Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

A unified framework for high-order numerical discretizations of variational inequalities

Jad Dabaghi *, Guillaume Delay

Sorbonne Université, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France

ARTICLE INFO

Keywords:
Variational inequalities
Complementarity conditions
Semismooth Newton method
Finite elements
Discontinuous Galerkin
Hybrid high-order method

ABSTRACT

We present in this work a unified framework for elliptic variational inequalities that gathers several problems in contact mechanics like the unilateral contact of one or two membranes or the Signorini problem. We study a family of Galerkin numerical schemes that discretize this framework. We prove the well-posedness of the discrete problem and we show that it is equivalent to a saddle-point mixed formulation containing complementarity constraints. To solve the arising nonlinear problem, we employ a semismooth Newton method and we prove local convergence properties. The abstract framework is then applied to the discretization of the unilateral contact between two membranes. We propose to discretize this problem with a finite element (FEM), a discontinuous Galerkin (dG), and a hybrid high-order (HHO) methods. We also adapt the semismooth Newton algorithm, including a static condensation procedure for the HHO method. Finally, we run numerical experiments for the FEM and HHO discretizations and compare their behavior.

1. Introduction

In the present study, we are interested in solving numerically a wide class of variational inequalities. The abstract framework that we use can be applied to several problems for instance in contact mechanics, see Section 1.2. Our goal is to develop a class of numerical schemes that can approximate this abstract framework. Numerical simulations are then considered for one of the examples given and two numerical schemes are compared.

1.1. Variational inequalities

Let $\widetilde{V}$ and $\widetilde{\Lambda}$ be two Hilbert spaces equipped respectively with the scalar products $(\cdot,\cdot)_{\widetilde{V}}$ and $(\cdot,\cdot)_{\widetilde{\Lambda}}$. Let V be a Hilbert subspace of $\widetilde{V}$ and $V^g := \{g\} + V$ be an affine subspace of $\widetilde{V}$ for some element $g \in \widetilde{V}$. As written below, g aims at representing Dirichlet boundary data. For any Hilbert space X, its corresponding dual space is denoted by X' and the duality pairing is denoted by $\langle \cdot,\cdot \rangle_{X',X}$. Let us consider a cone $\Lambda \subset \widetilde{\Lambda}$. This means that there exists $\check{\Lambda} \subset (\widetilde{\Lambda})'$ such that $\Lambda = \left\{ \chi \in \widetilde{\Lambda} \text{ s.t. } \langle \mu, \chi \rangle_{(\widetilde{\Lambda})',\widetilde{\Lambda}} \geqslant 0 \quad \forall \mu \in \check{\Lambda} \right\}$. The set Λ is clearly a nonempty closed convex set of $\widetilde{\Lambda}$. Let $\Phi : \widetilde{V} \to \widetilde{\Lambda}$ be a continuous linear and surjective mapping. Let $\mathcal{K}^g$ be the closed convex set of V^g defined by

$$\mathcal{K}^g := \{v \in V^g \text{ s.t. } \Phi(v) - \Psi \in \Lambda\},$$

where Ψ is a given element of $\widetilde{\Lambda}$. We further assume that $\mathcal{K}^g$ is nonempty. In practice, this assumption is fulfilled under some compatibility conditions on g and Ψ (see for instance problems (2) and (3) below).

Let $a : \widetilde{V} \times \widetilde{V} \to \mathbb{R}$ be a continuous bilinear form that is coercive on $V \times V$ and $\ell : \widetilde{V} \to \mathbb{R}$ be a continuous linear form. We are interested in the following variational inequality: Find $u \in \mathcal{K}^g$ such that

$$a(u, v-u) \geqslant \ell(v-u), \quad \forall v \in \mathcal{K}^g. \tag{1}$$

This problem is well posed as a result of the Lions–Stampacchia theorem, see [1, Theorem 2.1] or [2, Theorem 5.6]. Furthermore, it belongs to the wide range of variational inequalities of the first kind [3–5] and can be used to model several contact problems as we see in the next section.

1.2. Motivations

Let $\Omega \subset \mathbb{R}^2$ be a smooth connected domain. We denote by $(\cdot,\cdot)_\Omega$ the L^2-scalar product on Ω. Let us now give several applications of the abstract framework given above.

Obstacle problem: the unknown u represents the displacement of an elastic membrane that cannot penetrate a lower obstacle. We look for the solution of the following problem:

Find $u \in \mathcal{K}^g := \{v \in H^1(\Omega) \text{ s.t. } v = g \text{ on } \partial\Omega, \text{ and } \Phi(v) := v \geqslant \Psi \text{ a.e. in } \Omega\}$ such that

$$(\nabla u, \nabla(v-u))_\Omega \geqslant (f, v-u)_\Omega, \quad \forall v \in \mathcal{K}^g, \tag{2}$$

* Corresponding author.
E-mail addresses: jad.dabaghi@sorbonne-universite.fr (J. Dabaghi), guillaume.delay@sorbonne-universite.fr (G. Delay).

https://doi.org/10.1016/j.camwa.2021.03.011
Received 16 October 2020; Accepted 9 March 2021
Available online xxxx

where $\Psi \in H^1(\Omega)$ represents the position of the lower obstacle and $g \in H^{1/2}(\partial\Omega)$ is a Dirichlet boundary datum for u. They fulfill the compatibility condition $\Psi \leqslant g$ on $\partial\Omega$ in order to ensure that $\mathcal{K}^g$ is nonempty. Moreover, $f \in L^2(\Omega)$ represents a force acting on the membrane. Note that in this example $\Phi : \widetilde{V} := H^1(\Omega) \to \widetilde{\Lambda} := H^1(\Omega)$ is the identity function and $\Lambda := \left\{v \in H^1(\Omega) \text{ s.t. } v \geqslant 0 \text{ a.e. in } \Omega\right\}$. More details about this model can be found e.g. in [4].

Scalar Signorini problem: the boundary of the domain is partitioned as $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$. Dirichlet boundary conditions, Neumann boundary conditions and unilateral contact boundary conditions are respectively imposed on Γ_1, Γ_2 and Γ_3 to the unknown u that represents the position of an elastic membrane. The scalar Signorini problem reads:

Find $u \in \mathcal{K}^g := \{v \in H^1(\Omega) \text{ s.t. } v = g \text{ on } \Gamma_1, \text{ and } \Phi(v) := v \geqslant \Psi \text{ a.e. on } \Gamma_3\}$ such that

$$(\nabla u, \nabla(v-u))_\Omega \geqslant (f, v-u)_\Omega, \quad \forall v \in \mathcal{K}^g, \tag{3}$$

where $g \in H^{1/2}(\Gamma_1)$, $\Psi \in H^{1/2}(\Gamma_3)$ and $f \in L^2(\Omega)$. Here, $\Phi : \widetilde{V} := H^1(\Omega) \to \widetilde{\Lambda} := H^{1/2}(\Gamma_3)$ is the trace application and $\Lambda := \left\{v \in H^{1/2}(\Gamma_3) \text{ s.t. } v \geqslant 0 \text{ a.e. on } \Gamma_3\right\}$. We refer e.g. to [6] for a study of such a problem.

Contact between two membranes: two membranes are located one above the other and cannot penetrate each other. The unknown $\boldsymbol{u} := \left(u_1, u_2\right)$ is a vector that represents at every point of Ω the position of the two membranes. The problem reads:

Find $\boldsymbol{u} \in \mathcal{K}^g := \left\{\boldsymbol{v} := (v_1, v_2) \in H^1_{g_1}(\Omega) \times H^1_{g_2}(\Omega) \text{ s.t. } \Phi(\boldsymbol{v}) := v_1 - v_2 \geqslant 0 \text{ a.e. in } \Omega\right\}$ such that

$$\sum_{\alpha=1}^{2} \mu_\alpha (\nabla u_\alpha, \nabla(v_\alpha - u_\alpha))_\Omega \geqslant \sum_{\alpha=1}^{2} (f_\alpha, v_\alpha - u_\alpha)_\Omega, \quad \forall \boldsymbol{v} \in \mathcal{K}^g, \tag{4}$$

where $(f_1, f_2) \in (L^2(\Omega))^2$ represents the surface forces acting on the two membranes and μ_1, μ_2 are positive coefficients representing the tensions of the membranes. Moreover, $\left(g_1, g_2\right) \in (H^{1/2}(\partial\Omega))^2$ denotes the boundary datum for $\boldsymbol{u}$ fulfilling $g_1 \geqslant g_2$. We used the compact notation $H^1_{g_\alpha}(\Omega) := \left\{v \in H^1(\Omega) \text{ s.t. } v = g_\alpha \text{ on } \partial\Omega\right\}$ for $\alpha \in \{1, 2\}$. Here $\Phi : \widetilde{V} := (H^1(\Omega))^2 \to \widetilde{\Lambda} := H^1(\Omega)$ corresponds at each point of Ω to the signed distance function between the two membranes and $\Lambda := \left\{v \in H^1(\Omega) \text{ s.t. } v \geqslant 0 \text{ a.e. in } \Omega\right\}$. For further information about this problem, the reader can report for instance to [7,8].

Vector Signorini problem: the boundary of the domain is partitioned as $\partial\Omega = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$. A structure is clamped on Γ_1 (Dirichlet boundary conditions), it has a force acting on it on Γ_2 (Neumann boundary conditions) and it fulfills a unilateral contact on Γ_3. The goal is to find the displacement $\boldsymbol{u}$ of the structure knowing the displacement of the boundary on Γ_1 and the forces acting on Γ_2. It reads:

Find $\boldsymbol{u} \in \mathcal{K}^g := \{\boldsymbol{v} \in (H^1(\Omega))^2 \text{ s.t. } \boldsymbol{v} = \boldsymbol{g} \text{ on } \Gamma_1, \text{ and } \Phi(\boldsymbol{v}) := \boldsymbol{v} \cdot \mathbf{n} \leqslant 0 \text{ a.e. on } \Gamma_3\}$ such that

$$(\sigma(\boldsymbol{u}), \epsilon(\boldsymbol{v} - \boldsymbol{u}))_\Omega \geqslant (\boldsymbol{f}, \boldsymbol{v} - \boldsymbol{u})_\Omega + (\boldsymbol{g}_N, \boldsymbol{v} - \boldsymbol{u})_{\Gamma_2}, \quad \forall \boldsymbol{v} \in \mathcal{K}^g, \tag{5}$$

where $\epsilon(\boldsymbol{v}) := \frac{1}{2}\left(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^T\right)$ and $\sigma(\boldsymbol{v}) := \mathbb{K}\epsilon(\boldsymbol{v})$ are respectively the strain and stress tensors associated to the displacement $\boldsymbol{v}$ with $\mathbb{K}$ the fourth-order symmetric elasticity tensor. Moreover, $\boldsymbol{g} \in (H^{1/2}(\Gamma_1))^2$ and $\boldsymbol{g}_N \in (L^2(\Gamma_2))^2$ are respectively Dirichlet and Neumann boundary data and $\boldsymbol{f} \in (L^2(\Omega))^2$ is a volumic force acting on the structure. In this case, the mapping $\Phi : \widetilde{V} := (H^1(\Omega))^2 \to \widetilde{\Lambda} := H^{1/2}(\Gamma_3)$ is the normal trace on Γ_3 and $\Lambda := \left\{v \in H^{1/2}(\Gamma_3) \text{ s.t. } v \leqslant 0 \text{ a.e. in } \Gamma_3\right\}$. Such a problem has been studied for instance in [9].

Remark 1.1 (*Boundary Data*)**.** Note that in the previous problems, g (or $\boldsymbol{g}$) denotes the Dirichlet boundary data while in our abstract setting g is an element of the solution space (in fact we consider a lifting of those boundary data).

1.3. Scientific context

The numerical approximation of variational inequalities has been studied mainly in the context of the obstacle problem (2). Piecewise affine finite elements have been considered in several works, see [10–12] for a priori analysis and [13–17] for a posteriori analysis. Discontinuous Galerkin (dG) methods [18–20] and hybrid high-order (HHO) methods [21] have also been considered to solve the obstacle problem.

The other problems given in Section 1.2 have also been studied. The scalar Signorini problem has been discretized using e.g. finite elements [6], finite volumes [22] and the HHO method [23]. The contact between two membranes has been studied in [7,8,24,25] using finite elements. The vector Signorini problem (5) has been studied with the finite element method (FEM) [26,9,27], with the dG method [28], and with the HHO method [29].

Among these works, piecewise quadratic elements are considered in [11,12,30,6,20,27,29,21] and arbitrary high order elements are studied in [26,17,25,23]. The arbitrary high-order method we develop in the present work is inspired by [25].

Several strategies could be employed to solve the discretized nonlinear problem arising from any of these methods. We mention the interior point method [31], the active set strategy [32] and the primal–dual active set strategy [33,34]. In the present work we consider a semismooth Newton method [35–39].

1.4. Mixed formulation

Let us define the continuous bilinear form $b : \widetilde{V} \times (\widetilde{\Lambda})' \to \mathbb{R}$ by

$$b(v, \chi) := \langle \chi, \Phi(v) \rangle_{(\widetilde{\Lambda})', \widetilde{\Lambda}}, \quad \forall v \in \widetilde{V}, \quad \forall \chi \in (\widetilde{\Lambda})'. \tag{6}$$

We also define the adjoint cone to Λ by

$$\widehat{\Lambda} := \{\chi \in (\widetilde{\Lambda})' \text{ s.t. } \langle \chi, \mu \rangle_{(\widetilde{\Lambda})', \widetilde{\Lambda}} \geqslant 0 \quad \forall \mu \in \Lambda\}. \tag{7}$$

We can study the mixed problem: Find $(u, \lambda) \in V^g \times \widehat{\Lambda}$ such that

$$a(u, v) - b(v, \lambda) = \ell(v), \qquad \forall v \in V, \tag{8a}$$

$$b(u, \chi - \lambda) \geqslant \langle \chi - \lambda, \Psi \rangle_{(\widetilde{\Lambda})', \widetilde{\Lambda}}, \quad \forall \chi \in \widehat{\Lambda}. \tag{8b}$$

One can view $\lambda \in \widehat{\Lambda}$ as a Lagrange multiplier for the constraint $\Phi(u) - \Psi \in \Lambda$. It is standard to show that (8) rewrites as the following system of variational equalities with complementarity constraints: Find $(u, \lambda) \in V^g \times \widehat{\Lambda}$ such that

$$a(u, v) - b(v, \lambda) = \ell(v), \quad \forall v \in V, \tag{9a}$$

$$\Phi(u) - \Psi \in \Lambda, \quad \lambda \in \widehat{\Lambda}, \quad \langle \lambda, \Phi(u) - \Psi \rangle_{(\widetilde{\Lambda})', \widetilde{\Lambda}} = 0. \tag{9b}$$

Any solution to (8) is also solution to (1). Note however that the converse is not necessarily true in the general case but can be proven if we consider more regular data (see for instance [7,8] for the contact between two membranes).

1.5. Outline of the article

This contribution is organized as follows. In Section 2, we propose a discretization of the mixed problem (8) and we prove the well-posedness of the resulting nonlinear discretized problem. This problem is then rewritten under an algebraic formulation containing complementarity constraints. In Section 3, we introduce a semismooth Newton method to compute the numerical solution of the underlying nonlinear algebraic formulation. We also discuss convergence properties of that semismooth Newton method. We give in Section 4 several applications of our abstract framework: we discretize the elliptic contact problem between two membranes (4) with several numerical schemes. The results of Section 2 thus guarantee the well-posedness of these methods. In Section 5, we perform numerical simulations for this problem using two different schemes (FEM and HHO). Their behavior are then compared.

2. The nonlinear discretized problem

In this section, we mimic the framework of Section 1.1 to give a discretized variational inequality. We then propose a mixed formulation associated to this variational inequality. Furthermore, we prove the equivalence of these two problems. We then write the associated algebraic formulation.

2.1. Discrete variational inequality

In this section, for every space or operator X introduced in Section 1.1, we denote by X_h its discrete analogue. Let $\widetilde{V}_h$ and $\widetilde{\Lambda}_h$ be two finite dimensional Hilbert spaces respectively equipped with the scalar products $(\cdot,\cdot)_{\widetilde{V}_h}$ and $(\cdot,\cdot)_{\widetilde{\Lambda}_h}$. Let V_h and $\widehat{V}_h$ be linear subspaces of $\widetilde{V}_h$ such that $\widetilde{V}_h = V_h \oplus \widehat{V}_h$ and $V_h^g := \{g_h\} + V_h$ be an affine subspace of $\widetilde{V}_h$ where $g_h \in \widehat{V}_h$. Here the notation $\oplus$ stands for the direct sum between the spaces V_h and $\widehat{V}_h$. The linear mapping $\Phi_h : \widetilde{V}_h \to \widetilde{\Lambda}_h$ is assumed to be surjective. Let $\Lambda_h \subset \widetilde{\Lambda}_h$ be a cone, i.e. there exists $\check{\Lambda}_h \subset \widetilde{\Lambda}_h$ such that $\Lambda_h = \{\chi_h \in \widetilde{\Lambda}_h \text{ s.t. } (\chi_h, \mu_h)_{\widetilde{\Lambda}_h} \geqslant 0 \quad \forall \mu_h \in \check{\Lambda}_h\}$. Moreover we assume that

$$\chi_h \in \Phi_h(\widehat{V}_h) \iff (\chi_h \in \Lambda_h \text{ and } -\chi_h \in \Lambda_h), \tag{10}$$

$$\mathrm{Span}(\Lambda_h) = \widetilde{\Lambda}_h. \tag{11}$$

Remark 2.1. In the sequel, $\widehat{V}_h$ is the linear space generated by all the Lagrange basis functions associated to nodes where Dirichlet boundary conditions are imposed. The assumption (10) then means that the unilateral constraint is relaxed on the Dirichlet boundary nodes and that it is relaxed only on those nodes. The solution will still satisfy the constraints since the Dirichlet conditions are strongly imposed on the elements of $\mathcal{K}_h^g$ (see below). The assumption (11) means that the constraint is always unilateral, i.e. we never impose the elements of Λ_h to be 0-valued anywhere.

We define the discrete analogue to $\mathcal{K}^g$ by

$$\mathcal{K}_h^g := \{v_h \in V_h^g \text{ s.t. } \Phi_h(v_h) - \Psi_h \in \Lambda_h\}. \tag{12}$$

It is obviously a nonempty closed convex set of V_h^g where Ψ_h is a given element of $\widetilde{\Lambda}_h$. The bilinear form $a_h : \widetilde{V}_h \times \widetilde{V}_h \to \mathbb{R}$ is assumed to be continuous on $\widetilde{V}_h \times \widetilde{V}_h$ and coercive on $V_h \times V_h$. The linear form $\ell_h : \widetilde{V}_h \to \mathbb{R}$ is continuous.

We consider the following approximation to problem (1): Find $u_h \in \mathcal{K}_h^g$ such that

$$a_h(u_h, v_h - u_h) \geqslant \ell_h(v_h - u_h), \quad \forall v_h \in \mathcal{K}_h^g. \tag{13}$$

According to the Lions–Stampacchia theorem, the variational inequality (13) admits a unique solution.

2.2. Discrete mixed problem

Let us write a discrete mixed formulation associated to (13). Let $b_h : \widetilde{V}_h \times \widetilde{\Lambda}_h \to \mathbb{R}$ be the bilinear form defined by

$$b_h(v_h, \chi_h) := (\Phi_h(v_h), \chi_h)_{\widetilde{\Lambda}_h}, \quad \forall v_h \in \widetilde{V}_h, \quad \forall \chi_h \in \widetilde{\Lambda}_h. \tag{14}$$

We also define the adjoint cone of Λ_h:

$$\widehat{\Lambda}_h := \{\chi_h \in \widetilde{\Lambda}_h \text{ s.t. } (\chi_h, \mu_h)_{\widetilde{\Lambda}_h} \geqslant 0 \quad \forall \mu_h \in \Lambda_h\}. \tag{15}$$

This adjoint cone has been used for instance in the numerical discretization of the unilateral contact between two membranes with Lagrange finite elements [24,25] and with discontinuous Galerkin methods [18] for the obstacle problem. In the sequel, for any set $X_h \subset \widetilde{\Lambda}_h$, $X_h^\perp$ denotes its orthogonal space in $\widetilde{\Lambda}_h$ with respect to the scalar product $(\cdot,\cdot)_{\widetilde{\Lambda}_h}$.

Lemma 1. *We have the following characterizations for the cones Λ_h and $\widehat{\Lambda}_h$:*

$$\Lambda_h = \left\{ v_h \in \widetilde{\Lambda}_h \text{ s.t. } (v_h, \mu_h)_{\widetilde{\Lambda}_h} \geqslant 0 \quad \forall \mu_h \in \widehat{\Lambda}_h \right\}, \tag{16}$$

$$\widehat{\Lambda}_h = \left\{ \chi_h \in (\Phi_h(\widehat{V}_h))^\perp \text{ s.t. } (\mu_h, \chi_h)_{\widetilde{\Lambda}_h} \geqslant 0 \quad \forall \mu_h \in \Lambda_h \cap \Phi_h(V_h) \right\}. \tag{17}$$

Proof. For the proof of these characterizations, the reader can report to Section 2.3. They are a consequence of (25)–(26). □

We propose the following discretization to the mixed problem (8): Find $(u_h, \lambda_h) \in V_h^g \times \widehat{\Lambda}_h$ such that

$$a_h(u_h, v_h) - b_h(v_h, \lambda_h) = \ell_h(v_h), \quad \forall v_h \in V_h, \tag{18a}$$

$$b_h(u_h, \chi_h - \lambda_h) \geqslant (\Psi_h, \chi_h - \lambda_h)_{\widetilde{\Lambda}_h}, \quad \forall \chi_h \in \widehat{\Lambda}_h. \tag{18b}$$

We now show that the discrete variational inequality (13) is equivalent to the discrete mixed problem (18).

Theorem 2. *If $u_h \in \mathcal{K}_h^g$ is the solution to (13) then there exists a unique $\lambda_h \in \widehat{\Lambda}_h$ such that (u_h, λ_h) is a solution to (18); conversely if (u_h, λ_h) is a solution to (18) then $u_h \in \mathcal{K}_h^g$ and u_h is the unique solution to (13).*

Corollary 3. *Problem (18) is well-posed.*

Proof of Theorem 2. Let $(u_h, \lambda_h) \in V_h^g \times \widehat{\Lambda}_h$ be a solution to (18). According to the definition of $\widehat{\Lambda}_h$, for all $\chi_h \in \widehat{\Lambda}_h$ we have $\lambda_h + \chi_h \in \widehat{\Lambda}_h$. We then test (18b) with $\lambda_h + \chi_h$ and get $\forall \chi_h \in \widehat{\Lambda}_h$, $b_h(u_h, \chi_h) \geqslant (\Psi_h, \chi_h)_{\widetilde{\Lambda}_h}$, which implies $u_h \in \mathcal{K}_h^g$ by virtue of (16). Furthermore, testing equation (18b) with $\chi_h = 0 \in \widehat{\Lambda}_h$ gives $-b_h(u_h, \lambda_h) \geqslant -(\Psi_h, \lambda_h)_{\widetilde{\Lambda}_h}$ and since $\lambda_h \in \widehat{\Lambda}_h$, for all $v_h \in \mathcal{K}_h^g$, we have as a result of (16), $b_h(v_h, \lambda_h) \geqslant (\Psi_h, \lambda_h)_{\widetilde{\Lambda}_h}$. Then, $\forall v_h \in \mathcal{K}_h^g$, $b_h(v_h - u_h, \lambda_h) \geqslant 0$. Testing (18a) with $v_h - u_h \in V_h$ we get

$$a_h(u_h, v_h - u_h) - \ell_h(v_h - u_h) = b_h(v_h - u_h, \lambda_h) \geqslant 0, \quad \forall v_h \in \mathcal{K}_h^g,$$

which shows that u_h is the solution to (13).

Conversely, if $u_h \in \mathcal{K}_h^g$ is the solution to (13), we uniquely define $A_h \in V_h$ such that

$$(A_h, z_h)_{\widetilde{V}_h} = -\ell_h(z_h) + a_h(u_h, z_h), \quad \forall z_h \in V_h. \tag{19}$$

Then, $\forall v_h \in \mathcal{K}_h^g$, $u_h - v_h \in V_h$ and we have

$$(A_h, u_h - v_h)_{\widetilde{V}_h} = -\ell_h(u_h - v_h) + a_h(u_h, u_h - v_h) \leqslant 0, \quad \forall v_h \in \mathcal{K}_h^g. \tag{20}$$

Let us define the mapping $\widetilde{\Phi}_h : V_h \to \Phi_h(V_h) \subset \widetilde{\Lambda}_h$ such that for all $v_h \in V_h$, $\widetilde{\Phi}_h(v_h) := \Phi_h(v_h)$. We also define $\widetilde{\Phi}_h^* : (\Phi_h(\widehat{V}_h))^\perp \to \widehat{V}_h^\perp$ such that for all $v_h \in V_h$ and all $w_h \in (\Phi_h(\widehat{V}_h))^\perp$, $(v_h, \widetilde{\Phi}_h^*(w_h))_{\widetilde{\Lambda}_h} := (\widetilde{\Phi}_h(v_h), w_h)_{\widetilde{\Lambda}_h}$. We can prove that $\mathrm{Ker}(\widetilde{\Phi}_h)^\perp = V_h^\perp + \mathrm{Im}(\widetilde{\Phi}_h^*)$.

Now, for every $\chi_h \in \mathrm{Ker}(\widetilde{\Phi}_h) \subset V_h$, we can test (20) with $v_h = u_h \pm \chi_h \in \mathcal{K}_h^g$ and we get for every $\chi_h \in \mathrm{Ker}(\widetilde{\Phi}_h)$, $(A_h, \chi_h)_{\widetilde{V}_h} = 0$. Hence, $A_h \in \mathrm{Ker}(\widetilde{\Phi}_h)^\perp = \mathrm{Im}(\widetilde{\Phi}_h^*) + V_h^\perp$. Then, there exist $\lambda_h \in \Phi_h(\widehat{V}_h)^\perp \subset \widetilde{\Lambda}_h$ and $\hat{v}_h \in V_h^\perp$ such that $A_h = \widetilde{\Phi}_h^*(\lambda_h) + \hat{v}_h$. Considering Eq. (19) we then have

$$a_h(u_h, z_h) - \ell_h(z_h) = (\widetilde{\Phi}_h^*(\lambda_h), z_h)_{\widetilde{V}_h} = b_h(z_h, \lambda_h), \quad \forall z_h \in V_h.$$

Then (u_h, λ_h) satisfies (18a) and according to (13), we have

$$b_h(v_h - u_h, \lambda_h) \geqslant 0, \quad \forall v_h \in \mathcal{K}_h^g. \tag{21}$$

It remains to show that $\lambda_h \in \widehat{\Lambda}_h$ and that (18b) is valid. For all $\mu_h \in \Lambda_h \cap \Phi_h(V_h)$, $\exists v_\mu \in V_h$ such that $\Phi_h(v_\mu) := \mu_h$. Next, $\Phi_h(u_h + v_\mu) - \Psi_h = (\Phi_h(u_h) - \Psi_h) + \mu_h \in \Lambda_h$ and then $u_h + v_\mu \in \mathcal{K}_h^g$. Besides, testing (21) with $v_h = u_h + v_\mu$ we get $\forall \mu_h \in \Lambda_h \cap \Phi_h(V_h)$, $(\mu_h, \lambda_h)_{\widetilde{\Lambda}_h} \geqslant 0$ and thus with (17) it yields $\lambda_h \in \widehat{\Lambda}_h$.

In a similar way, there exist $v_\Psi \in V_h$ and $\widehat{\Psi}_h \in \Phi_h(\widehat{V}_h)$ such that $\Phi_h(v_\Psi) + \widehat{\Psi}_h := \Psi_h$. And since $\Phi_h(v_\Psi + g_h) - \Psi_h = -\widehat{\Psi}_h + \Phi_h(g_h) \in$

$\Phi_h(\widehat{V}_h) \subset \Lambda_h$ by assumption (10), we then have $v_\Psi + g_h \in \mathcal{K}_h^g$. We test (21) with $v_h = v_\Psi + g_h$ and get $b_h(g_h - u_h, \lambda_h) \geqslant -(\Psi_h - \widehat{\Psi}_h, \lambda_h)_{\widetilde{\Lambda}_h}$. Since $\lambda_h \in (\Phi_h(\widehat{V}_h))^\perp$, we have

$$b_h(-u_h, \lambda_h) \geqslant -(\Psi_h, \lambda_h)_{\widetilde{\Lambda}_h}. \tag{22}$$

Then, since $u_h \in \mathcal{K}_h^g$, the definition (15) yields

$$b_h(u_h, \chi_h) \geqslant (\Psi_h, \chi_h)_{\widetilde{\Lambda}_h}, \quad \forall \chi_h \in \widehat{\Lambda}_h. \tag{23}$$

Combining (22) and (23) we get (18b). Hence, (u_h, λ_h) is solution to (18).

We have proved the equivalence between the two problems. Let us now prove that λ_h is unique. Assume that (u_h, λ_h) and (u_h, μ_h) are two solutions to (18). Then (18a) gives $0 = b_h(v_h, \lambda_h - \mu_h) = (v_h, \Phi_h^*(\lambda_h - \mu_h))_{\widetilde{V}_h}$ for all $v_h \in V_h$, where Φ_h^* is the adjoint of Φ_h. Moreover, according to (17), $b_h(v_h, \lambda_h - \mu_h) = 0$ for all $v_h \in \widehat{V}_h$. Then, $\Phi_h^*(\lambda_h - \mu_h) = 0$.

Now, as $\mathrm{Im}(\Phi_h) = \widetilde{\Lambda}_h$, $\mathrm{Ker}(\Phi_h^*) = \mathrm{Im}(\Phi_h)^\perp = \{0\}$, Φ_h^* is invertible and $\lambda_h = \mu_h$. □

Remark 2.2 (*Consistency*)**.** In the present work we do not study the consistency of the scheme. This study is difficult to make in our general framework. Moreover, as we will see in Section 4, the discrete cones are nonconforming to the continuous ones (even for finite elements). This makes the consistency analysis even harder. The consistency will then be studied numerically only through the convergence of the schemes (see Section 5).

We observe that (18b) is equivalent to $\Phi_h(u_h) - \Psi_h \in \Lambda_h$ and $\big(\Phi_h(u_h) - \Psi_h, \lambda_h\big)_{\widetilde{\Lambda}_h} = 0$. This can be proven by evaluating (18b) with $\chi_h = \lambda_h + \mu_h$ $(\forall \mu_h \in \widehat{\Lambda}_h)$, $\chi_h = 2\lambda_h$ and $\chi_h = 0$ and using Lemma 1.

Therefore, system (18) rewrites as the following system of discrete equations with complementarity constraints: Find $(u_h, \lambda_h) \in V_h^g \times \widetilde{\Lambda}_h$ such that

$$a_h(u_h, v_h) - b_h(v_h, \lambda_h) = \ell_h(v_h), \quad \forall v_h \in V_h, \tag{24a}$$

$$\Phi_h(u_h) - \Psi_h \in \Lambda_h, \quad \lambda_h \in \widehat{\Lambda}_h, \quad \big(\Phi_h(u_h) - \Psi_h, \lambda_h\big)_{\widetilde{\Lambda}_h} = 0. \tag{24b}$$

2.3. Algebraic formulation

In this section, we rewrite system (24) under an algebraic formulation. For this purpose, we construct a basis $\left(\Gamma_l\right)_{1\leqslant l\leqslant m^\flat}$ of $\widetilde{\Lambda}_h$ such that Λ_h can be written as

$$\Lambda_h = \left\{ \sum_{1\leqslant i\leqslant m^\star} a_i \Gamma_i + \sum_{m^\star < i\leqslant m^\flat} b_i \Gamma_i \text{ s.t. } \forall i \in [1, m^\star], a_i \geqslant 0 \text{ and } \forall i \in [m^\star + 1, m^\flat], b_i \in \mathbb{R} \right\}, \tag{25}$$

where $m^\star := \dim(\Phi_h(V_h))$ and $m^\flat := \dim(\widetilde{\Lambda}_h)$.

Let us now comment on why such a basis exists. Let $(\Gamma_l)_{m^\star < l \leqslant m^\flat}$ be a basis of $\Phi_h(\widehat{V}_h)$. According to assumption (10), all these vectors belong to Λ_h. Moreover, as a consequence of (11), there exists a family $(\gamma_l)_{1\leqslant l\leqslant m^\star}$ of vectors of Λ_h such that $((\gamma_l)_{1\leqslant l\leqslant m^\star}, (\Gamma_l)_{m^\star < l\leqslant m^\flat})$ is a basis of $\widetilde{\Lambda}_h$. For all $1 \leqslant l \leqslant m^\star$, $\gamma_l = \tilde{\gamma}_l + \hat{\gamma}_l$ where $\hat{\gamma}_l \in \Phi_h(\widehat{V}_h)$ and $\tilde{\gamma}_l \in \Phi_h(V_h)$ and we set $\Gamma_l := \tilde{\gamma}_l$. The family $(\Gamma_l)_{1\leqslant l\leqslant m^\star}$ is then a basis of $\Phi_h(V_h)$ and $(\Gamma_l)_{1\leqslant l\leqslant m^\flat}$ is a basis of $\widetilde{\Lambda}_h$. Moreover, the identity (25) is a consequence of assumptions (10)–(11) and the way we constructed the family.

Let $\left(\xi_l\right)_{1\leqslant l\leqslant m^\flat}$ be the dual basis of $\left(\Gamma_l\right)_{1\leqslant l\leqslant m^\flat}$, i.e. we have $(\xi_i, \Gamma_j)_{\widetilde{\Lambda}_h} := \delta_{ij}$ where δ_{ij} is the Kronecker delta. We can then prove that

$$\widehat{\Lambda}_h = \left\{ \sum_{1\leqslant i\leqslant m^\star} a_i \xi_i \text{ s.t. } \forall i \in [1, m^\star], a_i \geqslant 0 \right\}. \tag{26}$$

Now let $(\phi_l)_{1\leqslant l\leqslant m^\#}$ be a basis of V_h. We decompose $u_h = u_h^0 + g_h$ (with $g_h \in \widehat{V}_h$) and we denote by $\boldsymbol{X}_{1h} \in \mathbb{R}^{m^\#}$ the vector of the components of u_h^0 in the basis $\left(\phi_l\right)_{1\leqslant l\leqslant m^\#}$. We also denote by $\boldsymbol{X}_{3h} \in \mathbb{R}^{m^\star}$ the vector of the components of λ_h in the family $\left(\xi_l\right)_{1\leqslant l\leqslant m^\star}$ (see (26)). We use the compact notation $\boldsymbol{X}_h := \left[\boldsymbol{X}_{1h}, \boldsymbol{X}_{3h}\right] \in \mathbb{R}^{m^\# + m^\star}$. The vector $\Psi_h \in \widetilde{\Lambda}_h$ can be decomposed as $\Psi_h = \widetilde{\Psi}_h + \widehat{\Psi}_h$ with $\widetilde{\Psi}_h \in \Phi_h(V_h)$ and $\widehat{\Psi}_h \in \Phi_h(\widehat{V}_h)$. We denote by $\boldsymbol{\Psi}_h \in \mathbb{R}^{m^\star}$ the vector of the components of $\widetilde{\Psi}_h$ in the basis $\left(\Gamma_l\right)_{1\leqslant l\leqslant m^\star}$. The problem (24) reads: Find $\boldsymbol{X}_h := \left[\boldsymbol{X}_{1h}, \boldsymbol{X}_{3h}\right] \in \mathbb{R}^{m^\# + m^\star}$ such that

$$\mathbb{E}\boldsymbol{X}_h = \boldsymbol{F}, \tag{27a}$$

$$\mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h \geqslant 0, \quad \boldsymbol{X}_{3h} \geqslant 0, \quad \left(\mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h\right) \cdot \boldsymbol{X}_{3h} = 0. \tag{27b}$$

Here, $\mathbb{E} \in \mathbb{R}^{m^\#, m^\# + m^\star}$ is a rectangular block matrix having the following structure

$$\mathbb{E} := [\, \mathbb{A} \quad -\mathbb{B} \,],$$

where the matrices $\mathbb{A} \in \mathbb{R}^{m^\#, m^\#}$ and $\mathbb{B} \in \mathbb{R}^{m^\#, m^\star}$ are defined by

$$\mathbb{A}_{l,k} := a_h(\phi_k, \phi_l), \quad \mathbb{B}_{k,j} := b_h(\phi_k, \xi_j), \quad \forall 1 \leqslant j \leqslant m^\star, \quad \forall 1 \leqslant l, k \leqslant m^\#. \tag{28}$$

The right-hand side vector $\boldsymbol{F} \in \mathbb{R}^{m^\#}$ is defined by

$$\boldsymbol{F}_k := \ell_h(\phi_k) - a_h(g_h, \phi_k), \quad \forall 1 \leqslant k \leqslant m^\#.$$

Note that (27b) is a consequence of (24b), (25) and (26).

3. Semismooth Newton method

In the present work, we consider a semismooth Newton algorithm to solve the nonlinear problem (27). We first present the method and then prove local convergence properties.

3.1. Presentation of the semismooth Newton method

Let us recall the definition of the class of complementarity functions (or C-functions).

Definition 4. We say that a function $f : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$, $m \geqslant 1$ is a *C*-function if

$$\forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^m \times \mathbb{R}^m, \quad f(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} \geqslant 0, \quad \mathbf{y} \geqslant 0, \quad \mathbf{x} \cdot \mathbf{y} = 0. \tag{29}$$

Possible C-functions are the min and the Fischer–Burmeister functions given by

$$\begin{aligned} &(\min(\mathbf{x}, \mathbf{y}))_l := \min\left(\mathbf{x}_l, \mathbf{y}_l\right), \text{ and} \\ &\left(f_{\mathrm{FB}}(\mathbf{x}, \mathbf{y})\right)_l := \sqrt{\mathbf{x}_l^2 + \mathbf{y}_l^2} - \left(\mathbf{x}_l + \mathbf{y}_l\right), \quad l \in [1, m], \end{aligned} \tag{30}$$

see also [36,37,40] and the references therein for further information about C-functions. Note that the min function is not differentiable for $\mathbf{x} = \mathbf{y}$ and the f_{FB} function is not differentiable in $(\mathbf{0}, \mathbf{0})$. Such *C*-functions are used to transform (27b) into algebraic equalities. Let $\tilde{C}$ be any C-function, we can then use Definition 4 to write

$$\begin{aligned} &\tilde{C}(\mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h, \boldsymbol{X}_{3h}) = 0 \\ &\quad \iff \mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h \geqslant 0, \ \boldsymbol{X}_{3h} \geqslant 0, \text{ and } \left(\mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h\right) \cdot \boldsymbol{X}_{3h} = 0. \end{aligned}$$

We introduce the function $C : \mathbb{R}^{m^\star + m^\#} \to \mathbb{R}^{m^\star}$ defined by $C(\boldsymbol{X}_h) := \tilde{C}(\mathbb{B}^T \boldsymbol{X}_{1h} - \boldsymbol{\Psi}_h, \boldsymbol{X}_{3h})$. Problem (27) can then be equivalently rewritten as: Find $\boldsymbol{X}_h := \left[\boldsymbol{X}_{1h}, \boldsymbol{X}_{3h}\right] \in \mathbb{R}^{m^\# + m^\star}$ such that

$$\begin{aligned} &\mathbb{E}\boldsymbol{X}_h = \boldsymbol{F}, \\ &C(\boldsymbol{X}_h) = \mathbf{0}. \end{aligned} \tag{31}$$

Note that since $\tilde{C}$ is not necessarily Fréchet differentiable at every point, we cannot use the standard results concerning the Newton algorithm. However, the weaker regularity of the C-functions that are locally Lipschitz can be still enough to prove convergence properties for the semismooth Newton algorithm (see Section 3.2). The proof is similar to the one of the convergence of the classical Newton method but with the Clarke subdifferential (or generalized Jacobian of C) instead of its classical Jacobian. In particular, when all the elements of the Clarke subdifferential are invertible, the Clarke subdifferential is said to be regular [37, Chapter 7]. The semismooth Newton algorithm is given in Algorithm 1. We denote by $\|\cdot\|_2$ the ℓ^2-norm, i.e. $\forall m \in \mathbb{N}^*$, $\forall \boldsymbol{X} \in \mathbb{R}^m$, $\|\boldsymbol{X}\|_2^2 := \sum_{l=1}^{m} (X_l)^2$.

Algorithm 1 Semismooth Newton algorithm

```
1. Choose an initial vector $\boldsymbol{X}_h^0 \in \mathbb{R}^{m^\star+m^\#}$ and set $k=1$. Let $\varepsilon_{\rm lin} > 0$ be a fixed (small) parameter.
while $\left\| \begin{pmatrix} \boldsymbol{F} - \mathbb{E}\boldsymbol{X}_h^{k-1} \\ C(\boldsymbol{X}_h^{k-1}) \end{pmatrix} \right\|_2 \geqslant \varepsilon_{\rm lin} \left\| \begin{pmatrix} \boldsymbol{F} - \mathbb{E}\boldsymbol{X}_h^{0} \\ C(\boldsymbol{X}_h^{0}) \end{pmatrix} \right\|_2$ do
    2. For $k \geqslant 1$, $\boldsymbol{X}_h^{k-1}$ is given. Compute the Jacobian matrix "in the sense of Clarke" $\mathbb{J}^{k-1} \in \mathbb{R}^{m^\star+m^\#, m^\star+m^\#}$ and the right-hand side vector $\boldsymbol{B}^{k-1} \in \mathbb{R}^{m^\star+m^\#}$ respectively by
       $\mathbb{J}^{k-1} := \begin{bmatrix} \mathbb{E} \\ \mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1}) \end{bmatrix}, \quad \boldsymbol{B}^{k-1} := \begin{bmatrix} \boldsymbol{F} \\ \mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})\boldsymbol{X}_h^{k-1} - C(\boldsymbol{X}_h^{k-1}) \end{bmatrix}.$   (32)
       Here, $\mathbf{J}_{\mathrm{C}}$ is the generalized Jacobian of $C$.
    3. Find $\boldsymbol{X}_h^k \in \mathbb{R}^{m^\#+m^\star}$ as the solution of the linear system
       $\mathbb{J}^{k-1}\boldsymbol{X}_h^k = \boldsymbol{B}^{k-1}.$   (33)
end while
```

3.2. The case of the min function: convergence properties

In this section, the C-function considered is the min function given in (30). The Clarke subdifferential $\mathbf{J}_{\mathrm{C}}(\boldsymbol{X})$ of C at point $\boldsymbol{X} := \left[\boldsymbol{X}_1, \boldsymbol{X}_3\right]^T$ can be computed in the following way. First, we construct the following block matrices $\mathbb{K} := \left[\mathbb{B}^T, \boldsymbol{0}\right] \in \mathbb{R}^{m^\star, m^\star+m^\#}$ and $\mathbb{G} := \left[\boldsymbol{0}, \mathbb{I}_{\mathrm{d}}\right] \in \mathbb{R}^{m^\star, m^\star+m^\#}$, where $\mathbb{I}_{\mathrm{d}}$ denotes the identity matrix. Then, the lth row of the Jacobian matrix $\mathbf{J}_{\mathrm{C}}(\boldsymbol{X})$ is either given by the lth row of $\mathbb{K}$ if $\left(\mathbb{B}^T\boldsymbol{X}_1 - \boldsymbol{\Psi}_h\right)_l \leqslant \left(\boldsymbol{X}_3\right)_l$ or the lth row of $\mathbb{G}$ if $\left(\boldsymbol{X}_3\right)_l < \left(\mathbb{B}^T\boldsymbol{X}_1 - \boldsymbol{\Psi}_h\right)_l$ for $1 \leqslant l \leqslant m^\star$. We provide in this section a local convergence result for the semismooth Newton Algorithm (Algorithm 1) when the C-function min is employed.

Theorem 5. *Let $\tilde{C}$ be the C-function* min *defined in (30). Then Algorithm 1 is well defined. Moreover, if the first guess $\boldsymbol{X}_h^0$ is close enough to the solution $\boldsymbol{X}_h^*$ to the nonlinear system (31), then the sequence $\left(\boldsymbol{X}_h^k\right)_{k\geqslant 1}$ converges to $\boldsymbol{X}_h^*$ with a finite number of semismooth iterations and the local convergence is quadratic.*

Proof. Let us first prove that there exists a unique solution to (32)–(33) for every $\boldsymbol{X}_h^{k-1} \in \mathbb{R}^{m^\#+m^\star}$ given. Since it is a finite dimensional square system, existence of a solution for every right-hand side is equivalent to uniqueness of this solution. Let

$$\mathcal{A} := \left\{ i \in \left[1, m^\star\right] \text{ s.t. } \left(\mathbb{B}^T\boldsymbol{X}_{1h}^{k-1} - \boldsymbol{\Psi}_h\right)_i \leqslant \left(\boldsymbol{X}_{3h}^{k-1}\right)_i \right\},$$

and $\mathcal{A}^c$ its complementarity set in $[1, m^\star]$.

Let $\boldsymbol{X} := \left[\boldsymbol{X}_1, \boldsymbol{X}_3\right] \in \mathbb{R}^{m^\#+m^\star}$ be the solution to the problem (33) with no right-hand side, i.e.

$$\begin{aligned} &\mathbb{A}\boldsymbol{X}_1 - \mathbb{B}\boldsymbol{X}_3 = 0, \\ &\mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})\boldsymbol{X} = 0. \end{aligned} \tag{34}$$

We want to prove that $\boldsymbol{X} = 0$. The relation $\mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})\boldsymbol{X} = 0$ implies that $\forall i \in \mathcal{A}$, $(\mathbb{B}^T\boldsymbol{X}_1)_i = 0$ and $\forall j \in \mathcal{A}^c$, $(\boldsymbol{X}_3)_j = 0$ (see the construction of $\mathbf{J}_{\mathrm{C}}$ at the beginning of Section 3.2). Then we have

$$\boldsymbol{X}_3^T\mathbb{B}^T\boldsymbol{X}_1 = \sum_{i\in\mathcal{A}} \left(\boldsymbol{X}_3\right)_i \left(\mathbb{B}^T\boldsymbol{X}_1\right)_i + \sum_{j\in\mathcal{A}^c} \left(\boldsymbol{X}_3\right)_j \left(\mathbb{B}^T\boldsymbol{X}_1\right)_j = 0. \tag{35}$$

Multiplying the first line of (34) by $\boldsymbol{X}_1^T$ and employing (35) we obtain $\boldsymbol{X}_1^T\mathbb{A}\boldsymbol{X}_1 = 0$. Since a_h is coercive on $V_h \times V_h$, the matrix $\mathbb{A}$ is positive definite and then we have $\boldsymbol{X}_1 = 0$.

Let us set $\tilde{\lambda}_h := \sum_{1\leqslant i\leqslant m^\star} (X_3)_i \xi_i \in \widetilde{\Lambda}_h$ the function associated to $\boldsymbol{X}_3$. Moreover, $(\mathbb{B}\boldsymbol{X}_3)_j = (\tilde{\lambda}_h, \boldsymbol{\Phi}_h(\phi_j))_{\widetilde{\Lambda}_h}$ and since $(\Gamma_i)_{1\leqslant i\leqslant m^\star}$ is a basis of $\boldsymbol{\Phi}_h(V_h)$, for all $i \in [1, m^\star]$, there exists $v_{\Gamma_i} \in V_h$ such that $\boldsymbol{\Phi}_h(v_{\Gamma_i}) := \Gamma_i$. Then $\mathbb{B}\boldsymbol{X}_3 = 0$ implies that $\forall i \in [1, m^\star]$, $0 = (\tilde{\lambda}_h, \boldsymbol{\Phi}_h(v_{\Gamma_i}))_{\widetilde{\Lambda}_h} = (\tilde{\lambda}_h, \Gamma_i)_{\widetilde{\Lambda}_h} = (X_3)_i$. Then $\boldsymbol{X}_3 = 0$.

This proves that all Jacobian matrices of the Clarke subdifferential are invertible and thus the latter is regular. Moreover, Algorithm 1 is well defined at every step.

Furthermore, since the C-function min is Lipschitz around $\boldsymbol{X}_h^{k-1} \in \mathbb{R}^{m^\#+m^\star}$, there exists $K > 0$ such that (see [37, Lemma 7.5.2])

$$\sup_{\boldsymbol{X}\in\mathbb{R}^{m^\#+m^\star}} \max\left\{ |||\mathbf{J}_{\mathrm{C}}(\boldsymbol{X})|||, \left|\left|\left| \left[\mathbf{J}_{\mathrm{C}}(\boldsymbol{X})\right]^{-1} \right|\right|\right| \right\} \leqslant K, \tag{36}$$

where $|||\cdot|||$ stands for the usual matrix norm: for any matrix $\mathbb{H} \in \mathbb{R}^{m^\star, m^\star+m^\#}$,

$$|||\mathbb{H}||| = \sup_{\boldsymbol{Y}\in\mathbb{R}^{m^\star+m^\#},\ \boldsymbol{Y}\neq\boldsymbol{0}} \frac{\|\mathbb{H}\boldsymbol{Y}\|_2}{\|\boldsymbol{Y}\|_2}.$$

We now prove the local quadratic convergence of the algorithm. Solving (33) is equivalent to computing $\boldsymbol{X}_h^k := \boldsymbol{X}_h^{k-1} + \boldsymbol{D}_h^k$ where $\boldsymbol{D}_h^k := \left[\boldsymbol{D}_{1h}^k, \boldsymbol{D}_{3h}^k\right] \in \mathbb{R}^{m^\#+m^\star}$ is the solution to the problem

$$\mathbb{J}^{k-1}\boldsymbol{D}_h^k = -\begin{pmatrix} \mathbb{E}\boldsymbol{X}_h^{k-1} - \boldsymbol{F} \\ C(\boldsymbol{X}_h^{k-1}) \end{pmatrix}.$$

We denote by $\boldsymbol{X}_h^*$ the solution to the nonlinear system (27). According to (33), for $k \geqslant 1$, we have $\mathbb{E}\boldsymbol{X}_h^k = \boldsymbol{F}$. We then have for all $k \geqslant 2$

$$\begin{aligned} \boldsymbol{X}_h^k - \boldsymbol{X}_h^* &= \boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^* - \left[\mathbb{J}^{k-1}\right]^{-1} \begin{pmatrix} 0 \\ C(\boldsymbol{X}_h^{k-1}) \end{pmatrix} \\ &= -\left[\mathbb{J}^{k-1}\right]^{-1} \left[\begin{pmatrix} 0 \\ C(\boldsymbol{X}_h^{k-1}) - C(\boldsymbol{X}_h^*) \end{pmatrix} - \mathbb{J}^{k-1}\left(\boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^*\right) \right]. \end{aligned} \tag{37}$$

Since the min function is strongly semismooth [37, Definition 7.4.2], when $\boldsymbol{X}_h^{k-1}$ is close enough to $\boldsymbol{X}_h^*$, we have

$$\left\| C(\boldsymbol{X}_h^{k-1}) - C(\boldsymbol{X}_h^*) - \mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})(\boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^*) \right\|_2 \leqslant K \left\| \boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^* \right\|_2^2, \tag{38}$$

where $K > 0$ is given in (36). Observe that for $k \geqslant 2$

$$\mathbb{J}^{k-1}\left(\boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^*\right) = \begin{pmatrix} 0 \\ \mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})(\boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^*) \end{pmatrix}.$$

Using (37), (36) and (38) we get

$$\begin{aligned} &\left\| \boldsymbol{X}_h^k - \boldsymbol{X}_h^* \right\|_2 \\ &\leqslant \left|\left|\left| \left[\mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})\right]^{-1} \right|\right|\right| \left\| C(\boldsymbol{X}_h^{k-1}) - C(\boldsymbol{X}_h^*) - \mathbf{J}_{\mathrm{C}}(\boldsymbol{X}_h^{k-1})(\boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^*) \right\|_2 \\ &\leqslant K^2 \left\| \boldsymbol{X}_h^{k-1} - \boldsymbol{X}_h^* \right\|_2^2, \end{aligned}$$

and we get quadratic convergence.

We end this proof by recalling that there is only a finite set of possible matrices for $\mathbb{J}$. Then if the algorithm converges, it does in a finite number of steps. □

3.3. Inexact resolution of the linear algebraic system

Solving (33) with a direct method can be expensive. An alternative is to employ an inexact Newton algorithm (see [41–43]) which is a popular approach to speed up the convergence. Suppose thus that some iterative algebraic solver is applied to the linearized system (33). Given an initial vector $\boldsymbol{X}_h^{k,0} \in \mathbb{R}^{m^\# + m^\star}$, often taken as $\boldsymbol{X}_h^{k,0} := \boldsymbol{X}_h^{k-1}$, this yields on step $i \geqslant 1$ an approximation $\boldsymbol{X}_h^{k,i}$ to $\boldsymbol{X}_h^k$ satisfying

$$\mathbb{J}^{k-1}\boldsymbol{X}_h^{k,i} = \boldsymbol{B}^{k-1} - \boldsymbol{R}_h^{k,i}, \tag{39}$$

where $\boldsymbol{R}_h^{k,i} := \boldsymbol{B}^{k-1} - \mathbb{J}^{k-1}\boldsymbol{X}_h^{k,i} \in \mathbb{R}^{m^\# + m^\star}$ is the algebraic residual vector. The algebraic solver can be stopped when the relative algebraic residual satisfies

$$\left\| \boldsymbol{R}_h^{k,i} \right\|_2 \leqslant \eta_k \times \left\| \boldsymbol{B}^{k-1} - \mathbb{J}^{k-1}\boldsymbol{X}_h^{k,0} \right\|_2 . \tag{40}$$

Here, η_k is called the "forcing term". We refer to [44] for more details.

4. Application to the discretization of the contact problem between two membranes

In this section, we discretize the contact problem between two membranes, see (4). Several schemes fulfilling the framework of Sections 2–3 are provided. We first present the continuous problem and then we give the discretization of this problem with FEM, dG and HHO. Finally, a static condensation procedure is proposed to speed up the resolution with the HHO method.

4.1. Continuous setting

In this section, we introduce the continuous problem and the associated unknowns and functional spaces. We are interested in solving the problem (4) with the method developed in Sections 2–3. Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain. The unknowns are the displacements $\boldsymbol{u} := (u_1, u_2)$ of two membranes that cannot penetrate each other and the force acting from the lower membrane onto the upper one represented by the Lagrange multiplier λ. We denote by $\mu_1 > 0$ and $\mu_2 > 0$ the tension of the membranes and $\left(f_1, f_2\right) \in \left(L^2(\Omega)\right)^2$ represents the surface forces acting on them. The system of PDE's modeling the contact between these two membranes is the following:

$$\left\{\begin{array}{l} -\mu_1 \Delta u_1 - \lambda = f_1 \quad \text{in} \quad \Omega, \\ -\mu_2 \Delta u_2 + \lambda = f_2 \quad \text{in} \quad \Omega, \\ u_1 - u_2 \geqslant 0, \quad \lambda \geqslant 0, \quad (u_1 - u_2)\lambda = 0 \quad \text{in} \quad \Omega, \\ u_1 = g_1, \quad u_2 = g_2 \quad \text{on} \quad \partial\Omega, \end{array}\right. \tag{41}$$

where g_1 and g_2 are Dirichlet boundary data fulfilling $g_1 \geqslant g_2$ on $\partial\Omega$.

4.2. Weak formulation

The problem (41) is equivalent to (1) with $\mathcal{K}^g := \{\boldsymbol{v} := \left(v_1, v_2\right) \in H^1_{g_1}(\Omega) \times H^1_{g_2}(\Omega) \text{ s.t. } v_1 - v_2 \geqslant 0 \text{ a.e in } \Omega\}$ where $H^1_{g_\alpha}(\Omega) := \{v \in H^1(\Omega) \text{ s.t. } v|_{\partial\Omega} = g_\alpha\}$ and with the bilinear and linear forms defined by

$$a(\boldsymbol{v}, \boldsymbol{w}) := \sum_{\alpha=1}^{2} \mu_\alpha (\boldsymbol{\nabla} v_\alpha, \boldsymbol{\nabla} w_\alpha)_\Omega, \quad \ell(\boldsymbol{w}) := \sum_{\alpha=1}^{2} (f_\alpha, w_\alpha)_\Omega, \quad \forall \boldsymbol{v}, \boldsymbol{w} \in (H^1(\Omega))^2.$$

As presented in Section 2, we are interested in discretizing the mixed problem (8) with $V := (H^1_0(\Omega))^2$, $V^g := H^1_{g_1}(\Omega) \times H^1_{g_2}(\Omega)$, $\widetilde{\Lambda} := H^1(\Omega)$, $\Lambda := \{\chi \in H^1(\Omega) \text{ s.t. } \chi \geqslant 0 \text{ a.e. in } \Omega\}$, $\widehat{\Lambda}$ defined by (7) and $\Psi := 0$. Moreover, the bilinear form b is defined by

$$b(\boldsymbol{v}, \chi) := \langle \chi, v_1 - v_2 \rangle_{(H^1(\Omega))', H^1(\Omega)}, \quad \forall \chi \in (H^1(\Omega))', \quad \forall \boldsymbol{v} \in (H^1(\Omega))^2.$$

We propose to follow Sections 2 and 3 to discretize this problem using a finite element method (FEM), a discontinuous Galerkin (dG) method, and a hybrid high-order method (HHO). As a consequence of Theorem 2, all these discrete formulations are well-posed.

4.3. Discrete setting

Let $\mathcal{T}_h$ be a conforming simplicial mesh of Ω, i.e $\mathcal{T}_h$ is a set of triangles verifying $\bigcup_{K \in \mathcal{T}_h} \overline{K} = \overline{\Omega}$, where the intersection of the closure of two elements of $\mathcal{T}_h$ is either an empty set, a vertex, or an edge. The number of elements composing the mesh $\mathcal{T}_h$ is denoted by $\mathcal{N}_{\mathcal{T}}$. We denote by $\mathcal{E}_h$ the set of mesh edges and by $\mathcal{N}_{\mathcal{E}}$ its cardinality.

Let $S \subset \Omega$, we denote by h_S the diameter of S. Moreover, we set $h := \max_{K \in \mathcal{T}_h} h_K$ and for all $p \geqslant 0$, we denote by $\mathbb{P}_p(S)$ the set of polynomials of total degree at most p on S.

Furthermore, we denote by $\mathcal{V}_p$ the set of the Lagrange nodes $\mathbf{x}_l$ and by $\mathcal{N}_p$ its cardinality. The interior nodes are collected in the set $\mathcal{V}_p^{\text{int}}$ (with $\mathcal{N}_p^{\text{int}}$ its cardinality) and the boundary ones are collected in the set $\mathcal{V}_p^{\text{ext}}$. The nodes of an element $K \in \mathcal{T}_h$ are collected in the set $\mathcal{V}_K$ and we denote respectively by $\mathcal{V}_K^{\text{int}}$ and $\mathcal{V}_K^{\text{ext}}$ the set of the Lagrange nodes in $K \cap \Omega$ and in $K \cap \partial\Omega$.

For all the numerical schemes that we propose, we discretize the bilinear form a with a classical discretization of the Laplace problem and the bilinear form b is discretized using the usual L^2-scalar product of Ω. That way, we expect to get consistent schemes.

4.4. Finite element method

In this section, for $p \geqslant 1$, we consider continuous piecewise $\mathbb{P}_p$-polynomial functions. We introduce the finite element spaces

$$\begin{aligned}
\widetilde{\Lambda}_h &:= \left\{ v_h \in C^0(\overline{\Omega}) \text{ s.t. } v_h|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T}_h \right\}, \quad \widetilde{V}_h := (\widetilde{\Lambda}_h)^2, \\
V_h &:= \widetilde{V}_h \cap (H^1_0(\Omega))^2, \\
V_h^g &:= \prod_{\alpha=1}^{2} V_h^{g_\alpha}, \\
V_h^{g_\alpha} &:= \left\{ v_h \in C^0(\overline{\Omega}) \text{ s.t. } v_h|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T}_h,\ v_h|_{\partial\Omega} = g_\alpha \right\}.
\end{aligned}$$

Note that here we abuse the notation by writing g_α instead of $g_{\alpha h}$. We define the Lagrange basis function $\Gamma_l \in \widetilde{\Lambda}_h$ associated to the Lagrange node $\mathbf{x}_l \in \mathcal{V}_p$ by $\Gamma_l(\mathbf{x}_k) := \delta_{kl}$. We also define the discrete nonempty closed convex sets

$$\begin{aligned}
\mathcal{K}_h^g &:= \left\{ \boldsymbol{v}_h := \left(v_{1h}, v_{2h}\right) \in V_h^g,\ \left(v_{1h} - v_{2h}\right)(\mathbf{x}_l) \geqslant 0\ \forall \mathbf{x}_l \in \mathcal{V}_p^{\text{int}} \right\}, \\
\Lambda_h &:= \left\{ v_h \in \widetilde{\Lambda}_h \text{ s.t. } v_h(\mathbf{x}_l) \geqslant 0\ \forall \mathbf{x}_l \in \mathcal{V}_p^{\text{int}} \right\}, \\
\widehat{\Lambda}_h &:= \left\{ v_h \in \widetilde{\Lambda}_h \text{ s.t. } \left(v_h, \Gamma_l\right)_\Omega \geqslant 0\ \forall \mathbf{x}_l \in \mathcal{V}_p^{\text{int}},\ \left(v_h, \Gamma_l\right)_\Omega = 0\ \forall \mathbf{x}_l \in \mathcal{V}_p^{\text{ext}} \right\}.
\end{aligned}$$

Note that $\mathcal{K}_h^g \subset \mathcal{K}^g$ when $p = 1$ and $\mathcal{K}_h^g \not\subset \mathcal{K}^g$ when $p \geqslant 2$. However, $\Lambda_h \not\subset \Lambda$ for all $p \geqslant 1$ since we may have $\chi_h \in \Lambda_h$ with $\chi_h(\mathbf{x}_l) < 0$ for some $\mathbf{x}_l \in \mathcal{V}_p^{\text{ext}}$. For all $\boldsymbol{v}_h, \boldsymbol{w}_h \in \widetilde{V}_h$, we define

$$a_h(\boldsymbol{v}_h, \boldsymbol{w}_h) := \sum_{\alpha=1}^{2} \mu_\alpha \left(\boldsymbol{\nabla} v_{\alpha h}, \boldsymbol{\nabla} w_{\alpha h}\right)_\Omega, \quad \ell_h(\boldsymbol{v}_h) := \sum_{\alpha=1}^{2} \left(f_\alpha, v_{\alpha h}\right)_\Omega.$$

Note that a_h is coercive on $V_h \times V_h$. We then consider the mixed problem (18) with

$$b_h(\boldsymbol{w}_h, \zeta_h) := (w_{1h} - w_{2h}, \zeta_h)_\Omega, \quad \forall \boldsymbol{w}_h \in \widetilde{V}_h, \quad \forall \zeta_h \in \widetilde{\Lambda}_h.$$

This discretization method has already been studied in [25] in the context of a posteriori estimates. The framework of Section 2 applies here with $m^\flat := \mathcal{N}_p$ and $2m^\star = m^\# = 2\mathcal{N}_p^{\text{int}}$. The linear problem with complementarity constraints that we solve is (27) with

$$\mathbb{E} := \begin{bmatrix} \mu_1 \mathbb{S} & \mathbf{0} & -\mathbb{I}_{\text{d}} \\ \mathbf{0} & \mu_2 \mathbb{S} & \mathbb{I}_{\text{d}} \end{bmatrix}, \quad \boldsymbol{F} := \left[\boldsymbol{F}_1, \boldsymbol{F}_2\right], \quad \boldsymbol{X}_h := \left[\boldsymbol{X}_{1h}^a, \boldsymbol{X}_{2h}^b, \boldsymbol{X}_{3h}\right]. \tag{42}$$

Here, $\mathbb{S} \in \mathbb{R}^{m^\star, m^\star}$ is the finite element stiffness matrix defined by $\mathbb{S}_{l,k} := \left(\boldsymbol{\nabla}\Gamma_l, \boldsymbol{\nabla}\Gamma_k\right)_\Omega$, $\forall 1 \leqslant k, l \leqslant m^\star$, and $\boldsymbol{F}_\alpha \in \mathbb{R}^{m^\star}$, $\alpha \in \{1, 2\}$, is defined by $\left(\boldsymbol{F}_\alpha\right)_l := \left(f_\alpha, \Gamma_l\right)_\Omega - \mu_\alpha \left(\boldsymbol{\nabla} g_\alpha, \boldsymbol{\nabla}\Gamma_l\right)_\Omega$ $\forall 1 \leqslant l \leqslant m^\star$. The vectors $\boldsymbol{X}_{1h}^a$, $\boldsymbol{X}_{2h}^b$ and $\boldsymbol{X}_{3h}$ are the coordinates (up to liftings) of u_{1h}, u_{2h} in the basis

$(\Gamma_l)_{1\leqslant l\leqslant m^\star}$ and of λ_h in $(\xi_l)_{1\leqslant l\leqslant m^\star}$. The complementarity constraints read

$$\boldsymbol{X}_{1h}^a - \boldsymbol{X}_{2h}^b \geqslant 0, \quad \boldsymbol{X}_{3h} \geqslant 0, \quad \left(\boldsymbol{X}_{1h}^a - \boldsymbol{X}_{2h}^b\right)\cdot \boldsymbol{X}_{3h} = 0. \tag{43}$$

Any C-function can be employed to transform the complementarity constraints (43) onto a system of algebraic equalities. In the sequel, we employ Algorithm 1 with the min C-function to compute the solution of the nonlinear discrete problem. We expect to get a p convergence rate in energy norm.

4.5. Discontinuous Galerkin method

In this section, we consider the discontinuous Galerkin method for problem (41). The discontinuous Galerkin spaces corresponding to Section 2 are defined by

$$\begin{aligned}
\widetilde{\Lambda}_h &:= \left\{v_h \in L^2(\Omega) \text{ s.t. } v_h|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T}_h\right\} \not\subset \widetilde{\Lambda},\\
\widetilde{V}_h &:= \left\{\boldsymbol{v}_h := (v_{1h}, v_{2h}) \in (L^2(\Omega))^2 \text{ s.t. } \boldsymbol{v}_h|_K \in (\mathbb{P}_p(K))^2\ \forall K \in \mathcal{T}_h\right\} \not\subset \widetilde{V},\\
V_h &:= \left\{\boldsymbol{v}_h \in \widetilde{V}_h \text{ s.t. } \boldsymbol{v}_h = 0 \text{ on } \partial\Omega\right\} \not\subset V,\\
V_h^g &:= \prod_{\alpha=1}^{2} V_h^{g_\alpha} \not\subset V^g,\\
V_h^{g_\alpha} &:= \{v_h \in L^2(\Omega) \text{ s.t. } v_h|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T}_h \text{ and } v_h = g_\alpha \text{ on } \partial\Omega\}.
\end{aligned}$$

The discrete convex sets are defined by

$$\begin{aligned}
\Lambda_h &:= \left\{v_h \in \widetilde{\Lambda}_h \text{ s.t. } v_h|_K(\mathbf{x}_l) \geqslant 0\ \forall \mathbf{x}_l \in \mathcal{V}_K^{\text{int}} \text{ and } \forall K \in \mathcal{T}_h\right\} \not\subset \Lambda,\\
\mathcal{K}_h^g &:= \left\{\boldsymbol{v}_h := \left(v_{1h}, v_{2h}\right) \in V_h^g \text{ s.t. } \left(v_{1h} - v_{2h}\right)|_K(\mathbf{x}_l) \geqslant 0 \right.\\
&\qquad\qquad \left. \forall \mathbf{x}_l \in \mathcal{V}_K^{\text{int}}\ \forall K \in \mathcal{T}_h\right\} \not\subset \mathcal{K}^g,\\
\widehat{\Lambda}_h &:= \left\{ \begin{array}{c} v_h \in \widetilde{\Lambda}_h \text{ s.t. } \left(v_h|_K, \Gamma_l|_K\right)_K \geqslant 0\ \forall K \in \mathcal{T}_h,\ \forall \mathbf{x}_l \in \mathcal{V}_K^{\text{int}}, \\ \text{and } \left(v_h|_K, \Gamma_l|_K\right)_K = 0\ \forall K \in \mathcal{T}_h\ \forall \mathbf{x}_l \in \mathcal{V}_K^{\text{ext}} \end{array} \right\},
\end{aligned}$$

where here, for $m^\flat := \frac{1}{2}(p+1)(p+2)\mathcal{N}_{\mathcal{T}}$, $(\Gamma_l)_{1\leqslant l\leqslant m^\flat}$ is the basis of $\widetilde{\Lambda}_h$ such that for all $l \in [1, m^\flat]$, there exists $K \in \mathcal{T}_h$ such that the support of Γ_l is in K and Γ_l takes value one at one Lagrange node of K and zero at the other Lagrange nodes.

Let us define for all $\boldsymbol{v}_h, \boldsymbol{w}_h$ in $\widetilde{V}_h$ the bilinear form:

$$\begin{aligned}
a_h(\boldsymbol{v}_h, \boldsymbol{w}_h) &:= \sum_{\alpha=1}^{2} \mu_\alpha \mathcal{A}_h(v_{\alpha h}, w_{\alpha h}),\\
\mathcal{A}_h(v_{\alpha h}, w_{\alpha h}) &:= \sum_{K\in\mathcal{T}_h} \left(\boldsymbol{\nabla} v_{\alpha h}, \boldsymbol{\nabla} w_{\alpha h}\right)_K + \delta_h(v_{\alpha h}, w_{\alpha h}).
\end{aligned}$$

Here, several choices are possible for the bilinear form δ_h in order to enforce $\mathcal{A}_h$ to be coercive. We mention the SIPG method [45,46]: for all $v_h \in \widetilde{\Lambda}_h$ and $w_h \in \widetilde{\Lambda}_h$,

$$\begin{aligned}
\delta_h(v_h, w_h) := &- \sum_{F\in\mathcal{E}_h} \left(\{\{\boldsymbol{\nabla} w_h\}\}_F\, [\![v_h]\!]_F + \{\{\boldsymbol{\nabla} v_h\}\}_F\, [\![w_h]\!]_F, 1\right)_F\\
&+ \sum_{F\in\mathcal{E}_h} \frac{\gamma}{h_F} \left([\![w_h]\!]_F, [\![v_h]\!]_F\right)_F,
\end{aligned} \tag{44}$$

and the NIPG method [47]: for all $v_h \in \widetilde{\Lambda}_h$ and $w_h \in \widetilde{\Lambda}_h$,

$$\begin{aligned}
\delta_h(v_h, w_h) := &- \sum_{F\in\mathcal{E}_h} \left(\{\{\boldsymbol{\nabla} w_h\}\}_F\, [\![v_h]\!]_F - \{\{\boldsymbol{\nabla} v_h\}\}_F\, [\![w_h]\!]_F, 1\right)_F\\
&+ \sum_{F\in\mathcal{E}_h} \frac{\gamma}{h_F} \left([\![w_h]\!]_F, [\![v_h]\!]_F\right)_F,
\end{aligned} \tag{45}$$

where $[\![v_h]\!]_F$ and $\left\{\{v_h\}\right\}_F$ denote respectively the jump and the mean value of v_h across $F \in \mathcal{E}_h$. For edges on the boundary, these values will both be taken as v_h. Note that in the SIPG method, the parameter $\gamma > 0$ has to be large enough to enforce the coercivity of $\mathcal{A}_h$, see [46, Lemma 4.12], while it can be taken arbitrarily in the NIPG method.

Next, we define the continuous linear form ℓ_h by

$$\ell_h(\boldsymbol{w}_h) := \sum_{\alpha=1}^{2} \sum_{K\in\mathcal{T}_h} (f_\alpha|_K, w_{\alpha h}|_K)_K, \quad \forall \boldsymbol{w}_h \in \widetilde{V}_h. \tag{46}$$

We then consider the mixed problem (18) with

$$b_h(\boldsymbol{w}_h, \zeta_h) := \sum_{K\in\mathcal{T}_h} \left(w_{1h}|_K - w_{2h}|_K, \zeta_h|_K\right)_K, \quad \forall \boldsymbol{w}_h \in \widetilde{V}_h, \quad \forall \zeta_h \in \widetilde{\Lambda}_h. \tag{47}$$

Following Section 2, we solve (27) with

$$\mathbb{E} := \begin{bmatrix} \mu_1 \mathbb{S} & \mathbf{0} & -\mathbb{I}_{\mathrm{d}} \\ \mathbf{0} & \mu_2 \mathbb{S} & \mathbb{I}_{\mathrm{d}} \end{bmatrix}, \quad \boldsymbol{F} := \left[\boldsymbol{F}_1, \boldsymbol{F}_2\right], \quad \boldsymbol{X}_h := \left[\boldsymbol{X}_{1h}^a, \boldsymbol{X}_{2h}^b, \boldsymbol{X}_{3h}\right].$$

Here, $m^\star$ is the number of internal DOFs and $\mathbb{S} \in \mathbb{R}^{m^\star, m^\star}$ with $\mathbb{S}_{l,k} := \mathcal{A}_h(\Gamma_k, \Gamma_l)$ and $(\boldsymbol{F}_\alpha)_l := \sum_{K\in\mathcal{T}_h} (f_\alpha, \Gamma_l)_K - \mu_\alpha \mathcal{A}_h(g_{\alpha h}, \Gamma_l)$. The complementarity constraints are this time given in all elements $K \in \mathcal{T}_h$ by

$$\boldsymbol{X}_{1h}^a|_K - \boldsymbol{X}_{2h}^b|_K \geqslant 0, \quad \boldsymbol{X}_{3h}|_K \geqslant 0, \quad \left(\boldsymbol{X}_{1h}^a|_K - \boldsymbol{X}_{2h}^b|_K\right)\cdot \boldsymbol{X}_{3h}|_K = 0, \tag{48}$$

where $\boldsymbol{X}_{1h}^a|_K$, $\boldsymbol{X}_{2h}^b|_K$ and $\boldsymbol{X}_{3h}|_K$ denote the coordinates of respectively $u_{1h}|_K$, $u_{2h}|_K$ and $\lambda_h|_K$ in the element $K \in \mathcal{T}_h$. We expect to get a p convergence rate in energy norm.

4.6. Hybrid high-order method

The hybrid high-order method (HHO) has been recently introduced in [48,49]. It is closely related to hybridizable discontinuous Galerkin (HDG) and to nonconforming virtual element methods (ncVEM) [50]. As other discontinuous skeletal methods, the unknowns are polynomial functions attached to the cells and the edges of the mesh. The polynomials attached to the edges are independent and do not necessarily correspond to the traces of the cell polynomials.

The polynomials attached to the cells can be eliminated through a static condensation procedure (see Section 4.7). The size of the linear system that is solved is then equal to the number of edge unknowns. The cell unknowns can finally be recovered by local solves in a post-processing step. For high-order polynomials, we expect this method to have fewer degrees of freedom than more classical methods such as FEM.

A HHO method for a contact problem has already been proposed in [21]. The constraint was that the mean value of $\boldsymbol{\Phi}_h(u_h)$ over each cell had to be nonnegative. In our approach, the constraint is expressed nodewise.

We present in this section, the HHO method without the static condensation procedure that is treated in Section 4.7. For $p \geqslant 1$, the displacement of each membrane is represented by a $\mathbb{P}_p$-polynomial function in every cell $K \in \mathcal{T}_h$ and a $\mathbb{P}_{p-1}$-polynomial function on every edge $F \in \mathcal{E}_h$. We introduce the following space

$$\hat{U}_h := \prod_{K\in\mathcal{T}_h} \mathbb{P}_p(K) \times \prod_{F\in\mathcal{E}_h} \mathbb{P}_{p-1}(F), \tag{49}$$

and for $K \in \mathcal{T}_h$, its local analogue

$$\hat{U}_K := \mathbb{P}_p(K) \times \prod_{F\in\mathcal{E}_K} \mathbb{P}_{p-1}(F),$$

that contains the polynomials attached to the cell K and its surrounding edges (the corresponding degrees of freedom are represented in Fig. 1). We denoted by $\mathcal{E}_K := \left\{F \in \mathcal{E}_h \text{ s.t. } F \subset \partial K\right\}$ the set of edges surrounding K.

As usual for HHO methods, for all $K \in \mathcal{T}_h$, an element $\hat{v}_K$ of $\hat{U}_K$ has a polynomial component attached to the cell K that we denote by v_K and a polynomial component attached to every edge surrounding K that we denote by $v_{\partial K} := (v_F)_{F\in\mathcal{E}_K}$. In a similar way, for every element $\hat{v}_h$ of $\hat{U}_h$, we denote by v_K the polynomial function attached to $K \in \mathcal{T}_h$ and by $v_{\partial K} := (v_F)_{F\in\mathcal{E}_K}$ the polynomial functions attached to the surrounding edges of K and $\hat{v}_K := (v_K, v_{\partial K})$.

We introduce the following vector spaces

$$\widetilde{V}_h := \left(\hat{U}_h\right)^2, \quad V_h := \left\{\boldsymbol{v}_h \in \widetilde{V}_h \text{ s.t. } \boldsymbol{v}_h|_{\partial\Omega} = 0\right\} \quad \text{and}$$

$$V_h^g := \left\{\boldsymbol{v}_h \in \widetilde{V}_h \text{ s.t. } \boldsymbol{v}_h|_{\partial\Omega} = \boldsymbol{g}\right\},$$

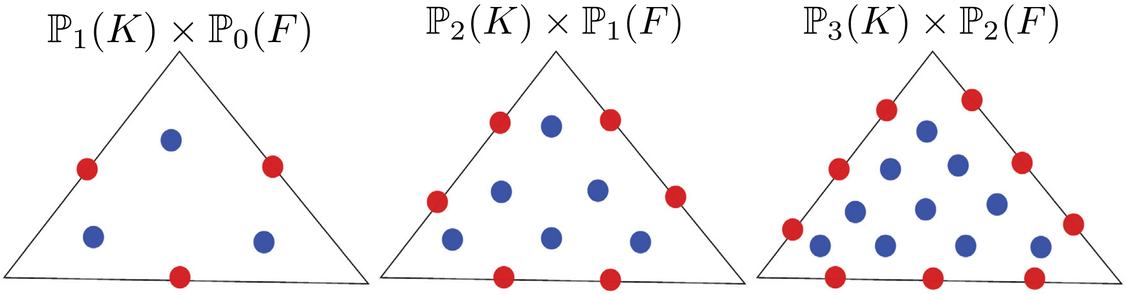
$\mathbb{P}_1(K) \times \mathbb{P}_0(F)$
$\mathbb{P}_2(K) \times \mathbb{P}_1(F)$
$\mathbb{P}_3(K) \times \mathbb{P}_2(F)$

Fig. 1. Representation of the degrees of freedom of $\hat{U}_K$. The cell DOFs are in blue, the edges DOFs are in red.

where here $\boldsymbol{v}_h|_{\partial\Omega}$ stands for the polynomials attached to the faces composing $\partial\Omega$. For $\boldsymbol{v}_h \in \widetilde{V}_h$, we denote by $\hat{v}_{1h}, \hat{v}_{2h} \in \hat{U}_h$ the components of $\boldsymbol{v}_h$, i.e. $\boldsymbol{v}_h := (\hat{v}_{1h}, \hat{v}_{2h})$.

The Lagrange multiplier is represented by a $\mathbb{P}_p$-polynomial function in every cell $K \in \mathcal{T}_h$, so that

$$\widetilde{\Lambda}_h := \prod_{K\in\mathcal{T}_h} \mathbb{P}_p(K). \tag{50}$$

For all χ_h of $\widetilde{\Lambda}_h$, we denote by χ_K the polynomial of χ_h attached to the cell $K \in \mathcal{T}_h$.

We define in every cell $K \in \mathcal{T}_h$ a gradient reconstruction operator $\mathbf{G}_K : \hat{U}_K \to \mathbb{P}_p(K;\mathbb{R}^2)$ such that for all $\hat{v}_K \in \hat{U}_K$ and for all $\mathbf{q} \in \mathbb{P}_p(K;\mathbb{R}^2)$ we have

$$(\mathbf{G}_K(\hat{v}_K), \mathbf{q})_K := (\boldsymbol{\nabla} v_K, \mathbf{q})_K + \sum_{F\in\mathcal{E}_K} (v_F - v_K, \mathbf{q}\cdot\mathbf{n}_K)_F,$$

where $\mathbb{P}_p(K;\mathbb{R}^2)$ denotes the set of vector-valued polynomials of degree at most p and $\mathbf{n}_K$ is the outward unit normal vector to the element K. This gradient reconstruction takes into account the value of the polynomials attached to the cell K and to the surrounding edges. Moreover it approximates the continuous gradient at optimal rate, see for instance [51, Lemma 8] for a proof in the context of unfitted meshes.

We consider problem (18) with the linear and bilinear forms defined such that for all $\boldsymbol{v}_h, \boldsymbol{w}_h \in \widetilde{V}_h$ and all $\chi_h \in \widetilde{\Lambda}_h$,

$$a_h(\boldsymbol{v}_h, \boldsymbol{w}_h) := \sum_{\alpha=1}^{2} \mu_\alpha \mathcal{A}_h(\hat{v}_{\alpha h}, \hat{w}_{\alpha h}),$$

$$\ell_h(\boldsymbol{w}_h) := \sum_{K\in\mathcal{T}_h} \sum_{\alpha=1}^{2} (f_\alpha, w_{\alpha K})_K, \quad b_h(\boldsymbol{w}_h, \chi_h) := \sum_{K\in\mathcal{T}_h} \left(w_{1K} - w_{2K}, \chi_K\right)_K,$$

where for all $\hat{v}_h, \hat{w}_h \in \hat{U}_h$

$$\mathcal{A}_h(\hat{v}_h, \hat{w}_h) := \sum_{K\in\mathcal{T}_h} \Bigg(\left(\mathbf{G}_K(\hat{v}_K), \mathbf{G}_K(\hat{w}_K)\right)_K + h_K^{-1} \sum_{F\in\mathcal{E}_K} \left(\Pi_F^{p-1}(v_K - v_F), w_K - w_F\right)_F \Bigg),$$

where Π_F^{p-1} is the L^2-projector onto $\mathbb{P}_{p-1}(F)$. A proof of the coercivity of a_h can be found for instance in [51, Corollary 7] in the context of unfitted meshes.

Note that in the present approach, we choose to impose the complementarity constraints on the cell unknowns only (we do not impose constraints on the polynomials attached to the edges), so that $\boldsymbol{\Phi}_h : (\hat{U}_h)^2 \to \prod_{K\in\mathcal{T}_h} \mathbb{P}_p(K)$ is defined by

$$\boldsymbol{\Phi}_h(\boldsymbol{v}_h)|_K := v_{1K} - v_{2K}, \quad \forall K \in \mathcal{T}_h, \quad \forall \boldsymbol{v}_h \in \widetilde{V}_h.$$

The nonempty closed convex set is then

$$\mathcal{K}_h^g := \{\boldsymbol{v}_h \in V_h^g \text{ s.t. } v_{1K}(\mathbf{x}_l) - v_{2K}(\mathbf{x}_l) \geqslant 0 \ \forall K \in \mathcal{T}_h, \ \forall \mathbf{x}_l \in \mathcal{V}_K\},$$

and we have

$$\Lambda_h := \left\{ v_h \in \widetilde{\Lambda}_h \text{ s.t. } v_h|_K(\mathbf{x}_l) \geqslant 0 \ \forall \mathbf{x}_l \in \mathcal{V}_K \text{ and } \forall K \in \mathcal{T}_h \right\} \not\subset \Lambda,$$

$$\hat{\Lambda}_h := \left\{ v_h \in \widetilde{\Lambda}_h \text{ s.t. } \left(v_h|_K, \Gamma_l|_K\right)_K \geqslant 0 \ \forall K \in \mathcal{T}_h, \ \forall \mathbf{x}_l \in \mathcal{V}_K \right\},$$

where $(\Gamma_l)_{1\leqslant l\leqslant m_c}$ is the basis defined in Section 4.5. Here, we have denoted by $m_c := m^\star = \frac{1}{2}(p+1)(p+2)\mathcal{N}_\mathcal{T}$ the number of basis functions attached to the cells.

According to the definitions (49) and (50), we can complete the basis of cell functions $(\Gamma_l)_{1\leqslant l\leqslant m_C}$ (basis of $\widetilde{\Lambda}_h$) with a basis of edge functions $(\beta_l)_{1\leqslant l\leqslant m_F}$ to get a basis of $\hat{U}_h$, where $m_F := \mathcal{N}_\mathcal{E}^{\text{int}} p$ with $\mathcal{N}_\mathcal{E}^{\text{int}}$ the number of internal edges. We then solve problem (27) with

$$\mathbb{E} := \begin{bmatrix} \mu_1\mathbb{S}_{CC} & \mu_1\mathbb{S}_{CF} & \mathbf{0} & \mathbf{0} & -\mathbb{I}_{\text{d}} \\ \mu_1\mathbb{S}_{FC} & \mu_1\mathbb{S}_{FF} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_2\mathbb{S}_{CC} & \mu_2\mathbb{S}_{CF} & \mathbb{I}_{\text{d}} \\ \mathbf{0} & \mathbf{0} & \mu_2\mathbb{S}_{FC} & \mu_2\mathbb{S}_{FF} & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{F} := \begin{bmatrix} \boldsymbol{F}_1 \\ \mathbf{0} \\ \boldsymbol{F}_2 \\ \mathbf{0} \end{bmatrix},$$

$$\boldsymbol{X}_h := \begin{bmatrix} \boldsymbol{X}_C^a \\ \boldsymbol{X}_F^a \\ \boldsymbol{X}_C^b \\ \boldsymbol{X}_F^b \\ \boldsymbol{X}_{3h} \end{bmatrix}.$$

Here, the matrices $\mathbb{S}_{CC} \in \mathbb{R}^{m_c,m_c}$, $\mathbb{S}_{CF} \in \mathbb{R}^{m_c,m_F}$, $\mathbb{S}_{FC} \in \mathbb{R}^{m_F,m_c}$, and $\mathbb{S}_{FF} \in \mathbb{R}^{m_F,m_F}$ are defined by

$$\begin{aligned} (\mathbb{S}_{CC})_{l,k} &:= \mathcal{A}_h(\Gamma_k, \Gamma_l), \quad \forall 1 \leqslant l,k \leqslant m_c, \\ (\mathbb{S}_{CF})_{l,k} &:= \mathcal{A}_h(\beta_k, \Gamma_l), \quad \forall 1 \leqslant l \leqslant m_c, \quad \forall 1 \leqslant k \leqslant m_F, \\ (\mathbb{S}_{FC})_{l,k} &:= \mathcal{A}_h(\Gamma_k, \beta_l), \quad \forall 1 \leqslant l \leqslant m_F, \quad \forall 1 \leqslant k \leqslant m_c, \\ (\mathbb{S}_{FF})_{l,k} &:= \mathcal{A}_h(\beta_k, \beta_l), \quad \forall 1 \leqslant l,k \leqslant m_F. \end{aligned}$$

Moreover, the right-hand side $\boldsymbol{F}_\alpha \in \mathbb{R}^{m_c}$ is defined by $(\boldsymbol{F}_\alpha)_l := (f_\alpha, \Gamma_l)_\Omega$ $\forall 1 \leqslant l \leqslant m_c$. The unknown vector $\boldsymbol{X}_h \in \mathbb{R}^{3m_c+2m_F}$ is composed of the cell DOFs and edge DOFs of the first displacement denoted by $\boldsymbol{X}_C^a \in \mathbb{R}^{m_c}$ and $\boldsymbol{X}_F^a \in \mathbb{R}^{m_F}$, the cell DOFs and edge DOFs of the second displacement denoted by $\boldsymbol{X}_C^b \in \mathbb{R}^{m_c}$ and $\boldsymbol{X}_F^b \in \mathbb{R}^{m_F}$, and the cell DOFS of the Lagrange multiplier denoted by $\boldsymbol{X}_{3h} \in \mathbb{R}^{m_c}$. The complementarity constraints are given by (48). We expect to get a convergence rate of order p in energy norm.

Remark 4.1. The discrete displacements $\hat{u}_{1h}$ and $\hat{u}_{2h}$ have cell and edge degrees of freedom contrary to the discrete Lagrange multiplier λ_h which only has cell degrees of freedom. We tested numerically a similar discretization with constraints on the edge unknowns then generating also edge degrees of freedom for the Lagrange multiplier. We observed similar results but involving a larger linear system.

4.7. Static condensation for skeletal methods

We describe in this section the static condensation procedure used to speed up skeletal methods such as HHO. In this procedure, we first rewrite the linear system as a system involving the edge unknowns only. Then, we recover the cell unknowns by means of local solves. For the sake of clarity, we first present how the cell unknowns can be recovered and then we give the linear problem fulfilled by the edge unknowns. This procedure is summed up in Algorithm 2. Note that a preprocessing step is needed before every semismooth Newton iteration. In this section, we treat only the min C-function as we proved its convergence properties in Section 3.2.

Let $k \geqslant 1$ be a semismooth Newton step and let $\boldsymbol{X}_h^{k-1} \in \mathbb{R}^{3m_c+2m_F}$ be the associated solution (computed e.g. by Algorithm 1). For every cell $K \in \mathcal{T}_h$, if we are given the solution attached to the surrounding edges, then we are able to recover the solution attached to K. More precisely, let $\boldsymbol{X}_K^{k-1} := [\boldsymbol{X}_{1K}^{k-1}, \boldsymbol{X}_{2K}^{k-1}, \boldsymbol{X}_{3K}^{k-1}] \in \mathbb{R}^{3\dim(\mathbb{P}_p(K))}$ be the components of the solution attached to K representing respectively u_{1K}^{k-1}, u_{2K}^{k-1} and λ_K^{k-1} in $\mathbb{P}_p(K)$ (at Newton step $k-1$) and let $\boldsymbol{X}_{\partial K}^k \in \mathbb{R}^{2\times 3\dim(\mathbb{P}_{p-1}(F))}$ be the components of $\boldsymbol{u}_{\partial K}^k := (u_1^k|_{\partial K}, u_2^k|_{\partial K})$ the solution attached to the surrounding edges at the semismooth Newton step k. Then, knowing $\boldsymbol{X}_{\partial K}^k$ and $\boldsymbol{X}_K^{k-1}$, we can recover the local cell unknowns $\boldsymbol{X}_K^k := [\boldsymbol{X}_{1K}^k \; \boldsymbol{X}_{2K}^k \; \boldsymbol{X}_{3K}^k]$ by solving the local problem: Find $\boldsymbol{X}_K^k \in \mathbb{R}^{3\dim(\mathbb{P}_p(K))}$ such that

$$\mathbb{J}_K^{k-1} \boldsymbol{X}_K^k = \boldsymbol{B}_K^k, \tag{51}$$

where the matrix $\mathbb{J}_K^{k-1} \in \mathbb{R}^{3\dim(\mathbb{P}_p(K)),3\dim(\mathbb{P}_p(K))}$ is the local contribution to the generalized jacobian $\mathbb{J}^{k-1}$ (see (32)), i.e.

$$\begin{aligned} \mathbb{J}_K^{k-1} &:= \begin{bmatrix} \mu_1 \mathbb{S}_{KK} & 0 & -\mathbb{I}_{\mathrm{d}} \\ 0 & \mu_2 \mathbb{S}_{KK} & +\mathbb{I}_{\mathrm{d}} \\ \mathbb{C}_{1K} & \mathbb{C}_{2K} & \mathbb{C}_{3K} \end{bmatrix}, \\ \boldsymbol{B}_K^k &:= \begin{bmatrix} \boldsymbol{F}_{1K} - \mu_1 \mathbb{S}_{KF} \boldsymbol{X}_{1\partial K}^k \\ \boldsymbol{F}_{2K} - \mu_2 \mathbb{S}_{KF} \boldsymbol{X}_{2\partial K}^k \\ 0 \end{bmatrix}. \end{aligned} \tag{52}$$

For $l \geqslant 1$ and $K \in \mathcal{T}_h$, let us denote by K_l the global index associated to the lth basis function attached to K. We also denote by $\mathcal{F}_l$ the global index corresponding to the lth basis function attached to ∂K. The last line-block $\left[\mathbb{C}_{1K}, \mathbb{C}_{2K}, \mathbb{C}_{3K}\right] \in \mathbb{R}^{\dim(\mathbb{P}_p(K)),3\dim(\mathbb{P}_p(K))}$ of matrix $\mathbb{J}_K^{k-1}$ is the local version of $\mathbf{J_C}(\boldsymbol{X}^{k-1})$ defined in Section 3.2, i.e. if $\left(u_{1K}^{k-1} - u_{2K}^{k-1}\right)(\mathbf{x}_{K_l}) \leqslant \lambda_K^{k-1}(\mathbf{x}_{K_l})$ the lth line of $\left[\mathbb{C}_{1K}, \mathbb{C}_{2K}, \mathbb{C}_{3K}\right]$ is defined by the lth line of the block matrix $\left[\mathbb{I}_{\mathrm{d}}, -\mathbb{I}_{\mathrm{d}}, \mathbf{0}\right] \in \mathbb{R}^{\dim(\mathbb{P}_p(K)),3\dim(\mathbb{P}_p(K))}$ and if $\left(u_{1K}^{k-1} - u_{2K}^{k-1}\right)(\mathbf{x}_{K_l}) > \lambda_K^{k-1}(\mathbf{x}_{K_l})$ the lth line of $\left[\mathbb{C}_{1K}, \mathbb{C}_{2K}, \mathbb{C}_{3K}\right]$ is defined by the lth line of the block matrix $\left[\mathbf{0}, \mathbf{0}, \mathbb{I}_{\mathrm{d}}\right] \in \mathbb{R}^{\dim(\mathbb{P}_p(K)),3\dim(\mathbb{P}_p(K))}$. Furthermore,

$$\begin{aligned} \left[\mathbb{S}_{KK}\right]_{l,l'} &:= \mathcal{A}_h(\Gamma_{K_{l'}}, \Gamma_{K_l}), \quad \forall 1 \leqslant l, l' \leqslant \dim(\mathbb{P}_p(K)), \\ \left[\mathbb{S}_{KF}\right]_{l,l'} &:= \mathcal{A}_h(\beta_{\mathcal{F}_{l'}}, \Gamma_{K_l}), \ \forall 1 \leqslant l \leqslant \dim(\mathbb{P}_p(K)), \ \forall 1 \leqslant l' \leqslant 3\dim(\mathbb{P}_{p-1}(F)), \end{aligned}$$

and $\boldsymbol{F}_{\alpha K}$ denotes the components of $\boldsymbol{F}_\alpha$ associated to $K \in \mathcal{T}_h$. In the previous expressions, we have used (Γ_l) the basis of functions attached to the cells and (β_l) the basis of functions attached to the edges.

Remark 4.2. The problem (51) actually corresponds to a local contact problem between two membranes: given the value of the displacements of the two membranes on the edges composing ∂K, we solve the contact problem inside the cell K.

In a similar way to the proof of Theorem 5, we can prove that problem (51) admits a unique solution $\boldsymbol{X}_K^k \in \mathbb{R}^{3\dim(\mathbb{P}_p(K))}$ and then

$$\boldsymbol{X}_K^k = \left[\mathbb{J}_K^{k-1}\right]^{-1} \boldsymbol{B}_K^k, \tag{53}$$

where

$$\left[\mathbb{J}_K^{k-1}\right]^{-1} =: \begin{bmatrix} \mathbb{D}_{11}^{k-1} & \mathbb{D}_{12}^{k-1} & \mathbb{D}_{13}^{k-1} \\ \mathbb{D}_{21}^{k-1} & \mathbb{D}_{22}^{k-1} & \mathbb{D}_{23}^{k-1} \\ \mathbb{D}_{31}^{k-1} & \mathbb{D}_{32}^{k-1} & \mathbb{D}_{33}^{k-1} \end{bmatrix}.$$

Here, the matrices $\mathbb{D}_{\alpha\gamma}^{k-1} \in \mathbb{R}^{\dim(\mathbb{P}_p(K)),\dim(\mathbb{P}_p(K))}$, α, $\gamma \in \{1,2,3\}$ are identified from the numerical computation of $\left[\mathbb{J}_K^{k-1}\right]^{-1}$. Now, let us present the problem satisfied by the edge unknowns. Let us denote by $\boldsymbol{X}_{hF}^k \in \mathbb{R}^{2m_F}$ the coordinates of all the polynomial unknowns attached to the edges (at the semismooth Newton step $k \geqslant 1$). These coordinates can be computed by solving the following problem: Find $\boldsymbol{X}_{hF}^k := [\boldsymbol{X}_a^k, \boldsymbol{X}_b^k] \in \mathbb{R}^{2m_F}$ such that

$$\begin{bmatrix} \widetilde{\mathbb{A}}_{11}^{k-1} & \widetilde{\mathbb{A}}_{12}^{k-1} \\ \widetilde{\mathbb{A}}_{21}^{k-1} & \widetilde{\mathbb{A}}_{22}^{k-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_a^k \\ \boldsymbol{X}_b^k \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{F}}_1^{k-1} \\ \widetilde{\boldsymbol{F}}_2^{k-1} \end{bmatrix}, \tag{54}$$

where for α, $\gamma \in \{1,2\}$, $\widetilde{\mathbb{A}}_{\alpha\gamma}^{k-1} \in \mathbb{R}^{m_F,m_F}$ and $\widetilde{\boldsymbol{F}}_\alpha^{k-1} \in \mathbb{R}^{m_F}$ are defined by

$$[\widetilde{\mathbb{A}}_{\alpha\gamma}^{k-1}]_{i,j} := \delta_{\alpha\gamma} \mu_\alpha [\mathbb{S}_{FF}]_{i,j} + \sum_{K \in \mathcal{T}_h} \left[-\mu_\alpha \mu_\gamma \mathbb{S}_{FK} \mathbb{D}_{\alpha\gamma}^{k-1} \mathbb{S}_{KF}\right]_{\tilde{\mathcal{F}}_i, \tilde{\mathcal{F}}_j}, \tag{55}$$

$$[\widetilde{\boldsymbol{F}}_\alpha^{k-1}]_i := \sum_{K \in \mathcal{T}_h} \left[-\mu_\alpha \mathbb{S}_{FK} (\mathbb{D}_{\alpha 1}^{k-1} \boldsymbol{F}_{1K} + \mathbb{D}_{\alpha 2}^{k-1} \boldsymbol{F}_{2K})\right]_{\tilde{\mathcal{F}}_i}, \tag{56}$$

where $[\mathbb{S}_{FF}]_{i,j} := \mathcal{A}_h(\beta_j, \beta_i)$ and $(\mathbb{S}_{FK})_{i,j} := \mathcal{A}_h(\Gamma_{K_j}, \beta_{F_i})$. Here, we denoted by $\tilde{\mathcal{F}}_i$ the local edge index associated to β_i. If i is not associated to a basis function in ∂K then we set $[\cdot]_{\tilde{\mathcal{F}}_i} = 0$.

Observe that constructing the matrices $\widetilde{\mathbb{A}}_{\alpha\gamma}^{k-1}$ follows an assembling. For every $K \in \mathcal{T}_h$, we compute the local contribution corresponding to the terms inside the brackets in (55)–(56) and we add it to the global contributions provided by the matrix $\mathbb{S}_{FF}$. The resulting stencil couples unknowns attached to neighboring edges (in the sense of cells).

The semismooth Newton algorithm with static condensation consists first in solving the global problem (54) in order to find the degrees of freedom attached to the edges of the mesh and then solve for every $K \in \mathcal{T}_h$ the local problem (51) to find the unknowns attached to the cells. We sum up these stages in Algorithm 2.

Algorithm 2 Newton-min algorithm with static condensation

```
1. Choose an initial vector $\boldsymbol{X}_h^0 := \left[\boldsymbol{X}_{hC}^0, \boldsymbol{X}_{hF}^0\right] \in \mathbb{R}^{3m_c+2m_F}$ and set $k = 1$.
2. Let $\varepsilon_{\mathrm{lin}} > 0$ be a fixed parameter. Consider the C-function min of (30).
while $\left\| \begin{pmatrix} \boldsymbol{F} - \mathbb{E}\boldsymbol{X}_h^{k-1} \\ C(\boldsymbol{X}_h^{k-1}) \end{pmatrix} \right\|_2 \geqslant \varepsilon_{\mathrm{lin}} \left\| \begin{pmatrix} \boldsymbol{F} - \mathbb{E}\boldsymbol{X}_h^0 \\ C(\boldsymbol{X}_h^0) \end{pmatrix} \right\|_2$ do
    3. For $k \geqslant 1$, $\boldsymbol{X}_h^{k-1} \in \mathbb{R}^{3m_C+2m_F}$ is given. In every cell $K \in \mathcal{T}_h$, we compute the local matrices of (52) and we can identify $\mathbb{D}_{\alpha\gamma}^{k-1}$, $1 \leqslant \alpha, \gamma \leqslant 3$ with (53). We then assemble the local contributions to get the matrices (55)–(56).
    4. We solve the linear problem (54) and get the coordinates $\boldsymbol{X}_{hF}^k \in \mathbb{R}^{2m_F}$ of the edges components.
    5. For every cell $K \in \mathcal{T}_h$, we solve the local problem (51) to recover the unknowns attached to the cells. We build a new vector $\boldsymbol{X}_h^k \in \mathbb{R}^{3m_C+2m_F}$ and we test the condition of the while loop.
end while
```

Lemma 6. *For $\boldsymbol{X}_h^{k-1} \in \mathbb{R}^{3m_c+2m_F}$ given, the vector $\boldsymbol{X}_h^k$ obtained by one Newton-min iteration of Algorithm 2 coincides with the one obtained by one Newton-min iteration of Algorithm 1.*

Proof. Let us denote by $\overline{\boldsymbol{X}_h^k}$ the solution obtained by Algorithm 1. We can easily show that the components of $\overline{\boldsymbol{X}_h^k}$ attached to K are solution to (51) (for all $K \in \mathcal{T}_h$). Moreover, by injecting (53) in (54), we show that the components of $\overline{\boldsymbol{X}_h^k}$ attached to the edges are solution to (54).

We have proved that for every right-hand side, the solution given by Algorithm 1 is solution to (54) (for its face components) and to (51) (for its cell components). These systems are squared, then the solution given by Algorithm 2 coincides with the one given by Algorithm 1. □

Remark 4.3. Compared to Algorithm 1, Algorithm 2 reduces the size of the linear problem to solve at each iteration of the while loop since only the degrees of freedom attached to the edges are considered. However, since the matrices $\mathbb{D}_{\alpha\gamma}^{k-1}$ depend on the previous state $\boldsymbol{X}_K^{k-1}$, the matrices (55)–(56) have to be (at least partially) assembled at every step of the while loop. This increases the assembly time of the method

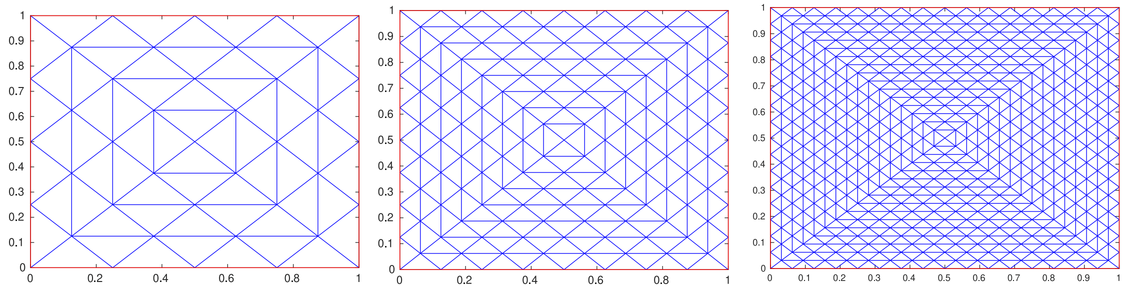

Fig. 2. Left: coarse mesh $\mathcal{T}_0$ containing 64 elements with $h_0 = 0.25$. Middle: first refined mesh $\mathcal{T}_1$ containing 256 elements with $h_1 = 0.125$. Right: second refined mesh $\mathcal{T}_2$ containing 1024 elements with $h_2 = 0.0625$.

Table 1
Number of DOFs for the finite element discretization (FEM).

Mesh	Number of mesh elements	Mesh size h	$\mathbb{P}_1$ DOFs	$\mathbb{P}_2$ DOFs	$\mathbb{P}_3$ DOFs	$\mathbb{P}_4$ DOFs
$\mathcal{T}_0$	64	0.25	75	339	795	1443
$\mathcal{T}_1$	256	0.125	339	1443	3315	5955
$\mathcal{T}_2$	1024	0.0625	1443	5955	13539	24195
$\mathcal{T}_3$	4096	0.03125	5955	24195	54723	97539
$\mathcal{T}_4$	16384	0.015625	24195	97539	220035	391683

Table 2
Number of DOFs for the HHO method with static condensation (SC) and no static condensation (no SC).

	$\mathbb{P}_1$ DOFs		$\mathbb{P}_2$ DOFs		$\mathbb{P}_3$ DOFs		$\mathbb{P}_4$ DOFs	
Mesh	no SC	SC	no SC	SC	no SC	SC	no SC	SC
$\mathcal{T}_0$	752	176	1504	352	2448	528	3584	704
$\mathcal{T}_1$	3040	736	6080	1472	9888	2208	14464	2944
$\mathcal{T}_2$	12224	3008	24448	6016	39744	9024	58112	12032
$\mathcal{T}_3$	49024	12160	98048	24320	159360	36480	232960	48640
$\mathcal{T}_4$	196352	48896	392704	97792	638208	146688	932864	195584

in order to decrease the time required to solve the linear system. A comparison of those times is made in Section 5 for a sequential code. Note however that for a parallelized code, it is much easier to distribute the assembly step than the solve of the linear system.

5. Numerical simulations

This section illustrates numerically our theoretical developments for the contact problem between two membranes proposed in Section 4. We compare the performances of the finite element method (see Section 4.4) and the hybrid high-order method (see Sections 4.6–4.7) for the polynomial degree $p \in \{1, 2, 3, 4\}$.

The problem is written in the unit square domain $\Omega := (0,1)\times(0,1)$. We start the computation with a coarse mesh $\mathcal{T}_0$ containing 64 elements with $h_0 := \max_{K\in\mathcal{T}_0} h_K = 0.25$. We consider four levels of uniform mesh refinement in the sense that the mesh $\mathcal{T}_j$ contains 4^{j+3} triangles for $j \in \{1, 2, 3, 4\}$ and that each element of the mesh $\mathcal{T}_j$ is partitioned by 4 elements in the subsequent mesh $\mathcal{T}_{j+1}$ (see Fig. 2).

In Tables 1 and 2, we compare the degrees of freedom for FEM and HHO. The ones of HHO are considered with and without static condensation. We observe that the HHO method needs the static condensation procedure to be competitive with FEM. Moreover, for low orders, FEM behaves better than HHO. On the contrary, for high order, the HHO method requires fewer degrees of freedom compared to FEM.

A first test case with a very regular solution and Lagrange multiplier aims at estimating the maximum convergence rate of the method. Since in practice the solutions associated to contact problems are not smooth, we consider a second test case with a discontinuous Lagrange multiplier.

5.1. First test case: a smooth solution

We propose an analytical solution to problem (41) given by

$$u_1(r) := -u_2(r) := \begin{cases} (r^2 - R^2)^N & \text{if } r \geqslant R, \\ 0 & \text{otherwise,} \end{cases}$$

$$\lambda(r) := \begin{cases} 0 & \text{if } r \geqslant R, \\ 1000r^3(R^2 - r^2)^3 & \text{otherwise,} \end{cases}$$

where $r := \sqrt{(x-0.5)^2 + (y-0.5)^2}$ is the distance to the center of the domain, $R := 1/3$ is the radius of the disk where contact occurs, and the parameter N is chosen as $N := 6$ to provide a smooth solution. This solution is associated to the right-hand sides f_1 and f_2 defined by

$$f_1(r) := -f_2(r) := \begin{cases} -4N(r^2 - R^2)^{N-2}(Nr^2 - R^2) & \text{if } r \geqslant R, \\ -1000r^3(R^2 - r^2)^3 & \text{otherwise.} \end{cases}$$

We set $\mu_1 := \mu_2 := 1$ for the sake of simplicity and we employ the semismooth Newton linearization of Section 3 with the min function (30) and a tolerance given by $\varepsilon_{\text{lin}} = 10^{-12}$. The solution to the HHO method is obtained using Algorithm 2. For both schemes, the errors are reported in the energy norm

$$|||\boldsymbol{u} - \boldsymbol{u}_h|||_\Omega := \left(\sum_{K\in\mathcal{T}_h} \mu_1 \left\| \boldsymbol{\nabla}(u_1 - u_{1K}) \right\|^2_{L^2(K)} + \mu_2 \left\| \boldsymbol{\nabla}(u_2 - u_{2K}) \right\|^2_{L^2(K)} \right)^{\frac{1}{2}}, \tag{57}$$

i.e. we use only the gradients of the cell unknowns.

Fig. 3 displays the behavior of the solution when the Newton-min solver has converged and when the $\mathbb{P}_2$ FEM discretization is employed. We observe from the shape of the Lagrange multiplier λ_h a contact zone for the two membranes in the area $r \leqslant \frac{1}{3}$. Furthermore, even at convergence, $\lambda_h < 0$ can occur with quadratic FEM, where small

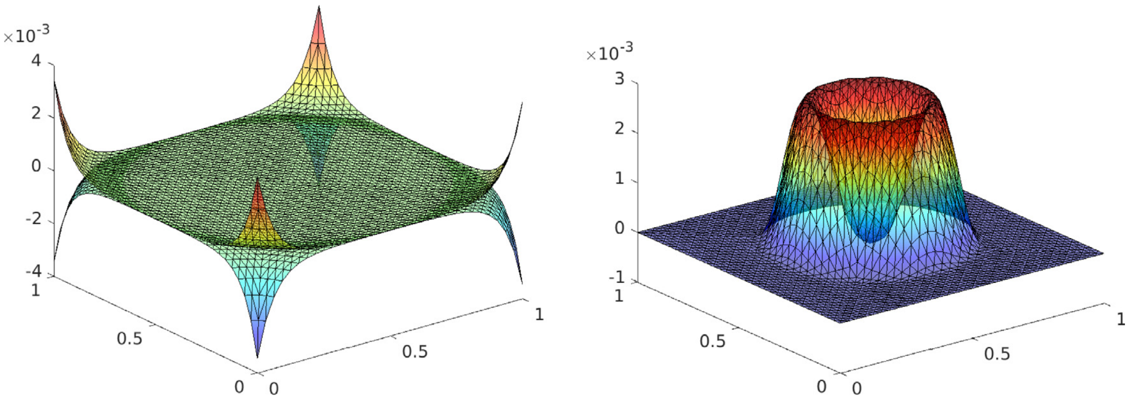

Fig. 3. Solution at convergence for $\mathbb{P}_2$ FEM and mesh $\mathcal{T}_3$. Left: position of the membranes (u_{1h}, u_{2h}). Right: discrete Lagrange multiplier (λ_h).

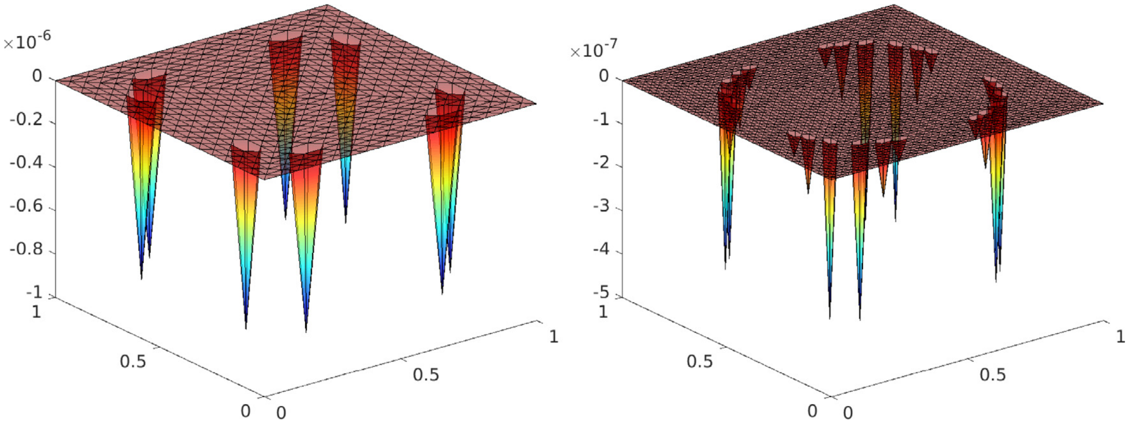

Fig. 4. Negative part of the Lagrange multiplier for FEM $\mathbb{P}_2$. Left: mesh $\mathcal{T}_2$. Right: mesh $\mathcal{T}_3$.

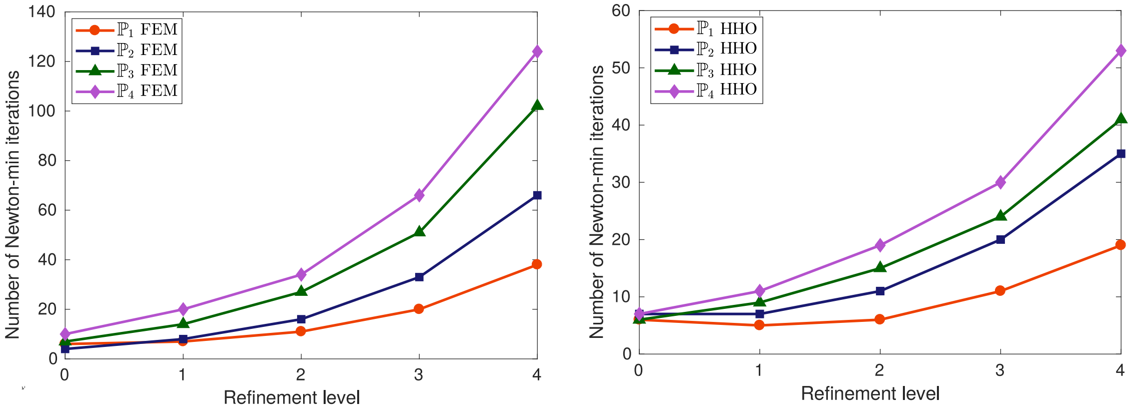

Fig. 5. Number of Newton-min iterations for each refinement level. Left: FEM method. Right: HHO method.

undershoots take place (see Fig. 4). In fact this phenomenon occurs for all $p \geqslant 2$. Note that in Fig. 4, we have represented the function $\lambda_h^{\text{neg}} := \min(\lambda_h, 0)$ at all Lagrange nodes for a better understanding. The discrete Lagrange multiplier λ_h is nonnegative everywhere (here $\lambda_h^{\text{neg}} = 0$) only when $p = 1$.

We report in Fig. 5 the required number of Newton-min iterations needed to reach convergence. We observe that this number increases when the number of mesh elements is increased. Furthermore, we observe that the HHO method with static condensation is less expensive in terms of Newton-min iterations than the FEM method with a gain factor roughly equal to 2.

Fig. 6 displays the shape of the energy norm $\|\!|\boldsymbol{u} - \boldsymbol{u}_h|\!\|_\Omega$ as a function of the refinement level. We get optimal convergence rate (i.e. roughly p) for $p \in \{1, 2, 3\}$. For $p = 4$ we observe a slower convergence rate (about 1) for the two schemes. This can be explained by the fact that the discrete convex set $\mathcal{K}_h^g$ is nonconforming with its continuous analogue $\mathcal{K}^g$. A full a priori analysis including consistency is required to have a better understanding of this problem. It will be explored in a future work.

The static condensation procedure is an important element of the HHO method. By using it, we expect to save some computation time. In Table 3, we report the assembly and solve time used by the HHO method. We consider CPU times with and without static condensation. The assembly time includes the computation of all the matrices of Sections 4.6–4.7 and the solve time includes all the time spent to solve the linear systems. We observe that the static condensation procedure

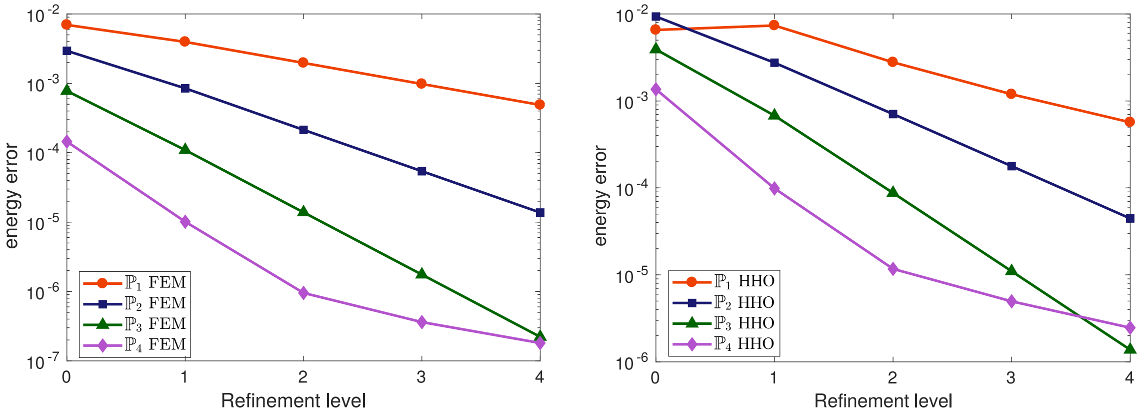

Fig. 6. Energy norm error for each refinement level. Left: FEM method. Right: HHO method.

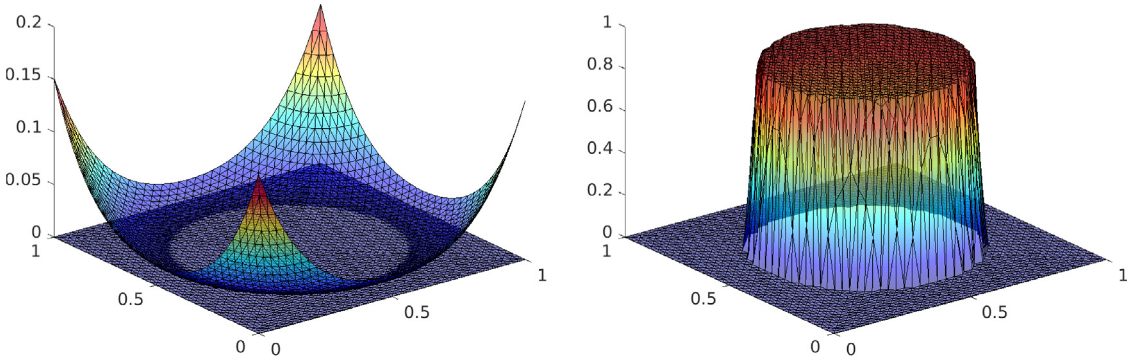

Fig. 7. Solution at convergence for $\mathbb{P}_2$ FEM and mesh $\mathcal{T}_3$. Left: position of the membranes (u_{1h}, u_{2h}). Right: discrete Lagrange multiplier (λ_h).

Table 3
Computation time for HHO (in seconds) for $p = 3$ with static condensation (SC) and without static condensation (no SC).

Mesh	HHO (no SC)			HHO (SC)		
	Assembly	Linear solve	Total	Assembly	Linear solve	Total
$\mathcal{T}_0$	2.87	0.40	3.27	3.09	0.086	3.18
$\mathcal{T}_1$	12.0	2.39	14.4	12.5	0.58	13.1
$\mathcal{T}_2$	76.9	19.4	96.3	73.6	4.26	77.9
$\mathcal{T}_3$	424	103	527	445	30	475
$\mathcal{T}_4$	2940	739	3679	2945	213	3158

drastically diminishes the time needed to solve the linear systems especially for refined meshes. Moreover, the overcost needed to compute the matrices (55)–(56) is negligible. We then recommend in practice the use of static condensation.

5.2. *A test case with a jump for the multiplier*

In practice, the Lagrange multiplier often presents discontinuities. In this second test case, we depict the lack of efficiency of high-order methods in such situations. We consider the following analytical solution,

$$u_1(r) := \begin{cases} 0 & \text{if } r \leqslant R, \\ (r^2 - R^2)^2 & \text{if } r > R, \end{cases} \quad u_2(r) := 0, \quad \lambda(r) := \begin{cases} 8R^2 & \text{if } r \leqslant R, \\ 0 & \text{if } r > R, \end{cases}$$

where $r^2 := (x-0.5)^2 + (y-0.5)^2$. This triple is the solution of (41) for the data f_1 and f_2 given by

$$f_1(r) := \begin{cases} -8R^2 & \text{if } r \leqslant R, \\ 8R^2 - 16r^2 & \text{if } r > R, \end{cases} \quad f_2(r) := \begin{cases} 8R^2 & \text{if } r \leqslant R, \\ 0 & \text{if } r > R. \end{cases}$$

Once more, we use $\mu_1 := \mu_2 := 1$, $\varepsilon_{\text{lin}} := 10^{-12}$ and $R := 1/3$. In Fig. 7 we represent the shape of the discrete solution for the $\mathbb{P}_2$ FEM discretization and for the mesh $\mathcal{T}_3$. This time, we obtain a nonnegative discrete Lagrange multiplier λ_h in the domain Ω and a contact zone in the area $r \leqslant R$.

The number of required Newton-min iterations to reach convergence is reported in Fig. 8. We observe that the HHO resolution with static condensation is faster than the classical FEM resolution and the gain factor in terms of Newton-min iterations is roughly equal to 2.5.

We reported in Fig. 9 the energy error of the two schemes. For $p = 1$, we observe similar results than the ones obtained in Section 5.1. For $p = 2, 3$, the solution converges with a reduced rate (about 1.5). This is a consequence of the fact that the solution is less regular compared with the one used in Section 5.1. For $p = 4$, the method converges with a rate equal to 1 which agrees with the results of Section 5.1.

6. Conclusion

In this work, we presented a unified framework to study the numerical approximation of several variational inequalities. We proposed to discretize a mixed formulation associated to this framework and to use a semismooth Newton algorithm to solve the arising nonlinear system. We proved local convergence properties for this algorithm.

This framework was then applied to compare the behavior of the finite element method (FEM) and the hybrid high-order (HHO) method on the elliptic contact problem between two membranes. A static condensation procedure was given to reduce the size of the system arising for HHO. We considered two test cases: one with a very smooth reference solution and another more realistic with a discontinuous Lagrange multiplier. We observed that the HHO method is faster in terms of semismooth Newton iterations and that the scheme converges with optimal rate for orders $p = 1, 2, 3$ but not for order $p = 4$ even for smooth solutions. This can be due to the nonconforming cones we considered. We think that the HHO method is a viable alternative to FEM if static condensation is used.

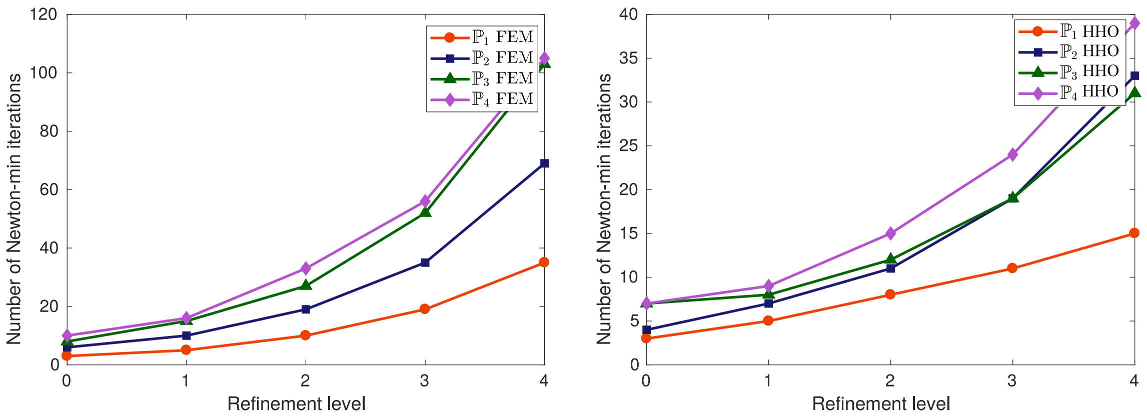

Fig. 8. Number of Newton-min iterations for each refinement level. Left: FEM method. Right: HHO method.

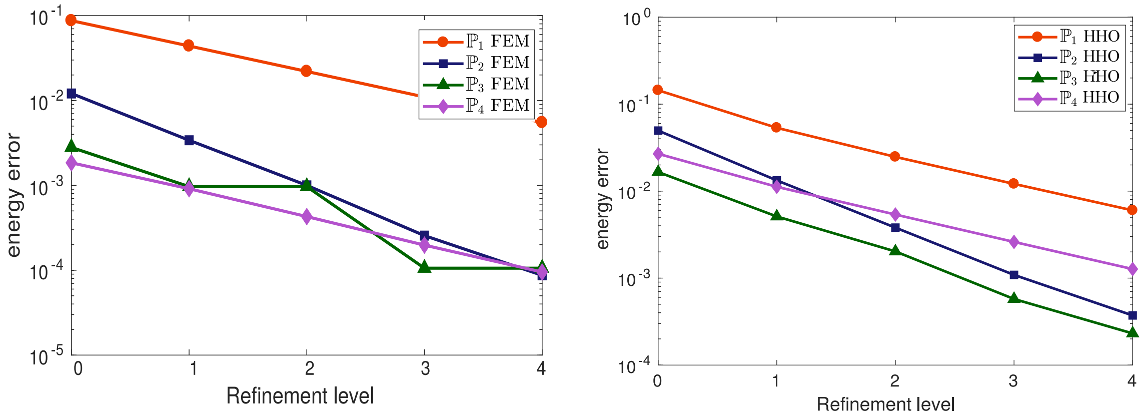

Fig. 9. Energy norm error for each refinement level. Left: FEM method. Right: HHO method.

References

[1] J.-L. Lions, G. Stampacchia, Variational inequalities, Comm. Pure Appl. Math. 20 (1967) 493–519, http://dx.doi.org/10.1002/cpa.3160200302.

[2] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, in: Universitext, Springer, New York, 2011, p. xiv+599.

[3] R. Glowinski, J.-L. Lions, R. Trémolières, Numerical Analysis of Variational Inequalities, in: Studies in Mathematics and its Applications, vol. 8, North-Holland Publishing Co., Amsterdam-New York, 1981, p. xxix+776, Translated from the French.

[4] J.-F. Rodrigues, Obstacle Problems in Mathematical Physics, in: North-Holland Mathematics Studies, vol. 134, North-Holland Publishing Co., Amsterdam, 1987, p. xvi+352, Notas de Matemática [Mathematical Notes], 114.

[5] I. Hlaváček, J. Haslinger, J. Nečas, J. Lovíšek, Solution of Variational Inequalities in Mechanics, in: Applied Mathematical Sciences, vol. 66, Springer-Verlag, New York, 1988, p. x+275, http://dx.doi.org/10.1007/978-1-4612-1048-1, Translated from the Slovak by J. Jarník.

[6] Z. Belhachmi, F. Ben Belgacem, Quadratic finite element approximation of the Signorini problem, Math. Comp. 72 (241) (2003) 83–104, http://dx.doi.org/10.1090/S0025-5718-01-01413-2.

[7] F. Ben Belgacem, C. Bernardi, A. Blouza, M. Vohralík, A finite element discretization of the contact between two membranes, M2AN Math. Model. Numer. Anal. 43 (1) (2009) 33–52, http://dx.doi.org/10.1051/m2an/2008041.

[8] F. Ben Belgacem, C. Bernardi, A. Blouza, M. Vohralík, On the unilateral contact between membranes. I. Finite element discretization and mixed reformulation, Math. Model. Nat. Phenom. 4 (1) (2009) 21–43, http://dx.doi.org/10.1051/mmnp/20094102.

[9] F. Chouly, P. Hild, On convergence of the penalty method for unilateral contact problems, Appl. Numer. Math. 65 (2013) 27–40, http://dx.doi.org/10.1016/j.apnum.2012.10.003.

[10] R.S. Falk, Error estimates for the approximation of a class of variational inequalities, Math. Comp. 28 (1974) 963–971, http://dx.doi.org/10.2307/2005358, URL https://www.jstor.org/stable/2005358?seq=1.

[11] F. Brezzi, W.W. Hager, P.-A. Raviart, Error estimates for the finite element solution of variational inequalities, Numer. Math. 28 (4) (1977) 431–443, http://dx.doi.org/10.1007/BF01404345.

[12] F. Brezzi, W.W. Hager, P.-A. Raviart, Error estimates for the finite element solution of variational inequalities. II. Mixed methods, Numer. Math. 31 (1) (1978) 1–16, http://dx.doi.org/10.1007/BF01396010.

[13] M. Ainsworth, J.T. Oden, C.-Y. Lee, Local a posteriori error estimators for variational inequalities, Numer. Methods Partial Differential Equations 9 (1) (1993) 23–33, http://dx.doi.org/10.1002/num.1690090104.

[14] R. Kornhuber, A posteriori error estimates for elliptic variational inequalities, Comput. Math. Appl. 31 (8) (1996) 49–60, http://dx.doi.org/10.1016/0898-1221(96)00030-2.

[15] Z. Chen, R.H. Nochetto, Residual type a posteriori error estimates for elliptic obstacle problems, Numer. Math. 84 (4) (2000) 527–548, http://dx.doi.org/10.1007/s002110050009.

[16] A. Veeser, Efficient and reliable a posteriori error estimators for elliptic obstacle problems, SIAM J. Numer. Anal. 39 (1) (2001) 146–167, http://dx.doi.org/10.1137/S0036142900370812.

[17] M. Bürg, A. Schröder, A posteriori error control of *hp*-finite elements for variational inequalities of the first and second kind, Comput. Math. Appl. 70 (12) (2015) 2783–2802, http://dx.doi.org/10.1016/j.camwa.2015.08.031.

[18] T. Gudi, K. Porwal, A posteriori error control of discontinuous Galerkin methods for elliptic obstacle problems, Math. Comp. 83 (286) (2014) 579–602, http://dx.doi.org/10.1090/S0025-5718-2013-02728-7.

[19] T. Gudi, K. Porwal, A remark on the a posteriori error analysis of discontinuous Galerkin methods for the obstacle problem, Comput. Methods Appl. Math. 14 (1) (2014) 71–87, http://dx.doi.org/10.1515/cmam-2013-0015.

[20] F. Wang, W. Han, X.-L. Cheng, Discontinuous Galerkin methods for solving elliptic variational inequalities, SIAM J. Numer. Anal. 48 (2) (2010) 708–733, http://dx.doi.org/10.1137/09075891X.

[21] M. Cicuttin, A. Ern, T. Gudi, Hybrid high-order methods for the elliptic obstacle problem, J. Sci. Comput. 83 (1) (2020) http://dx.doi.org/10.1007/s10915-020-01195-z, Paper No. 8, 18.

[22] R. Herbin, E. Marchand, Finite volume approximation of a class of variational inequalities, IMA J. Numer. Anal. 21 (2) (2001) 553–585, http://dx.doi.org/10.1093/imanum/21.2.553.

[23] K.L. Cascavita, F. Chouly, A. Ern, Hybrid high-order discretizations combined with Nitsche's method for Dirichlet and Signorini boundary conditions, HAL Preprint 02016378, 2019, URL https://hal.archives-ouvertes.fr/hal-02016378.

[24] F. Ben Belgacem, C. Bernardi, A. Blouza, M. Vohralík, On the unilateral contact between membranes. Part 2: *a posteriori* analysis and numerical experiments, IMA J. Numer. Anal. 32 (3) (2012) 1147–1172, http://dx.doi.org/10.1093/imanum/drr003.

[25] J. Dabaghi, V. Martin, M. Vohralík, Adaptive Inexact semismooth Newton methods for the contact problem between two membranes, J. Sci. Comput. 84 (2) (2020) 28, http://dx.doi.org/10.1007/s10915-020-01264-3.

[26] A. Schröder, A posteriori error estimates of higher-order finite elements for frictional contact problems, Comput. Methods Appl. Mech. Engrg. 249/252 (2012) 151–157, http://dx.doi.org/10.1016/j.cma.2012.02.001.

[27] F. Chouly, M. Fabre, P. Hild, J. Pousin, Y. Renard, Residual-based *a posteriori* error estimation for contact problems approximated by Nitsche's method, IMA J. Numer. Anal. 38 (2) (2018) 921–954, http://dx.doi.org/10.1093/imanum/drx024.

[28] T. Gudi, K. Porwal, A posteriori error estimates of discontinuous Galerkin methods for the Signorini problem, J. Comput. Appl. Math. 292 (2016) 257–278, http://dx.doi.org/10.1016/j.cam.2015.07.008.

[29] F. Chouly, A. Ern, N. Pignet, A hybrid high-order discretization combined with Nitsche's method for contact and Tresca friction in small strain elasticity, SIAM J. Sci. Comput. 42 (4) (2020) A2300–A2324, http://dx.doi.org/10.1137/19M1286499.

[30] L.-h. Wang, On the quadratic finite element approximation to the obstacle problem, Numer. Math. 92 (4) (2002) 771–778, http://dx.doi.org/10.1007/s002110100368.

[31] S.J. Wright, Primal-Dual Interior-Point Methods, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997, p. xx+289, http://dx.doi.org/10.1137/1.9781611971453.

[32] C. Kanzow, An active set-type Newton method for constrained nonlinear systems, in: Complementarity: Applications, Algorithms and Extensions (Madison, WI, 1999), in: Appl. Optim, vol. 50, Kluwer Acad. Publ., Dordrecht, 2001, pp. 179–200, http://dx.doi.org/10.1007/978-1-4757-3279-5_9.

[33] M. Hintermüller, K. Ito, K. Kunisch, The primal-dual active set strategy as a semismooth Newton method, SIAM J. Optim. 13 (3) (2002) 865–888 (2003), http://dx.doi.org/10.1137/S1052623401383558.

[34] S. Hüeber, G. Stadler, B.I. Wohlmuth, A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction, SIAM J. Sci. Comput. 30 (2) (2008) 572–596, http://dx.doi.org/10.1137/060671061, URL https://epubs.siam.org/doi/10.1137/060671061.

[35] K. Ito, K. Kunisch, Semi-smooth Newton methods for variational inequalities of the first kind, M2AN Math. Model. Numer. Anal. 37 (1) (2003) 41–62, http://dx.doi.org/10.1051/m2an:2003021.

[36] F. Facchinei, J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. I, Springer-Verlag, New York, 2003, Springer Series in Operations Research.

[37] F. Facchinei, J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. II, in: Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[38] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces, in: MOS-SIAM Series on Optimization, vol. 11, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011, http://dx.doi.org/10.1137/1.9781611970692.

[39] I. Ben Gharbia, J.C. Gilbert, Nonconvergence of the plain Newton-min algorithm for linear complementarity problems with a P-matrix, Math. Program. 134 (2, Ser. A) (2012) 349–364, http://dx.doi.org/10.1007/s10107-010-0439-6.

[40] J.F. Bonnans, J.C. Gilbert, C. Lemaréchal, C.A. Sagastizábal, Numerical Optimization, second ed., in: Universitext, Springer-Verlag, Berlin, 2006, p. xiv+490, Theoretical and practical aspects.

[41] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, SIAM J. Numer. Anal. 19 (2) (1982) 400–408, http://dx.doi.org/10.1137/0719025.

[42] S.C. Eisenstat, H.F. Walker, Globally convergent inexact Newton methods, SIAM J. Optim. 4 (2) (1994) 393–422, http://dx.doi.org/10.1137/0804022.

[43] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, in: Frontiers in Applied Mathematics, vol. 16, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995, p. iv+165, http://dx.doi.org/10.1137/1.9781611970944, With separately available software.

[44] S.C. Eisenstat, H.F. Walker, Choosing the forcing terms in an inexact Newton method, SIAM J. Sci. Comput. 17 (1) (1996) 16–32, http://dx.doi.org/10.1137/0917003, Special issue on iterative methods in numerical linear algebra (Breckenridge, CO, 1994).

[45] D.N. Arnold, An interior penalty finite element method with discontinuous elements, SIAM J. Numer. Anal. 19 (4) (1982) 742–760, http://dx.doi.org/10.1137/0719052.

[46] D.A. Di Pietro, A. Ern, Mathematical Aspects of Discontinuous Galerkin Methods, in: Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 69, Springer, Heidelberg, 2012, p. xviii+384, http://dx.doi.org/10.1007/978-3-642-22980-0.

[47] B. Rivière, M.F. Wheeler, V. Girault, A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems, SIAM J. Numer. Anal. 39 (3) (2001) 902–931, http://dx.doi.org/10.1137/S003614290037174X.

[48] D.A. Di Pietro, A. Ern, A hybrid high-order locking-free method for linear elasticity on general meshes, Comput. Methods Appl. Mech. Engrg. 283 (2015) 1–21, http://dx.doi.org/10.1016/j.cma.2014.09.009, URL https://www.sciencedirect.com/science/article/pii/S0045782514003181?via%3Dihub.

[49] D.A. Di Pietro, A. Ern, S. Lemaire, An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators, Comput. Methods Appl. Math. 14 (4) (2014) 461–472, http://dx.doi.org/10.1515/cmam-2014-0018, URL https://www.degruyter.com/view/journals/cmam/14/4/article-p461.xml.

[50] B. Cockburn, D.A. Di Pietro, A. Ern, Bridging the hybrid high-order and hybridizable discontinuous Galerkin methods, ESAIM Math. Model. Numer. Anal. 50 (3) (2016) 635–650, http://dx.doi.org/10.1051/m2an/2015051, URL https://www.esaim-m2an.org/articles/m2an/abs/2016/03/m2an150026/m2an150026.html.

[51] E. Burman, M. Cicuttin, G. Delay, A. Ern, An unfitted hybrid high-order method with cell agglomeration for elliptic interface problems, SIAM J. Sci. Comput. 43 (2) (2021) 859–882, http://dx.doi.org/10.1137/19M1285901.