A PENALTY METHOD FOR SOME NONLINEAR VARIATIONAL OBSTACLE PROBLEMS*

HAYDEN SCHAEFFER †

Abstract. We formulate a penalty method for the obstacle problem associated with a nonlinear variational principle. It is proven that the solution to the relaxed variational problem (in both the continuous and discrete settings) is exact for finite parameter values above some calculable quantity. To solve the relaxed variational problem, an accelerated forward-backward method is used, which ensures convergence of the iterates, even when the Euler-Lagrange equation is degenerate and non-differentiable. Several nonlinear examples are presented, including quasi-linear equations, degenerate and singular elliptic operators, discontinuous obstacles, and a nonlinear two-phase membrane problem.

Key words. Nonlinear obstacle problem, sparsity, penalty method, variational methods.

AMS subject classifications. Primary: 35A15. Secondary: 35R35, 35J60

1. Introduction

The obstacle problem is a classical model for the equilibrium state of an elastic membrane which is displaced by a physical obstacle. Typically, these problems are modeled via a variational principle along with an inequality constraint that represents the physical obstacle. Obstacle problems arise from many physical systems including elastic-plastic torsion, phase transition, membrane-fluid interaction, and shallow ice sheets [40, 33, 39, 70].

One general form for the obstacle problem is:

$$\min_{u\geq\varphi}\int_\Omega L(x,u,\nabla u)\,\mathrm{dx}.$$

where $\Omega\subset\mathbb{R}^N$, φ is the obstacle, and L is the Lagrangian. The Euler-Lagrange equation for the constrained problem yields a variational inequality, and the boundary of the contact set $\{u=\varphi\}$ is a free boundary, see for example [40, 33]. Classical theory on the obstacle and related free boundary problems can be found in [19, 45, 17, 14, 16, 30, 15, 18, 54]. Some recent theory on the obstacle problem for fractional Laplacian include [65, 20] and integro-differential operators [21].

Numerical methods for the obstacle problem typically focus on solving the variational inequality that appears from the first variation of the constrained problem. In [36], the variational inequality is solved by using a multigrid algorithm. In [37], a piecewise linear finite element method with a multilevel preconditioner is used to solve the problem. A predictor-corrector continuation method to solve the variational inequality is presented in [28]. Another piecewise linear finite element solver for elliptic variational inequalities using monotone multigrid methods can be found in [41, 42]. In [26], a piecewise linear finite element method applied to elliptic obstacle problems is constructed with the use of positivity preserving interpolation. In [43], *a posteriori* error estimates for self-adjoint elliptic obstacle problems are presented. The authors included an extra term, the sum of local residuals on certain nodes, to measure the discrepancy associated with the discrete free boundary.

Another popular method for solving the variational inequality is via the Schwarz domain decomposition, for example [67, 3, 2]. For the variational inequality related to

*H.S. acknowledges the support of AFOSR, FA9550-17-1-0125

† Department of Mathematical Sciences, Wean Hall 6113, Carnegie Mellon University, Pittsburgh, PA 15213, (schaeffer@cmu.edu)

the one-body contact problem, two error estimators are proposed and analyzed in [71]. In [27], the authors propose an adaptive finite element method using bisection grids for Poisson's equation with obstacles. It is shown that the method yields an (almost) uniform convergence rate in terms of the objective function. An alternative approach is presented in [46], which uses the level set method to capture the free boundary (the contact set $\{u=\varphi\}$) thereby avoiding the variational inequality. The Euler-Lagrange equation for the clamped plates model results in biharmonic equations, making the finite element discretization more challenging.

The obstacle problem can also be relaxed to an unconstrained problem with the addition of a penalty term. One example can be found in [64], which uses an L^2-like penalty parameterized with a coefficient equal to $\frac{1}{\epsilon}$. Error rates are established in relation to the parameter ϵ (which is coupled to the grid spacing h) and requires that the penalty parameter be $\mathcal{O}(h^{-2})$. Therefore, the relaxed problem is not exact even in the smooth obstacle case. In [31], a finite element approximation with an L^2 penalty is analyzed for the Laplacian obstacle problem. As expected the relaxed problem, using an L^2 penalty, requires that the parameter goes to infinity in order for the solution to be exact. In [35], the obstacle constraint is included in the objective function by using a Lagrange multiplier, and the resulting saddle point problem is smoothed in order to be solved numerically. In [7], two numerical schemes are proposed for the two-phase membrane problem, which involves non-differentiable terms similar to a double-penalty. In one of their proposed methods, the non-differentiable terms are relaxed using a smooth approximation.

Similar to the penalty methods, we relax the constrained minimization by introducing a term in the energy which encourages the solution to satisfy the constraint. In this work, we focus on the nonlinear (possibly degenerate) elliptic variational problem. For related theory and numerics for the linear elliptic case, see [69]. In a related work [72], the L^1 penalty from [69] was applied to the obstacle problem associated with minimal surfaces. The penalty used in [69] is related to the L^1 norm of the difference between the obstacle and the solution. Unlike the penalties found in the literature, the L^1-like penalty is non-differentiable; however, it is exact in the sense that the solution of the relaxed problem agrees with the constrained minimization.

Variational methods using the L^1 and related norms for scientific computing and PDEs have appeared recently in the literature, for example [58, 38] for multiscale PDEs, [48, 52, 53] for quantum models, and in [10, 1, 12, 13, 56] for low-rank approximations to dynamic systems. The difficulty in the theory is due to the fact that the L^1 term yields a subdifferential in the Euler-Lagrange equation. For some related theoretical results on the interaction between PDEs and subgradients of the L^1 norm, see for example [22, 8, 9]. Also, recently L^1 and L^0 penalties have been used for sparse recovery of governing equations from dynamic data [11, 68, 59, 55, 60, 61, 62].

1.1. Contribution of this work. In this work, we present a formulation of the obstacle problem for nonlinear variational models using an L^1 penalty from [69]. Specifically, we generalize the method to non-smooth variational problems and investigate the applicability of L^1 penalties to degenerate elliptic equations. The L^1 penalty formulation and motivation is given in Section 2. In Section 3, solutions of the relaxed variational problem are shown to be exact for penalty parameters larger than a known (finite) lower bound. A characterization of the minimizers with respect to the penalty parameter is done for the continuous and discrete cases. This shows some of the advantages of this approach over other methods; in particular, the penalty parameter may be fixed relative to the grid spacing and the initial data does not need to be in the

feasible set. In Section 4, a numerical scheme for solving nonlinear variational problems is presented with results shown in Section 5. The algorithm and approach are robust and easy to implement. It is worth noting that the proximal operators related to the various L^1 penalties are derived in the Appendix.

2. Variational Problem Consider the following functional, $F(u)$, defined over the set of functions $C=\{u\in B,\ u=g \text{ on } \partial\Omega\}$ (where B is an appropriate Banach space and $\Omega\subset\mathbb{R}^n$):

$$F(u):=\int_\Omega L(x,u,\nabla u)\,\mathrm{dx}. \tag{2.1}$$

The Lagrangian $L(x,u,z)$ is assumed to be jointly convex in (u,z), proper, and lower semi-continuous. We write the obstacle problem as a constrained minimization:

$$u^*=\operatorname*{argmin}_{u\in K} F(u),$$

where $K=\{u\geq\varphi, u\in C\}$. For simplicity, it is assumed that the obstacle remains below the data on the boundary, *i.e.* $\varphi\leq g$ on $\partial\Omega$. Let DF be the derivative associated with the Gâteaux differentiable functional F, *i.e.*

$$\frac{d}{d\epsilon}F(u+\epsilon v)\big|_{\epsilon=0}=\langle DF(u),v\rangle.$$

Then the minimization problem is equivalent to finding $u^*\in K$ such that:

$$\langle DF(u^*),u^*-v\rangle\geq 0,\ \forall v\in K. \tag{2.2}$$

For more details on the variational inequality, Equation 2.2, see for example [40].

We relax Equation 2.1 by introducing a continuous (but non-differentiable) penalty term:

$$\begin{aligned}E_\mu(u):=F(u)+\mu P(u)&=F(u)+\mu||(\varphi-u)_+||_{L^1(\Omega)}\\&=\int_\Omega L(x,u,\nabla u)\,\mathrm{dx}+\mu\int_\Omega(\varphi-u)_+\,\mathrm{dx}\end{aligned} \tag{2.3}$$

where $\mu>0$, $P(u)=||(\varphi-u)_+||_{L^1(\Omega)}$ is the penalty term, and $x_+=\max(x,0)$. The minimizer of Equation 2.3 is denoted by

$$u_\mu:=\operatorname*{argmin}_{u\in C} E_\mu(u).$$

We refer to Equation 2.3 as the relaxed problem, in the sense that the obstacle constraint has been removed and incorporated into the functional (thereby "relaxing" the feasible set). Since the penalty term has a set-valued subdifferential, the Euler-Lagrange equation of Equation 2.3 is a differential inclusion:

$$-DF(u)\in\mu\partial(\varphi-u)_+. \tag{2.4}$$

The subdifferential on the RHS of Equation 2.4 can be identified as follows:

$$\partial(\varphi-u)_+=\begin{cases}0, \text{ if } \varphi<u,\\ [-1,0], \text{ if } \varphi=u,\\ -1, \text{ if } \varphi>u,\end{cases}$$

Note that the proposed relaxation is consistent in the sense that: when $u \in K$, the differential inclusion (and the corresponding variational problem) is equivalent to the obstacle problem, Equation 2.1. In the region where the obstacle constraint is not satisfied, the penalty term contributes a constant force of $-\mu$ to the differential equation.

Although the relaxed problem is non-differentiable, it is of practical interest since it has a simple proximal operator [5, 24, 34] (also known as the resolvent $(I+\lambda\partial P)^{-1}$):

$$\begin{aligned} u = \operatorname{prox}_{\lambda P} v &= (I+\lambda\partial P)^{-1}(v) \\ &= \operatorname*{argmin}_{u\in B} \int_\Omega (\varphi-u)_+ + \frac{1}{2\lambda}(u-v)^2 \mathrm{dx} \\ &= \begin{cases} v+\lambda, & \text{if } \varphi > v+\lambda \\ \varphi, & \text{if } v \le \varphi \le v+\lambda \\ v, & \text{if } \varphi < v \end{cases} \end{aligned}$$

Since the proximal operator has a simple, explicit form, this lends itself well to many convex optimization algorithms, which typically rely on applying the proximal operator within an iterative scheme (see Section 4).

2.1. Example

In Section 3, we will show that if $\mu \ge ||DF(\varphi)||_{L^\infty}$, then the minimizer of the penalty problem is exact, under some assumptions on the unconstrained problem. As a motivating example, consider the simple obstacle problem:

$$\min_{u\ge\frac{1}{2}} \int_{-1}^{1} \frac{1}{4}\left|u-x^2\right|^4 \mathrm{dx}.$$

The solution to the unconstrained problem is $u(x)=x^2$, and the solution to the constrained problem is:

$$u^*(x) = \begin{cases} x^2, & \text{if } |x| \ge \frac{1}{\sqrt{2}} \\ \frac{1}{2}, & \text{if } |x| < \frac{1}{\sqrt{2}}. \end{cases}$$

Relaxing the problem by introducing the penalty yields the unconstrained minimization:

$$\min_u \int_{-1}^{1} \frac{1}{4}\left|u-x^2\right|^4 + \mu\left(\frac{1}{2}-u\right)_+ \mathrm{dx}. \tag{2.5}$$

To derive the minimizer of Equation 2.5, denoted by u_μ, we consider three regimes depending on the constraint. In the first case, consider the points in the set $\left\{u(x) > \frac{1}{2}\right\}$, then the derivative of Equation 2.5 is zero only if $u(x)=x^2$. In the second case, consider the points $\left\{u(x) < \frac{1}{2}\right\}$, then setting the derivative of Equation 2.5 to zero yields:

$$\left(u-x^2\right)^3 - \mu = 0$$

The solution in this regime is equal to $u(x)=x^2+\mu^{\frac{1}{3}}$. The last case to consider is the contact set where $u(x)=\frac{1}{2}$, which occurs when $\sqrt{\left(\frac{1}{2}-\mu^{\frac{1}{3}}\right)_+} < |x| < \frac{1}{\sqrt{2}}$. Altogether the

solution to the relaxed problem is:

$$u_\mu^*(x) = \begin{cases} x^2, & \text{if } |x| \geq \frac{1}{\sqrt{2}} \\ \frac{1}{2}, & \text{if } \sqrt{\left(\frac{1}{2} - \mu^{\frac{1}{3}}\right)_+} < |x| < \frac{1}{\sqrt{2}} \\ x^2 + \mu^{\frac{1}{3}}, & \text{if } \sqrt{\left(\frac{1}{2} - \mu^{\frac{1}{3}}\right)_+} \geq |x|. \end{cases}$$

Examples for various μ values are provided in Figure 2.1. Comparing the relaxed u_μ and exact u^* solutions, we see that the region in which $u_\mu(x) \neq u^*(x)$ is given by the set $\left\{|x| < \sqrt{(\frac{1}{2} - \mu^{\frac{1}{3}})_+}\right\}$, which vanishes when $\mu \geq \frac{1}{8}$. By examining the bound for μ:

$$\mu \geq \left\| DF\left(\frac{1}{2}\right) \right\|_{L^\infty(-1,1)} = \left\| \left(\frac{1}{2} - x^2\right)^3 \right\|_{L^\infty(-1,1)} = \frac{1}{8},$$

we can conclude that the bound for μ is sharp.

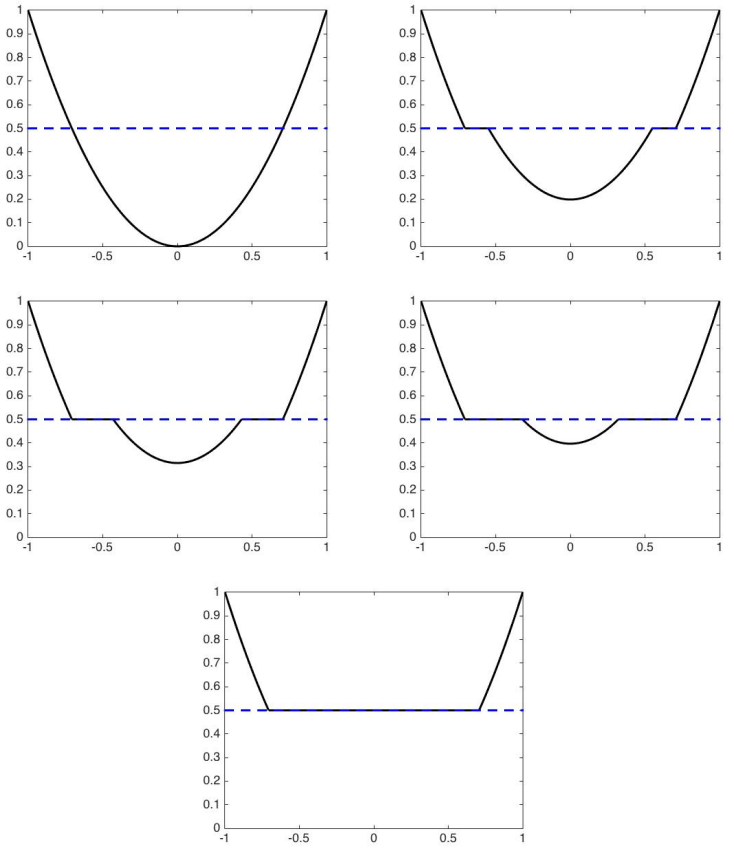

FIG. 2.1. *Solution u_μ to Equation 2.5 for various μ, in order: $\mu = 0$, $\frac{1}{128}$, $\frac{1}{32}$, $\frac{1}{16}$, and $\frac{1}{8}$.*

3. Characterization of minimizers

Consider the following relaxed variational problem:

$$E_\mu(u) := \int_\Omega L(x,u,\nabla u)\,\mathrm{dx} + \mu \int_\Omega (\varphi - u)_+\,\mathrm{dx}. \tag{3.1}$$

We will show that there exists a finite value for μ, depending on the obstacle φ and DF, such that the minimizer of Equation 3.1 agrees with the minimizer of Equation 2.1 (over $u \in K$). Before doing so, we will assume several properties of the obstacle problem hold. First, assume that:

(A1) Both the constrained problem, Equation 2.1, and the relaxed problem, Equation 3.1, have unique minimizers, respectively.

(A2) The functional, $F(u)$, is Gâteaux differentiable and convex, with the associated derivative denoted by $DF(u)$, *i.e.*

$$\frac{d}{d\epsilon} F(u+\epsilon v)\big|_{\epsilon=0} = \langle DF(u), v\rangle.$$

where the inner product is the dual pairing of $L^\infty(\Omega)$ and $L^1(\Omega)$.

The second assumption ensures that for any two functions $v, w \in C \subset L^1(\Omega)$, we have:

$$F(v) - F(w) - \langle DF(w), v - w\rangle \geq 0.$$

The arguments below could include the case where DF is set-valued; however, for simplicity, we assume that it is a singleton. This is not essential to the proof. In addition to (A1-A2), we assume that $C = \{u \in B,\ u = g \text{ on } \partial\Omega \subset \mathbb{R}^n\}$ where

(A3) B is a Banach space with the property: if $u \in B$, then $u_+ \in B$.

Many Banach spaces of functions satisfy (A3). For example, let Ω be a Lipschitz and open set in $\mathbb{R}^n$, then (A3) holds for the following spaces:

- $L^p(\Omega)$ and $W^{1,p}(\Omega)$ for $p \in [1,\infty]$, see [66],
- $H^s(\Omega)$ for $s < \frac{3}{2}$, see [6],
- Sobolev Spaces $W^{s,p}(\Omega)$ for $p \in [1,\infty)$ and $s \in [1, 1+p^{-1})$, see [6],
- Besov Spaces $B^s_{q,p}(\Omega)$ for $p,q \in [1,\infty)$ and $s \in [1, 1+p^{-1})$, see [57],
- Bounded Variation $BV(\Omega)$ and Bounded Hessian $BH(\Omega)$, see [57].

Note that (A3) does not hold for $W^{2,p}(\Omega)$ due to jumps in the gradient.

THEOREM 3.1. *Let u^* be the unique minimizer to the constrained problem, Equation 2.1, and u_μ be the unique minimizer to the relaxed problem, Equation 3.1. Suppose that assumptions (A1-A3) hold and that DF is bounded in $L^\infty(\Omega)$ at $\varphi \in L^1(\Omega)$. If $\mu \geq \|DF(\varphi)\|_\infty$, then $u_\mu = u^*$.*

Before detailing the proof, it is important to note that the assumption of smoothness, *i.e.* that $\varphi \in L^1(\Omega)$ and that $DF(\varphi) \in L^\infty(\Omega)$ is essential for this argument. However, it may be possible to provide a more general result.

Proof. For all $v \in C$, define the auxiliary function:

$$w := v + (\varphi - v)_+ = \begin{cases} v, & \text{if } v \geq \varphi \\ \varphi, & \text{if } v < \varphi. \end{cases}$$

By (A3), since $v, \varphi \in C$, we have that $w \in C$ and by construction $w \geq \varphi$. Therefore, the relaxed functional is simply:

$$E_\mu(w) = F(w) + \mu \int_\Omega (\varphi - w)_+\,\mathrm{dx} = F(w).$$

By (A2), since F is Gâteaux differentiable and convex, we have that:

$$F(v) - F(w) - \langle DF(w), v - w \rangle \geq 0 \tag{3.2}$$

holds. Therefore, the functional $E_\mu(u)$ can be bounded from above using Equation 3.2:

$$\begin{aligned} E_\mu(w) = F(w) &\leq F(v) - \langle DF(w), v - w \rangle \\ &= F(v) - \langle DF(w), -(\varphi - v)_+ \rangle \\ &= F(v) + \langle DF(w), (\varphi - v)_+ \rangle \end{aligned}$$

where the equality holds since $v - w = -(\varphi - v)_+$. We must remove the dependence of the derivative, DF, on the unknown function, w. To do so, we note that the second term is non-zero, $(\varphi - v)_+ \neq 0$, precisely when $\varphi > v$. We can substitute $w = v + (\varphi - v)_+ = \varphi$ within the dual pairing. Therefore,

$$\begin{aligned} E_\mu(w) &\leq F(v) + \langle DF(\varphi), (\varphi - v)_+ \rangle \\ &\leq F(v) + ||DF(\varphi)||_{L^\infty(\Omega)} \int_\Omega (\varphi - v)_+ \, \mathrm{dx} \\ &\leq F(v) + \mu \int_\Omega (\varphi - v)_+ \, \mathrm{dx} \\ &= E_\mu(v). \end{aligned}$$

The inequality above holds for all $v \in C$ as long as $\mu \geq ||DF(\varphi)||_{L^\infty(\Omega)}$. Taking $v = u_\mu$, we have:

$$E_\mu(u_\mu + (\varphi - u_\mu)_+) \leq E_\mu(u_\mu).$$

Since u_μ is the unique minimizer, then $(\varphi - u_\mu)_+ = 0$ must hold. Therefore, the minimizer to the relaxed problem satisfies the constraint, $u_\mu \geq \varphi$, and we have exactness in the sense that $u_\mu = u^*$.

□

Theorem 3.1 shows that the relaxed problem is exact for finite values of the parameter, which can be determined completely from the given information. A similar result holds for the discrete problem. Suppose we discretize the problem by taking N uniformly distributed points with grid spacing h, contained in the domain $\Omega \subset \mathbb{R}^n$. For simplicity, we take piecewise linear interpolants for all discrete functions. The discretization of Equation 2.1 is denoted by $F^h(u^h)$ and the minimizer is denoted by:

$$u^{*,h} = \underset{u^h \in K^h}{\operatorname{argmin}} \; F^h(u^h) \tag{3.3}$$

where $K^h = \left\{ u^h \geq \varphi^h, u \in C^h \right\}$, C^h is an appropriate discretization of the continuous Banach space, and $h > 0$ (the grid is taken to be uniform). The discrete obstacle, denoted by φ^h, is simply the linear interpolant of the continuous obstacle restricted to the discrete grid. Let G^h be a consistent, stable, and monotone numerical approximation to DF.

The relaxed problem, Equation 3.1, is discretized by:

$$E^h_\mu(u^h) := F^h(u^h) + \mu h^n \sum_{j=1}^{N} (\varphi^h_j - u^h_j)_+ \tag{3.4}$$

where the penalty is approximated using piecewise constant quadrature over a uniform grid indexed by j. We denote the minimizer of Equation 3.4 by:

$$u_\mu^h = \underset{u^h \in C^h}{\operatorname{argmin}}\ E_\mu^h(u^h)$$

With these assumptions, we can guarantee a similar result for the discrete obstacle problem. The theorem below provides the main condition needed for the numerical experiments in this work.

THEOREM 3.2. *Suppose u_μ^h and $u^{*,h}$ are the unique minimizers to the relaxed and constrained discrete obstacle problems, respectively. Let $F^h : \mathbb{R}^N \to \mathbb{R}$ be a convex function whose derivative G^h is monotone in the sense that for any $v^h, w^h \in \mathbb{R}^N$ we have:*

$$\left\langle G^h[w^h] - G^h[v^h], w^h - v^h \right\rangle_h \geq 0$$

where $\langle \cdot, \cdot \rangle_h$ is the discrete inner product. If

$$\mu^h \geq \|G^h[\varphi^h]\|_{\infty,h} := \max_{j=1,2,\ldots,N} \left| G_j^h[\varphi^h] \right|,$$

then $u_\mu^h = u^{*,h}$.

Proof. For all $v^h \in C^h$, define the auxiliary function:

$$w^h := v^h + (\varphi^h - v^h)_+.$$

The energy at w^h reduces to:

$$E_\mu^h(w^h) = F^h(w^h) + \mu h^n \sum_{j=1}^N (\varphi_j^h - v_j^h)_+ = F^h(w^h)$$

By the assumptions that F^h is convex on $\mathbb{R}^N$ and has a monotone derivative, we have the discrete inequality:

$$F^h(v^h) - F^h(w^h) - \left\langle G^h[w^h], v^h - w^h \right\rangle_h \geq 0$$

Therefore, a similar bound for the relaxed function holds:

$$\begin{aligned} E_\mu^h(w^h) &\leq F^h(v^h) - \left\langle G^h[w^h], v^h - w^h \right\rangle_h \\ &= F^h(v^h) + \left\langle G^h[w^h], (\varphi^h - v^h)_+ \right\rangle_h \end{aligned}$$

Since $(\varphi^h - v^h)_+ \neq 0$ only when $\varphi^h \geq v^h$, we have $w^h = v^h + (\varphi^h - v^h)_+ = \varphi^h$ within the discrete inner product. Therefore,

$$\begin{aligned} E_\mu^h(w^h) &\leq F^h(v^h) + \left\langle G^h[\varphi^h], (\varphi^h - v^h)_+ \right\rangle_h \\ &= F^h(v^h) + h^n \sum_{j=1}^N G_j^h[\varphi^h] (\varphi_j^h - v_j^h)_+ \\ &\leq F^h(v^h) + \left(\max_j \left| G_j^h[\varphi^h] \right| \right) \left(h^n \sum_{j=1}^N (\varphi_j^h - v_j^h)_+ \right) \\ &\leq F^h(v^h) + \mu h^n \sum_{j=1}^N (\varphi_j^h - v_j^h)_+ \\ &= E_\mu^h(v^h) \end{aligned}$$

If we take $v^h = u_\mu^h$, then the following inequality holds:

$$E_\mu^h(u_\mu^h + (\varphi^h - u_\mu^h)_+) \leq E_\mu(u_\mu^h).$$

Since u_μ^h is the unique minimizer of E_μ^h, $(\varphi^h - u_\mu^h)_+ = 0$. Therefore the constraint holds: $u_\mu^h \geq \varphi^h$ and the relaxed minimizer is exact $u_\mu^h = u^{*,h}$.
□

Since Theorem 3.2 requires applying G^h to obstacle φ^h, the obstacle must be sufficiently smooth. In the case where the obstacle is discontinuous, the parameter μ^h will grow as $h \to 0$. In particular, let G^h be a Lipschitz function with Lipschitz constant L^h and let the obstacle and solution be bounded by k in the infinity norm, then it can be shown that as long as $\mu^h \geq 2kL^h$, the relaxed minimizer u_μ^h is exact. Note that L^h will typically scale like $\frac{1}{h^2}$. The results presented here in the finite dimension case are similar to those found in [47, 32].

4. Numerical Scheme

For the numerical scheme, we use the convexity of the discrete energy, along with the simple proximal operator to construct an iterative method to solve the discrete optimization. The derivative of F^h is approximated explicitly, while the penalty is approximated implicitly, leading to what is known as the Forward-Backward algorithm. For each problem below we provide a monotone (and convergent) approximation to G^h (the derivative of F^h). Monotone discretization of partial differential equations are typically non-differentiable [51], therefore treating this term in an explicit manner is expected. However, if one has an implicit solver for G^h, then it is possible to solve the optimization using a fully implicit method, for example using the Douglas-Rachford algorithm. The algorithm used here is stated below.

Forward-Backward Algorithm
Given: u^0 and parameters μ and dt

while $||u^n - u^{n-1}||_\infty > tol$ **do**

$u^{n+1} = \text{prox}_{\mu\,\text{dt}\,P}(u^n - \text{dt}\,G^h(u^n))$

end while

It is known that the foward-backward algorithms decreases the energy at a rate of $\mathcal{O}(n^{-1})$ (where n is the iteration index), see [50]. A slight modification of the algorithm is typically preformed by introducing an extrapolation step with extrapolation parameter $\alpha^n = \dfrac{t_n - 1}{t_{n+1}}$, where $t_n = \frac{n+a-1}{a}$, see [50, 5]. As long as $a > 2$ and the energy is convex, lower semi-continuous, and proper, then the sequence of iterates u^n generated by the accelerated foward-backward algorithm decreases the energy by:

$$E_\mu^h(u^n) - E_\mu^h(u_\mu^{*,h}) \leq \frac{c}{n^2}$$

and the discrete iterates converge weakly to the minimizer [25]. In both algorithms, the time-step is taken to be $0 < \text{dt} < \frac{1}{L^h}$, where L^h is the Lipschitz constant of G^h.

Accelerated Forward-Backward Algorithm

Given: u^0 and parameters μ, tol, and, a dt

while $||u^n - u^{n-1}||_\infty > tol$ **do**

$\alpha^n = \dfrac{t_n - 1}{t_{n+1}}$, where $t_n = \frac{n+a-1}{a}$

$v^n = u^n + \alpha^n(u^n - u^{n-1})$

$u^{n+1} = \text{prox}_{\mu\,\text{dt}\,P}(v^n - \text{dt}\,G^h(v^n))$

end while

The only parameter that is needed is a; however, any value above 2 produces convergent results. Other possible algorithms that could be used to solve the optimization include [24, 49, 34, 23]. Note that the accelerated algorithm was applied to degenerate elliptic problems in [63].

5. Results In each of the examples presented below, the value for μ^h is chosen to ensure that it satisfies the bound for exactness from Theorems 3.1 and 3.2. The time-step, dt, is chosen to be less than the Lipschitz constant of G^h. We also provide a monotone and convergent discretization for G^h. The accelerated forward-backward method terminates when the maximum difference between two consecutive iterates is less than a set tolerance, denoted tol, which will be specified for each example. In all of the examples presented here, the computational time is on the order of a few seconds or less.

5.1. Semi-Linear Problem For a semi-linear example, we use the following functional:

$$F_1(u) = \int_\Omega \frac{1}{2}|\nabla u|^2 + \frac{1}{4}u^4 + u f \,\text{dx}$$

which is approximated by using forward differences and piecewise constant quadrature:

$$F_1^h(u) = \sum_{i,j} \frac{1}{2}\,|u_{i+1,j} - u_{i,j}|^2 + \frac{1}{2}\,|u_{i,j+1} - u_{i,j}|^2 + \frac{h^2}{4}u_{i,j}^4 + h^2\, u_{i,j}\, f_{i,j}$$

The derivative of F_1^h is:

$$G_1^h(u) = -\frac{1}{h^2}\,(u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1} - 4u_{i,j}) + u_{i,j}^3 + f_{i,j}$$

which is consistent with the continuous equation $DF_1(u) = -\Delta u + u^3 + f$.

Example 1: The first example is solved over the domain $[-1,1]^2$, with obstacle $\varphi = 5e^{-5(x^2+y^2)}$, forcing term $f = 1$, and the boundary condition $u = 0$. The penalty parameter is set to $\mu^h = 250$, the tolerance is set to $tol = 6 \times 10^{-7}$, and the time-step is set to $\text{dt} = \frac{h^2}{10}$. The over-relaxation parameter for the algorithm is assigned as $a = 5$. The calculations are done using various grid sizes, and the computed solution using $h = 0.0078$ (grid size is 256 by 256) is shown in Figure 5.1 along with a plot of the *log* of the maximum difference between two consecutive iterates (which is used as the stopping criterion for the iterative algorithm). In Table 5.1, the accuracy of the discretization applied to this problem is found to be approximately $\mathcal{O}(h^{1.02})$.

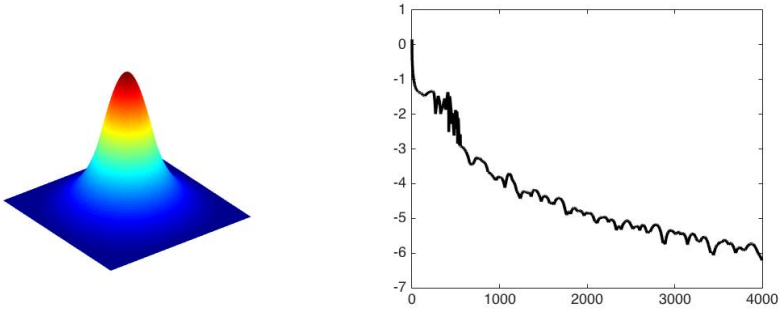

FIG. 5.1. *Left: The numerical solution for Example 1 with* $h=0.0078$ *(grid size is* 256 *by* 256*),* $\mu=250$*,* $tol=6\times10^{-7}$*, obstacle* $\varphi=5e^{-5(x^2+y^2)}$*, and* $f=1$*. Right: The log of the maximum difference between consecutive iterates (i.e* $\log_{10}(||u^n-u^{n-1}||_\infty)$*).*

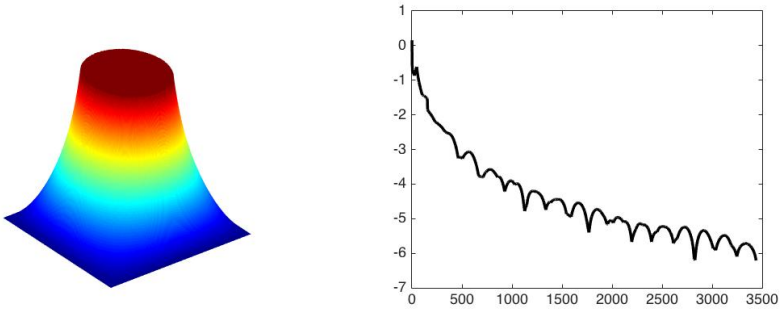

FIG. 5.2. *Left: The numerical solution for Example 2 on a grid of* $h=0.0078$ *(grid size is 256 by 256),* $\mu=5000$*,* $tol=6\times10^{-7}$*, obstacle* $\varphi=5\chi_{x^2+0.75y^2\leq0.25}$*, and* $f=1$*. Right: The log of the maximum difference between consecutive iterates (i.e* $\log_{10}(||u^n-u^{n-1}||_\infty)$*).*

Example 2: For the second example, we solve the problem on $[-1,1]^2$, setting the obstacle to $\varphi=5\chi_{x^2+0.75y^2\leq0.25}$. Since the obstacle is discontinuous, the parameter μ^h grows unbounded as $h\to0^+$. To handle this case, the penalty parameter is set to $\mu=5000$ to ensure that the parameter remains above the discrete estimate (see Theorem 3.2) for all grid sizes tested here. The forcing term and boundary condition are both set to zero, the tolerance $tol=6\times10^{-7}$, and the time-step $\mathrm{dt}=\frac{h^2}{10}$. The over-relaxation parameter for the case is set to $a=5$. In Figure 5.2, the numerical solution

h	0.1333	0.0645	0.0317	0.0157	0.0078
Error Ratio in L^∞	1.0564	1.0477	0.9979	1.0030	1.0039

TABLE 5.1. *Error analysis for Example 1. The error ratio* $=\log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu=250$*,* $tol=\frac{h^2}{100}$*,* $dt=\frac{h^2}{10}$*, and* $a=5$*. The order of accuracy for this example is approximately* $\mathcal{O}(h^{1.02})$*.*

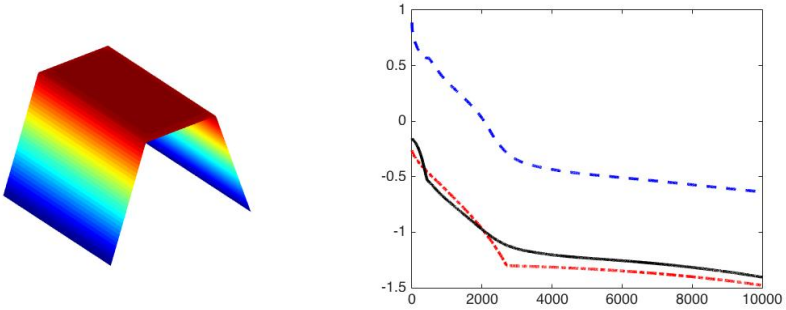

FIG. 5.3. *The numerical solution for Example 3 with* $h=0.0156$ *(grid size is* 128 *by* 128*),* $p=4$*,* $\mu=1000$*,* $tol=\frac{h^p}{250}$*, and the obstacle is* $\varphi=\chi_{|x|\leq 0.5}$*. Right: A log plot of the error in* L^∞ *(red),* L^2 *(black), and* H^1 *(blue) between the exact solution and the iterates for a fixed grid size.*

is graphed on a 256 by 256 sized grid along with the *log* of the maximum difference between consecutive iterates. The oscillatory behavior found in the plot of the difference between consecutive iterates is directly related to the extrapolation used in the iterative scheme. The accuracy of the discretization applied to this problem is found to be approximately $\mathcal{O}(h^{0.95})$, see Table 5.2. This example shows that even when the obstacle is discontinuous, this approach is applicable.

5.2. Quasi-Linear Problem

For a quasi-linear example, we use the anisotropic $W^{1,p}$ semi-norm:

$$F_2(u)=\int_\Omega \frac{1}{p}|\nabla u|_1^p + u f \,\mathrm{dx}$$

for $p>1$ and for a given force f. The functional F_2 is approximated using half-step centered differences and piecewise constant quadrature:

$$F_2^h(u)=\sum_{i,j}\frac{h^{p-2}}{p}|u_{i+\frac{1}{2},j}-u_{i-\frac{1}{2},j}|^p+\frac{h^{p-2}}{p}|u_{i,j+\frac{1}{2}}-u_{i,j-\frac{1}{2}}|^p+h^2 u_{i,j} f_{i,j}$$

h	0.0317	0.0157	0.0078	0.0039
Error Ratio in L^∞	0.7438	0.8100	1.0576	1.2056

TABLE 5.2. *Error analysis for Example 2. The error ratio* $=\log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu=5000$*,* $tol=\frac{h^2}{100}$*,* $dt=\frac{h^2}{10}$*, and* $a=5$*. The order of accuracy for this example is approximately* $\mathcal{O}(h^{0.95})$*.*

h	0.2857	0.1333	0.0645	0.0317	0.0157	0.0078
Error in L^∞	0.0733	0.0509	0.0286	0.0151	0.0077	0.0039

TABLE 5.3. *Error analysis for Example 3. The error* $=||u^h-u^h_{exact}||_\infty$ *for several grid sizes is examined. The parameters are fixed at* $p=4$*,* $tol=\frac{h^p}{250}$*,* $dt=\frac{h^p}{10}$*, and* $a=15$*. The order of accuracy for this example is approximately* $\mathcal{O}(h^{0.83})$*.*

The corresponding derivative is:

$$G_2^h(u) = -\frac{1}{h^p}|u_{i+1,j}-u_{i,j}|^{p-2}(u_{i+1,j}-u_{i,j}) - \frac{1}{h^p}|u_{i-1,j}-u_{i,j}|^{p-2}(u_{i-1,j}-u_{i,j})$$
$$-\frac{1}{h^p}|u_{i,j+1}-u_{i,j}|^{p-2}(u_{i,j+1}-u_{i,j}) - \frac{1}{h^p}|u_{i,j-1}-u_{i,j}|^{p-2}(u_{i,j-1}-u_{i,j}) + f_{i,j}$$

The equation above is consistent with the continuous derivative:

$$DF_2(u) = -(|u_x|^{p-2}u_x)_x - (|u_y|^{p-2}u_y)_y + f$$

which is known as the anisotropic p-Laplacian, [29]. The discrete equation is monotone, consistent, and stable (thus convergent by [4, 44, 51]). The obstacle problem associated with the p-Lapace's equation is related to shallow ice sheet models, see [39, 70].

Example 3: In Figure 5.3, we solve the $p=4$ problem on the domain $[0,2]^2$, setting the obstacle to the indicator function $\varphi = \chi_{|x|\le 0.5}$ and the forcing term $f=1$. It can be shown that the exact solution is:

$$u^*(x,y) \approx \begin{cases} \frac{3}{4}|x+7.75086|^{\frac{4}{3}} - 11.50434, & \text{if } x < 0.5 \\ 1, & \text{if } 0.5 \le x \le 1.5 \\ \frac{3}{4}|-x+9.75086|^{\frac{4}{3}} - 11.50434, & \text{if } x > 1.5. \end{cases}$$

with the corresponding boundary condition. The penalty parameter is set to $\mu = 1000$, with the same argument as in Example 2. The tolerance is set to $tol = \frac{h^4}{250}$ and the time-step is set to $\mathrm{dt} = \frac{h^4}{10}$. The over-relaxation parameter for the algorithm is set to $a = 15$. In Figure 5.4, the numerical solution is computed for a 256 by 256 sized grid along with a plot of the L^∞ (red, dotted), L^2 (black, solid), and H^1 (blue, lined) errors between the iterates and the exact minimizer u^*. Using the exact minimizer as a reference, the order of accuracy is estimated to be $\mathcal{O}(h^{0.83})$, see Table 5.3. This order of accuracy is expected, since the Euler-Lagrange equation contains a degenerate second order elliptic operator. It is important to note that this problem poses some difficulties. The p-Laplacian is a degenerate elliptic operator and thus a monotone numerical scheme is needed for convergence (as $h \to 0^+$). The obstacle is discontinuous, which can effect the regularity of the solution. However, the proposed method applied to this example still converges to the exact solution and achieves (nearly) first order accuracy.

Example 4: For a second degenerate example, we solve the $p=3$ problem on $[0,2]^2$, setting the obstacle to $\varphi = 2e^{-15((x-1)^2+(y-1)^2)}$. The force and boundary conditions are both set to zero. The parameters are fixed at: $\mu = 500$, $tol = \frac{h^3}{100}$, and $\mathrm{dt} = \frac{h^3}{10}$. The over-relaxation parameter for the algorithm is set to $a = 15$. In Figure 5.4, the numerical solution is plotted on a 256 by 256 grid along with the *log* of the maximum difference between consecutive iterates. The oscillatory behavior found once again is due to the over-relaxation term over-estimating the descent direction in the iterative algorithm. The accuracy of the numerical scheme is estimated to be $\mathcal{O}(h^{1.13})$, see Table 5.4. Since the obstacle and solution are smooth, this improved accuracy can be expected.

Example 5: For a singular elliptic obstacle problem, we consider the case when $p = 1.25$. To avoid dividing by zero, terms like $|u_{i+1,j}-u_{i,j}|^{p-2}$ are replaced by

$$\left(|u_{i+1,j}-u_{i,j}|^2 + 10^{-10}\right)^{p/2-1}.$$

The domain is $[-1.5,1.5]^2$, the obstacle is taken to be four randomly sampled bump functions (see Figure 5.5), and the force and boundary conditions are set to zero. The

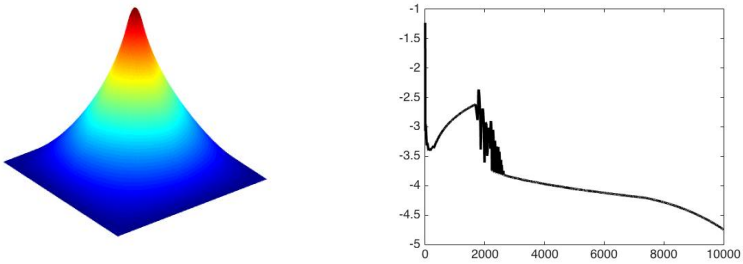

FIG. 5.4. *Left: The numerical solution for Example 4 on a grid of* $h=0.0078$ *(grid size is* 256 *by* 256*),* $\mu=250$*,* $p=3$*, and obstacle* $\varphi=2e^{-15((x-1)^2+(y-1)^2)}$*. Right: The log of the maximum difference between consecutive iterates (i.e* $\log_{10}(||u^n-u^{n-1}||_\infty)$*).*

parameters are fixed at: $\mu=500$, $tol=10^{-4}$, and $\mathrm{dt}=\frac{h^2}{15}$. The over-relaxation parameter for the algorithm is set to $a=15$. In Figure 5.5, the numerical solution is plotted on a 256 by 256 grid (bottom) and the corresponding obstacle (top). The accuracy of the numerical scheme is estimated to be $\mathcal{O}(h^{0.67})$, see Table 5.5. This accuracy is expected, since the operator is singular. In terms of the convergence, although the operator is singular and the obstacles form sharp peaks (discretely), the method is still able to construct a solution which respects the constraint (there is no cross-over between the solution and the obstacle).

5.3. Double Obstacle Problem We also consider the double obstacle problem:

$$\min_{\varphi_1\le u\le\varphi_2}\left\{F(u)=\int_\Omega L(x,u,\nabla u)\,\mathrm{dx}\right\} \tag{5.1}$$

h	0.0645	0.0317	0.0157	0.0078	0.0039
Error Ratio in L^∞	1.8111	0.7579	1.0715	1.0103	0.9885

TABLE 5.4. *Error analysis for Example 4. The error ratio* $=\log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu=250$*,* $tol=\frac{h^2}{100}$*,* $dt=\frac{h^2}{10}$*, and* $a=5$*. The order of accuracy for this example is approximately* $\mathcal{O}(h^{1.13})$*.*

h	0.0968	0.0476	0.0236	0.0118
Error Ratio in L^∞	0.6268	0.7091	0.8255	0.5212

TABLE 5.5. *Error analysis for Example 5. The error ratio* $=\log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu=500$*,* $tol=10^{-4}$*,* $dt=\frac{h^2}{15}$*, and* $a=15$*. The order of accuracy for this example is approximately* $\mathcal{O}(h^{0.67})$*.*

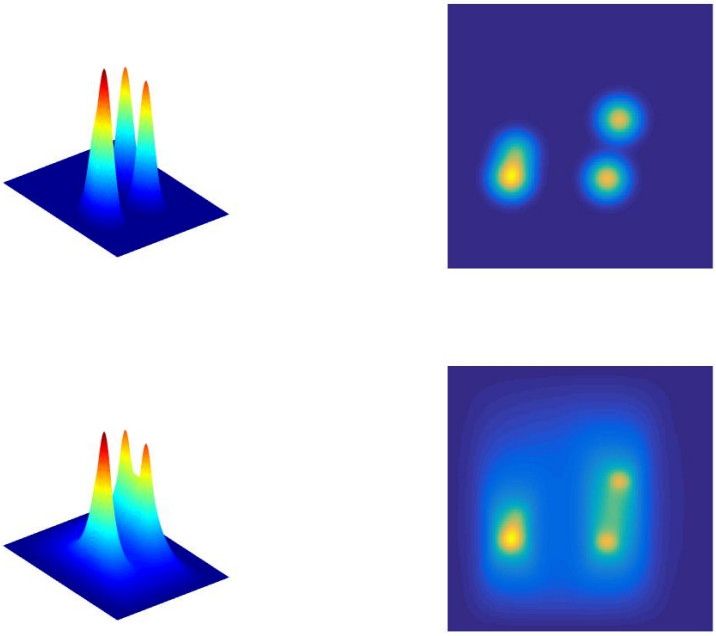

FIG. 5.5. *Graphs for Example 5. Top: The obstacle is chosen as the linear combination of four random bump functions. Note that two of the bumps are close to each other. Bottom: The numerical solution with* $h=0.0118$ *(grid size is* 256 *by* 256*),* $\mu=500$*,* $p=1.25$*, and* $tol=10^{-4}$*.*

where the solution is constrained between two functions. The penalty takes the form:

$$\int_\Omega \mu_1(\varphi_1-u)_+ + \mu_2(u-\varphi_2)_+ \mathrm{dx},$$

and the algorithm stays the same, except that the proximal operator from the previous examples is replaced by the proximal operator for the penalty above (see Appendix 7.2).

Example 6 & 7: Using F_2 with $p=4$, we compute the numerical solution to the double obstacle problem. The domain is $[-1,1]^2$, the obstacles for Example 6 are $\varphi_1 = 1.5-15((x-0.5)^2+y^2)$ and $\varphi_2=-1.5+15((x+0.5)^2+y^2)$ and, for Example 7, are $\varphi_1 = 1.5-10\sqrt{(x-0.5)^2+y^2}$ and $\varphi_2=-1.5+10\sqrt{(x+0.5)^2+y^2}$. The parameters are fixed at: $\mu=5000$, $tol=10^{-7}$, and $\mathrm{dt}=\frac{h^4}{10}$. The over-relaxation parameter for the algorithm is set to $a=15$. In Figure 5.6, the obstacles are plotted in blue and the numerical solutions are displayed as gray meshes. The accuracy for Example 6 is estimated to be $\mathcal{O}(h^{0.90})$, see Table 5.6. For Example 7, it is important to note that even though the cones form sharp peaks, the computed solution remains between the two obstacles (no cross-over is observed).

5.4. Nonlinear Two-Phase Membrane Problem

An anisotropic nonlinear version of the two-phase problem [7] is as follows:

$$\min_u \int_\Omega \frac{1}{p}(|u_x|^p+|u_y|^p)+\mu_1 u_+ + \mu_2(-u)_+ \mathrm{dx} \tag{5.2}$$

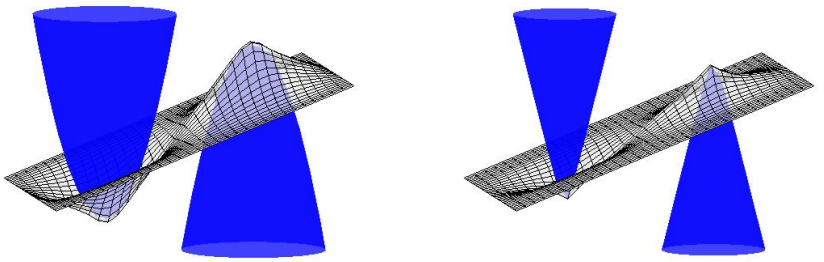

FIG. 5.6. *Double obstacle problem. Obstacles are in blue, the solutions are displayed as gray meshes. Example 6 is on the left and Example 7 is on the right. The computational times are similar between the two examples (Example 6 took* 0.241 *seconds and Example 7 took* 0.254 *seconds), even though the regularities differ.*

where μ_1 and μ_2 are positive parameters. This can be viewed as a soft obstacle problem since the parameters are prescribed and not directly used to enforce the constraint. The two-phase membrane problem shows an additional motivation for the penalty term used here, in particular, that it arrises naturally in physical models.

Example 8: In Figure 5.7, we consider the problem with $p=3$ (right) and $p=4$ (left). The parameters are set to $h=0.0078$, $\mu_1=\mu_2=1$, $tol=10^{-8}$, $\mathrm{dt}=\frac{h^4}{10}$, and $a=15$. Three regions form: the positive values of the solution (warm), the negative (cold), and the regime where u is identically zero (green). The regions are separated by triple-junction like structures. As p is varied, the numerical method remains stable.

In Figure 5.8, the parameter p is varied: $p=2,3,4$, and an image of the solution is shown (left column). Zoomed in images of the triple-junction-like structure are plotted as well (right column). We observe the "opening" of the triple-junction-like structure seen in the linear problem (top). Note that when $p=4$, the gap between the two signed regions are on the order of the grid spacing.

h	0.0645	0.0317	0.0157	0.0078
Error Ratio L^∞	0.9851	0.6528	0.9693	0.9954

TABLE 5.6. *Error analysis for Example 6. The error ratio* $= \log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu_1=\mu_2=5000$, $tol=10^{-7}$, $dt=\frac{h^4}{10}$, *and* $a=15$. *The order of accuracy for the problem is approximately* $\mathcal{O}(h^{0.90})$.

h	0.0645	0.0317	0.0157	0.0078
Error Ratio in L^∞	1.00	1.00	1.00	1.00

TABLE 5.7. *Error analysis for Example 8. The error ratio* $= \log_2\left(\frac{||u^h-u^{2h}||_\infty}{||u^{2h}-u^{4h}||_\infty}\right)$ *for several grid sizes is examined. The parameters are fixed at* $\mu_1=\mu_2=1$, $tol=10^{-8}$, $dt=\frac{h^4}{10}$, *and* $a=15$. *The order of accuracy for the problem is approximately* $\mathcal{O}(h^{1.00})$.

FIG. 5.7. *The numerical solution for Example 8 on a grid of* $h=0.0078$ *(grid size is 256 by 256),* $\mu_1=\mu_2=1$, $tol=10^{-8}$, $dt=\frac{h^4}{10}$, *and* $a=15$, *where the "obstacles" are identically zero. Left:* $p=3$, *Right:* $p=4$.

6. Conclusion We proposed an exact relaxation for the obstacle problem associated with nonlinear variational models. In both the continuous and discrete settings, the relaxation parameter μ can be calculated from the given data, and thus the method can be considered as "parameter-free". To compute the solution numerically, an accelerated variant of the forward-backward algorithm is used. This is made possible by the simplicity of the proximal operators associated with the penalty terms. Since the derivatives of the unconstrained problem may be degenerate or singular, they are taken explicitly in the algorithm as a forward iteration. This can be made implicit by using, for example, the Douglas-Rachford algorithm, as was done in the linear case. Several examples were presented, including a quasi-linear equation, constrained anisotropic p-Laplacians, the nonlinear double obstacle problem, and the nonlinear two-phase membrane problem.

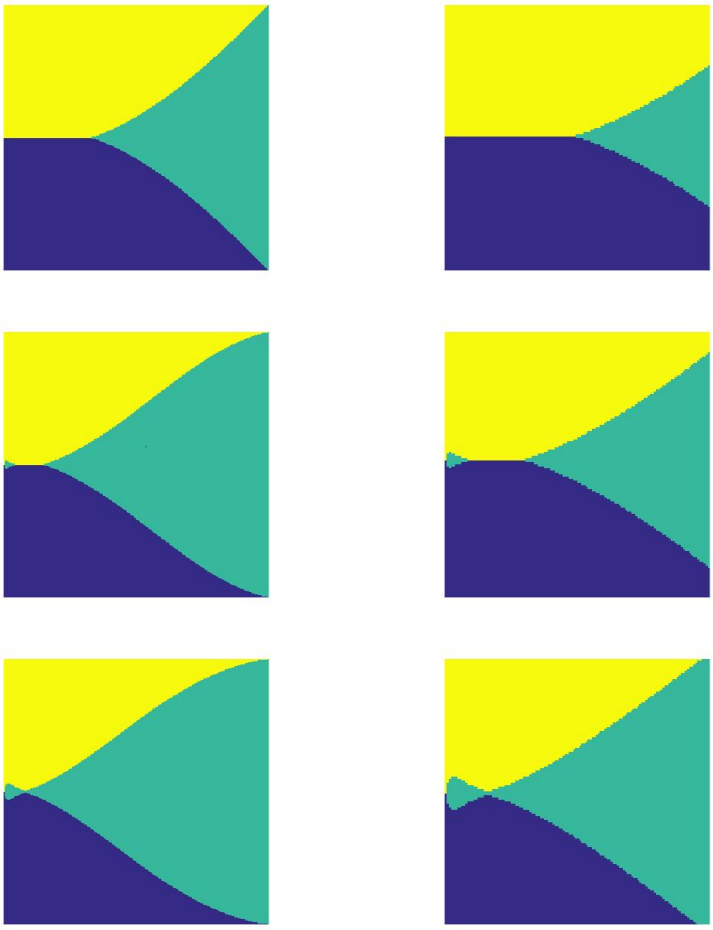

FIG. 5.8. *The numerical observation of the "opening" of the triple junction as a function of p for Example 8. The first row p = 2, second row p = 3, and third row p = 4. The left column is the solution, the right column is the zoomed-in images.*

7. Appendix For completeness, we include a brief derivation of the proximal operators for the penalty terms. These operators are related to the "shrink" function found in L^1 optimization.

7.1. Proximal Operator for a Single Obstacle The proximal operator for the penalty term with a single obstacle is defined as:

$$\begin{aligned} u &= \operatorname{prox}_{\lambda P} v \\ &= \operatorname*{argmin}_u \int_\Omega (\varphi - u)_+ + \frac{1}{2\lambda}(u-v)^2 \mathrm{dx}. \end{aligned}$$

To derive an explicit form for the minimizer, first consider the two regimes where $u > \varphi$ and $u < \varphi$. In the first case, where $u > \varphi$, the minimization can be solved directly:

$$\begin{aligned} u &= \operatorname*{argmin}_u \int_\Omega (\varphi - u)_+ + \frac{1}{2\lambda}(u-v)^2 \mathrm{dx} \\ &= \operatorname*{argmin}_u \int_\Omega \frac{1}{2\lambda}(u-v)^2 \mathrm{dx} \\ &= v. \end{aligned}$$

In the second case, where $u < \varphi$, the minimization becomes:

$$\begin{aligned} u &= \operatorname*{argmin}_u \int_\Omega (\varphi - u)_+ + \frac{1}{2\lambda}(u-v)^2 \mathrm{dx} \\ &= \operatorname*{argmin}_u \int_\Omega \varphi - u + \frac{1}{2\lambda}(u-v)^2 \mathrm{dx}. \end{aligned}$$

The functional is differentiable (within the domain $u < \varphi$) and thus by taking the derivative and setting it to zero it can be shown that:

$$u = v + \lambda.$$

The contact set case, $u = \varphi$, is trivial. Altogether, the proximal operator for a single obstacle is:

$$u = \begin{cases} v + \lambda, & \text{if } \varphi > v + \lambda \\ \varphi, & \text{if } v \le \varphi \le v + \lambda \\ v, & \text{if } \varphi < v. \end{cases}$$

7.2. Proximal Operator for the Double Obstacle The proximal operator for the penalty term for the double obstacle problem is defined as:

$$u = \operatorname*{argmin}_u \int_\Omega \mu_1(\varphi_1 - u)_+ + \mu_2(u - \varphi_2)_+ + \frac{1}{2}(u-v)^2 \,\mathrm{dx},$$

where $\varphi_2 > \varphi_1$. We consider several cases relating the solution to the obstacles. For the first case, when $u > \varphi_2$, the minimization can be solved directly:

$$\begin{aligned} u &= \operatorname*{argmin}_u \int_\Omega \mu_1(\varphi_1 - u)_+ + \mu_2(u - \varphi_2)_+ + \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= \operatorname*{argmin}_u \int_\Omega \mu_2(u - \varphi_2) + \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= v - \mu_2. \end{aligned}$$

For the second case, when $u < \varphi_1$ the minimization can also be solved directly:

$$\begin{aligned} u &= \operatorname*{argmin}_u \int_\Omega \mu_1(\varphi_1 - u)_+ + \mu_2(u - \varphi_2)_+ + \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= \operatorname*{argmin}_u \int_\Omega \mu_1(\varphi_1 - u) + \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= v + \mu_1. \end{aligned}$$

The third case $\varphi_1 < u < \varphi_2$, results in:

$$\begin{aligned} u &= \underset{u}{\operatorname{argmin}} \int_\Omega \mu_1(\varphi_1 - u)_+ + \mu_2(u-\varphi_2)_+ + \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= \underset{u}{\operatorname{argmin}} \int_\Omega \frac{1}{2}(u-v)^2 \,\mathrm{dx} \\ &= v. \end{aligned}$$

The contact set cases, $u = \varphi_1$ or $u = \varphi_2$, are trivial. Altogether, we have:

$$u = \begin{cases} v - \mu_2, & \text{if } \varphi_2 + \mu_2 \leq v \\ \varphi_2, & \text{if } \varphi_2 \leq v < \varphi_2 + \mu_2 \\ v, & \text{if } \varphi_1 \leq v \leq \varphi_2 \\ \varphi_1, & \text{if } \varphi_1 - \mu_1 \leq v < \varphi_1 \\ v + \mu_1, & \text{if } v < \varphi_1 - \mu_1. \end{cases}$$

REFERENCES

[1] Alessandro Alla and J. Nathan Kutz Nonlinear model order reduction via Dynamic Mode Decomposition. *arXiv preprint* , arXiv:1602.05080, 2016.

[2] Lori Badea. Convergence rate of a multiplicative schwarz method for strongly nonlinear variational inequalities. In *Analysis and optimization of differential systems*, pages 31–41. Springer, 2003.

[3] Lori Badea, Xue-Cheng Tai, and Junping Wang. Convergence rate analysis of a multiplicative Schwarz method for variational inequalities. *SIAM Journal on Numerical Analysis*, 41(3):1052–1073, 2003.

[4] Guy Barles and Panagiotis E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations *Asymptotic Analysis*, 4(3):271–283, 1991.

[5] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009.

[6] Gérard Bourdaud and Yves Meyer. Fonctions qui operent sur les espaces de Sobolev *Journal of functional analysis*, 97(2): 351–360, 1991.

[7] Farid Bozorgnia. Numerical solutions of a two-phase membrane problem. *Applied Numerical Mathematics*, 61(1):92–107, 2011.

[8] Haim Brezis. Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert *Elsevier*, 5, 1973.

[9] Haim Brezis. *Monotone Operators Non Linear Semi-groups and Applications*. Université Pierre et Marie Curie, Laboratoire d'Analyse Numérique, 1974.

[10] Ido Bright, Guang Lin, and J Nathan Kutz. Compressive sensing based machine learning strategy for characterizing the flow around a cylinder with limited pressure measurements. *Physics of Fluids (1994-present)*, 25(12):127102, 2013.

[11] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[12] Steven L. Brunton, Jonathan H. Tu, Ido Bright, and J. Nathan Kutz. Compressive sensing and low-rank libraries for classification of bifurcation regimes in nonlinear dynamical systems *SIAM Journal on Applied Dynamical Systems*, 13(4):1716-1732, 2014.

[13] Steven L. Brunton, Bingni W. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Koopman observable subspaces and finite linear representations of nonlinear dynamical systems for control. *arXiv preprint*, arXiv:1510.03007, 2015

[14] Luis A Caffarelli. The obstacle problem. lezioni fermiane [fermi lectures], 1998.

[15] Luis A Caffarelli. A remark on the Hausdorff measure of a free boundary, and the convergence of coincidence sets. *Bollettino della Unione Matematica Italiana A*, 18-A(5):109–113, 1991.

[16] Luis A Caffarelli. Compactness methods in free boundary problems. *Communications in Partial Differential Equations*, 5(4):427–448, 1980.

[17] Luis A Caffarelli. The regularity of elliptic and parabolic free boundaries. *Bulletin of the American Mathematical Society*, 82(4):616–618, 1976.

[18] Luis A Caffarelli. The regularity of free boundaries in higher dimensions *Acta Mathematica*, 139(1):155–184,1977.

[19] Luis A Caffarelli. The obstacle problem revisited. *Journal of Fourier Analysis and Applications*, 4(4):383–402, 1998.

[20] Luis A Caffarelli, Sandro Salsa, and Luis Silvestre. Regularity estimates for the solution and the free boundary of the obstacle problem for the fractional Laplacian. *Inventiones mathematicae*, 171(2):425–461, 2008.

[21] Luis A Caffarelli, Xavier Ros-Oton, and Joaquim Serra Obstacle problems for integro-differential operators: Regularity of solutions and free boundaries. *arXiv preprint* , arXiv:1601.05843, 2016.

[22] Russel E Caflisch, Stanley J Osher, Hayden Schaeffer, and Giang Tran. PDEs with compressed solutions. *arXiv preprint arXiv:1311.5850*, 2013.

[23] Jian-Feng Cai, Stanley Osher, and Zuowei Shen. Linearized bregman iterations for compressed sensing. *Mathematics of Computation*, 78(267):1515–1536, 2009.

[24] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[25] Antonin Chambolle and Charles Dossal. On the weak convergence of the iterates of "FISTA" *hal.inria.fr*, 2014.

[26] Zhiming Chen and Ricardo H. Nochetto. Residual type a posteriori error estimates for elliptic obstacle problems." *Numerische Mathematik* 84(4): 527–548, 2000.

[27] Long Chen, Ricardo H. Nochetto, and Chen-Song Zhang Multigrid methods for elliptic obstacle problems on 2D bisection grids *Domain decomposition methods in science and engineering XIX.*,Springer Berlin Heidelberg, 229–236, 2011.

[28] Francis Conrad, Raphaele Herbin, and Hans D Mittelmann. Approximation of obstacle problems by continuation methods. *SIAM Journal on Numerical Analysis*, 25(6):1409–1431, 1988.

[29] Abderrahim Elmoataz, Matthieu Toutain, and Daniel Tenbrinck. On the p-Laplacian and ∞-Laplacian on Graphs with Applications in Image and Data Processing. *SIAM Journal on Imaging Sciences*, 8(4):2412–2451, 2015.

[30] Jens Frehse On the regularity of solutions of a second order variational inequality. *Bollettino della Unione Matematica Italiana A*, 6(4), 1972.

[31] Donald A. French, Stig Larsson, and Ricardo H. Nochetto. Pointwise a posteriori error analysis for an adaptive penalty finite element method for the obstacle problem. *Computational Methods in Applied Mathematics*, 1(1):18–38, 2001.

[32] Michael P Friedlander and Paul Tseng. Exact regularization of convex programs. *SIAM Journal on Optimization*, 18(4):1326–1350, 2007.

[33] Avner Friedman. *Variational principles and free-boundary problems*. Courier Dover Publications, 2010.

[34] Tom Goldstein and Stanley Osher. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[35] Michael Hintermüller, VA Kovtunenko, and Karl Kunisch. Obstacle problems with cohesion: a hemivariational inequality approach and its efficient numerical solution. *SIAM Journal on Optimization*, 21(2):491–516, 2011.

[36] Ronald HW Hoppe. Multigrid algorithms for variational inequalities. *SIAM Journal on Numerical Analysis*, 24(5):1046–1065, 1987.

[37] Ronald HW Hoppe and Ralf Kornhuber. Adaptive multilevel methods for obstacle problems. *SIAM Journal on Numerical Analysis*, 31(2):301–323, 1994.

[38] Thomas Y. Hou, Qin Li, and Hayden Schaeffer. Sparse+ low-energy decomposition for viscous conservation laws. *Journal of Computational Physics*, 288:150–166, 2015.

[39] Guillaume Jouvet and Ed Bueler. Steady, shallow ice sheets as obstacle problems: well-posedness and finite element approximation. *SIAM Journal on Applied Mathematics*, 72(4): 1292–1314, 2012.

[40] David Kinderlehrer and Guido Stampacchia. *An introduction to variational inequalities and their applications*, volume 31. SIAM, 2000.

[41] Ralf Kornhuber. Monotone multigrid methods for elliptic variational inequalities I. *Numerische Mathematik*, 69(2):167–184, 1994.

[42] Ralf Kornhuber. Monotone multigrid methods for elliptic variational inequalities II. *Numerische Mathematik*, 72(4): 481–499, 1996.

[43] Ralf Kornhuber and Qingsong Zou. Efficient and reliable hierarchical error estimates for the discretization error of elliptic obstacle problems. *Mathematics of Computation*, 80(273): 69–88, 2011.

[44] Hung-Ju Kuo and Neil S. Trudinger. Discrete methods for fully nonlinear elliptic equations. *SIAM Journal on Numerical Analysis*, 29(1): 123–135, 1992.

[45] Peter Lindqvist. Regularity for the gradient of the solution to a nonlinear obstacle problem with degenerate ellipticity. *Nonlinear Analysis: Theory, Methods & Applications*, 12(11):1245–1255,

1988.

[46] Kirsi Majava and Xue-Cheng Tai. A level set method for solving free boundary problems associated with obstacles. *Int. J. Numer. Anal. Model*, 1(2):157–171, 2004.

[47] OL Mangasarian. Sufficiency of exact penalty minimization. *SIAM Journal on Control and Optimization*, 23(1):30–37, 1985.

[48] Lance J Nelson, Gus LW Hart, Fei Zhou, and Vidvuds Ozoliņš. Compressive sensing as a paradigm for building physics models. *Physical Review B*, 87(3):035125, 2013.

[49] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, 2004.

[50] Yurii Nesterov. A method of solving a convex programming problem with convergence rate $O(\frac{1}{k^2})$. *Soviet Mathematics Doklady*, 27(2):372–376. 1983.

[51] Adam M. Oberman. Convergent difference schemes for degenerate elliptic and parabolic equations: Hamilton–Jacobi equations and free boundary problems. *SIAM Journal on Numerical Analysis*, 44(2):879–895, 2006.

[52] Vidvuds Ozoliņš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed modes for variational problems in mathematics and physics. *Proceedings of the National Academy of Sciences*, 110(46):18368–18373, 2013.

[53] Vidvuds Ozolinš, Rongjie Lai, Russel Caflisch, and Stanley Osher. Compressed plane waves-compactly supported multiresolution basis for the laplace operator. *Proceedings of the National Academy of Sciences*, 2013.

[54] Arshak Petrosyan. *Regularity of free boundaries in obstacle-type problems*, volume 136. American Mathematical Soc., 2012.

[55] Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4): e1602614, 2017.

[56] Syuzanna Sargsyan, Steven L. Brunton, and J. Nathan Kutz. Nonlinear model reduction for dynamical systems using sparse sensor locations from learned libraries *Physical Review E*, 92(3):033304, 2015.

[57] Giuseppe Savaré. On the regularity of the positive part of functions. *Nonlinear Analysis: Theory, Methods & Applications*, 27(9): 1055–1074, 1996.

[58] Hayden Schaeffer, Russel Caflisch, Cory D Hauck, and Stanley Osher. Sparse dynamics for partial differential equations. *Proceedings of the National Academy of Sciences*, 110(17):6634–6639, 2013.

[59] Hayden Schaeffer. Learning partial differential equations via data discovery and sparse optimization. *Proc. R. Soc. A*, 473(2197):20160446, 2017.

[60] Hayden Schaeffer and Scott G. McCalla. Sparse model selection via integral terms. *Phys. Rev. E*, 96(2):023302, 2017.

[61] Hayden Schaeffer, Giang Tran, and Rachel Ward. Extracting Sparse High-Dimensional Dynamics from Limited Data *arXiv preprint* arXiv:1707.08528, 2017.

[62] Hayden Schaeffer, Giang Tran, and Rachel Ward. Learning Dynamical Systems and Bifurcation via Group Sparsity *arXiv preprint* arXiv:1709.01558, 2017.

[63] Hayden Schaeffer and Thomas Y. Hou An accelerated method for nonlinear elliptic PDE. *Journal of Scientific Computing*, 69(2): 556–580, 2016.

[64] Reinhard Scholz. Numerical solution of the obstacle problem by the penalty method. *Computing*, 32(4):297–306, 1984.

[65] Luis Silvestre. Regularity of the obstacle problem for a fractional power of the Laplace operator. *Communications on pure and applied mathematics*, 60(1):67–112, 2007.

[66] Guido Stampacchia Le probleme de Dirichlet pour les équations elliptiques du second ordre á coefficients discontinus, *Ann. Inst. Fourier (Grenoble)* 15(1):189–258, 1965.

[67] Xue-Cheng Tai. Rate of convergence for some constraint decomposition methods for nonlinear variational inequalities. *Numerische Mathematik*, 93(4):755–786, 2003.

[68] Giang Tran and Rachel Ward. Exact recovery of chaotic systems from highly corrupted data. *Multiscale Modeling & Simulation, SIAM*, 15(3):1108–1129, 2017.

[69] Giang Tran, Hayden Schaeffer, William M. Feldman, and Stanley J. Osher An L^1 Penalty Method for General Obstacle Problems *SIAM Journal on Applied Mathematics*, 75(4):1424–1444, 2015.

[70] C. Vázquez, E. Schiavi, J. Durany, J. I. Díaz, and N. Calvo. On a doubly nonlinear parabolic obstacle problem modelling ice sheet dynamics. *SIAM Journal on Applied Mathematics*, 63(2): 683–707, 2003.

[71] Alexander Weiss and Barbara Wohlmuth. A posteriori error estimator and error control for contact problems. *Mathematics of Computation*, 78(267):1237–1267, 2009.

[72] Dominique Zosso, Braxton Osting, Mandy Mengqi Xia and Stanley J. Osher. An Efficient Primal-Dual Method for the Obstacle Problem *Journal of Scientific Computing*, 73(1):416–437, 2017.