Lemma 4.1, Eq. (66), Lemma 4.4, Eq. (92), and Eq. (108) leads to

$$\begin{aligned} ||| \mathbf{M}^\mathrm{e} - \mathbf{M}_\mathrm{h} |||_1 &\leq ||| \mathbf{M}^\mathrm{e} - \mathrm{I_h}\mathbf{M} |||_1 + ||| \mathrm{I_h}\mathbf{M} - \mathbf{M}_\mathrm{h} |||_1 \leq ||| \mathbf{M}^\mathrm{e} - \mathrm{I_h}\mathbf{M} |||_1 + \mathrm{C}^{\mathrm{k}'} ||| \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e} |||_1 \\ &\leq (1 + \mathrm{C}^{\mathrm{k}'}) ||| \mathbf{M}^\mathrm{e} - \mathrm{I_h}\mathbf{M} |||_1 \leq \mathrm{C}^\mathrm{k}(1 + \mathrm{C}^{\mathrm{k}'})\mathrm{h}_\mathrm{s}^{\mu-1} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H^s}(\Omega_\mathrm{h})} \leq \mathrm{C}^{\mathrm{k}''}\mathrm{h}_\mathrm{s}^{\mu-1} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H^s}(\Omega_\mathrm{h})}, \end{aligned} \tag{109}$$

where $\mu = \min\{\mathrm{s}, \mathrm{k}+1\}$, and $\mathrm{C}^{\mathrm{k}''} = \mathrm{C}^\mathrm{k}(1 + \mathrm{C}^{\mathrm{k}'})$. This shows that the convergence of the error estimate is optimal in $\mathrm{h_s}$.

4.3.4. Error estimate in L^2-norm

Since the linearized problem (85) is adjoint consistent, an optimal order of convergence in the L^2-norm is obtained by applying the duality argument.

To this end, let us consider the following dual problem

$$\begin{aligned} -\nabla^\mathrm{T}(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}) + \mathbf{j}_\mathbf{M}^\mathrm{T}(\mathbf{M}^\mathrm{e}, \nabla\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi} = \mathbf{e} \quad &\text{on} \quad \Omega, \\ \boldsymbol{\psi} = \mathbf{g} \quad &\text{on} \quad \partial\Omega, \end{aligned} \tag{110}$$

which is assumed to satisfy the elliptic regularity condition since $\mathbf{j}_{\nabla\mathbf{M}}$ is positive definite, with $\boldsymbol{\psi} \in \mathrm{H}^{2\mathrm{m}}(\Omega_\mathrm{h}) \times \mathrm{H}^{2\mathrm{m}}(\Omega_\mathrm{h})$ for $\mathrm{p} \geq 2\mathrm{m}$ and

$$\| \boldsymbol{\psi} \|_{\mathrm{H^p}(\Omega_\mathrm{h})} \leq \mathrm{C}\left(\| \mathbf{e} \|_{\mathrm{H}^{\mathrm{p}-2\mathrm{m}}_{(\Omega_\mathrm{h})}} + \| \mathbf{g} \|_{\mathrm{H}^{\mathrm{p}-\frac{1}{2}}_{(\partial\Omega_\mathrm{h})}} \right), \tag{111}$$

if $\mathbf{e} \in \mathrm{H}^{\mathrm{p}-2\mathrm{m}}(\Omega_\mathrm{h}) \times \mathrm{H}^{\mathrm{p}-2\mathrm{m}}(\Omega_\mathrm{h})$.

Considering $\mathbf{e} = \mathbf{M}^\mathrm{e} - \mathbf{M}_\mathrm{h} \subset \mathrm{L}^2(\Omega_\mathrm{h}) \times \mathrm{L}^2(\Omega_\mathrm{h})$ the error and $\mathbf{g} = 0$, multiplying Eq. (110) by $\mathbf{e}$, and integrating over Ω_h, result in

$$\int_{\Omega_\mathrm{h}} [\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}]^\mathrm{T} \nabla\mathbf{e}\mathrm{d}\Omega + \int_{\Omega_\mathrm{h}} \left[\mathbf{j}_\mathbf{M}^\mathrm{T}(\mathbf{M}^\mathrm{e}, \nabla\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}\right]^\mathrm{T} \mathbf{e}\mathrm{d}\Omega - \sum_\mathrm{e} \int_{\partial\Omega^\mathrm{e}} [\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}]^\mathrm{T} \mathbf{e}_\mathbf{n}\mathrm{dS} = \| \mathbf{e} \|^2_{\mathrm{L}^2(\Omega_\mathrm{h})}, \tag{112}$$

with

$$\parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})} \leq \mathrm{C} \parallel \mathbf{e} \parallel_{\mathrm{L}^2(\Omega_\mathrm{h})} . \tag{113}$$

Since $[\![\boldsymbol{\psi}]\!] = [\![\nabla\boldsymbol{\psi}]\!] = 0$ on $\partial_\mathrm{I}\Omega_\mathrm{h}$ and $[\![\boldsymbol{\psi}]\!] = -\boldsymbol{\psi} = 0$ on $\partial_\mathrm{D}\Omega_\mathrm{h}$, we have by comparison with Eqs. (82-83), that

$$\begin{cases} \int_{\Omega_\mathrm{h}} \left[\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}\right]^\mathrm{T} \nabla\mathbf{e}\mathrm{d}\Omega + \int_{\partial_\mathrm{I}\Omega_\mathrm{h}\cup\partial_\mathrm{D}\Omega_\mathrm{h}} \left[\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^\mathrm{e})\nabla\boldsymbol{\psi}\right]^\mathrm{T} [\![\mathbf{e}_\mathbf{n}]\!]\, \mathrm{dS} & = \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi}), \\ \int_{\Omega_\mathrm{h}} \left[\mathbf{j}_\mathbf{M}(\mathbf{M}^\mathrm{e}, \nabla\mathbf{M}^\mathrm{e})\mathbf{e}\right]^\mathrm{T} \nabla\boldsymbol{\psi}\mathrm{d}\Omega & = \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi}), \end{cases} \tag{114}$$

since $\mathbf{j}_{\nabla\mathbf{M}}$ is symmetric. Therefore, Eq. (112) reads

$$\parallel \mathbf{e} \parallel^2_{\mathrm{L}^2(\Omega_\mathrm{h})} = \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi}). \tag{115}$$

From Eq. (81), one has

$$\mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{M}^\mathrm{e} - \mathbf{M}_\mathrm{h}, \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{M}^\mathrm{e} - \mathbf{M}_\mathrm{h}, \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) = \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{M}_\mathrm{h}; \mathrm{I}_\mathrm{h}\boldsymbol{\psi}), \tag{116}$$

since $\mathbf{M}^\mathrm{e}$ is the exact solution and $\mathrm{I}_\mathrm{h}\boldsymbol{\psi} \in \mathrm{X}^\mathrm{k}$, and Eq. (115) is rewritten

$$\parallel \mathbf{e} \parallel^2_{\mathrm{L}^2(\Omega_\mathrm{h})} = \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi} - \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi} - \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) + \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{M}_\mathrm{h}; \mathrm{I}_\mathrm{h}\boldsymbol{\psi}). \tag{117}$$

First, using Lemma 4.3, Eq. (89), Lemma 4.4, Eq. (92), and Eq. (109), leads to

$$\begin{aligned} \mid \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi} - \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{e}, \boldsymbol{\psi} - \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) \mid & \leq \mathrm{C} \mid\parallel \mathbf{e} \parallel\mid_1 \mid\parallel \boldsymbol{\psi} - \mathrm{I}_\mathrm{h}\boldsymbol{\psi} \parallel\mid_1 \\ & \leq \mathrm{C}^\mathrm{k} \mid\parallel \mathbf{e} \parallel\mid_1 \ \mathrm{h}_\mathrm{s} \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})} \\ & \leq \mathrm{C}^{\mathrm{k}''} \mathrm{h}_\mathrm{s}^\mu \parallel \mathbf{M}^\mathrm{e} \parallel_{\mathrm{H}^\mathrm{s}(\Omega_\mathrm{h})} \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})}, \end{aligned} \tag{118}$$

with $\mu = \min\{\mathrm{s}, \mathrm{k}+1\}$. Then proceeding as for establishing Lemma 4.7 and using the a priori error estimate (108-109), see details in [24], lead to

$$\mid \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{M}_\mathrm{h}; \mathrm{I}_\mathrm{h}\boldsymbol{\psi}) \mid \leq \mathrm{C}^{\mathrm{k}''} \mathrm{C}_\mathrm{y} \mathrm{h}_\mathrm{s}^{2\mu-3} \parallel \mathbf{M}^\mathrm{e} \parallel^2_{\mathrm{H}^\mathrm{s}(\Omega_\mathrm{h})} \mid\parallel \mathrm{I}_\mathrm{h}\boldsymbol{\psi} \parallel\mid . \tag{119}$$

Finally, using Lemma 4.4, Eq. (92), remembering $[\![\boldsymbol{\psi}]\!] = 0$ in Ω, we deduce that

$$\mid\parallel \mathrm{I}_\mathrm{h}\boldsymbol{\psi} \parallel\mid \leq \mid\parallel \mathrm{I}_\mathrm{h}\boldsymbol{\psi} - \boldsymbol{\psi} \parallel\mid_1 + \mid\parallel \boldsymbol{\psi} \parallel\mid \leq \mathrm{C}^\mathrm{k}\mathrm{h}_\mathrm{s} \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})} + \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^1(\Omega_\mathrm{h})} \leq \mathrm{C}^\mathrm{k}(\mathrm{h}_\mathrm{s} + 1) \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})} . \tag{120}$$

Combining Eqs. (118-120), Eq. (117) becomes

$$\parallel \mathbf{e} \parallel^2_{\mathrm{L}^2(\Omega_\mathrm{h})} \leq \mathrm{C}^{\mathrm{k}''} \mathrm{h}_\mathrm{s}^\mu \left(1 + \mathrm{h}_\mathrm{s}^{\mu-3} \parallel \mathbf{M}^\mathrm{e} \parallel_{\mathrm{H}^s(\Omega_\mathrm{h})}\right) \parallel \mathbf{M}^\mathrm{e} \parallel_{\mathrm{H}^s(\Omega_\mathrm{h})} \parallel \boldsymbol{\psi} \parallel_{\mathrm{H}^2(\Omega_\mathrm{h})}, \tag{121}$$

with $\mu = \min\{\mathrm{s}, \mathrm{k}+1\}$, or using Eq. (113), the final result for $\mathrm{k} \geq 2$ ($\mu \geq 3$)

$$\parallel \mathbf{e} \parallel_{\mathrm{L}^2(\Omega_\mathrm{h})} \leq \mathrm{C}^{\mathrm{k}'''} \mathrm{h}_\mathrm{s}^\mu \parallel \mathbf{M}^\mathrm{e} \parallel_{\mathrm{H}^s(\Omega_\mathrm{h})} . \tag{122}$$

This result demonstrates the optimal convergence rate of the method with the mesh-size for cases in which $\mathrm{k} \geq 2$.

5. Numerical examples

We present 1-, 2-, and 3-dimensional simulations to verify the DG numerical properties for electro-thermal problems on shape regular and shape irregular meshes. First the method is compared to analytical results and continuous Galerkin formulation on simple 1D-tests, then the method is applied on 2D-tests to verify the optimal convergence rates. Finally, a 3D unit cell model is presented. In the applications, the Dirichlet boundary conditions have been enforced strongly for simplicity.

5.1. 1-D example with one material

The first test is inspired from [5], where the boundary condition induces an electric current density, with a constrained temperature on the two opposite faces, as shown in Fig. 2. The target of this test is to find the distribution of the temperature, electric potential, and their corresponding fluxes, when considering the material properties, i.e. $\mathbf{l}, \mathbf{k}$, and α, as reported in Table 1. The simulation is performed using a quadratic polynomial approximation, with 12 elements, and the value of the stabilization parameter is $\mathcal{B} = 100$.

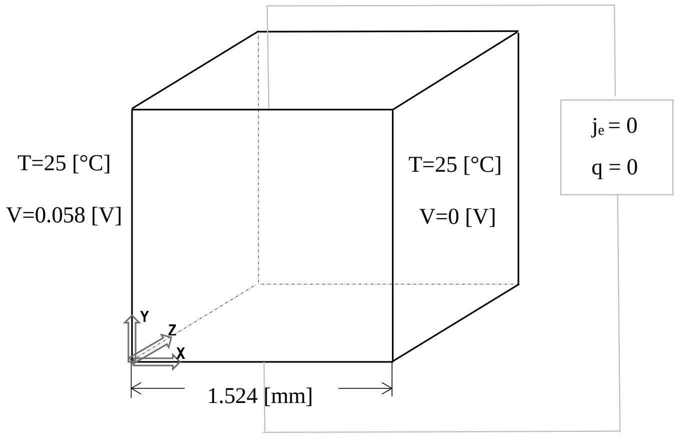
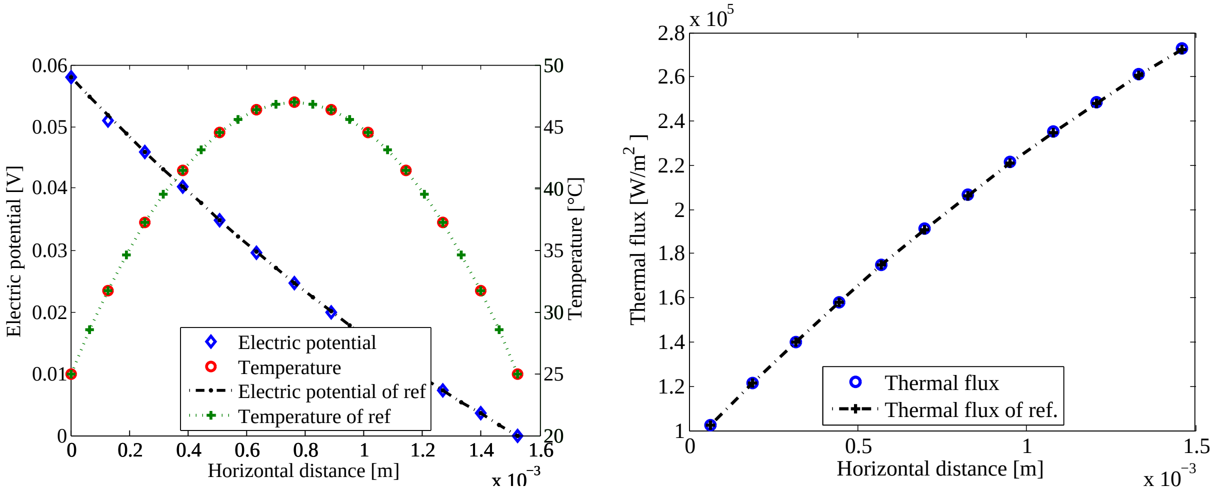


Figure 2: One-material electro-thermal problem and the boundary conditions

Table 1: Material parameters for the one-material electro-thermal problem

Material	$\mathbf{l}$ [S/m]	$\mathbf{k}$ [W/(K · m)]	α [V/K]
Bismuth telluride ($\mathbf{B}\mathrm{i}_2\mathbf{T}\mathrm{e}_3$)	diag(8.422×10^4)	diag(1.612)	1.941×10^{-4}

As it can be seen in Fig. 3(a), the electric potential distribution is close to linear but the temperature distribution is almost quadratic with a maximum value of 47 [°C] due to the volumetric Joule effect. This shows that this electro-thermal domain acts as a heat pump. Then Fig. 3(b) presents the distribution of the thermal flux which is almost linear with an electric current of about 3.2×10^6 [A/m^2]. The results of the



(a)

(b)

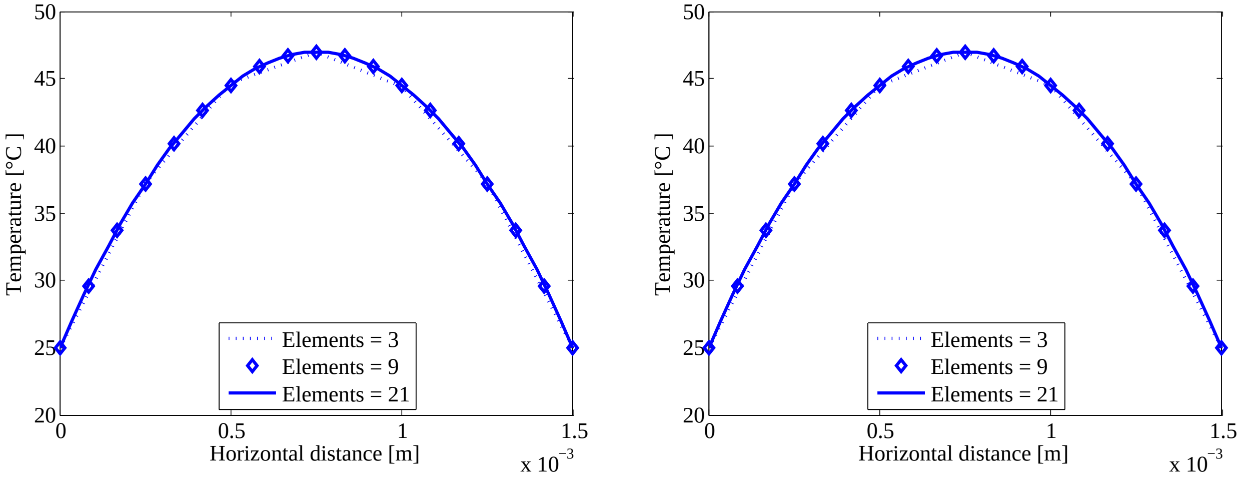
Figure 3: (a) The distributions of the electrical potential and temperature in the electro-thermal domain for one material, (b) the distribution of the thermal flux in the electro-thermal domain for one material. Ref.-curves are extracted from [5].

present DG method agree with the analytical approximation provided in [5] –the difference being due to the approximations required to derive the analytical solution.

Then the same test is simulated with the same boundary conditions, polynomial degree approximation, and value of $\mathcal{B}$, but with successively 3, 9, and 21 elements. Figure 4 presents the comparison of the results obtained with a Continuous Galerkin (CG) and the Discontinuous Galerkin (DG) formulations. As the distributions are almost parabolic, three elements already capture the solution, which does not make this test fit to study the convergence rate. Figure 5 illustrates the comparison of the thermal flux (one value per element is reported) with different mesh sizes between the CG and DG formulations and shows that the same thermal flux distribution is recovered. We also note from Figs. 4(a and b) and Figs. 5(a and b), that the results of the present DG formulation are in agreement with those obtained by the CG method.

5.2. 1-D example with two materials

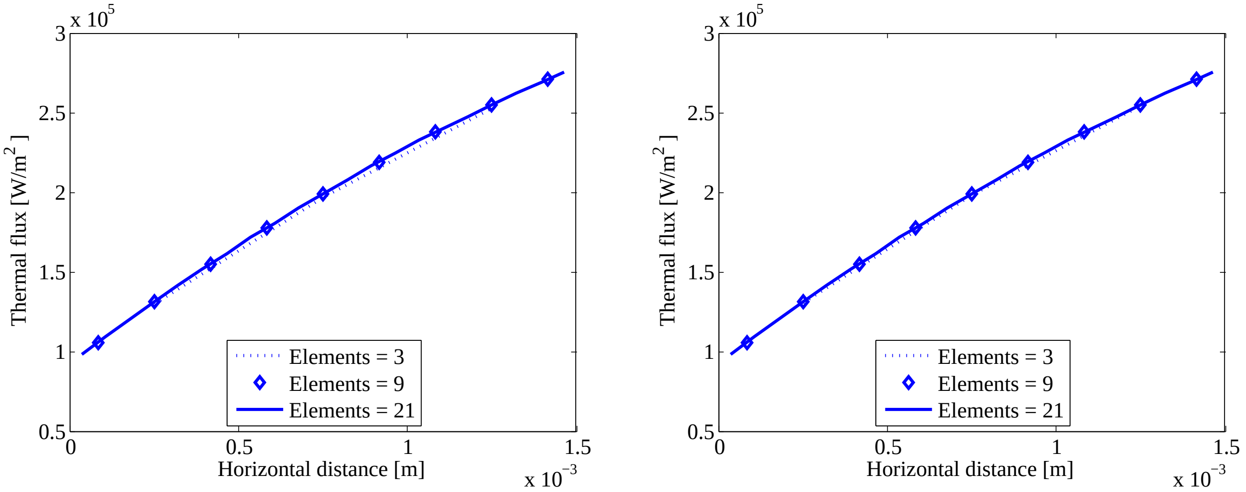
By applying the same kind of boundary conditions but for a combination of two materials –matrix (i.e., polymer) which is a non-conductive material and conductive fillers (i.e., carbon fiber)– as shown in Fig. 6, we can study the effect of the DG formulation in case of material interfaces. The electrical and thermal material properties are considered constant and reported in Table. 2, for the carbon fiber and the polymer matrix.



(a) Discontinuous Galerkin

(b) Continuous Galerkin

Figure 4: Comparison between the distributions of the temperature in the electro-thermal domain for different numbers of elements between (a) the DG formulation, and (b) the CG formulation



(a) Discontinuous Galerkin

(b) Continuous Galerkin

Figure 5: Comparison between the distributions of the thermal flux in the electro-thermal domain for different numbers of elements between (a) the DG formulation, and (b) the CG formulation

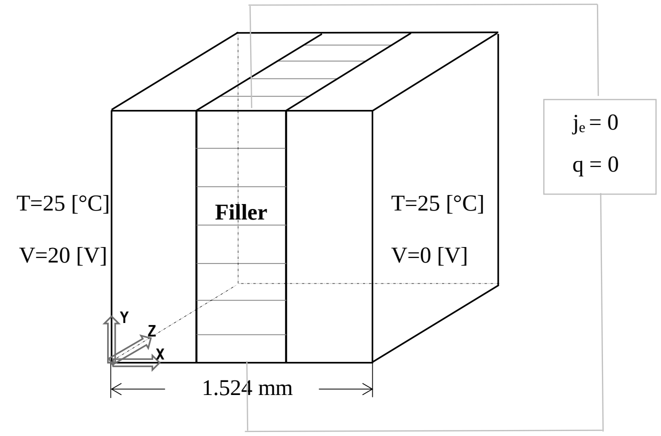
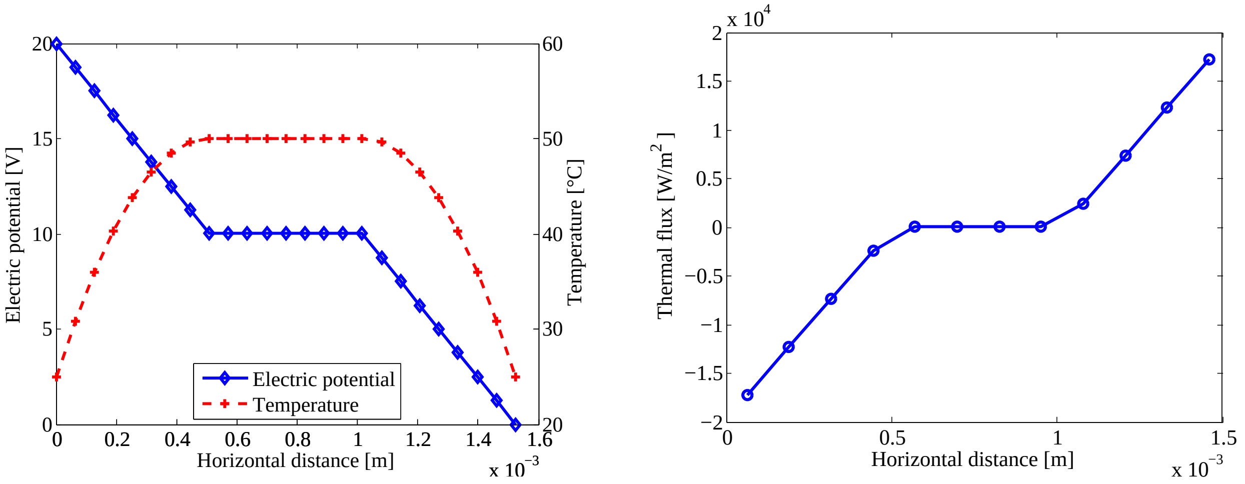


Figure 6: Electro-thermal composite domain and the boundary conditions

Table 2: Composite material phases parameters

Material	$\mathbf{l}$ [S/m]	$\mathbf{k}$ [W/(K $\cdot$ m)]	α [V/K]
Carbon fiber	diag(100000)	diag(40)	3 $\times 10^{-6}$
Polymer	diag(0.1)	diag(0.2)	3 $\times 10^{-7}$

Second order polynomial approximations, 12 elements, and the value of $\mathcal{B} = 100$, are still considered in this test. An electric potential difference of 20 [V] is applied, which is higher than in the previous test in order to reach a similar increase in temperature. Figure 7(a) shows the distributions of the voltage and of the temperature in this electro-thermal composite domain, and Fig. 7(b) the distribution of the thermal flux. We can see that the temperature, electric potential, and thermal flux fields are almost constant in the filler (the conductive material), since its electrical conductivity is high, and transient gradually in the polymer matrix (non conductive material). The resulting electric current is of about 1.96×10^3 [A/m^2].



(a)

(b)

Figure 7: (a) The distributions of the electrical potential and temperature in the electro-thermal composite domain, and (b) the distribution of the thermal flux in the electro-thermal composite domain

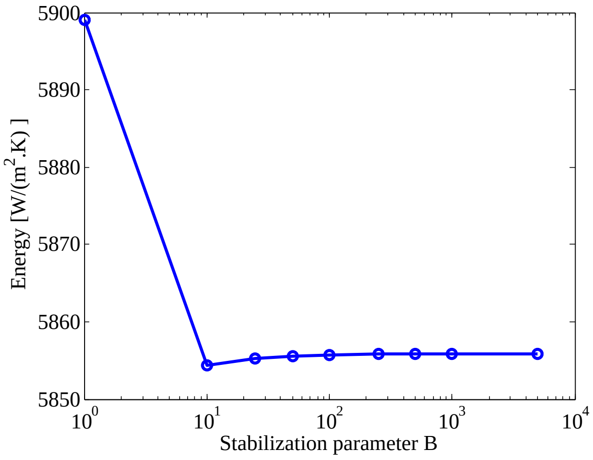


Figure 8: The internal energy of the electro-thermal composite domain for different values of the stabilization parameter $\mathcal{B}$

Then, we carry out the study of the stabilization parameter effect on the quality of the approximation in Fig. 8, where the internal energy per unit section is presented in terms of the stabilization parameter. The test is simulated with different values of the stabilization parameter $\mathcal{B}$ =1, 10, 25, 50, 100, 250, 500, 1000, and 5000. Although for the lower value of the stability parameter, the energy is overestimated, sign of an instability, the energy converges from below for stabilization parameters $\mathcal{B}$ ≥10, which proves that if $\mathcal{B}$ is large enough, the method is stable.

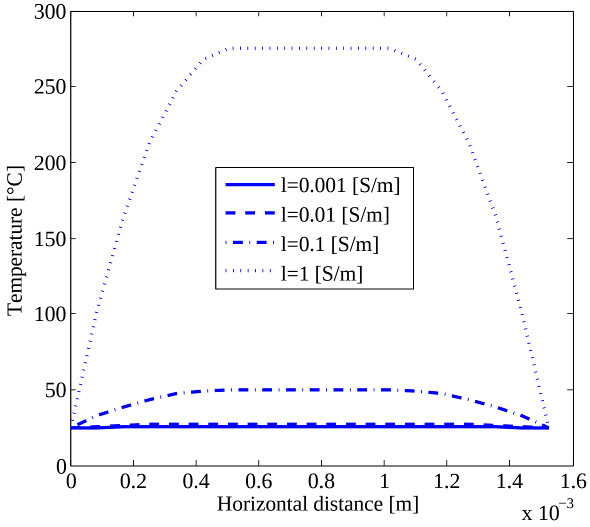


Figure 9: The temperature distributions in the electro-thermal composite domain for different values of electrical conductivity of the matrix material

Figure 9 compares the results obtained on the composite domain for different electrical conductivity values of the matrix material, all the other parameters being the same as before. This figure shows the difference in the maximum temperature reached when different values of the electrical conductivity are applied. This result indicates that the present DG formulation can be used for composite materials with high or low contrast.

5.3. 1-D The variation of electric potential with temperature difference

The following test converts heat energy into electricity, in Bismuth Telluride with the set up of Fig. 10 and the material parameters as presented in Table 1.

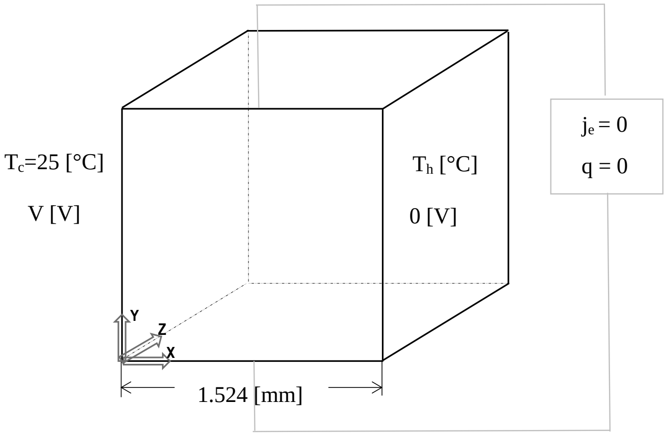


Figure 10: Electro-thermal unit cell and applied temperature difference as boundary condition

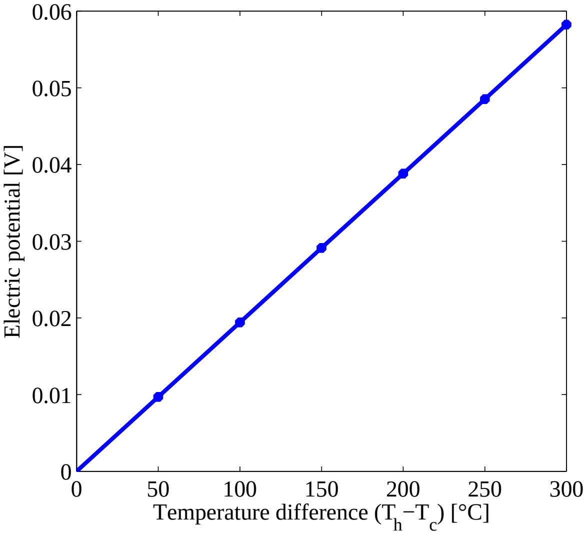


Figure 11: The variation of electric potential with temperature difference

The result in Fig. 11 shows the relation between the electric potential and temperature difference. It can be seen that the output electric potential, according to Seebeck coefficient, increases as the temperature difference increases. This proves that our formulation is effective and works in the two directions, production of electricity from temperature difference, as showed on this test and production of temperature difference by applying electric current, as showed in the previous examples.

5.4. 2-D study of convergence order

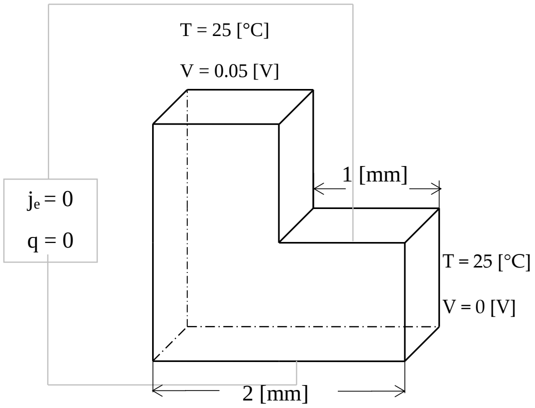
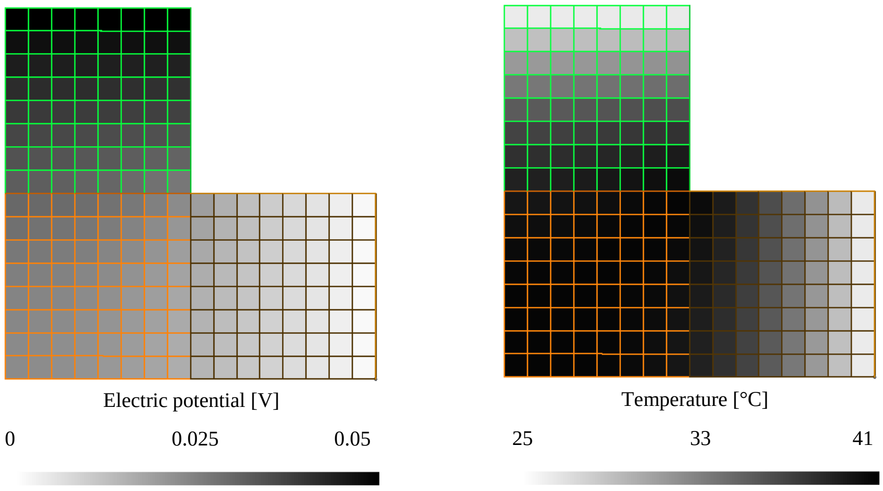


Figure 12: L-shaped electro-thermal problem and the boundary conditions



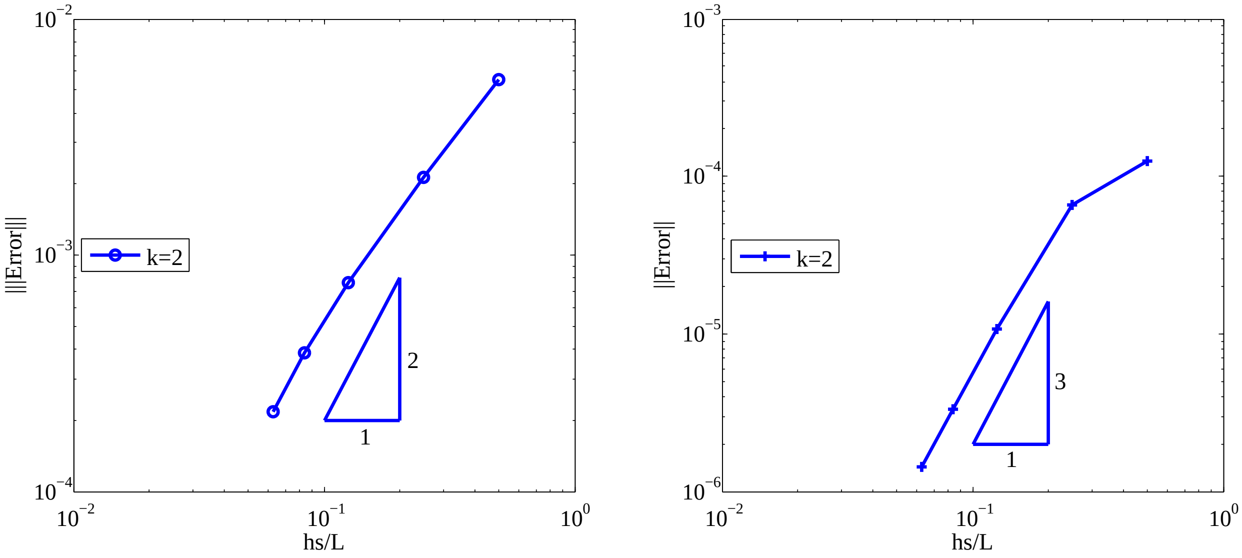
(a)

(b)

Figure 13: The distribution in the L-shaped electro-thermal problem of (a) the electrical potential, and (b) the temperature

In order to generate 2D gradients, we consider an L-shaped domain with the boundary condition illustrated in Fig. (12), and with the material properties reported in Table. 1. To prove the optimal rate of convergence in the L^2-norm and H^1-norm, a uniform h_s refinement is considered. A second order polynomial approximation is considered with $\mathcal{B} = 100$. The resulting distributions of temperature and electrical potential are illustrated respectively in Fig. 13(a) and in Fig. 13(b).

First, the convergence rate of the energy error $||| \mathbf{M}^e - \mathbf{M}_h |||$ –error in the H^1-norm– with respect to the mesh size is reported in Fig. 14(a). The reference solution is obtained with a refined mesh of $h_s/L = 1/32$. It can be seen that as the mesh is refined the error in the energy decreases quadratically for quadratic elements, once the mesh size is small enough. Thereby that confirms the prior error estimate derived in Section 4.3.3. Second, the error in the L^2-norm in terms of the mesh size h_s is illustrated in Fig. 14(b). The computed order of convergence is $k + 1$ for $k = 2$, which is optimal, once the mesh size is small enough, in agreement with the theory predicted by Section 4.3.4.



(a) Error in the H^1 norm

(b) Error in the L^2 norm

Figure 14: Error with respect to the mesh size: (a) in the H^1 norm, (b) in the L^2 norm

5.5. 3-D unit cell simulation

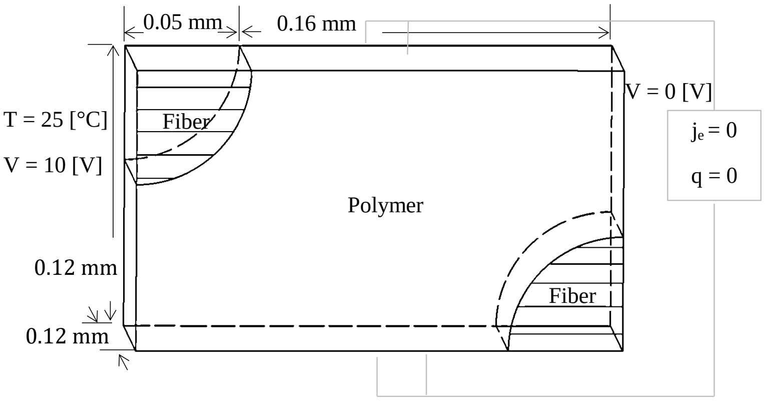
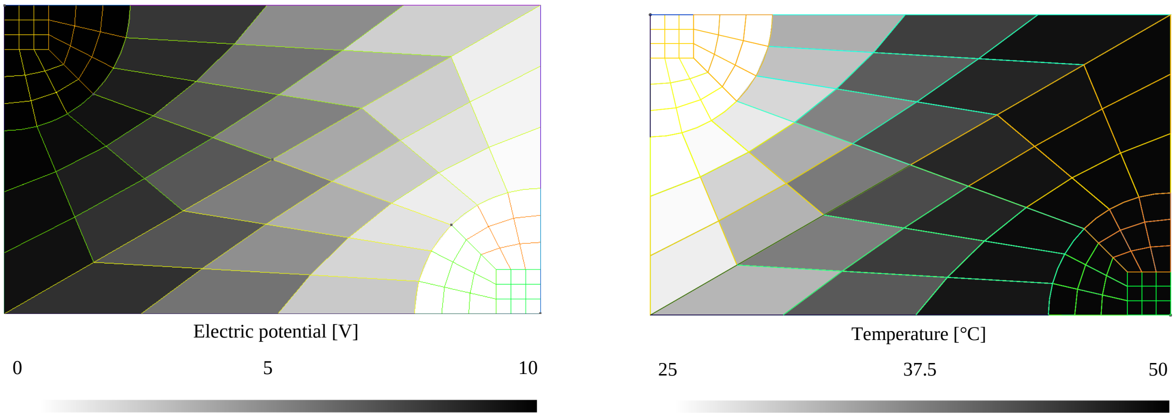


Figure 15: Electro-thermal unit cell and boundary condition

The third test illustrates the electrical thermal behavior of a composite material i.e., carbon fiber reinforced polymer matrix, which is heated by pplying an electric current. The studied unit cell and the boundary conditions are illustrated in Fig. 15, and the materials properties are reported in Table 2. A finite element mesh of 90 quadratic bricks is considered (the test is thus run in 3D). The initial temperature of the cell is 25 [°C].



(a)

(b)

Figure 16: The distributions in the unit cell of (a) the electrical potential, and (b) the temperature

Figure 16 presents the distributions of the temperature and of the electric potential in the unit cell. When the electric potential of 10 [V] is applied on one side, the temperature of the other side increases from 25 [°C] to 50.06 [°C]. This shows the applicability of the present formulation when different (irregular) mesh sizes are used simultaneously.

6. Conclusions

In this work, starting from the continuum theory for electro-thermal coupled problems, based on continuum mechanics and thermodynamic laws, a weak discontinuous Galerkin (DG) form has been formulated using conjugated fluxes and fields gradients.

As the weak discontinuous form is derived in terms of those energy conjugated fluxes and fields gradients, the resulting DG finite element method is consistent and stable. The numerical properties of the DG method for nonlinear elliptic problems, such as the consistency and uniqueness of the solution have been analyzed in 2D by reformulating the problem in a linearized fixed point form, following the methodology set by previous works [17, 18] for non-linear elliptic problems, herein particularized for thermo-electrical problems.

The numerical verifications have been undertaken to demonstrate the theoretical results. In particular, the convergence rates in the L^2-norm and the H^1-norm with respect to the mesh size are optimal and agree with the error analysis that was derived in the theory.

Finally, a unit cell problem has been solved numerically to illustrate the capability of the algorithm.

In further work, the method will be extended to thermo-electrical-mechanics with a view to the study of hybrid shape memory composites by using the electric current to stimulate the shape memory polymers.

Acknowledgement

This project has been funded with support of the European Commission under the grant number 2012-2624/001-001-EM. This publication reflects the view only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Appendix A. Stiffness matrix

For the stiffness matrix, by recalling the internal forces, Eq. (58), we have

$$
\begin{aligned}
\frac{\partial \mathbf{F}^{\mathrm{a}}_{\mathrm{int}}}{\partial \mathbf{M}^{\mathrm{b}}_{\mathrm{h}}} &= \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \nabla \mathbf{N}^{\mathrm{a}^{\mathrm{T}}} \frac{\partial \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}})}{\partial \mathbf{M}_{\mathrm{h}}} \mathbf{N}^{\mathrm{b}} \mathrm{d}\Omega + \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \nabla \mathbf{N}^{\mathrm{a}^{\mathrm{T}}} \frac{\partial \mathbf{j}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}})}{\partial \nabla \mathbf{M}_{\mathrm{h}}} \nabla \mathbf{N}^{\mathrm{b}} \mathrm{d}\Omega \\
&+ \sum_{\mathrm{e}} \int_{\Omega^{\mathrm{e}}} \mathbf{N}^{\mathrm{a}} \frac{\partial \mathbf{i}}{\partial \mathbf{M}_{\mathrm{h}}} \mathbf{N}^{\mathrm{b}} \mathrm{d}\Omega.
\end{aligned} \tag{A.1}
$$

For the terms related to the interface $\partial_{\mathrm{I}}\Omega_{\mathrm{h}}$[2], Eqs. (60, 61, and 62), we have

$$
\begin{aligned}
\frac{\partial \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I1}}}{\partial \mathbf{M}^{\mathrm{b}\pm}_{\mathrm{h}}} &= \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm \mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-^{\mathrm{T}}} \frac{\partial \mathbf{j}^{\pm}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}})}{\partial \mathbf{M}^{\pm}_{\mathrm{h}}} \mathbf{N}^{\mathrm{b}\pm} \mathrm{dS} \\
&+ \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm \mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-^{\mathrm{T}}} \frac{\partial \mathbf{j}^{\pm}(\mathbf{M}_{\mathrm{h}}, \nabla \mathbf{M}_{\mathrm{h}})}{\partial \nabla \mathbf{M}^{\pm}_{\mathrm{h}}} \nabla \mathbf{N}^{\mathrm{b}\pm} \mathrm{dS},
\end{aligned} \tag{A.2}
$$

$$
\frac{\partial \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I2}}}{\partial \mathbf{M}^{\mathrm{b}\pm}_{\mathrm{h}}} = \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \nabla \mathbf{N}^{\mathrm{a}\pm^{\mathrm{T}}} \mathbf{Z}^{\pm}(\mathbf{M}_{\mathrm{h}}) \bar{\mathbf{n}}^{-} \left(\pm \mathbf{N}^{\mathrm{b}\pm}\right) \mathrm{dS} + \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \nabla \mathbf{N}^{\mathrm{a}\pm^{\mathrm{T}}} \mathcal{Z}^{\pm}(\mathbf{M}_{\mathrm{h}}) \left[\!\left[\mathbf{M}_{\mathrm{h}_{\mathbf{n}}} \right]\!\right] \mathbf{N}^{\mathrm{b}\pm} \mathrm{dS}, \tag{A.3}
$$

where $\mathcal{Z}^{\pm} = \frac{\partial \mathbf{Z}^{\pm}(\mathbf{M}_{\mathrm{h}})}{\partial \mathbf{M}^{\pm}_{\mathrm{h}}}$ is a matrix of size ($2\mathrm{d} \times 2\mathrm{d} \times 2$), which results in a $2\mathrm{d} \times 2$ matrix after multiplying it by $[\![\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!]$, and $(\mathcal{Z}\,[\![\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!])_{\mathrm{IK}} = \sum_{\mathrm{J}=1}^{2\mathrm{d}} (\mathcal{Z}_{\mathrm{IJK}}\,[\![\mathbf{M}_{\mathrm{h}_{\mathbf{n}}}]\!]_{\mathrm{J}})$, and

$$
\begin{aligned}
\frac{\partial \mathbf{F}^{\mathrm{a}\pm}_{\mathrm{I3}}}{\partial \mathbf{M}^{\mathrm{b}\pm}_{\mathrm{h}}} &= \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm \mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-^{\mathrm{T}}} \left\langle \frac{\mathbf{Z}(\mathbf{M}_{\mathrm{h}})\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \right\rangle \bar{\mathbf{n}}^{-} (\pm \mathbf{N}^{\mathrm{b}\pm}) \mathrm{dS} \\
&+ \frac{1}{2} \sum_{\mathrm{s}} \int_{(\partial_{\mathrm{I}}\Omega)^{\mathrm{s}}} \left(\pm \mathbf{N}^{\mathrm{a}\pm}\right) \bar{\mathbf{n}}^{-^{\mathrm{T}}} \mathcal{Z}^{\pm}(\mathbf{M}_{\mathrm{h}}) \frac{\mathcal{B}}{\mathrm{h}_{\mathrm{s}}} \left[\!\left[\mathbf{M}_{\mathrm{h}_{\mathbf{n}}} \right]\!\right] \mathbf{N}^{\mathrm{b}\pm} \mathrm{dS}.
\end{aligned} \tag{A.4}
$$

Appendix B. General properties for finite element space

Lemma Appendix B.1 (Interpolation inequality)**.** *For $\boldsymbol{M} \in H^s(\Omega^e) \times H^s(\Omega^e)$ there exists a sequence $\boldsymbol{M}^h \in \mathbb{P}^k(\Omega^e) \times \mathbb{P}^k(\Omega^e)$and a positive constant $C^k_{\mathcal{D}}$ depending on s and k but independent of $\boldsymbol{M}$ and h_s, such that*

1. *for any $0 \leq n \leq s$*

$$
\parallel \boldsymbol{M} - \boldsymbol{M}^h \parallel_{H^n(\Omega^e)} \leq C^k_{\mathcal{D}} h_s^{\mu - n} \parallel \boldsymbol{M} \parallel_{H^s(\Omega^e)}, \tag{B.1}
$$

2. *for any $0 \leq n \leq s - 1 + \frac{2}{r}$*

$$
\parallel \boldsymbol{M} - \boldsymbol{M}^h \parallel_{W^n_r(\Omega^e)} \leq C^k_{\mathcal{D}} h_s^{\mu - n - 1 + \frac{2}{r}} \parallel \boldsymbol{M} \parallel_{H^s(\Omega^e)}, \quad \textit{if } d = 2\,, \tag{B.2}
$$

[2] The contributions on $\partial_{\mathrm{D}}\Omega_{\mathrm{h}}$ can be directly deduced by removing the factor (1/2) accordingly to the definition of the average flux on the Dirichlet boundary and $\mathbf{Z}(\bar{\mathbf{M}})$, which is constant with respect to $\mathbf{M}_{\mathrm{h}}$, instead of $\mathbf{Z}(\mathbf{M}_{\mathrm{h}})$.

3. *for any* $s > n + \frac{1}{2}$

$$\parallel \boldsymbol{M} - \boldsymbol{M}^h \parallel_{H^n(\partial\Omega^e)} \leq C_{\mathcal{D}}^k h_s^{\mu - n - \frac{1}{2}} \parallel \boldsymbol{M} \parallel_{H^s(\Omega^e)}, \tag{B.3}$$

where $\mu = min\{s, k+1\}$.

The proof of the first and third properties can be found in [26], and the proof of the second property in the particular case of d = 2 can be found in [27, 28], see also the discussion by [18].

Remarks

i) The approximation property in (2) is still valid for r = ∞, see [29].

ii) For $\mathbf{M} \in \mathrm{X_s}$, let us define the interpolant $\mathrm{I_h}\mathbf{M} \in \mathrm{X^k}$ by $\mathrm{I_h}\mathbf{M}|_{\Omega^e} = \mathbf{M}^\mathrm{h}(\mathbf{M}|_{\Omega^e})$, which means $\mathrm{I_h}\mathbf{M}$ satisfies the interlation inequality property provided in Lemma Appendix B.1, see [23].

Lemma Appendix B.2 (Trace inequality)**.** *For all* $\boldsymbol{M} \in H^{s+1}(\Omega^e) \times H^{s+1}(\Omega^e)$, *there exists a positive constant* $C_{\mathcal{T}}$, *such that*

$$\parallel \boldsymbol{M} \parallel^r_{W^s_r(\partial\Omega^e)} \leq C_{\mathcal{T}} \left(\frac{1}{h_s} \parallel \boldsymbol{M} \parallel^r_{W^s_r(\Omega^e)} + \parallel \boldsymbol{M} \parallel^{r-1}_{W^s_{2r-2}(\Omega^e)} \parallel \nabla^{s+1} \boldsymbol{M} \parallel_{L^2(\Omega^e)} \right), \tag{B.4}$$

where $s = 0, 1$ *and* $r = 2, 4$, *or in other words*

$$\begin{aligned} \parallel \boldsymbol{M} \parallel^2_{L^2(\partial\Omega^e)} &\leq C_{\mathcal{T}} \left(\frac{1}{h_s} \parallel \boldsymbol{M} \parallel^2_{L^2(\Omega^e)} + \parallel \boldsymbol{M} \parallel_{L^2(\Omega^e)} \parallel \nabla \boldsymbol{M} \parallel_{L^2(\Omega^e)} \right), \\ \parallel \boldsymbol{M} \parallel^4_{L^4(\partial\Omega^e)} &\leq C_{\mathcal{T}} \left(\frac{1}{h_s} \parallel \boldsymbol{M} \parallel^4_{L^4(\Omega^e)} + \parallel \boldsymbol{M} \parallel^3_{L^6(\Omega^e)} \parallel \nabla \boldsymbol{M} \parallel_{L^2(\Omega^e)} \right). \end{aligned} \tag{B.5}$$

The first equation, s = 0 and r = 2, is proved in [8], and the second one, r = 4 and s = 0, is proved in [30].

Lemma Appendix B.3 (Trace inequality on the finite element space)**.** *For* $\boldsymbol{M}_h \in \mathbb{P}^k(\Omega^e) \times \mathbb{P}^k(\Omega^e)$ *Then there is a constant* $C^k_{\mathcal{K}} > 0$ *depending on k, such that*

$$\parallel \nabla^l \boldsymbol{M}_h \parallel_{L^2(\partial\Omega^e)} \leq C^k_{\mathcal{K}} h_s^{-\frac{1}{2}} \parallel \nabla^l \boldsymbol{M}_h \parallel_{L^2(\Omega^e)} \qquad l = 0,\ 1, \tag{B.6}$$

where $C^k_{\mathcal{K}} = sup_{\upsilon \in P_K(\Omega^e)} \frac{h_s \|\nabla \boldsymbol{M}_h\|^2_{L^2(\partial\Omega^e)}}{\|\nabla \boldsymbol{M}_h\|^2_{L^2(\Omega^e)}}$ *is a constant which depends on the degree of the polynomial approximation only with* $h_s = \frac{|\Omega^e|}{|\partial\Omega^e|}$.

We refer to [31] for more details.

Lemma Appendix B.4 (Inverse inequality)**.** *For* $\boldsymbol{M}_h \in \mathbb{P}^k(\Omega^e) \times \mathbb{P}^k(\Omega^e)$ *and* $r \geq 2$, *there exists* $C^k_{\mathcal{I}} > 0$, *such that*

$$\parallel \boldsymbol{M}_h \parallel_{L^r(\Omega^e)} \leq C^k_{\mathcal{I}} h_s^{\frac{d}{r} - \frac{d}{2}} \parallel \boldsymbol{M}_h \parallel_{L^2(\Omega^e)}, \tag{B.7}$$

$$\parallel \boldsymbol{M}_h \parallel_{L^r(\partial\Omega^e)} \leq C^k_{\mathcal{I}} h_s^{\frac{d-1}{r} - \frac{d-1}{2}} \parallel \boldsymbol{M}_h \parallel_{L^2(\partial\Omega^e)}, \tag{B.8}$$

$$\parallel \nabla \boldsymbol{M}_h \parallel_{L^2(\Omega^e)} \leq C^k_{\mathcal{I}} h_s^{-1} \parallel \boldsymbol{M}_h \parallel_{L^2(\Omega^e)}. \tag{B.9}$$

The proof of these properties can be found in [32, Theorem 3.2.6]. Note that Eqs. (B.7-B.8) involve the space dimension d.

Appendix C. $\mathbf{S_h}$ maps the ball $\mathbf{O_{\boldsymbol{\sigma}}(I_hM)}$ into itself

Let $\mathbf{y} \in \mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$ and $\mathrm{S_h}(\mathbf{y}) = \mathbf{M_y}$ be a solution of the problem given by Eq. (86). Then using Lemma 4.2, Eq. (88), Lemma 4.3, Eq. (90), Lemma 4.7, Eq. (104), and the definition of the ball (100), we successively find that

$$\begin{aligned}
&\mathrm{C}_1^{\mathrm{k}} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,|||^2 - \mathrm{C}_2^{\mathrm{k}} \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} \\
&\leq \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M_y}, \mathrm{I_h}\mathbf{M} - \mathbf{M_y}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M_y}, \mathrm{I_h}\mathbf{M} - \mathbf{M_y}) \\
&\leq \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e}, \mathrm{I_h}\mathbf{M} - \mathbf{M_y}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e}, \mathrm{I_h}\mathbf{M} - \mathbf{M_y}) + \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}, \mathrm{I_h}\mathbf{M} - \mathbf{M_y}) \\
&\leq \mathrm{C}^{\mathrm{k}} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e} \,|||_1 \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| + \mathrm{C}^{\mathrm{k}}\mathrm{C_y} \,\|\, \mathbf{M}^\mathrm{e} \,\|_{\mathrm{H}^s(\Omega_\mathrm{h})} \, \mathrm{h_s}^{\mu-2-\varepsilon}\sigma \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| \\
&\leq (\mathrm{C}^{\mathrm{k}}\mathrm{h_s^\varepsilon} + \mathrm{C}^{\mathrm{k}}\mathrm{C_y} \,\|\, \mathbf{M}^\mathrm{e} \,\|_{\mathrm{H}^s(\Omega_\mathrm{h})} \, \mathrm{h_s}^{\mu-2-\varepsilon})\sigma \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| \,.
\end{aligned} \tag{C.1}$$

Let us define $\mathrm{C}^{\mathrm{k}'}(\mathrm{C}^{\mathrm{k}}, \mathrm{C_y}, \mathrm{C_M})$ a constant, that can depend on C^{k}, $\mathrm{C_y}$ and $\mathrm{C_M}$, then, as $0 < \varepsilon < \frac{1}{4}$, the last expression can be rewritten for $k \geq 2$:

$$\mathrm{C}_1^{\mathrm{k}} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,|||^2 - \mathrm{C}_2^{\mathrm{k}} \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} \leq \mathrm{C}^{\mathrm{k}'}\sigma\mathrm{h_s^\varepsilon} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| \,. \tag{C.2}$$

Then, in order to estimate $\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|_{\mathrm{L}^2(\Omega_\mathrm{h})}$, we consider the auxiliary problem defined in Lemma 4.5. Choosing $\boldsymbol{\xi} = \delta\mathbf{M}_\mathrm{h} = \mathrm{I_h}\mathbf{M} - \mathbf{M_y}$, there exists $\boldsymbol{\phi}_\mathrm{h}$ such that, $|||\, \boldsymbol{\phi}_\mathrm{h} \,||| \leq \mathrm{C}^{\mathrm{k}} \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|_{\mathrm{L}^2(\Omega)}$ with

$$\begin{aligned}
\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} &= \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M_y}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M_y}, \boldsymbol{\phi}_\mathrm{h}) \\
&\leq \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}; \boldsymbol{\phi}_\mathrm{h}) \\
&\leq \mathrm{C}^{\mathrm{k}} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M}^\mathrm{e} \,|||_1 |||\, \boldsymbol{\phi}_\mathrm{h} \,||| + \mathrm{C}^{\mathrm{k}}\mathrm{C_y} \,\|\, \mathbf{M}^\mathrm{e} \,\|_{\mathrm{H}^s(\Omega_\mathrm{h})} \, \mathrm{h_s}^{\mu-2-\varepsilon}\sigma \,|||\, \boldsymbol{\phi}_\mathrm{h} \,||| \\
&\leq (\mathrm{C}^{\mathrm{k}}\sigma\mathrm{h_s^\varepsilon} + \mathrm{C}^{\mathrm{k}}\mathrm{C_y} \,\|\, \mathbf{M}^\mathrm{e} \,\|_{\mathrm{H}^s(\Omega_\mathrm{h})} \, \sigma\mathrm{h_s}^{\mu-2-\varepsilon}) \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|_{\mathrm{L}^2(\Omega_\mathrm{h})} \\
&\leq \mathrm{C}^{\mathrm{k}'}\sigma\mathrm{h_s^\varepsilon} \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|_{\mathrm{L}^2(\Omega_\mathrm{h})} \ \text{ if k} \geq 2.
\end{aligned} \tag{C.3}$$

Substituting Eq. (C.3) in Eq. (C.2) gives

$$\begin{aligned}
\mathrm{C}_1^{\mathrm{k}} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,|||^2 &\leq \mathrm{C}^{\mathrm{k}'}\sigma\mathrm{h_s^\varepsilon} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| + \mathrm{C}_2^{\mathrm{k}} \,\|\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,\|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} \\
&\leq \mathrm{C}^{\mathrm{k}'}\sigma\mathrm{h_s^\varepsilon} \,|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| + \mathrm{C}_2^{\mathrm{k}}(\mathrm{C}^{\mathrm{k}'})^2\sigma^2\mathrm{h_s^{2\varepsilon}} \text{ if k} \geq 2.
\end{aligned} \tag{C.4}$$

Hence, we get

$$|||\, \mathrm{I_h}\mathbf{M} - \mathbf{M_y} \,||| \leq \mathrm{C}^{\mathrm{k}'}\sigma\mathrm{h_s^\varepsilon} \text{ if k} \geq 2, \tag{C.5}$$

and for a mesh size $\mathrm{h_s}$ small enough and a given ball size σ, $\mathrm{I_h}\mathbf{M} - \mathbf{M_y} \longrightarrow 0$, hence $\mathrm{S_h}$ maps $\mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M})$ to itself.

Appendix D. The continuity of the map $\mathbf{S_h}$ in the ball $\mathbf{O_{\boldsymbol{\sigma}}(I_hM)}$

For $\mathbf{y}_1, \mathbf{y}_2 \in \mathrm{O}_\sigma(\mathrm{I_h}\mathbf{M}) \subset \mathrm{X}^{\mathrm{k}^+}$, let $\mathbf{M}_{\mathbf{y}_1} = \mathrm{S_h}(\mathbf{y}_1)$, $\mathbf{M}_{\mathbf{y}_2} = \mathrm{S_h}(\mathbf{y}_2)$ be the solutions of the linearized problem stated by Eq. (86), which satisfy

$$\begin{aligned}
&\mathcal{A}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_1}, \delta\mathbf{M}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathrm{I_h}\mathbf{M} - \mathbf{M}_{\mathbf{y}_1}, \delta\mathbf{M}_\mathrm{h}) \\
&= \mathcal{A}(\mathbf{M}^\mathrm{e}; \boldsymbol{\eta}, \boldsymbol{\delta}\mathbf{M}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \boldsymbol{\eta}, \delta\mathbf{M}_\mathrm{h}) + \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}_1; \delta\mathbf{M}_\mathrm{h}) \ \ \forall\delta\mathbf{M}_\mathrm{h} \in \mathrm{X}^\mathrm{k},
\end{aligned} \tag{D.1}$$

and

$$\begin{aligned}&\mathcal{A}(\mathbf{M}^{\mathrm{e}};\mathrm{I_h}\mathbf{M}-\mathbf{M}_{\mathbf{y}_2},\delta\mathbf{M}_{\mathrm{h}})+\mathcal{B}(\mathbf{M}^{\mathrm{e}};\mathrm{I_h}\mathbf{M}-\mathbf{M}_{\mathbf{y}_2},\delta\mathbf{M}_{\mathrm{h}})\\&=\mathcal{A}(\mathbf{M}^{\mathrm{e}};\boldsymbol{\eta},\delta\mathbf{M}_{\mathrm{h}})+\mathcal{B}(\mathbf{M}^{\mathrm{e}};\boldsymbol{\eta},\delta\mathbf{M}_{\mathrm{h}})+\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_2;\delta\mathbf{M}_{\mathrm{h}})\ \ \forall\delta\mathbf{M}_{\mathrm{h}}\in\mathrm{X}^{\mathrm{k}},\end{aligned}\tag{D.2}$$

where $\boldsymbol{\eta}=\mathrm{I_h}\mathbf{M}-\mathbf{M}^{\mathrm{e}}$. By subtracting Eq. (D.1) from Eq. (D.2), we have

$$\mathcal{A}(\mathbf{M}^{\mathrm{e}};\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1},\delta\mathbf{M}_{\mathrm{h}})+\mathcal{B}(\mathbf{M}^{\mathrm{e}};\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1},\delta\mathbf{M}_{\mathrm{h}})=\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_2;\delta\mathbf{M}_{\mathrm{h}})-\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_1;\delta\mathbf{M}_{\mathrm{h}}).\tag{D.3}$$

Choosing $\boldsymbol{\zeta}_1=\mathbf{M}^{\mathrm{e}}-\mathbf{y}_1\in\mathrm{X}$ and $\boldsymbol{\zeta}_2=\mathbf{M}^{\mathrm{e}}-\mathbf{y}_2\in\mathrm{X}$, the right hand side of Eq. (D.3) can be rewritten as follows:

$$\begin{aligned}\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_2;\delta\mathbf{M}_{\mathrm{h}})-\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_1;\delta\mathbf{M}_{\mathrm{h}})=&\int_{\Omega_{\mathrm{h}}}\nabla\delta\mathbf{M}_{\mathrm{h}}^{\mathrm{T}}\left(\bar{\mathbf{R}}_{\mathbf{j}}(\boldsymbol{\zeta}_2,\nabla\boldsymbol{\zeta}_2)-\bar{\mathbf{R}}_{\mathbf{j}}(\boldsymbol{\zeta}_1,\nabla\boldsymbol{\zeta}_1)\right)\mathrm{d}\Omega\\&+\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}\cup\partial_{\mathrm{D}}\Omega_{\mathrm{h}}}[\![\delta\mathbf{M}_{\mathrm{h_n}}^{\mathrm{T}}]\!]\left\langle\bar{\mathbf{R}}_{\mathbf{j}}(\boldsymbol{\zeta}_2,\nabla\boldsymbol{\zeta}_2)-\bar{\mathbf{R}}_{\mathbf{j}}(\boldsymbol{\zeta}_1,\nabla\boldsymbol{\zeta}_1)\right\rangle\mathrm{dS}\\&+\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}}\left[\!\!\left[\mathbf{M}^{\mathrm{e}^{\mathrm{T}}}-\mathbf{y}_{2_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right]\langle(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}})-\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{y}_2))\,\nabla\delta\mathbf{M}_{\mathrm{h}}\rangle\,\mathrm{dS}\\&-\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}}\left[\!\!\left[\mathbf{M}^{\mathrm{e}^{\mathrm{T}}}-\mathbf{y}_{1_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right]\langle(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}})-\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{y}_1))\,\nabla\delta\mathbf{M}_{\mathrm{h}}\rangle\,\mathrm{dS}\\&+\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}}\left[\!\!\left[\mathbf{M}^{\mathrm{e}^{\mathrm{T}}}-\mathbf{y}_{2_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right]\left\langle\frac{\mathcal{B}}{\mathrm{h_s}}(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}})-\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{y}_2))\right\rangle[\![\delta\mathbf{M}_{\mathrm{h_n}}]\!]\,\mathrm{dS}\\&-\int_{\partial_{\mathrm{I}}\Omega_{\mathrm{h}}}\left[\!\!\left[\mathbf{M}^{\mathrm{e}^{\mathrm{T}}}-\mathbf{y}_{1_{\mathbf{n}}}^{\mathrm{T}}\right]\!\!\right]\left\langle\frac{\mathcal{B}}{\mathrm{h_s}}(\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{M}^{\mathrm{e}})-\mathbf{j}_{\nabla\mathbf{M}}(\mathbf{y}_1))\right\rangle[\![\delta\mathbf{M}_{\mathrm{h_n}}]\!]\,\mathrm{dS}.\end{aligned}\tag{D.4}$$

By applying Taylor series, Eqs. (72-75), to rewrite the right hand side, every term will be either in $\mathbf{y}_1-\mathbf{y}_2$ or in $\nabla(\mathbf{y}_1-\mathbf{y}_2)$. So proceeding as to establish Lemma 4.7 and using Lemma 4.1, Eq. (66), lead to [18]

$$\mid\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_2;\delta\mathbf{M}_{\mathrm{h}})-\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_1;\delta\mathbf{M}_{\mathrm{h}})\mid\leq\mathrm{C^kC_y}\parallel\mathbf{M}^{\mathrm{e}}\parallel_{\mathrm{H^s}(\Omega_{\mathrm{h}})}\mathrm{h_s^{\mu-2-\varepsilon}}\mid\mid\mid\mathbf{y}_1-\mathbf{y}_2\mid\mid\mid\,\mid\mid\mid\delta\mathbf{M}_{\mathrm{h}}\mid\mid\mid.\tag{D.5}$$

Choosing $\delta\mathbf{M}_{\mathrm{h}}=\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}$, and using Eq. (88), Eq. (D.3) becomes:

$$\begin{aligned}\mathrm{C_1^k}\mid\mid\mid\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\mid\mid\mid^2-\mathrm{C_2^k}\parallel\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\parallel^2_{\mathrm{L^2}(\Omega_{\mathrm{h}})}\leq\ &\mathcal{A}(\mathbf{M}^{\mathrm{e}};\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1},\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1})\\&+\mathcal{B}(\mathbf{M}^{\mathrm{e}};\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1},\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1})\\&\leq\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_2;\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1})-\mathcal{N}(\mathbf{M}^{\mathrm{e}},\mathbf{y}_1;\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}).\end{aligned}\tag{D.6}$$

Similarly, setting $\delta\mathbf{M}_{\mathrm{h}}=\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}$ in Eq. (D.5), Eq (D.6) becomes:

$$\mid\mid\mid\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\mid\mid\mid^2\leq\mathrm{C_1^kC_y}\parallel\mathbf{M}^{\mathrm{e}}\parallel_{\mathrm{H^s}(\Omega_{\mathrm{h}})}\mathrm{h_s^{\mu-2-\varepsilon}}\mid\mid\mid\mathbf{y}_2-\mathbf{y}_1\mid\mid\mid\,\mid\mid\mid\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\mid\mid\mid+\mathrm{C_2^k}\parallel\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\parallel^2_{\mathrm{L^2}(\Omega_{\mathrm{h}})}.\tag{D.7}$$

Since $\parallel\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\parallel^2_{\mathrm{L^2}(\Omega_{\mathrm{h}})}\leq\mid\mid\mid\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\mid\mid\mid\,\parallel\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\parallel_{\mathrm{L^2}(\Omega_{\mathrm{h}})}$, this last relation becomes

$$\mid\mid\mid\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\mid\mid\mid\leq\mathrm{C_1^kC_y}\parallel\mathbf{M}^{\mathrm{e}}\parallel_{\mathrm{H^s}(\Omega_{\mathrm{h}})}\mathrm{h_s^{\mu-2-\varepsilon}}\mid\mid\mid\mathbf{y}_2-\mathbf{y}_1\mid\mid\mid+\mathrm{C_2^k}\parallel\mathbf{M}_{\mathbf{y}_2}-\mathbf{M}_{\mathbf{y}_1}\parallel_{\mathrm{L^2}(\Omega_{\mathrm{h}})}.\tag{D.8}$$

In order to estimate $\| \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1} \|^2_{\mathrm{L}^2(\Omega_\mathrm{h})}$, we consider $\boldsymbol{\xi} = \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1}$ in Lemma 4.5. Therefore, there exists a unique $\boldsymbol{\phi}_\mathrm{h}$ satisfying Eq. (95) $\forall \delta\mathbf{M}_\mathrm{h} \in \mathrm{X}^\mathrm{k}$. In particular for $\delta\mathbf{M}_\mathrm{h} = \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1}$, this implies

$$\begin{aligned} \| \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1} \|^2_{\mathrm{L}^2(\Omega_\mathrm{h})} &= \mathcal{A}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1}, \boldsymbol{\phi}_\mathrm{h}) + \mathcal{B}(\mathbf{M}^\mathrm{e}; \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1}, \boldsymbol{\phi}_\mathrm{h}) \\ &= \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}_2; \boldsymbol{\phi}_\mathrm{h}) - \mathcal{N}(\mathbf{M}^\mathrm{e}, \mathbf{y}_1; \boldsymbol{\phi}_\mathrm{h}) \\ &\leq \mathrm{C}^\mathrm{k}\mathrm{C}_\mathrm{y} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^s(\Omega_\mathrm{h})} \mathrm{h}_\mathrm{s}^{\mu-2-\varepsilon} \, ||| \, \mathbf{y}_2 - \mathbf{y}_1 \, |||||| \, \boldsymbol{\phi}_\mathrm{h} \, ||| \\ &\leq \mathrm{C}^\mathrm{k}\mathrm{C}_\mathrm{y} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^s(\Omega_\mathrm{h})} \mathrm{h}_\mathrm{s}^{\mu-2-\varepsilon} \, ||| \, \mathbf{y}_2 - \mathbf{y}_1 \, ||| \, \| \mathbf{M}_{\mathbf{y}_2} - \mathbf{M}_{\mathbf{y}_1} \|_{\mathrm{L}^2(\Omega_\mathrm{h})}, \end{aligned} \tag{D.9}$$

where we have used Eq (D.3), Eq. (D.5), and Eq. (96). Therefore, substituting Eq. (D.9) in Eq. (D.8) yields

$$||| \, \mathbf{M}_{\mathbf{y}_1} - \mathbf{M}_{\mathbf{y}_2} \, ||| \leq \mathrm{C}^\mathrm{k}\mathrm{C}_\mathrm{y} \| \mathbf{M}^\mathrm{e} \|_{\mathrm{H}^s(\Omega_\mathrm{h})} \mathrm{h}_\mathrm{s}^{\mu-2-\varepsilon} \, ||| \, \mathbf{y}_1 - \mathbf{y}_2 \, ||| \, . \tag{D.10}$$

References

[1] D. Ebling, M. Jaegle, M. Bartel, A. Jacquot, H. Böttner, Multiphysics simulation of thermoelectric systems for comparison with experimental device performance, Journal of electronic materials 38 (7) (2009) 1456–1461.

[2] Y. Y. Hsiao, W. C. Chang, S. L. Chen, A Mathematic Model of Thermoelectric Module with Applications on Waste Heat Recovery from Automobile Engine, Energy 35 (3) (2010) 1447–1454.

[3] J. L. Pérez-Aparicio, R. Palma, R. L. Taylor, Finite element analysis and material sensitivity of Peltier thermoelectric cells coolers, International Journal of Heat and Mass Transfer 55 (4) (2012) 1363–1374.

[4] G. D. Mahan, Density variations in thermoelectrics, Journal of Applied Physics 87 (10) (2000) 7326–7332.

[5] J. L. Pérez-Aparicio, R. L. Taylor, D. Gavela, Finite element analysis of nonlinear fully coupled thermoelectric materials, Computational Mechanics 40 (1) (2007) 35–45.

[6] L. Liu, A continuum theory of thermoelectric bodies and effective properties of thermoelectric composites, International Journal of Engineering Science 55 (2012) 35–53.

[7] Y. Yang, S. Xie, F. Ma, J. Li, On the effective thermoelectric properties of layered heterogeneous medium, Journal of Applied Physics 111 (1) (2012) 013510.

[8] S. Prudhomme, F. Pascal, J. Oden, A. Romkes, Review of a priori error estimation for discontinuous Galerkin methods, Tech. Rep., TICAM, UTexas, 2000.

[9] L. Noels, R. Radovitzky, A general discontinuous Galerkin method for finite hyperelasticity. Formulation and numerical applications, International Journal for Numerical Methods in Engineering 68 (1) (2006) 64–97.

[10] A. Romkes, S. Prudhomme, J. Oden, A priori error analyses of a stabilized discontinuous Galerkin method, Computers & Mathematics with Applications 46 (8) (2003) 1289–1311.

[11] W. H. Reed, T. Hill, Triangular mesh methods for the neutron transport equation, Los Alamos Report LA-UR-73-479, `http://www.osti.gov/scitech/servlets/purl/4491151`, 1973.

[12] B. Cockburn, G. E. Karniadakis, C. W. Shu, The development of discontinuous Galerkin methods, Springer, 2000.

[13] J. Douglas, T. Dupont, Interior Penalty Procedures for Elliptic and Parabolic Galerkin Methods, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-37550-0, 207–216, doi:\bibinfo{doi}{10.1007/BFb0120591}, URL `http://dx.doi.org/10.1007/BFb0120591`, 1976.

[14] M. F. Wheeler, An elliptic collocation-finite element method with interior penalties, SIAM Journal on Numerical Analysis 15 (1) (1978) 152–161.

[15] D. N. Arnold, F. Brezzi, B. Cockburn, L. D. Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, SIAM journal on numerical analysis 39 (5) (2002) 1749–1779.

[16] E. H. Georgoulis, Discontinuous Galerkin methods on shape-regular and anisotropic meshes, University of Oxford D. Phil. Thesis, 2003.

[17] S. Yadav, A. Pani, E. J. Park, Superconvergent discontinuous Galerkin methods for nonlinear elliptic equations, Mathematics of Computation 82 (283) (2013) 1297–1335.

[18] T. Gudi, N. Nataraj, A. K. Pani, hp-Discontinuous Galerkin methods for strongly nonlinear elliptic boundary value problems, Numerische Mathematik 109 (2) (2008) 233–268.

[19] S. Sun, M. F. Wheeler, Discontinuous Galerkin methods for coupled flow and reactive transport problems, Applied Numerical Mathematics 52 (2) (2005) 273–298.

[20] X. P. Zheng, D. H. Liu, Y. Liu, Thermoelastic coupling problems caused by thermal contact resistance: A discontinuous Galerkin finite element approach, Science China Physics, Mechanics and Astronomy 54 (4) (2011) 666–674.

[21] S. Kesavan, Topics in functional analysis and applications, John Wiley & Sons, 1989.

[22] O. C. Zienkiewicz, R. L. Taylor, The finite element method: Solid mechanics, vol. 2, Butterworth-heinemann, 2000.

[23] P. Houston, J. Robson, E. Süli, Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems I: The scalar case, IMA journal of numerical analysis 25 (4) (2005) 726–749.

[24] L. Homsi, Development of non-linear Electro-Thermo-Mechanical Discontinuous Galerkin formulations, Ph.D. thesis, University of Liege, Belgium, 2017.

[25] D. Gilbarg, N. S. Trudinger, Elliptic partial differential equations of second order, Springer, 1983.

[26] I. Babuška, M. Suri, The *hp* version of the finite element method with quasiuniform meshes, RAIRO-Modélisation mathématique et analyse numérique 21 (2) (1987) 199–238.

[27] M. Ainsworth, D. Kay, The approximation theory for the p-version finite element method and application to non-linear elliptic PDEs, Numerische Mathematik 82 (3) (1999) 351–388.

[28] M. Ainsworth, D. Kay, Approximation theory for the hp-version finite element method and application to the non-linear Laplacian, Applied numerical mathematics 34 (4) (2000) 329–344.

[29] D. K. M. Ainsworth, The approximation theory for the p-version finite element method and application to the nonlinear elliptic PDEs, Numer. Math 82 (1999) 351–388.

[30] T. Gudi, Discontinuous Galerkin Methods for nonlinear elliptic problems, Ph.D. thesis, Indian Institute of Technology, Bombay, 2006.

[31] P. Hansbo, M. G. Larson, Discontinuous Galerkin methods for incompressible and nearly incompressible elasticity by Nitsche's method, Computer methods in applied mechanics and engineering 191 (17) (2002) 1895–1908.

[32] P. Ciarlet, Conforming Finite Element Methods for Second-Order Problems, chap. 3, SIAM, 110–173, doi:\bibinfo{doi}{10.1137/1.9780898719208.ch3}, URL `http://epubs.siam.org/doi/abs/10.1137/1.9780898719208.ch3`, 2002.