French-German-Spanish Conference on Optimization
Gijón, June 18-21, 2024
(pp. 108-115)

A numerical solution approach for non-smooth optimal control problems based on the Pontryagin maximum principle

Daniel Wachsmuth
daniel.wachsmuth@uni-wuerzburg.de University of Würzburg

Abstract

We consider nonsmooth optimal control problems subject to a linear elliptic partial differential equation with homogeneous Dirichlet boundary conditions. It is well-known that local solutions satisfy the celebrated Pontryagin maximum principle. In this note, we will investigate an optimization method that is based on the maximum principle. We prove that the discrepancy in the maximum principle vanishes along the resulting sequence of iterates. Numerical experiments confirm the theoretical findings.

1. Introduction

In this note, we consider the following optimal control problem: Minimize

$$J(y,u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \int_\Omega g(u(x))\,\mathrm{d}x \tag{1.1}$$

over all $u \in L^2(\Omega)$ and $y \in H^1_0(\Omega)$ satisfying

$$\begin{aligned} -\Delta y &= u \quad \text{in } \Omega,\\ y &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

Here, $\Omega \subset \mathbb{R}^d$ is a bounded domain, and $g : \mathbb{R} \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is assumed to be proper and lower semicontinuous. In addition, we require

$$\lim_{|v|\to\infty} \frac{g(v)}{|v|} = +\infty. \tag{1.2}$$

Note, that we assume neither continuity nor convexity of g. Hence, it is impossible to prove existence of solutions of (1.1). In fact, one can construct problems without solution, see [18, Section 4.5]. In this note, we will work with the example

$$g(u) := \frac{\alpha}{2}u^2 + I_{\mathbb{Z}}(u) = \begin{cases} \frac{\alpha}{2}u^2 & \text{if } u \in \mathbb{Z} \\ +\infty & \text{otherwise,} \end{cases} \tag{1.3}$$

where $\alpha > 0$. If g is assumed to be convex and continuous, then existence of solutions of (1.1) can be proven by the direct method of the calculus of variations [17]. Let us remark that by the above assumptions g is bounded from below.

If solutions exist, then the Pontryagin maximum principle [11] is a necessary optimality condition. Its main feature is that no differentiability with respect to the controls is needed, and so it is perfectly suited for the problems considered here. In fact, due to the structure of the problem (linear state equation, convexity of J with respect to y), the maximum principle is sufficient. We refer to [2, 4, 5, 12] for the Pontryagin maximum principle applied to optimal control problems for partial differential equations. The goal of this note is to construct an algorithm to solve the maximum principle. We will comment on related work in Section 4.

2. Sensitivity analysis

In this section, we will perform a sensitivity analysis with respect to perturbations of the control with characteristic functions. The setup is as follows: Let $u, \tilde{u} \in L^2(\Omega)$ be feasible controls, i.e., the integrals $\int_\Omega g(u)\,\mathrm{d}x$ and $\int_\Omega g(\tilde{u})\,\mathrm{d}x$ exist. Let $B \subset \Omega$ be measurable. We define

$$u_B := u + \chi_B(\tilde{u} - u).$$

Let y, y_B be the uniquely determined weak solutions of

$$\begin{aligned} -\Delta y &= u & -\Delta y_B &= u_B && \text{in } \Omega,\\ y &= 0 & y_B &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Let $p \in H_0^1(\Omega)$ be the weak solution of the adjoint equation

$$\begin{aligned} -\Delta p &= y - y_d && \text{in } \Omega,\\ p &= 0 && \text{on } \partial\Omega. \end{aligned}$$

The goal is now to estimate $J(y_B, u_B) - J(y, u)$ in terms of $u, \tilde{u}, p$ and the Lebesgue measure $|B|$ of B. Here, we have the following result.

Lemma 2.1 *Under the assumptions above, we have*

$$J(y_B, u_B) - J(y, u) = \int_B (\tilde{u} - u)p + g(\tilde{u}) - g(u)\,\mathrm{d}x + \frac{1}{2}\|y_B - y\|^2_{L^2(\Omega)}.$$

Proof This follows directly from the definition of p and u_B:

$$\begin{aligned} J(y_B, u_B) - J(y, u) &= \frac{1}{2}\|y_B - y_d\|^2_{L^2(\Omega)} + \int_\Omega g(u_B)\,\mathrm{d}x - \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} - \int_\Omega g(u)\,\mathrm{d}x\\ &= \int_\Omega (y_B - y)(y - y_d) + \frac{1}{2}(y_B - y)^2\,\mathrm{d}x + \int_B g(\tilde{u}) - g(u)\,\mathrm{d}x\\ &= \int_B (\tilde{u} - u)p + g(\tilde{u}) - g(u)\,\mathrm{d}x + \frac{1}{2}\|y_B - y\|^2_{L^2(\Omega)}. \end{aligned}$$

□

We will now prove that $\|y_B - y\|^2_{L^2(\Omega)}$ is of higher order with respect to the Lebesgue measure $|B|$ of B.

Lemma 2.2 *There are constants $c > 0$ and $\nu > 1/2$ independent of $u, \tilde{u}, B$ such that*

$$\|y_B - y\|_{L^2(\Omega)} \le c\,|B|^\nu \cdot \|\tilde{u} - u\|_{L^\infty(\Omega)},$$

where $|B|$ denotes the Lebesgue measure of B. The constant ν can be chosen as

$$\nu = \begin{cases} 1 & \text{if } d \le 3,\\ 1 - \epsilon & \text{if } d = 4 \text{ for } \epsilon > 0,\\ \frac{1}{2} + \frac{2}{d} & \text{if } d > 4. \end{cases}$$

Proof We prove the claim by a well-known duality argument. Assume $d \le 3$. Let $w \in L^2(\Omega)$ be given. Let $z, q \in H_0^1(\Omega)$ be the weak solutions of

$$\begin{aligned} -\Delta z &= w & -\Delta q &= z && \text{in } \Omega,\\ z &= 0 & q &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Due to [15], there is $c > 0$ independent of w, z such that

$$\|z\|_{L^\infty(\Omega)} \le c\|w\|_{L^2(\Omega)}.$$

Testing the weak formulations with z and q yields

$$\|z\|^2_{L^2(\Omega)} = \int_\Omega wq\,\mathrm{d}x \le \|w\|_{L^1(\Omega)}\|q\|_{L^\infty(\Omega)} \le c\|w\|_{L^1(\Omega)}\|z\|_{L^2(\Omega)}.$$

This proves $\|z\|_{L^2(\Omega)} \le c\|w\|_{L^1(\Omega)}$. Applying this estimate to $z := y_B - y$ and $w := u_B - u$ yields the claim with

$$\|y_B - y\|_{L^2(\Omega)} \le c\|u_B - u\|_{L^1(\Omega)} \le c\,|B| \cdot \|\tilde{u} - u\|_{L^\infty(\Omega)}.$$

In case $d > 3$ one can use the estimates from [3, Theorem 18]. □

Combining these results proves the following theorem.

Theorem 2.3 *Let $u, \tilde{u} \in L^\infty(\Omega)$. Let $B \subset \Omega$ be measurable. Let $\tilde{u}, y_B, y, p$ be defined as above. Then there are $\gamma > 0$ and $c > 0$ independent of $u, \tilde{u}, B$ such that*

$$J(y_B, u_B) - J(y, u) \le \int_B (\tilde{u} - u)p + g(\tilde{u}) - g(u)\,\mathrm{d}x + c\,|B|^{1+\gamma}\|\tilde{u} - u\|^2_{L^\infty(\Omega)}.$$

3. Pontryagin maximum principle

With the help of Theorem 2.3 we can prove the Pontryagin maximum principle.

Theorem 3.1 *Let $\bar{u} \in L^\infty(\Omega)$ be locally optimal with respect to $L^1(\Omega)$ topology for the control problem* (1.1)*. Let $\bar{y}, \bar{p} \in H_0^1(\Omega)$ be the optimal state and adjoint solving*

$$\begin{aligned} -\Delta\bar{y} &= \bar{u} & -\Delta\bar{p} &= \bar{y} - y_d & &\text{in } \Omega, \\ \bar{y} &= 0 & \bar{p} &= 0 & &\text{on } \partial\Omega. \end{aligned}$$

Let $v \in \mathbb{R}$ be such that $g(v) < +\infty$. Then

$$\bar{u}(x)\bar{p}(x) + g(\bar{u}(x)) \le v\bar{p}(x) + g(v) \text{ for almost all } x \in \Omega. \tag{3.1}$$

Proof Let $v \in \mathbb{R}$ be such that $g(v) < +\infty$. Applying Theorem 2.3 with $u := \bar{u}, \tilde{u} := v$ yields

$$0 \le J(y_B, u_B) - J(\bar{y}, \bar{u}) = \int_B (v - \bar{u})\bar{p} + g(\tilde{u}) - g(\bar{u})\,\mathrm{d}x + o(|B|).$$

By standard arguments based on the Lebesgue differentiation theorem, see, e.g., [10, Theorem 2.1], the claim follows. □

The maximum principle is a sufficient condition for the problem considered here.

Corollary 3.2 *Let $\bar{u} \in L^2(\Omega)$ satisfy the conclusion* (3.1) *of Theorem 3.1. Then $\bar{u}$ is global optimal for* (1.1)*.*

Proof Let $\tilde{u} \in L^2(\Omega)$ be an admissible control with associated state $\tilde{y}$. Then Lemma 2.1 with $B = \Omega$ yields

$$J(\tilde{y}, \tilde{u}) - J(\bar{y}, \bar{u}) = \int_\Omega (\tilde{u} - \bar{u})\bar{p} + g(\tilde{u}) - g(\bar{u})\,\mathrm{d}x + \frac{1}{2}\|\tilde{y} - \bar{y}\|^2_{L^2(\Omega)}.$$

Since $\bar{u}$ satisfies (3.1), the first expression is non-negative, which implies $J(\tilde{y}, \tilde{u}) - J(\bar{y}, \bar{u}) \ge \frac{1}{2}\|\tilde{y} - \bar{y}\|^2_{L^2(\Omega)} \ge 0$. □

4. Construction of an algorithm

We will now apply Theorem 2.3 with $u := u_k$ and $\tilde{u} := \tilde{u}_k$, where u_k is the current iterate of the algorithm to be devised. Let y_k and p_k be the associated state and adjoint. The control $\tilde{u}_k$ has to be computed in each iteration. Let B_k be measurable. Then we have

$$J(y_{B_k}, u_{B_k}) - J(y_k, u_k) = \int_{B_k} (\tilde{u}_k - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x + o(|B_k|). \tag{4.1}$$

The idea is now to choose $\tilde{u}_k$ and B_k such that $J(y_{B_k}, u_{B_k}) - J(y_k, u_k)$ is negative and to define the new iterate by

$$u_{k+1} = u_k + \chi_{B_k}(\tilde{u}_k - u_k).$$

In view of the maximum principle, Theorem 3.1, it is natural to choose $\tilde{u}_k$ as a function satisfying

$$\tilde{u}_k(x) \in \arg\min_{v\in\mathbb{R}} vp_k + g(v). \tag{4.2}$$

In addition, B_k will be chosen to get sufficient descent.

Let us comment on related work. The classic algorithm of [8] chooses $B_k := \Omega$, resulting in a fixed-point scheme to solve the maximum principle. The min-h method of [7] uses the update $u_{k+1} := u_k + t(\tilde{u}_k - u_k)$ with $t \in (0, 1]$, and is thus only suited for convex functions g. In the monograph [14], a method similar to ours is presented to solve optimal control problems with ODEs. Let us also also mention the review papers [6, 16]. In [9] binary control problems are solved with a similar approach: there a trust-region globalization is proposed, whereas we use an Armijo line-search to globalize.

As motivated above, we will compute $\tilde{u}_k$ as a result of the pointwise minimization

$$\tilde{u}_k(x) \in \arg\min_{v\in\mathbb{R}} vp_k + g(v).$$

Due to (1.2) this problem is solvable for all x. A measurable selection of this argmin-map exists [1]. For the example of g proposed in (1.3), we get

$$\tilde{u}_k(x) \in \operatorname{round}\left(-\frac{1}{\alpha}p_k(x)\right).$$

It remains to describe how B_k is chosen. Here, we are faced with two competing goals: In order to make the first term in (4.1) as small as possible, B_k has to be chosen as large as possible. However, to control the remainder term in (4.1), $|B_k|$ has to be chosen sufficiently small.

We propose the following line-search. Given $t \in (0, 1]$, choose B_t such that

$$\begin{gathered}\int_{B_t} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x \le t \int_{\Omega} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x,\\ |B_t| \le t \cdot |\Omega|.\end{gathered} \tag{4.3}$$

Due to the celebrated Lyapunov convexity theorem, see, e.g., [13, Theorem 5.5], a measurable set B_t satisfying (4.3) exists. Given t and B_t, we set $u_t := u_k + \chi_{B_t}(\tilde{u}_k - u_k)$. Let y_t be the associated state.

The parameter t_k is determined by the following procedure: Let t_k be the largest number in $\{\beta^l : l \in \mathbb{N} \cup \{0\}\}$, where $\beta \in (0, 1)$, that satisfies the descent condition

$$J(y_t, u_t) - J(y_k, u_k) \le \sigma \int_{B_t} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x \tag{4.4}$$

where $\sigma \in (0, 1)$, and B_t is a measurable set satisfying (4.3). This condition is inspired by the well-known Armijo line-search in nonlinear optimization. If u_k does not satisfy the maximum principle, there is an admissible step-size t_k, and the resulting algorithm produces a new iterate with smaller value of the objective.

Lemma 4.1 *Suppose that*

$$\int_{\Omega} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x < 0.$$

There is $t_0 > 0$ such that for all $t \in (0, t_0)$ condition (4.4) *is satisfied.*

Proof Due to Theorem 2.3, we have

$$\begin{aligned}J(y_t, u_t) - J(y_k, u_k) &- \sigma \int_{B_t} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x\\ &\le (1 - \sigma) \int_{B_t} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x + o(t)\\ &\le t(1 - \sigma) \int_{\Omega} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x + o(t),\end{aligned}$$

which proves the claim. □

The resulting algorithm is sketched in Algorithm 1.

Let us now turn to the convergence analysis of Algorithm 1. Here, we follow the related analysis in [19]. Let us define

$$\rho_k := \int_{\Omega} (\tilde{u}_k - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x.$$

Due to the choice of $\tilde{u}_k$ in (4.2), it follows $\rho_k \le 0$. If $\rho_k = 0$ then u_k satisfies the maximum principle Theorem 3.1, and the corresponding control u_k is optimal by Corollary 3.2.

Lemma 4.2 *Let (u_k) be an infinite sequence generated by Algorithm 1. Then*

$$\sum_{k=0}^{\infty} t_k \|\rho_k\|_{L^1(\Omega)} < +\infty.$$

Algorithm 1 Maximum-principle based descent algorithm

Choose $\beta \in (0,1)$, $\sigma \in (0,1)$, u_0 with $\int_\Omega g(u_0)\,\mathrm{d}x < \infty$, $\delta_{\text{tol}} \geq 0$. Set $k := 0$.
loop ▷ Gradient descent
  Compute state y_k and adjoint p_k associated to u_k.
  Compute $\tilde{u}_k$ as in (4.2).
  if $\left|\int_\Omega (\tilde{u}_k - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x\right| \leq \delta_{\text{tol}}$ **then** ▷ Termination criterion
    return u_k
  end if
  $t := 1$.
  loop ▷ Armijo line-search
    Compute $B_{k,t}$ satisfying (4.3).
    Compute $J(y_t, u_t)$.
    if (4.4) is satisfied **then**
      break
    end if
    $t := \beta \cdot t$.
  end loop
  $t_k := t$. ▷ Update
  $u_{k+1} := u_k + \chi_{B_{k,t_k}}(\tilde{u}_k - u_k)$.
  $k := k + 1$.
end loop

Proof Using conditions (4.4) and (4.3) shows

$$J(y_{k+1}, u_{k+1}) - J(y_k, u_k) \leq \sigma \int_{B_{t_k}} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x \leq t_k \int_\Omega (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x = -t_k \|\rho_k\|_{L^1(\Omega)}.$$

Due to (1.2), g has a global minimum and is bounded from below, so that J is bounded from below by some $M \in \mathbb{R}$. Summing this inequality over $k \in \mathbb{N}$ and using $J \geq M$ proves $\sum_{k=1}^{\infty} t_k \|\rho_k\|_{L^1(\Omega)} \leq J(y_0, u_0) - M < \infty$. □

For simplicity, we assume for the subsequent convergence analysis that

$$\operatorname{dom} g := \{v : \; g(v) < \infty\} \tag{4.5}$$

is compact. Then the set of iterates (u_k) and $(\tilde{u}_k)$ is uniformly bounded in $L^\infty(\Omega)$.

Corollary 4.3 *Assume* (4.5)*. Let* $M > 0$ *such that* $\operatorname{dom} g \subset [-M, +M]$*. Then* $\|u_k\|_{L^\infty(\Omega)} \leq M$ *and* $\|\tilde{u}_k\|_{L^\infty(\Omega)} \leq M$ *for all k.*

Theorem 4.4 *Assume* (4.5)*. Either the Algorithm 1 stops after finitely many steps with*

$$\left|\int_\Omega (\tilde{u}_k - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x\right| \leq \delta_{\text{tol}}$$

(so that u_k satisfies the maximum principle if $\delta_{\text{tol}} = 0$), or

$$\int_\Omega (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x \to 0,$$

i.e., the residual in the maximum principle tends to zero, and (u_k) is a minimizing sequence.

Proof We follow the proof of the related result [19, Theorem 6.7]. Let us assume the algorithm generates an infinite sequence of iterates. Let k be such that $t_k < 1$. Due to the line-search procedure of Algorithm 1, it follows that $t := \beta^{-1} t_k \leq 1$ violates the descent condition (4.4), that is

$$0 < J(y_t, u_t) - J(y_k, u_k) - \sigma \int_{B_t} (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x.$$

As in the proof of Lemma 4.1, we get from Theorem 2.3

$$0 < t(1-\sigma)\int_\Omega (\tilde{u} - u_k)p_k + g(\tilde{u}) - g(u_k)\,\mathrm{d}x + c\,|t|^{1+\gamma}\|\tilde{u} - u\|_{L^\infty(\Omega)}.$$

Together with Corollary 4.3, we get

$$0 < -t(1-\sigma)\|\rho_k\|_{L^1(\Omega)} + c|t|^{1+\gamma},$$

where c is independent of k. This implies

$$\|\rho_k\|_{L^1(\Omega)} \le c t_k^\gamma$$

for all k such that $t_k < 1$. With Lemma 4.2, we get

$$+\infty > \sum_{k=0}^\infty t_k\|\rho_k\|_{L^1(\Omega)} = \left(\sum_{k:\,t_k=1} \|\rho_k\|_{L^1(\Omega)}\right) + \left(\sum_{k:\,t_k<1} t_k\|\rho_k\|_{L^1(\Omega)}\right) \ge \left(\sum_{k:\,t_k=1} \|\rho_k\|_{L^1(\Omega)}\right) + c\left(\sum_{k:\,t_k<1} \|\rho_k\|_{L^1(\Omega)}^{1+\frac{1}{\gamma}}\right),$$

which results in $\lim_{k\to\infty}\|\rho_k\|_{L^1(\Omega)} = 0$. Hence, the algorithm stops after finitely many iterations if $\delta_{\mathrm{tol}} > 0$. □

5. Numerical results

Let us now report about numerical experiments with Algorithm 1. Here, we consider the optimal control problem

$$J(y,u) := \frac{1}{2}\|y - y_d\|_{L^2(\Omega)}^2 + \frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2 + I_{\mathbb{Z}\cap[-b,b]}(u)$$

over all $u \in L^2(\Omega)$ and $y \in H_0^1(\Omega)$ satisfying

$$\begin{aligned} -\Delta y &= u \quad \text{in } \Omega,\\ y &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

This fits into the setting of the paper with the choice

$$g(v) := \frac{\alpha}{2}v^2 + I_{\mathbb{Z}\cap[-b,b]}$$

Here, we chose $\Omega = (0,1)^2$,

$$y_d(x_1,x_2) = 10x_1\sin(5x_1)\cos(7x_2), \quad \alpha = 0.01, \quad \beta = 0.01, \quad b = 10.$$

We discretized the problem with piecewise linear finite elements on a regular mesh for state and adjoint variables, while the control was discretized with piecewise constant finite elements. We report the results for a sequence of different meshes, where the finest mesh has mesh-size $h = 1.41\cdot10^{-3}$ resulting in $\approx 2\cdot10^6$ degrees of freedom for the control variables, which results in a mixed-integer optimization problem with $\approx 2\cdot10^6$ integer variables. In the implementation of Algorithm 1 a greedy strategy was used to determine B_t. The loop in Algorithm 1 was terminated if in the inner loop $t|\Omega|$ was smaller than any of the elements in the grid.

Now let us report about some of the results. The optimal control can be seen in the left plot of Figure 1. In the right plot, we report about the iteration history of the residual $\|\rho_k\|_{L^1(\Omega)}$. Surprisingly, the iterations seem to be mesh independent. In addition, for this particular problem a very small number of iterations was needed to optimize over $2\cdot10^6$ discrete control variables.

This is underlined by the results in Table 1. It shows for different discretizations the final value of the objective J and the final value of the residual $\|\rho\|_{L^1(\Omega)}$. As can be seen from the last column of this table, very few outer iterations are needed. In conclusion, this new algorithm seems to be capable of solving quite challenging mixed-inter programs.

Acknowledgements

This research was supported by the German Research Foundation (DFG) under grant number WA 3626/3-2 within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962). The author thanks Anna Lentz for comments on an earlier version of this manuscript.

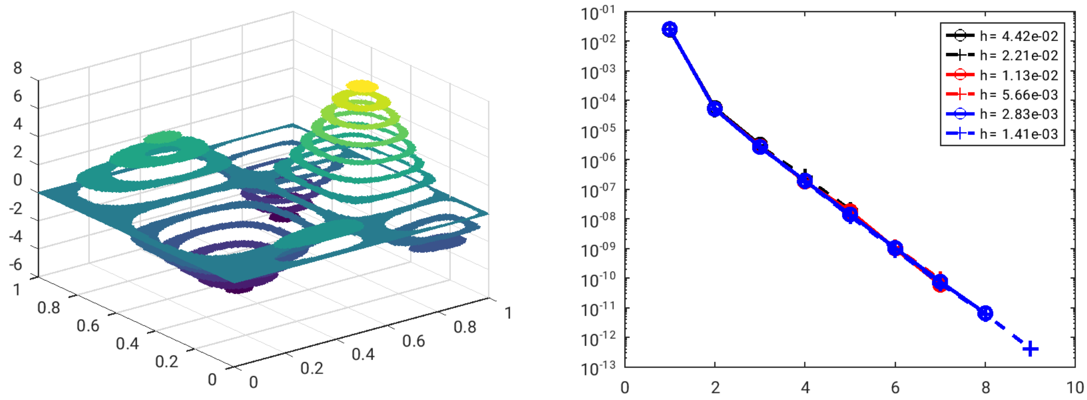

Fig. 1 Optimal control (left), iteration history (right)

h	J	$\|\rho\|_{L^1(\Omega)}$	It
$4.42 \cdot 10^{-2}$	4.706	$3.20 \cdot 10^{-6}$	4
$2.21 \cdot 10^{-2}$	5.048	$2.02 \cdot 10^{-8}$	6
$1.13 \cdot 10^{-2}$	5.210	$6.00 \cdot 10^{-11}$	8
$5.66 \cdot 10^{-3}$	5.293	$8.91 \cdot 10^{-11}$	8
$2.83 \cdot 10^{-3}$	5.334	$6.46 \cdot 10^{-12}$	9
$1.41 \cdot 10^{-3}$	5.354	$4.11 \cdot 10^{-13}$	10

Tab. 1 Iteration history

References

[1] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*, volume 2 of *Systems & Control: Foundations & Applications*. Birkhäuser Boston, Inc., Boston, MA, 1990. doi:10.1007/978-0-8176-4848-0.

[2] F. Bonnans and E. Casas. An extension of Pontryagin's principle for state-constrained optimal control of semilinear elliptic equations and variational inequalities. *SIAM J. Control Optim.*, 33(1):274–298, 1995. doi:10.1137/S0363012992237777.

[3] H. Brézis and W. A. Strauss. Semi-linear second-order elliptic equations in L^1. *J. Math. Soc. Japan*, 25:565–590, 1973. doi:10.2969/jmsj/02540565.

[4] E. Casas, J.-P. Raymond, and H. Zidani. Pontryagin's principle for local solutions of control problems with mixed control-state constraints. *SIAM J. Control Optim.*, 39(4):1182–1203, 2000. doi:10.1137/S0363012998345627.

[5] E. Casas. Pontryagin's principle for optimal control problems governed by semilinear elliptic equations. In *Control and estimation of distributed parameter systems: nonlinear phenomena (Vorau, 1993)*, volume 118 of *Internat. Ser. Numer. Math.*, pages 97–114. Birkhäuser, Basel, 1994. doi:10.1007/978-3-0348-8530-0_6.

[6] F. L. Chernous'ko and A. A. Lyubushin. Method of successive approximations for solution of optimal control problems. *Optimal Control Appl. Methods*, 3(2):101–114, 1982. doi:10.1002/oca.4660030201.

[7] R. G. Gottlieb. Rapid convergence to optimum solutions using a min-h strategy. *AIAA Journal*, 5(2):322–329, 1967. doi:10.2514/3.3960.

[8] I. A. Krylov and F. L. Černous'ko. The method of successive approximations for solving optimal control problems. *Ž. Vyčisl. Mat i Mat. Fiz.*, 2:1132–1139, 1962. (In Russian).

[9] P. Manns, M. Hahn, C. Kirches, S. Leyffer, and S. Sager. On convergence of binary trust-region steepest descent. *J. Nonsmooth Anal. Optim.*, 4, 2023. doi:10.46298/jnsao-2023-10164.

[10] C. Natemeyer and D. Wachsmuth. A proximal gradient method for control problems with non-smooth and non-convex control cost. *Comput. Optim. Appl.*, 80(2):639–677, 2021. doi:10.1007/s10589-021-00308-0.

[11] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The mathematical theory of optimal processes*. Interscience Publishers John Wiley & Sons, Inc., New York-London, 1962. Translated from the Russian by K. N. Trirogoff; edited by L. W. Neustadt.

[12] J. P. Raymond and H. Zidani. Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. *SIAM J. Control Optim.*, 36(6):1853–1879, 1998. doi:10.1137/S0363012996302470.

[13] W. Rudin. *Functional analysis*. McGraw-Hill Series in Higher Mathematics. McGraw-Hill Book Co., New York-Düsseldorf-Johannesburg, 1973.

[14] V. A. Srochko. *Iterative methods for solving optimal control problems*. Fismatlit, Moscow, 2000. (In Russian).

[15] G. Stampacchia. Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Ann. Inst. Fourier (Grenoble)*, 15(fasc. 1):189–258, 1965. URL: http://www.numdam.org/item?id=AIF_1965__15_1_189_0.

[16] A. S. Strekalovsky. Modern methods for solving nonconvex optimal control problems. *The Bulletin of Irkutsk State University. Series Mathematics*, 8:141–163, 2014. (In Russian).

[17] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels. doi:10.1090/gsm/112.

[18] D. Wachsmuth. Iterative hard-thresholding applied to optimal control problems with $L^0(\Omega)$ control cost. *SIAM J. Control Optim.*, 57(2):854–879, 2019. doi:10.1137/18M1194602.

[19] D. Wachsmuth. A topological derivative-based algorithm to solve optimal control problems with $L^0(\Omega)$ control cost. *J. Nonsmooth Anal. Optim.*, 5, 2024. doi:10.46298/jnsao-2024-12366.