SIAM J. NUMER. ANAL.
Vol. 47, No. 2, pp. 1126–1148

NUMERICAL ANALYSIS OF A FINITE ELEMENT/VOLUME PENALTY METHOD*

BERTRAND MAURY†

Abstract. We present here some contributions to the numerical analysis of the penalty method in the finite element context. We are especially interested in the ability provided by this approach to use Cartesian, non boundary-fitted meshes to solve elliptic problems in complicated domain. In the spirit of fictitious domains, the initial problem is replaced by a penalized one, posed over a simply shaped domain which covers the original one. This method relies on two parameters, namely h (space-discretization parameter) and ε (penalty parameter). We propose here a general strategy to estimate the error in both parameters, and we present how it can be applied to various situations. We pay special attention to a scalar version of the rigid motion constraint for fluid-particle flows.

Key words. finite element method, penalty, Poisson's problem, error estimate

AMS subject classifications. 65N30, 65N12, 49M30

DOI. 10.1137/080712799

1. Introduction.

Because of its conceptual simplicity and the fact that it is straightforward to implement, the penalty method has been widely used to incorporate constraints in numerical optimization. The general principle can been seen as a relaxed version of the following fact: given a proper functional J over a set X, and K a subset of X, minimizing J over K is equivalent to minimizing $J_K = J + I_K$ over X, where I_K is the indicatrix of K:

$$I_K(x) = \left| \begin{array}{lll} 0 & \text{if} & x \in K, \\ +\infty & \text{if} & x \notin K. \end{array} \right.$$

Assume now that K is defined as $K = \{x \in X\,,\ \Psi(x) = 0\}$, where Ψ is a nonnegative function; the penalty method consists in considering relaxed functionals J_ε defined as

$$J_\varepsilon = J + \frac{1}{\varepsilon}\Psi\,, \qquad \varepsilon > 0.$$

By definition of K, the function Ψ/ε approaches I_K pointwisely:

$$\frac{1}{\varepsilon}\Psi(x) \longrightarrow I_K(x) \text{ as } \varepsilon \text{ goes to } 0 \quad \forall x \in X.$$

If J_ε admits a minimum u^ε, for any ε, one can expect u^ε to approach a (or *the*) minimizer of J over K, if it exists.

In the finite element context, some u_h^ε is computed as the solution to a finite dimensional problem, where h is a space-discretization parameter. The work we present here is motivated by the fact that, even if the penalty method for the continuous problem is convergent and the discretization procedure is sound, the rate of convergence of u_h^ε toward the exact solution is not straightforward to obtain. A huge literature is

*Received by the editors January 9, 2008; accepted for publication (in revised form) November 6, 2008; published electronically February 19, 2009.
http://www.siam.org/journals/sinum/47-2/71279.html
†Laboratoire de Mathématiques, Université Paris-Sud, 91405 Orsay Cedex, France (Bertrand.Maury@math.u-psud.fr).

dedicated to the situation where the constraint is distributed over the domain, like the divergence-free constraint for incompressible Stokes flows (see [BF91, GR79]). In this context, the penalty approach makes it possible to use mixed finite element methods which do not fulfill the so-called Babuska–Brezzi–Ladyzhenskaya (or inf-sup) condition. The penalty approach is also commonly used to prescribe (possibly nonhomogeneous) Dirichlet boundary conditions on a boundary. The pioneering papers [Nit71] and [Bab73] already addressed in the early 70's the problem of error estimation with respect to both parameters h and ε. Those works have been widely used since then, and this area has recently experienced a regain of interest, triggered by problems arising in domain decomposition (see, e.g., [BHS03]), discontinuous Galerkin methods [BE07], or handling of discontinuities for elliptic problems with discontinuous coefficients [HH02].

We will focus here on another type of constraints, namely geometrical ones: we are interested in solving an elliptic problem on a domain $\Omega \setminus \overline{\mathcal{O}}$, where Ω is a simply shaped domain (e.g., a rectangle) and $\mathcal{O}$ a set of holes, and we aim at replacing it by a new problem posed over the global domain Ω. The simplest situation one may consider consists in solving a Poisson problem in a perforated, rectangular domain Ω, with homogeneous Dirichlet boundary conditions on the holes and over the external boundary. In the purpose of using a Cartesian mesh which covers the whole domain (which can be of great interest if the holes are intended to move), it is natural to consider the penalized version of the problem, which consists in minimizing ($\mathcal{O}$ designs the subdomain covered by the holes)

$$\frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv + \frac{1}{2\varepsilon}\int_{\mathcal{O}} \left(v^2 + |\nabla v|^2\right)$$

over $H_0^1(\Omega)$. Another situation where the penalty approach has already proved to be quite efficient is the modeling of fluid-particle flows (see [RPVC05] or [JLM05]). The scalar version of this problem, which we shall address in detail in the following pages, consists in minimizing the standard functional

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 - \int_\Omega fv$$

over all those functions which are constant on each connected component of the set of holes $\mathcal{O}$. Again, the constraint is easily relaxed by adding to J a term which penalizes the H^1 seminorm of v over $\mathcal{O}$.

Two points advocate for the use of this approach:

1. The use of a Cartesian mesh makes this approach quite easy to implement: both cases reduce to a few lines of instructions within user-friendly finite element solvers like Freefem++ [FFp] for two-dimensional problems, or Freefem3D [FFp] for three-dimensional ones. Note that the penalty terms do not preserve the spectrum of the discrete Laplacian matrix, which prevents us from using standard fast solvers like fast Fourier transform (to the contrary of Lagrange multiplier based fictitious domain methods [PG02, GG95], which do preserve the structure of the matrix, at the price of an iterative algorithm on the Lagrange multipliers). A harmful effect upon the condition number of the solution matrix is furthermore to be expected. Yet, as the penalty parameter does not need to be taken too small, the method remains quite competitive for reasonably sized problems.

2. This method provides, with no extra computational cost, an approximation of the Lagrange multiplier associated with the constraint, which is of great significance from the modeling standpoint in many situations. For example, in the first situation we considered, which can be seen as the stationary heat equation, it is quite straightforward that, if we denote by u^ε the solution to the discretized problem, $\xi^\varepsilon \in H^{-1}$ defined as

$$\langle \xi^\varepsilon \, , v\rangle = \frac{1}{\varepsilon}\int_{\mathcal{O}} (u^\varepsilon v + \nabla u^\varepsilon \cdot \nabla v)$$

approximates the heat source which is necessary to fulfill the constraint. We shall establish that this natural outcome of the method is still provided by the discretized/penalized version. Note that this property has already been used to handle numerically the motion of a three-dimensional turbine in a Navier–Stokes fluid (see [DPM07]).

As for the theoretical analysis of the method, the error due to the fact that the mesh is not boundary fitted is analyzed in [AR08, RAB07]. See also [SMSTT05] for similar estimates used to establish the convergence of a method to handle the motion of a rigid motion in the limit $\varepsilon = 0$. Yet, to the best of our knowledge, a full error estimate (simultaneous convergence of h and ε toward 0) has not yet been provided for the type of volume penalty approach we propose here. We aim here at showing that the global error can be controlled, as expected, by the sum of the penalty error and the space-discretization error, under quite general assumptions.

This paper is organized as follows: in section 2, we recall some standard properties of the penalty method in the framework of constrained quadratic minimization, including some general facts about the space discretization of those problems. Section 3 is devoted to the main result: an abstract estimate for the primal and the dual parts of the discretized/penalized problem. The next section is concerned with a model problem, in the spirit of fluid-particle flows, for which we present in detail how the abstract estimate can be applied. Finally, we present in section 5 some other typical situations where the abstract estimate can be used.

2. Preliminaries, abstract framework.

2.1. Continuous problem. We recall here some standard properties concerning the penalty method applied to infinite dimensional problems. Most of those properties are established in [BF91], with a slightly different formalism. We consider the following set of assumptions:

$$\left|\begin{array}{l} V \text{ is a Hilbert space}, \varphi \in V', \\ a(\cdot,\cdot) \text{ bilinear, symmetric, continuous, elliptic } (a(v,v) \geq \alpha\,|v|^2), \\ b(\cdot,\cdot) \text{ bilinear, symmetric, continuous, nonnegative}, \\ K = \{u \in V \, , \; b(u,u) = 0\} = \ker b, \\ J(v) = \dfrac{1}{2}a(v,v) - \langle \varphi \, , v\rangle, \quad u = \arg\min_K J, \\ J_\varepsilon(v) = \dfrac{1}{2}a(v,v) + \dfrac{1}{2\varepsilon}b(v,v) - \langle \varphi \, , v\rangle, \quad u^\varepsilon = \arg\min_V J_\varepsilon. \end{array}\right. \tag{2.1}$$

PROPOSITION 2.1. *Under assumptions* (2.1), *the solution* u^ε *to the penalized problem converges to* u.

Proof. As the family (J_ε) is uniformly elliptic, $|u^\varepsilon|$ is bounded. We extract a subsequence, still denoted by (u^ε), which converges weakly to some $z \in V$. As $J_\varepsilon \geq J$ and $b(u,u) = 0$, we have

$$J(u^\varepsilon) \leq J_\varepsilon(u^\varepsilon) \leq J_\varepsilon(u) = J(u) \quad \forall \varepsilon > 0, \tag{2.2}$$

so that (J is convex and continuous) $J(z) \leq \liminf J(u^\varepsilon) \leq J(u)$. As

$$J(u^\varepsilon) + \frac{1}{2\varepsilon} b(u^\varepsilon, u^\varepsilon) \leq J(u),$$

$b(u^\varepsilon, u^\varepsilon)/\varepsilon$ is bounded, so that $b(u^\varepsilon, u^\varepsilon)$ goes to 0 with ε. Consequently, it holds that $0 \leq b(z,z) \leq \liminf b(u^\varepsilon, u^\varepsilon) = 0$, which implies $z \in K$, so that $z = u$.

To establish the strong character of the convergence, we show that u^ε converges toward u for the norm associated with $a(\cdot,\cdot)$, which is equivalent to the original norm. As u^ε converges weakly to u for this scalar product ($a(u^\varepsilon, v) \to a(u,v)$ for any $v \in V$), it is sufficient to establish the convergence of $|u^\varepsilon|_a = a(u^\varepsilon, u^\varepsilon)^{1/2}$ toward $|u|_a$. First, $|u|_a \leq \liminf |u^\varepsilon|_a$, and the other inequality comes from (2.2):

$$\frac{1}{2} a(u^\varepsilon, u^\varepsilon) - \langle \varphi \, , u^\varepsilon \rangle \leq \frac{1}{2} a(u,u) - \langle \varphi \, , u \rangle,$$

so that $\limsup |u^\varepsilon|_a \leq |u|_a$. ∎

The proposition does not say anything about the rate of convergence, and it can be very poor, as the following example illustrates.

Example 2.1. Consider $I =]0,1[$, $V = H^1(I)$, and the problem which consists in minimizing the functional

$$J(v) = \frac{1}{2} \int_I |u'|^2$$

over $K = \{ v \in V \, , \ v(x) = 0 \text{ a.e. in } \mathcal{O} =]0, 1/2[\}$. The solution to that problem is obviously $u = \max(0, 2(x - 1/2))$. Now let us denote by u^ε the minimum of the penalized functional

$$J_\varepsilon = \frac{1}{2} \int_I |u'|^2 + \frac{1}{2\varepsilon} \int_{\mathcal{O}} |u|^2 .$$

The solution to the penalized problem can be computed exactly:

$$u^\varepsilon = k_\varepsilon(x) \operatorname{sh}\left(\frac{x}{\sqrt{\varepsilon}} \right) \text{ in }]0, 1/2[\text{ with } k_\varepsilon(x) = \left(\operatorname{sh}\left(\frac{x}{\sqrt{\varepsilon}} \right) + \frac{1}{2\sqrt{\varepsilon}} \operatorname{ch}\left(\frac{x}{\sqrt{\varepsilon}} \right) \right)^{-1},$$

and u^ε affine in $]1/2, 1[$, continuous at $1/2$. This makes it possible to estimate $|u^\varepsilon - u|$, which turns out to behave like $\varepsilon^{1/4}$.

Yet, in many situations, convergence can be shown to be of order 1, given some assumptions are fulfilled. Let us introduce $\xi \in V'$ as the unique linear functional such that

$$a(u,v) + \langle \xi \, , v \rangle = \langle \varphi \, , v \rangle \quad \forall v \in V. \tag{2.3}$$

Before stating the first order convergence result, we show here that the penalty method provides an approximation of ξ.

PROPOSITION 2.2. *Let $\xi^\varepsilon \in V'$ be defined by*

$$v \in V \longmapsto \langle \xi^\varepsilon \, , v\rangle = \frac{1}{\varepsilon} b(u^\varepsilon, v).$$

Then ξ^ε converges (strongly) to ξ in V', at least as fast as u^ε converges to u.

Proof. The variational formulation of the penalized problem reads

$$a(u^\varepsilon, v) + \frac{1}{\varepsilon} b(u^\varepsilon, v) = \langle \varphi \, , v\rangle \quad \forall v \in V. \tag{2.4}$$

The result is then a direct consequence of the identity which we obtain by substracting (2.3) and (2.4):

$$\langle \xi \, , v\rangle - \frac{1}{\varepsilon} b(u^\varepsilon, v) = a(u - u^\varepsilon, v) \quad \forall v \in V,$$

which yields $\|\xi - \xi^\varepsilon\|_{V'} \leq C\,|u - u^\varepsilon|$. ∎

Let us now establish the first order convergence, provided an extra compatibility condition between $b(\cdot,\cdot)$ and ξ is met.

PROPOSITION 2.3. *Under assumptions* (2.1)*, we assume in addition that there exists $\tilde{\xi} \in V$ such that $b(\tilde{\xi}, v) = \langle \xi \, , v\rangle$ for all $v \in V$. Then $|u^\varepsilon - u| = \mathcal{O}(\varepsilon)$.*

Proof. First of all, notice that it is possible to pick $\tilde{\xi}$ in $K^\perp$ (if not, we project it onto $K^\perp$). Now following the idea which is proposed in [Bab73] in a slightly different context (see the proof of Thm. 3.2 therein), we introduce

$$R_\varepsilon(v) = \frac{1}{2} a(u - v, u - v) + \frac{1}{2\varepsilon} b(\varepsilon\tilde{\xi} - v, \varepsilon\tilde{\xi} - v),$$

which can be written

$$R_\varepsilon(v) = \frac{1}{2} a(u,u) + \frac{\varepsilon}{2} b(\tilde{\xi}, \tilde{\xi}) + \frac{1}{2} a(v,v) + \frac{1}{2\varepsilon} b(v,v) - a(u,v) - b(\tilde{\xi}, v).$$

As $b(\tilde{\xi}, v) = \langle \xi \, , v\rangle$ and $-a(u,v) - \langle \xi \, , v\rangle = -\langle \varphi \, , v\rangle$, the functional R_ε is equal to J_ε up to a constant. Therefore minimizing R_ε amounts to minimizing J_ε. Let us now introduce $w = \varepsilon\tilde{\xi} + u$. We have

$$R_\varepsilon(w) = \frac{\varepsilon^2}{2} a(\tilde{\xi}, \tilde{\xi}) + 0 \quad \text{because } u \in K = \ker b,$$

so that $|R_\varepsilon(w)| \leq C\varepsilon^2$. As u^ε minimizes R_ε,

$$0 \leq R_\varepsilon(u^\varepsilon) = \frac{1}{2} a(u - u^\varepsilon, u - u^\varepsilon) + \frac{1}{2\varepsilon} b(\varepsilon\tilde{\xi} - u^\varepsilon, \varepsilon\tilde{\xi} - u^\varepsilon) \leq C\varepsilon^2,$$

from which we deduce, as $a(\cdot,\cdot)$ is elliptic, $|u - u^\varepsilon| = \mathcal{O}(\varepsilon)$. ∎

COROLLARY 2.4. *Under assumptions* (2.1)*, we assume in addition that $b(\cdot,\cdot)$ can be written $b(u,v) = (Bu, Bv)$, where B is a linear continuous operator onto a Hilbert space Λ, with closed range. Then $|u^\varepsilon - u| = \mathcal{O}(\varepsilon)$.*

Proof. Let us show that the assumption of Proposition 2.3 is met. It is sufficient to prove that any $\xi \in V'$ which vanishes over K identifies through $b(\cdot,\cdot)$ with some $\tilde{\xi} \in V$; i.e., there exists $\tilde{\xi} \in V$ such that

$$\langle \xi \, , v\rangle = b(\tilde{\xi}, v) \quad \forall v \in V.$$

Note that, as ξ vanishes over K, it can be seen as a linear functional defined on $K^\perp$, so that it is equivalent to establish that $T : V \longrightarrow (K^\perp)'$ defined by

$$\tilde{\xi} \longmapsto \xi \ : \ \langle \xi \, , v \rangle = b(\tilde{\xi}, v) \quad \forall v \in K^\perp$$

is surjective. We denote by $T^\star \in \mathcal{L}\,(K^\perp, V)$ the adjoint of T. For all $w \in K^\perp$,

$$|T^\star w| = \sup_{v \neq 0} \frac{(T^\star w, v)}{|v|} = \sup_{v \neq 0} \frac{b(w, v)}{|v|} = \sup_{v \neq 0} \frac{(Bw, Bv)}{|v|} \geq \frac{|Bw|^2}{|w|}.$$

As B has closed range, $|Bw| \geq C\,|w|$ for all w in $(\ker B)^\perp = K^\perp$, so that

$$|T^\star w| \geq C^2\,|w| \quad \forall w \in K^\perp,$$

from which we conclude that T is surjective. □

Remark 2.1. Note that Proposition 2.3 is strictly stronger than its corollary. Indeed, consider the handling of homogeneous Dirichlet boundary conditions by penalty: $V = H^1(\Omega)$, where Ω is a smooth, bounded domain, $a(u, v) = \int \nabla u \cdot \nabla v$, and $\langle \varphi \, , v \rangle = \int f v$, where f is in $L^2(\Omega)$, and $b(v, v) = \int_{\partial\Omega} v^2$. In this situation the corollary cannot be used, because the trace operator from $H^1(\Omega)$ onto $L^2(\partial\Omega)$ does not have a close range. On the other hand one can establish that

$$\langle \xi \, , v \rangle = \int_{\partial\Omega} \frac{\partial u}{\partial n} v,$$

and, as the solution u is regular ($u \in H^2(\Omega)$), its normal derivative (in $H^{1/2}(\partial\Omega)$) can be built as the trace of a function $\tilde{\xi}$ in $H^1(\Omega)$, so that Proposition 2.3 holds true.

We conclude this section by some considerations concerning the saddle-point formulation of the constrained problem, which will be useful in the following. We consider again the closed situation.

PROPOSITION 2.5. *Under the assumptions of Corollary* 2.4*, there exists* $\lambda \in \Lambda$ *such that*

$$a(u, v) + (\lambda, Bv) = \langle \varphi \, , v \rangle \quad \forall v \in V. \tag{2.5}$$

The solution is unique in $B(V)$ *(which identifies with* $\Lambda / \ker B^\star$*).*

Proof. The proof of this standard property can be found in [BF91]. In fact, it has just been established in the proof of Corollary 2.4: λ is simply $B\tilde{\xi}$. Uniqueness is straightforward. □

PROPOSITION 2.6. *Under the assumptions of Proposition* 2.5 *(assumptions* (2.1) *and* $B(V)$ *is closed), we introduce*

$$\lambda^\varepsilon = \frac{1}{\varepsilon} B u^\varepsilon.$$

Then $|\lambda^\varepsilon - \lambda| = \mathcal{O}(\varepsilon)$*, where* λ *is the unique solution of* (2.5) *in* $B(V)$*.*

Proof. Substracting the variational formulations for u and u^ε, we get

$$(\lambda^\varepsilon - \lambda, Bv) = a(u^\varepsilon - u, v) \quad \forall v \in V.$$

Now, as the range of B is closed, and $\lambda^\varepsilon - \lambda \in B(V) = (\ker B^\star)^\perp$, we have the inf-sup condition (see, e.g., [BF91])

$$\sup_{v \in V} \frac{(\lambda^\varepsilon - \lambda, Bv)}{|v|} \geq \beta\,|\lambda^\varepsilon - \lambda|\,,$$

so that

$$\beta\,|\lambda^\varepsilon-\lambda| \le \sup\frac{(\lambda^\varepsilon-\lambda,Bv)}{|v|} = \sup\frac{a(u^\varepsilon-u,v)}{|v|} \le \|a\|\,|u^\varepsilon-u|\,,$$

which ensures the first order convergence thanks to Corollary 2.4. ∎

COROLLARY 2.7. *For any $z\in V$ such that $Bz=\lambda$, there exists a sequence (v^ε) in $\ker B$ such that*

$$\left|\frac{u^\varepsilon}{\varepsilon}-v^\varepsilon-z\right| = \mathcal{O}(\varepsilon).$$

Proof. This is a direct consequence of the fact that, $B(V)$ being closed, the restriction of B to $\ker B^\perp$ is a bicontinuous bijection between $\ker B^\perp$ and $B(V)$. The convergence is therefore obtained by taking $v^\varepsilon = P_{\ker B}(u^\varepsilon/\varepsilon - z)$. ∎

2.2. Discretized problem. We consider now a family $(V_h)_h$ of inner approximation spaces ($V_h\subset V$) and the associated penalized/discretized problems

$$\left\{\begin{array}{l} \text{Find } u_h^\varepsilon\in V_h \text{ such that } J^\varepsilon(u_h^\varepsilon) = \inf\limits_{v_h\in V_h} J^\varepsilon(v_h)\,,\\[2ex] J^\varepsilon(v_h) = \dfrac{1}{2}a(v_h,v_h)+\dfrac{1}{2\varepsilon}b(v_h,v_h)-\langle\varphi\,,v_h\rangle. \end{array}\right. \tag{2.6}$$

As far as we know, there does not exist any general theory which would give an upper bound for the error $|u-u_h^\varepsilon|$ as the sum of a discretization error (typically h of $h^{1/2}$ for volume penalty, depending on whether the mesh is boundary-fitted or not), and a penalty error (typically ε for closed-range penalty terms, possibly poorer in general situations, as in Example 2.1). We propose here two general properties which are direct consequences of standard arguments. They are suboptimal in the sense that neither of them is optimal from both standpoints (discretization and penalty), but, at least, they make it possible to recover the behavior in extreme situations (when ε goes to 0 much quicker than h, and the opposite).

The first proposition uses the following lemma.

LEMMA 2.8. *Under assumptions* (2.1)*, there exists $C>0$ such that*

$$b(u^\varepsilon,u^\varepsilon) \le C\varepsilon\,|u-u^\varepsilon|\,.$$

Proof. By definition of u^ε,

$$J_\varepsilon(u^\varepsilon) = \frac{1}{2}a(u^\varepsilon,u^\varepsilon)-\langle\varphi\,,u^\varepsilon\rangle+\frac{1}{2\varepsilon}b(u^\varepsilon,u^\varepsilon) \le J_\varepsilon(u) = \frac{1}{2}a(u,u)-\langle\varphi\,,u\rangle,$$

so that

$$\begin{aligned} 0 \le \frac{1}{2\varepsilon}b(u_\varepsilon,u_\varepsilon) &\le \frac{1}{2}a(u,u)-\frac{1}{2}a(u^\varepsilon,u^\varepsilon)+\langle\varphi\,,u^\varepsilon-u\rangle\\ &\le \frac{1}{2}a(u+u^\varepsilon,u-u^\varepsilon)+\langle\varphi\,,u^\varepsilon-u\rangle, \end{aligned}$$

which yields the estimate by continuity of $a(\cdot,\cdot)$ and φ. ∎

PROPOSITION 2.9. *Under assumptions* (2.1)*, we denote by u_h^ε the solution to problem* (2.6)*. Then*

$$|u_h^\varepsilon-u| \le C\left(\min_{v_h\in V_h\cap K}|v_h-u|+\sqrt{|u^\varepsilon-u|}\right).$$

Proof. As u_h^ε minimizes $a(v-u^\varepsilon, v-u^\varepsilon) + b(v-u^\varepsilon, v-u^\varepsilon)/\varepsilon$ over V_h,

$$\begin{aligned}
\alpha \left|u_h^\varepsilon - u^\varepsilon\right|^2 &\le a(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) \\
&\le a(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) + \frac{1}{\varepsilon} b(u_h^\varepsilon - u^\varepsilon, u_h^\varepsilon - u^\varepsilon) \\
&\le \min_{v_h \in V_h} \left(a(v_h - u^\varepsilon, v_h - u^\varepsilon) + \frac{1}{\varepsilon} b(v_h - u^\varepsilon, v_h - u^\varepsilon) \right) \\
&\le \min_{v_h \in V_h \cap K} \left(a(v_h - u^\varepsilon, v_h - u^\varepsilon) + \frac{1}{\varepsilon} b(v_h - u^\varepsilon, v_h - u^\varepsilon) \right).
\end{aligned}$$

As v_h is in K, the second term is $b(u^\varepsilon, u^\varepsilon)/\varepsilon$, which is bounded by $C\left|u^\varepsilon - u\right|$ (by Lemma 2.8). Finally, we get

$$\left|u_h^\varepsilon - u^\varepsilon\right| \le C \left(\min_{v_h \in V_h \cap K} \left|v_h - u^\varepsilon\right| + \sqrt{\left|u^\varepsilon - u\right|} \right),$$

from which we conclude. ∎

PROPOSITION 2.10. *Under assumptions* (2.1), $V_h \subset V$, *and* u_h^ε *being the solution to* (2.6), *it holds that*

$$\left|u_h^\varepsilon - u\right| \le \frac{C}{\sqrt{\varepsilon}} \inf_{v_h \in V_h} \left|u^\varepsilon - v_h\right| + \left|u^\varepsilon - u\right|.$$

Proof. One has

$$\left|u_h^\varepsilon - u\right| \le \left|u_h^\varepsilon - u^\varepsilon\right| + \left|u^\varepsilon - u\right|,$$

and we control the first term by Céa's lemma applied to the bilinear form $a + b/\varepsilon$, whose ellipticity constant behaves like $1/\varepsilon$. ∎

The following example illustrates how those estimates can be used in practice.

Example 2.2. The simplest example of penalty formulation one may think about is the following: the constraint to vanish on the boundary of a subdomain $\mathcal{O} \subset\subset \Omega$ is handled by minimizing the functional

$$J_\varepsilon(v) = \frac{1}{2} \int_\Omega |\nabla v|^2 - \int_\Omega f v + \frac{1}{2\varepsilon} \int_{\mathcal{O}} u^2. \tag{2.7}$$

Now considering the L^2 penalty method in $\mathcal{O}$, if we admit the $\varepsilon^{1/4}$ convergence of $|u^\varepsilon - u|$, Proposition 2.9 provides an estimate in $h^{1/2} + \varepsilon^{1/8}$. This estimate is optimal in h: the natural space discretization order is obtained if ε is small enough ($\varepsilon = h^4$ in the present case).

Symmetrically, the natural order in ε can be recovered if h is small enough: Indeed, if we admit that u^ε can be approximated at the same order as u over Ω, which is $1/2$, then the choice $\varepsilon = h^{4/3}$ in Proposition 2.10 gives

$$\left|u_h^\varepsilon - u\right| \le \frac{C}{\varepsilon^{1/2}} \varepsilon^{3/4} + \varepsilon^{1/4} = \mathcal{O}(\varepsilon^{1/4}).$$

Note that if we replace u^2 by $u^2 + |\nabla u|^2$ in the integral over $\mathcal{O}$ in (2.7), assumptions of Corollary 2.4 are fulfilled, so that convergence holds at the first order in ε. As a consequence, $|u - u_h^\varepsilon|$ is bounded by $C(h^{1/2} + \varepsilon^{1/2})$ (by Proposition 2.9), which suggests the choice $\varepsilon = h$.

3. Full error estimate. As shall be made clear below, a full and optimal error estimate calls for a uniform discrete inf-sup condition. In the case of a nonconforming mesh, it appears immediately that the penalty term has to be modified. To anticipate this difficulty, we introduce a modified version of B, namely B_h, in this abstract approach. No assumption is made a priori on B_h in terms of approximation properties, but the estimate we establish below will not express any convergence property unless B_h approaches B in some sense.

Besides (2.1), we consider the following set of additional assumptions and notation:

$$\left|\begin{array}{l} b(v,v) = (Bv, Bv), \text{ where } B \in \mathcal{L}(V,\Lambda) \text{ has a closed range,} \\ (V_h)_h \text{ family of approximation spaces}, \ V_h \subset V, \\ B_h \in \mathcal{L}(V,\Lambda), \ \ker B \subset \ker B_h, \ \|B_h\| \text{ bounded}, \ \Lambda_h = B_h(V_h), \\ J_h^\varepsilon(v_h) = J(v_h) + \dfrac{1}{\varepsilon}(B_h v_h, B_h v_h), \\ u_h^\varepsilon = \arg\min\limits_{V_h} J_h^\varepsilon, \ \lambda_h^\varepsilon = \dfrac{1}{\varepsilon} B_h u_h^\varepsilon \in \Lambda_h, \\ \sup\limits_{v_h \in V_h} \dfrac{(B_h v_h, \lambda_h)}{|v_h|} \geq \beta\, |\lambda_h|_{\Lambda_h} \quad \forall \lambda_h \in \Lambda_h. \end{array}\right. \tag{3.1}$$

THEOREM 3.1 (primal/dual error estimate). *Under assumptions* (2.1) *and* (3.1), *we have the following error estimate:*

$$|u - u_h^\varepsilon| + |\lambda - \lambda_h^\varepsilon| \tag{3.2}$$

$$\leq C\left(\varepsilon + \inf_{\tilde{u}_h \in V_h} |\tilde{u}_h - u| + \inf_{\tilde{\lambda}_h \in \Lambda_h} \left|\tilde{\lambda}_h - \lambda\right| + |(B_h^\star - B^\star)\lambda| + |(B_h - B)z|\right),$$

where z is such that $\lambda = Bz$.

Proof. The proof relies on some general properties of the continuous penalty method which we established in the beginning of this section, and an abstract stability estimate for saddle-point-like problems with stabilization (see Proposition 3.2 below).

First of all, note that, as the range of B is closed, the convergence of u^ε toward u holds at the first order (by Corollary 2.4). As another consequence, $\lambda^\varepsilon = Bu^\varepsilon/\varepsilon$ is such that $|\lambda - \lambda^\varepsilon| = \mathcal{O}(\varepsilon)$ (by Proposition 2.6).

We write the continuous penalized problem

$$\left\{\begin{array}{rcrcll} a(u^\varepsilon, v) & + & (\lambda^\varepsilon, Bv) & = & \langle \varphi\,, v\rangle & \forall v \in V, \\ (Bu^\varepsilon, \mu) & - & \varepsilon(\lambda^\varepsilon, \mu) & = & 0 & \forall \mu \in \Lambda \end{array}\right.$$

and the discrete penalized problem in a saddle-point form

$$\left\{\begin{array}{rcrcll} a(u_h^\varepsilon, v_h) & + & (\lambda_h^\varepsilon, B_h v_h) & = & \langle \varphi\,, v_h\rangle & \forall v_h \in V_h, \\ (B_h u_h^\varepsilon, \mu_h) & - & \varepsilon(\lambda_h^\varepsilon, \mu_h) & = & 0 & \forall \mu_h \in \Lambda_h. \end{array}\right.$$

As Λ_h is exactly $B_h(V_h)$, this problem admits a unique solution $(u_h^\varepsilon, \lambda_h^\varepsilon)$ (see Proposition 2.5). For any $(\tilde{u}_h, \tilde{\lambda}_h) \in V_h \times \Lambda_h$, $v_h \in V_h$, $\mu_h \in \Lambda_h$,

$$\left\{\begin{array}{rcl} a(\tilde{u}_h - u_h^\varepsilon, v_h) + (\tilde{\lambda}_h - \lambda_h^\varepsilon, B_h v_h) & = & a(\tilde{u}_h - u^\varepsilon, v_h) + (\tilde{\lambda}_h - \lambda^\varepsilon, B_h v_h) \\ & & + \langle (B_h^\star - B^\star)\lambda^\varepsilon\,, v_h\rangle, \\ (B_h(\tilde{u}_h - u_h^\varepsilon), \mu_h) - \varepsilon(\tilde{\lambda}_h - \lambda_h^\varepsilon, \mu_h) & = & (B_h(\tilde{u}_h - u^\varepsilon), \mu_h) - \varepsilon(\tilde{\lambda}_h - \lambda^\varepsilon, \mu_h) \\ & & + \langle (B_h - B)u^\varepsilon\,, \mu_h\rangle. \end{array}\right.$$

Our purpose is to use Proposition 3.2 (V_h and Λ_h play the role of V and Λ in the proposition, respectively) with

$$\langle \varphi\,,\, v_h\rangle = a(\tilde{u}_h - u^\varepsilon, v_h) + (\tilde{\lambda}_h - \lambda^\varepsilon, B_h v_h) + \langle (B_h^\star - B^\star)\lambda^\varepsilon\,,\, v_h\rangle, \tag{3.3}$$

$$\langle \Psi\,,\, \mu_h\rangle = (B_h(\tilde{u}_h - u^\varepsilon), \mu_h) - \varepsilon(\tilde{\lambda}_h - \lambda^\varepsilon, \mu_h) + ((B_h - B)u^\varepsilon, \mu_h). \tag{3.4}$$

The last term of (3.3) is transformed as follows:

$$(B_h^\star - B^\star)\lambda^\varepsilon = (B_h^\star - B^\star)\lambda + (B_h^\star - B^\star)(\lambda^\varepsilon - \lambda),$$

where $\lambda \in B(V)$ is the exact Lagrange multiplier defined by Proposition 2.5. So, defining

$$c(\mu, \mu') = \varepsilon(\mu, \mu')\,,\ w = \tilde{u}_h - u^\varepsilon\,,\ \gamma = -(\tilde{\lambda}_h - \lambda^\varepsilon) + (B_h - B)\,\frac{u^\varepsilon}{\varepsilon}$$

(see (3.7) for the meaning of w and γ), Proposition 3.2 ensures existence of a constant $C > 0$ (which does not depend on h) such that $|\tilde{u}_h - u_h^\varepsilon| + |\tilde{\lambda}_h - \lambda_h^\varepsilon|$ is less than

$$C\left(|\tilde{u}_h - u^\varepsilon| + \left|\tilde{\lambda}_h - \lambda^\varepsilon\right| + \|(B_h^\star - B^\star)\lambda\| + |\gamma|\right).$$

The second contribution to γ can be written, thanks to Corollary 2.7 and the fact that $\ker B \subset \ker B_h$,

$$(B_h - B)\,\frac{u^\varepsilon}{\varepsilon} = (B_h - B)\left(\frac{u^\varepsilon}{\varepsilon} - v^\varepsilon - z\right) + (B_h - B)\,z,$$

where $v^\varepsilon \subset \ker B$, and z is such that $Bz = \lambda$, which yields

$$|\gamma| \le \left|\tilde{\lambda}_h - \lambda^\varepsilon\right| + \mathcal{O}(\varepsilon) + |(B_h - B)\,z|\,.$$

We finally obtain that $|u^\varepsilon - u_h^\varepsilon| + |\lambda^\varepsilon - \lambda_h^\varepsilon|$ is less than

$$C\left(\inf_{\tilde{u}_h \in V_h} |\tilde{u}_h - u^\varepsilon| + \inf_{\tilde{\lambda}_h \in \Lambda_h} \left|\tilde{\lambda}_h - \lambda^\varepsilon\right| + |(B_h^\star - B^\star)\lambda| + \varepsilon + |(B_h - B)z|\right),$$

so that, by eliminating u^ε in the left-hand side, and again using $|u^\varepsilon - u| = \mathcal{O}(\varepsilon)$ and $|\lambda^\varepsilon - \lambda| = \mathcal{O}(\varepsilon)$ (see Corollary 2.4 and Proposition 2.6), we obtain the error estimate. □

PROPOSITION 3.2 (abstract stability estimate). *Let V and Λ be two Hilbert spaces, $B \in \mathcal{L}(V, \Lambda)$, $a(\cdot,\cdot)$ and $c(\cdot,\cdot)$ bilinear continuous functionals, which we suppose elliptic. Then the problem*

$$\left\{\begin{array}{rcrcll} a(u,v) & + & (\lambda, Bv) & = & \langle \varphi\,,\, v\rangle & \forall v \in V, \\ (Bu, \mu) & - & c(\lambda, \mu) & = & \langle \Psi\,,\, \mu\rangle & \forall \mu \in \Lambda \end{array}\right. \tag{3.5}$$

admits a unique solution $(u, \lambda) \in V \times \Lambda$. We assume furthermore that there exists a constant $\beta > 0$ such that[1]

$$\beta\left|P_{(\ker B)^\perp} v\right| \le |Bv|\,, \quad \sup_{v \in V} \frac{(\mu, Bv)}{|v|} \ge \beta\,\|\mu\|_{\Lambda/\ker B^\star}\,, \tag{3.6}$$

[1] As the second inequality of (3.6) is a direct consequence of the first one, it could be suppressed. We keep both assumptions for clarity reasons.

that Ψ can be written

$$\langle \Psi\ , \mu \rangle = (Bw, \mu) + c(\gamma, \mu), \tag{3.7}$$

and finally that $c(\cdot,\cdot)$ verifies

$$\mu_1 \perp \mu_2 \longrightarrow c(\mu_1, \mu_2) = 0. \tag{3.8}$$

Then we have the following estimate:

$$|u| + |\lambda| \leq C(\|\varphi\| + |w| + |\gamma|), \tag{3.9}$$

where C is a locally bounded expression of $\|a\|$, $1/\alpha$, $1/\beta$, $\|B\|$, $\|c\|$ (α is the coercivity constant of $a(\cdot,\cdot)$). Note that C does not depend upon the coercivity constant of $c(\cdot,\cdot)$.

Proof. The first part of the proposition is trivial. With obvious notation, problem (3.5) can be written

$$\left\{ \begin{array}{rcrcl} Au & + & B^\star \lambda & = & \varphi, \\ Bu & - & M\lambda & = & \Psi, \end{array} \right. \tag{3.10}$$

so that (u, λ) is uniquely determined as

$$u = (A + B^\star M^{-1} B)^{-1} \left(\varphi + B^\star M^{-1} \Psi\right), \quad \lambda = M^{-1}\left(Bu - \Psi\right).$$

In order to get an upper bound of $|u|$ which does not degenerate with $c(\cdot,\cdot)$, we introduce, following [BF91],

$$u = \underbrace{u_0}_{\in \ker B} + \underbrace{u^\perp}_{\in (\ker B)^\perp}, \quad \lambda = \underbrace{\lambda_0}_{\in \ker B^\star} + \underbrace{\lambda^\perp}_{\in (\ker B^\star)^\perp}. \tag{3.11}$$

From (3.6) and the first line of (3.5), we have

$$\beta \left|\lambda^\perp\right| = \beta \|\lambda\|_{\Lambda/\ker B^\star} \leq \sup \frac{(\lambda, Bv)}{|v|} \leq \|a\| \, |u| + \|\varphi\|. \tag{3.12}$$

From (3.6) again and the second line of (3.5), we get

$$\beta \left|u^\perp\right| = \beta \left|P_{(\ker B)^\perp} u\right| \leq |Bu| = \sup \frac{(Bu, \mu)}{|\mu|} \leq \|\Psi\| + \|c\|^{1/2} \, c(\lambda, \lambda)^{1/2}. \tag{3.13}$$

From the ellipticity of $a(\cdot,\cdot)$ and the first line of (3.5),

$$\begin{aligned} \alpha \, |u_0| \leq a\left(u_0, \frac{u_0}{|u_0|}\right) &\leq \sup_{v_0 \in \ker B} \frac{a(u_0, v_0)}{|v_0|} = \sup_{v_0 \in \ker B} \frac{a(u, v_0) - a(u^\perp, v_0)}{|v_0|} \\ &\leq \|\varphi\| + \|a\| \left|u^\perp\right|. \end{aligned} \tag{3.14}$$

From (3.13) and (3.14), we have

$$\begin{aligned} |u| \leq \left|u^\perp\right| + |u_0| &\leq \frac{1}{\beta}\left(\|\Psi\| + \|c\|^{1/2} \, c(\lambda, \lambda)^{1/2}\right) + \frac{1}{\alpha}\left(\|\varphi\| + \|a\| \left|u^\perp\right|\right) \\ &\leq \frac{1}{\beta}\left(\|\Psi\| + \|c\|^{1/2} \, c(\lambda, \lambda)^{1/2}\right)\left(1 + \frac{\|a\|}{\alpha}\right) + \frac{\|\varphi\|}{\alpha}. \end{aligned} \tag{3.15}$$

Now substracting the two lines of (3.5) with $v = u$ and $\mu = \lambda$, we obtain

$$\begin{aligned} a(u,u) + c(\lambda,\lambda) &= \langle \varphi \, , u \rangle - \langle \Psi \, , \lambda \rangle = \langle \varphi \, , u \rangle - (Bw, \lambda) - c(\gamma, \lambda) \\ &\leq \|\varphi\| \, |u| + \|B\| \, |w| \, \left|\lambda^{\perp}\right| + c(\gamma,\gamma)^{1/2} c(\lambda,\lambda)^{1/2}, \end{aligned}$$

so that, from (3.15) and (3.12),

$$\begin{aligned} a(u,u) + c(\lambda,\lambda) &\leq \left(\|\varphi\| + \frac{\|B\|}{\beta} \, |w| \, \|a\| \right) \left(\frac{\|\Psi\|}{\beta} \left(1 + \frac{\|a\|}{\alpha} \right) + \frac{\|\varphi\|}{\alpha} \right) \\ &+ c(\lambda,\lambda)^{1/2} \left(c(\gamma,\gamma)^{1/2} + \frac{1}{\beta} \, \|c\|^{1/2} \left(1 + \frac{\|a\|}{\alpha} \right) \left(\|\varphi\| + \frac{\|B\|}{\beta} \, |w| \, \frac{\|a\|}{\alpha} \right) \right), \end{aligned} \tag{3.16}$$

which can be written

$$a(u,u) + c(\lambda,\lambda) \leq P_0(\|\varphi\| \, , \|\Psi\| \, , |w| \, , |\gamma|_c) + c(\lambda,\lambda)^{1/2} P_1(\|\varphi\| \, , \|\Psi\| \, , |w| \, , |\gamma|_c),$$

where P_0 (resp., P_1) is an homogeneous polynomial of degree 2 (resp., 1) in its four variables. The coefficients of those polynomials are polynomial in $\|B\|$, $\|a\|$, $1/\beta$, $1/\alpha$, $\|c\|^{1/2}$ with positive coefficients. We write $X = c(\lambda,\lambda)^{1/2}$, so that $X^2 \leq P_1 X + P_0$, which implies $|X| \leq P_1 + \sqrt{P_0}$, and finally

$$c(\lambda,\lambda) = X^2 \leq 2P_1^2 + 2P_0 = P_2(\|\varphi\| \, , \|\Psi\| \, , |w| \, , |\gamma|_c),$$

where P_2 is an homogeneous polynomial of degree 2. It is dominated by the square of the sum of the modulus of its variables, so that

$$c(\lambda,\lambda)^{1/2} \leq C(\|\varphi\| + \|\Psi\| + |w| + |\gamma|_c).$$

Again using (3.16) (we keep C to design a generic constant, or more precisely a polynomial in $\|B\|$, $\|a\|$, $1/\beta$, $1/\alpha$, $\|c\|^{1/2}$), we obtain immediately

$$|u| \leq C(\|\varphi\| + \|\Psi\| + |w| + |\gamma|_c).$$

Finally, we write the second line of (3.5) with $\mu \in \ker B^{\star}$. As $c(\cdot,\cdot)$ verifies (3.8), it yields $\lambda_0 = P_{\ker B^{\star}} \gamma$, so that $|\lambda_0| \leq |\gamma|$. As $|\gamma|_c \leq \|c\|^{1/2} \, |\gamma|$, and $\|\Psi\| \leq |w| + |\gamma|$, estimate (3.9) is obtained. □

4. Application. This section is dedicated to the application of Theorem 3.1 to a particular problem, namely a scalar version of the rigidity constraint for fluid-particle flows.

4.1. Model problem. In order to present explicit constructions when needed, we consider a particular situation. We introduce $\Omega =]-2,2[^2$, and $\mathcal{O} = B(0,1) \subset\subset \Omega$ (see Figure 4.1). The case of more general situations is addressed in Remark 4.2, at the end of this paper. We consider the following problem:

$$\left\{ \begin{array}{rcll} -\triangle u & = & f & \text{in} \quad \Omega \setminus \overline{\mathcal{O}}, \\ u & = & 0 & \text{on} \quad \partial\Omega, \\ u & = & U & \text{on} \quad \partial\mathcal{O}, \\ \displaystyle\int_{\partial\mathcal{O}} \frac{\partial u}{\partial n} & = & 0, & \end{array} \right. \tag{4.1}$$

where U is an unknown constant, and $f \in L^2(\Omega \setminus \overline{\mathcal{O}})$. The scalar field u can be seen as a temperature and $\mathcal{O}$ as a zone with infinite conductivity.

DEFINITION 4.1. *We say that u is a weak solution to* (4.1) *if $u \in V = H^1_0(\Omega)$, there exists $U \in \mathbb{R}$ such that $u = U$ a.e. in $\mathcal{O}$, and*

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv \quad \forall v \in \mathcal{D}_{\mathcal{O}}(\Omega),$$

where $\mathcal{D}_{\mathcal{O}}(\Omega)$ is the set of all those functions which are compactly supported, C^∞ on Ω, and which are constant over $\mathcal{O}$.

PROPOSITION 4.2. *Problem* (4.1) *admits a unique weak solution $u \in V = H^1_0(\Omega)$, which is characterized as the solution to the minimization problem*

$$\left\{ \begin{array}{l} \text{Find } u \in K \text{ such that} \\ J(u) = \inf\limits_{v \in K} J(v)\,, \quad \text{with } J(v) = \dfrac{1}{2}\displaystyle\int_\Omega |\nabla u|^2 - \int_\Omega fv, \\ K = \left\{ v \in H^1_0(\Omega), \nabla v = 0 \text{ a.e. in } \mathcal{O} \right\}, \end{array} \right. \tag{4.2}$$

where f has been extended by 0 inside $\mathcal{O}$. Furthermore the restriction of u to the domain $\Omega \setminus \overline{\mathcal{O}}$ is in $H^2(\Omega \setminus \overline{\mathcal{O}})$.

Proof. Existence and uniqueness are direct consequences of the Lax–Milgram theorem applied in $K = \{v \in V\,,\ \nabla v = 0 \text{ a.e. in } \mathcal{O}\}$, which gives in addition the characterization of u as the solution to (4.2). Now $u_{|\Omega\setminus\overline{\mathcal{O}}}$ satisfies $-\triangle u = f$, with regular Dirichlet boundary conditions on the boundary of $\Omega \setminus \overline{\mathcal{O}}$ which decomposes as $\partial\mathcal{O} \cup \partial\Omega$. As Ω is a convex polygon and $\partial\mathcal{O}$ is smooth, standard theory ensures that $u_{|\Omega\setminus\overline{\mathcal{O}}} \in H^2(\Omega \setminus \overline{\mathcal{O}})$. ∎

PROPOSITION 4.3 (saddle-point formulation). *Let u be the weak solution to* (4.1). *There exists a unique $\lambda \in \Lambda = L^2(\mathcal{O})^2$ such that λ is a gradient, and*

$$\int_\Omega \nabla u \cdot \nabla v + \int_{\mathcal{O}} \lambda \cdot \nabla v = \int_\Omega fv \quad \forall v \in V.$$

In addition λ is in $H^1(\mathcal{O})^2$.

Proof. The first part is a consequence of Proposition 2.5, where B is defined by

$$B \ : \ v \in H^1_0(\Omega) \longmapsto \nabla v \in L^2(\mathcal{O})^2.$$

Let us prove that B has a closed range. Considering $\mu \in \Lambda$ with $\mu = \nabla v$, we define $w \in H^1_0(\mathcal{O})$ as $w = v - m(v)$, where $m(v)$ is the mean value of v over $\mathcal{O}$. By the Poincaré–Wirtinger inequality, one has

$$\|w\|_{H^1(\mathcal{O})} \leq C \|\mu\|_{L^2(\mathcal{O})^2}\,.$$

Now, as $\mathcal{O} \subset\subset \Omega$, there exists a continuous extension operator from $H^1(\mathcal{O})$ to $H^1_0(\Omega)$, so that we can extend w to obtain $\tilde{w} \in H^1_0(\Omega)$ with a norm controlled by $\|\mu\|_{L^2(\mathcal{O})^2}$, which proves the closed character of $B(V)$, and consequently the existence of $\lambda \in \Lambda$, and its uniqueness in $B(V)$.

Let us now describe λ. We have

$$\int_\Omega \nabla u \cdot \nabla v + \int_{\mathcal{O}} \lambda \cdot \nabla v = \int_\Omega fv,$$

so that, by taking test functions in $\mathcal{D}(\mathcal{O})$, we get $\lambda \in H_{\rm div}(\mathcal{O})$ with $\nabla \cdot \lambda = 0$. Taking now test functions which do not vanish on the boundary of $\mathcal{O}$, we identify the normal trace of λ with $\partial u/\partial n \in H^{1/2}(\partial\mathcal{O})$. Therefore λ is defined as the unique divergence-free vector field in $\mathcal{O}$, with normal derivative equal to $\partial u/\partial n$ on $\partial\mathcal{O}$, which, in addition, is a gradient. In other words $\lambda = \nabla\Phi$, with

$$\begin{cases} \triangle\Phi &= 0 & \text{in } \mathcal{O}, \\ \dfrac{\partial\Phi}{\partial n} &= \dfrac{\partial u}{\partial n} & \text{on } \partial\mathcal{O}. \end{cases}$$

As $\mathcal{O}$ is smooth, $\Phi \in H^2(\mathcal{O})$, so that $\lambda = \nabla\Phi \in H^1(\mathcal{O})^2$. □

We introduce the penalized version of problem (4.2)

$$\begin{cases} \text{Find } u^\varepsilon \in V \text{ such that } J^\varepsilon(u^\varepsilon) = \inf\limits_{v\in V} J^\varepsilon(v)\,, \\ J^\varepsilon(v) = \dfrac{1}{2}\displaystyle\int_\Omega |\nabla v|^2 + \frac{1}{2\varepsilon}\int_{\mathcal{O}} |\nabla v|^2 - \int_\Omega fv. \end{cases} \tag{4.3}$$

Now we consider the family of Cartesian triangulations (T_h) of the square Ω (see Figure 4.1), and we denote by V_h the standard finite element space of continuous, piecewise affine function with respect to T_h:

$$V_h = \left\{ v_h \in V\,,\ V_{|T} \text{ is affine } \forall T \in T_h \right\}.$$

It is tempting to define the fully discretized problem as the problem which consists in minimizing J^ε over V_h. But this straightforward approach (which does not correspond to what is done in actual computations; see Remark 4.1) raises some problems in relation to the discrete inf-sup condition which we need to establish the error estimate (see Proposition 4.7). It is related to the fact that we cannot control the size of intersections of triangles with $\mathcal{O}$ (relative to the size of the whole triangle, which is $h^2/2$). To overcome this problem, many strategies can be adopted, all of them leading to change B onto a new discrete operator B_h. We propose here a radical method, which simply consists in removing in the penalty integral all squares (two-triangle sets) which intersect the boundary of $\mathcal{O}$. It will be made clear that the convergence

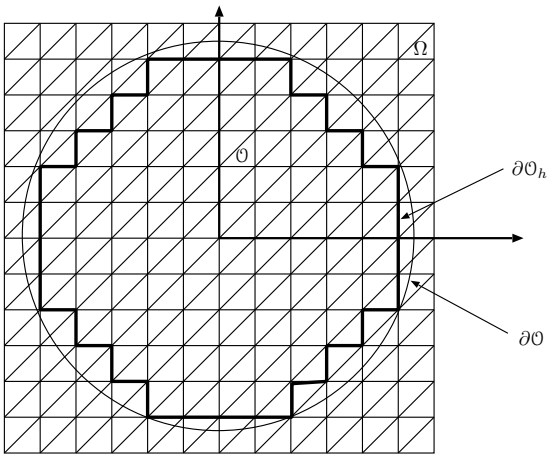

FIG. 4.1. *Domains Ω, $\mathcal{O}$, $\mathcal{O}_h$, and the mesh T_h.*

result is not sensitive to what is actually done in the neighborhood of $\partial\Omega$. The proof simply requires that the reduced obstacle is included in the exact one, and that the difference set $\mathcal{O} \setminus \mathcal{O}_h$ lies in a narrow band whose width goes to 0 like h.

DEFINITION 4.4. *The reduced obstacle $\mathcal{O}_h \subset \mathcal{O}$ is defined as the union of the triangles which belong to an elementary square which is contained in the disk $\mathcal{O}$ (see Figure 4.1).*

DEFINITION 4.5. *We recall that $V = H^1_0(\Omega)$, Λ is $L^2(\mathcal{O})^2$, and $B \in \mathcal{L}(V, \Lambda)$ is the gradient operator (see Proposition 4.3). We define $B_h \in \mathcal{L}(V, \Lambda)$ as*

$$v \in V \longmapsto \mu = B_h v = \mathbb{1}_{\mathcal{O}_h} \nabla v,$$

where $\mathbb{1}_{\mathcal{O}_h}$ is the characteristic function of $\mathcal{O}_h$ (see Definition 4.4). Finally, the discretization space $\Lambda_h \subset \Lambda = L^2(\mathcal{O})^2$ is the set of all those vector fields μ_h such that their restriction to $\mathcal{O}_h$ is the gradient of a scalar field $v_h \in V_h$, and which vanish a.e. in $\mathcal{O} \setminus \mathcal{O}_h$, which we can express as

$$\Lambda_h = \{\mu_h \in \Lambda\,,\ \exists v_h \in V_h\,,\ \mu_h = B_h v_h\} = B_h(V_h).$$

The fully discretized problem reads

$$\left\{ \begin{array}{l} \text{Find } u_h^\varepsilon \in V_h \text{ such that } J_h^\varepsilon(u^\varepsilon) = \inf\limits_{v_h \in V_h} J_h^\varepsilon(v_h)\,, \\ \\ J_h^\varepsilon(v_h) = \dfrac{1}{2}\displaystyle\int_\Omega |\nabla v_h|^2 + \dfrac{1}{2\varepsilon}\int_{\mathcal{O}_h} |\nabla v_h|^2 - \int_\Omega f v_h. \end{array} \right. \tag{4.4}$$

4.2. Error estimate for the model problem.

PROPOSITION 4.6 (primal/dual error estimate for (4.1), nonconforming case). *Let u be the weak solution to (4.1), u_h^ε the solution to (4.4), and λ the Lagrange multiplier (see Proposition 4.3), and let $\lambda_h^\varepsilon = B_h u_h^\varepsilon/\varepsilon$ (see Definition 4.5). We have the following error estimate:*

$$|u - u_h^\varepsilon| + |\lambda - \lambda_h^\varepsilon| \leq C(h^{1/2} + \varepsilon). \tag{4.5}$$

Proof. The proof is based on the abstract estimate in Theorem 3.1. All technical ingredients are put off until the end of the section. We shall simply refer here to the corresponding properties. The crucial requirement is the discrete inf-sup condition, which can be established for this choice of B_h (see Proposition 4.7). The terms

$$\inf_{\tilde{u}_h \in V_h} |\tilde{u}_h - u| \quad \text{and} \quad \inf_{\tilde{\lambda}_h \in \Lambda_h} \left|\tilde{\lambda}_h - \lambda\right|$$

can be shown to behave like $h^{1/2}$ (see Propositions 4.8 and 4.9, respectively). The last two terms can be handled the same way as $|\tilde{\lambda}_h - \lambda|$. Indeed,

$$|(B_h^\star - B^\star)\lambda| \leq |\lambda|_{0,\mathcal{O}\setminus\overline{\mathcal{O}_h}},$$

which is a $\mathcal{O}(h^{1/2})$ (it is the L^2 norm of a function with H^1 regularity, on a h-neighborhood of $\partial\mathcal{O}$). The very same argument holds for $|(B_h - B)z|$ (in our case, both quantities are the same). ∎

PROPOSITION 4.7 (discrete inf-sup condition). *Let Ω and $\mathcal{O}$ be defined as in the beginning of section 4. We introduce $h = 1/N$, $N \in \mathbb{N}$, and T_h is the regular triangulation with step h, so that the center of $\mathcal{O}$ is a vertex of T_h. According to*

Definitions 4.4 *and* 4.5, $\mathcal{O}_h$ *is the reduced obstacle, and* $\Lambda_h \subset L^2(\mathcal{O})^2 = \Lambda$ *is the set of all those vector fields which are the gradient of a piecewise affine function in* $\mathcal{O}_h$, *and which vanish in* $\mathcal{O} \setminus \mathcal{O}_h$.

There exists $\beta > 0$ *such that, for all* h $(= 1/N)$,

$$\beta \left| P_{(\ker B_h)^\perp} v_h \right| \le |B_h v_h| \quad \forall v_h \in V_h\,, \ \sup_{v_h \in V_h} \frac{(B_h v_h, \lambda_h)}{|v_h|} \ge \beta \|\lambda_h\|_{\Lambda_h}\,. \tag{4.6}$$

Proof. Let $v_h \in V_h$ be given. If we are able to build $w_h \in V_h$ such that $B_h w_h = B_h v_h$, with $\|w_h\| \le C \|B_h v_h\|$, we obtain

$$\left| P_{(\ker B_h)^\perp} v_h \right| = \inf_{\tilde{v}_h \in \ker B_h} |v_h - \tilde{v}_h| \le |v_h - (w_h - v_h)| = |w_h| \le C\,|B_h v_h|\,,$$

and the first inequality is proven. Let us describe how this $w_h \in V_h$ can be built in five steps. First, we introduce $w_h^1 = v_h - \overline{v}_h$, where $\overline{v}_h$ is the mean value of w_h over $\mathcal{O}_h$. Note that w_h^1 is not in V_h (it does not vanish on $\partial\Omega$), but we consider only its restriction to $\mathcal{O}_h$. We have $B_h w_h^1 = B_h v_h$, and the norm of w_h^1 is controlled: $\left\|w_h^1\right\|_{H^1(\mathcal{O}_h)} \le C_1 \|B_h v_h\|_{L^2(\mathcal{O}_h)^2}$ by the Poincaré–Wirtinger inequality (with a constant which does not depend on h, as can be checked easily).

We shall now describe how we plan to extend w_h^1 in the first quadrant, the three others being done the same way. This construction is illustrated by Figure 4.2. The first step consists in extending w_h^1 in the polygonal domain $CA_3A_2'A_1$ on each horizontal segment by symmetry (see Figure 4.2). A similar construction extends w_h^1 in

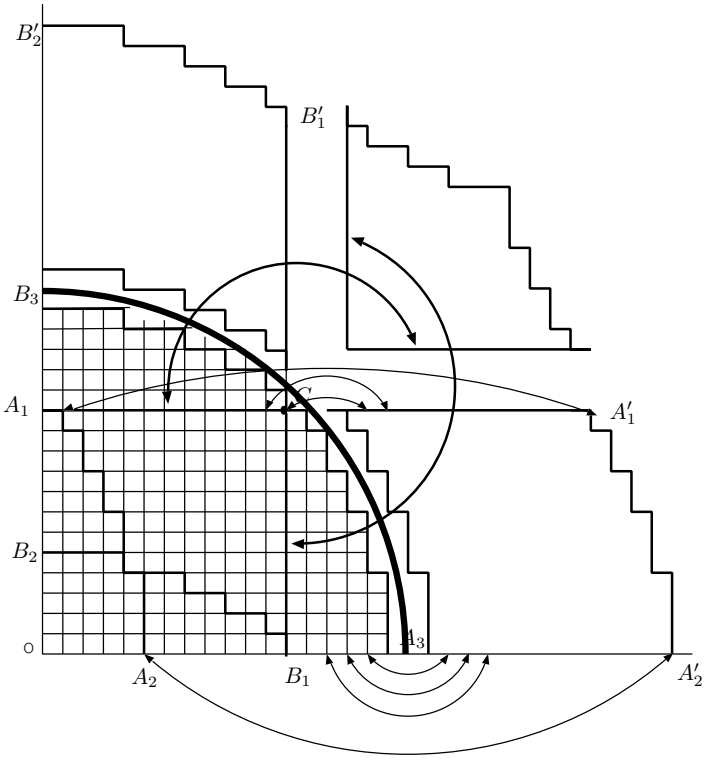

FIG. 4.2. *Construction of* w_h^2.

FIG. 4.3. *Stretching of w_h^2 (detail).*

$CB_1'B_2'B_3$. Now the function is simply extended in the upper right zone by symmetry around C. To show that the H^1 seminorm of the newly defined function w_h^2 is under control, we first remark that the shift between two consecutive lines does not exceed one cell. Now consider the detail in Figure 4.3. On the left we represented a detail of the triangulated domain in $\mathcal{O}$ where w_h^2 is already defined; the u_i's and v_i's represent the values of w_h^2 at some vertices. Now by applying the "symmetry" described previously, we obtain the stretched function which we represent on a single element. To control the effect of this stretching, we use Lemma 4.10 in the following way: The square of the H^1 seminorm of the new function is a quadratic nonnegative form q_1 in the six variables, and the square of the H^1 seminorm corresponding to the left-hand situation itself is a scale invariant quadratic, nonnegative form q_2 in the same variables, so that Lemma 4.10 ensures the existence of a universal constant C such that $q_1 \leq Cq_2$. As a consequence, the H^1 seminorm of the stretched function (in $CA_3A_2'A_1'$) is controlled by the H^1 seminorm of the initial function (in $CA_1A_2A_3$). As the new function in $CA_1'B_1'$ is obtained by standard symmetry, the H^1 seminorm identifies with the one of the initial function in CA_1B_1.

This leads to a new function w_h^2 defined on $\mathcal{O}_h^2$, subtriangulation of T_h, with $\left|w_h^2\right|_{1,\mathcal{O}_h^4} \leq C_2 \left\|B_h v_h\right\|_{L^2(\Omega)^2}$. As w_h^2 has zero mean value in $B(0,1/2)$, one has

$$\left\|w_h^2\right\|_{H^1(\mathcal{O}_h^2)} \leq C_2' \left\|B_h v_h\right\|_{L^2(\Omega)^2}.$$

Finally, $\mathcal{O}_h^2$ contains a ball strictly larger than $\mathcal{O}$, say $B(0, 1+\sqrt{2}/4)$. Considering now a smooth function ρ which is equal to 1 in $B(0,(1+r)/2)$, and 0 outside $B(0,r)$, we define w_h^3 as $I_h(\rho w_h^2)$ on $\mathcal{O}_h^2$, and 0 in $\Omega \setminus \mathcal{O}_h^2$, where I_h is the standard interpolation operator. This function is in $V_h \cap H_0^1(\Omega)$, and it verifies

$$B_h w_h^3 = \lambda_h, \quad \left\|w_h^3\right\|_{H^1(\Omega)} \leq C_3 \left\|B_h v_h\right\|_{L^2(\Omega)^2},$$

so that the first inequality of (4.6) holds, with $\beta = 1/C_3$.

The second one is a direct consequence of the first one: given $\lambda_h = B_h u_h$, one considers $w_h = P_{(\ker B_h)^\perp} v_h$, so that

$$\sup_{v_h \in V_h} \frac{(B_h v_h, \lambda_h)}{|v_h|} \geq \frac{(B_h w_h, \lambda_h)}{|w_h|} = \frac{|B_h w_h|^2}{|w_h|} \geq \beta \left|B_h w_h\right| = \beta \left\|\lambda_h\right\|_{\Lambda_h},$$

which ends the proof. □

PROPOSITION 4.8 (approximation of u). *We make the same assumptions as in Proposition* 4.7, *and we consider $u \in H_0^1(\Omega)$ such that $u = U \in \mathbb{R}$ a.e. in $\mathcal{O}$, $u_{\Omega \setminus \overline{\mathcal{O}}} \in H^2(\Omega \setminus \overline{\mathcal{O}})$. There exists $C > 0$ such that*

$$\inf_{\tilde{u}_h \in V_h} \left\|u - \tilde{u}_h\right\|_{H^1(\Omega)} \leq Ch^{1/2}.$$

Proof. We recall that I_h is the standard interpolation operator from $C(\Omega)$ onto V_h. Let us assume here that the constant value U on $\mathcal{O}$ is O (which can be achieved by substracting a smooth extension of this constant outside $\mathcal{O}$). Now we define $\tilde{\mathcal{O}}_h$ as the union of all those triangles of T_h which have a nonempty intersection with $\mathcal{O}$. We define $\tilde{u}_h$ as the function in V_h which is 0 in $\tilde{\mathcal{O}}_h$ and which identifies with $I_h u$ at all other vertices. We introduce a narrow band around $\mathcal{O}$:

$$\omega_h = \left\{ x \in \Omega\,,\ x \notin \overline{\mathcal{O}}\,,\ d(x, \overline{\mathcal{O}}) < 2\sqrt{2}h \right\}. \tag{4.7}$$

As $u_{|\Omega\setminus\overline{\mathcal{O}}} \in H^2(\Omega \setminus \overline{\mathcal{O}})$, standard finite element estimates give

$$|u - \tilde{u}_h|_{0,L^2(\Omega\setminus(\mathcal{O}\cup\overline{\omega}_h))} \leq Ch^2 \, |u|_{H^2(\Omega\setminus\overline{\mathcal{O}})}\,, \tag{4.8}$$

$$|u - \tilde{u}_h|_{1,L^2(\Omega\setminus(\mathcal{O}\cup\overline{\omega}_h))} \leq Ch \, |u|_{H^2(\Omega\setminus\overline{\mathcal{O}})}\,. \tag{4.9}$$

By construction, both L^2 and H^1 errors in $\mathcal{O}$ are zero. There remains to estimate the error in the band ω_h. The principle is the following: $\tilde{u}_h$ is a poor approximation of u in ω_h, but it is not very harmful because ω_h is small. Note that similar estimates are proposed in [SMSTT05] or [AR08] . For the sake of completeness, and because it is essential to understand why a better order than 1/2 cannot be expected, we shall detail here the proof. First of all, we write

$$\|u - \tilde{u}_h\| \leq |u|_{0,\omega_h} + |u|_{1,\omega_h} + |u_h|_{0,\omega_h} + |u_h|_{1,\omega_h} = A + B + C + D. \tag{4.10}$$

Lemma 4.13 ensures $B \leq Ch^{1/2}$, and $A \leq Ch^{3/2}$. As for $\tilde{u}_h$ (terms C and D in (4.10)), the proof is less trivial. It relies on the technical lemmas (Lemmas 4.11, 4.12, and 4.14 (see section 4.3)) which can be used as follows. The problematic triangles are those on which $\tilde{u}_h$ identifies neither with 0, nor with $I_h u$. On such triangles, $\tilde{u}_h$ sticks to $I_h u$ at 1 or 2 vertices, and vanishes at 2 or 1 vertices. As a consequence, the L^∞ norm of $\tilde{u}_h$ is less than the L^∞ norm of $I_h u$. Let T be such a triangle. We write (using Lemma 4.11, the latter remark, the fact that I_h is a contraction from L^∞ onto L^∞, Lemma 4.11 again, and Lemma 4.14)

$$\begin{aligned} |\tilde{u}_h|^2_{L^2(T)} &\leq C' \, |T| \, \|\tilde{u}_h\|^2_{L^\infty(T)} \leq C' \, |T| \, \|I_h u\|^2_{L^\infty(T)} \\ &\leq \frac{C'}{C} \, \|I_h u\|^2_{L^2(T)} \leq C'' \left(\|u\|^2_{L^2(T)} + h^4 \, |u|^2_{2,T} \right). \end{aligned}$$

By summing up all these contributions over all triangles which intersect ω_h, and using the fact that the L^2 norm of u on ω_h behaves like $h^{3/2} \, |u|_{2,T}$, we obtain

$$\|\tilde{u}_h\|^2_{L^2(\omega_h)} \leq \sum_{T\cap\omega_h\neq\emptyset} \|\tilde{u}_h\|^2_{L^2(T)} \leq h^3 \, |u|^2_{2,T}\,,$$

which gives the expected $h^{3/2}$ estimate for C. The last term of (4.10) is directly obtained by the previous estimate combined with the inverse inequality expressed by Lemma 4.12. ∎

PROPOSITION 4.9 (approximation of λ). *Let $\lambda \in H^1(\mathcal{O})^2$ be given, with $\lambda = \nabla w$, $w \in H^2(\mathcal{O})$. There exists a constant $C > 0$ such that*

$$\inf_{\tilde{\lambda}_h \in \Lambda_h} \left\| \lambda - \tilde{\lambda}_h \right\|_{L^2(\mathcal{O})} \leq Ch^{1/2} \, |\lambda|_{1,\mathcal{O}}\,,$$

where Λ_h is defined in section 3 (see Definition 4.5).

Proof. First of all, we extend w on $\Omega \setminus \overline{\mathcal{O}}$, to obtain a function (still denoted by w) in $H^1_0(\Omega) \cap H^2(\Omega)$. Let us define w_h as the standard interpolate of w over T_h. One has $|w - w_h|_{1,\mathcal{O}} \leq Ch$. We define $\tilde{\lambda}_h \in \Lambda_h$ as the piecewise constant function which identifies with ∇w_h on $\mathcal{O}_h$ (see Definition 4.4), and which vanishes in $\mathcal{O} \setminus \mathcal{O}_h$. One has

$$\left\| \nabla w_h - \tilde{\lambda}_h \right\|_{L^2(\mathcal{O})} = \left\| \nabla w_h - \tilde{\lambda}_h \right\|_{L^2(\mathcal{O} \setminus \mathcal{O}_h)} = \left\| \nabla w_h \right\|_{L^2(\mathcal{O} \setminus \mathcal{O}_h)} \leq C \left\| \nabla w \right\|_{L^2(\mathcal{O} \setminus \mathcal{O}_h)},$$

which is the H^1 seminorm of a function in H^2, in a narrow domain. Therefore it behaves like $h^{1/2}$ times the H^2 seminorm of u (see Lemma 4.13 and Remark 4.3), which is the H^1 seminorm of λ. Finally, one gets

$$\left\| \lambda - \tilde{\lambda}_h \right\|_{L^2(\mathcal{O})} \leq |w - w_h|_{1,\mathcal{O}} + \left\| \nabla w_h - \tilde{\lambda}_h \right\|_{L^2(\mathcal{O})} \leq C(h + h^{1/2}) \, |\lambda|_{1,\mathcal{O}},$$

which ends the proof. ∎

Remark 4.1 (boundary fitted meshes). Although it is somewhat in contradiction with its original purpose, the penalty method can be used together with a discretization based on a boundary fitted mesh. In that case, the approximation error behaves no longer like $h^{1/2}$ but like h.

Remark 4.2 (technical assumptions). Some assumptions we made are only technical and can surely be relaxed without changing the convergence results. For example the inclusion, which we supposed circular, could be a collection of smooth domains. Note that a convex polygon is not acceptable, as it is seen from the outside, so that u may no longer be in H^2, which rules out some of the approximation properties we made. Concerning the mesh, we have good confidence in the fact that the result generalizes to any kind of unstructured mesh, but the proof of Proposition 4.7 in the general case can no longer be based on an explicit construction.

4.3. Technical lemmas. We gather here some elementary properties which are used in the proofs of Propositions 4.6, 4.7, 4.8, and 4.9.

LEMMA 4.10. *Let E be a finite dimensional real vector space, with q_1 and q_2 two nonnegative quadratic forms with $\ker q_2 \subset \ker q_1$. There exists $C > 0$ such that $q_1 \leq C q_2$.*

Proof. As q_2 is nonnegative, $\tilde{v} \mapsto |\tilde{v}|_{q_2(v)} = \sqrt{q_2(v)}$ is a norm for $E/\ker q_2$. Now we define

$$\tilde{q}_1 : \tilde{v} \in E/\ker q_2 \longmapsto \tilde{q}_1(\tilde{v}) = q_1(v) \in \mathbb{R}.$$

As $\ker q_1$ contains $\ker q_2$, this functional is well defined. As it is quadratic over a finite dimensional space, it is continuous for the norm $\sqrt{q_2}$, so that

$$q_1(v) = \tilde{q}_1(\tilde{v}) \leq C \, |v|^2_{q_2} = q_2(v),$$

which ends the proof. ∎

LEMMA 4.11. *There exist constants C and C' such that, for any nondegenerated triangle T, for any function w_h affine in T,*

$$C \, |T| \, \|w_h\|^2_{L^\infty(T)} \leq \|w_h\|^2_{L^2(T)} \leq C' \, |T| \, \|w_h\|^2_{L^\infty(T)}. \tag{4.11}$$

Proof. It is a consequence of the fact that, when deforming the supporting triangle T, the L^∞ norm is unchanged whereas the L^2 norm scales like $|T|^{1/2}$. ∎

LEMMA 4.12. *There exists a constant C such that, for any nondegenerated triangle T, for any function w_h affine in T,*

$$|w_h|^2_{1,K} \leq C\frac{|T|}{\rho_K^2}\,\|w_h\|^2_{L^\infty(T)}\,,$$

where ρ_K is the diameter of the inscribed circle.

Proof. Again, it is a straightforward consequence of the fact that, when deforming the supporting triangle T, the L^∞ norm is unchanged whereas the gradient (which is constant over the triangle) scales like $1/\rho_k$, so that the H^1 seminorm scales like $|T|^{1/2}/\rho_K$. ∎

The next lemma establishes some Poincaré-like inequalities in narrow domains.

LEMMA 4.13. *Let $\mathcal{O} \subset \mathbb{R}^2$ be the unit disk, strongly included in a domain Ω, and let ω_η be the narrow band (note that this definition differs slightly from* (4.7)*, which is of no consequence):*

$$\omega_\eta = \left\{x \in \Omega\,,\ x \notin \overline{\mathcal{O}}\,,\ d(x,\overline{\mathcal{O}}) < \eta\right\},\quad \textit{with } \eta > 0.$$

Denoting by $|\cdot|_{p,\omega}$ the H^p seminorm over ω, we have the following estimates:

$$|\varphi|_{0,\omega_\eta} \leq C\eta^{1/2}\,|\varphi|_{1,\Omega\setminus\overline{\mathcal{O}}} \quad \forall\varphi \in H^1(\Omega\setminus\overline{\mathcal{O}}),\quad \varphi_{|\partial\Omega} = 0,$$

$$|\varphi|_{1,\omega_\eta} \leq C\eta^{1/2}\,|\varphi|_{2,\Omega\setminus\overline{\mathcal{O}}} \quad \forall\varphi \in H^2(\Omega\setminus\overline{\mathcal{O}}),\quad \varphi_{|\partial\Omega} = 0,$$

$$|\varphi|_{0,\omega_\eta} \leq C\eta^{3/2}\,|\varphi|_{2,\Omega\setminus\overline{\mathcal{O}}} \quad \forall\varphi \in H^2(\Omega\setminus\overline{\mathcal{O}}),\quad \varphi_{|\partial\Omega} = 0,\quad \varphi_{|\partial\mathcal{O}} = 0.$$

Proof. We assume here that φ is C^1 in $\Omega\setminus\overline{\mathcal{O}}$ (the general case is obtained immediately by density). Using polar coordinates, we write $u(r,\theta) = u(1,\theta) + \int_1^r \partial_r u\,dr$, so that

$$\begin{aligned}|u|^2_{0,\omega_h} &\leq 2\int_0^{2\pi}\int_1^{1+\eta} |u(1,\theta)|^2\,r\,dr\,d\theta + 2\int_0^{2\pi}\int_1^{1+\eta}\left|\int_1^r \partial_r\varphi\,ds\right|^2 r\,dr\,d\theta\\ &\leq C\left(\eta\,|\varphi|^2_{0,\partial\mathcal{O}} + \eta^2\,|\varphi|^2_{1,\omega_\eta}\right) \leq C\eta\,|\varphi|^2_{1,\Omega\setminus\overline{\mathcal{O}}}\,,\end{aligned}$$

from which we deduce the first estimate.

This same approach can be applied to $\partial_i\varphi$ for $\varphi \in H^2$. As φ is supposed to vanish over $\partial\Omega$, one has

$$|\partial_i\varphi| \leq C\,\|\nabla\varphi\|_{H^1(\Omega\setminus\overline{\mathcal{O}})} \leq C'\,|\varphi|^2_{2,\Omega\setminus\overline{\mathcal{O}}}\,,$$

which leads to the second estimate. As for the third one, simply notice that the boundary term (L^2 norm over $\partial\mathcal{O}$) vanishes in the equation above:

$$|\varphi|_{0,\omega_\eta} \leq \eta\,|\varphi|_{1,\omega_\eta} \leq \eta^{3/2}\,|\varphi|_{2,\omega_\eta}\,,$$

which ends the proof. ∎

Remark 4.3. The previous lemma extends straightforwardly to the case of any smooth inclusion (C^2 regularity of the boundary is sufficient) strongly included in a

domain Ω (for a detailed proof of a similar property, see [GLM06]) or to the case where the function is defined within the subdomain (in that case, ω_η is defined as an inner narrow band).

The last lemma quantifies how one can control the L^2 norm of the interpolate of a regular function on a triangle, by means of the L^2 norm and the H^2 seminorm of the function.

LEMMA 4.14. *There exists a constant C such that, for any regular triangle T (see below), for any $u \in H^2(T)$,*

$$\|I_h u\|^2_{L^2(T)} \leq C \left(\|u\|^2_{L^2(T)} + h^4 \, |u|^2_{2,T} \right).$$

By regular we mean that T runs over a set of triangles such that the flatness $\mathrm{diam}(T)/\rho_K$ *is bounded.*

Proof. The interpolation operator $I_h : H^2(T) \longrightarrow L^2(T)$ is continuous, and $|u|_{2,T}$ scales like $h/\rho_K^2 \approx 1/h$ whereas the L^2 norms scale like h. ∎

5. Additional examples, concluding remarks. The approach can be checked to be applicable to some standard situations, like the constraint to vanish in an inclusion $\mathcal{O} \subset\subset \Omega$ (see Example 2.2), as soon as H^1-penalty is used. The functional to minimize is then

$$J_\varepsilon(v) = \frac{1}{2} \int_\Omega |\nabla v|^2 - \int_\Omega f v + \frac{1}{2\varepsilon} \int_{\mathcal{O}} \left(u^2 + |\nabla u|^2 \right),$$

so that B identifies with the restriction operator from $H^1_0(\Omega)$ to $H^1(\mathcal{O})$. The discrete inf-sup condition, as well as the approximation properties, are essentially the same as in the case of an inclusion with infinite conductivity.

Another straightforward application of the abstract framework presented in section 3 is the numerical modeling of a rigid inclusion in a material which obeys Lamé's equations of linear elasticity. The penalized functional is then

$$J_\varepsilon(\mathbf{v}) = \frac{1}{2} \int_\Omega \mu \, |e(\mathbf{v})|^2 + \frac{1}{2} \int_\Omega \lambda \, |\nabla \cdot \mathbf{v}|^2 - \int_\Omega \mathbf{f} \cdot \mathbf{v} + \frac{1}{2\varepsilon} \int_{\mathcal{O}} |e(\mathbf{v})|^2,$$

where $e(\mathbf{v}) = \left(\nabla \mathbf{v} + (\nabla \mathbf{v})^T \right)/2$ is the strain tensor.

We conclude this section by some remarks on the proof itself and on possible extensions of this approach.

Remark 5.1 (conditioning issues). The fact that there is no need to choose ε too small (both errors balance for ε of the order of $\sqrt{h}$) is of particular importance in terms of conditioning. Indeed, considering the matrix A_h^ε resulting from the two-dimensional discrete minimization problem (4.4), it can be checked easily that its smallest eigenvalue scales like h^2, whereas its largest eigenvalue behaves like $1/\varepsilon$, leading to a condition number of the order of $1/\varepsilon h^2$. Following the ε-h balance suggested by the error estimates, the condition number finally scales like $1/h^{5/2}$, which compares reasonably to the standard $1/h^2$. Note also that some special fixed point algorithms, recently proposed in [BFM08], can be used to circumvent the problem of ill-conditioning.

Remark 5.2 (convergence in space). The poor rate of convergence in h is optimal for a uniform mesh, at least if we consider the H^1 error over all Ω. Indeed, as the solution is constant inside $\mathcal{O}$, nonconstant outside with a jump in the normal derivative, the error within each element intersecting $\partial\mathcal{O}$ is a $\mathcal{O}(1)$ in this L^∞ norm. By summing

up over all those triangles, which cover a zone whose measure scales like h, we end up with this $h^{1/2}$ error. Note that a better convergence could be expected, in theory, if one considers only the error in the domain of interest $\Omega \setminus \overline{\mathcal{O}}$, the question now being whether the bad convergence in the neighborhood of $\partial\mathcal{O}$ pollutes the overall approximation. Our feeling is that this pollution actually occurs, because nothing is done in the present approach to distinguish both sides of $\partial\mathcal{O}$, so that the method tends to balance the errors on both sides. An interesting way to give priority to the side of interest is proposed in [DP02] for a boundary penalty method; it consists in having the diffusion coefficient vanish within Ω. Note that other methods have been proposed to reach the optimal convergence rate on nonboundary fitted mesh (see [Mau01]), but they are less straightforward to implement.

The simplest way to improve the actual order of convergence is to carry out a local refinement strategy in the neighborhood of $\partial\mathcal{O}$, as proposed in [RAB07].

Remark 5.3 (nonregular domains). The method can be implemented straightforwardly to nonregular domains (e.g., with corners or cusps), but the numerical analysis presented here is no longer valid. In particular, the inf-sup condition established in Proposition 4.7 and approximation properties for u (see Proposition 4.8) may no longer hold. Notice that Propositions 2.9 and 2.10 do not require any regularity assumption, so that convergence can be established for some sequences (h, ε) tending to $(0, 0)$, but the optimal order of convergence is lost. Practical tests suggest a reasonably good behavior of the method is such situations, like in the case where $\mathcal{O}$ consists of two tangent discs (this situation is of special interest for practical applications in the context of fluid particle flows, when two particles are in contact; see, for example, [Lef07]).

Remark 5.4. Note that having ε go to 0 for any $h > 0$ leads to an estimate for a fictitious domain method (à la Glowinski, i.e., based on the use of Lagrange multipliers). In [GG95], an error estimate is obtained for such a method; it relies on two independent meshes for the primal and dual components of the solution (conditionally to some compatibility conditions between the sizes of the two meshes). We recover this estimate in the situation where the local mesh is simply the restriction of the covering mesh to the obstacle (to the reduced obstacle $\mathcal{O}_h$, to be more precise).

REFERENCES

[AR08] P. Angot and I. Ramière, *Convergence analysis of the Q1-finite element method for elliptic problems with non boundary-fitted meshes*, Internat. J. Numer. Methods Engrg., 75 (2008), pp. 1007–1052.

[Bab73] I. Babuška, *The finite element method with penalty*, Math. Comp., 27 (1973), pp. 221–228.

[BE07] E. Burman and A. Ern, *A continuous finite element method with face penalty to approximate Friedrichs' systems*, M2AN Math. Model. Numer. Anal., 41 (2007), pp. 55–76.

[BF91] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Ser. Comput. Math. 15, Springer-Verlag, New York, 1991.

[BFM08] T. T. C. Bui, P. Frey, and B. Maury, *Méthode du second membre modifié pour la gestion de rapports de viscosité importants dans le problème de Stokes bifluide*, C. R. Mécanique, 336 (2008), pp. 524–529.

[BHS03] R. Becker, P. Hansbo, and R. Stenberg, *A finite element method for domain decomposition with non-matching grids*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 209–225.

[DP02] S. Del Pino, *Une méthode d'éléments finis pour la résolution d'EDP dans des domaines décrits par géométrie constructive*, Ph.D. thesis, Université Pierre et Marie Curie, Paris, France, 2002.

[DPM07] S. Del Pino and B. Maury, *2d/3d turbine simulations with freefem*, in Numerical Analysis and Scientific Computing for PDEs and Their Challenging Applications, J. Haataja, R. Stenberg, J. Periaux, P. Raback, and P. Neittaanmaki, eds., CIMNE, Barcelona, Spain, 2008.

[FFp] freeFEM++; http://www.freefem.org/.

[GG95] V. Girault and R. Glowinski, *Error analysis of a fictitious domain method applied to a Dirichlet problem*, Japan J. Indust. Appl. Math., 12 (1995), pp. 487–514.

[GLM06] V. Girault, H. López, and B. Maury, *One time-step finite element discretization of the equation of motion of two-fluid flows*, Numer. Methods Partial Differential Equations, 22 (2006), pp. 680–707.

[GR79] V. Girault and P.-A. Raviart, *Finite Element Approximation of the Navier-Stokes Equations*, Lecture Notes in Math. 749, Springer-Verlag, Berlin, 1979.

[HH02] A. Hansbo and P. Hansbo, *An unfitted finite element method, based on Nitsche's method, for elliptic interface problems*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 5537–5552.

[JLM05] J. Janela, A. Lefebvre, and B. Maury, *A penalty method for the simulation of fluid-rigid body interaction*, in CEMRACS 2004—Mathematics and Applications to Biology and Medicine, ESAIM Proc. 14, EDP Sciences, Les Ulis, France, 2005, pp. 115–123.

[Lef07] A. Lefebvre, *Fluid-particle simulations with FreeFem++*, in Paris-Sud Working Group on Modelling and Scientific Computing 2006–2007, ESAIM Proc. 18, EDP Sciences, Les Ulis, France, 2007, pp. 120–132.

[Mau01] B. Maury, *A fat boundary method for the Poisson problem in a domain with holes*, J. Sci. Comput., 16 (2001), pp. 319–339.

[Nit71] J. Nitsche, *Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind*, Abh. Math. Sem. Univ. Hamburg, 36 (1971), pp. 9–15.

[PG02] T.-W. Pan and R. Glowinski, *Direct simulation of the motion of neutrally buoyant circular cylinders in plane Poiseuille flow*, J. Comput. Phys., 181 (2002), pp. 260–279.

[RAB07] I. Ramière, P. Angot, and M. Belliard, *A fictitious domain approach with spread interface for elliptic problems with general boundary conditions*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 766–781.

[RPVC05] T. N. Randrianarivelo, G. Pianet, S. Vincent, and J. P. Caltagirone, *Numerical modelling of solid particle motion using a new penalty method*, Internat. J. Numer. Methods Fluids, 47 (2005), pp. 1245–1251.

[SMSTT05] J. San Martín, J.-F. Scheid, T. Takahashi, and M. Tucsnak, *Convergence of the Lagrange–Galerkin method for the equations modelling the motion of a fluid-rigid system*, SIAM J. Numer. Anal., 43 (2005), pp. 1536–1571.