Deutsche Forschungsgemeinschaft
Priority Programme 1962

Global Minima for Optimal Control of the Obstacle Problem

Ahmad Ali, Klaus Deckelnick, Michael Hinze

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-095

received on October 29, 2018

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

Global minima for optimal control of the obstacle problem

Ahmad Ali[*], Klaus Deckelnick[†] & Michael Hinze[‡]

Abstract

An optimal control problem subject to an elliptic obstacle problem is studied. We obtain a numerical approximation of this problem by discretising the PDE obtained via a Moreau–Yosida type penalisation. For the resulting discrete control problem we provide a condition that allows to decide whether a solution of the necessary first order conditions is a global minimum. In addition we show that the corresponding result can be transferred to the limit problem provided that the above condition holds uniformly in the penalisation and discretisation parameters. Numerical examples with unique global solutions are presented.

Key words. Optimal control, obstacle problem, Moreau–Yosida penalisation, finite elements, global solution.

Mathematics Subject Classification. 49J20, 49M05, 49M20, 65M15, 65M60.

1 Introduction

In this paper we are concerned with the following distributed optimal control problem for the elliptic obstacle problem

$$(\mathbb{P}) \qquad \min_{u \in L^2(\Omega)} J(u) = \frac{1}{2}\|y - y_0\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$

subject to

$$y \in K, \quad \int_\Omega \nabla y \cdot \nabla(\phi - y)dx \geq \int_\Omega (f + u)(\phi - y)dx \qquad \forall \phi \in K, \tag{1.1}$$

where

$$K = \{\phi \in H^1_0(\Omega) \,|\, \phi(x) \geq \psi(x) \text{ a.e. in } \Omega\}.$$

Furthermore, $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ is a bounded polyhedral domain, $\psi \in H^1_0(\Omega) \cap H^2(\Omega)$ is the given obstacle, $y_0, f \in L^2(\Omega)$ and $\alpha > 0$.
It is well–known that for a given function $u \in L^2(\Omega)$ the variational inequality (1.1) has a unique solution $y \in K$ and using standard arguments (cf. [14, Theorem 2.1]) one obtains

[*]Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany.

[†]Institut für Analysis und Numerik, Otto–von–Guericke–Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

[‡]Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany.

the existence of a solution of ($\mathbb{P}$). However, a major issue in the analysis and numerical approximation of ($\mathbb{P}$) is the fact that the mapping $u \mapsto y$ is in general not Gâteaux differentiable, so that the derivation of necessary first order optimality conditions becomes a difficult task. A common approach in order to handle this difficulty consists in approximating (1.1) by a sequence of penalised or regularised problems and then to pass to the limit, see e.g. [3], [14], [8], [6], [7], [9], [15] and [12]. In our work we employ a Moreau–Yosida type penalisation of the obstacle problem resulting in the following optimal control problem depending on the parameter $\gamma \gg 1$:

$$(\mathbb{P}^\gamma) \qquad \min_{u \in L^2(\Omega)} J^\gamma(u) = \frac{1}{2}\|y - y_0\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$

subject to

$$\int_\Omega \nabla y \cdot \nabla \phi \, dx + \gamma^3 \int_\Omega [(y-\psi)^-]^3 \, \phi \, dx = \int_\Omega (f+u)\phi \, dx \qquad \forall \phi \in H^1_0(\Omega). \tag{1.2}$$

Here, $a^- = \min(a, 0)$. Existence of a solution of ($\mathbb{P}^\gamma$) and a detailed convergence analysis as $\gamma \to \infty$ can be found in [12, Section 3]. For numerical purposes we shall discretise the PDE (1.2) with the help of continuous, piecewise linear finite elements giving rise to a discrete optimisation problem, whose solutions satisfy standard first order optimality conditions. Due to the lack of convexity of the underlying problem it is however not clear whether a computed discrete stationary point is actually a global minimum. Our first main result of this paper establishes for a fixed penalisation parameter and a discrete stationary solution a condition that ensures global optimality, see Theorem 2.2. This condition has the form of an inequality involving the state and the adjoint variable as well as the obstacle. Furthermore, the minimum is unique in case that the inequality is strict. A similar kind of result for control problems subject to a class of semilinear elliptic PDEs has been obtained in [1,2], but the form of the condition presented here is adapted to the obstacle problem and entirely different from the one proposed in [1,2]. In Section 3 we consider a sequence of approximate control problems ($\mathbb{P}_h^{\gamma_h}$) with $h \to 0$ and $\gamma_h \to \infty$ as $h \to 0$, where h denotes the grid size. As our second main result we shall prove that a corresponding sequence of discrete stationary points which are uniformly bounded in a suitable sense has a subsequence that converges to a limit satisfying a system of first order optimality conditions, see Theorem 3.1. It turns out that a solution of this system is strongly stationary in the sense of [14] and that we obtain a continuous analogue of the condition mentioned above guaranteeing global optimality of a stationary point, see Theorem 3.2. The only other sufficient condition for optimality which we are aware of can be found in [13, Theorem 5.4], where global optimality is derived from the condition that $y_0 \leq \psi$, see also [8, Section 5.2]. A sufficient second order optimality condition giving local optimality is derived in [10]. In Section 4 we briefly outline how to apply our theory to a direct discretisation of (1.1). Finally, several numerical examples with unique global solutions are presented in Section 5.

2 Discretisation of ($\mathbb{P}$)

Let $\mathcal{T}_h$ be an admissible triangulation of Ω so that

$$\bar{\Omega} = \bigcup_{T \in \mathcal{T}_h} \bar{T}.$$

We denote by $x_1, \ldots, x_n$ the interior and by $x_{n+1}, \ldots, x_{n+m}$ the boundary vertices of $\mathcal{T}_h$. Next, let

$$X_h := \{v_h \in C^0(\bar{\Omega}) : v_h \text{ is a linear polynomial on each } T \in \mathcal{T}_h\}$$

be the space of linear finite elements as well as $X_{h0} := X_h \cap H_0^1(\Omega)$. The standard nodal basis functions are defined by $\phi_i \in X_h$ with $\phi_i(x_j) = \delta_{ij}, i, j = 1, \ldots, n+m$. In particular, $\{\phi_1, \ldots, \phi_n\}$ is a basis of X_{h0}. We shall make use of the Lagrange interpolation operator

$$I_h : C^0(\bar{\Omega}) \to X_h \quad I_h y := \sum_{j=1}^{n+m} y(x_j)\phi_i.$$

Let us approximate (1.2) as follows: for a given $u \in L^2(\Omega)$, find $y_h \in X_{h0}$ such that

$$\int_\Omega \nabla y_h \cdot \nabla \phi_h \, dx + \gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^3 \, \phi_h\} \, dx = \int_\Omega (f+u)\phi_h \, dx \quad \forall \, \phi_h \in X_{h0}. \tag{2.1}$$

It is not difficult to show that (2.1) has a unique solution $y_h \in X_{h0}$. The variational discretization of Problem $(\mathbb{P}^\gamma)$ now reads:

$$(\mathbb{P}_h^\gamma) \qquad \begin{array}{l} \min\limits_{u \in L^2(\Omega)} J_h^\gamma(u) := \frac{1}{2}\|y_h - y_0\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ \text{subject to } y_h \text{ solves (2.1)}\ . \end{array}$$

Using standard arguments one obtains:

Lemma 2.1 *Suppose that $u_h \in L^2(\Omega)$ is a local solution of $(\mathbb{P}_h^\gamma)$ with corresponding state $y_h \in X_{h0}$. Then there exists an adjoint state $p_h \in X_{h0}$ such that*

$$\int_\Omega \nabla y_h \cdot \nabla \phi_h dx + \gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^3 \, \phi_h\} dx = \int_\Omega (f + u_h)\phi_h \, dx \quad \forall \, \phi_h \in X_{h0} \tag{2.2}$$

$$\int_\Omega \nabla p_h \cdot \nabla \phi_h dx + 3\gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^2 p_h \phi_h\} dx = \int_\Omega (y_h - y_0)\phi_h \, dx \quad \forall \, \phi_h \in X_{h0} \tag{2.3}$$

$$\alpha u_h + p_h = 0 \quad in\ \Omega. \tag{2.4}$$

Note that (2.4) implicitly yields a discretisation of the control variable so that (2.2)–(2.4) is a finite–dimensional system that can be solved using classical nonlinear programming algorithms. However, due to the non-convexity of the problem, it is not clear whether a solution of (2.2)–(2.4) is actually a global minimum of $(\mathbb{P}_h^\gamma)$. Our first main result provides a sufficient condition for a discrete stationary point that guarantees that this is the case. In order to formulate the corresponding condition we introduce λ_1 as the smallest eigenvalue of $-\Delta$ in Ω subject to homogeneous Dirichlet boundary conditions, i.e.

$$\lambda_1 = \inf_{\phi \in H_0^1(\Omega)\setminus\{0\}} \frac{\int_\Omega |\nabla \phi|^2 \, dx}{\int_\Omega \phi^2 \, dx}. \tag{2.5}$$

In what follows we shall abbreviate $y_j = y_h(x_j), p_j = p_h(x_j)$ and $\psi_j = \psi(x_j)$, $j = 1, \ldots, n$.

Theorem 2.2 *Suppose that (u_h, y_h, p_h) is a solution of (2.2)–(2.4), which satisfies*

$$p_k \geq 0 \quad \forall k \in \{1, \ldots, n\} \text{ with } y_k - \psi_k = 0 \tag{2.6}$$

and define

$$\eta := \min\Big(\min_{y_j > \psi_j} \frac{p_j}{y_j - \psi_j}, \min_{y_j < \psi_j} \frac{3p_j}{\psi_j - y_j}, 0\Big). \tag{2.7}$$

If

$$|\eta| \leq \alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}, \tag{2.8}$$

then u_h is a global minimum for Problem $(\mathbb{P}_h^\gamma)$. If the inequality (2.8) is strict, then u_h is the unique global minimum.

Proof: Let $v \in L^2(\Omega)$ be arbitrary and denote by $\tilde{y}_h \in X_{h0}$ the solution of

$$\int_\Omega \nabla \tilde{y}_h \cdot \nabla \phi_h dx + \gamma^3 \int_\Omega I_h\{[(\tilde{y}_h - \psi)^-]^3 \, \phi_h\} dx = \int_\Omega (f+v)\phi_h \, dx \quad \forall\, \phi_h \in X_{h0}. \tag{2.9}$$

A straightforward calculation shows that

$$\begin{aligned} J_h^\gamma(v) - J_h^\gamma(u_h) &= \frac{1}{2}\|\tilde{y}_h - y_h\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u_h\|^2_{L^2(\Omega)} \\ &\quad + \int_\Omega (y_h - y_0)(\tilde{y}_h - y_h)\, dx + \alpha \int_\Omega u_h(v - u_h)\, dx. \end{aligned} \tag{2.10}$$

We deduce from (2.3), (2.2) and (2.9) that

$$\begin{aligned} \int_\Omega (y_h - y_0)(\tilde{y}_h - y_h)\, dx &= \int_\Omega \nabla p_h \cdot \nabla(\tilde{y}_h - y_h)\, dx + 3\gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^2 p_h(\tilde{y}_h - y_h)\} dx \\ &= -\gamma^3 \int_\Omega I_h\{[(\tilde{y}_h - \psi)^-]^3 p_h\} dx + \gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^3 p_h\} dx + \int_\Omega p_h(v - u_h)\, dx \\ &\quad + 3\gamma^3 \int_\Omega I_h\{[(y_h - \psi)^-]^2 p_h(\tilde{y}_h - y_h)\} dx \\ &= \gamma^3 \int_\Omega I_h\{p_h r_h\}\, dx + \int_\Omega p_h(v - u_h)\, dx, \end{aligned} \tag{2.11}$$

where

$$\begin{aligned} r_h &= -[(\tilde{y}_h - \psi)^-]^3 + [(y_h - \psi)^-]^3 + 3[(y_h - \psi)^-]^2(\tilde{y}_h - y_h) \\ &= -[(\tilde{y}_h - \psi)^-]^3 - 2[(y_h - \psi)^-]^3 + 3[(y_h - \psi)^-]^2(\tilde{y}_h - \psi). \end{aligned}$$

Using Young's inequality we find that

$$\begin{aligned} [(y_h - \psi)^-]^2(\tilde{y}_h - \psi) &= [(y_h - \psi)^-]^2\big((\tilde{y}_h - \psi)^- + (\tilde{y}_h - \psi)^+\big) \\ &\geq [(y_h - \psi)^-]^2(\tilde{y}_h - \psi)^- \geq \frac{2}{3}[(y_h - \psi)^-]^3 + \frac{1}{3}[(\tilde{y}_h - \psi)^-]^3 \end{aligned}$$

so that

$$r_h \geq 0 \quad \text{in } \bar{\Omega}. \tag{2.12}$$

Inserting (2.11) into (2.10) and applying (2.4) we obtain

$$J_h^\gamma(v) - J_h^\gamma(u_h) = \frac{1}{2}\|\tilde{y}_h - y_h\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u_h\|^2_{L^2(\Omega)} + \gamma^3 \sum_{j=1}^n p_j r_j m_j, \tag{2.13}$$

where we have abbreviated $r_j := r_h(x_j) \geq 0$ and $m_j = \int_\Omega \phi_j \, dx$.

Let us decompose

$$\{1,\ldots,n\} = \{j \,|\, y_j > \psi_j\} \cup \{j \,|\, y_j = \psi_j\} \cup \{j \,|\, y_j < \psi_j\} =: \mathcal{N}^+ \cup \mathcal{N}^0 \cup \mathcal{N}^-.$$

(i) $j \in \mathcal{N}^+$: In this case we have in view of (2.7) that $p_j \geq \eta(y_j - \psi_j)$ and hence

$$\begin{aligned} p_j r_j &\geq \eta(y_j - \psi_j) r_j = -\eta(y_j - \psi_j)[(\tilde{y}_j - \psi_j)^-]^3 \\ &= \eta(\tilde{y}_j - y_j)[(\tilde{y}_j - \psi_j)^-]^3 - \eta[(\tilde{y}_j - \psi_j)^-]^4 \\ &\geq \eta(\tilde{y}_j - y_j)\{[(\tilde{y}_j - \psi_j)^-]^3 - [(y_j - \psi_j)^-]^3\} \end{aligned} \tag{2.14}$$

since $(y_j-\psi_j)^-=0, j\in\mathcal{N}^+$ and $\eta\le 0$.
(ii) $j\in\mathcal{N}^-$: In this case we have $p_j\ge -\frac{\eta}{3}(y_j-\psi_j)$ so that

$$\begin{aligned} p_jr_j &\ge -\frac{\eta}{3}(y_j-\psi_j)r_j \qquad (2.15)\\ &= \frac{\eta}{3}(y_j-\psi_j)[(\tilde y_j-\psi_j)^-]^3+\frac{2}{3}\eta[(y_j-\psi_j)^-]^4-\eta[(y_j-\psi_j)^-]^3(\tilde y_j-\psi_j)\\ &= \eta(\tilde y_j-y_j)\{[(\tilde y_j-\psi_j)^-]^3-[(y_j-\psi_j)^-]^3\}m_j\\ &\quad -\eta\{[(\tilde y_j-\psi_j)^-]^4+\frac{1}{3}[(y_j-\psi_j)^-]^4-\frac{4}{3}(y_j-\psi_j)[(\tilde y_j-\psi_j)^-]^3\}\\ &\ge \eta(\tilde y_j-y_j)\{[(\tilde y_j-\psi_j)^-]^3-[(y_j-\psi_j)^-]^3\}, \qquad (2.16)\end{aligned}$$

since $\eta\le 0$ and $\frac{4}{3}|a|^3|b|\le a^4+\frac{1}{3}b^4$ in view of Young's inequality.
(iii) $j\in\mathcal{N}^0$: In this case we have $p_j\ge 0$ by (2.6) and therefore

$$p_jr_j\ge 0\ge \eta(\tilde y_j-y_j)\{[(\tilde y_j-\psi_j)^-]^3-[(y_j-\psi_j)^-]^3\}. \qquad (2.17)$$

Combining (2.14)–(2.17) with (2.1), (2.9) and the definition of λ_1 we derive

$$\begin{aligned}\gamma^3\sum_{j=1}^n p_jr_jm_j &\ge \eta\gamma^3\sum_{j=1}^n(\tilde y_j-y_j)\{[(\tilde y_j-\psi_j)^-]^3-[(y_j-\psi_j)^-]^3\}m_j\\ &= \eta\gamma^3\int_\Omega I_h\{([(\tilde y_h-\phi)^-]^3-[(y_h-\psi)^-]^3)(\tilde y_h-y_h)\}dx\\ &= -\eta\int_\Omega|\nabla(\tilde y_h-y_h)|^2\,dx+\eta\int_\Omega(v-u_h)(\tilde y_h-y_h)\,dx\\ &\ge |\eta|\lambda_1\int_\Omega|\tilde y_h-y_h|^2\,dx-|\eta|\int_\Omega(v-u_h)(\tilde y_h-y_h)\,dx.\end{aligned}$$

If we insert this bound into (2.13) we deduce that

$$J_h^\gamma(v)-J_h^\gamma(u_h)\ge\int_\Omega[(\frac{1}{2}+|\eta|\lambda_1)|\tilde y_h-y_h|^2+\frac{\alpha}{2}|v-u_h|^2-|\eta|(v-u_h)(\tilde y_h-y_h)]\,dx.$$

It is not difficult to verify that the bilinear form $(x_1,x_2)\mapsto(\frac{1}{2}+|\eta|\lambda_1)x_1^2+\frac{\alpha}{2}x_2^2-|\eta|x_1x_2$ is positive semidefinite (positive definite) if

$$\frac{\alpha}{4}+\frac{\alpha}{2}\lambda_1|\eta|-\frac{|\eta|^2}{4}\ge 0\quad(>0).$$

This is the case, if $|\eta|\le\mu$ ($|\eta|<\mu$), where $\mu=\alpha\lambda_1+\sqrt{\alpha^2\lambda_1^2+\alpha}$ is the positive root of the quadratic equation $x^2-2\alpha\lambda_1x-\alpha=0$. This completes the proof of the theorem. ■

3 Convergence

Let $(\mathcal{T}_h)_{0<h\le h_0}$ be a sequence of triangulations of $\bar\Omega$ with mesh size $h:=\max_{T\in\mathcal{T}_h}h_T$, where $h_T=\mathrm{diam}(T)$. We suppose that the sequence is regular in the sense that there exists $\rho>0$ such that

$$r_T\ge\rho h_T\quad\forall T\in\mathcal{T}_h,\ 0<h\le h_0,$$

where r_T denotes the radius of the largest ball contained in T. For a sequence $(\gamma_h)_{0<h\le h_0}$ satisfying

$$\gamma_h \to \infty \quad \text{as} \quad h \to 0 \tag{3.1}$$

we now consider the corresponding sequence of control problems $(\mathbb{P}_h^{\gamma_h})$. Our first result is concerned with the question how the stationarity conditions at the discrete level transfer to the continuous level under the assumption that the quantity defined in (2.7) is uniformly bounded.

Theorem 3.1 *Let $(\bar u_h, \bar y_h, \bar p_h)_{0<h\le h_0} \subset L^2(\Omega)\times X_{h0}\times X_{h0}$ be a sequence of solutions of (2.2)–(2.4) with corresponding $\eta_h \le 0$ given by (2.7) and suppose that*

$$\|\bar u_h\|_{L^2(\Omega)} \le C, \quad |\eta_h| \le C, \qquad 0<h\le h_0 \tag{3.2}$$

for some $C\ge 0$. Then there exists a subsequence $h\to 0$ and $(\bar u,\bar y,\bar p)\in L^2(\Omega)\times H_0^1(\Omega)\times H_0^1(\Omega)$ as well as $\bar\xi,\bar\mu\in H^{-1}(\Omega)$ such that

$$\bar u_h \to \bar u \text{ in } L^2(\Omega), \quad \bar y_h \to \bar y \text{ in } H^1(\Omega), \quad \bar p_h \rightharpoonup \bar p \text{ in } H^1(\Omega), \quad \eta_h \to \eta$$

and

$$\int_\Omega \nabla\bar y\cdot\nabla\phi\,dx = \int_\Omega (f+\bar u)\phi\,dx + \langle\bar\xi,\phi\rangle \quad \forall\phi\in H_0^1(\Omega) \tag{3.3}$$

$$\bar y\ge\psi \text{ a.e. in } \Omega,\ \bar\xi\ge 0,\ \langle\bar\xi,\bar y-\psi\rangle = 0 \tag{3.4}$$

$$\int_\Omega \nabla\bar p\cdot\nabla\phi\,dx = \int_\Omega (\bar y-y_0)\phi\,dx - \langle\bar\mu,\phi\rangle \quad \forall\phi\in H_0^1(\Omega) \tag{3.5}$$

$$\langle\bar\xi,\bar p\rangle = 0,\ \langle\bar\mu,\bar y-\psi\rangle = 0 \tag{3.6}$$

$$\alpha\bar u+\bar p = 0 \quad \text{a.e. in } \Omega \tag{3.7}$$

$$\bar p\ge\eta(\bar y-\psi) \text{ a.e. in } \Omega,\ \bar\mu\ge\eta\,\bar\xi. \tag{3.8}$$

Proof: Let us first derive on upper bound on $\bar y_h$ in $H^1(\Omega)$. Inserting $\phi_h=\bar y_h - I_h\psi$ into (2.1) we derive

$$\int_\Omega \nabla\bar y_h\cdot\nabla(\bar y_h-I_h\psi)\,dx + \gamma_h^3\int_\Omega I_h\{[(\bar y_h-\psi)^-]^4\}\,dx = \int_\Omega (f+\bar u_h)(\bar y_h-I_h\psi)\,dx,$$

from which we infer with the help of Poincaré's inequality

$$\|\bar y_h\|^2_{H^1(\Omega)} + \gamma_h^3\int_\Omega I_h\{[(\bar y_h-\psi)^-]^4\}\,dx \le c\big(\|f\|^2_{L^2(\Omega)} + \|I_h\psi\|^2_{H^1(\Omega)} + \|\bar u_h\|^2_{L^2(\Omega)}\big) \le c \tag{3.9}$$

in view of (3.2) and the fact that $\psi\in H^2(\Omega)$. Next, using $\phi_h=p_h$ in (2.3) we derive

$$\int_\Omega |\nabla\bar p_h|^2\,dx + 3\gamma_h^3\int_\Omega I_h\{[(\bar y_h-\psi)^-]^2\bar p_h^2\}\,dx = \int_\Omega (\bar y_h-y_0)\bar p_h\,dx,$$

which combined with (3.9) yields

$$\|\bar p_h\|^2_{H^1(\Omega)} + \gamma_h^3\int_\Omega I_h\{[(\bar y_h-\psi)^-]^2\bar p_h^2\}\,dx \le c. \tag{3.10}$$

Thus, there exists a subsequence $h\to 0$ and $\bar u\in L^2(\Omega)$, $\bar y\in H_0^1(\Omega)$, $\bar p\in H_0^1(\Omega)$ as well as $\eta\le 0$ such that

$$\bar y_h \rightharpoonup \bar y \quad \text{in } H_0^1(\Omega), \quad \bar y_h\to\bar y \quad \text{in } L^2(\Omega), \tag{3.11}$$

$$\begin{aligned} \bar{p}_h &\rightharpoonup \bar{p} \quad \text{in } H_0^1(\Omega), \quad \bar{p}_h \to \bar{p} \quad \text{in } L^2(\Omega), && (3.12)\\ \bar{u}_h &\to \bar{u} \quad \text{in } L^2(\Omega), && (3.13)\\ \eta_h &\to \eta. && (3.14) \end{aligned}$$

Note that the strong convergence in (3.13) is a consequence of (3.12) and (2.4). Let us first verify that $\bar{y} \in K$. Using the convexity of $s \mapsto (s^-)^4$ we derive that

$$[(\bar{y}_h - I_h\psi)^-]^4 \le I_h\{[(\bar{y}_h - \psi)^-]^4\} \quad \text{on } T \text{ for all } T \in \mathcal{T}_h,$$

which together with (3.9) yields

$$\|(\bar{y}_h - I_h\psi)^-\|_{L^1(\Omega)} \le c\|(\bar{y}_h - I_h\psi)^-\|_{L^4(\Omega)} \le c\gamma_h^{-\frac{3}{4}} \to 0, \; h \to 0.$$

We then have

$$\|(\bar{y} - \psi)^-\|_{L^1(\Omega)} \le \|(\bar{y}_h - I_h\psi)^-\|_{L^1(\Omega)} + \|\bar{y}_h - \bar{y}\|_{L^1(\Omega)} + \|I_h\psi - \psi\|_{L^1(\Omega)}$$

so that we deduce that $\|(\bar{y} - \psi)^-\|_{L^1(\Omega)} = 0$ by sending $h \to 0$. Thus $\bar{y} \ge \psi$ a.e. in Ω and hence $\bar{y} \in K$.

We next show that $\bar{y}$ is the solution of (1.1) with $u = \bar{u}$ by verifying that $\bar{y}$ minimizes the functional

$$y \mapsto \frac{1}{2}\int_\Omega |\nabla y|^2\,dx - \int_\Omega (f + \bar{u})y\,dx \quad \text{over } K.$$

To see this, let $y \in K$ be arbitrary. Arguing as in the proof of Proposition 5.2 in [4] we obtain a sequence $(y_k)_{k\in\mathbb{N}}$ such that $y_k \in K \cap H^2(\Omega), k \in \mathbb{N}$ and $y_k \to y$ in $H^1(\Omega)$ as $k \to \infty$. Since $\bar{y}_h$ is a solution of (2.1) it satisfies $Q(\bar{y}_h) = \min_{z_h \in X_{h0}} Q(z_h)$, where

$$Q(z_h) = \frac{1}{2}\int_\Omega |\nabla z_h|^2\,dx + \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(z_h - \psi)^-]^4\}\,dx - \int_\Omega (f + \bar{u}_h)z_h\,dx. \tag{3.15}$$

Therefore,

$$\begin{aligned} \frac{1}{2}\int_\Omega |\nabla \bar{y}_h|^2\,dx + \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(\bar{y}_h - \psi)^-]^4\}\,dx - \int_\Omega (f + \bar{u}_h)\bar{y}_h\,dx \\ \le \frac{1}{2}\int_\Omega |\nabla I_h y_k|^2\,dx - \int_\Omega (f + \bar{u}_h)I_h y_k\,dx, \end{aligned} \tag{3.16}$$

since $y_k(x_j) \ge \psi(x_j), j = 1, \dots, n$. Letting first $h \to 0$ and then $k \to \infty$ we infer that

$$\frac{1}{2}\int_\Omega |\nabla \bar{y}|^2\,dx - \int_\Omega (f + \bar{u})\bar{y}\,dx \le \frac{1}{2}\int_\Omega |\nabla y|^2\,dx - \int_\Omega (f + \bar{u})y\,dx$$

for all $y \in K$, so that $\bar{y}$ solves (1.1) with $u = \bar{u}$. If we use (3.16) for a sequence $\bar{y}_k \in K \cap H^2(\Omega)$ with $\bar{y}_k \to \bar{y}$ in $H^1(\Omega)$ we obtain

$$\frac{1}{2}\int_\Omega |\nabla \bar{y}_h|^2\,dx + \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(\bar{y}_h - \psi)^-]^4\}\,dx \le \frac{1}{2}\int_\Omega |\nabla I_h\bar{y}_k|^2\,dx + \int_\Omega (f + \bar{u}_h)(\bar{y}_h - I_h\bar{y}_k)\,dx,$$

from which we deduce that

$$\limsup_{h\to 0}\Big(\frac{1}{2}\int_\Omega |\nabla \bar{y}_h|^2\,dx + \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(\bar{y}_h - \psi)^-]^4\}\,dx\Big) \le \frac{1}{2}\int_\Omega |\nabla \bar{y}_k|^2\,dx + \int_\Omega (f + \bar{u})(\bar{y} - \bar{y}_k)\,dx$$

and hence after sending $k \to \infty$

$$\limsup_{h\to 0}\Big(\frac{1}{2}\int_\Omega |\nabla \bar{y}_h|^2\,dx + \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^4\}\,dx\Big) \le \frac{1}{2}\int_\Omega |\nabla \bar{y}|^2\,dx. \tag{3.17}$$

In particular, $\limsup_{h\to 0}\|\nabla \bar{y}_h\|^2_{L^2(\Omega)} \le \|\nabla \bar{y}\|^2_{L^2(\Omega)}$ and hence $\|\nabla \bar{y}_h\|^2_{L^2(\Omega)} \to \|\nabla \bar{y}\|^2_{L^2(\Omega)}$. Thus, we obtain together with (3.17) that

$$\bar{y}_h \to \bar{y} \text{ in } H^1(\Omega) \quad \text{and} \quad \frac{\gamma_h^3}{4}\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^4\}\,dx \to 0. \tag{3.18}$$

Next, let us introduce $\bar{\xi}, \bar{\mu} \in H^{-1}(\Omega)$ by

$$\begin{aligned}\langle \bar{\xi}, \phi\rangle &= \int_\Omega \nabla \bar{y}\cdot\nabla\phi\,dx - \int_\Omega (f+\bar{u})\phi\,dx,\\ \langle \bar{\mu}, \phi\rangle &= -\int_\Omega \nabla \bar{p}\cdot\nabla\phi\,dx + \int_\Omega (\bar{y}-y_0)\phi\,dx.\end{aligned}$$

Obviously, (3.3) and (3.5) are satisfied by definition, while (3.4) follows from the fact that $\bar{y}$ is a solution of (1.1). Let us next show that (3.6) holds. In view of (3.18), (3.12) and (2.3) we have

$$\begin{aligned}|\langle \bar{\mu}, \bar{y}-\psi\rangle| &\leftarrow \Big| -\int_\Omega \nabla \bar{p}_h\cdot\nabla(\bar{y}_h - I_h\psi)\,dx + \int_\Omega (\bar{y}_h-y_0)(\bar{y}_h - I_h\psi)\,dx\Big|\\ &= \Big|3\gamma_h^3\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^3\bar{p}_h\}\,dx\Big|\\ &\le 3\Big(\gamma_h^3\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^4\}\,dx\Big)^{\frac{1}{2}}\Big(\gamma_h^3\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^2\bar{p}_h^2\}\,dx\Big)^{\frac{1}{2}}\\ &\to 0\end{aligned}$$

by (3.18) and (3.10). Hence $\langle \bar{\mu}, \bar{y}-\psi\rangle = 0$ and in the same way we can show that $\langle \bar{\xi}, \bar{p}\rangle = 0$. Furthermore, (3.7) is an immediate consequence of (2.4). It remains to prove (3.8). Note that (2.6) and the definition of η_h imply

$$\bar{p}_h \ge \eta_h I_h[(\bar{y}_h-\psi)^+] - \frac{\eta_h}{3}I_h[(\bar{y}_h-\psi)^-] = \eta_h(\bar{y}_h - I_h\psi) - \frac{4}{3}\eta_h I_h[(\bar{y}_h-\psi)^-].$$

Letting $h \to 0$ we find that $\bar{p} \ge \eta(\bar{y}-\psi)$ a.e. in Ω, since $I_h[(\bar{y}_h-\psi)^-] \to 0$ in $L^1(\Omega)$ in view of (3.18) and (3.2). Finally, let $\phi \in C_0^\infty(\Omega), \phi \ge 0$ be arbitrary. We then have by (2.3) and (2.2) that

$$\begin{aligned}\langle \bar{\mu} - \eta\bar{\xi}, \phi\rangle &\leftarrow -\int_\Omega \nabla \bar{p}_h\cdot\nabla I_h\phi\,dx + \int_\Omega (\bar{y}_h - y_0)I_h\phi\,dx\\ &\quad -\eta_h\int_\Omega \nabla\bar{y}_h\cdot\nabla I_h\phi\,dx + \eta_h\int_\Omega (f+\bar{u}_h)I_h\phi\,dx\\ &= 3\gamma_h^3\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^2\bar{p}_h I_h\phi\}\,dx + \eta_h\gamma_h^3\int_\Omega I_h\{[(\bar{y}_h-\psi)^-]^3 I_h\phi\}\,dx\\ &= \gamma_h^3\sum_{\bar{y}_j<\psi_j}[(\bar{y}_j-\psi_j)^-]^2\{3\bar{p}_j + \eta_h(\bar{y}_j-\psi_j)\}\phi(x_j)m_j\\ &\ge 0,\end{aligned}$$

by the definition of η_h and since $\phi(x_j) \ge 0$. Hence $\bar{\mu} - \eta\bar{\xi} \ge 0$ and (3.8) holds. ■

In order to relate the system (3.3)–(3.8) to known stationarity concepts we briefly recall the notion of strong stationarity:

Definition 1 The point $(u, y, \xi) \in L^2(\Omega) \times H_0^1(\Omega) \times H^{-1}(\Omega)$ is called strongly stationary if there exists $p \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla y \cdot \nabla \phi \, dx = \int_\Omega (f + u)\phi \, dx + \langle \xi, \phi \rangle \quad \forall \phi \in H_0^1(\Omega) \tag{3.19}$$

$$y \geq \psi \text{ a.e. in } \Omega, \; \xi \geq 0, \; \langle \xi, y - \psi \rangle = 0 \tag{3.20}$$

$$p \in S_y, \quad \int_\Omega \nabla p \cdot \nabla \phi \, dx \leq \int_\Omega (y - y_0)\phi \, dx \quad \forall \phi \in S_y \tag{3.21}$$

$$\alpha u + p = 0 \quad \text{ a.e. in } \Omega, \tag{3.22}$$

where $S_y = \{\phi \in H_0^1(\Omega) \,|\, \phi \geq 0 \text{ q.e. on } Z_y, \langle \xi, \phi \rangle = 0\}$ and $Z_y = \{x \in \Omega \,|\, y(x) = \psi(x)\}$ (defined up to sets of zero capacity). Here, q.e. stands for quasi-everywhere.

It is shown in [14, Theorem 2.2] that a solution of $(\mathbb{P})$ is strongly stationary. This result was extended to the case of control constraints in [16]. In our next result we show that a solution of the system (3.3)–(3.8) is strongly stationary. Furthermore, we prove a continuous analogue of Theorem 2.2.

Theorem 3.2 *Suppose that $(u, y, p, \xi, \mu) \in L^2(\Omega) \times H_0^1(\Omega) \times H_0^1(\Omega) \times H^{-1}(\Omega) \times H^{-1}(\Omega)$ is a solution of (3.3)–(3.8) for some $\eta \leq 0$. Then there holds:*
a) (u, y, ξ) is strongly stationary.
b) If

$$|\eta| \leq \alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}, \tag{3.23}$$

then u is a global minimum for Problem $(\mathbb{P})$. If the inequality (3.23) is strict, then u is the unique global minimum.

Proof: a) We only have to prove (3.21). To do so, we shall make use of some basic results from capacity theory for which we refer the reader to [16, Section 2]. Since $p \geq \eta(y - \psi)$ a.e. in Ω we deduce from [16, Lemma 2.3] that $p \geq \eta(y - \psi)$ q.e. on Ω and hence that $p \geq 0$ q.e. on Z_y. Combining this relation with (3.6) we infer that $p \in S_y$. Next, (3.8) implies that $\lambda := \mu - \eta\xi$ is a nonnegative functional in $H^{-1}(\Omega)$. Hence there exists a regular Borel measure (also denoted by λ) such that

$$\langle \lambda, \phi \rangle = \int_\Omega \phi \, d\lambda, \quad \phi \in H_0^1(\Omega),$$

where we integrate the quasi-continuous representative of ϕ. Lemma 2.4 in [16] yields that $y \geq \psi$ λ-a.e. in Ω so that we have for every compact set $C \subset \Omega$

$$0 \leq \int_C (y - \psi) \, d\lambda \leq \int_\Omega (y - \psi) \, d\lambda = \langle \lambda, y - \psi \rangle = 0$$

in view of (3.4) and (3.6). Thus $\lambda(C \cap \{y > \psi\}) = 0$ for all compact sets $C \subset \Omega$ and hence $\lambda(\{y > \psi\}) = 0$. We deduce for every $\phi \in S_y$

$$\langle \mu, \phi \rangle = \langle \mu - \eta\xi, \phi \rangle = \int_\Omega \phi \, d\lambda = \int_{\{y>\psi\}} \phi \, d\lambda + \int_{\{y=\psi\}} \phi \, d\lambda = \int_{Z_y} \phi \, d\lambda \geq 0,$$

since $\phi \geq 0$ q.e. (and hence also λ-a.e.) on Z_y. Combining this result with (3.5) we infer

$$\int_\Omega \nabla p \cdot \nabla \phi \, dx = \int_\Omega (y - y_0)\phi \, dx - \langle \mu, \phi \rangle \leq \int_\Omega (y - y_0)\phi \, dx \qquad \forall \phi \in S_y,$$

and hence (3.20) is satisfied.

b) Let $v \in L^2(\Omega)$ be arbitrary and $\tilde{y} \in K$ the solution of

$$\int_\Omega \nabla \tilde{y} \cdot \nabla(\phi - \tilde{y}) dx \geq \int_\Omega (f + v)(\phi - \tilde{y}) dx \qquad \forall \phi \in K.$$

Defining $\tilde{\xi} \in H^{-1}(\Omega)$ by

$$\langle \tilde{\xi}, \phi \rangle := \int_\Omega \nabla \tilde{y} \cdot \nabla \phi \, dx - \int_\Omega (f + v)\phi \, dx,$$

we have that $\tilde{\xi} \geq 0$ and $\langle \tilde{\xi}, \tilde{y} - \psi \rangle = 0$. Similarly as in the proof of Theorem 2.2 we calculate

$$J(v) - J(u) = \frac{1}{2}\|\tilde{y} - y\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u\|^2_{L^2(\Omega)} + \int_\Omega (y - y_0)(\tilde{y} - y) \, dx + \alpha \int_\Omega u(v - u) \, dx. \quad (3.24)$$

We infer from (3.5), (3.3) and the definition of $\tilde{\xi}$ that

$$\begin{aligned} \int_\Omega (y - y_0)(\tilde{y} - y) \, dx &= \int_\Omega \nabla p \cdot \nabla(\tilde{y} - y) \, dx + \langle \mu, \tilde{y} - y \rangle \\ &= \int_\Omega p(v - u) \, dx + \langle \tilde{\xi} - \xi, p \rangle + \langle \mu, \tilde{y} - y \rangle. \end{aligned}$$

Inserting this relation into (3.24) and recalling (3.4), (3.6) and (3.8) we derive

$$\begin{aligned} J(v) - J(u) &= \frac{1}{2}\|\tilde{y} - y\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u\|^2_{L^2(\Omega)} + \langle \tilde{\xi}, p \rangle + \langle \mu, \tilde{y} - \psi \rangle \\ &\geq \frac{1}{2}\|\tilde{y} - y\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u\|^2_{L^2(\Omega)} + \eta \, \langle \tilde{\xi}, y - \psi \rangle + \eta \, \langle \xi, \tilde{y} - \psi \rangle \quad (3.25) \\ &= \frac{1}{2}\|\tilde{y} - y\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|v - u\|^2_{L^2(\Omega)} - \eta \langle \tilde{\xi} - \xi, \tilde{y} - y \rangle, \end{aligned}$$

since $\tilde{y} - \psi \geq 0$ and $\langle \tilde{\xi}, \tilde{y} - \psi \rangle = 0$. Using once more (3.3), the definition of $\tilde{\xi}$ and recalling (2.5) we may write

$$\begin{aligned} \langle \tilde{\xi} - \xi, \tilde{y} - y \rangle &= \int_\Omega |\nabla(\tilde{y} - y)|^2 \, dx - \int_\Omega (\tilde{y} - y)(v - u) \, dx \\ &\geq \lambda_1 \int_\Omega |\tilde{y} - y|^2 \, dx - \int_\Omega (\tilde{y} - y)(v - u) \, dx. \end{aligned}$$

If we multiply this relation by $-\eta = |\eta|$ and insert it into (3.25) we obtain

$$J(v) - J(u) \geq \int_\Omega [(\frac{1}{2} + |\eta|\lambda_1)|\tilde{y} - y|^2 + \frac{\alpha}{2}|v - u|^2 - |\eta|(v - u)(\tilde{y} - y)] \, dx$$

and the result follows in the same way as in the proof of Theorem 2.2. ■

As an immediate consequence we have:

Corollary 3.3 *Let $(\bar{u}_h, \bar{y}_h, \bar{p}_h)_{0<h\le h_0}$ be a sequence of solutions of (2.2)–(2.4) with corresponding $\eta_h \le 0$ given by (2.7) and suppose that*

$$|\eta_h| \le \alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}, \quad 0 < h \le h_0. \tag{3.26}$$

Then

$$\bar{u}_h \to \bar{u} \text{ in } L^2(\Omega) \text{ for a subsequence } h \to 0,$$

where $\bar{u}$ is a global minimum for Problem $(\mathbb{P})$. If

$$|\eta_h| \le \kappa\alpha\big(\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}\big), \quad 0 < h \le h_0, \tag{3.27}$$

for some $0 < \kappa < 1$, then $\bar{u}$ is the unique global solution of $(\mathbb{P})$ and the whole sequence $(\bar{u}_h)_{0<h\le h_0}$ converges to $\bar{u}$.

Proof: Let us denote by $\hat{y}_h$ the solution of (2.1) with $u \equiv 0$. In the same way as at the beginning of the proof of Theorem 3.1 we can show that $\|\hat{y}_h\|_{H^1(\Omega)} \le c$. It follows from (3.26) and Theorem 2.2 that $\bar{u}_h$ is a solution of $(\mathbb{P}_h^{\gamma_h})$, so that in particular $J_h^{\gamma_h}(\bar{u}_h) \le J_h^{\gamma_h}(0)$, $0 < h \le h_0$ and therefore

$$\frac{\alpha}{2}\|\bar{u}_h\|^2_{L^2(\Omega)} \le \frac{1}{2}\|\hat{y}_h - y_0\|^2_{L^2(\Omega)} \le c. \tag{3.28}$$

Combining (3.28) with (3.26) we may infer from Theorem 3.1 that there exists a subsequence $h \to 0$ and a solution $(\bar{u}, \bar{y}, \bar{p}, \bar{\xi}, \bar{\mu})$ of (3.3)–(3.8) such that

$$\bar{u}_h \to \bar{u} \text{ in } L^2(\Omega),\ \bar{y}_h \to \bar{y} \text{ in } H^1(\Omega),\ \eta_h \to \eta \le 0.$$

Since $|\eta| \le \alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}$ it follows from Theorem 3.2 b) that $\bar{u}$ is a global minimum for Problem $(\mathbb{P})$. If (3.27) holds, then the above inequality is strict and the minimum is unique. ■

4 The unpenalised case

In this section we briefly discuss how our theory can be adapted to the case when the state is approximated by a discrete version of the variational inequality (1.1), namely we consider the following discrete control problem:

$$(\mathbb{P}_h) \qquad \min_{u\in L^2(\Omega)} J_h(u) = \frac{1}{2}\|y_h - y_0\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}$$

subject to

$$y_h \in K_h, \quad \int_\Omega \nabla y_h \cdot \nabla(\phi_h - y_h)dx \ge \int_\Omega (f+u)(\phi_h - y_h)dx \qquad \forall \phi_h \in K_h, \tag{4.1}$$

where

$$K_h = \{\phi_h \in X_{h0} \,|\, \phi_h(x) \ge (I_h\psi)(x), x \in \Omega\}.$$

Existence of a solution of $(\mathbb{P}_h)$ is shown in [11, Section 3]. We shall formulate the necessary first order optimality conditions in matrix/vector form. To do so, let us define the mass matrix $\mathcal{M}$ and the stiffness matrix $\mathcal{A}$, i.e.

$$\mathcal{M}_{ij} := \int_\Omega \phi_i\phi_j\,dx, \quad \mathcal{A}_{ij} := \int_\Omega \nabla\phi_i \cdot \nabla\phi_j\,dx, \quad i,j = 1,\dots,n.$$

Introducing a slack variable $\boldsymbol{\xi} \in \mathbb{R}^n$, problem (4.1) can be written as

$$\mathcal{A}\mathbf{y} = \Big(\int_\Omega (f+u)\phi_i \, dx\Big)_{i=1}^n + \boldsymbol{\xi}$$
$$y_j \geq \psi_j, \xi_j \geq 0, \xi_j(y_j - \psi_j) = 0, j = 1, \dots, n.$$

Here, $y_h = \sum_{j=1}^n y_j \phi_j$ and $\mathbf{y} = (y_j)_{j=1}^n$. The following result is proved in [11, Theorem 4.1]. Note that the system given below slightly differs from the one in [11] in that we have replaced $\boldsymbol{\mu}$ by $-\boldsymbol{\mu}$.

Theorem 4.1 *Let $u_h \in L^2(\Omega)$ be a local optimal solution of $(\mathbb{P}_h)$ with associated state $y_h \in X_{h0}$ and slack variable $\boldsymbol{\xi} \in \mathbb{R}^n$. Then there exist an adjoint state $p_h \in X_{h0}$ and a multiplier $\boldsymbol{\mu} \in \mathbb{R}^n$ such that the following strong stationarity system is satisfied:*

$$\mathcal{A}\mathbf{y} = \Big(\int_\Omega (f+u_h)\phi_i \, dx\Big)_{i=1}^n + \boldsymbol{\xi} \tag{4.2}$$
$$y_j \geq \psi_j, \xi_j \geq 0, \xi_j(y_j - \psi_j) = 0, j = 1, \dots, n \tag{4.3}$$
$$\mathcal{A}^T\mathbf{p} = \mathcal{M}\mathbf{y} - \Big(\int_\Omega y_0\phi_i \, dx\Big)_{i=1}^n - \boldsymbol{\mu} \tag{4.4}$$
$$(y_j - \psi_j)\mu_j = 0, \ \xi_j p_j = 0, j = 1, \dots, n \tag{4.5}$$
$$\alpha u_h + p_h = 0 \quad \text{in } \Omega \tag{4.6}$$
$$\mu_k \geq 0, \ p_k \geq 0 \text{ for all } k \in \{1, \dots, n\} \text{ with } y_k - \psi_k = \xi_k = 0. \tag{4.7}$$

Here, $y_h = \sum_{j=1}^n y_j\phi_j, p_h = \sum_{j=1}^n p_j \phi_j$.

In practice, the system (4.2)–(4.6) can be solved with the help of a primal–dual active set strategy, see [8, Section 6]. The corresponding numerical experiments indicate that this method typically enters into a cycle in the presence of bi-active sets. This is one of the reasons to base our numerical treatment on the Moreau-Yosida relaxed version $(\mathbb{P}_h^\gamma)$ of the original optimal control problem $(\mathbb{P})$.

The following result is the analogue of Theorem 2.2. We remark that the quantity η in (4.8) has been used in [5] (see (4.19)) in order to show the equivalence between certain optimality systems.

Theorem 4.2 *Suppose that $(u_h, y_h, p_h, \boldsymbol{\xi}, \boldsymbol{\mu}) \in L^2(\Omega) \times X_{h0} \times X_{h0} \times \mathbb{R}^n \times \mathbb{R}^n$ is a solution of (4.2)-(4.7) and define*

$$\eta := \min\Big(\min_{y_k > \psi_k} \frac{p_k}{y_k - \psi_k}, \min_{\xi_k > 0} \frac{\mu_k}{\xi_k}, 0 \Big). \tag{4.8}$$

If

$$|\eta| \leq \alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha}, \tag{4.9}$$

then u_h is a global minimum for Problem $(\mathbb{P}_h)$. If the inequality (4.9) is strict, then u_h is the unique global minimum.

Proof: The proof is essentially a discrete version of the proof of Theorem 3.2 b). Note that (4.7) and the definition of η imply that

$$p_j \geq \eta(y_j - \psi_j), \ \mu_j \geq \eta \, \xi_j, \ j = 1, \dots, n$$

so that $(u_h, y_h, p_h, \boldsymbol{\xi}, \boldsymbol{\mu})$ satisfies a discrete analogue of (3.3)–(3.8). ■

It is not difficult to see that the result of Theorem 3.1 also holds for a sequence of solutions $(\bar{u}_h, \bar{y}_h, \bar{p}_h)_{0<h\le h_0}$ of (4.2)-(4.7) satisfying the bounds (3.2). The corresponding arguments in fact become a little bit easier because the penalisation term is no longer present. Furthermore, the convergence result in Corollary 3.3 holds as well. We omit the details.

5 Numerical examples

In this section we apply Theorem 2.2 to some numerical examples taken from [8], [11] and [12]. In Examples 1–3 the computational domain is given by $\Omega := (0,1)\times(0,1)$, and consequently, the constant λ_1 from (2.5) has the value $\lambda_1 = 2\pi^2$. On the other hand, Example 4 is formulated on the L-shaped domain $\Omega := (-1,0)\times(-1,1)\cup[0,1)\times(0,1)$ and λ_1 is approximated by solving a generalized eigenvalue problem leading to $\lambda_1 \approx 9.63977851$. In all of the examples, the domain Ω is partitioned using a uniform triangulation with mesh size $h = 2^{-6}\sqrt{2}$.

We solve (2.2),(2.3),(2.4) using Newton's method with the stopping criterion

$$\frac{1}{\alpha}\|p_h^{(k)} - p_h^{(k+1)}\|_{L^2(\Omega)} \le 10^{-15},$$

where $p_h^{(k)}$ denotes the discrete adjoint variable corresponding to the k-th iteration. We take the zero point as an initial guess for Newton's method and initialize our $\gamma-$homotopy with $\gamma = 1$. As the value of γ increases we take the solution of the system (2.2),(2.3),(2.4) at the preceding value of γ as starting value in the current Newton iteration.

We introduce the sets of nodes

$$\begin{aligned}\mathcal{N}^+ &:= \{k \in \{1,\dots,n\} : y_k - \psi_k > 0\},\\ \mathcal{N}^0 &:= \{k \in \{1,\dots,n\} : y_k - \psi_k = 0\}, \text{ and}\\ \mathcal{N}^- &:= \{k \in \{1,\dots,n\} : y_k - \psi_k < 0\}.\end{aligned}$$

Then, one can see from the quantity $\min\limits_{k\in\mathcal{N}^-}(y_k - \psi_k)$ the amount by which the state violates the obstacle constraint, which typically should tend to zero as the parameter γ increases. We point out that the equality $y_k = \psi_k$ is often difficult to be observed on computers. In fact, it has never been detected when performing computations for the considered examples. Hence, we consider directly condition (2.8) as the set $\mathcal{N}^0$ is empty. We shall report for each example the values of the quantities η, $\min\limits_{k\in\mathcal{N}^-}(y_k - \psi_k)$ and the number of Newton iterations, denoted by $\#N$, as we increase the value of the penalization parameter γ. We stop increasing the parameter γ once the linear system in Newton's method becomes too ill-conditioned. All the computations are done using MATLAB R2018a.

In what follows we refer to the Lagrange interpolations of

$$-\gamma^3[(y_h - \psi)^-]^3 \text{ and } -3\gamma^3[(y_h-\psi)^-]^2 p_h$$

as multipliers ξ and μ, respectively.

Example 1 This is the Example 6.4 from [8] with f, y_0, ψ replaced by $-f$, $-y_0$ and $-\psi$. Here we choose the following data for $(\mathbb{P})$:

$$\alpha = 10^{-1}, \quad y_0(x) = -(5x_1 + x_2 - 1) \text{ in } \Omega, \quad f(x) = -0.1 \text{ in } \Omega,$$

$$\psi(x) = -4\big(x_1(x_1-1) + x_2(x_2-1)\big) - 1.5 \text{ in } \Omega.$$

We have

$$\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 3.9730.$$

The numerical results are reported in Table 1.

We see that the condition (2.8) holds for the considered values of γ, which indicates that the computed solution is the unique global minimum of $(\mathbb{P}_h^\gamma)$. Graphical illustration of the solution is provided in Figure 1 for $\gamma = 10^8$. We also observe that the violation of the obstacle constraint satisfies $\min_{k\in\mathcal{N}^-}(y_k - \psi_k) \sim -\gamma^{-1}$.

Table 1 Example 1 where $\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 3.9730$.

γ	$\min_{k\in\mathcal{N}^+} \frac{p_k}{y_k-\psi_k}$	$\min_{k\in\mathcal{N}^-} \frac{3p_k}{\psi_k-y_k}$	η	$\min_{k\in\mathcal{N}^-}(y_k-\psi_k)$	$\#N$
1.0e+00	-4.10942407e-05	6.49648857e-01	-4.10942407e-05	-5.80096766e-01	5
1.0e+01	-3.04816156e-04	2.56513024e-01	-3.04816156e-04	-2.28879959e-01	8
1.0e+02	-5.20715816e-04	1.23095522e-01	-5.20715816e-04	-2.52527352e-02	11
1.0e+03	-5.46497352e-04	9.94266004e-02	-5.46497352e-04	-2.52516115e-03	12
1.0e+04	-5.49770680e-04	9.06835069e-02	-5.49770680e-04	-2.52508949e-04	13
1.0e+05	-5.50113221e-04	8.63210169e-02	-5.50113221e-04	-2.52508176e-05	15
1.0e+06	-5.50622631e-04	8.39380400e-02	-5.50622631e-04	-2.52524125e-06	12
1.0e+07	-5.50679810e-04	8.34709569e-02	-5.50679810e-04	-2.52538920e-07	11
1.0e+08	-5.50685642e-04	8.34241524e-02	-5.50685642e-04	-2.52542988e-08	11
1.0e+09	-5.50686226e-04	8.34194702e-02	-5.50686226e-04	-2.52543397e-09	11
1.0e+10	-5.50686284e-04	8.34190005e-02	-5.50686284e-04	-2.52543431e-10	10
1.0e+11	-5.50686290e-04	8.34189542e-02	-5.50686290e-04	-2.52543542e-11	10
1.0e+12	-5.50686291e-04	8.34185828e-02	-5.50686291e-04	-2.52542431e-12	10
1.0e+13	-5.50686291e-04	8.34167493e-02	-5.50686291e-04	-2.52520227e-13	9
1.0e+14	-5.50686291e-04	8.33984159e-02	-5.50686291e-04	-2.52575738e-14	9
1.0e+15	-5.50686291e-04	8.43368891e-02	-5.50686291e-04	-2.49800181e-15	8

Example 2 This is Example 6.5 from [8] with lack of strict complementarity, where f, y_0 are replaced by $-f$ and $-y_0$. Here we choose for $(\mathbb{P})$ the data

$$\alpha = 10^{-1}, \quad y_0(x) = -(5x_1 + x_2 - 1) \text{ in } \Omega, \quad f(x) = -(x_1 - \frac{1}{2}) \text{ in } \Omega, \quad \psi(x) = 0 \text{ in } \Omega.$$

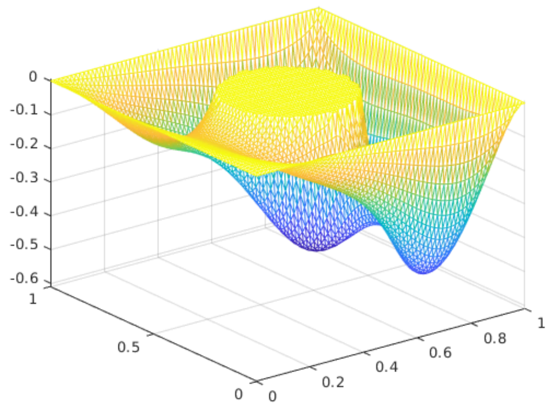
We have

$$\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 3.9730.$$

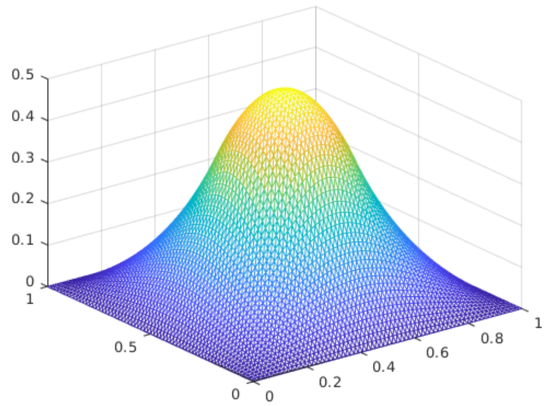
The numerical results are reported in Table 2. We see that the condition (2.8) is satisfied for the considered values of γ, and hence, the unique global solution has been computed, which is presented in Figure 2 for $\gamma = 10^8$. We again observe that the violation of the obstacle constraint satisfies $\min_{k\in\mathcal{N}^-}(y_k - \psi_k) \sim -\gamma^{-1}$. This is also true in the subsequent examples.

Example 3 The data of this Example 1 is taken from [11], where an exact solution for $(\mathbb{P})$ is constructed as follows;

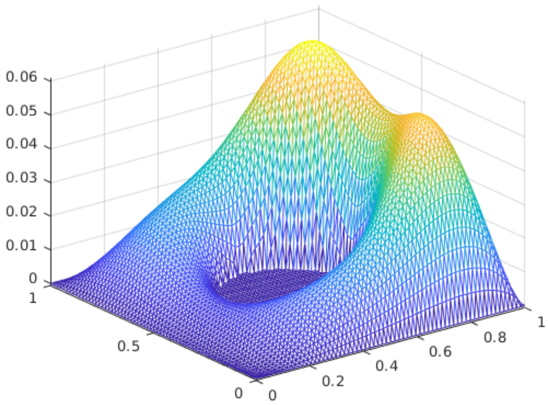
$$\alpha = 1, \quad \psi = 0 \text{ in } \Omega,$$



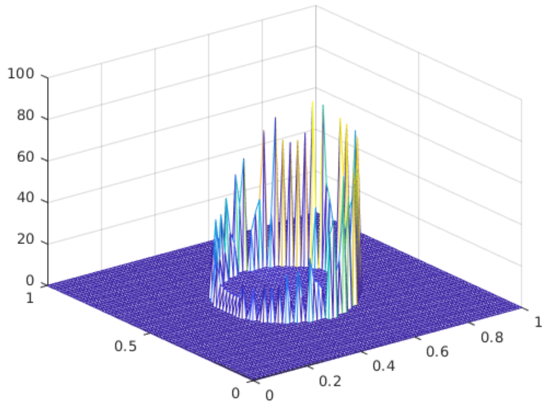
(a) The optimal control



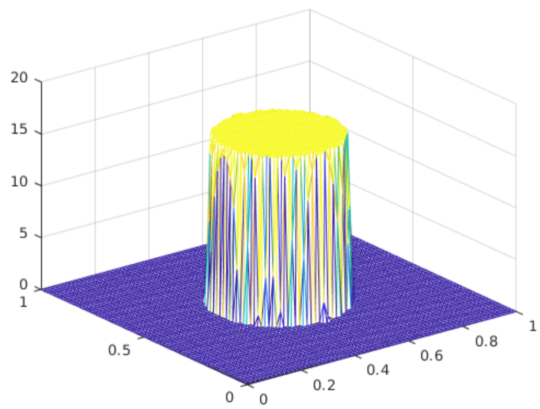
(b) The optimal state



(c) The optimal adjoint state

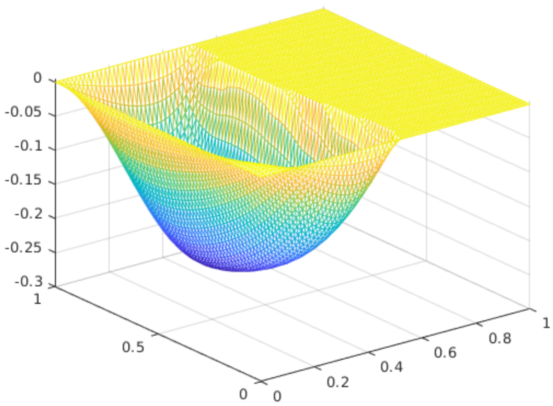


(d) The multiplier μ

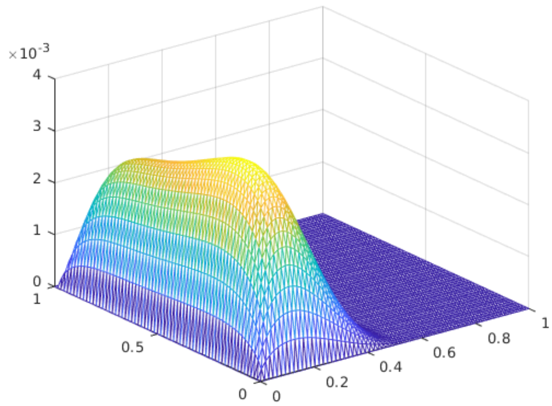


(e) The multiplier ξ

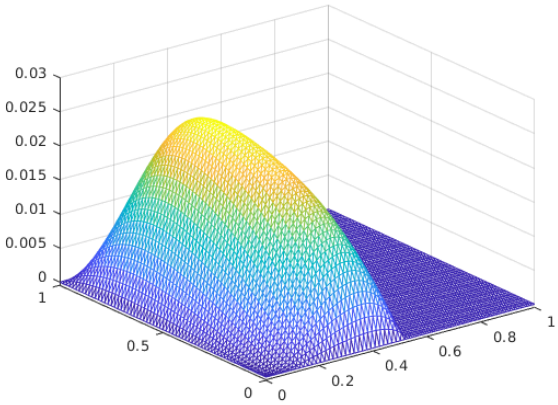
Figure 1 Example 1: The optimal control, state, adjoint state for $\gamma = 10^8$.



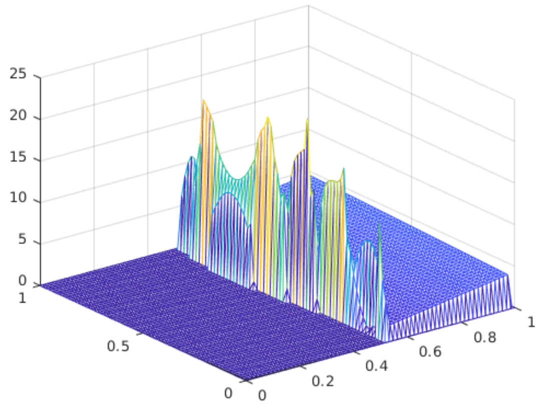
(a) The optimal control



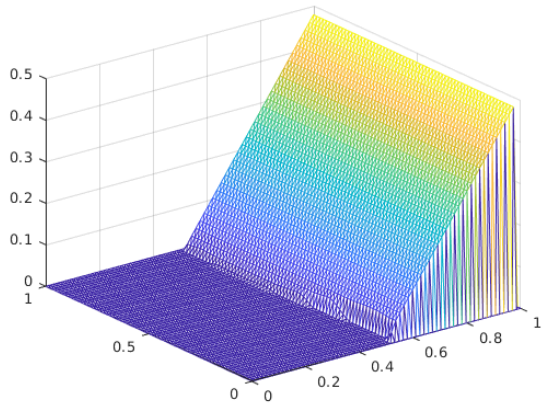
(b) The optimal state



(c) The optimal adjoint state



(d) The multiplier μ



(e) The multiplier ξ

Figure 2 Example 2: The optimal control, state and adjoint state for $\gamma = 10^8$.

Table 2 Example 2 where $\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 3.9730$.

γ	$\min_{k\in\mathcal{N}^+} \frac{p_k}{y_k - \psi_k}$	$\min_{k\in\mathcal{N}^-} \frac{3p_k}{\psi_k - y_k}$	η	$\min_{k\in\mathcal{N}^-} (y_k - \psi_k)$	#N
1.0e+00	-9.18374766e-01	5.31348016e+00	-9.18374766e-01	-8.34645660e-02	4
1.0e+01	-9.72376118e-01	5.75321935e+00	-9.72376118e-01	-5.91117294e-02	6
1.0e+02	-1.01940110e+00	8.79601682e+00	-1.01940110e+00	-7.28493127e-03	11
1.0e+03	-1.02120448e+00	9.45008887e+00	-1.02120448e+00	-7.34501977e-04	12
1.0e+04	-1.02054664e+00	8.61664669e+00	-1.02054664e+00	-7.65785826e-05	12
1.0e+05	-1.02044429e+00	8.29383499e+00	-1.02044429e+00	-7.76577329e-06	13
1.0e+06	-1.02030934e+00	8.19782461e+00	-1.02030934e+00	-7.83603514e-07	15
1.0e+07	-1.02028019e+00	8.13627555e+00	-1.02028019e+00	-7.85170648e-08	13
1.0e+08	-1.02027779e+00	8.12975658e+00	-1.02027779e+00	-7.85327708e-09	11
1.0e+09	-1.02027759e+00	8.12910469e+00	-1.02027759e+00	-7.85343418e-10	11
1.0e+10	-1.02027757e+00	8.12903950e+00	-1.02027757e+00	-7.85344989e-11	10
1.0e+11	-1.02027757e+00	8.12903298e+00	-1.02027757e+00	-7.85345146e-12	10
1.0e+12	-1.02027757e+00	8.12903233e+00	-1.02027757e+00	-7.85345161e-13	10

$$f(x) = -\Delta y(x) - \xi(x) + \frac{1}{\alpha} p(x) \text{ in } \Omega$$
$$y_0(x) = \begin{cases} y(x) + \Delta p_1(Q^t x) & \text{in } \Omega_1, \\ y(x) & \text{otherwise.} \end{cases}$$

Here y denotes the optimal state with the corresponding adjoint state p and slackness variable ξ, which are defined by

$$y(x_1, x_2) = \begin{cases} y_1(x_1) \cdot y_2(x_2) & \text{in } (0, 0.5) \times (0, 0.8), \\ 0 & \text{otherwise,} \end{cases}$$
$$p(x) = \begin{cases} p_1(Q^t x) & \text{in } \Omega_1, \\ 0 & \text{otherwise,} \end{cases}$$
$$\xi(x_1, x_2) = \begin{cases} y_1(x_1 - 0.5) \cdot y_2(x_2) & \text{in } (0.5, 1) \times (0, 0.8), \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, y_1, y_2, and p_1 are the functions

$$y_1(x_1) = -4096x_1^6 + 6144x_1^5 - 3072x_1^4 + 512x_1^3,$$
$$y_2(x_2) = -244.140625x_2^6 + 585.9375x_2^5 - 468.75x_2^4 + 125x_2^3,$$
$$p_1(x_1, x_2) = (-200(x_1 - 0.8)^2 + 0.5)(-200(x_2 - 0.9)^2 + 0.5),$$

and Ω_1 is the square with midpoint (0.8,0.9) and edge length 0.1 after being rotated by the matrix

$$Q = \begin{bmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{bmatrix}$$

around its midpoint. This example contains the biactive set

$$\{x \in \Omega; y(x) = \xi(x) = 0\} \equiv [0, 1] \times [0.8, 1],$$

which makes its numerical treatment challenging. Furthermore should we note that in this example the cost functional is of the form

$$J(y,u) = \frac{1}{2}\|y - y_0\|^2_{L^2(\Omega)} + \frac{1}{2}\|u - u_d\|^2_{L^2(\Omega)}$$

with $u_d = u + \frac{1}{\alpha}p$. We account for u_d in the setting of f above. Our theory also is valid in this situation without further modifications, see the proof of Lemma 2.1.

From the previous data we have

$$\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 39.5037.$$

We provide the numerical results in Table 3. Again, the unique global minimum has been computed and the corresponding graphs are illustrated in Figure 3 when $\gamma = 10^8$.

Table 3 Example 3 with $\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 39.5037$

γ	$\min_{k\in\mathcal{N}^+} \frac{p_k}{y_k - \psi_k}$	$\min_{k\in\mathcal{N}^-} \frac{3p_k}{\psi_k - y_k}$	η	$\min_{k\in\mathcal{N}^-} (y_k - \psi_k)$	#N
1.0e+00	1.63643751e-02	3.74767766e+00	0.00000000e+00	-2.10533852e-02	3
1.0e+01	1.58450462e-02	3.61352870e+00	0.00000000e+00	-2.07865748e-02	4
1.0e+02	1.97319240e-03	1.04106448e-01	0.00000000e+00	-8.25007721e-03	8
1.0e+03	-1.25202529e-02	-3.51210686e-01	-3.51210686e-01	-9.83097618e-04	11
1.0e+04	-2.42618027e-01	-3.37078927e-01	-3.37078927e-01	-9.97061633e-05	12
1.0e+05	-1.36127707e-01	-1.12485742e-01	-1.36127707e-01	-1.02473476e-05	13
1.0e+06	-9.75060053e-02	-3.06984382e-02	-9.75060053e-02	-1.03346212e-06	15
1.0e+07	-8.26915590e-02	-3.18139489e-02	-8.26915590e-02	-1.03449193e-07	13
1.0e+08	-1.81555278e-01	-3.47707710e-02	-1.81555278e-01	-1.03459974e-08	14
1.0e+09	-2.20032200e-01	-4.44457854e-02	-2.20032200e-01	-1.03461058e-09	15
1.0e+10	-2.24336956e-01	-4.55306972e-02	-2.24336956e-01	-1.03461166e-10	17
1.0e+11	-2.24772813e-01	-4.56405644e-02	-2.24772813e-01	-1.03461177e-11	16

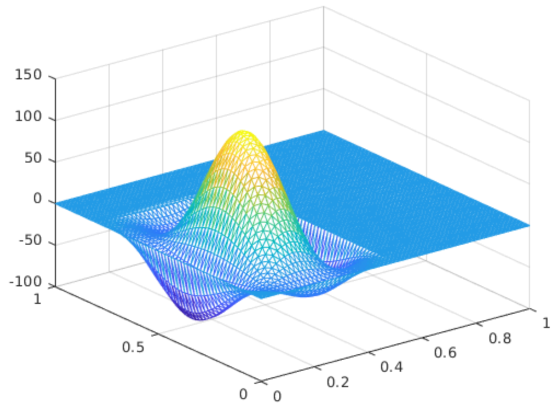
Example 4 The data for this problem is taken from Example 6.2 in [12] and for ($\mathbb{P}$) reads

$$\alpha = 1, \quad \psi(x) = 0 \text{ in } \Omega, \quad f(x) = \frac{1}{2} + \frac{1}{2}(x_1 - x_2) \text{ in } \Omega,$$
$$y_0(x) = \begin{cases} -1 & \text{if } |x| \geq 0.1, \\ 1 - 100x_1^2 - 50x_2^2 & \text{otherwise.} \end{cases}$$

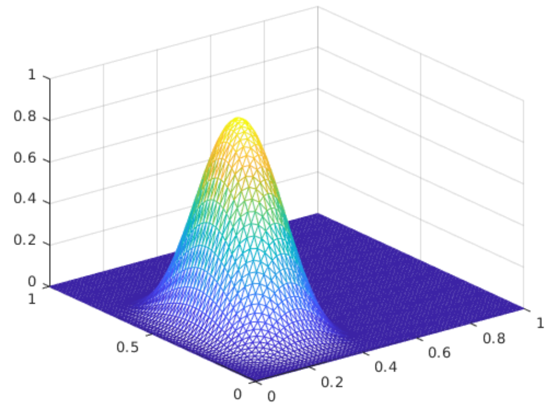
We have

$$\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 19.3313.$$

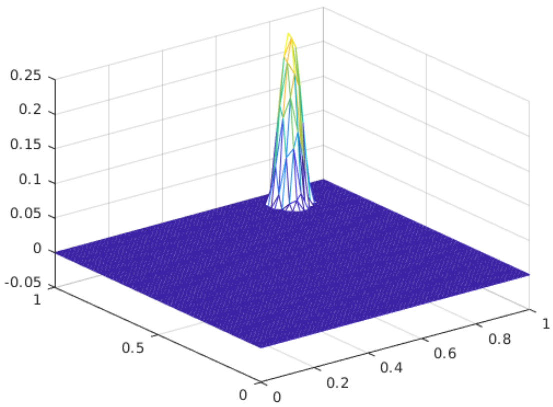
The numerical results are reported in Table 4. The condition (2.8) is satisfied, and hence, the unique global solution has been computed, which we illustrate in Figure 4 for $\gamma = 10^8$.



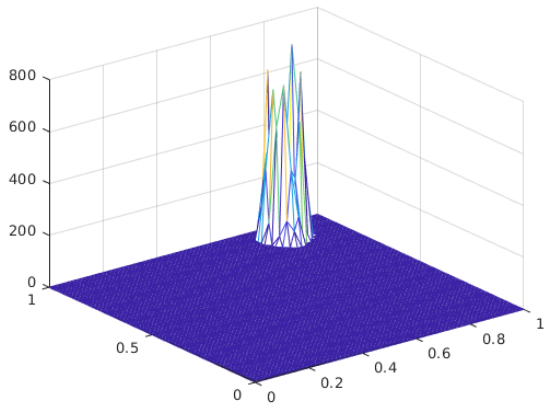
(a) The optimal control



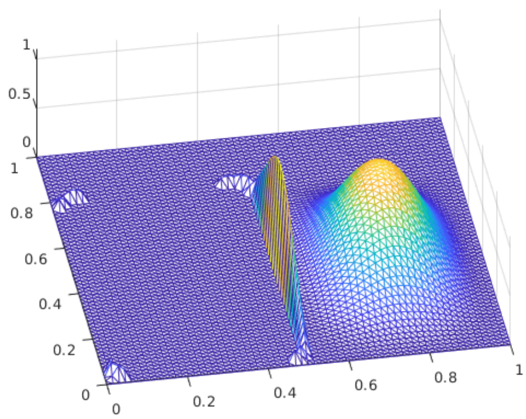
(b) The optimal state



(c) The optimal adjoint state

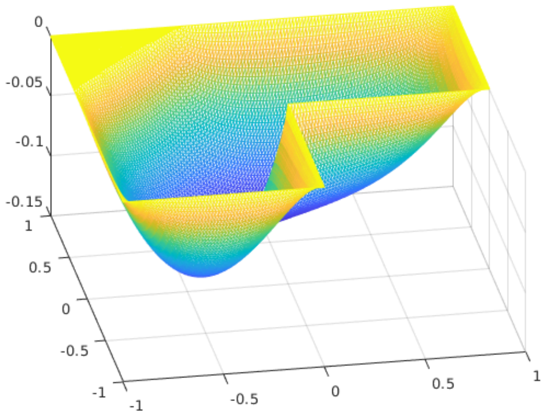


(d) The multiplier μ

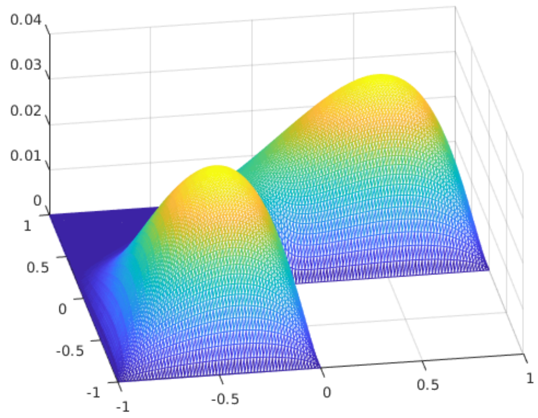


(e) The multiplier ξ

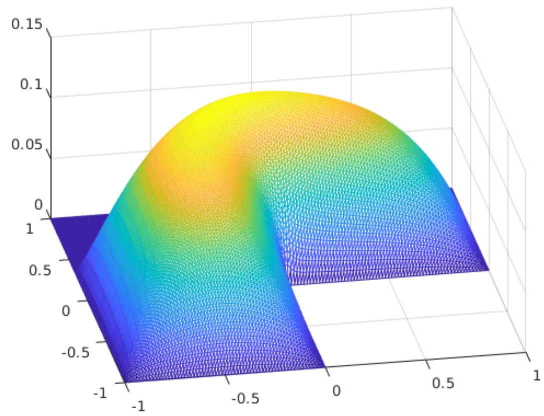
Figure 3 Example 3: The optimal control, state, adjoint state, and the multipliers for $\gamma = 10^8$.



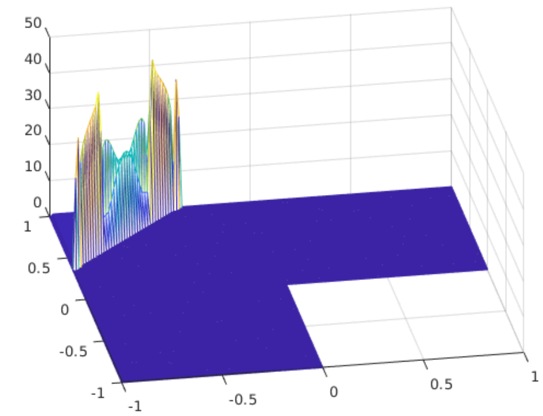
(a) The optimal control



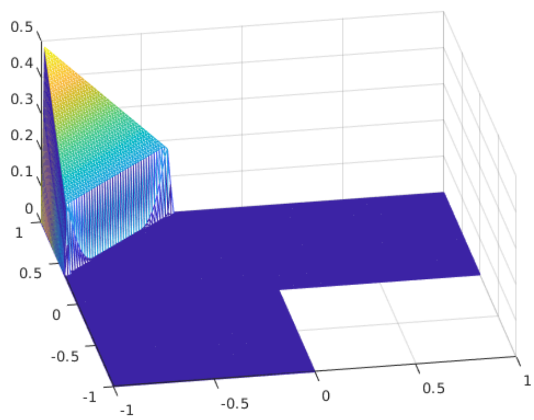
(b) The optimal state



(c) The optimal adjoint state



(d) The multiplier μ



(e) The multiplier ξ

Figure 4 Example 4: The optimal control, state, adjoint state, and the multipliers for $\gamma = 10^8$.

Table 4 Example 4 where $\alpha\lambda_1 + \sqrt{\alpha^2\lambda_1^2 + \alpha} \approx 19.3313$.

γ	$\min_{k\in\mathcal{N}^+} \frac{p_k}{y_k - \psi_k}$	$\min_{k\in\mathcal{N}^-} \frac{3p_k}{\psi_k - y_k}$	η	$\min_{k\in\mathcal{N}^-} (y_k - \psi_k)$	#N
1.0e+00	1.28191544e+00	9.12899685e+00	0.00000000e+00	-7.61042379e-03	3
1.0e+01	1.28189321e+00	9.11848759e+00	0.00000000e+00	-7.59555549e-03	3
1.0e+02	1.27854041e+00	7.14870263e+00	0.00000000e+00	-4.77178798e-03	7
1.0e+03	1.27598179e+00	2.95273878e+00	0.00000000e+00	-7.05858478e-04	11
1.0e+04	1.27568823e+00	2.30861182e+00	0.00000000e+00	-7.60379296e-05	12
1.0e+05	1.27561164e+00	2.14128368e+00	0.00000000e+00	-7.76063518e-06	12
1.0e+06	1.27562640e+00	2.08304474e+00	0.00000000e+00	-7.81857173e-07	13
1.0e+07	1.27563830e+00	2.06636877e+00	0.00000000e+00	-7.84995677e-08	15
1.0e+08	1.27563972e+00	2.06470139e+00	0.00000000e+00	-7.85310208e-09	11
1.0e+09	1.27563987e+00	2.06453466e+00	0.00000000e+00	-7.85341667e-10	11
1.0e+10	1.27563988e+00	2.06451798e+00	0.00000000e+00	-7.85344814e-11	10
1.0e+11	1.27563988e+00	2.06451631e+00	0.00000000e+00	-7.853345128e-12	10
1.0e+12	1.27563989e+00	2.06451615e+00	0.00000000e+00	-7.85345160e-13	10

References

[1] A. Ahmad Ali, K. Deckelnick and M. Hinze, *Global minima for semilinear optimal control problems*. Comput. Optim. Appl. **65**, 261–288 (2016).

[2] A. Ahmad Ali, K. Deckelnick and M. Hinze, *Error analysis for global minima of semilinear optimal control problems*. Mathematical Control and related Fields **8**, 195–215 (2018).

[3] V. Barbu, *Optimal control of variational inequalities*. Res. Notes Math. **100**, Pitman, Boston, 1984.

[4] S. Bartels, *Numerical methods for nonlinear partial differential equations*. Springer Series in Computational Mathematics, **47**. Springer, Cham, 2015.

[5] M. Bergounioux and F. Mignot, *Optimal control of obstacle problems: existence of Lagrange multipliers*. ESAIM Control Optim. Calc. Var. **5**, 45–70 (2000).

[6] M. Hintermüller, *Inverse coefficient problems for variational inequalities: Optimality conditions and numerical realization*. ESAIM Math. Model. Numer. Anal. **35**, 129–152 (2001).

[7] M. Hintermüller and I. Kopacka, *A smooth penalty approach and a nonlinear multigrid algorithm for elliptic MPECs*. Comput. Optim. Appl. **50**, 111–145 (2011).

[8] K. Ito and K. Kunisch, *Optimal control of elliptic variational inequalities*. Appl. Math. Optim. **41**, 343–364 (2000).

[9] K. Kunisch and D. Wachsmuth, *Path–following for optimal control of stationary variational inequalities*. Comput. Opt. Appl. **51**, 1345–1373 (2012).

[10] K. Kunisch and D. Wachsmuth, *Sufficient optimality conditions and semi–smooth Newton methods for optimal control of stationary variational inequalities*. ESAIM Control Optim. Calc. Var. **18**, no. 2, 520—547 (2012).

[11] C. Meyer and O. Thoma, *A priori finite element error analysis for optimal control of the obstacle problem.* SIAM J. Numer. Anal. **51**, 605–628 (2013).

[12] C. Meyer, A. Rademacher, and W. Wollner, *Adaptive optimal control of the obstacle problem.* SIAM J. Sci. Comput. **37**, 918–945 (2015).

[13] F. Mignot, *Contrôle dans les inéquations variationelles elliptiques.* J. Funct. Anal. **22**, 130–185 (1976).

[14] F. Mignot and J.-P. Puel, *Optimal control in some variational inequalities.* SIAM J. Control Optim. **22** (3), 466–476 (1984).

[15] A. Schiela and D. Wachsmuth, *Convergence analysis of smoothing methods for optimal control of stationary variational inequalities with control constraints.* ESAIM Math. Model. Numer. Anal. **47**, no. 3, 771–787 (2013).

[16] G. Wachsmuth, *Strong stationarity for optimal control of the obstacle problem with control constraints.* SIAM J. Optim. **24** (1), 1914–1932 (2014).