DG APPROACH TO LARGE BENDING PLATE DEFORMATIONS WITH ISOMETRY CONSTRAINT

ANDREA BONITO[1], RICARDO H. NOCHETTO[2], AND DIMITRIOS NTOGKAS[3]

ABSTRACT. We propose a new discontinuous Galerkin (dG) method for a geometrically nonlinear Kirchhoff plate model for large isometric bending deformations. The minimization problem is nonconvex due to the isometry constraint. We present a practical discrete gradient flow that decreases the energy and computes discrete minimizers that satisfy a prescribed discrete isometry defect. We prove Γ-convergence of the discrete energies and discrete global minimizers. We document the flexibility and accuracy of the dG method with several numerical experiments.

1. INTRODUCTION

Large bending deformations of thin plates is a critical feature for many modern engineering applications due to the extensive use of plate actuators in a variety of systems like thermostats, nano-tubes, micro-robots and micro-capsules [9, 24, 27, 33, 34]. From the mathematical viewpoint, there is an increasing interest in the modeling and the numerical treatment of such plates. A rigorous analysis of large bending deformations of plates was conducted by Friesecke, James and Müller [23], who derived the geometrically non-linear Kirchhoff model from three dimensional hyperelasticity via Γ-convergence. Since then, there have been various other interesting models, such as the models of prestrained plates derived in [10, 28]. Previous work on the numerical treatment of large bending deformations includes the single layer problem by Bartels [5], the bilayer problem by Bartels, Bonito and Nochetto [8] and the modeling and simulation of thermally actuated bilayer plates by Bartels, Bonito, Muliana and Nochetto [7]. In all three approaches [5, 7, 8] the model involves minimizing an energy functional that is dominated by the Hessian of the deformation $y : \Omega \to \mathbb{R}^3$ of the mid-plane $\Omega \subset \mathbb{R}^2$ of the undeformed plate. Given functions (g, Φ), the minimization takes place under Dirichlet boundary conditions for $y = g$ and $\nabla y = \Phi$ on part of the boundary $\partial_D \Omega$ of $\partial \Omega$ and the isometry constraint

$$(1.1) \qquad \nabla y^T \nabla y = I \quad \text{a.e. in } \Omega,$$

where I stands for the identity matrix in $\mathbb{R}^2$. The authors of [5, 7, 8] employ Kirchhoff elements in order to impose the isometry constraint at the nodes of the

triangulation and rely on an H^2- gradient flow that allows them to construct solutions of decreasing discrete energy. In [5, 8] it is also proved that the discrete energy Γ-converges to the continuous one. Since for fourth order problems a conforming approach would be very costly, the Kirchhoff elements offer a natural non-conforming space for the model that allows the imposition of (1.1) nodewise.

1.1. **Our contribution.** In this paper we focus on the single layer problem, as in [5], in order to investigate the applicability of a more flexible approach that hinges on a non-conforming space of discontinuous functions. We use interior penalty terms for the discrete energy along with a Nitsche's approach to enforce the boundary conditions in the limit. We start with the Dirichlet and forcing data

$$(1.2) \qquad g \in [H^1(\Omega)]^3, \quad \Phi \in [H^1(\Omega)]^{3\times 2}, \quad f \in [L^2(\Omega)]^3,$$

and the affine manifold of $[H^2(\Omega)]^3$

$$(1.3) \qquad \mathbb{V}(g, \Phi) := \left\{ v \in [H^2(\Omega)]^3 : v = g, \ \nabla v = \Phi \text{ on } \partial_D\Omega \right\},$$

where $\partial_D\Omega$ is an open non-empty subset of the boundary $\partial\Omega$. We wish to approximate a minimizer $y : \Omega \to \mathbb{R}^3$ of the continuous energy

$$(1.4) \qquad E[y] = \frac{1}{2}\int_\Omega |D^2 y|^2 - \int_\Omega f \cdot y$$

in the nonconvex set of admissible functions

$$(1.5) \qquad \mathbb{A}(g, \Phi) := \left\{ y \in \mathbb{V}(g, \Phi) : \ (\nabla y)^T \nabla y = I \text{ a.e. in } \Omega \right\},$$

where $|\cdot|$ denotes the Frobenius norm. To avoid the costly use of a conforming finite element subspace of $[H^2(\Omega)]^3$, we resort to a space $[\mathbb{V}_h^k]^3$ of discontinuous piecewise polynomials of degree $k \geq 2$ over a *shape-regular* but possibly *graded* mesh $\mathcal{T}_h$ (constructed either using the reference unit triangle or unit square). Since our estimates below are all local, hereafter h stands for a mesh density function locally equivalent to the element size. However, to de-emphasize this aspect of our approach in favor of others (non-convexity, Hessian reconstruction, Γ-convergence) and to simplify notation, h written as a parameter signifies the meshsize of $\mathcal{T}_h$ (i.e. the largest element size). We denote by $E_h : [\mathbb{V}_h^k]^3 \to \mathbb{R}$ the discrete energy that approximates E and accounts for the discontinuities of the functions $v_h \in [\mathbb{V}_h^k]^3$ and of their broken (i.e. piecewise) gradients $\nabla_h v_h$; see (2.16).

It is important to notice that the energy E in (1.4) is convex but the isometry constraint (1.1) is not. Therefore, we must approximate (1.1), and thus the admissible set $\mathbb{A}(g, \Phi)$ in a way amenable to computation, as well as construct an algorithm able to find critical points. We define the discrete admissible set $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ to be the set of functions $v_h \in [\mathbb{V}_h^k]^3$ whose boundary jumps include g and Φ (see (2.9) and (2.10) below) and whose *discrete isometry defect* $D_h[v_h]$ satisfies

$$(1.6) \qquad D_h[v_h] := \sum_{T\in\mathcal{T}_h} \left| \int_T (\nabla_h v_h)^T \nabla_h v_h - I \right| \leq \varepsilon,$$

where $\varepsilon = \varepsilon(h) \to 0$ as $h \to 0$. We then search for $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ that minimizes the discrete energy $E_h[y_h]$. To this end, we propose a discrete H^2-gradient flow with fictitious time step τ that can be made arbitrarily small. We show that it gives rise to iterates $\{y_h^n\}_{n=1}^\infty \subset \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ with decreasing discrete energy $E_h[y_h^{n+1}] < E_h[y_h^n]$, whenever $y_h^{n+1} \neq y_h^n$, and guarantees the discrete isometry defect (1.6) for all $n \geq 1$ provided the initial guess $y_h^0 \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ is an approximate isometry such

that $D_h[y_h^0] \leq \tau$. This is achieved by selecting ε proportional to τ, depending on $E_h[y_h^0], g, \Phi, f$.

We also prove Γ-convergence of the discrete energy E_h to the continuous energy E and that global minimizers $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ of E_h converge in $L^2(\Omega)$ to global minimizers $y \in \mathbb{A}(g, \Phi)$ of E as $h \to 0$. A key ingredient for Γ-convergence is reconstruction of a suitable discrete Hessian $H_h[y_h]$ of y_h, which uses the broken Hessian $D_h^2 y_h$ and the jumps $[y_h]$ and $[\nabla_h y_h]$ of y_h and $\nabla_h y_h$ across interelement boundaries. Such $H_h[y_h]$ is an L^2-function in Ω that converges weakly to $D^2 y$ under suitable mesh conditions. A similar approach is employed in [20] for second order problems and later in [31] for the $p-$biharmonic equation, where $p = 2$. We refer to Section 4 for a discussion of properties of $H_h[y_h]$ and critical differences with [20, 31].

Our approach is motivated by the flexibility of dG compared to Kirchhoff elements [5, 6]. First of all, Kirchhoff elements require polynomial degree $k = 3$ and suffer from a complicated implementation that involves a discrete gradient that maps the gradient of the discrete deformation to another space. Since this is not implemented in standard finite element libraries, the above difficulty hinders the impact of the method in the engineering community. In contrast, the proposed dG approach works for $k \geq 2$, does not require such map (or more precisely, the map is the trivial elementwise differentiation) and its implementation is standard. Moreover, Kirchhoff elements are not amenable to adaptively refined partitions, at least in theory; the current dG theory, instead, allows for graded meshes $\mathcal{T}_h$. It is worth mentioning as well that imposing Dirichlet boundary conditions weakly, instead of directly in the admissible set as in [5, 6, 8, 7], allows for more flexibility. In particular, this alleviates the constraints on the construction of the recovery sequence necessary for the Γ-convergence of E_h towards E; see Section 5.3. Upon dropping the penalty term of jumps of ∇y_h across a prescribed curve, dG naturally accommodates configurations with kinks, as is the case of origami. Lastly, imposing the isometry constraint numerically with Kirchhoff elements at each vertex seems to be very rigid at the expense of approximation accuracy. In contrast, our experiments with dG indicate that this approach is more accurate and adjusts better to the geometry of the problem; see Section 6.1.

1.2. **Outline.** We first construct our discrete energy functional E_h in Section 2, and prove key properties for functions $v_h \in [\mathbb{V}_h^k]^3$, including the coercivity of E_h with respect to an appropriate mesh-dependent energy norm. We introduce in Section 3 the discrete H^2-gradient flow and show it is able to find discrete minimizers of E_h that satisfy the desired discrete isometry defect (1.6); this sets the stage for relations between ε, h and τ in (1.6). In Section 4 we define the discrete Hessian $H_h[y_h]$ and derive a bound in L^2 and its weak convergence in L^2. This leads to the proof of Γ-convergence of the discrete energy E_h to the exact energy E, in Section 5, as well as the convergence of global minimizers of E_h to global minimizers of E. In Section 6 we present experiments that corroborate numerically the excellent properties of our method and compare it with the Kirchhoff element approach. In Section 7 we provide some details for our implementation and justify the choice of a discontinuous versus a continuous space of functions $\mathbb{V}_h^k$. We draw conclusions in Section 8 and present some technical estimates for isoparametric maps in the appendix of Section 9.

2. Discrete Energy and Properties of dG

We start this section by providing intuition on the derivation of the discrete energy E_h, but without presenting all the details. We then introduce the discrete dG space $\mathbb{V}_h^k$ for scalar functions along with an interpolation operator $\Pi_h : \prod_{T \in \mathcal{T}_h} H^1(T) \to \mathbb{V}_h^k \cap H^1(\Omega)$, and discuss its properties. We finally prove coercivity of E_h.

2.1. **Continuous energy.** Starting from a three-dimensional hyperelasticity model, dimension reduction as the thickness of the plate decreases to zero leads to the two-dimensional energy functional [5, 6, 23]

$$(2.1) \qquad E[y] = \frac{1}{2} \int_\Omega |H|^2 - \int_\Omega f \cdot y,$$

up to a multiplicative constant for the first term. Hereafter, $y \in \mathbb{A}(g, \Phi)$ is an isometric deformation from $\Omega \subset \mathbb{R}^2$ into $\mathbb{R}^3$, $\nu := \partial_1 y \times \partial_2 y$ is the unit normal and $H = (h_{ij})_{i,j=1}^2 := (\partial_{ij} y \cdot \nu)_{i,j=1}^2$ is the second fundamental form of the deformed plate $y(\Omega)$. The connection between (2.1) and (1.4) follows from (1.1), namely

$$\partial_i y \cdot \partial_j y = \delta_{ij}, \quad i, j = 1, 2,$$

where δ_{ij} is the Kronecker delta. Differentiating with respect to the cartesian coordinates x_1 and x_2 and using simple algebraic manipulations, these relations imply

$$\partial_k y \cdot \partial_{ij} y = 0 \quad \Longrightarrow \quad \partial_{ij} y \parallel \nu = \partial_1 y \times \partial_2 y, \quad i, j, k = 1, 2.$$

This, combined with the definition $h_{ij} = \partial_{ij} y \cdot \nu$, leads in turn to

$$|h_{ij}|^2 = |\partial_{ij} y|^2, \quad i, j = 1, 2.$$

Similarly, using that

$$\partial_1 (\partial_{12} y \cdot \partial_2 y) = 0 \quad \text{and} \quad \partial_2 (\partial_{11} y \cdot \partial_2 y) = 0,$$

we obtain

$$|\partial_{12} y|^2 = \partial_{11} y \cdot \partial_{22} y.$$

Therefore, the isometry property (1.1) yields the pointwise relations

$$(2.2) \qquad |H|^2 = |D^2 y|^2 = |\Delta y|^2,$$

so that the nonlinear expression (2.1) of $E[y]$ coincides with (1.4), namely

$$(2.3) \qquad E[y] = \frac{1}{2} \int_\Omega |D^2 y|^2 - \int_\Omega f \cdot y.$$

The Euler-Lagrange equation for a minimizer $y \in \mathbb{V}(g, \Phi)$ of (2.3) reads

$$(2.4) \qquad \int_\Omega D^2 y : D^2 v = \int_\Omega f \cdot v \quad \forall v \in \mathbb{V}(0, \mathbf{0}),$$

where $\mathbb{V}(g, \Phi)$ is defined in (1.3). If $y = (y_n)_{n=1}^3$, then the strong form of (2.4) is

$$(2.5) \qquad \text{div div } D^2 y_n = \Delta^2 y_n = \sum_{i,j=1}^2 \partial_{jiij} y_n = f_n \quad \text{in } \Omega, \quad n = 1, 2, 3,$$

whereas the natural boundary conditions imposed on $\partial\Omega \setminus \partial_D\Omega$ are

$$\partial_\mu \nabla y_n = D^2 y_n\,\mu = \sum_{i=1}^{2} \partial_{ij} y_n\,\mu_i = 0, \qquad n = 1, 2, 3,$$

(2.6)

$$\partial_\mu \Delta y_n = (\operatorname{div} D^2 y_n)\mu = \sum_{i,j=1}^{2} \partial_{iij} y_n\,\mu_j = 0, \qquad n = 1, 2, 3.$$

Hereafter, μ denotes the outward unit normal to Ω.

2.2. **Discrete energy.** We denote by $\mathbb{P}_k$ (resp. $\mathbb{Q}_k$) the space of polynomial functions of degree at most $k \geq 0$ (resp. at most k on each variable). Moreover, $\widehat{T}$ stands for the reference element, which is either the unit triangle for $\mathbb{P}_k$ or the unit square for $\mathbb{Q}_k$. A generic element $T := F_T(\widehat{T})$ is given by a map $F_T \in [\mathbb{P}^k]^2$ (resp. $[\mathbb{Q}^k]^2$) provided $\widehat{T}$ is the unit triangle (resp. square). Notice that when $k = 1$, F_T is affine for triangles T and bi-linear for quadrilaterals T.

We consider a sequence of meshes $\{\mathcal{T}_h\}_{h>0}$ of Ω made of closed shape regular but possibly graded elements T [19]. We assume that Ω can be exactly represented by such subdivisions, i.e. we do not account in the analysis below for variational crimes induced by the approximation of the boundary $\partial\Omega$. Hereafter, we denote by h a mesh density function which is equivalent locally to the size h_T of T and h_e of an edge e; h written as a parameter also signifies the largest value of h_T. In order to handle hanging nodes (necessary for graded meshes based on quadrilaterals), we assume that all the elements within each domain of influence have comparable diameters (independently of h). We refer to Sections 2.2.4 and 6 of Bonito-Nochetto [11] for precise definitions and properties. At this point, we only point out that sequences of meshes made of quadrilaterals with at most one hanging node per side satisfy this assumption. We append a sub-index h to differential operators to indicate that they are applied element-wise. For instance, $\nabla_h v$ is defined by $\nabla_h v|_T := \nabla v|_T$, $T \in \mathcal{T}_h$.

From now on c and C are generic constants independent of h but perhaps depending on the shape-regularity constant of the mesh sequence $\{\mathcal{T}_h\}_{h>0}$. Also, we use the notation $A \lesssim B$ to indicate $A \leq cB$, where c is a constant independent of h, A and B.

Let $\mathbb{V}_h^k$ be the space of discontinuous functions over the mesh $\mathcal{T}_h$

(2.7) $$\mathbb{V}_h^k := \left\{ v_h \in L^2(\Omega) \;:\; v_h \circ F_T \in \mathbb{P}_k \text{ (resp. } \mathbb{Q}_k) \quad \forall T \in \mathcal{T}_h \right\}$$

of degree $k \geq 2$. We point out that the discrete functions $v_h \in \mathbb{V}_h^k$ are not polynomials on the physical elements T unless $F_T \in [\mathbb{P}^1]^2$. This is consistent with the implementation in the deal.ii library[1] used to obtain the numerical illustrations of Section 6. Similarly, the Hessians of $v_h|_T$ and $v_h \circ F_T$ are not proportional because $D^2 F_T \neq \mathbf{0}$, except when $F_T \in [\mathbb{P}^1]^2$. We account for this delicate issue in Section 9.

We denote by $\mathcal{E}_h^0$ the collection of edges of $\mathcal{T}_h$ contained in Ω and by $\mathcal{E}_h^b$ those contained in $\partial_D\Omega$; hence $\mathcal{E}_h = \mathcal{E}_h^0 \cup \mathcal{E}_h^b$ is the set of active interelement boundaries. We further denote the interior skeleton and the boundary counterpart by

(2.8) $$\Gamma_h^0 := \cup\{e : e \in \mathcal{E}_h^0\}, \quad \Gamma_h^b := \cup\{e : e \in \mathcal{E}_h^b\},$$

and by $\Gamma_h := \Gamma_h^0 \cup \Gamma_h^b$ the full skeleton.

As is customary for dG methods, we need to introduce jumps and averages on edges. For $e \in \mathcal{E}_h^0$ we fix $\mu := \mu_e$ to be one of the two unit normals to e in Ω; this choice is irrelevant for the discussion below. Given $v_h \in \mathbb{V}_h^k$, we denote its broken gradient by $\nabla_h v_h$ and the *jumps* of v_h and $\nabla_h v_h$ across interior edges by

$$(2.9) \qquad [v_h]_e := v_h^- - v_h^+, \quad [\nabla_h v_h]_e := \nabla_h v_h^- - \nabla_h v_h^+, \quad \forall\, e \in \mathcal{E}_h^0,$$

where $v_h^\pm(x) = \lim_{s \to 0^+} v_h(x \pm s\,\mu_e)$ and $x \in e$. In order to deal with Dirichlet boundary data (g, Φ) we resort to a Nitsche's approach; hence we do not impose essential restrictions on the discrete space $[\mathbb{V}_h^k]^3$. However, to simplify the notation later it turns out to be convenient to introduce the discrete sets $\mathbb{V}_h^k(g, \Phi)$ and $\mathbb{V}_h^k(0, \mathbf{0})$ which mimic the continuous counterparts $\mathbb{V}(g, \Phi)$ and $\mathbb{V}(0, \mathbf{0})$ but coincide with $[\mathbb{V}_h^k]^3$. In fact, we say that $v_h \in [\mathbb{V}_h^k]^3$ belongs to $\mathbb{V}_h^k(g, \Phi)$ provided the boundary jumps of v_h are defined to be

$$(2.10) \qquad [v_h]_e := v_h - g, \quad [\nabla_h v_h]_e := \nabla_h v_h - \Phi, \quad \forall\, e \in \mathcal{E}_h^b.$$

We stress that $\|[v_h]\|_{L^2(\Gamma_h^b)}$ and $\|[\nabla_h v_h]\|_{L^2(\Gamma_h^b)} \to 0$ imply $v_h \to g$ and $\nabla_h v_h \to \Phi$ in $L^2(\partial_D \Omega)$ as $h \to 0$; hence the connection between $\mathbb{V}_h^k(g, \Phi)$ and $\mathbb{V}(g, \Phi)$. Therefore, the sets $[\mathbb{V}_h^k]^3$ and $\mathbb{V}_h^k(g, \Phi)$ coincide but the latter is not a space because it carries the additional information of boundary jumps, namely

$$(2.11) \quad \mathbb{V}_h^k(g, \Phi) := \left\{ v_h \in [\mathbb{V}_h^k]^3 : \ [v_h]_e, [\nabla_h v_h]_e \text{ given by } (2.10) \text{ for all } e \in \mathcal{E}_h^b \right\}.$$

In the same vein, we will deal with discrete *test* functions $v_h \in \mathbb{V}_h^k(0, \mathbf{0})$ for which boundary jumps are given by $[v_h]_e := v_h$ and $[\nabla_h v_h]_e := \nabla_h v_h$ for all $e \in \mathcal{E}_h^b$, which is consistent with (2.10) for $g = 0$, $\Phi = \mathbf{0}$. We observe again that the sets $\mathbb{V}_h^k(0, \mathbf{0})$ and $[\mathbb{V}_h^k]^3$ are the same. We will not write the subscript e whenever no confusion arises. On the other hand, the definiton of *average* of $v_h \in [\mathbb{V}_h^k]^3$ across an edge $e \in \mathcal{E}_h$ is independent of Dirichlet conditions and is thus given by

$$(2.12) \quad \{v_h\} := \begin{cases} \frac{1}{2}(v_h^+ + v_h^-) & e \in \mathcal{E}_h^0 \\ v_h^- & e \in \mathcal{E}_h^b \end{cases}, \quad \{\nabla_h v_h\} := \begin{cases} \frac{1}{2}(\nabla_h v_h^+ + \nabla_h v_h^-) & e \in \mathcal{E}_h^0 \\ \nabla_h v_h^- & e \in \mathcal{E}_h^b \end{cases}.$$

Definitions (2.9), (2.10) and (2.12) extend to the *broken energy space*

$$(2.13) \qquad\qquad \mathbb{E}(\mathcal{T}_h) := \prod_{T \in \mathcal{T}_h} H^1(T).$$

Before introducing the discrete energy E_h, we derive the corresponding bilinear form $a_h(\cdot, \cdot)$ in the usual manner. This is the first instance where the definitions of $\mathbb{V}_h^k(g, \Phi)$ and $\mathbb{V}_h^k(0, \mathbf{0})$ become critical. We integrate by parts twice the strong equation (2.5) over elements $T \in \mathcal{T}_h$ against a test function $v_h \in \mathbb{V}_h^k(0, \mathbf{0})$. We assume y to be smooth, whence the jumps $[\partial_\mu \nabla y]_e$ and $[\partial_\mu \Delta y]_e$ vanish on edges $e \in \mathcal{E}_h$, to arrive at

$$\left(f, v_h\right)_{L^2(\Omega)} = \left(D^2 y, D_h^2 v_h\right)_{L^2(\Omega)} - \left(\{\partial_\mu \nabla y\}, [\nabla v_h]\right)_{L^2(\Gamma_h)} + \left(\{\partial_\mu \Delta y\}, [v_h]\right)_{L^2(\Gamma_h)}.$$

We next use that $[y]_e = 0$ and $[\nabla y]_e = 0$ for all edges $e \in \mathcal{E}_h$: for interior edges $e \in \mathcal{E}_h^0$ this is because y is smooth, whereas for boundary edges $e \in \mathcal{E}_h^b$ it is a consequence of $y = g$ and $\nabla y = \Phi$ on $\partial_D \Omega$ and definition (2.10). We can thus

symmetrize the previous equality and add vanishing penalty terms

$$
\begin{aligned}
(f, v_h)_{L^2(\Omega)} = {} & \left(D^2 y, D_h^2 v_h\right)_{L^2(\Omega)} \\
& - \left(\{\partial_\mu \nabla y\}, [\nabla v_h]\right)_{L^2(\Gamma_h)} - \left(\{\partial_\mu \nabla_h v_h\}, [\nabla y]\right)_{L^2(\Gamma_h)} \\
& + \left(\{\partial_\mu \Delta y\}, [v_h]\right)_{L^2(\Gamma_h)} + \left(\{\partial_\mu \Delta_h v_h\}, [y]\right)_{L^2(\Gamma_h)} \\
& + \gamma_1 \left(h^{-1}[\nabla y], [\nabla_h v_h]\right)_{L^2(\Gamma_h)} + \gamma_0 \left(h^{-3/2}[y], [v_h]\right)_{L^2(\Gamma_h)} \\
=: {} & a_h(y, v_h),
\end{aligned}
$$
(2.14)

with penalty parameters $\gamma_0, \gamma_1 > 0$ to be determined. Hereafter, we denote by $(\cdot, \cdot)_{L^2(\mathcal{M})}$ the $L^2(\mathcal{M})$ scalar product and vector versions of it. Since (2.14) reduces to (2.4) for $v_h \in \mathbb{V}(0, \mathbf{0})$, we resort to (2.14) to define the discrete equation for y_h

$$
(2.15) \qquad y_h \in \mathbb{V}_h^k(g, \Phi): \quad a_h(y_h, v_h) = (f, v_h)_{L^2(\Omega)} \quad \forall v_h \in \mathbb{V}_h^k(0, \mathbf{0}).
$$

We point out that the Dirichlet conditions in (1.3) are enforced in the Nitsche's sense. Since a_h is symmetric by construction, we define the discrete energy to be

$$
(2.16) \qquad E_h[y_h] := \frac{1}{2} a_h(y_h, y_h) - (f, y_h)_{L^2(\Omega)},
$$

The first variation $\delta E_h[y_h; v_h]$ of $E_h[y_h]$ in the direction $v_h \in \mathbb{V}_h^k(0, \mathbf{0})$ yields (2.15).

In order for (2.16) to be meaningful with respect to the original minimization problem in (1.4)-(1.5), we define the *discrete admissible set* $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$, a discrete analogue of $\mathbb{A}(g, \Phi)$ in (1.5) that involves the discrete isometry defect D_h, to be
(2.17)

$$
\mathbb{A}_{h,\varepsilon}^k(g, \Phi) := \left\{ y_h \in \mathbb{V}_h^k(g, \Phi): \ D_h[y_h] = \sum_{T \in \mathcal{T}_h} \left| \int_T (\nabla_h y_h)^T \nabla_h y_h - I \right| \le \varepsilon \right\}
$$

with parameter $0 < \varepsilon = \varepsilon(h) \to 0$ as $h \to 0$. We will see in Section 3 that the discrete gradient flow used to construct discrete solutions yields $\varepsilon = Ch$, where C depends on $E_h[y_h^0], g, \Phi, f, y_h^0$ being the initial iterate, and other geometric constants.

2.3. Interpolation onto Continuous Piecewise Polynomials.

For several reasons we need to interpolate from the broken energy space $\mathbb{E}(\mathcal{T}_h)$ defined in (2.13) onto the space $\mathring{\mathbb{V}}_h^k := \mathbb{V}_h^k \cap H^1(\Omega)$ of continuous functions which are polynomials of degree $\le k$ over the reference element $\widehat{T}$. We refer to [3, 11, 13, 15, 17] for such interpolation estimates. We now construct a Clément type interpolation operator $\Pi_h : \mathbb{E}(\mathcal{T}_h) \to \mathring{\mathbb{V}}_h^k$, thereby extending [3, 11] to $\mathbb{E}(\mathcal{T}_h)$, because of its simplicity and fitness with our application of it.

We construct Π_h in two steps. We first compute the local L^2-projection $P_h : \mathbb{E}(\mathcal{T}_h) \to \mathbb{V}_h^k$, which for every $v \in \mathbb{E}(\mathcal{T}_h)$ and $T \in \mathcal{T}_h$ reduces to the equation

$$
(2.18) \qquad P_h v \in \mathbb{V}_h^k(T): \quad \int_T (P_h v - v) w = 0 \quad \forall w \in \mathbb{V}_h^k(T),
$$

where $\mathbb{V}_h^k(D)$ (resp. $\mathring{\mathbb{V}}_h^k(D)$) stands for the restriction of functions in $\mathbb{V}_h^k$ (resp. $\mathring{\mathbb{V}}_h^k$) to $D \subset \Omega$. We next define the Clément interpolation operator $I_h : \mathbb{V}_h^k \to \mathring{\mathbb{V}}_h^k$ of [3, 11] as follows. Given the canonical basis functions $\{\phi_i\}_{i=1}^N$ of $\mathring{\mathbb{V}}_h^k$ with supports

$\{\omega_i\}_{i=1}^N$ associated with nodes $\{x_i\}_{i=1}^N$, we compute the L^2-projection of $v \in \mathbb{V}_h^k$ on stars ω_i

$$(2.19) \qquad v_i \in \mathring{\mathbb{V}}_h^k(\omega_i): \quad \int_{\omega_i} (v - v_i)w = 0 \quad \forall w \in \mathring{\mathbb{V}}_h^k(\omega_i),$$

and define $I_h v := \sum_{i=1}^N v_i(x_i)\phi_i \in \mathring{\mathbb{V}}_h^k$ and

$$(2.20) \qquad\qquad\qquad \Pi_h := I_h \circ P_h : \mathbb{E}(\mathcal{T}_h) \to \mathring{\mathbb{V}}_h^k.$$

If $\mathcal{T}_h$ has hanging nodes, the stars ω_i are related to the notion of domain of influence; we refer to Section 6 of [11] for details. We make the same assumptions as [11] on $\mathcal{T}_h$, which in turn imply that all elements of $\mathcal{T}_h$ within ω_i possess comparable size.

Lemma 2.1 (interpolation). *The interpolation operator Π_h defined in (2.20) is invariant in the space $\mathring{\mathbb{V}}_h^k$ and satisfies the following estimate for all $v \in \mathbb{E}(\mathcal{T}_h)$*

$$(2.21) \quad \|\nabla\Pi_h v\|_{L^2(\Omega)} + \|h^{-1}(v - \Pi_h v)\|_{L^2(\Omega)} \lesssim \|\nabla_h v\|_{L^2(\Omega)} + \|h^{-1/2}[v]\|_{L^2(\Gamma_h^0)},$$

where we recall that Γ_h^0 stands for the interior skeleton of $\mathcal{T}_h$ defined in (2.8).

Proof. The operator Π_h is invariant in $\mathring{\mathbb{V}}_h^k$ because so are P_h and I_h. We next examine properties of P_h, I_h and Π_h separately. Recall that all $T \in \mathcal{T}_h$ are closed.

Step 1: Operator P_h. Since P_h is an elementwise L^2-projection, we easily deduce

$$\|P_h v - v\|_{L^2(T)} + h_T^{1/2}\|P_h v - v\|_{L^2(\partial T)} + h_T\|\nabla P_h v\|_{L^2(T)} \lesssim h_T\|\nabla v\|_{L^2(T)}.$$

Given an interior edge $e \in \mathcal{E}_h^0$, let $T^\pm \in \mathcal{T}_h$ be the two adjacent elements that satisfy $e = T^+ \cap T^-$ and $v^\pm$ be the restrictions of v to $T^\pm$. If $\omega(e) := T^+ \cup T^-$, a simple calculation now shows

$$\|[P_h v]\|_{L^2(e)} \le \|P_h v^+ - v^+\|_{L^2(e)} + \|P_h v^- - v^-\|_{L^2(e)} + \|v^+ - v^-\|_{L^2(e)}$$
$$\lesssim h_e^{1/2}\|\nabla_h v\|_{L^2(\omega(e))} + \|[v]\|_{L^2(e)}.$$

Step 2: Operator I_h. Let $\omega(T)$ be the union of stars ω_i alluded to in (2.19) that intersect T; note that all elements within $\omega(T)$ have comparable size [11]. Moreover, let $\gamma^0(\omega(T))$ and $\gamma_i^0 = \gamma^0(\omega_i)$ be the skeletons of $\omega(T)$ and ω_i (interelement boundaries internal to these sets). We recall (see e.g. [11]) that to prove the estimate

$$\|\nabla I_h v\|_{L^2(T)} + \|h^{-1}(I_h v - v)\|_{L^2(T)} \lesssim \|\nabla_h v\|_{L^2(\omega(T))} + \|h^{-1/2}[v]\|_{L^2(\gamma^0(\omega(T)))}$$

for all $v \in \mathbb{V}_h^k$, it suffices to derive the bounds

$$\|\nabla v_i\|_{L^2(\omega_i)} + \|h^{-1}(v - v_i)\|_{L^2(\omega_i)} \lesssim \|\nabla_h v\|_{L^2(\omega_i)} + \|h^{-1/2}[v]\|_{L^2(\gamma_i^0)},$$

where v_i is defined in (2.19). Since the dimension of the space $\mathbb{V}_h^k(\omega_i)$ is finite and depends only on shape regularity, all norms in $\mathbb{V}_h^k(\omega_i)$ are equivalent and independent of meshsize upon rescaling ω_i to unit size. We further observe that if the right hand side of the last inequality vanishes, then v is constant in ω_i. The definition of v_i thus gives $v_i = v$ and the left hand side vanishes, thereby showing that the desired estimate is valid in the rescaled ω_i. The powers of meshsize result from a standard scaling argument from unit size back to actual size.

Step 3: Operator Π_h. Combining the estimates for P_h and I_h yields

$$\|\nabla\Pi_h v\|_{L^2(\Omega)} \lesssim \|\nabla P_h v\|_{L^2(\Omega)} + \|h^{-1/2}[P_h v]\|_{L^2(\Gamma_h^0)} \lesssim \|\nabla_h v\|_{L^2(\Omega)} + \|h^{-1/2}[v]\|_{L^2(\Gamma_h^0)}.$$

A similar estimate is valid for $\|h^{-1}(v - \Pi_h v)\|_{L^2(\Omega)}$. This concludes the proof. $\qquad\square$

We have written (2.21) in a convenient form for our later application. Note that it only requires that $\mathbb{V}_h^k$ contains piecewise constants and make no reference to the actual polynomial degree k. However, since Π_h is invariant on $\mathring{\mathbb{V}}_h^k$, we may apply (2.21) to $v - p$ where $p \in \mathring{\mathbb{V}}_h^k$ is the best H^1-approximation of $v \in H^{k+1}(\Omega)$ to get

$$\|v - \Pi_h v\|_{L^2(\Omega)} = \|(v - p) - \Pi_h(v - p)\|_{L^2(\Omega)} \lesssim h\|\nabla(v - p)\|_{L^2(\Omega)} \lesssim h^{k+1}\|v\|_{H^{k+1}(\Omega)};$$

see Lemma 9.4 (error estimates for curved quadrilaterals). Another consequence of the proof of Lemma 2.1 is the *local* interpolation estimate

$$\|\nabla \Pi_h v\|_{L^2(T)} + \|h^{-1}(v - \Pi_h v)\|_{L^2(T)} \lesssim \|\nabla_h v\|_{L^2(\omega(T))} + \|h^{-1/2}[v]\|_{L^2(\gamma^0(\omega(T)))}$$

for all $T \in \mathcal{T}_h$; recall that $\omega(T)$ and $\gamma^0(\omega(T))$ are defined in Step 2 of Lemma 2.1.

Corollary 2.1 (boundary error estimate)**.** *The following estimate is valid*

$$(2.22) \quad \|h^{-1/2}(v - \Pi_h v)\|_{L^2(\partial\Omega)} \lesssim \|\nabla_h v\|_{L^2(\Omega)} + \|h^{-1/2}[v]\|_{L^2(\Gamma_h^0)} \quad \forall v \in \mathbb{E}(\mathcal{T}_h).$$

Proof. Let e be a generic boundary edge, not necessarily in $\mathcal{E}_h^b$, and $T \in \mathcal{T}_h$ be an adjacent element (i.e. $e \subset \partial T$). The scaled trace inequality reads

$$\|h^{-1/2}(v - \Pi_h v)\|_{L^2(e)} \lesssim \|h^{-1}(v - \Pi_h v)\|_{L^2(T)} + \|\nabla(v - \Pi_h v)\|_{L^2(T)}$$

Adding over e, the desired estimate follows from (2.21). $\square$

The following Friedrichs inequality is another straightforward application of (2.21); similar estimates are proved in [13, 15]. We observe that the jumps $[v]$ of $v \in \mathbb{E}(\mathcal{T}_h)$ in (2.21) are computed on the interior skeleton Γ_h^0, but the desired estimate requires control of the trace of v on $\partial_D\Omega$.

Corollary 2.2 (discrete Friedrichs inequality)**.** *There exists a constant $C_F > 0$, depending on Ω and $\partial_D\Omega$, such that for all $v \in \mathbb{E}(\mathcal{T}_h)$ there holds*

$$(2.23) \qquad \|v\|_{L^2(\Omega)} \leq C_F\Big(\|\nabla_h v\|_{L^2(\Omega)} + \|h^{-1/2}[v]\|_{L^2(\Gamma_h^0)} + \|v\|_{L^2(\partial_D\Omega)}\Big).$$

Proof. In view of (2.21), it suffices to prove (2.23) for $\Pi_h v$. The standard form of the Friedrichs inequality is valid for $\Pi_h v \in H^1(\Omega)$, namely

$$\|\Pi_h v\|_{L^2(\Omega)} \lesssim \|\nabla \Pi_h v\|_{L^2(\Omega)} + \|\Pi_h v\|_{L^2(\partial_D\Omega)},$$

where the hidden constant depends only on Ω and $\partial_D\Omega$. Let $e \in \mathcal{E}_h^b$ be a boundary edge in $\partial_D\Omega$ and $T \in \mathcal{T}_h$ be an element so that $e \subset \partial T$. The scaled trace inequality

$$\|w\|_{L^2(e)} \lesssim h_T^{-1/2}\|w\|_{L^2(T)} + h_T^{1/2}\|\nabla w\|_{L^2(T)}$$

for $w = v - \Pi_h v$, in conjunction with (2.21), yields

$$\|\nabla \Pi_h v\|_{L^2(\Omega)} + \|\Pi_h v\|_{L^2(\partial_D\Omega)} \lesssim \|\nabla_h v\|_{L^2(\Omega)} + \|h^{-1/2}[v]\|_{L^2(\Gamma_h^0)} + \|v\|_{L^2(\partial_D\Omega)},$$

and shows (2.23) as desired. $\square$

2.4. **Coercivity of the Discrete Energy.** We now prove that the discrete energy (2.16) is coercive with respect to a suitable dG quantity that substitutes the H^2-norm. As motivation, we start with a similar coercivity estimate for the continuous case.

Lemma 2.2 (coercivity of E). *Let data (g, Φ, f) satisfy (1.2). For any $y \in \mathbb{V}(g, \Phi)$ there holds*

$$\|y\|^2_{H^2(\Omega)} \lesssim E[y] + \|g\|^2_{H^1(\Omega)} + \|\Phi\|^2_{H^1(\Omega)} + \|f\|^2_{L^2(\Omega)},$$

where $E[y]$ is defined in (1.4).

Proof. Since $y = g$ and $\nabla y = \Phi$ on $\partial_D \Omega$, the Friedrichs inequality implies

$$\|y - g\|_{L^2(\Omega)} \lesssim \|\nabla(y - g)\|_{L^2(\Omega)}, \quad \|\nabla y - \Phi\|_{L^2(\Omega)} \lesssim \|D^2 y - \nabla \Phi\|_{L^2(\Omega)},$$

whence

$$(2.24) \qquad \|y\|^2_{H^1(\Omega)} \leq C \left(\|D^2 y\|^2_{L^2(\Omega)} + \|g\|^2_{H^1(\Omega)} + \|\Phi\|^2_{H^1(\Omega)} \right).$$

Consequently, we use (1.4) to find that for any $\rho > 0$

$$\|D^2 y\|^2_{L^2(\Omega)} = 2E[y] + 2\int_\Omega f \cdot y \leq 2E[y] + 2\|f\|_{L^2(\Omega)}\|y\|_{L^2(\Omega)}$$

$$\leq 2E[y] + \frac{1}{\rho}\|f\|^2_{L^2(\Omega)} + \rho\|y\|^2_{L^2(\Omega)}$$

$$\leq 2E[y] + \frac{1}{\rho}\|f\|^2_{L^2(\Omega)} + C\rho(\|D^2 y\|^2_{L^2(\Omega)} + \|g\|^2_{H^1(\Omega)} + \|\Phi\|^2_{H^1(\Omega)}).$$

Setting $\rho = \frac{1}{2C}$, we deduce that

$$\|D^2 y\|^2_{L^2(\Omega)} \lesssim E[y] + \|g\|^2_{H^1(\Omega)} + \|\Phi\|^2_{H^1(\Omega)} + \|f\|^2_{L^2(\Omega)},$$

which combined with (2.24) gives the asserted estimate. $\qquad\square$

We observe that the Friedrichs inequality plays a crucial role in the previous proof to control $y - g$ and $\nabla y - \Phi$, which vanish on $\partial_D \Omega$. At the discrete level we face two difficulties: the lack of regularity that leads to interior jumps for y_h and the Nitsche's approach that circumvents the explicit imposition of Dirichlet boundary conditions. To deal with the first issue, we introduce a discrete H^2-scalar product

$$(2.25) \qquad \begin{aligned} \left(w_h, v_h\right)_{H^2_h(\Omega)} :=& \left(D^2_h w_h, D^2_h v_h\right)_{L^2(\Omega)} \\ &+ \left(h^{-1}[\nabla_h w_h], [\nabla_h v_h]\right)_{L^2(\Gamma_h)} + \left(h^{-3}[w_h], [v_h]\right)_{L^2(\Gamma_h)} \end{aligned}$$

and corresponding discrete H^2-norm $\|v_h\|_{H^2_h(\Omega)} := (v_h, v_h)^{1/2}_{H^2_h(\Omega)}$ for all $v_h, w_h \in \mathbb{V}^k_h(0, \mathbf{0})$. We point out that the jumps are computed in the full skeleton Γ_h and for boundary edges $e \in \mathcal{E}^b_h$ we have $[v_h] = v_h$ and $[\nabla_h v_h] = \nabla_h v_h$ according to the definition of $\mathbb{V}^k_h(0, \mathbf{0})$ in (2.11) for $g = 0$ and $\Phi = \mathbf{0}$.

In order to account for the second issue, the Nitsche's approach, we need to allow $v_h \in \mathbb{V}^k_h(g, \Phi)$ and thus utilize the convention that boundary jumps in

$$(2.26) \quad \|v_h\|^2_{H^2_h(\Omega)} := \|D^2_h v_h\|^2_{L^2(\Omega)} + \|h^{-1/2}[\nabla_h v_h]\|^2_{L^2(\Gamma_h)} + \|h^{-3/2}[v_h]\|^2_{L^2(\Gamma_h)}$$

are defined as in (2.11) for $v_h \in \mathbb{V}^k_h(g, \Phi)$. This convention will simplify notation and will not lead to confusion because we will always specify the membership of v_h in the sequel. Notice that $\|.\|_{H^2_h(\Omega)}$ is not a norm on $\mathbb{V}^k_h(g, \Phi)$.

In addition, we can derive a Friedrichs-type estimate for any $v_h \in \mathbb{V}_h^k(g, \Phi)$. We simply apply (2.23) to v_h and $\nabla_h v_h$, and add and subtract g and Φ to the boundary terms $\|v_h\|_{L^2(\partial_D \Omega)}$ and $\|\nabla_h v_h\|_{L^2(\partial_D \Omega)}$, to obtain

$$(2.27) \qquad \|v_h\|_{L^2(\Omega)} + \|\nabla_h v_h\|_{L^2(\Omega)} \le C_F \Big(\|v_h\|_{H_h^2(\Omega)} + \|g\|_{H^1(\Omega)} + \|\Phi\|_{H^1(\Omega)} \Big),$$

where C_F (not relabeled) is proportional to the constant in (2.23).

Lemma 2.3 (coercivity of E_h)**.** *Let data* (g, Φ, f) *satisfy* (1.2) *and set*

$$(2.28) \qquad R(g, \Phi, f) := \|g\|_{H^1(\Omega)}^2 + \|\Phi\|_{H^1(\Omega)}^2 + \|f\|_{L^2(\Omega)}^2.$$

If the penalty parameters γ_0, γ_1 *in the discrete energy* $E_h[y_h]$ *defined in* (2.16) *are sufficiently large, then there exists a constant* C_F^* *depending on* C_F *in* (2.27) *such that*

$$\|y_h\|_{H_h^2(\Omega)}^2 \lesssim E_h[y_h] + C_F^* R(g, \Phi, f) \quad \forall y_h \in \mathbb{V}_h^k(g, \Phi).$$

Proof. We examine the various terms in $a_h(y_h, y_h)$ defined in (2.14) with $y_h \in \mathbb{V}_h^k(g, \Phi)$. We start with $\big(\{\partial_\mu \nabla_h y_h\}, [\nabla_h y_h] \big)_{L^2(\Gamma_h)}$, recall the definition (2.12) of averages, and use the inverse estimate (9.8) for isoparametric elements to obtain

$$\|\{\partial_\mu \nabla_h y_h\}\|_{L^2(e)} \le \|\{D_h^2 y_h\}\|_{L^2(e)} \lesssim h_e^{-1/2} \|D_h^2 y_h\|_{L^2(\omega(e))}$$

for all $e \in \mathcal{E}_h$, where $\omega(e)$ is either the union of two elements containing e for $e \in \mathcal{E}_h^0$ or a single element for $e \in \mathcal{E}_h^b$. Consequently, Young's inequality yields for any $\rho > 0$

$$\left| \big(\{\partial_\mu \nabla_h y_h\}, [\nabla_h y_h] \big)_{L^2(\Gamma_h)} \right| \le \rho \|D_h^2 y_h\|_{L^2(\Omega)}^2 + \frac{C_1}{\rho} \|h^{-1/2} [\nabla_h y_h]\|_{L^2(\Gamma_h)}^2.$$

We next consider $\big(\{\partial_\mu \Delta_h y_h\}, [y_h] \big)_{L^2(\Gamma_h)}$. The inverse inequality (9.8) again implies

$$\|\{\partial_\mu \Delta_h y_h\}\|_{L^2(e)} \le \|\{D_h^3 y_h\}\|_{L^2(e)} \lesssim h_e^{-3/2} \|D_h^2 y_h\|_{L^2(\omega(e))} \quad \forall \, e \in \mathcal{E}_h,$$

whence Young's inequality gives

$$\left| \big(\{\partial_\mu \Delta_h y_h\}, [y_h] \big)_{L^2(\Gamma_h)} \right| \lesssim \rho \|D_h^2 y_h\|_{L^2(\Omega)}^2 + \frac{C_0}{\rho} \|h^{-3/2} [y_h]\|_{L^2(\Gamma_h)}^2.$$

In light of definition (2.26) of $\| \cdot \|_{H_h^2(\Omega)}$, we choose $\rho = 1/8$ and γ_1, γ_0 sufficiently large, depending on the geometric constants C_1, C_0, to absorb the above terms into $\frac{1}{2} \|D_h^2 y_h\|_{L^2(\Omega)}^2$, $\frac{\gamma_1}{2} \|h^{-1/2} [\nabla_h y_h]\|_{L^2(\Gamma_h)}^2$ and $\frac{\gamma_0}{2} \|h^{-3/2} [y_h]\|_{L^2(\Gamma_h)}^2$ of $a_h(y_h, y_h)$. We thus deduce coercivity of a_h and, according to (2.16), we find

$$\|y_h\|_{H_h^2(\Omega)}^2 \lesssim \frac{1}{2} a_h(y_h, y_h) = E_h[y_h] + \big(f, y_h \big)_{L^2(\Omega)}.$$

We apply the Friedrichs-type estimate (2.27) to deal with the forcing term, namely

$$\left| \int_\Omega f \cdot y_h \right| \le \|f\|_{L^2(\Omega)} \|y_h\|_{L^2(\Omega)} \le C_F \|f\|_{L^2(\Omega)} \Big(\|y_h\|_{H_h^2(\Omega)} + \|g\|_{H^1(\Omega)} + \|\Phi\|_{H^1(\Omega)} \Big).$$

We finally see that $\|y_h\|_{H_h^2(\Omega)}$ can be hidden on the left-hand side of the previous coercivity estimate via Young's inequality, and conclude the proof. $\qquad \square$

Lemma 2.3 and its proof yield the following bounds on $E_h[y_h]$

$$(2.29) \qquad -C_F^* R(g, \Phi, f) \le E_h[y_h] \lesssim \|y_h\|_{H_h^2(\Omega)}^2 + R(g, \Phi, f).$$

3. Discrete Gradient Flow

We now design a discrete H^2-gradient flow to construct discrete minimizers. A key aspect of this flow is to guarantee the discrete isometry defect (1.6), namely

$$D_h[y_h] = \sum_{T \in \mathcal{T}_h} \left| \int_T (\nabla_h v_h)^T \nabla_h v_h - I \right| \le \varepsilon.$$

We start with the first variation of the isometry constraint (1.1) evaluated at $y \in \mathbb{V}(g, \Phi)$ in the direction v, the so-called *linearized isometry constraint*, which reads

$$(3.1) \qquad L[y; v] = (\nabla v)^T \nabla y + (\nabla y)^T \nabla v = 0 \quad \forall \, v \in \mathbb{V}(0, \mathbf{0}).$$

This serves to describe the tangent manifold to (1.1) at $y \in \mathbb{V}(g, \Phi)$

$$\mathcal{F}[y] := \big\{ v \in \mathbb{V}(0, \mathbf{0}) : \quad L[y; v] = 0 \big\}.$$

In view of (3.1), let the *discrete linearized isometry constraint* at $y_h \in \mathbb{V}_h^k(g, \Phi)$ be

$$(3.2) \qquad L_T[y_h; v_h] := \int_T (\nabla_h v_h)^T \nabla_h y_h + (\nabla_h y_h)^T \nabla_h v_h = 0 \quad \forall \, v_h \in \mathbb{V}_h^k(0, \mathbf{0}),$$

which imposes the pointwise equality (3.1) on average over each element $T \in \mathcal{T}_h$. This in turn leads to the subspace $\mathcal{F}_h[y_h]$ of $\mathbb{V}_h^k(0, \mathbf{0})$ defined as

$$\mathcal{F}_h[y_h] := \big\{ v_h \in \mathbb{V}_h^k(0, \mathbf{0}) : \quad L_T[y_h; v_h] = 0 \quad \forall \, T \in \mathcal{T}_h \big\}.$$

We are now in a position to describe the *discrete H^2-gradient flow*. Let $y_h^0 \in \mathbb{V}_h^k(g, \Phi)$ be a suitable initial guess with energy $E_h[y_h^0]$ as small as possible and isometry defect $D_h[y_h^0] \le \tau$, where $\tau > 0$ is a fictituous time-step to be determined later. Given an iterate $y_h^n \in \mathbb{V}_h^k(g, \Phi)$ for $n \ge 0$, we seek to minimize the functional

$$\mathcal{F}_h[y_h^n] \ni w_h \mapsto \frac{1}{2\tau} \|w_h\|_{H_h^2(\Omega)}^2 + E_h[y_h^n + w_h],$$

where the discrete H^2-norm $\|w_h\|_{H_h^2(\Omega)}$ is defined in (2.26). We thus minimize $E_h[y_h^n + w_h]$ but penalizing the deviation of w_h from zero in the norm $\|w_h\|_{H_h^2(\Omega)}$. If $\delta y_h^{n+1} \in \mathcal{F}_h[y_h^n]$ is a minimizer, then it satisfies the optimality condition

$$(3.3) \qquad \tau^{-1} \big(\delta y_h^{n+1}, v_h \big)_{H_h^2(\Omega)} + \delta E_h[y_h^n + \delta y_h^{n+1}; v_h] = 0 \quad \forall v_h \in \mathcal{F}_h[y_h^n],$$

where $\delta E_h[y_h^{n+1}; v_h]$ is the variational derivative of E_h at $y_h^{n+1} := y_h^n + \delta y_h^{n+1}$ in the direction of v_h. In view of (2.15) and (2.16), the Euler-Lagrange equation (3.3) is equivalent to

$$(3.4) \qquad \begin{aligned} \tau^{-1}(\delta y_h^{n+1}, v_h)_{H_h^2(\Omega)} &+ a_h(\delta y_h^{n+1}, v_h) \\ &= -a_h(y_h^n, v_h) + (f, v_h)_{L^2(\Omega)} =: F^n(y_h^n; v_h) \quad \forall v_h \in \mathcal{F}_h[y_h^n]. \end{aligned}$$

Remark 3.1 (solvability of (3.4)). Note that problem (3.4) is linear in δy_h^{n+1} and Lemma 2.3 (coercivity of E_h) with vanishing data (g, Φ, f) yields coercivity of a_h

$$(3.5) \qquad a_h(v_h, v_h) \ge \alpha \|v_h\|_{H_h^2(\Omega)}^2 \quad \forall \, v_h \in \mathbb{V}_h^k(0, \mathbf{0})$$

for some constant α independent of h. Since $0 \in \mathcal{F}_h[y_h^n]$ we infer that $\mathcal{F}_h[y_h^n] \ne \emptyset$ and the Lax-Milgram theorem implies existence and uniqueness of a solution $\delta y_h^{n+1} \in \mathcal{F}_h[y_h^n]$ to each step of the discrete gradient flow (3.3).

Remark 3.2 (saddle point formulation). Instead of solving (3.4) within the manifold $\mathcal{F}_h[y_h^n] \subset \mathbb{V}_h^k(0, \mathbf{0})$ we could pose the problem in the entire space $\mathbb{V}_h^k(0, \mathbf{0})$ provided we append the constraint $\delta y_h^{n+1}, v_h \in \mathcal{F}_h[y_h^n]$ via a Lagrange multiplier. To this end, consider the bilinear form $\ell_h(y_h^n; \cdot, \cdot) : [\mathbb{V}_h^k]^3 \times [\mathbb{V}_h^0]^{2\times2}$ given by

$$(3.6) \qquad \ell_h(y_h^n; v_h, \mu_h) := \sum_{T \in \mathcal{T}_h} \int_T \mu_n : \left((\nabla_h y_h^n)^T \nabla_h v_h + (\nabla_h v_h)^T \nabla_h y_h^n \right).$$

The saddle point system equivalent to (3.4) reads: find $(\delta y_h^{n+1}, \lambda_h^{n+1}) \in [\mathbb{V}_h^k]^3 \times [\mathbb{V}_h^0]^{2\times2}$ such that for all $(v_h, \mu_h) \in [\mathbb{V}_h^k]^3 \times [\mathbb{V}_h^0]^{2\times2}$ there holds

$$(3.7) \qquad \tau^{-1}(\delta y_h^{n+1}, v_h)_{H_h^2(\Omega)} + a_h(\delta y_h^{n+1}, v_h) + \ell_h(y_h^n; v_h, \lambda_h^{n+1}) = F^n(y_h^n; v_h)$$
$$\ell_h(y_h^n; \delta y_h^{n+1}, \mu_h) = 0.$$

It is an open question whether a uniform inf-sup condition is valid for (3.7). We will get back to (3.7) in Section 7 as a practical procedure to find δy_h^{n+1}.

We now show that the discrete H^2-gradient flow (3.3) reduces the energy E_h.

Lemma 3.3 (energy decay). *Let $y_h^n \in \mathbb{V}_h^k(g, \Phi)$ be the n-th iterate of the discrete H^2-gradient flow (3.3) with data (g, Φ, f) obeying (1.2). If $\delta y_h^{n+1} \neq 0$ is the solution of (3.3), then the next iterate $y_h^{n+1} = y_h^n + \delta y_h^{n+1}$ satisfies*

$$E_h[y_h^{n+1}] < E_h[y_h^n].$$

Moreover, if α is the coercivity constant in (3.5) and τ is the time step, then

$$(3.8) \qquad \left(\frac{\alpha}{2} + \frac{1}{\tau} \right) \sum_{n=0}^{N-1} \|\delta y_h^{n+1}\|_{H_h^2(\Omega)}^2 + E_h[y_h^N] \leq E_h[y_h^0].$$

Proof. We set $v_h = \delta y_h^{n+1}$ in (3.3) and use the fact that E_h is quadratic to obtain

$$\delta E_h[y_h^n + \delta y_h^{n+1}, \delta y_h^{n+1}] = E_h[y_h^{n+1}] - E_h[y_h^n] + \frac{1}{2} a_h(\delta y_h^{n+1}, \delta y_h^{n+1}).$$

Invoking the coercivity property (3.5) we thus get

$$\left(\frac{\alpha}{2} + \frac{1}{\tau} \right) \|\delta y_h^{n+1}\|_{H_h^2(\Omega)}^2 + E_h[y_h^{n+1}] \leq E_h[y_h^n],$$

whence $E_h[y_h^{n+1}] < E_h[y_h^n]$ if $\delta y_h^{n+1} \neq 0$. We finally sum over n to deduce (3.8). $\square$

We now show that the discrete H^2-gradient flow guarantees the discrete isometry defect (1.6) for any $\varepsilon > 0$ provided the time step τ in (3.3) is suitably chosen. This is due to the presence of the *dissipative* leftmost term in (3.8).

Lemma 3.4 (discrete isometry defect). *Let the initial guess $y_h^0 \in \mathbb{V}_h^k(g, \Phi)$ for (3.4) be an approximate isometry in the sense that*

$$(3.9) \qquad D_h[y_h^0] = \sum_{T \in \mathcal{T}_h} \left| \int_T (\nabla_h y_h^0)^T \nabla_h y_h^0 - I \right| \leq \tau,$$

let $R(g, \Phi, f)$ and C_F^* be given in Lemma 2.3 (coercivity of E_h), let C_F be the Friedrichs constant in (2.27), and set

$$(3.10) \qquad \varepsilon_0 := \left(1 + C_F^2 \big(E_h[y_h^0] + C_F^* R(g, \Phi, f) \big) \right) \tau,$$

Then every iterate y_h^n of (3.3) *for $n \geq 1$ satisfies the discrete isometry defect* (1.6)

$$(3.11) \qquad D_h[y_h^n] = \sum_{T \in \mathcal{T}_h} \left| \int_T (\nabla_h y_h^n)^T \nabla_h y_h^n - I \right| \leq \varepsilon \quad \forall \varepsilon \geq \varepsilon_0.$$

Proof. We start by quantifying the increase of the discrete isometry defect in each iteration of the discrete gradient flow. Combining $y_h^{n+1} = y_h^n + \delta y_h^{n+1}$ with the property $L_T[y_h^n; \delta y_h^{n+1}] = 0$, which is (3.2) for $v_h = \delta y_h^{n+1}$, we get the identity

$$\int_T (\nabla_h y_h^{n+1})^T \nabla_h y_h^{n+1} = \int_T (\nabla_h y_h^n)^T \nabla_h y_h^n + \int_T \left(\nabla_h \delta y_h^{n+1} \right)^T \nabla_h \delta y_h^{n+1}.$$

Summing for $n = 0$ to $n = N - 1$, and exploiting telescopic cancellation, we obtain

$$\int_T (\nabla_h y_h^N)^T \nabla_h y_h^N = \int_T (\nabla_h y_h^0)^T \nabla_h y_h^0 + \sum_{n=0}^{N-1} \| \nabla_h \delta y_h^{n+1} \|_{L^2(T)}^2.$$

Consequently, adding over $T \in \mathcal{T}_h$ and employing (3.9), we deduce

$$D_h[y_h^N] = \sum_{T \in \mathcal{T}_h} \left| \int_T (\nabla_h y_h^N)^T \nabla_h y_h^N - I \right| \leq \tau + \sum_{n=0}^{N-1} \| \nabla_h \delta y_h^{n+1} \|_{L^2(\Omega)}^2.$$

We finally combine the Friedrichs-type inequality (2.27) for $\delta y_h^{n+1} \in \mathbb{V}_h^k(0, \mathbf{0})$ (i.e. corresponding to Dirichlet boundary data $g = 0$ and $\Phi = \mathbf{0}$) with the energy decay relation (3.8) and the lower bound in (2.29) to get

$$\sum_{n=0}^{N-1} \| \nabla_h \delta y_h^{n+1} \|_{L^2(\Omega)}^2 \leq C_F^2 \sum_{n=0}^{N-1} \| \delta y_h^{n+1} \|_{H_h^2(\Omega)}^2 < C_F^2 \left(E_h[y_h^0] + C_F^* R(g, \Phi, f) \right) \tau.$$

Inserting this in the bound for $D_h[y_h^N]$ yields the asserted estimate (3.11). $\qquad \square$

Remark 3.5 (choice of initial guess y_h^0). We realize from Lemma 3.4 (discrete isometry defect) that choosing y_h^0 might be tricky, unless we can make the discrete isometry defect $D_h[y_h^0]$ small without increasing much the discrete energy $E_h[y_h^0]$ due to the boundary penalty terms. We will revisit this issue in Section 6.

4. Discrete Hessian

In this section we provide a suitable definition of discrete Hessian $H_h[y_h] \in [L^2(\Omega)]^{3 \times 2 \times 2}$ for $y_h \in [\mathbb{V}_h^k]^3$. Central to this concept is the following question: if the discontinuous function y_h converges strongly in $[L^2(\Omega)]^3$ to $y \in [H^2(\Omega)]^3$, under what conditions could $H_h[y_h]$ converge weakly in $[L^2(\Omega)]^{3 \times 2 \times 2}$ to $D^2 y$, namely

$$(4.1) \qquad\qquad H_h[y_h] \rightharpoonup D^2 y \quad \text{in} \quad [L^2(\Omega)]^{3 \times 2 \times 2}?$$

It is apparent that the information contained in the broken Hessian $D_h^2 y_h \in [L^2(\Omega)]^{3 \times 2 \times 2}$ is insufficient for this purpose, and that the jumps $[y_h] \in [L^2(\Gamma_h)]^3$ and $[\nabla_h y_h] \in [L^2(\Gamma_h)]^{3 \times 2}$ do not provide directly the missing information because they are singular with respect to the Lebesgue measure. To bridge this gap, in Section 4.1 we introduce lifting operators of $[y_h]$ and $[\nabla_h y_h]$ [] and use them to construct $H_h[y_h]$. Another important property of $H_h[y_h]$ critical for the lim-inf argument is

$$(4.2) \qquad\qquad \frac{1}{2} \| H_h[y_h] \|_{L^2(\Omega)}^2 - (f_h, y_h)_{L^2(\Omega)} \leq E_h[y_h],$$

with constant 1 on the right-hand side. We will prove (4.1) and (4.2) in Section 4.2.

Our approach is similar in spirit to that developed by Pryer[31], who in turn is inspired by Di Pietro and Ern[20] for second order problems. However, an important distinction is worth pointing out. The finite element spaces considered in [20, 31] consist of polynomial functions (of total degree $\leq k$) on the physical elements $T \in \mathcal{T}_h$. Instead, the dG method proposed in this work uses polynomial functions on the reference element $\widehat{T}$ (see (2.7)), which is consistent with the implementation within the deal.ii library[1, 4] responsible for the numerical experiments of Section 6. Consequently, several challenges arise when dealing with isoparametric elements (in particular quadrilaterals elements): (i) the mappings $F_T : \widehat{T} \to T$ between $\widehat{T}$ and $T \in \mathcal{T}_h$ are non-affine and do not produce polynomial functions on T; (ii) $\nabla_h \mathbb{V}_h^k$ does not embed in $[\mathbb{V}_h^{k'}]^2$ for some k' as required, for example, for the analyses in [20, 16, 18]; (iii) the Hessians of $D^2 v_h|_T$ and $D^2(v_h \circ F_T)$ are not proportional because $D^2 F_T \neq \mathbf{0}$. As a consequence, the space of Hessians

$$(4.3) \qquad \mathbb{H}_h^k := \{\tau_h \in L^2(\Omega)^{2\times 2} \ : \ \tau_h|_T = D_h^2 w_h, \quad w_h \in \mathbb{V}_h^k(T), \ T \in \mathcal{T}_h\},$$

where the reconstructed Hessian $H_h[v_h]$ of a function $v_h \in \mathbb{V}_h^k$ belongs to, is not a subspace of $[\mathbb{V}_h^k]^{2\times 2}$ in general; $\tau_h \in \mathbb{H}_h^k$ might not be piecewise polynomial.

4.1. Lifting operators and definition of $H_h[y_h]$.
Given a scalar-valued function $v_h \in \mathbb{V}_h^k$, we define two lifting operators $R_h([\nabla_h v_h])$ and $B_h([v_h])$ that extend the jumps $[\nabla_h v_h]$ and $[v_h]$ of $\nabla_h v_h$ and v_h from the full skeleton Γ_h to the bulk Ω, as in Brezzi et al.[16]. It turns out that of the many ways to achieve this, there is only one that leads to (4.1) and (4.2). We describe this next.

Given an edge $e \in \mathcal{E}_h$, let $\omega(e)$ be the patch associated with e (i.e. the union of two elements sharing e for interior edges $e \in \mathcal{E}_h^0$ or just one single element for boundary edges $e \in \mathcal{E}_h^b$), and let $\mathbb{H}_h^k(\omega(e))$ stand for the restriction of functions in $\mathbb{H}_h^k$ to $\omega(e)$. The first lifting operator $R_h := \sum_{e \in \mathcal{E}_h} r_e : [L^2(\Gamma_h)]^2 \to \mathbb{H}_h^k$ hinges on the local liftings $r_e : [L^2(e)]^2 \to \mathbb{H}_h^k$ which, for all $e \in \mathcal{E}_h$ and $\phi \in [L^2(e)]^2$, are defined by

$$(4.4) \qquad r_e(\phi) \in \mathbb{H}_h^k(\omega(e)) : \int_{\omega(e)} r_e(\phi) : \tau_h = \int_e \phi \cdot \{\tau_h\} \, \mu_e \quad \forall \tau_h \in \mathbb{H}_h^k(\omega(e)),$$

and vanish outside $\omega(e)$. Notice that upon taking $\tau_h = D_h^2 w_h$, we get

$$(4.5) \qquad \left(R_h([\nabla_h v_h]), D_h^2 w_h\right)_{L^2(\Omega)} = \left([\nabla_h v_h], \{\partial_\mu(\nabla_h w_h)\}\right)_{L^2(\Gamma_h)} \quad \forall w_h \in \mathbb{V}_h^k.$$

The second lifting operator $B_h := \sum_{e \in \mathcal{E}_h} b_e : L^2(e) \to \mathbb{H}_h^k$ relies on the local liftings $b_e : L^2(e) \to \mathbb{H}_h^k$ which, for all $e \in \mathcal{E}_h$ and $\phi \in L^2(e)$, are given by

$$(4.6) \quad b_e(\phi) \in \mathbb{H}_h^k(\omega(e)) : \int_{\omega(e)} b_e(\phi) : \tau_h = \int_e \phi \, \{\text{div } \tau_h\} \cdot \mu_e \quad \forall \tau_h \in \mathbb{H}_h^k(\omega(e)),$$

and vanish outside $\omega(e)$. In this case, taking $\tau_h = D_h^2 w_h$ again and observing that $\text{div}(D_h^2 v_h) = \nabla_h(\Delta_h v_h)$ elementwise, we obtain

$$(4.7) \qquad \left(B_h([v_h]), D_h^2 w_h\right)_{L^2(\Omega)} = \left([v_h], \{\partial_\mu(\Delta_h w_h)\}\right)_{L^2(\Gamma_h)} \quad \forall w_h \in \mathbb{V}_h^k.$$

To deal with vector-valued functions, we apply (4.4) and (4.6) componentwise.

Definition 4.1 (discrete Hessian). We define $H_h : \mathbb{V}_h^k(g, \Phi) \to [\mathbb{H}_h^k]^3$ to be

$$(4.8) \qquad\qquad H_h[y_h] := D_h^2 y_h - R_h([\nabla_h y_h]) + B_h([y_h]),$$

where $R_h = \sum_{e \in \mathcal{E}_h} r_e$ and $B_h = \sum_{e \in \mathcal{E}_h} b_e$ are defined in (4.4) and (4.6).

It is worth realizing that, in view of (4.5) and (4.7), we readily find
(4.9)
$$\left(H_h[y_h], \tau_h\right)_{L^2(\Omega)} = \left(D_h^2 y_h, D_h^2 w_h\right)_{L^2(\Omega)}$$
$$- \left([\nabla_h y_h], \{\partial_\mu(\nabla_h w_h)\}\right)_{L^2(\Gamma_h)} + \left([y_h], \{\partial_\mu(\Delta_h w_h)\}\right)_{L^2(\Gamma_h)}$$

for all $\tau_h = D_h^2 w_h \in [\mathbb{H}_h^k]^3$ and $y_h \in \mathbb{V}_h^k(g, \Phi)$, and that the boundary jumps on Γ_h^b are given by $[y_h] = y_h - g$ and $[\nabla_h y_h] = \nabla_h y_h - \Phi$, according to (2.10) and (2.11). Relation (4.9) is a key tool in the analysis of our dG method and justifies using the Hessian space (4.3) for the definition of reconstructed Hessian $H_h[y_h]$. In contrast, Pryer [31] assumes that y_h is piecewise polynomial of total degree $\le k$ in Ω and $H_h[y_h]$ is piecewise polynomial of total degree $k-2$ in Ω.

4.2. **Properties of $H_h[y_h]$.** We start with L^2 a priori bounds for R_h, B_h and H_h.

Lemma 4.2 (L^2-bounds of lifting operators). *Let $y_h \in \mathbb{V}_h^k(g, \Phi)$ with data (g, Φ) satisfying (1.2). Then, for all $e \in \Gamma_h$, there holds*

$$\|r_e([\nabla_h y_h])\|_{L^2(\omega(e))} \lesssim \left\|h^{-1/2}[\nabla_h y_h]\right\|_{L^2(e)}$$

and

$$\|b_e([y_h])\|_{L^2(\omega(e))} \lesssim \left\|h^{-3/2}[y_h]\right\|_{L^2(e)}.$$

Proof. We argue as in [16] but with emphasis on boundary edges $e \in \mathcal{E}_h^b$ because they contain the Dirichlet data (g, Φ) in view of (2.10); see also [11, 20, 31]. We prove the first bound because the other one is identical. Definition (4.4) of r_e yields

$$\|r_e(\nabla_h y_h - \Phi)\|_{L^2(\omega(e))}^2 = \int_e (\nabla_h y_h - \Phi) : \{r_e(\nabla_h y_h - \Phi)\}\,\mu_e.$$

Since $\{r_e(\nabla_h y_h - \Phi)\}_e = r_e(\nabla_h y_h - \Phi)$ for all $e \in \mathcal{E}_h^b$, according to (2.12), invoking the inverse estimate (9.8) for $D^2 v_h = r_e(\nabla_h y_h - \Phi)$ implies

$$\|r_e(\nabla_h y_h - \Phi)\|_{L^2(\omega(e))}^2 \le \|h^{-1/2}(\nabla_h y_h - \Phi)\|_{L^2(e)} \|h^{1/2} r_e(\nabla_h y_h - \Phi)\|_{L^2(e)}$$
$$\lesssim \|h^{-1/2}(\nabla_h y_h - \Phi)\|_{L^2(e)} \|r_e(\nabla_h y_h - \Phi)\|_{L^2(\omega(e))},$$

and yields the desired bound for $e \in \mathcal{E}_h^b$. For interior edges $e \in \mathcal{E}_h^0$, we replace $\nabla y_h - \Phi$ with $[\nabla_h y_h]_e$ given by (2.9), and apply the same argument. $\qquad\square$

Since each element $T \in \mathcal{T}_h$ intersects at most 4 sets $\omega(e)$, we immediately get

(4.10)
$$\|R_h([\nabla_h y_h])\|_{L^2(\Omega)} \lesssim \|h^{-1/2}[\nabla_h y_h]\|_{L^2(\Gamma_h)},$$
$$\|B_h([y_h])\|_{L^2(\Omega)} \lesssim \|h^{-3/2}[y_h]\|_{L^2(\Gamma_h)},$$

as well as the following corollary.

Corollary 4.1 (L^2-bound of discrete Hessian). *If $y_h \in \mathbb{V}_h^k(g, \Phi)$ with data (g, Φ) satisfying (1.2), then the following bound holds*

$$\|H_h[y_h]\|_{L^2(\Omega)} \lesssim \|y_h\|_{H_h^2(\Omega)}.$$

Proof. Combine Definition 4.1 (discrete Hessian) with (4.10) and (2.26). $\qquad\square$

While there is some flexibility in the definitions (4.4) and (4.6) of local lifts r_e and b_e for Lemma 4.2 (L^2-bounds of lifting operators) to be valid, the following result reveals that these definitions are just right for the weak convergence of $H_h[y_h]$ provided the following condition holds on the sequence of meshes $\{\mathcal{T}_h\}_{h>0}$

$$(4.11) \qquad \frac{\|D^m F_T^{-1}\|_{L^\infty(T)}}{\|DF_T^{-1}\|_{L^\infty(T)}^m} \le \xi(h) \to 0 \quad \text{as } h_T \to 0,$$

for all $T = F_T(\widehat{T}) \in \mathcal{T}_h$ and $m = 2,3$. Note that (4.11) is valid for affine equivalent elements because F_T^{-1} is affine and $\xi(h) = 0$. However, (4.11) is more restrictive for non-affine mappings F_T (i.e. $F_T \in [\mathbb{P}^k]^2$ for $k \ge 2$ or $F_T \in [\mathbb{Q}^k]^2$ for $k \ge 1$).

Proposition 4.3 (weak convergence of $H_h[y_h]$)**.** *Let* (4.11) *hold and* $y_h \in \mathbb{V}_h^k(g, \Phi)$ *with data* (g, Φ) *satisfying* (1.2). *If* $\|y_h\|_{H_h^2(\Omega)} \le \Lambda$, *with* $\Lambda > 0$ *independent of* h, *and* y_h *converges to a function* $y \in [H^2(\Omega)]^3$ *in* $[L^2(\Omega)]^3$ *as* $h \to 0$, *then*

$$H_h[y_h] \rightharpoonup D^2 y \quad \text{in} \quad [L^2(\Omega)]^{3\times2\times2}.$$

Proof. Since we know $D^2 y \in [L^2(\Omega)]^{3\times2\times2}$ by assumption, we need to show that

$$\int_\Omega H_h[y_h] : \tau \to \int_\Omega D^2 y : \tau \quad \text{as } h \to 0$$

for all $\tau \in [C_0^\infty(\Omega)]^{3\times2\times2}$. First we show this property relying on a suitable approximation $\tau_h \in [\mathbb{H}_h^k]^3$ of τ, and next we construct such τ_h using (4.11).

Step 1: Weak convergence. We argue with each component of y_h, integrate by parts elementwise and utilize the definition (4.8) of $H_h[y_h]$ and elementary calculations to deduce

$$\int_\Omega H_h[y_h] : \tau = \sum_{T \in \mathcal{T}_h} \int_T y_h \cdot \operatorname{div} \operatorname{div} \tau$$

$$- \sum_{e \in \mathcal{E}_h} \int_{\omega(e)} r_e([\nabla_h y_h]) : (\tau - \tau_h) + \sum_{e \in \mathcal{E}_h} \int_{\omega(e)} b_e([y_h]) : (\tau - \tau_h)$$

$$+ \sum_{e \in \mathcal{E}_h^0} \int_e [\nabla_h y_h] : \{\tau - \tau_h\} \mu_e - \sum_{e \in \mathcal{E}_h^0} \int_e [y_h] \cdot \{\operatorname{div}_h(\tau - \tau_h)\} \mu_e$$

$$- \sum_{e \in \mathcal{E}_h^b} \int_e (\nabla_h y_h - \Phi) : \tau_h \, \mu_e + \sum_{e \in \mathcal{E}_h^b} \int_e (y_h - g) \cdot \operatorname{div}_h \tau_h \, \mu_e,$$

provided $\tau_h = D_h^2 w_h \in [\mathbb{H}_h^k]^3$ is a discrete Hessian. We point out that there is a contribution of boundary terms $e \in \mathcal{E}_h^b$ because they appear in Definition 4.1 (discrete Hessian). Combining the bound $\|y_h\|_{H_h^2(\Omega)} \le \Lambda$ with (2.26) and (4.10) gives

$$(4.12) \qquad \begin{aligned} \|R_h([\nabla_h y_h])\|_{L^2(\Omega)} + \|B_h([y_h])\|_{L^2(\Omega)} &\lesssim \Lambda, \\ \|h^{-1/2}[\nabla_h y_h]\|_{L^2(\Gamma_h)} + \|h^{-3/2}[y_h]\|_{L^2(\Gamma_h)} &\lesssim \Lambda. \end{aligned}$$

Suppose that there exists $w_h \in [\mathbb{V}_h^k]^3$ such that $\tau_h = D_h^2 w_h$ satisfies

$$(4.13) \quad \|\tau - \tau_h\|_{L^2(\Omega)} + \|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)} + \|h^2 D_h^2(\tau - \tau_h)\|_{L^2(\Omega)} \to 0 \quad \text{as } h \to 0.$$

The scaled trace inequalities

$$\|h^{1/2}(\tau - \tau_h)\|_{L^2(\Gamma_h)} \lesssim \|\tau - \tau_h\|_{L^2(\Omega)} + \|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)}$$

$$\|h^{3/2}\mathrm{div}_h(\tau - \tau_h)\|_{L^2(\Gamma_h)} \lesssim \|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)} + \|h^2 D_h^2(\tau - \tau_h)\|_{L^2(\Omega)},$$

together with (4.12) and (4.13), readily yield

$$\Big| \int_\Omega H_h[y_h] : \tau - \int_\Omega y_h \cdot \mathrm{div}\,\mathrm{div}\,\tau \Big| \lesssim \Lambda \|\tau - \tau_h\|_{L^2(\Omega)}$$
$$+ \Lambda\Big(\|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)} + \|h^2 D_h^2(\tau - \tau)h)\|_{L^2(\Omega)} \Big) \to 0.$$

Since $y_h \to y$ in $[L^2(\Omega)]^3$, we obtain

$$(4.14) \qquad \int_\Omega H_h[y_h] : \tau \to \int_\Omega y \cdot \mathrm{div}\,\mathrm{div}\,\tau = \int_\Omega D^2 y : \tau \quad \text{as } h \to 0,$$

because $y \in [H^2(\Omega)]^3$ and $\tau \in [C_0^\infty(\Omega)]^{3 \times 2 \times 2}$. This is the asserted limit.

Step 2: Proof of (4.13). We argue componentwise and construct a scalar $w_h \in \mathbb{V}_h^k$ such that $\tau_h = D_h^2 w_h \in \mathbb{H}_h^k$ satisfies (4.13) for $\tau \in [C_0^\infty(\Omega)]^{2 \times 2}$. To this end, we choose an arbitrary element $T \in \mathcal{T}_h$ with isoparametric map $F = F_T : \widehat{T} \to T$ from the reference element $\widehat{T}$; see Section 9. We assume that $\widehat{T}$ contains the origin, and let $x_T := F(0) \in T$. If quadratic polynomials were contained in $\mathbb{V}_h^k(T)$, the obvious idea would be to construct a quadratic approximation of τ in T by Taylor expansion. Since this is not true for isoparametric maps F, we build a quadratic approximation in $\widehat{T}$ instead. We denote by $\overline{\tau}_T$ the meanvalue of τ within T, set

$$\widehat{\tau} := DF^T(0)\, \overline{\tau}_T\, DF(0) \in \mathbb{R}^{2 \times 2},$$

and note that $\|\widehat{\tau}\|_{L^2(\widehat{T})} \lesssim h_T \|\overline{\tau}_T\|_{L^2(T)}$. Let $w_h := \widehat{w} \circ F^{-1} \in \mathbb{V}_h^k(T)$ where $\widehat{w}$ is the quadratic function $\widehat{w}(\widehat{x}) = \frac{1}{2}\widehat{x}^T \widehat{\tau} \widehat{x}$ in $\widehat{T}$; hence $D^2 \widehat{w}(\widehat{x}) = \widehat{\tau}$ for all $\widehat{x} \in \widehat{T}$. According to (9.4), the Hessian of w_h within T reads

$$\partial_{ij} w_h(x) = \sum_{m,n} \partial_{mn}\widehat{w}(\widehat{x})\, \partial_j F_n^{-1}(x)\partial_i F_m^{-1}(x) + \sum_m \partial_m\widehat{w}(\widehat{x})\, \partial_{ij} F_m^{-1}(x).$$

Since $\nabla\widehat{w}(0) = 0$, we infer that

$$D^2 w_h(x_T) = DF^{-T}(x_T)\, \widehat{\tau}\, DF^{-1}(x_T) = \overline{\tau}_T,$$

whence Poincaré inequality implies

$$\|\tau - D^2 w_h(x_T)\|_{L^2(T)} \lesssim h_T \|D\tau\|_{L^2(T)}.$$

We need to control the deviation of $D^2 w_h$ from constant due to F; for affine equivalent elements, $D^2 w_h = D^2 w_h(x_T)$ turns out to be constant. It is here that we exploit (4.11) to deal with F. Differentiating $\partial_{ij} w_h$ and using $D^3\widehat{w} = 0$, we discover

$$\|h D^3 w_h\|_{L^2(T)} \lesssim h_T^2 \|DF^{-1}\|_{L^\infty(T)}^3 \|D^3\widehat{w}\|_{L^2(\widehat{T})}$$
$$+ h_T^2 \|D^2 F^{-1}\|_{L^\infty(T)} \|DF^{-1}\|_{L^\infty(T)} \|D^2\widehat{w}\|_{L^2(\widehat{T})}$$
$$+ h_T^2 \|D^3 F^{-1}\|_{L^\infty(T)} \|\nabla\widehat{w}\|_{L^2(\widehat{T})} \lesssim \xi(h) h_T^{-1} \|\widehat{\tau}\|_{L^2(\widehat{T})} \lesssim \xi(h) \|\overline{\tau}_T\|_{L^2(T)},$$

whence $\delta D^2 w_h = D^2 w_h - D^2 w_h(x_T)$ satisfies

$$\|\delta D^2 w_h\|_{L^2(T)} \lesssim h_T \|\delta D^2 w_h\|_{L^\infty(T)}$$
$$\lesssim h_T^2 \|D^3 w_h\|_{L^\infty(T)} \lesssim h_T \|D^3 w_h\|_{L^2(T)} \lesssim \xi(h)\|\bar\tau_T\|_{L^2(T)}$$

and $\|\tau - \tau_h\|_{L^2(\Omega)} \to 0$ as $h \to 0$ because $\sum_{T \in \mathcal{T}_h} \|\bar\tau_T\|_{L^2(T)}^2 \leq \|\tau\|_{L^2(\Omega)}^2$. Moreover, we also see that

$$\|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)} \lesssim \|h\nabla\tau\|_{L^2(\Omega)} + \|hD_h^3 w_h\|_{L^2(\Omega)} \to 0 \quad \text{as } h \to 0.$$

Arguing with $D_h^4 w_h$ similarly to $D_h^3 w_h$, we easily derive

$$\|h^2 D^4 w_h\|_{L^2(T)} \lesssim \xi(h)\|\bar\tau_T\|_{L^2(T)},$$

as well as $\|h^2 D_h^2(\tau - \tau_h)\|_{L^2(\Omega)} \to 0$ as $h \to 0$. This shows (4.13) as desired. $\qquad\square$

We reiterate that condition (4.11) is innocuous for affine equivalent elements because $D^2 F_T^{-1} = 0$, but is rather restrictive otherwise (e.g quadrilateral meshes). In order to avoid (4.11), we content ourselves with a weaker statement yet sufficient for Γ-convergence of E_h towards E. We state this next.

Proposition 4.4 (lim-inf property of the reconstructed Hessian). *Let* $y_h \in \mathbb{V}_h^k(g, \Phi)$ *and* $y \in \mathbb{V}(g, \Phi)$ *with data* (g, Φ) *satisfying* (1.2). *If* $\|y_h\|_{H_h^2(\Omega)} \leq \Lambda$, *with* $\Lambda > 0$ *independent of* h, *and* $y_h \to y$ *in* $[L^2(\Omega)]^3$, *and* $y_h \to y$, $\nabla_h y_h \to \nabla y$ *in* $L^2(\partial\Omega)$ *as* $h \to 0$, *then there exists* $H \in [L^2(\Omega)]^{3\times2\times2}$ *such that* $H_h[y_h] \rightharpoonup H$ *in* $[L^2(\Omega)]^{3\times2\times2}$ *and*

$$(4.15) \qquad \int_\Omega (H - D^2 y) : D^2 w = 0, \qquad \forall w \in [H^2(\Omega)]^3.$$

Furthermore, the following estimate holds

$$(4.16) \qquad \|D^2 y\|_{L^2(\Omega)} \leq \liminf_{h\to0} \|H_h[y_h]\|_{L^2(\Omega)}.$$

Proof. We assume that the reference element $\widehat{T}$ is the unit square and omit the somewhat simpler case of the unit triangle for which interpolations estimates such as (9.14) are standard. We split the proof into two steps.

Step 1: Proof of (4.15). Thanks to Corollary 4.1 (L^2-bound on the Hessian) and the boundedness assumption $\|y_h\|_{H_h^2(\Omega)} \leq \Lambda$, there exists $H \in L^2(\Omega)^{3\times2\times2}$ such that

$$(4.17) \quad H_h[y_h] \rightharpoonup H \quad \text{in } [L^2(\Omega)]^{3\times2\times2} \quad \Rightarrow \quad \|H\|_{L^2(\Omega)} \leq \liminf_{h\to0} \|H_h[y_h]\|_{L^2(\Omega)}.$$

Proceeding as in Step 1 of the proof of Proposition 4.3 with $\tau \in [C^\infty(\overline\Omega)]^{2\times2}$ we get boundary terms involving both τ and $\text{div}\,\tau$ which result in the last line of the expression for $\int_\Omega H_h[y_h] : \tau$ being replaced by $I + II + III$, where

$$I = -\int_{\partial_D\Omega} (\nabla_h y_h - \Phi) : (\tau_h - \tau)\mu + \int_{\partial_D\Omega} (y_h - g) \cdot \text{div}_h(\tau_h - \tau)\mu,$$

$$II = \int_{\partial\Omega\setminus\partial_D\Omega} (\nabla_h y_h - \nabla y) : \tau\mu - \int_{\partial\Omega\setminus\partial_D\Omega} (y_h - y) \cdot \text{div}\,\tau\mu,$$

$$III = \int_{\partial\Omega} \nabla y : \tau\mu - \int_{\partial\Omega} y \cdot \text{div}\,\tau\mu.$$

Let $\tau = D^2 w$ for $w \in [C^\infty(\overline{\Omega})]^3$ and $\tau_h = D_h^2 w_h$ where w_h is chosen as follows: given $T \in \mathcal{T}_h$, let $\widehat{w}_T = w|_T \circ F$, $\widehat{P}\widehat{w}_T$ be the L^2-projection of $\widehat{w}_T$ onto the space $\mathbb{Q}_k$, and $w_h|_T = \widehat{P}\widehat{w}_T \circ F^{-1}$. We intend to prove the estimates

$$(4.18) \qquad \|\tau - \tau_h\|_{L^2(\Omega)} + \|h\nabla_h(\tau - \tau_h)\|_{L^2(\Omega)} + \|h^2 D_h^2(\tau - \tau_h)\|_{L^2(\Omega)} \lesssim h,$$

which are refinements of (4.13) for $w \in [C^\infty(\overline{\Omega})]^3$. In view of (4.18), employing the same argument as in Step 1 of Proposition 4.3, together with the assumption that $y_h \to y$ and $\nabla_h y_h \to \nabla y$ in $L^2(\partial\Omega)$, we infer that $|I|, |II| \to 0$ as well as

$$\int_\Omega H_h[y_h] : D^2 w \to \int_\Omega y \cdot \operatorname{div}\operatorname{div} D^2 w - \int_{\partial\Omega} y \cdot \operatorname{div} D^2 w\, \mu + \int_{\partial\Omega} \nabla y : D^2 w\, \mu,$$

as $h \to 0$ for all $w \in [C^\infty(\overline{\Omega})]^3$. Integrating twice back by parts, we deduce

$$(4.19) \qquad \int_\Omega H_h[y_h] : D^2 w \to \int_\Omega D^2 y : D^2 w \quad \text{as } h \to 0.$$

Upon approximating any $H^2(\Omega)$ by functions in $[C^\infty(\overline{\Omega})]$, relation (4.19) is also valid for any $w \in [H^2(\Omega)]^3$. This combined with (4.17) yields the orthogonality property (4.15). The latter in turn implies $\|D^2 y\|_{L^2(\Omega)} \leq \|H\|_{L^2(\Omega)}$ as well as (4.16).

Step 2: Proof of (4.18). We recall that $F : \widehat{T} \to T$ is the isoparametric map between $\widehat{T}$ and $T \in \mathcal{T}_h$. The first estimate in (4.18) is similar to (9.14):

$$\|\tau - \tau_h\|_{L^2(T)} = \|D^2(w - w_h)\|_{L^2(T)} \lesssim h_T \|w\|_{H^3(T)}$$

upon replacing the Lagrange interpolant $I_h w$ by w_h defined in Step 1. To prove the middle estimate in (4.18) we observe that if $k \geq 3$, then (9.14) yields

$$\|\nabla(\tau - \tau_h)\|_{L^2(T)} = \|D^3(w - w_h)\|_{L^2(T)} \lesssim h_T \|w\|_{H^4(T)}.$$

However, this estimate is inadequate for $k = 2$ for which we revisit the proof of Lemma 9.3 (error estimate for quadrilaterals) and add and subtract the L^2-projection $\widehat{P}_3 \widehat{w}$ onto the space of polynomials of degree ≤ 3 to arrive at

$$\|D^3(w - w_h)\|_{L^2(T)} \lesssim h_T^{-2} \sum_{j=1}^3 \left(\|D^j(\widehat{w} - \widehat{P}_3\widehat{w})\|_{L^2(\widehat{T})} + \|D^j(\widehat{P}_3\widehat{w} - \widehat{P}\widehat{w})\|_{L^2(\widehat{T})} \right).$$

For the first term, we use the Bramble-Hilbert estimate in $\mathbb{Q}_3$ to deduce

$$\sum_{j=1}^3 \|D^j(\widehat{w} - \widehat{P}_3\widehat{w})\|_{L^2(\widehat{T})} \lesssim \|[D^4\widehat{w}]\|_{L^2(\widehat{T})} \lesssim h_T^3 \|w\|_{H^4(T)}$$

in view of (9.15). For the second term, instead, we use an inverse estimate

$$\|D^3(\widehat{P}_3\widehat{w} - \widehat{P}\widehat{w})\|_{L^2(\widehat{T})} \lesssim \|D^2(\widehat{P}_3\widehat{w} - \widehat{P}\widehat{w})\|_{L^2(\widehat{T})},$$

and next add and subtract $\widehat{w}$. We first realize that $\|D^2(\widehat{w} - \widehat{P}_3\widehat{w})\|_{L^2(\widehat{T})}$ may be tackled as in the case $k \geq 3$ whereas

$$\sum_{j=1}^2 \|D^j(\widehat{w} - \widehat{P}\widehat{w})\|_{L^2(\widehat{T})} \lesssim \|[D^3\widehat{w}]\|_{L^2(\widehat{T})} \lesssim h_T^2 \|w\|_{H^3(T)}$$

according to (9.15) again. Altogether, we obtain the bound

$$\|D^3(w - w_h)\|_{L^2(T)} \lesssim h_T \|w\|_{H^4(T)} + \|w\|_{H^3(T)} \lesssim \|w\|_{H^4(T)},$$

which implies the desired middle estimate in (4.18). Exactly the same procedure leads to the remaining estimate for $\|D^4(w-w_h)\|_{L^2(T)}$ provided $w \in H^5(T)$, except that this time we invoke the L^2-projection $\widehat{P}_4$ onto polynomials of degree ≤ 4 and an inverse estimate to relate $\|D^4\widehat{v}\|_{L^2(\widehat{T})}$ to $\|D^2\widehat{v}\|_{L^2(\widehat{T})}$ for $\widehat{v}$ polynomial; we thus deduce $h_T\|D^4(w-w_h)\|_{L^2(T)} \lesssim \|w\|_{H^5(T)}$. This concludes the proof. $\qquad\square$

Another consequence of Definition 4.1 (discrete Hessian) and Lemma 4.2 (L^2-bounds of lifting operators) is (4.2) with constant 1. This is crucial for the rest.

Proposition 4.5 (relation between $E_h[y_h]$ and $H_h[y_h]$)**.** *Let data* (g, Φ, f) *satisfy* (1.2) *and* $y_h \in \mathbb{V}_h^k(g, \Phi)$*. Then the discrete energy* $E_h[y_h]$ *defined in* (2.16) *and the discrete Hessian* $H_h[y_h]$ *defined in* (4.8) *satisfy*

$$\frac{1}{2}\|H_h[y_h]\|_{L^2(\Omega)}^2 - (f, y_h)_{L^2(\Omega)} \leq E_h[y_h],$$

provided the penalty parameters γ_0 *and* γ_1 *are chosen sufficiently large.*

Proof. We expand the expression for $H_h[y_h]$ and utilize (4.5) and (4.7) to obtain

$$(4.20) \qquad \frac{1}{2}\int_\Omega |H_h[y_h]|^2 - \int_\Omega f \cdot y_h = E_h[y_h] - J_h[y_h],$$

where

$$\begin{aligned} J_h[y_h] := &-\frac{1}{2}\big\|B_h([y_h]) - R_h([\nabla_h y_h])\big\|_{L^2(\Omega)}^2 \\ &+ \gamma_0\big\|h^{-3/2}[y_h]\big\|_{L^2(\Gamma_h)}^2 + \gamma_1\big\|h^{-1/2}[\nabla_h y_h]\big\|_{L^2(\Gamma_h)}^2. \end{aligned}$$

In view of (4.10) there is a constant C independent of h such that

$$(4.21) \quad J_h[y_h] \geq (\gamma_0 - C)\big\|h^{-3/2}[y_h]\big\|_{L^2(\Gamma_h)}^2 + (\gamma_1 - C)\big\|h^{-1/2}[\nabla_h y_h]\big\|_{L^2(\Gamma_h)}^2 \geq 0$$

for γ_0 and γ_1 sufficiently large. This is what we intended to prove. $\qquad\square$

It is convenient to point out that γ_0, γ_1 sufficiently large in (4.21) also yields

$$(4.22) \qquad \big\|h^{-3/2}[y_h]\big\|_{L^2(\Gamma_h)}^2 + \big\|h^{-1/2}[\nabla_h y_h]\big\|_{L^2(\Gamma_h)}^2 \lesssim J_h[y_h].$$

5. Γ-CONVERGENCE OF E_h

In this section we prove the Γ-convergence of E_h to E. This consists of three parts. We first show equi-coercivity and compactness for $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ in Section 5.1. We next prove a lim-inf inequality for $E_h[y_h]$ in Section 5.2 and a lim-sup property in Section 5.3; the former uses lower semicontinuity of the L^2-norm whereas the latter hinges on constructing a recovery sequence for any $y \in \mathbb{A}(g, \Phi)$ via regularization and interpolation. These three results are responsible for Γ-convergence of E_h as well as convergence (up to a subsequence) of global minimizers y_h of E_h in $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ to global minimizers of E in $\mathbb{A}(g, \Phi)$; this is the topic of Section 5.4. Lastly, we combine these results in Section 5.5 to show that the stabilization terms

$$\big\|h^{-3/2}[y_h]\big\|_{L^2(\Gamma_h)} \to 0, \quad \big\|h^{-1/2}[\nabla_h y_h]\big\|_{L^2(\Gamma_h)} \to 0 \quad \text{as } h \to 0,$$

for such y_h and that $H_h[y_h]$ in Proposition 4.4 (lim-inf property of the reconstructed Hessian) as well as $D_h^2 y_h$ converge to $D^2 y$ strongly in $[L^2(\Omega)]^{3\times2\times2}$.

5.1. **Equi-coercivity and compactness.** If data (g, Φ, f) satisfies (1.2) and $y_h \in \mathbb{V}_h^k(g, \Phi)$ possesses a uniform bound $E_h[y_h] \leq \Lambda$ for all $h > 0$, then Lemma 2.3 (coercivity of E_h) guarantees *equi-coercivity*

$$(5.1) \qquad \|y_h\|_{H_h^2(\Omega)}^2 \lesssim E_h[y_h] + C_F^* R(g, \Phi, f) \leq C,$$

where $C > 0$ is independent of h. This, together with Friedrichs-type estimate (2.27), leads to weakly converging subsequences.

We now establish the L^2-compactness property.

Proposition 5.1 (compactness in $L^2(\Omega)$). *Let data (g, Φ, f) satisfy (1.2) and let the sequence $\{y_h\}_{h>0} \subset \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ satisfy the uniform bound $E_h[y_h] \leq \Lambda$ for all $h > 0$ with Λ independent of h. Then there exists a subsequence (not relabeled) and a function $y \in \mathbb{A}(g, \Phi)$ such that $y_h \to y$ in $[L^2(\Omega)]^3$, $\nabla_h y_h \to \nabla y$ in $[L^2(\Omega)]^{3\times2}$, $y_h \to y$ in $[L^2(\partial\Omega)]^3$ and $\nabla_h y_h \to \nabla y$ in $[L^2(\partial\Omega)]^{2\times3}$ as $h, \varepsilon \to 0$.*

Proof. We proceed in several steps.

Step 1: Weak convergence in L^2. We first use (5.1) to deduce that $\|y_h\|_{H_h^2(\Omega)} \leq C$ for all h. We next employ the Friedrichs-type inequality (2.27) to obtain $\|y_h\|_{L^2(\Omega)} \leq C$ for all h. Consequently, there exists a subsequence of $\{y_h\}_{h>0}$ (not relabeled) converging weakly in $[L^2(\Omega)]^3$ to some $y \in [L^2(\Omega)]^3$. We must show that $y \in \mathbb{V}(g, \Phi)$.

Step 2: H^1-regularity of y and L^2-strong convergence. To prove that $y \in [H^1(\Omega)]^3$ we need to regularize y_h. To this end, we employ the smoothing interpolation operator Π_h defined in Section 2.3 and let $z_h := \Pi_h y_h \in [\mathring{\mathbb{V}}_h^k]^3 = [\mathbb{V}_h^k \cap H_0^1(\Omega)]^3$. In view of the stability bound (2.21) and the Friedrichs-type estimate (2.27), we find that

$$\|z_h\|_{L^2(\Omega)} + \|\nabla z_h\|_{L^2(\Omega)} \lesssim \|y_h\|_{L^2(\Omega)} + \|\nabla_h y_h\|_{L^2(\Omega)} + \|h^{-1/2}[y_h]\|_{L^2(\Gamma_h^0)}$$
$$\lesssim \|y_h\|_{H_h^2(\Omega)} + \|g\|_{H^1(\Omega)} + \|\Phi\|_{H^1(\Omega)} \leq C,$$

because of equi-coercivity property (5.1). Since z_h is uniformly bounded in H^1, (a subsequence of) z_h converges weakly to some $z \in [H^1(\Omega)]^3$ and strongly to z in $[L^2(\Omega)]^3$ and $[L^2(\partial\Omega)]^3$. Properties (2.21) and (2.22) of Π_h yield

$$\|y_h - \Pi_h y_h\|_{L^2(\Omega)} + \|y_h - \Pi_h y_h\|_{L^2(\partial\Omega)} \lesssim h^{\frac{1}{2}} \left(\|\nabla_h y_h\|_{L^2(\Omega)} + \|h^{-\frac{1}{2}}[y_h]\|_{L^2(\Gamma_h^0)} \right) \lesssim h^{\frac{1}{2}},$$

whence as $h \to 0$

$$\|y_h - z\|_{L^2(\Omega)} \leq \|y_h - \Pi_h y_h\|_{L^2(\Omega)} + \|z_h - z\|_{L^2(\Omega)} \to 0,$$
$$\|y_h - z\|_{L^2(\partial\Omega)} \leq \|y_h - \Pi_h y_h\|_{L^2(\partial\Omega)} + \|z_h - z\|_{L^2(\partial\Omega)} \to 0.$$

The uniqueness of the weak $L^2(\Omega)$ limit implies that $y = z \in [H^1(\Omega)]^3$ and that y_h converges to y strongly in $[L^2(\Omega)]^3$. Regarding the trace, we first observe that

$$\|h^{-3/2}(y_h - g)\|_{L^2(\partial_D\Omega)} = \|h^{-3/2}[y_h]\|_{L^2(\Gamma_h^b)} \leq \|y_h\|_{H_h^2(\Omega)} \leq C,$$

according to (2.26) and (5.1); hence $\|y_h - g\|_{L^2(\partial_D\Omega)} \to 0$ as $h \to 0$. Since

$$z - g = (z - z_h) + (\Pi_h y_h - y_h) + (y_h - g),$$

and $\|\Pi_h y_h - y_h\|_{L^2(\partial_D\Omega)} \to 0$ as $h \to 0$, in view of Corollary 2.1 (boundary error estimate) and (5.1), we infer that $y = z = g$ on $\partial_D\Omega$.

Step 3: H^2-regularity of y and H^1-strong convergence. This entails repeating Step 2 with $\nabla_h y_h$. We thus define $Z_h := \Pi_h(\nabla_h y_h) \in [\mathring{\mathbb{V}}_h^k]^{3\times 2}$. Applying again the stability bound (2.21) in conjunction with (2.26) and (5.1) gives the uniform bound

$$\|Z_h\|_{L^2(\Omega)} + \|\nabla Z_h\|_{L^2(\Omega)} \lesssim \|\nabla_h y_h\|_{L^2(\Omega)} + \|D_h^2 y_h\|_{L^2(\Omega)} + \|h^{-1/2}[\nabla_h y_h]\|_{L^2(\Gamma_h^0)} \le C.$$

Hence, (a subsequence of) Z_h converges weakly in $[H^1(\Omega)]^{3\times 2}$ and strongly in $[L^2(\Omega)]^{3\times 2}$ and $[L^2(\partial\Omega)]^{3\times 2}$ to a function $Z \in [H^1(\Omega)]^{3\times 2}$. Moreover, an argument similar to Step 2, again relying on Corollary 2.1, yields $Z = \Phi$ on $\partial_D\Omega$ and $\|\nabla_h y_h - Z\|_{L^2(\Omega)}, \|\nabla_h y_h - Z\|_{L^2(\partial\Omega)} \to 0$ as $h \to 0$.

It remains to show that $Z = \nabla y$, whence $y \in [H^2(\Omega)]^3$. For any test function $\phi \in [C_0^\infty(\Omega)]^{3\times 2}$, elementwise integration by parts leads to

$$\int_\Omega \nabla_h y_h : \phi = -\int_\Omega y_h \cdot \mathrm{div}\phi + \int_{\Gamma_h^0} [y_h] \cdot \phi\,\mu.$$

We next show that the last term tends to 0 as $h \to 0$. Note that

$$\left| \int_{\Gamma_h^0} [y_h] \cdot \phi\,\mu \right| \lesssim \|h^{-1/2}[y_h]\|_{L^2(\Gamma_h^0)}\, \|h^{1/2}\phi\|_{L^2(\Gamma_h^0)} \to 0 \qquad \text{as } h \to 0,$$

because $\|h^{-1/2}[y_h]\|_{L^2(\Gamma_h)} \le h\|y_h\|_{H_h^2(\Omega)} \to 0$ while $\|h^{1/2}\phi\|_{L^2(\Gamma_h^0)} \lesssim |\Omega|\|\phi\|_{L^\infty(\Omega)} \le C$. This in turn implies that as $h \to 0$

$$\int_\Omega Z : \phi = -\int_\Omega y \cdot \mathrm{div}\phi,$$

or equivalently $Z = \nabla y$.

Step 4: Isometry constraint. To show that y satisfies (1.1), we combine the discrete isometry defect (1.6) with a bound for the broken Hessian $D_h^2 y_h$. The former controls the isometry constraint mean over elements, whereas the latter controls oscillations of $\nabla_h y_h$. In fact, we prove

$$(5.2) \qquad \sum_{T\in\mathcal{T}_h} \left\|(\nabla_h y_h)^T \nabla_h y_h - I\right\|_{L^1(T)} \lesssim h + \varepsilon.$$

Let $M_T[y_h] \in \mathbb{R}^{2\times 2}$ denote the *mean* of the isometry constraint over $T \in \mathcal{T}_h$

$$M_T[y_h] := |T|^{-1} \int_T \left((\nabla_h y_h)^T \nabla_h y_h - I\right),$$

and write

$$\left\|(\nabla_h y_h)^T \nabla_h y_h - I\right\|_{L^1(T)} \le \left\|((\nabla_h y_h)^T \nabla_h y_h - I) - M_T[y_h]\right\|_{L^1(T)} + \left\|M_T[y_h]\right\|_{L^1(T)}.$$

For the second term, we use that $y_h \in \mathbb{A}_{h,\varepsilon}^k(g,\Phi)$, defined in (2.17)), to deduce that

$$\left\|M_T[y_h]\right\|_{L^1(T)} = |T|\,|M_T[y_h]| \quad \Rightarrow \quad \sum_{T\in\mathcal{T}_h} |T|\,|M_T[y_h]| = D_h[y_h] \le \varepsilon.$$

For the first term, instead, we apply the Poincaré inequality in $L^1(T)$ to obtain

$$\left\|((\nabla_h y_h)^T \nabla_h y_h - I) - M_T[y_h]\right\|_{L^1(T)} \lesssim h_T \left\|\nabla_h\left((\nabla_h y_h)^T \nabla_h y_h\right)\right\|_{L^1(T)}$$

$$\lesssim h_T \|D_h^2 y_h\|_{L^2(T)}\|\nabla_h y_h\|_{L^2(T)}.$$

Summing over $T \in \mathcal{T}_h$ and using the Friedrichs-type inequality (2.27), together with Lemma 2.3 (coercivity of $E_h[y_h]$) and the uniform bound on $E_h[y_h] \leq \Lambda$, we get

$$\sum_{T \in \mathcal{T}_h} \left\| ((\nabla_h y_h)^T \nabla_h y_h - I) - M_T[y_h] \right\|_{L^1(T)} \lesssim h \, \|D_h^2 y_h\|_{L^2(\Omega)} \|\nabla_h y_h\|_{L^2(\Omega)} \lesssim h,$$

where the hidden constant depends on Λ and data (g, Φ, f). This shows (5.2).

With this at hand, we now show that $\|\nabla y^T \nabla y - I\|_{L^1(\Omega)} = 0$. We observe

$$\nabla y^T \nabla y - I = \nabla y^T (\nabla y - \nabla_h y_h) + (\nabla y^T - (\nabla_h y_h)^T) \nabla_h y_h + (\nabla_h y_h)^T \nabla_h y_h - I,$$

which implies

$$\|\nabla y^T \nabla y - I\|_{L^1(\Omega)} \lesssim \left(\|\nabla y\|_{L^2(\Omega)} + \|\nabla_h y_h\|_{L^2(\Omega)} \right) \|\nabla_h y_h - \nabla y\|_{L^2(\Omega)} + h + \varepsilon \to 0$$

as $h, \varepsilon \to 0$, upon recalling the strong convergence of $\nabla_h y_h$ to ∇y in $[L^2(\Omega)]^{3 \times 2 \times 2}$ from Step 3 and the uniform bound of $\|\nabla_h y_h\|_{L^2(\Omega)}$ from (2.27). The proof is thus complete. $\qquad\square$

5.2. **Lim-inf property of E_h.**

The lim-inf property easily follows from the preceding results, but we prove it here for completeness. Instead of looking at a general sequence $\{y_h\}_{h>0}$ with $y_h \to y$ in $[L^2(\Omega)]^3$, we assume $E_h[y_h] \leq \Lambda$ for all h and uniform constant $\Lambda > 0$, since otherwise $\liminf_{h \to 0} E_h[y_h] = \infty$ and the lim-inf inequality is trivial.

Proposition 5.2 (lim-inf property)**.** *Let data (g, Φ, f) satisfy (1.2). Let the penalty parameters γ_0 and γ_1 in (2.16) be chosen sufficiently large, and let the discrete isometry defect parameter $\varepsilon = \varepsilon(h) \to 0$ as $h \to 0$. Let $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ and $E_h[y_h] \leq \Lambda$ for all h with Λ independent of h. Then, there exists $y \in \mathbb{A}(g, \Phi)$ such that $y_h \to y$ in $[L^2(\Omega)]^3$, $\nabla_h y_h \to \nabla y$ in $[L^2(\Omega)]^{3 \times 2}$, $y_h \to y$ in $[L^2(\partial\Omega)]^3$ and $\nabla_h y_h \to \nabla y$ in $[L^2(\partial\Omega)]^{2 \times 3}$ as $h \to 0$ and*

$$E[y] \leq \liminf_{h \to 0} E_h[y_h].$$

Proof. Since $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ and $E_h[y_h] \leq \Lambda$, we invoke Proposition 5.1 (compactness in $L^2(\Omega)$) to get $y \in \mathbb{A}(g, \Phi)$, namely y satisfies the Dirichlet boundary conditions on $\partial_D \Omega$ and the isometry constraint (1.1), whereas y_h satisfies $y_h \to y$ in $[L^2(\Omega)]^3$, $\nabla_h y_h \to \nabla y$ in $[L^2(\Omega)]^{3 \times 2}$, $y_h \to y$ in $[L^2(\partial\Omega)]^3$ and $\nabla_h y_h \to \nabla y$ in $[L^2(\partial\Omega)]^{2 \times 3}$ as $h, \varepsilon \to 0$. Furthermore, the equi-coercivity bound (5.1) in conjunction with Proposition 4.4 (lim inf for the reconstructed Hessian) guarantees that

$$\|D^2 y\|_{L^2(\Omega)} \leq \liminf_{h \to 0} \|H_h[y_h]\|_{L^2(\Omega)}.$$

This, together with $(f, y_h)_{L^2(\Omega)} \to (f, y)_{L^2(\Omega)}$ as $h \to 0$, yields

$$\frac{1}{2} \int_\Omega |D^2 y|^2 - \int_\Omega f \cdot y \leq \frac{1}{2} \liminf_{h \to 0} \left(\int_\Omega |H_h(y_h)|^2 - \int_\Omega f \cdot y_h \right).$$

We next employ Proposition 4.5 (relation between $E_h[y_h]$ and $H_h[y_h]$), i.e.

$$\frac{1}{2} \int_\Omega |H_h(y_h)|^2 - \int_\Omega f \cdot y_h \leq E_h[y_h],$$

and combine the last two inequalities to deduce the asserted estimate. $\qquad\square$

5.3. **Lim-sup property of E_h.** Since we are interested in minimizers of E within the admissible set $\mathbb{A}(g, \Phi)$, to show the lim-sup property we construct a recovery sequence in $[L^2(\Omega)]^3$ for a function $y \in \mathbb{A}(g, \Phi)$ via regularization and Lagrange interpolation. Since the isometry and Dirichlet boundary contraints are relaxed via (1.6) and the Nitsche's approach, we do not need to preserve them in the regularization and interpolation procedures. This extra flexibility is an improvement over [5, 8], where the lim-sup property is proven under the assumption that both procedures preserve those constraints at the nodes using an intricate approximation argument by P. Hornung [26].

Proposition 5.3 (lim-sup property). *If the parameters γ_0 and γ_1 in (2.14) are sufficiently large, then for any $y \in \mathbb{A}(g, \Phi)$, there exists a recovery sequence $\{y_h\}_h \subset \mathbb{A}_{h,\varepsilon}^k(g, \Phi) \cap [H^1(\Omega)]^{3 \times 2}$ with $\|y_h\|_{H_h^2(\Omega)} \lesssim \|y\|_{H^2(\Omega)}$ uniformly in h such that*

$$y_h \to y \quad in \quad [L^2(\Omega)]^3, \qquad \nabla y_h \to \nabla y \quad in \quad [L^2(\Omega)]^{3 \times 2} \qquad as \ h \to 0,$$

and

$$\limsup_{h, \varepsilon \to 0} E_h[y_h] \le E[y],$$

provided $h \le 1$ and $\varepsilon := \varepsilon(h) \ge \varepsilon_1$, where

$$(5.3) \qquad \varepsilon_1 := Ch\|y\|_{H^2(\Omega)}^2$$

and C is an interpolation constant that depends on the shape regularity of $\{\mathcal{T}_h\}_h$.

Proof. We proceed in several steps.

Step 1: Recovery sequence. Let $\Lambda := \|y\|_{H^2(\Omega)}$ and let $y_h := I_h y \in [\mathring{\mathbb{V}}_h^k]^3 = [\mathbb{V}_h^k \cap H^1(\Omega)]^3$ be the standard Lagrange interpolant of y, which is well-defined in $\mathbb{R}^2$ because $H^2(\Omega) \subset C^0(\overline{\Omega})$. We first resort to the local estimate (9.9), namely

$$(5.4) \qquad \|D_h^2 y_h\|_{L^2(T)} \lesssim |y|_{H^2(T)} \quad \forall T \in \mathcal{T}_h,$$

which is valid for isoparametric maps. Moreover, $\|[y_h]\|_{L^2(e)} = 0$ for all $e \in \mathcal{E}_h^0$ because $y_h \in C^0(\overline{\Omega})$, whereas combining a trace inequality with (9.10) yields

$$\|[y_h]\|_{L^2(e)} = \|I_h y - y\|_{L^2(e)} \lesssim h_e^{3/2} |y|_{H^2(\omega(e))} \quad \forall e \in \mathcal{E}_h^b,$$

because $y = g$ on $\partial_D\Omega$ and (2.10). Similarly, for all $e \in \mathcal{E}_h^0$ we have

$$\left\|[\nabla y_h]\right\|_{L^2(e)} \le \left\|\nabla(I_h y^+ - y)\right\|_{L^2(e)} + \left\|\nabla(I_h y^- - y)\right\|_{L^2(e)} \lesssim h_e^{1/2} |y|_{H^2(\omega(e))},$$

where $y^\pm = y|_{T^\pm}$ and $T^\pm$ are adjacent elements such that $e = T^+ \cap T^-$ and $\omega(e) = T^+ \cup T^-$. Moreover, for $e \in \mathcal{E}_h^b$ we have $[\nabla y_h]_e = \nabla(I_h y - y)$ according to (2.10), because $\nabla y = \Phi$ on $\partial_D\Omega$, and we proceed as before. Therefore, we infer that

$$\|y_h\|_{H_h^2(\Omega)} \lesssim |y|_{H^2(\Omega)} \quad \Rightarrow \quad E_h[y_h] \lesssim \Lambda^2 + R(g, \Phi, f),$$

where the last inequality is a consequence of the upper bound in (2.29).

Step 2: Discrete isometry defect. We claim that y_h satisfies (1.6) for $\varepsilon \ge \varepsilon_1$. Since

$$\|\nabla y_h\|_{L^2(T)} \le \|\nabla y\|_{L^2(T)} + \|\nabla(y_h - y)\|_{L^2(T)} \le \|\nabla y\|_{L^2(T)} + Ch_T |y|_{H^2(T)},$$

in light of (9.10), adding and substracting y and recalling $\nabla y^T \nabla y = I$ we obtain

$$\nabla y_h^T \nabla y_h - I = \nabla(y_h - y)^T \nabla y_h + \nabla y^T \nabla(y_h - y),$$

whence $D_h[y_h]$ defined in (1.6) satisfies for $h \leq 1$

$$\begin{aligned} D_h[y_h] &\leq \sum_{T \in \mathcal{T}_h} \left(\|\nabla y_h\|_{L^2(T)} + \|\nabla y\|_{L^2(T)} \right) \|\nabla(y - y_h)\|_{L^2(T)} \\ &\lesssim \sum_{T \in \mathcal{T}_h} h_T \|\nabla y\|_{L^2(T)} \, |y|_{H^2(T)} + \sum_{T \in \mathcal{T}_h} h_T^2 |y|_{H^2(T)}^2 \\ &\leq h \|\nabla y\|_{L^2(\Omega)} |y|_{H^2(\Omega)} + h^2 |y|_{H^2(\Omega)}^2 \leq 2\Lambda^2 h. \end{aligned}$$

Therefore, if $\varepsilon_1 = Ch\Lambda^2 = Ch\|y\|_{H^2(\Omega)}^2$, we deduce that $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ for $\varepsilon \geq \varepsilon_1$.

Step 3: Convergence of $E_h[y_h]$. It remains to show that as $h, \varepsilon \to 0$

$$E_h[y_h] \to E[y],$$

with $\varepsilon \geq \varepsilon_1$. We focus on two critical terms in $E_h[y_h]$ in (2.16), namely

$$\|D_h^2 y_h\|_{L^2(\Omega)} \to \|D^2 y\|_{L^2(\Omega)}, \quad \|h^{-1/2} [\nabla_h y_h]\|_{L^2(\Gamma_h)} \to 0,$$

because similar arguments yield $\|h^{-3/2} [y_h]\|_{L^2(\Gamma_h)} \to 0$ as well as convergence of the remaining terms in $E_h[y_h]$ upon proceeding as in Lemma 2.3 (coercivity of E_h).

Since y is merely in $[H^2(\Omega)]^3$, we argue by density. Let $\{y^\sigma\}_{\sigma>0} \subset [C^\infty(\mathbb{R}^2)]^3$ be a sequence of regularizations of y such that $y^\sigma \to y$ in $[H^2(\Omega)]^3$ as $\sigma \to 0$, and let $y_h^\sigma = I_h y^\sigma \in [\mathring{\mathbb{V}}_h^k]^3$ be the Lagrange interpolant of y^σ. Estimate (5.4) implies

$$\|D^2(y_h - y_h^\sigma)\|_{L^2(T)} \lesssim |y - y^\sigma|_{H^2(T)},$$

whence writing $y - y_h = (y - y^\sigma) - I_h(y - y^\sigma) + (y^\sigma - y_h^\sigma)$ gives

$$h_T^{-1} \|\nabla(y - y_h)\|_{L^2(T)} + \|D^2(y - y_h)\|_{L^2(T)} \lesssim |y - y^\sigma|_{H^2(T)} + h_T \|y^\sigma\|_{H^3(T)},$$

because the polynomial degree is $k \geq 2$. This can be made arbitrarily small upon choosing first the coarse scale σ and next h, and implies $\|D_h^2 y_h\|_{L^2(\Omega)} \to \|D^2 y\|_{L^2(\Omega)}$ as $h \to 0$ by the triangle inequality.

We finally examine the stabilization term $\|h^{-1/2} [\nabla_h y_h]\|_{L^2(\Gamma_h)}$. We recall that $[\nabla y] = 0$ for all interior edges $e \in \mathcal{E}_h^0$ and $\nabla y = \Phi$ for boundary edges $e \in \mathcal{E}_h^b$. We next utilize the scaled trace inequality in conjunction with (2.10) to write

$$\begin{aligned} \left\| h^{-1/2} [\nabla_h y_h] \right\|_{L^2(e)} &= \left\| h^{-1/2} [\nabla_h (y - y_h)] \right\|_{L^2(e)} \\ &\lesssim h_e^{-1} \|\nabla_h(y - y_h)\|_{L^2(\omega(e))} + \|D_h^2(y - y_h)\|_{L^2(\omega(e))}, \end{aligned}$$

whence $\|h^{-1/2} [\nabla_h y_h]\|_{L^2(\Gamma_h)} \to 0$ as $h \to 0$. This concludes the proof. $\square$

We point out the occurrence of the H^2-seminorm on the right-hand side of (5.4) for both affine equivalent and isoparametric elements (including subdivisions made of quadrilaterals). We present a proof in Section 9 and refer to [19] and Chapter 13 of [22] for a comprehensive discussion of isoparametric elements.

5.4. **Convergence of global minimizers.** We now show that cluster points of global minimizers of E_h are global minimizers of E, without assuming the existence of the latter. The proof combines Propositions 5.1 (compactness in $L^2(\Omega)$), 5.2 (lim-inf property) and 5.3 (lim-sup property).

We now collect properties of the *nonconvex* discrete admissible set $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ of (2.17), the discrete counterpart of the nonconvex set $\mathbb{A}(g, \Phi)$ of (1.5). Given an initial guess $y_h^0 \in \mathbb{V}_h^k(g, \Phi)$ with isometry defect $D_h[y_h^0] \leq \tau$ and energy $E_h[y_h^0] \leq$

Λ_0 independently of h, Lemma 3.4 (discrete isometry defect) guarantees that the discrete H^2-gradient flow produces a sequence $\{y_h^n\}_n \subset \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ with $\varepsilon \geq \varepsilon_0$ and

$$\varepsilon_0 = \Big(1 + C_F^2\big(\Lambda_0 + C_F^* R(g, \Phi, f)\big)\Big)\tau.$$

This shows that $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ is non-empty provided $\varepsilon \geq \varepsilon_0$ and that ε_0 can be made arbitrarily small as $\tau \to 0$. We could take τ proportional to h, which is the choice in Section 6. Moreover, Lemma 3.3 (energy decay) shows that $E_h[y_h] \leq \Lambda_0$ whereas Lemma 2.3 (coercivity of E_h) gives

$$\|y_h\|_{H_h^2(\Omega)}^2 \leq \Lambda_0 + C_F^* R(g, \Phi, f).$$

This is the setting for Γ-convergence and is discussed next.

Theorem 5.4 (convergence of global minimizers). *Let data* (g, Φ, f) *satisfy* (1.2). *Let the penalty parameters* γ_0 *and* γ_1 *in* (2.14) *be chosen sufficiently large. Let* $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ *be a sequence of functions such that* $E_h[y_h] \leq \Lambda_0$, *for a constant* Λ_0 *independent of* h, *and let the discrete isometry defect parameter* ε *satisfy*

$$\varepsilon \geq \varepsilon_1 := C\big(\Lambda_0 + R(g, \Phi, f)\big)h,$$

where C *is an interpolation constant depending on the shape regularity of* $\{\mathcal{T}_h\}$. *If* $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ *is an almost global minimizer of* E_h *in the sense that*

$$E_h[y_h] \leq \inf_{w_h \in \mathbb{A}_{h,\varepsilon}^k(g,\Phi)} E_h[w_h] + \sigma,$$

where $\sigma, \varepsilon \to 0$ *as* $h \to 0$, *then* $\{y_h\}_h$ *is precompact in* $[L^2(\Omega)]^3$ *and every cluster point* y *of* y_h *belongs to* $\mathbb{A}(g, \Phi)$ *and is a global minimizer of* E, *namely* $E[y] = \inf_{w \in \mathbb{A}(g,\Phi)} E[w]$. *Moreover, up to a subsequence (not relabeled), the energies converge*

$$\lim_{h \to 0} E_h[y_h] = E[y].$$

Proof. This proof is standard and given for completeness. Since $E_h[y_h] \leq \Lambda_0$, we invoke Proposition 5.2 (lim-inf property) to deduce the existence of $y \in \mathbb{A}(g, \Phi)$ such that a (non-relabeled) subsequence of $y_h \to y$ in $[L^2(\Omega)]^3$ and

$$E[y] \leq \liminf_{h \to 0} E_h[y_h].$$

This means that $\mathbb{A}(g, \Phi)$ is non-empty and that

$$\inf_{z \in \mathbb{A}(g,\Phi)} E[z] \leq E[y] \leq \Lambda_0.$$

To show that y is a global minimizer of E, we let $0 < \eta < 1$ and $w \in \mathbb{A}(g, \Phi)$ satisfy

$$E[w] \leq \inf_{z \in \mathbb{A}(g,\Phi)} E[z] + \eta \leq \eta + \Lambda_0.$$

Lemma 2.2 (coercivity of E) yields $\|w\|_{H^2(\Omega)}^2 \lesssim \eta + \Lambda_0 + R(g, \Phi, f) =: \Lambda_1$ whence, if $\varepsilon \geq \varepsilon_1 = Ch\Lambda_1$, Proposition 5.3 (lim-sup property) gives a recovery sequence $w_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ of w such that $w_h \to w$ in $[L^2(\Omega)]^3$ as $h, \varepsilon \to 0$ and

$$\limsup_{h,\varepsilon \to 0} E_h[w_h] \leq E[w].$$

We next utilize that $E_h[y_h] \leq E_h[w_h] + \sigma$, because y_h is an almost global minimizer of E_h within $\mathbb{A}_{h,\varepsilon}^k(g, \Phi)$, to deduce

$$E[y] \leq \liminf_{h \to 0} E_h[y_h] \leq \limsup_{h \to 0} \big(E_h[w_h] + \sigma\big) \leq E[w] \leq \inf_{z \in \mathbb{A}(g,\Phi)} E[z] + \eta.$$

Taking $\eta \to 0$ implies that y is an global minimizer of E and

$$\lim_{h \to 0} E_h[y_h] = E[y].$$

This concludes the proof.   □

5.5. **Strong Convergence of $H_h[y_h]$ and Scaled Jumps.** We now exploit The-
orem 5.4 (convergence of global minimizers) to strengthen the assertions of Propo-
sition 5.1 (compactness in $L^2(\Omega)$), in the spirit of [20, 31]. In fact, we show strong
convergence of the scaled jump terms to zero and of the discrete Hessian $H_h[y_h]$ as
well as broken Hessian $D_h^2 y_h$ of y_h to $D^2 y$ as $h \to 0$.

Corollary 5.1 (strong convergence of Hessian and scaled jumps). *Let* (g, Φ, f)
satisfy (1.2) *and the penalty parameters* γ_0 *and* γ_1 *in* (2.14) *be chosen sufficiently
large so that* (4.22) *holds. Let* $y_h \in \mathbb{A}_{h,\varepsilon}^k(g, \Phi)$ *be a subsequence of almost global
minimizers of* E_h *converging to a global minimizer* $y \in \mathbb{A}(g, \Phi)$ *of* E, *as established
in Theorem 5.4. Then, the following statements are valid as* $h \to 0$

(i) $H_h[y_h] \to D^2 y$ *strongly in* $[L^2(\Omega)]^{3 \times 2 \times 2}$;
(ii) $\|h^{-1/2}[\nabla_h y_h]\|_{L^2(\Gamma_h)} + \|h^{-3/2}[y_h]\|_{L^2(\Gamma_h)} \to 0$;
(iii) $D_h^2 y_h \to D^2 y$ *strongly in* $[L^2(\Omega)]^{3 \times 2 \times 2}$.

Proof. We first observe that Theorem 5.4 (convergence of global minimizers) yields

$$E_h[y_h] \to E[y] = \frac{1}{2} \int_\Omega |D^2 y|^2 - \int_\Omega f \cdot y \qquad \text{as } h \to 0.$$

We apply Propositions 4.5 (relation between $E_h[y_h]$ and $H_h[y_h]$) to deduce

$$\frac{1}{2} \limsup_{h \to 0} \|H_h[y_h]\|_{L^2(\Omega)}^2 \le \limsup_{h \to 0} \left(E_h[y_h] + \int_\Omega f \cdot y_h \right) = \frac{1}{2} \|D^2 y\|_{L^2(\Omega)}^2.$$

We now resort to Proposition 4.4 (lim-inf property of the reconstructed Hessian).
In fact, combining (4.16) with the previous inequality gives the convergence in norm

$$\|H_h[y_h]\|_{L^2(\Omega)} \to \|D^2 y\|_{L^2(\Omega)}.$$

In addition, Proposition 4.4 guarantees that $H_h[y_h] \rightharpoonup H$ in $[L^2(\Omega)]^{3 \times 2 \times 2}$ where H
satisfies the orthogonality property (4.15). This leads to

$$\left(H_h[y_h], D^2 y \right)_{L^2(\Omega)} \to \left(H, D^2 y \right)_{L^2(\Omega)} = \|D^2 y\|_{L^2(\Omega)}^2, \quad \text{as } h \to 0.$$

The above property and the convergence of norms imply strong convergence, namely

$$\left\| H_h[y_h] - D^2 y \right\|_{L^2(\Omega)}^2 = \left\| H_h[y_h] \right\|_{L^2(\Omega)}^2 + \|D^2 y\|_{L^2(\Omega)}^2 - 2 \left(H_h[y_h], D^2 y \right)_{L^2(\Omega)} \to 0.$$

This in turn yields (i). To prove (ii) we make use of (i) to infer that as $h \to 0$

$$J_h[y_h] = E_h[y_h] - \frac{1}{2} \int_\Omega |H_h[y_h]|^2 + \int_\Omega f \cdot y_h \to E[y] - \frac{1}{2} \int_\Omega |D^2 y|^2 + \int_\Omega f \cdot y = 0,$$

where we have used the definition (4.20) of $J_h[y_h]$. In view of (4.22), we thus get

$$\|h^{-3/2}[y_h]\|_{L^2(\Gamma_h)} + \|h^{-1/2}[\nabla_h y_h]\|_{L^2(\Gamma_h)} \to 0 \qquad \text{as } h \to 0.$$

This not only establishes (ii), but combined with (i), the definition (4.8) of $H_h[y_h]$,
and the bounds (4.10) for $R_h([\nabla_h y_h])$ and $B_h([y_h])$, directly implies (iii).   □

6. NUMERICAL EXPERIMENTS

In this section we explore and compare the performance of our method with that of Kirchhoff elements [5, 6, 7, 8]. We are interested in the speed and accuracy of the method, as well as its ability to capture the essential physics and geometry of the problems appropriately. We observe that our dG approach seems to be more flexible with comparable or better speed. We present below specific examples computed within the platform deal.ii with polynomial degree $k = 2$ and uniform meshes $\mathcal{T}_h$ made of rectangles [1, 4]; hence we use $\mathbb{Q}_2(T)$ for all $T \in \mathcal{T}_h$. Moreover, one might notice that we consistently use rather large penalty parameters γ_0, γ_1 relative to the second-order case [1, 2, 4, 11, 16, 32]. These choices hinge mostly on experiments performed for the vertical load example of Subsection 6.1 and are not dictated by stability considerations exclusively. In fact, they are a compromise between the discrete initial energy $E_h[y_h^0]$ and the fictitious time step τ of the gradient flow, which obey the relations (3.9) and (3.10) and control the discrete isometry defect $D_h[y_h]$ according to (3.11). We point out that enforcing Dirichlet conditions via the Nitsche's approach depends on the magnitude of γ_0 and γ_1, which in turn affects $E_h[y_h^0]$. We examined a very wide range of γ_0 and γ_1 and compared $D_h[y_h]$ with the rate at which it decreases for each tested pair. We selected those values that lead to the smallest defect and the largest rate of convergence. We made similar choices for all subsequent examples without exhaustive testing for each of them. We note that our theory does not explicitly predict why the best convergence behavior manifests for such large values of γ_0 and γ_1. We do not provide our computational study leading to γ_0 and γ_1, for the sake of brevity.

6.1. **Vertical load on a square domain.** This is a simple example of bending due to a vertical load. We use the same configuration as in [5] in order to present a faithful comparison of the two methods. We deal with a square domain with two of the non-parallel sides being clamped.

Example 6.1. Let $\Omega = (0, 4) \times (0, 4)$ and $\partial_D \Omega = \{0\} \times [0, 4] \cup [0, 4] \times \{0\}$ be the part of the boundary where we enforce the Dirichlet boundary conditions

(6.1)
$$g = (x_1, x_2, 0) \quad \text{and} \quad \Phi = [I_2, 0]^T.$$

We apply a vertical force $f = (0, 0, F)$ of magnitude $F = 2.5 \times 10^{-2}$.

No. Cells	DoFs	$\tau = h$	$E_h[y_h]$	$D_h[y_h]$	Iterations
256	7680	$\sqrt{2}\, 2^{-2}$	-7.53e-3	4.02e-3	13
1024	30,720	$\sqrt{2}\, 2^{-3}$	-5.76e-3	1.63e-3	28
4096	122,880	$\sqrt{2}\, 2^{-4}$	-4.26e-3	6.07e-4	76
16384	491,520	$\sqrt{2}\, 2^{-5}$	-3.30e-3	2.28e-4	140

TABLE 1. Example 6.1: Number of cells, degrees of freedom, discrete energy $E_h[y_h]$, isometry defect $D_h[y_h]$ and number of gradient flow iterations for the square plate $\Omega = [0, 4]^2$, clamped in two adjacent sides with vertical forcing. We observe super-linear rates for the isometry defect, while theory predicts linear rates for the case $\tau = h$.
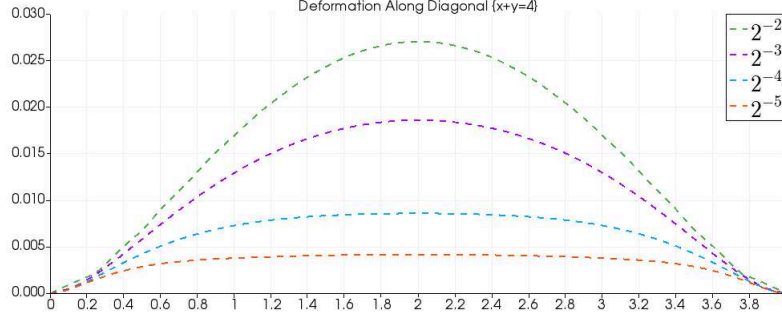
FIGURE 1. Example 6.1: Deformation along the diagonal $x_1 + x_2 = 4$. We observe smaller deformation y than with the Kirchhoff elements, up to one order of magnitude. For example, $h = 2^{-3}$ yields $|y| \approx 0.018$ for dG, while $h = 2^{-6}$ gives $|y| \approx 0.02$ for Kirchhoff elements.

We first illustrate the convergence of the energy $E_h[y_h]$ and the isometry defect $D_h[y_h]$ as h decreases. We set $\gamma_0 = 5,000$, $\gamma_1 = 1100$ and choose $\tau = h$ and observe that the isometry defect decays super-linearly with τ, which is better than the linear convergence predicted in Lemma 3.4 (discrete isometry defect). Compared to [5], we obtain a more clear rate, while the defect itself is up to one order of magnitude smaller. The number of gradient flow iterations is similar for both methods.

We now explore the geometric behavior of our method. More precisely, it was observed in [5] that there was an artificial displacement along the diagonal $x_1 + x_2 = 4$, which does not correspond to the actual physics of the problem, namely $y = 0$ for $x_1 + x_2 \le 4$. This displacement decreases with smaller mesh size h. Our method introduces the same artificial deformation. However, we can see in Figure 1 that this displacement is smaller now, up to one order of magnitude for the last refinement.

6.2. **Obstacle Problem.** In Example 6.1 we document the flexibility of the dG approach to capture the correct bending behavior of plates. To explore this further, we introduce an extra element to the deformation: a rigid obstacle. We use a square plate clamped on one side and we exert a vertical force. We require that the plate does not penetrate the obstacle. This example is motivated by [7], where the deformation of the plate is the result of thermal actuation of bilayer hinges connected to the plate.

Example 6.2. Let $\Omega = (-1, 1) \times (-1, 1)$ and $\partial_D \Omega = \{-1\} \times [-1, 1]$ be the Dirichlet boundary where we enforce clamped boundary conditions (6.1). We apply the vertical force $f = (0, 0, 1)$. We choose a simple case where the obstacle is a rigid flat plate at height $z = 0.2$. From a mathematical viewpoint, this obstacle problem can be treated by introducing the convex set of admissible deformations

$$K = \left\{ y \in [L^2(\Omega)]^3 \ : \quad y_3 \le 0.2 \right\},$$

along with splitting of variables. We introduce another deformation s, $s \approx y$ such that $s \in K$ always and penalize the $L^2(\Omega)$-distance between y and s. If σ is an obstacle penalty parameter, we add the following extra term to the energy $E[y]$

$$\frac{1}{\sigma} \|y - s\|_{L^2(\Omega)}^2.$$

At the discrete level, this affects the gradient flow at each step, where we use as s_h the L^2-projection of the previous solution y_h^n in K. We refer to [7] for more details about the variable splitting and the projection.

The plate configurations before contact are unaffected by the obstacle presence and dictated by pure bending. Contact occurs with minor penetration of the obstacle because we do not force the deformation y_h to belong to the set K, but rather penalize its L^2-distance from its projection to K. As time evolves, the plate bends backwards trying to decrease the obstacle crossing because it is energetically costly. This behavior can be enhanced by further mesh refinement and by choosing a smaller obstacle parameter σ. To illustrate the obstacle-plate interaction, we provide six different steps of the gradient flow in Figure 2. The illustration corresponds to 1024 cells, $\sigma = 3 \times 10^{-4}$ and time step $\tau = 5 \times 10^{-4}$. The penalty parameters are $\gamma_0 = \gamma_1 = 5000$, and the isometry defect at the end of this simulation is 9×10^{-5}.

(A) t=500ts

(B) t=700ts

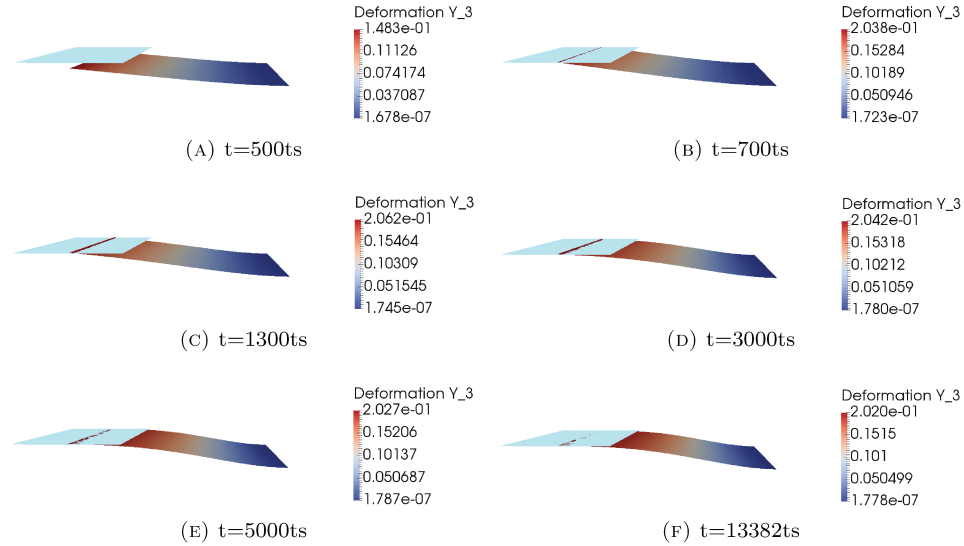(C) t=1300ts

(D) t=3000ts

(E) t=5000ts

(F) t=13382ts

FIGURE 2. Example 6.2 (deformation of the plate with obstacle dictated only by forcing): (a) Configuration before contact, (b) First contact with the obstacle, (c) Penetration of the obstacle, (d) Plate starts adjusting shape to decrease crossing, (e) Stronger bending with smaller crossing, (f) Final stage with obtacle crossing of about 0.002.

6.3. **Compressive Case: Buckling.** We finally investigate the geometrically and physically interesting case of buckling. We use a rectangular plate with a small vertical load, that induces a bias for bending, and we impose compressive boundary conditions on two opposite sides. The latter reveals the delicate interplay between the Nitsche's approach for enforcement of Dirichlet boundary conditions and the choice of suitable initial configurations, for the performance of the gradient flow. This example is also motivated by [5].

Example 6.3. Let $\Omega = (-2, 2) \times (0, 1)$ and $\partial_D \Omega = \{-2, 2\} \times [0, 1]$ be the two sides where we impose the compressive boundary conditions

$$g = (x_1 \pm 1.4, x_2, 0) \quad \text{and} \quad \Phi = [I_2, 0]^T.$$

We apply a vertical force f of magnitude 10^{-2}. Given the compressive nature of the problem, the plate could bend either upwards or downwards (buckling), resulting in two deformations which are the reflection of each other with respect to the $x_1 - x_2$ plane and have the same minimal energy. This is why we apply a small force upwards to select the deformation with positive third component. Since our initial configuration is flat, there is a boundary mismatch that leads to large boundary penalty terms in the Nitsche's approach and correspondingly large initial energy $E_h[y_h^0]$. This in turn gives rise to either large isometry defects, because of (3.10) and (3.11), or tiny time steps τ and very slow evolution. This might explain why we need a stronger force than the one in [5] (i.e. 10^{-5}) because $E_h[y_h^0]$ is dominated by the boundary terms. Moreover, to prevent a large $E_h[y_h^0]$ from creating very abrupt and non-physical deformations during the gradient flow for moderate τ, we employ a quasi-static approach: we enforce the boundary conditions gradually thereby avoiding a large mismatch (parameter continuation). We use a parameter α that starts from zero and increases throughout the flow until it reaches the value 1. In order to achieve a gradual adjustment to the boundary conditions, we let

$$\varphi(\alpha) := (1 - \alpha) \, id + \alpha \, g, \qquad 0 \leq \alpha \leq 1,$$

and $\Phi(\alpha) = \Phi$ be the Dirichlet boundary conditions for y and ∇y, where id stands for the identity function. As the mesh $\mathcal{T}_h$ becomes finer, the growth of parameter α must be slower to compensate for the larger initial energy associated with mesh-dependent boundary terms as well as to allow for smooth flow evolutions.

(A) $\alpha = 0.05$

(B) $\alpha = 0.1$

(C) $\alpha = 0.25$

(D) $\alpha = 0.6$
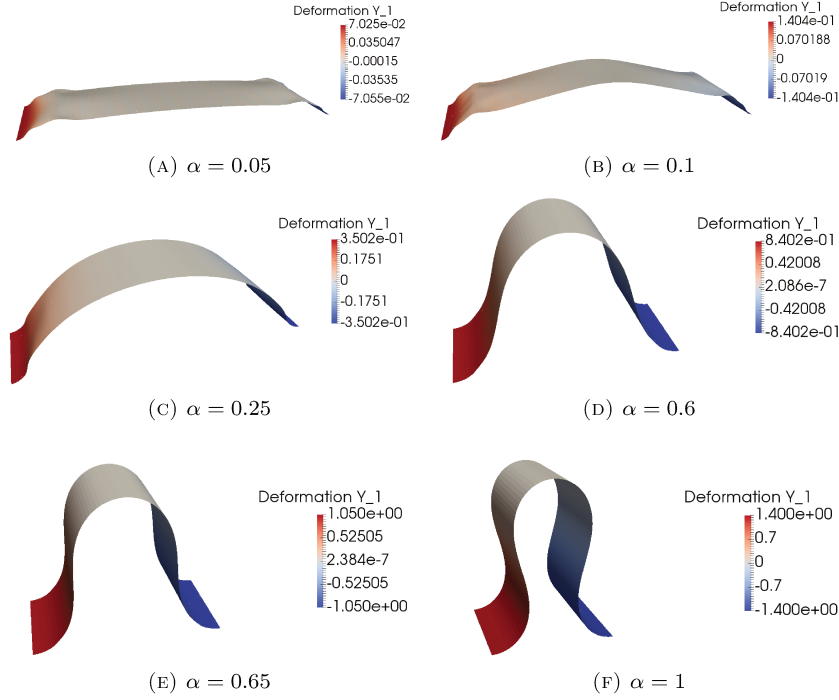
(E) $\alpha = 0.65$

(F) $\alpha = 1$

FIGURE 3. Example 6.3 (buckling of a rectangular plate at six different stages): (a)-(c) Initial stages with small deformations dictated mostly by the forcing, (d)-(e) The compressive boundary conditions dominate the evolution and induce large bending deformations, (f) Final deformation with attained compressive boundary condition.

To illustrate the effect of α, we depict deformations in Figure 3 that corresponds to various stages of the gradient flow with $\gamma_0 = \gamma_1 = 10^4$ and time step $\tau = 0.04625$. The parameter α increases linearly by the amount 5×10^{-5} in each iteration. We see that the compressive nature of the boundary conditions becomes more apparent after α exceeds the value $1/2$ and that the boundary conditions are attained at the end of the deformation. The final isometry defect is $D_h[y_h] = 5.04 \, 10^{-3}$, one order of magnitude smaller than observed in [5].

7. IMPLEMENTATION

We now make some implementation remarks and connect them with the theory in Sections 2 and 3. If $(\boldsymbol{\delta Y}_h^{n+1}, \boldsymbol{\Lambda}_h^{n+1})$ are the nodal values of functions $(\delta y_h^{n+1}, \lambda_h^{n+1})$, then the matrix form of the saddle point problem (3.7) reads

$$(7.1) \qquad \begin{bmatrix} A & B_n^T \\ B_n & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta Y}_h^{n+1} \\ \boldsymbol{\Lambda}_h^{n+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_n \\ \boldsymbol{0} \end{bmatrix},$$

where A is the matrix corresponding to the left-hand side of (3.4) and B_n is the matrix associated with (3.6), which depends on y_h^n. Since A does not depend on y_h^n we can assemble it and perform its LU decomposition once and subsequently use a direct solver whenever we need the action of A^{-1}. For the full system we use

the Schur complement approach with a conjugate gradient iterative solver, to first solve for $\mathbf{\Lambda}_h^{n+1}$ and next recover $\boldsymbol{\delta Y}_h^{n+1}$. The numerical experiments of Section 6 reveal the existence of solution $(\delta y_h^{n+1}, \lambda_h^{n+1})$ of the discrete gradient flow (3.7), which exhibits small isometry defects asymptotically as n grows and confirm the validity of an inf-sup condition for (3.7). However, this issue remains open (see Remark 3.2).

Lastly, it is important to mention that our choice in Section 2 of a space $\mathbb{V}_h^k$ of discontinuous polynomials is strongly motivated by the structure of (7.1). More precisely, if δy_h^{n+1} were continuous then the inf-sup condition for ℓ_n in (3.6) would not be local and be harder to achieve. In fact, computational experiments for continuous functions (not reported here) indicate that the conjugate gradient method becomes significantly slower if the functions of $\mathbb{V}_h^k$ are required to be continuous; this justifies our choice of a fully discontinuous space $\mathbb{V}_h^k$. We refer to [29] for details.

8. Conclusions

In this work we design, analyze and implement a dG approach to construct minimizers for large bending deformations under a nonlinear isometry constraint. The problem is nonconvex and falls within the nonlinear Kirchhoff plate theory. We propose a discrete energy functional and provide a flexible approximation of the isometry constraint. We devise a discrete gradient flow for computing discrete minimizers and enforcing a discrete isometry defect. We construct a discrete approximation of the Hessian inspired by [20, 31] that turns out to be instrumental to prove $\Gamma-$convergence: convergence of the discrete energy to the continuous one as well as L^2-convergence of global minimizers of the discrete energy to global minimizers of the continuous energy. The existence of the latter is not assumed a-priory, but is rather a consequence of our analysis. Our dG approach simplifies some implementation details and theoretical constructions needed in [5, 8] for the Kirchhoff elements. The dG formulation is also valid for graded isoparametric elements of degree $k \geq 2$, which is considerably more general than [5, 8]. Moreover, we present numerical experiments that indicate that the dG approach also captures the physics of some problems better than the Kirchhoff approach, while also giving rise to a more accurate approximation of the isometry constraint.

9. Appendix: Estimates for Isoparametric Mappings

In this appendix we present inverse and error estimates involving the Hessian for elements obtained by isoparametric transformations; this includes quadrilaterals mapped by bilinear maps $\mathbb{Q}_1$. The issue at stake is that the isoparametric map $F_T : \widehat{T} \to T$ from the reference element $\widehat{T}$ to a generic element $T \in \mathcal{T}_h$ is no longer affine but rather F_T belongs to $[\mathbb{P}_k(\widehat{T})]^2$ or $[\mathbb{Q}_k(\widehat{T})]^2$, whence $D^m F_T \neq \mathbf{0}$ for $2 \leq m \leq k$. Since the considerations below are local to a single element T, we drop the sub-index T in F_T. Moreover, we denote by $\widehat{v} := v \circ F$ the pullback of a function v defined on T. Property $D^m F \neq \mathbf{0}$ does not affect the first derivatives of a function v but it does influence its higher derivatives. To quantify this effect we recall that we assume that $\{\mathcal{T}_h\}_{h>0}$ is shape-regular so that $\|DF\|_{L^\infty(\widehat{T})} \lesssim h_T$ and $\|DF^{-1}\|_{L^\infty(T)} \lesssim h_T^{-1}$, where the hidden constants depend on shape regularity of $\mathcal{T}_h$. Furthermore, because $\|D^2 F\|_{L^\infty(\widehat{T})} \lesssim \|DF\|_{L^\infty(\widehat{T})}$, Lemma 13.4 in Ern and

Guermond[22] (see also Ciarlet and Raviart[19]) guarantees that

$$(9.1) \qquad \|D^m F^{-1}\|_{L^\infty(T)} \lesssim \|DF^{-1}\|_{L^\infty(T)}^m \quad 2 \le m \le k+1.$$

Our estimates below rely on the following key property of isoparametric maps F:

$$(9.2) \qquad p \in \mathbb{P}_1(T) \quad \Rightarrow \quad \widehat{p} = p \circ F \in \mathbb{V}_h^k(\widehat{T}).$$

If $\widehat{I} : C^0(\widehat{T}) \to \mathbb{V}_h^k(\widehat{T})$ is the Lagrange interpolation operator over $\widehat{T}$ of degree $k \ge 1$ and $I_h : C^0(T) \to \mathbb{V}_h^k(T)$ is the corresponding operator induced by the map F, then (9.2) translates into

$$(9.3) \qquad I_h p(x) = \widehat{I}\widehat{p}(\widehat{x}) = \widehat{p}(\widehat{x}) = p(x) \quad \forall\, x := F_T(\widehat{x}) \in T, \widehat{x} \in \widehat{T}.$$

If $v \in H^2(T)$, $\widehat{v} \in H^2(\widehat{T})$, then the chain rule gives

$$(9.4) \qquad \partial_{ij}^2 v(x) = \sum_{m,n=1}^{2} \partial_{mn}^2 \widehat{v}(\widehat{x})\, \partial_i F_m^{-1}(x)\, \partial_j F_n^{-1}(x) + \sum_{m=1}^{2} \partial_m \widehat{v}(\widehat{x})\, \partial_{ij} F_m^{-1}(x),$$

$$(9.5) \qquad \partial_{ij}^2 \widehat{v}(\widehat{x}) = \sum_{m,n=1}^{2} \partial_{mn}^2 v(x)\, \partial_i F_m(\widehat{x})\, \partial_j F_n(\widehat{x}) + \sum_{m=1}^{2} \partial_m v(x)\, \partial_{ij} F_m(\widehat{x}).$$

This, together with (9.1), yields

$$(9.6) \qquad \|D^2 v\|_{L^2(T)} \lesssim h_T \|DF^{-1}\|_{L^\infty(T)}^2 \|\widehat{v}\|_{H^2(\widehat{T})} \lesssim h_T^{-1} \|\widehat{v}\|_{H^2(\widehat{T})},$$

$$(9.7) \qquad \|D^2 \widehat{v}\|_{L^2(\widehat{T})} \lesssim h_T \|D^2 v\|_{L^2(T)} + \|\nabla v\|_{L^2(T)}.$$

This gives relations between the Hessians of v and $\widehat{v}$ that involve lower order terms. The next two estimates connect higher order derivatives of isoparametric maps.

Lemma 9.1 (inverse estimate). *Let $T \in \mathcal{T}_h$ and $e \in \mathcal{E}_h$ be an edge of T. Then the following estimates are valid for all $v_h \in \mathbb{V}_h^k$*

$$(9.8) \qquad \|D^2 v_h\|_{L^2(e)} \lesssim h_e^{-1/2} \|D^2 v_h\|_{L^2(T)}, \quad \|D^3 v_h\|_{L^2(e)} \lesssim h_e^{-3/2} \|D^2 v_h\|_{L^2(T)}.$$

Proof. Let $p \in \mathbb{P}_1(T)$ be a linear polynomial to be chosen later. Combining (9.4) and (9.1) with an inverse estimate for $\widehat{v}_h - \widehat{p} \in \mathbb{V}_h^k(\widehat{T})$, according to (9.2), yields

$$\|D^2 v_h\|_{L^2(e)} = \|D^2(v_h - p)\|_{L^2(e)} \lesssim h_e^{-3/2} \|\widehat{v}_h - \widehat{p}\|_{H^2(\widehat{e})} \lesssim h_e^{-3/2} \|\widehat{v}_h - \widehat{p}\|_{H^2(\widehat{T})}.$$

We now map back to T using (9.7) for $D^2(\widehat{v}_h - \widehat{p})$ to get

$$\|\widehat{v}_h - \widehat{p}\|_{H^2(\widehat{T})} \lesssim h_T \|D^2 v_h\|_{L^2(T)} + \|\nabla(v_h - p)\|_{L^2(T)} + h_T^{-1}\|v_h - p\|_{L^2(T)}.$$

We finally choose p as the best linear approximation of v_h in T, whence the Bramble-Hilbert lemma yields $\|v_h - p\|_{L^2(T)} + h_T\|\nabla(v_h - p)\|_{L^2(T)} \lesssim h_T^2 \|D^2 v_h\|_{L^2(T)}$ and gives the first assertion. The remaining assertion follows along the same lines upon differenciating (9.4) once more and using (9.1) with $m = 2$ together with the inverse estimate : $\|\widehat{v}_h - \widehat{p}\|_{H^3(\widehat{e})} \lesssim \|\widehat{v}_h - \widehat{p}\|_{H^2(\widehat{T})}$. The proof is complete. $\square$

Lemma 9.2 (H^2-stability). *Let $v \in H^2(T)$ for $T \in \mathcal{T}_h$ and $I_h v \in \mathbb{V}_h^k(T)$ be the Lagrange interpolant of v of degree $k \ge 1$. The following bound is then valid*

$$(9.9) \qquad \|D^2 I_h v\|_{L^2(T)} \lesssim \|D^2 v\|_{L^2(T)}.$$

Proof. We first note that $I_h v$ is well defined because $H^2(T) \subset C^0(T)$ (recall that T is closed). In view of (9.3), we see that $D^2 I_h v = D^2 (I_h v - p) = D^2 I_h(v-p)$ for any $p \in \mathbb{P}_1(T)$. We utilize (9.6) to deduce

$$\|D^2 I_h v\|_{L^2(T)} \lesssim h_T^{-1} \|\widehat{I}(\widehat{v} - \widehat{p})\|_{H^2(\widehat{T})}.$$

We choose p to equal v at three vertices of T, and observe that $\widehat{w} = \widehat{v} - \widehat{p}$ vanishes at the corresponding three vertices of $\widehat{T}$ and the associated linear Lagrange interpolant $\widehat{I}_1 \widehat{w} = 0$ vanishes as well. Concatenating an inverse estimate with the interpolation error estimate $\|\widehat{w} - \widehat{I}_1 \widehat{w}\|_{L^\infty(\widehat{T})} \lesssim |\widehat{w}|_{H^2(\widehat{T})}$, valid because $H^2(T) \subset L^\infty(T)$, yields

$$\|\widehat{I}(\widehat{v} - \widehat{p})\|_{H^2(\widehat{T})} \lesssim \|\widehat{I}(\widehat{v} - \widehat{p})\|_{L^\infty(\widehat{T})} \leq \|\widehat{v} - \widehat{p}\|_{L^\infty(\widehat{T})} \lesssim \|D^2(\widehat{v} - \widehat{p})\|_{L^2(\widehat{T})},$$

We finally invoke (9.7) to infer that

$$\|D^2(\widehat{v} - \widehat{p})\|_{L^2(\widehat{T})} \lesssim h_T \|D^2(v - p)\|_{L^2(T)} + \|\nabla(v-p)\|_{L^2(T)} \lesssim h_T \|D^2 v\|_{L^2(T)}.$$

This leads to the asserted estimate (9.9). $\square$

We point out that the usual interpolation estimate

(9.10) $$\|v - I_h v\|_{L^2(T)} + h_T \|\nabla(v - I_h v)\|_{L^2(T)} \lesssim h_T^2 |v|_{H^2(T)},$$

with H^2-seminorm on the right-hand side is valid for isoparametric elements with polynomial degree $k = 1$. In fact, transforming to $\widehat{T}$ and back to T via (9.7) gives

$$\|v - I_h v\|_{L^2(T)} + h_T \|\nabla(v - I_h v)\|_{L^2(T)} \lesssim h_T^2 \|D^2 v\|_{L^2(T)} + h_T |\nabla v|_{L^2(T)}.$$

Applying this estimate to $v - p$, with $p \in \mathbb{P}_1(T)$, and recalling (9.3) leads to (9.10). However, this argument fails for $k \geq 2$. We now state, and prove for completeness, a key error estimates for quadrilaterals valid for $k \geq 1$ due to Ciarlet and Raviart (see Examples 7 and 8 in [19]). We also refer to (13.27) in Ern and Guermond [22].

Lemma 9.3 (error estimate for quadrilaterals). *Let $T \in \mathcal{T}_h$ be so that $T = F(\widehat{T})$ with $F \in \mathbb{Q}_1$ bilinear. If $v \in H^{k+1}(T)$ and $I_h v \in \mathbb{V}_h^k(T)$ is the Lagrange interpolant of v with $k \geq 1$, then for $0 \leq m \leq k+1$ there holds*

(9.11) $$|v - I_h v|_{H^m(T)} \lesssim h_T^{k+1-m} |v|_{H^{k+1}(T)}.$$

Proof. Expression (9.4) in conjunction with (9.1) reveals that mapping $D^m(v - I_h v)$ from T to $\widehat{T}$ involves computing all derivatives $D^j(\widehat{v} - \widehat{I}\widehat{v})$ according to

$$\|D^m(v - I_h v)\|_{L^2(T)} \lesssim h_T^{1-m} \sum_{j=1}^{m} \|D^j(\widehat{v} - \widehat{I}\widehat{v})\|_{L^2(\widehat{T})} \lesssim h_T^{1-m} \|[D^{k+1}\widehat{v}]\|_{L^2(\widehat{T})},$$

where $[D^{k+1}\widehat{v}] = (\partial_i^{k+1}\widehat{v})_{i=1}^2$ stands for all *pure* derivatives of order $k+1$. The latter inequality is a consequence of the Bramble-Hilbert estimate for $\mathbb{Q}_k$ elements; see Theorem 1 in [12]. We now resort to a variant of (9.5) involving $k+1$ derivatives but simplified by the fact that pure derivatives $\partial_i^j F = 0$ if $j \geq 2$ because F is bilinear:

$$\partial_i^{k+1}\widehat{v}(\widehat{x}) = \sum_{n_1,\cdots,n_{k+1}=1}^{2} \partial_{n_1,\cdots,n_{k+1}}^{k+1} v(x)\, \partial_i F_{n_1}(x) \cdots \partial_i F_{n_{k+1}}(x).$$

Since $\|DF\|_{L^\infty(\widehat{T})} \lesssim h_T$, this yields

(9.12) $$\|[D^{k+1}\widehat{v}]\|_{L^2(\widehat{T})} \lesssim h_T^k |v|_{H^{k+1}(T)}$$

and combined with the previous estimate gives the asserted estimate (9.11). $\qquad\square$

Lemma 9.3 extends to isoparametric maps $F \in \mathbb{Q}_k$ for $k \geq 1$. We quote here Theorem 6 of Ciarlet and Raviart [19]; see also (13.30) in Ern and Guermond [22].

Lemma 9.4 (error estimates for curved quadrilaterals)**.** *Let $T \in \mathcal{T}_h$ be so that $T = F(\widehat{T})$ with $F \in \mathbb{Q}_k$ and let $\widetilde{F} \in \mathbb{Q}_1$ be the bilinear function that maps the vertices of $\widehat{T}$ to those of T. Let F satisfy*

$$(9.13) \qquad \|[D^m F]\|_{L^\infty(\widehat{T})} \lesssim \|D\widetilde{F}\|_{L^\infty(\widehat{T})}^m \quad \forall\, 2 \leq m \leq k+1.$$

If $v \in H^{k+1}(T)$ and $I_h v \in \mathbb{V}_h^k(T)$ is the Lagrange interpolant of v with $k \geq 1$, then for $0 \leq m \leq k+1$ there holds

$$(9.14) \qquad |v - I_h v|_{H^m(T)} \lesssim h_T^{k+1-m} \|v\|_{H^{k+1}(T)}.$$

The proof of Lemma 9.4 is similar to that of Lemma 9.3 except that

$$(9.15) \qquad \|[D^{k+1}\widehat{v}]\|_{L^2(\widehat{T})} \lesssim h_T^k \|v\|_{H^{k+1}(T)}$$

replaces (9.12). This is because $\partial_i^j F \neq 0$ for $j \geq 2$ and (9.13) is used instead.

References

1. D. Arndt, W. Bangerth, D. Davydov, T. Heister, L. Heltai, M. Kronbichler, M. Maier, J.-P. Pelteret, B. Turcksin, D. Wells, *The `deal.II` Library, Version 8.5*, Journal of Numerical Mathematics, 25(3):137–146, 2017.
2. D. N. Arnold, F. Brezzi, B. Cockburn, L. D. Marini, *Unified analysis of discontinuous Galerkin methods for elliptic problems*, SIAM J. Numer. Anal., 39(5), 1749–1779, 2002.
3. E. Bänsch, P. Morin, R.H. Nochetto, *An adaptive Uzawa FEM for the Stokes problem: Convergence without the inf-sup condition*, SIAM J. Numer. Anal. 40, 1207–1229, 2002.
4. W. Bangerth, R. Hartmann, G. Kanschat, *deal.II – a general purpose object oriented finite element library*, ACM Trans. Math. Softw., 33(4):24/1–24/27, 2017.
5. S. Bartels, *Finite element approximation of large bending isometries*, Numer. Math. 124, 3, 415-440, 2013.
6. S. Bartels, *Numerical Methods for Nonlinear Partial Differential Equations*, Springer International Publishing, 2015.
7. S. Bartels, A. Bonito, Anastasia H. Muliana, R. H. Nochetto, *Modeling and simulation of thermally actuated bilayer plates*, J. Comp. Phys. 354, 512–528, 2018.
8. S. Bartels, A. Bonito, R. H. Nochetto, *Bilayer plates: Model reduction, Γ-convergent finite element approximation and discrete gradient flow*, Commun. Pure Appl. Math. 70, 3, 547--589, 2017.
9. N. Bassik, B. Abebe, K. Laflin, D. Gracias, *Photolithographically patterned smart hydrogel based bilayer actuators*, Polymer 51, 6093-6098, 2010.
10. K. Bhattacharya, M. Lewicka, M. Schaffner, *Plates with incompatible prestrain*, Arch. Rational Mech. Anal. 221, 143-–181, 2016.
11. A. Bonito, R. H. Nochetto, *Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method*, SIAM J. Numer. Anal., 48(2), 734-771, 2010.
12. J.H. Bramble, S.R. Hilbert, *Bounds for a class of linear functionals with applications to Hermite interpolation*, Numer. Math., 16, 362–369, 1970/71.
13. S. C. Brenner, *Poincaré-Friedrichs inequalities for piecewise H^1 functions*, SIAM J. Numer. Anal., 41(1), 306-324, 2003.
14. S. C. Brenner, *Two-level additive Schwarz preconditioners for nonconforming finite elements*, Math. Comp., 65(215), 897-921, 1996.
15. S. C. Brenner, K. Wang, J. Zhao, *Poincaré-Friedrichs inequalities for piecewise H^2 functions*, Numer. Funct. Anal. Optim., 25, 463-478, 2004.
16. F. Brezzi, G. Manzini, D. Marini, P. Pietra, A. Russo, *Discontinuous Galerkin approximations for elliptic problems*, Numer. Methods Partial Differential Equations, 16(4), 365-378, 2000.

17. A. Buffa, C. Ortner,*Compact embeddings of broken Sobolev spaces and applications*, IMA Numer. Anal., 29, 827-855, 2009.

18. P. Castillo, B. Cockburn, I. Perugia, D. Schötzau,*An a priori error analysis of the local discontinuous Galerkin method for elliptic problems*, SIAM J. Numer. Anal., 38, 1676–1706, 2000.

19. P.G. Ciarlet, P.-A. Raviart,*Interpolation theory over curved elements, with applications to finite element methods*, Comput. Methods Appl. Mech. Engrg., 1, 217–249, 1972.

20. D. A. Di Pietro, A. Ern, *Discrete functional fnalysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations*, Math. Comp., 79(271), 1303-1330, 2010.

21. A. Ern, J-L Guermond, *Finite element quasi-interpolation and best approximation*, ESAIM: M2AN 51, 1367-1385, 2017.

22. A. Ern, J-L Guermond, *Finite Elements I: Approximation and interpolation*, to appear.

23. G. Friesecke, R.D. James, S. Müller, *A theorem on geometric rigidity and the derivation of nonlinear plate theory from three-dimensional elasticity*, Comm. Pure Appl. Math. 55, 11, 1461-1506, 2002.

24. E. Jager, E. Smela, O. Inganäs, *Microfabricating conjugated polymer actuators*, Science 290, 1540-1545, 2000.

25. R.H.W. Hoppe, B. Wohlmuth, *Element-oriented and edge-oriented local error estimators for nonconforming finite element methods*, RAIRO Modél. Math. Anal. Numér., 30(2), 237-263, 1996.

26. P. Hornung, *Approximation of flat $W^{2,2}$ isometric immersions by smooth ones*, Arch. Ration. Mech. Anal. 199, 1015- 1067, 2011.

27. J.-N. Kuo, G.-B. Lee, W.-F. Pan, H.-L. Lee, *Shape and thermal effects of metal films on stress-induced bending of micromachined bilayer cantilever*, Japanese Journal of Applied Physics 44, 5R, 3180, 2005.

28. M. Lewicka, P. Ochoa, M.-R. Pakzad, *Variational models for prestrained plates with Monge-Ampère constraint*, Differential Integral Equations 28, no. 9/10, 861–898, 2015.

29. D. Ntogkas, *Non-linear geometric PDEs: algorithms, numerical analysis and computation*, PhD Thesis, University of Maryland, College Park, 2018.

30. P. Oswald, *On a BPX-preconditioner for* P1 *elements*, Computing, 51(2), 125-133, 1993.

31. T. Pryer, *Discontinuous Galerkin methods for the p−biharmonic equation from a discrete variational perspective*, Electronic Transactions of Numerical Analysis, 2014.

32. B. Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*, Society for Industrial and Applied Mathematics, 2008.

33. O. Schmidt, K. Eberl, *Thin solid films roll up into nanotubes*, Nature 410, 168, 2001.

34. E. Smela, O. Inganös, I. Lundström, *Controlled folding of micrometer-size structures*, Science 268, 5218, 1735–1738, 1995.

Department of Mathematics, Texas A&M University, College Station, TX 77843
Email address: `bonito@math.tamu.edu`

Department of Mathematics and Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742
Email address: `rhn@math.umd.edu`

Department of Mathematics, University of Maryland, College Park, Maryland 20742
Email address: `dimnt@math.umd.edu`