Numerische Mathematik 141.2 (2019): 455-493

DOI: https://doi.org/10.1007/s00211-018-1005-z

Greedy optimal control for elliptic problems and its application to turnpike problems

Víctor Hernández-Santamaría · **Martin Lazar** · **Enrique Zuazua**

Received: date / Accepted: date

Abstract In this paper, we deal with the approximation of optimal controls for parameter-dependent elliptic and parabolic equations. We adapt well-known results on greedy algorithms to approximate in an efficient way the optimal controls for parameterized elliptic control problems. Our results yield an optimal approximation procedure that, in particular, performs better than simply sampling the parameter-space to compute controls for each parameter value. The same method can be adapted for parabolic control problems, but this leads to greedy selections of the realizations of the parameters that depend on the initial datum under consideration. To avoid this difficulty we employ the turnpike property for time evolution control problems that ensures the asymptotic simplification of optimal control problems for evolution equations towards the elliptic steady-state ones in long time horizons $[0, T]$. The combination of the turnpike property and greedy methods allows us to develop efficient methods for the approximation of the parameter-dependent parabolic optimals too. We present various numerical experiments discussing the efficiency of our methodology and its application to turnpike control problems.

Keywords parameterized PDEs · optimal control · turnpike property · greedy algorithms · elliptic equations · parabolic equations

Mathematics Subject Classification (2000) 49J20 · 49K20 · 93C20 · 49N05 · 65K10

1 Introduction

Optimal control problems play a major role in many fields of science and engineering. These problems are commonly subject to partial differential equations (PDE) depending on several parameters. While the PDE describes the underlying system, the parameters are used to identify or specify particular configurations such as material properties, the position of sensors and actuators, initial conditions, among others. In applications, it is of interest to explore within different parameter configurations. However, from the control

The work of the first author was partially supported by the Advanced Grant DYCON (Dynamic Control) of the European Research Council Executive Agency. The second author was also partly supported by Croatian Science Foundation under ConDyS Project, IP-2016-06-2468. The work of the third author was partially supported by the Advanced Grant DYCON (Dynamic Control) of the European Research Council Executive Agency, FA9550-15-1-0027 of AFOSR, FA9550-14-1-0214 of the EOARD-AFOSR, the MTM2014-52347 Grant of the MINECO (Spain) and ICON of the French ANR.

V. Hernández-Santamaría
DeustoTech, University of Deusto, 48007 Bilbao, Basque Country, Spain.
Facultad Ingeniería, Universidad de Deusto, Avda. Universidades, 24, 48007, Bilbao, Basque Country, Spain.
E-mail: victor.santamaria@deusto.es

M. Lazar
Department of Electrical Engineering and Computing, University of Dubrovnik, Dubrovnik, Croatia.
E-mail: mlazar@unidu.hr

E. Zuazua
DeustoTech, University of Deusto, 48007 Bilbao, Basque Country, Spain.
Departamento de Matemáticas, Universidad Autónoma de Madrid, 28049 Madrid, Spain.
Facultad Ingeniería, Universidad de Deusto, Avda. Universidades, 24, 48007, Bilbao, Basque Country, Spain.
Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France.
E-mail: enrique.zuazua@deusto.es

point of view, this leads to solving a different problem for each new desired configuration which at a computational level may be rather expensive or prohibitive.

To overcome this expensive task, different techniques have been developed in the past years in order to speed up the solution of parameterized optimal control problems. First examples of PDE constrained optimization problems solved by computational reduction techniques have been addressed in [22] using reduced basis (RB) methods or [33] by means of proper orthogonal decomposition (POD).

In the context of parameterized control problems, POD techniques have been successfully applied in [2, 26, 27, 36]. In particular, in [36], the authors estimate the distance between the optimal control and an approximation using perturbation arguments, obtaining an efficient algorithm to solve the optimal control problem. However, the evaluation of the a posteriori error bounds requires a forward-backward solution of the associated state and andjoint equations and, as pointed in [36], this is computationally expensive.

In the same way, RB techniques have been used in the context of parameter-dependent control problems, see, for instance, [3, 12, 23, 20, 25, 30]. In such works, reliable error estimates of the reduced order optimal controls are presented.

More recently, in [28], the authors develop greedy methods in the context of controllability problems. Their analysis is based on the ideas of reduced modelling and (weak) greedy algorithms for parameter dependent PDE and abstract equations in Banach spaces (see, e.g., [7, 10]). The method is shown to be computationally efficient, but it has the drawback that the greedy choice of the parameters depends on the data to be controlled. The problem of developing greedy methods for controllability problems in a way that the greedy choice of the parameters is independent of the data to be controlled is an open problem.

Here, we address this issue through the *turnpike* theory (see, for instance, [31]) which roughly states that optimal control and controllability problems, under suitable conditions, have the property that in long time-horizons optimal, the optimal trajectories are exponentially close to the steady state ones during most of the time duration.

Thanks to the turnpike property, the long-time horizon problem associated to a parabolic equation can be reduced to study the steady elliptic one. However, in spite of this asymptotic simplification, when the system depends on one or more parameters, exploring many different parameter configurations becomes prohibitive and this justifies the relevance of applying efficient approximation algorithms to the optimal control of elliptic control problems.

The main contributions of this paper can be summarized as follows. On one hand, greedy and weak greedy algorithms have been developed recently in a systematic manner to be applied to yield optimal approximations (in the sense of the Kolmogorov width) for PDEs which solutions depend on the parameters in an analytic manner. Although the optimal control problem for parameterized elliptic equations is well known and has been studied in numerous papers (see, for instance, [3, 20, 24, 30]), as far as the authors knowledge, this greedy and weak greedy algorithms have not been developed in the context of optimal control problems.

On the other hand, it is well known that solutions of optimal control problems for elliptic PDEs can be characterized as the solutions of the corresponding optimality system, which is constituted by the coupling of two elliptic equations, the one fulfilled by the state and the one corresponding to the adjoint state. Accordingly, the optimal control problem can be identified with an elliptic system to which the existing methods are applied. The second contribution consists in developing cheap surrogates allowing to diminish the computational cost that the greedy algorithms would require if applied directly using the residual as error indicator. This allows to circumvent the forward-backward expensive solution, as for instance, discussed in [36].

The paper is organized as follows. In section 2, we introduce the optimal control problem for a parabolic equation and then, by means of the turnpike property, we reduce this problem to study a steady state system. This will be the starting point to formulate the greedy approach for elliptic optimal control systems. In Section 3, we present the main results of this paper: firstly, we present a brief summary on greedy algorithms and then we analyze their application in the optimal control of parameterized elliptic equations. Section 4 contains several numerical examples and experiments for the greedy approach of finite-difference discretizations to 2-D elliptic control problems. We devote Section 5 to present the applicability of the (greedy) steady control in time-dependent problems, while in Section 6 we make a detailed analysis of the computational cost of our greedy algorithm. Finally, in Section 7 we make some concluding remarks.

2 Problem formulation

Let $\Omega \subset \mathbb{R}^d$ be an open and bounded Lipschitz Domain and consider the parameter dependent parabolic equation with Dirichlet boundary conditions

$$\begin{cases} y_t - \operatorname{div}(a(x,\nu)\nabla y) + c(x,\nu)y = \chi_\omega u & \text{in } Q = \Omega \times (0,T), \\ y = 0 & \text{on } \Sigma = \partial\Omega \times (0,T), \\ y(x,0) = y^0(x) & \text{in } \Omega, \end{cases} \tag{1}$$

where $y = y(x,t;\nu)$ is the state, $u = u(x,t)$ is a control function and y_0 is given initial datum. Here ω is an open subset of Ω and χ_ω denotes the characteristic function of the set ω where the control is being applied.

In (1), $a(x,\nu)$ is an elliptic, scalar $L^\infty(\Omega)$ coefficient depending on some parameter $\nu \in \mathbb{R}^d$ and $c = c(x,\nu) \in L^\infty(\Omega)$ is parameter dependent potential. To abridge the notation, we will denote $a_\nu = a(x,\nu)$ and $c_\nu = c(x,\nu)$

It is well-known (see, for instance, [15]), that for any initial data $y_0 \in L^2(\Omega)$ and $u \in L^2(\omega \times (0,T))$, systems (1) admits a unique weak solution $y \in W(0,T)$, where $W(0,T)$ stands for the Hilbert space

$$W(0,T) := \left\{ y \in L^2(0,T;H_0^1(\Omega)),\ y_t \in L^2(0,T;H^{-1}(\Omega)) \right\},$$

equipped with the norm

$$\|y\|_{W(0,T)} = \left(\|y\|^2_{L^2(0,T;H_0^1(\Omega))} + \|y_t\|^2_{L^2(0,T;H^{-1}(\Omega))} \right)^{1/2}.$$

Moreover, y satisfies an estimate of the form

$$\|y\|^2_{W(0,T)} \le C \left(\|y_0\|_{L^2(\Omega)} + \|u\|_{L^2(\omega\times(0,T))} \right),$$

where C depends on Ω, T, a_ν and $\|c_\nu\|_\infty$ only.

Now, consider the following associated optimal control problem

$$\min_{u\in L^2(0,T;L^2(\omega))} J^T(u) = \frac{1}{2}\int_0^T |u(t)|^2_{L^2(\omega)} + \frac{\beta}{2}\int_0^T \|y(t) - y_d\|^2_{L^2(\Omega)}, \tag{2}$$

where y is the solution to (1), $y_d \in L^2(\Omega)$ is a desired observation and $\beta > 0$ is given.

It is classical to prove (see e.g., [29]) that the minimization problem (2) has a unique optimal solution that hereinafter we denote by (u^T, y^T). This is due to the fact that the functional J^T is strictly convex, continuous and coercive.

Moreover, the optimal control is given by

$$u^T = -\chi_\omega p^T$$

where q^T can be found from the solution (y^T, p^T) of the optimality system

$$\begin{cases} y_t^T - \operatorname{div}(a_\nu \nabla y^T) + c_\nu y^T = -\chi_\omega p^T, & \text{in } Q, \\ -p_t^T - \operatorname{div}(a_\nu \nabla p^T) + c_\nu p^T = \beta\,(y^T - y_d), & \text{in } Q, \\ y^T = q^T = 0, & \text{on } \Sigma, \\ y^T(x,0) = y^0, \quad q^T(x,T) = 0, & \text{in } \Omega. \end{cases} \tag{3}$$

From here, it is clear that the optimal control u^T also depends on the parameter ν since the state y equally depends on it. As mentioned before, when studying parameter-dependent problems from the control point of view, this means that one has to solve the minimization problem (2) for each new choice of the parameter ν. Although theoretically feasible, the computational effort to compute a control function for every new selection of the parameter is rather expensive and undesirable.

Moreover, the effective computation of optimal controls for problems posed in a long time horizon can be very expensive since it requires iterative methods to solve the coupled optimality system (3) combining the forward controlled state equation and the backward adjoint one. Particularly, adjoint equation methods for the solution of control problems posed on long-time intervals may led to the storage of huge quantities of information. Memory saving methods, as for instance discussed in [21] and references within, may help to alleviate this problem.

In this work, we use a combination of (weak) greedy methods and the so-called *turnpike* property to determine the most relevant values of a parameter-space (to be precised below) and provide the best possible approximation of the set of parameter dependent optimal controls.

Firstly, we will use the *turnpike* property to reduce the problem of computing the time-dependent optimal controls u^T by computing asymptotic simplifications.

To this end, we begin by considering the stationary version of the state equation

$$\begin{cases} -\mathrm{div}\,(a_\nu \nabla y) + c_\nu y = \chi_\omega u & \text{in } \Omega, \\ y = 0 & \text{on } \partial\Omega. \end{cases} \tag{4}$$

Using classical Fredholm theory (see, e.g. [6]), one can prove that for fixed $a_\nu, c_\nu \in L^\infty(\Omega)$, system (4) has a unique weak solution $y \in H_0^1(\Omega)$ verifying

$$\alpha_1 \|y\|_{H_0^1(\Omega)} \leq \|u\|_{H^{-1}(\Omega)} \leq \alpha_2 \|y\|_{H_0^1(\Omega)}. \tag{5}$$

for some constants α_1 and α_2 (not depending on u), provided the following condition holds

$$\ker(L_\nu) = \{0\}, \tag{6}$$

where L_ν is the map defined as

$$L_\nu := -\mathrm{div}\,(a_\nu \,\cdot\,) + cI : H_0^1(\Omega) \to H^{-1}(\Omega).$$

Hereinafter, we will briefly discuss the *turnpike* theory from the point of view where condition (6) holds. Observe that when this condition is not fulfilled, the solution to (4) is not longer unique and some modifications to the discussion presented below are necessary (see Section 7 and [31, Remark 3.7]), but the results remain valid.

Let us consider the corresponding minimization problem

$$\min_{u \in L^2(\omega)} J(u) = \frac{1}{2}|u|^2_{L^2(\omega)} + \frac{\beta}{2}\|y - y_d\|^2_{L^2(\Omega)}. \tag{7}$$

As before, one can prove that, since J is strictly convex, continuous and coercive, the minimization problem (7) has a unique optimal solution $(\bar{u}, \bar{y})$, where the optimal control is characterized by

$$\bar{u} = -\chi_\omega \bar{p},$$

and $(\bar{y}, \bar{p})$ solve the optimality system

$$\begin{cases} -\mathrm{div}\,(a_\nu \nabla \bar{y}) + c_\nu \bar{y} = -\chi_\omega \bar{p}, & \text{in } \Omega, \\ -\mathrm{div}\,(a_\nu \nabla \bar{p}) + c_\nu \bar{p} = \beta\,(\bar{y} - y_d), & \text{in } \Omega, \\ \bar{y} = \bar{p} = 0, & \text{on } \partial\Omega. \end{cases} \tag{8}$$

Also, in this case, it is clear the the optimal control $\bar{u}$ depends on ν. Notice that, thanks to condition (6), each control control $\bar{u}$ can be uniquely determined. This is not the case where $\ker(L_\nu)$ is not trivial, since the solution $\bar{p}$ to the adjoint equation is defined up to elements of $\ker(L_\nu)$.

Since each control $\bar{u}$ can be uniquely determined, in what follows, we shall write $\bar{u}_\nu$ to denote the dependence of the optimal control with respect to the parameter, and similarly for the optimal states y_ν and p_ν.

A natural question that arises in this context is to which extent the long horizon optimal controls and states $(u^T(t), y^T(t))$ approach the steady ones $(\bar{u}, \bar{y})$ as $T \to +\infty$. According to [31], we have the following result:

Theorem 1 *Let us consider the control problem* (2) *and let* (u^T, y^T) *be the optimal control and state. Then, there exists* $\mu > 0$ *such that*

$$\|y^T(t) - \bar{y}\|_{L^2(\Omega)} + \|u^T(t) - \bar{u}\|_{L^2(\Omega)} \leq K\left(e^{-\mu t} + e^{-\mu(T-t)}\right), \quad \forall t \in [0, T], \tag{9}$$

where $(\bar{u}, \bar{y})$ *is the optimal control and state corresponding to* (8).

This means that we have exponential convergence of the finite horizon control problems to their steady state version as T tends to infinity. In order to prove (9), additional assumptions on stabilizability and observability for (1) and its adjoint are required. Indeed, although these assumptions are not required when considering solely the optimal control problem (2), they are essential for the analysis when the time horizon tends to infinity.

Remark 1 It is important to mention that for Theorem 1 to hold, only the controllability of system (1) plays a fundamental role. In other words, positivity or smallness conditions on $c(x)$ guaranteeing the stability of the parabolic problem are not necessary to prove the exponential convergence of the finite time horizon control problem to the steady version.

As pointed out in [31], this kind of stationary behavior in the transient time for long horizon control problems is commonly referred in the literature as turnpike property. Such property has been mostly investigated in the finite-dimensional case, as well as in connection to Calculus of Variations. We refer the interested reader to the survey [37] and references therein and to [8,38] for some results on control problems in the infinite-dimensional case. We also refer to [35] for a more recent and systemic discussion on this topic.

Thanks to Theorem 1, we can now focus our attention on the steady system (4) and the approximation of the family of parameterized controls. We begin by assuming the following hypotheses:

H1 The parameter ν ranges within a compact set $\mathcal{K} \subseteq \mathbb{R}^d$.
H2 The coefficient functions a_ν and c_ν depend on ν in an analytic manner.
H3 Condition (5) holds uniformly for all $\nu \in \mathcal{K}$.

Then, the main idea is to propose a methodology to determine an optimal selection of a finite number of realizations of the parameter ν so that all controls, for all possible values of ν, are optimally approximated. More precisely, the problem can be formulated as follows:

Problem 1 Let us consider the set of controls verifying (7) for each possible value $\nu \in \mathcal{K}$. That is,

$$\overline{\mathcal{U}} = \{\bar{u}_\nu : \nu \in \mathcal{K}\}. \tag{10}$$

This control set is compact in $L^2(\omega)$.

Given $\varepsilon > 0$, we seek to determine a family of parameters $\{\nu_1, \ldots, \nu_n\}$ in $\mathcal{K}$, whose cardinality n depends on ε, so that the corresponding controls, denoted by $u_{\nu_1}, \ldots, u_{\nu_n}$ are such that for every $\nu \in \mathcal{K}$, there exists $u^\star_\nu \in \operatorname{span}\{u_{\nu_1}, \ldots, u_{\nu_n}\}$ such that

$$\|u^\star_\nu - \bar{u}_\nu\| \leq \varepsilon.$$

This problem is motivated by the practical issue of avoiding the computation of a control function u_ν for each new parameter value ν. By developing suitable greedy algorithms to solve Problem 1, we look for the most representative values of ν providing a fast and reliable way to compute approximated optimal controls for any other value of ν.

As noted in [28], Hypotheses **H1** and **H2** makes this goal feasible and, in particular, allows to implement a *naive* approach where the parameter set $\mathcal{K}$ is uniformly sampled in a very fine mesh. Nevertheless, the objective is to minimize the number of spanning controls n and derive the most efficient approximation. As we will see below, **H3** enable us to implement a (weak) greedy approach for solving Problem 1 in an optimal way.

Remark 2 We can readily distinguish two classes of coefficients verifying Hypothesis **H3**. These will be of interest in the discussion of Sections 4 and 5.

1. The case where $a_1 \leq a_\nu \leq a_2$ and $c_\nu \geq 0$, for some positive constants a_1, a_2 and all $\nu \in \mathcal{K}$. Under these assumptions we can recover (5) by Lax-Milgram theorem and classical energy estimates.
2. The case where $a_\nu \equiv \mathit{const}$ and $c_\nu \in [-\lambda_{i+1} + \epsilon, -\lambda_i - \epsilon]$, where λ_i, λ_{i+1} are two adjacent eigenvalues of the constant Laplacian. In this case, (6) holds and since we are away from the eigenvalues, L_ν is invertible with bounded inverse.

3 Greedy optimal control for elliptic problems

3.1 Preliminaries on (weak) greedy algorithms

In this section, we present a short introduction about linear approximation theory of parametric problems based on (weak) greedy algorithms. For a more detailed read, we refer, for instance, to [11,13]. We will apply this theory systematically to deal with the optimal control problem of parameterized elliptic equations.

In general, the goal is to approximate a compact set $\mathcal{K}$ in a Banach space X by a sequence of finite dimensional subspaces V_n of dimension n. Increasing n improves the accuracy of the approximation.

Although constructing this subspace within a given precision error commonly implies a high computational effort, this process is performed only once as an *offline* procedure. Then, with a good approximated subspace at hand, one can easily compute *online* approximations for every vector from $\mathcal{K}$.

Vectors x_i, $i = 1, ..., n$ spanning the space V_n are called *snapshots* of $\mathcal{K}$.

The goal of (weak) greedy algorithms is to construct a family of finite dimensional spaces $V_n \leq X$ approximating the set $\mathcal{K}$ in the best possible way. The algorithm reads as follows.

Algorithm 1: Weak greedy algorithm

initialize: fix $\gamma \in (0, 1]$ and a tolerance parameter $\varepsilon > 0$;

In the first step, choose $x_1 \in \mathcal{K}$ such that

$$\|x_1\|_X \geq \gamma \max_{x \in \mathcal{K}} \|x\|_X ;$$

At the general step, having found $x_1, \dots, x_n$ denote

$$V_n = \operatorname{span}\{x_1, \dots, x_n\} \quad \text{and} \quad \sigma_n(\mathcal{K}) := \max_{x \in \mathcal{K}} \operatorname{dist}(x, V_n) \; ; \tag{11}$$

repeat

  choose x_{n+1} such that

$$\operatorname{dist}(x_{n+1}, V_n) \geq \gamma \sigma_n(\mathcal{K});$$

until $\sigma_n(\mathcal{K}) < \varepsilon$;

The algorithm produces a finite dimensional space V_n that approximates the set $\mathcal{K}$ within precision ε. As mentioned in [13], it is important to notice that the weak greedy algorithm does not give unique sequences $(x_n)_{n\geq 1}$ and $(\sigma_n(\mathcal{K}))_{n\geq 1}$. However, every chosen sequence decays at the same rate, which under certain assumptions (see Theorem 2), is close to the optimal one. Therefore, the algorithm optimizes the number of steps required to satisfy the given tolerance as well as the dimension of the approximated space V_n.

The pure greedy algorithm corresponds to the case $\gamma = 1$. As noted in [13], the relaxation of the pure greedy method ($\gamma = 1$) to a weak greedy one ($\gamma \in (0, 1]$) will not significantly reduce the efficiency of the algorithm, making it, by the contrary, much easier for implementation.

When using the (weak) greedy algorithm one has to choose the elements of the approximation space by exploring the distance defined in (11) for all possible values $x \in \mathcal{K}$. Two main difficulties arises:

1. the set $\mathcal{K}$ in general consists of infinitely many vectors,
2. in practical implementations, the set $\mathcal{K}$ is often unknown, e.g., it represents the family of solutions to parameter dependent problems.

The first problem can be avoided by searching over some finite discrete subset of $\mathcal{K}$. Here we use the fact that $\mathcal{K}$, being a compact set, can be covered by a finite number of balls of an arbitrary small radius. To circumvent the second one, instead of considering the distance appearing in (11), one uses some *surrogate*, which is easier to compute.

In order to estimate the efficiency of the weak greedy algorithm we compare its approximation rates $\sigma_n(\mathcal{K})$ with the best possible ones.

The best choice of an approximating space V_n is the one producing the smallest approximation error. This smallest error for a compact set $\mathcal{K}$ is called the *Kolmogorov n-width of* $\mathcal{K}$, and is defined as

$$d_n(\mathcal{K}) := \inf_{\dim Y = n} \sup_{x \in \mathcal{K}} \inf_{y \in Y} \|x - y\|_X \, .$$

It measures how well $\mathcal{K}$ can be approximated by a subspace in X of a fixed dimension n.

In the sequel we want to compare $\sigma_n(\mathcal{K})$ with the Kolmogorov width $d_n(\mathcal{K})$, which represents the best possible approximation of $\mathcal{K}$ by a n dimensional subspace of the referent Banach space X. A precise estimate in that direction was provided by [4] in the Hilbert space setting, and later improved and extended to the case of a general Banach space X in [14].

Theorem 2 (Corollary 3.3 of [14]) *For the weak greedy algorithm with constant γ in a Hilbert space X we have the following: If the compact set $\mathcal{K}$ is such that, for some $\alpha > 0$ and $C_0 > 0$*

$$d_n(\mathcal{K}) \leq C_0 n^{-\alpha}, \; n \in \mathbb{N},$$

then

$$\sigma_n(\mathcal{K}) \le C_1 n^{-\alpha},\ n \in \mathbb{N},$$

where $C_1 := \gamma^{-2} 2^{5\alpha+1} C_0$.

This theorem implies that the weak greedy algorithms preserve the polynomial decay rates of the approximation errors. A similar estimate also holds for exponential decays (cf. [14]).

3.2 Definition of the residual

As mentioned before, one of the main goals of this paper is to apply the greedy approach described above to the family of parameter dependent steady state optimal control problems

$$(\mathcal{P}) \qquad \min_{u \in L^2(\omega)} \{J_\nu(u)\}, \tag{12}$$

where J is the cost functional given by

$$J_\nu(u) = \frac{1}{2}|u|^2_{L^2(\omega)} \ + \ \frac{1}{2}\|y_\nu(u) - y^d\|^2_{L^2(\Omega)},$$

while $y_\nu(u)$ is the solution to (4).

As stated in Problem 1, the aim is to choose the most representative set of parameter values ν_i, whose associated controls u_i will provide a good approximation of the control manifold (10). Essential for an effective application of a greedy algorithm is the construction of a residual by which one can estimate the distance between two (possible unknowns) controls by some easily computable quantity.

To construct an appropriate residual, we begin by computing the optimality condition to the minimization problem (12). More precisely, for given $\nu \in K$, the optimality condition read as follows

$$\bar{u}_\nu + S_\nu^*(S_\nu \bar{u}_\nu - y^d) = 0 \tag{13}$$

where $S_\nu : L^2(\omega) \to L^2(\Omega)$ is an operator that assigns to the control u the solution $y_\nu(u)$ of the problem (4), while $S_\nu^* : L^2(\Omega) \to L^2(\omega)$ is its adjoint operator defined by

$$S_\nu^* \, y = \chi_\omega p$$

where p is the solution to the adjoint problem with the right-hand side equal to βy. As mentioned in the previous section, this leads to the following optimality system

$$\begin{aligned} \bar{u}_\nu &= -\chi_\omega \bar{p}_\nu, \\ -\mathrm{div}\,(a_\nu \nabla \bar{y}_\nu) + c_\nu \bar{y}_\nu &= \chi_\omega \bar{u}_\nu, \\ -\mathrm{div}\,(a_\nu \nabla \bar{p}_\nu) + c_\nu \bar{p}_\nu &= \beta\,(\bar{y}_\nu - y^d), \end{aligned} \tag{14}$$

where $\bar{y}_\nu$ stands for the optimal state, while $\bar{p}_\nu$ for the optimal dual variable. As before, **H3** plays a key role to define uniquely thesolution to the optimality system (14).

To this effect one can consider ∇J_ν as a candidate for the residual. Indeed, one can readily verify that

$$\|\bar{u}_\nu - u\|_{L^2(\omega)} \approx \|\nabla J_\nu(u)\|_{L^2(\omega)}.$$

This allows to estimate the distance between the given control u from the (unknown) optimal one $\bar{u}$ by checking the minimization performance of the first , i.e, by calculating $\nabla J_\nu(u)$. However, as ∇J_ν is expressed in terms of S and S^*, that calculation requires solving both the primal and dual problem, which, in general, is expensive (c.f. [36]).

In order to reduce the cost of computing the surrogate, we can rewrite the cost functional in terms of the dual variable p and state variable y. Namely, the gradient of the cost functional is expressed as

$$\nabla J_\nu(u) = u + \chi_\omega p(u),$$

where the dual variable p is related to the control by the relation $\mathbf{1}_\omega p(u) = S_\nu^*(S_\nu u - y^d)$, i.e.

$$\begin{aligned} -\mathrm{div}\,(a_\nu \nabla y(u)) + c_\nu y(u) &= \chi_\omega u \\ -\mathrm{div}\,(a_\nu \nabla p(u)) + c_\nu p(u) &= \beta\left(y(u) - y^d\right). \end{aligned}$$

In what follows, to abridge the notation, for any $z \in H_0^1(\Omega)$ and any parameter $\nu \in K$, we will use the notation

$$L_\nu z := -\mathrm{div}\,(a_\nu \nabla z) + c_\nu z$$

In this way, we introduce the residual operator as

$$R_\nu(p,y) := \begin{pmatrix} L_\nu y + \chi_\omega p \\ L_\nu p - \beta\left(y - y^d\right) \end{pmatrix}. \tag{15}$$

Of course, for optimal variables $(\bar{p}_\nu, \bar{y}_\nu)$ the value of the residual equals zero, and the residual can be equivalently written as

$$R_\nu(p,y) := \begin{pmatrix} L_\nu(y-\bar{y}_\nu) + \chi_\omega(p-\bar{p}_\nu) \\ L_\nu(p-\bar{p}_\nu) - \beta\,(y-\bar{y}_\nu) \end{pmatrix}. \tag{16}$$

By means of the introduced operator we shall try to perform greedy approximation of the manifold

$$\bar{\mathcal{P}} \times \bar{\mathcal{Y}} = \{(\bar{p}_\nu, \bar{y}_\nu) : \nu \in K\} \subset H_0^1(\Omega) \times H_0^1(\Omega).$$

We will show that for an arbitrary $(p,y) \in H_0^1(\Omega) \times H_0^1(\Omega)$ the residual $R_\nu(p,y)$ provides a good measure of the distance of (p,y) from the optimal pair $(\bar{p}_\nu, \bar{y}_\nu)$. To this end, from definition (16), we easily obtain that

$$\begin{aligned} {}_{H^{-1}(\Omega)}\Big\langle R_\nu(p,y), \begin{pmatrix} p-\bar{p}_\nu \\ -(y-\bar{y}_\nu) \end{pmatrix} \Big\rangle_{H^1(\Omega)} &= |\chi_\omega(p-\bar{p}_\nu)|^2_{L^2(\Omega)} + \beta|y-\bar{y}_\nu|^2_{L^2(\Omega)} \\ &\geq 2\sqrt{\beta}\langle \chi_\omega(p-\bar{p}_\nu), y-\bar{y}_\nu\rangle_{L^2(\Omega)}. \end{aligned} \tag{17}$$

Similarly, one gets

$$\left| {}_{H^{-1}(\Omega)}\Big\langle R_\nu(p,y), \begin{pmatrix} y-\bar{y}_\nu \\ 0 \end{pmatrix} \Big\rangle_{H^1(\Omega)} \right| \geq \alpha_1 \|y-\bar{y}_\nu\|^2_{H^1(\Omega)} - \left|\langle \chi_\omega(p-\bar{p}_\nu), y-\bar{y}_\nu\rangle_{L^2(\Omega)}\right| \tag{18}$$

where α_1 is the uniform bound from (5).

Combining the last inequality with (17) we obtain

$$\tilde{c}_1 \|y-\bar{y}_\nu\|^2_{H^1(\Omega)} \leq \|R_\nu(p,y)\|_{H^{-1}(\Omega)} (\|p-\bar{p}_\nu\|_{H^1(\Omega)} + \|y-\bar{y}_\nu\|_{H^1(\Omega)}),$$

where $\tilde{c}_1$ depends on α_1 and β only.

In order to obtain a similar bound for the dual variable, we consider the following product

$$\begin{aligned} {}_{H^{-1}(\Omega)}\Big\langle R_\nu(p,y), \begin{pmatrix} y-\bar{y}_\nu \\ p-\bar{p}_\nu \end{pmatrix} \Big\rangle_{H^1(\Omega)} &= {}_{H^{-1}(\Omega)}\Big\langle L_\nu(y-\bar{y}_\nu), (y-\bar{y}_\nu) \Big\rangle_{H^1(\Omega)} + \langle \chi_\omega(p-\bar{p}_\nu), y-\bar{y}_\nu\rangle_{L^2(\Omega)} \\ &+ {}_{H^{-1}(\Omega)}\Big\langle L_\nu(p-\bar{p}_\nu), (p-\bar{p}_\nu) \Big\rangle_{H^1(\Omega)} - \beta\langle y-\bar{y}_\nu, p-\bar{p}_\nu\rangle_{L^2(\Omega)} \\ &\geq \alpha_1\|y-\bar{y}\|^2_{H^1(\Omega)} - \Big|\langle \chi_\omega(p-\bar{p}_\nu), y-\bar{y}_\nu\rangle_{L^2(\Omega)}\Big| \\ &+ \alpha_1\|p-\bar{p}_\nu\|^2_{H^1(\Omega)} - \beta|\langle y-\bar{y}_\nu, p-\bar{p}_\nu\rangle_{L^2(\Omega)}| \end{aligned}$$

In order to estimate the last term, note that for an arbitrary $\varepsilon > 0$ we have

$$|\langle y-\bar{y}_\nu, p-\bar{p}_\nu\rangle_{L^2(\Omega)}| \leq \frac{1}{2\varepsilon}\|y-\bar{y}_\nu\|^2_{L^2(\Omega)} + \frac{\varepsilon}{2}\|p-\bar{p}_\nu\|_{L^2(\Omega)}.$$

By taking $\varepsilon = \frac{\alpha_1}{\beta}$ from (18), we obtain

$$\begin{aligned} \frac{\alpha_1}{2}(\|y-\bar{y}_\nu\|^2_{H^1(\Omega)} + \|p-\bar{p}_\nu\|^2_{H^1(\Omega)}) &\leq \|R_\nu(p,y)\|_{H^{-1}(\Omega)}(\|p-\bar{p}_\nu\|_{H^1(\Omega)} + \|y-\bar{y}_\nu\|_{H^1(\Omega)}) \\ &+ \Big|\langle \chi_\omega(p-\bar{p}_\nu), y-\bar{y}_\nu\rangle_{L^2(\Omega)}\Big| + \frac{1}{2\alpha_1}\|y-\bar{y}_\nu\|^2_{H^1(\Omega)}. \end{aligned}$$

Finally, combining the last estimate with (17) and (18) we obtain

$$c_1(\|y-\bar{y}_\nu\|_{H^1(\Omega)} + \|p-\bar{p}_\nu\|_{H^1(\Omega)}) \leq \|R_\nu(p,y)\|_{H^{-1}(\Omega)}, \tag{19}$$

for some constant $c_1 > 0$ only depending on α_1 and β.

This shows, in particular, that the introduced residual provides a good estimate of the distance of (p, y) from the optimal pair $(\bar{p}_\nu, \bar{y}_\nu)$.

In order to obtain a reverse type inequality, we explore the regularity estimates (5). Combining it with the inclusion of $H_0^1(\Omega)$ into $H^{-1}(\Omega)$, we obtain

$$\|R_\nu(p, y)\|_{H^{-1}(\Omega)} \leq c_2(\|y - \bar{y}_\nu\|_{H^1(\Omega)} + \|p - \bar{p}_\nu\|_{H^1(\Omega)}), \tag{20}$$

with the constant $c_2 > 0$ depending on α_2 and β only.

The last two inequalities enable us to apply a weak greedy procedure for the approximation of the manifold $\bar{\mathcal{P}} \times \bar{\mathcal{Y}}$ (see Algorithm 2 below).

Algorithm 2: Greedy control algorithm - offline part

Initialize: Fix the approximation error $\varepsilon > 0$.

STEP 1: (discretisation)

Choose a finite subset $\tilde{\mathcal{K}}$ such that

$$(\forall\, \nu \in \mathcal{K}) \quad \mathrm{dist}(\nu, \tilde{\mathcal{K}}) < \delta,$$

where $\delta > 0$ is a constant determined by (25).

STEP 2: (Choosing ν_1)

Check the inequality

$$\max_{\tilde{\nu} \in \tilde{\mathcal{K}}} \|\beta y_\nu^d\|_{H^{-1}(\Omega)} < \frac{\varepsilon}{2}. \tag{21}$$

If it is satisfied, stop the algorithm. Otherwise, choose the first parameter value as

$$\nu_1 \in \operatorname*{argmax}_{\tilde{\nu} \in \tilde{\mathcal{K}}} \{\|y_\nu^d\|_{H^{-1}(\Omega)}\}. \tag{22}$$

and find corresponding optimal primal and dual states $\bar{y}_1, \bar{p}_1$;

STEP 3: (Choosing ν_{j+1})

Having chosen $\nu_1, \ldots, \nu_j$ calculate $R_{\tilde{\nu}}(\bar{p}_j, 0)$ and $R_{\tilde{\nu}}(0, \bar{y}_j)$ for each $\tilde{\nu} \in \tilde{\mathcal{K}}$.

Check the approximation criteria

$$\max_{\tilde{\nu} \in \tilde{\mathcal{K}}} || \inf_{(p,y) \in (\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)} R_{\tilde{\nu}}(p, y) ||_{H^{-1}(\Omega)} < \frac{\varepsilon}{2}. \tag{23}$$

If the inequality is satisfied, stop the algorithm. Otherwise, determine the next parameter value as

$$\nu_{j+1} \in \operatorname*{argmax}_{\tilde{\nu} \in \tilde{\mathcal{K}}} || \inf_{(p,y) \in (\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)} R_{\tilde{\nu}}(p, y) ||_{H^{-1}(\Omega)}, \tag{24}$$

Find the corresponding optimal primal and dual states $\bar{y}_{j+1}, \bar{p}_{j+1}$ and repeat Step 3.

3.3 The greedy algorithm

The residual $R_\nu(p, y)$ introduced in the previous subsection enable us to construct a weak greedy algorithm for an effective construction of an approximating linear space of the control manifold $\overline{\mathcal{U}}$.

The precise description of the offline part of the algorithm is given below in Algorithm 2. This algorithm results in the approximating space $(\overline{\mathcal{P}}_n, \overline{\mathcal{Y}}_n) = \mathrm{span}\{(\bar{p}_1, \bar{y}_1) \ldots, (\bar{p}_n, \bar{y}_n)\}$, where n is a number of chosen snapshots (specially $(\overline{\mathcal{P}}_n, \overline{\mathcal{Y}}_n) = \{0\}$ for $n = 0$). In particular, the obtained bounds on the residual (19)–(20) ensure that Algorithm 2 satisfies the requirements of the weak greedy procedure. More precisely, the following result holds.

Theorem 3 *For a given $\varepsilon > 0$ take the discretisation constant δ such that*

$$\delta \leq \varepsilon/(2C_L c_1). \tag{25}$$

Then the algorithm 2 provides a week greedy approximation of the manifold $(\overline{\mathcal{P}}, \overline{\mathcal{Y}})$ with the constant

$$\gamma = \frac{c_1}{2c_2}, \tag{26}$$

and the approximation error less than ε/c_1. Here C_L is the Lipschitz constant of the mapping $\nu \to (\bar{p}_\nu, \bar{y}_\nu)$, while c_1 and c_2 are constants appearing in estimates (19) and (20), respectively.

Remark 3 Note that the above algorithm definitely stops after finite, $n \in N_0$ number of iterations due to the compactness assumption on the parameter set $\mathcal{K}$. In the sequel we exclude the case $n = 0$, occurring when inequality (21) holds, which results in a null approximating space for which the statement trivially holds

Remark 4 Note that the infimum appearing in (23) can be expressed as a distance from a suitable space determined by the residual R_ν. More precisely, denote by G_ν the linear part of the residual R_ν, i.e.

$$G_\nu(p, y) := \begin{pmatrix} L_\nu y + \chi_\omega p \\ L_\nu p - \beta y \end{pmatrix}, \tag{27}$$

implying

$$R_\nu(p, y) = G_\nu(p, y) + \begin{pmatrix} 0 \\ \beta y^d \end{pmatrix}.$$

Denoting by $G_\nu(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)$ the space spanned by $G_\nu(\bar{p}_i, \bar{y}_i), i = 1, \ldots, j$, we obtain

$$\| \inf_{(p,y)\in(\overline{\mathcal{P}}_j,\overline{\mathcal{Y}}_j)} R_{\tilde{\nu}}(p, y)\|_{H^{-1}(\Omega)} = \operatorname{dist}\left(\begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}, G_\nu(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)\right). \tag{28}$$

The last relation enables practical computation of the infimum appearing in (23) which can be performed by projecting vector $\begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}$ to the space $G_\nu(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)$.

Proof (Proof of Theorem 3) We will adapt the weak greedy procedure presented in Algorithm 1. For this particular problem, we have to show that

i) the selected optimal states $(\bar{p}_i, \bar{y}_i)$ associated to the parameter values determined by (22) and (24) satisfy

$$\|(\bar{p}_{\nu_1}, \bar{y}_{\nu_1})\|_{H^1(\Omega)} \geq \gamma \max_{\nu\in\mathcal{K}} \|(\bar{p}_\nu, \bar{y}_\nu)\|_{H^1(\Omega)}$$

$$\operatorname{dist}\Big((\bar{p}_{\nu_{j+1}}, \bar{y}_{\nu_{j+1}}), (\bar{\mathcal{P}}_i, \bar{\mathcal{Y}}_i)\Big) \geq \gamma \max_{\nu\in\mathcal{K}} \operatorname{dist}\Big((\bar{p}_\nu, \bar{y}_\nu), (\bar{\mathcal{P}}_i, \bar{\mathcal{Y}}_i)\Big), \quad j = 1, \ldots, n-1$$

with the constant γ given by (26),

ii) the approximation error, defined as

$$\sigma\Big((\bar{p}, \bar{y})(\mathcal{K})\Big) := \max_\nu \operatorname{dist}\Big((\bar{y}_\nu, \bar{p}_\nu), (\bar{\mathcal{P}}_i, \bar{\mathcal{Y}}_i)\Big)$$

obtained at the end of the algorithm is less than ε/c_1.

In order to prove i), we begin by noting that from estimate (19) and the definition of the residual R_ν, the following inequality holds

$$\|(\bar{p}_\nu, \bar{y}_\nu)\|_{H^1(\Omega)} = \operatorname{dist}\Big((\bar{p}_\nu, \bar{y}_\nu), (0, 0)\Big) \leq \left\|\frac{1}{c_1} R_\nu(0, 0)\right\|_{H^{-1}(\Omega)} = \left\|\frac{1}{c_1}\begin{pmatrix} 0 \\ \beta y^d_\nu \end{pmatrix}\right\|_{H^{-1}(\Omega)},$$

for all $\nu \in \mathcal{K}$. From the choice of the first snapshot (22) for every parameter $\tilde{\nu}$ from the discretised set $\tilde{\mathcal{K}}$ the above inequality readily implies

$$\begin{aligned} \|(\bar{p}_{\tilde{\nu}}, \bar{y}_{\tilde{\nu}})\|_{H^1(\Omega)} &\leq \left\|\frac{1}{c_1}\begin{pmatrix} 0 \\ \beta y^d_{\nu_1}, \end{pmatrix}\right\|_{H^{-1}(\Omega)}, \\ &= \left\|\frac{1}{c_1} R_{\nu_1}(0, 0)\right\|_{H^{-1}(\Omega)} \leq \frac{c_2}{c_1}\|(\bar{p}_{\nu_1}, \bar{y}_{\nu_1})\|_{H^1(\Omega)}, \end{aligned} \tag{29}$$

where in the last line we have used the estimate (20).

In order to obtain a similar inequality for every $\nu \in \mathcal{K}$, by taking $\tilde{\nu} \in \tilde{\mathcal{K}}$ such that $|\nu - \tilde{\nu}| < \delta$, by means of (25) and (29) we have

$$\begin{aligned} \|(\bar{p}_\nu, \bar{y}_\nu)\|_{H^1(\Omega)} &\leq \|(\bar{p}_\nu - \bar{p}_{\tilde{\nu}}, \bar{y}_\nu - \bar{y}_{\tilde{\nu}})\|_{H^1(\Omega)} + \|(\bar{p}_{\tilde{\nu}}, \bar{y}_{\tilde{\nu}})\|_{H^1(\Omega)} \\ &\leq \frac{\varepsilon}{2c_1} + \frac{c_2}{c_1}\|(\bar{p}_{\nu_1}, \bar{y}_{\nu_1})\|_{H^1(\Omega)}. \end{aligned}$$

Having excluded the case (21), we have

$$\frac{\varepsilon}{2} \leq \|R_{\nu_1}(0,0)\|_{H^{-1}(\Omega)} \leq c_2 \|(\bar{p}_{\nu_1}, \bar{y}_{\nu_1})\|_{H^1(\Omega)},$$

implying the first inequality in i).

The same arguments can be employed for the general j-th iteration. Recall that the stopping criterion (23) is not fulfilled for iterations $j < n$. Suppose that we have selected $\nu_1, \ldots, \nu_i$ different parameters. From inequalities (19), (20) and the definition of the next snapshot (see eq. (24)), for each $\tilde{\nu} \in \tilde{\mathcal{K}}$ we readily obtain

$$\begin{aligned}\operatorname{dist}\Big((\bar{p}_{\tilde{\nu}}, \bar{y}_{\tilde{\nu}}), (\bar{\mathcal{P}}_i, \bar{\mathcal{Y}}_i)\Big) &\leq \|\frac{1}{c_1} \inf_{(p,y)\in(\overline{\mathcal{P}}_i, \overline{\mathcal{Y}}_i)} R_{\tilde{\nu}}(p,y)\|_{H^{-1}(\Omega)} \\ &\leq \|\frac{1}{c_1} \inf_{(p,y)\in(\overline{\mathcal{P}}_i, \overline{\mathcal{Y}}_i)} R_{\nu_{j+1}}(p,y)\|_{H^{-1}(\Omega)} \leq \frac{c_2}{c_1} \operatorname{dist}\Big((\bar{p}_{\nu_{j+1}}, \bar{y}_{\nu_{j+1}}), (\bar{\mathcal{P}}_i, \bar{\mathcal{Y}}_i)\Big).\end{aligned}$$

As before, the above inequality can be generalized to all $\nu \in \mathcal{K}$ by employing Lipschitz continuity of the mapping $\nu \to (\bar{p}, \bar{y})$, particularly implying the second inequality in i).

To prove ii), it is straightforward to see from equation (19) that after n iterations and for any $\nu \in \mathcal{K}$ by selecting $\tilde{\nu} \in \tilde{\mathcal{K}}$ such that $|\nu - \tilde{\nu}| < \delta$, the following holds

$$\begin{aligned}\operatorname{dist}\Big((\bar{p}_{\nu}, \bar{y}_{\nu}), (\bar{\mathcal{P}}_n, \bar{\mathcal{Y}}_n)\Big) &\leq \|(\bar{p}_{\nu} - \bar{p}_{\tilde{\nu}}, \bar{y}_{\nu} - \bar{y}_{\tilde{\nu}})\|_{H^1(\Omega)} + \operatorname{dist}\Big((\bar{p}_{\tilde{\nu}}, \bar{y}_{\tilde{\nu}}), (\bar{\mathcal{P}}_n, \bar{\mathcal{Y}}_n)\Big) \\ &\leq \frac{\varepsilon}{2c_1} + \frac{1}{c_1}\|R_{\tilde{\nu}}(\bar{\mathcal{P}}_n, \bar{\mathcal{Y}}_n)\|_{H^{-1}(\Omega)} < \frac{\varepsilon}{c_1},\end{aligned}$$

since the stopping condition (23) has been reached. This ends the proof of Theorem 3.

It remains to check the effect of the approximation of the manifold $\bar{\mathcal{P}} \times \bar{\mathcal{Y}}$ on the control manifold $\bar{\mathcal{U}}$ and, in the end, the approximation of the cost functional J_ν.

In the next step we want to explore the approximation space constructed in the offline part of the algorithm with the aim of an effective construction of an approximating states and controls for an arbitrary given parameter value. This is the step of the greedy method usually referred to as the *online part*.

To this effect we propose the following algorithm for the online part of the greedy procedure.

Algorithm 3: Greedy control algorithm - online part

Initialize: A parameter value $\nu \in \mathcal{K}$ is given.

STEP 1:

Project the vector $\begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}$ to $G_\nu(\overline{\mathcal{P}}_n, \overline{\mathcal{Y}}_n)$, the space spanned by $G_\nu(\overline{p}_i, \overline{y}_i), i = 1, \ldots, n$.

Here $\{(\overline{p}_i, \overline{y}_i), i = 1, \ldots, n\}$ are snapshots selected in the offline part of the algorithm, while G_ν is the linear part of the residual given by (27).

Denote the projection by $P_n \begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}$.

STEP 2:

Solve the system

$$P_n \begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix} = \sum \alpha_i G_\nu(\overline{p}_i, \overline{y}_i).$$

STEP 3:

Define the approximating dual state as

$$p^*_\nu = \sum \alpha_i \overline{p}_i.$$

STEP 4:

Define the approximating optimal control as

$$u^*_\nu = -\chi_\omega p^*_\nu. \tag{30}$$

STEP 5:

Define the approximating optimal state $y^\star_\nu$ as the solution to

$$L_\nu y^\star_\nu = \chi_\omega u^\star_\nu. \tag{31}$$

Using the above proposed definition of the approximative optimal control (30) and the approximation performance of the offline part of the algorithm we immediately obtain

$$\|u_\nu^\star - \bar{u}_\nu\|_{L^2(\omega)} \le \|p_\nu^\star - \bar{p}_\nu\|_{H^1(\Omega)} \le \frac{\varepsilon}{c_1}, \tag{32}$$

which provides the required estimate proposed by Problem 1 (up to a scaling factor c_1).

Similarly, from the regularity bound (5) it follows

$$\|y_\nu^\star - \bar{y}_\nu\|_{H^1(\Omega)} \le \frac{1}{\alpha_1}\|u_\nu^\star - \bar{u}_\nu\|_{L^2(\omega)} \le \frac{\varepsilon}{\alpha_1 c_1}, \tag{33}$$

which provides the estimate on the approximative optimal state obtained by the greedy procedure.

Finally, we obtain the estimate for the error in the cost functional. Note that

$$|J(u_\nu^\star) - J(\bar{u}_\nu)| = \frac{1}{2}\left(\|\bar{u}_\nu\|^2_{L^2(\Omega)} - \|u_\nu^\star\|^2_{L^2(\Omega)} + \|\bar{y}_\nu - y^d\|^2_{L^2(\Omega)} - \|y_\nu^\star - y^d\|^2_{L^2(\Omega)}\right). \tag{34}$$

The first two terms on the right hand side can be bounded as follows:

$$\begin{aligned}\int_\Omega |\bar{u}_\nu|^2 - |u_\nu^\star|^2 dx &\le \int_\Omega (|\bar{u}_\nu| - |u_\nu^\star|)(|\bar{u}_\nu| + |u_\nu^\star|)dx \\ &\le \int_\Omega |\bar{u}_\nu - u_\nu^\star|(|\bar{u}_\nu - u_\nu^\star| + 2|\bar{u}_\nu|)dx \\ &\le \|\bar{u}_\nu - u_\nu^\star\|_{L^2(\Omega)}\left(\|\bar{u}_\nu - u_\nu^\star\|_{L^2(\Omega)} + 2\|\bar{u}_\nu\|_{L^2(\Omega)}\right).\end{aligned} \tag{35}$$

By using optimality condition (13) one easily obtains the bound

$$\|\bar{u}_\nu\|_{L^2(\Omega)} \le \|S^* y^d\|_{L^2(\Omega)}, \tag{36}$$

which together with (32) enables the first difference in (35) to be estimated in terms of the approximation error ε and the given problem data.

Similarly, for the last two terms in (34) one gets

$$\begin{aligned}\|\bar{y}_\nu - y^d\|^2_{L^2(\Omega)} - \|y_\nu^\star - y^d\|^2_{L^2(\Omega)} &\le \|\bar{y}_\nu - y_\nu^\star\|_{L^2(\Omega)}\left(\|\bar{y}_\nu - y_\nu^\star\|_{L^2(\Omega)} + 2\|\bar{y}_\nu - y^d\|_{L^2(\Omega)}\right) \\ &\le \|\bar{y}_\nu - y_\nu^\star\|_{L^2(\Omega)}\left(\|\bar{y}_\nu - y_\nu^\star\|_{L^2(\Omega)} + \frac{2}{\alpha_1}\|S^* y^d\|_{L^2(\Omega)} + 2\|y^d\|_{L^2(\Omega)}\right)\end{aligned}$$

where the last inequality follows from the regularity assumptions (5), the definition of the optimal state $(14)_2$ and the estimate (36).

Thus, combining estimates on the error on approximate control (32) and the approximate state (33), one obtains bound on the approximation of the minimal cost functional value in terms of the approximation error ε and the desired trajectory y^d.

We can summarise the obtained results in the following theorem

Theorem 4 *The proposed greedy algorithm provides the following estimates on the approximate control, approximate optimal state and approximate minimal value of the cost functional*

- $\|u_\nu^\star - \bar{u}_\nu\|_{L^2(\Omega)} \le \varepsilon,$
- $\|y_\nu^\star - \bar{y}_\nu\|_{H^1(\Omega)} \le \left(\frac{\varepsilon}{\alpha_1}\right),$
- $|J(u_\nu^\star) - J(\bar{u}_\nu)| \le \frac{\varepsilon}{c_1}\left(\frac{\varepsilon}{c_1}\left(1 + \frac{1}{\alpha_1^2}\right) + 2\left(1 + \frac{1}{\alpha_1^2}\right)\|S^* y_d\|_{L^2(\Omega)} + \frac{2}{\alpha_1}\|y_d\|_{L^2(\Omega)}\right),$

where $u_\nu^\star$ and $y_\nu^\star$ are given by (30) *and* (31)*, respectively, while α_1 is the elliptic bound appearing in* (5)*.*

3.4 Optimal approximation rates

One of main advantages of the greedy procedure is that it provides optimal approximation rates in terms of Kolmogorov widths. However, although the Kolmogorov widths of a set of admissible parameters (set K in our problem) is usually easy to estimate, this is not the case for the corresponding set of controls ($\bar{\mathcal{U}}$ in our setting).

Fortunately, a result in that direction has been provided recently for holomorphic mappings ([10]) under the assumption of a polynomial decay of Kolmogorov widths.

Theorem 5 *For a pair of complex Banach spaces X and V assume that u is a holomorphic map from an open set $\mathcal{O} \subset X$ into V with uniform bound. If $\mathcal{K} \subset \mathcal{O}$ is a compact subset of X then for any $\alpha > 1$ and $C_0 > 0$*

$$d_n(\mathcal{K}) \le C_0 n^{-\alpha} \quad \Longrightarrow \quad d_n(u(\mathcal{K})) \le c_1 n^{-\beta}, \quad n \in \mathbb{N},$$

for any $\beta < \alpha - 1$ and the constant C_1 depending on C_0, α and the mapping u.

Remark 5 The proof of the theorem provides an explicit estimate of the constant C_1 in dependence on C_0, α and the mapping u. However, due to its rather complicated form we do not expose it here.

Going back to our problem, Theorem 5 can be applied if we proved that the mapping $\nu \to \bar{u}_\nu$ is analytic, which in terms of the optimality equations (14) provides that the mapping K to $\bar{\mathcal{P}} \times \bar{\mathcal{Y}}$ is analytic as well.

To this effect we explore analytic version of the implicit function theorem and apply it to the mapping

$$D_u J_\nu(u) = u + S_\nu^*(S_\nu u - y_d)\,.$$

Of course, for an optimal control it holds $D_u J_\nu(\bar{u}_\nu) = 0$, and in order to employ the implicit function theorem we have to check if the operator $D_u(D_u J_\nu)(\bar{u}_\nu)$ is an isomorphism. As

$$D_u(D_u J_\nu)(\bar{u}_\nu) = I + S_\nu^* S_\nu$$

and using that $S_\nu^* S_\nu$ is positive semi-definite it follows that $D_u(D_u J_\nu)(\bar{u}_\nu)$ is coercive, bounded operator, thus allowing for the inverse.

Furthermore, the operator S_ν inhertis regulrity properties of the coefficients a_ν, c_ν in its dependence with respect on ν, and this on all levels. Specially, assuming the mappings $\nu \to a_\nu$ and $\nu \to c_\nu$ are analytic the same holds for the map $\nu \to S_\nu$, and similarly for S_ν^*. Thus the implicit function theorem implies that the unique mapping $\nu \to \bar{u}_\nu$ is analytic.

Combining this result with Theorem 5 we obtain the following one.

Corollary 1 *Let $\mathcal{O} \in \mathbb{R}^d$ be an open domain containing the parameter set $\mathcal{K}$ and let the mappings $a_\nu, c_\nu : \mathcal{O} \to L^\infty(\Omega)$ be analytic. Then the greedy control algorithm ensures a polynomial decay of arbitrary order of the approximation rates.*

The last result can be extend to the case of infinite number of parameters as well, provided the Kolmogorov widths of the set $\mathcal{K}$ decay polynomially.

4 Numerical experiments

We devote this section to present two numerical examples where the greedy approach is applied. To illustrate the procedure, we consider the two dimensional domain $\Omega = (0,1)^2$. In order to approximate the solution of the elliptic problem (4) we use the elementary finite difference (FD) method.

We shall use uniform meshes, i.e., meshes with constant discretization steps in each direction. That is, for given $N_{x_1}, N_{x_2} \in \mathbb{N}^*$, we set

$$x_{1,i} = ih_{x_1}, \quad i \in [0, N_{x_1}+1], \quad h_{x_1} = \frac{1}{N_{x_1}+1},$$
$$x_{2,j} = jh_{x_2} \quad j \in [0, N_{x_2}+1], \quad h_{x_2} = \frac{1}{N_{x_2}+1}.$$

Thus, by abuse of notation, we denote by $y_{i,j}$ an approximate value of the solution to (4) at the grid points (x_i, y_j).

We will approximate the elliptic operator $\mathcal{A}_\nu = -\text{div}(a_\nu \nabla \cdot)$ with homogeneous Dirichlet boundary conditions, by using the standard 5-point discretization given by

$$(\mathcal{A}_h^\nu y)_{i,j} = -\,a^\nu_{i+\frac{1}{2},j}\frac{y_{i+1,j}-y_{i,j}}{h_{x_1}^2} + a^\nu_{i-\frac{1}{2},j}\frac{y_{i,j}-y_{i-1,j}}{h_{x_1}^2} \\ -\,a^\nu_{i,j+\frac{1}{2}}\frac{y_{i,j+1}-y_{i,j}}{h_{x_2}^2} + a^\nu_{i,j-\frac{1}{2}}\frac{y_{i,j}-y_{i,j-1}}{h_{x_2}^2} \tag{37}$$

where we set $a^\nu_{i\pm\frac{1}{2},j} = a_\nu(x_i \pm h_{x_1}/2, y_j)$ and so on. Observe that the boundary conditions are taken into account in those formulas by imposing that $y_{0,j} = y_{N_{x_1}+1,j} = 0$, $1 \leq j \leq N_{x_2}$ and $y_{i,0} = y_{i,N_y+1} = 0$, $1 \leq i \leq N_{x_1}$.

The linear system of algebraic equations derived from the FD discretization (37) in $N \times N$ grid points for an elliptic PDE has $O(N^2)$ equations, so the coefficient matrix has $O(N^2 \times N^2)$ entries. Even for a small number $N = 200$, the resulting matrix cannot be stored and handled in most computers. However, the matrix for a self adjoint elliptic operator as in (4) obtained from (37) is sparse with number of non-zero entries $\sim 5N^2$. Thus, sparse matrix techniques will be used in what follows. We refer to Section 6 for a more detailed description of the computational cost of implementing the greedy approach.

4.1 Greedy test # 1

Let us consider a coefficient a_ν of the form

$$a_\nu(x) = 1 + \nu(x_1^2 + x_2^2)$$

where we assume that

$$\nu \in [1, 10] = \mathcal{K}. \tag{38}$$

In this way, we can readily verify that a_ν satisfies **H1** and **H2**.

On the other hand, we take the coefficient c as

$$c(x) = \sin(\pi x_1)\sin(\pi x_2).$$

which clearly fulfills the condition $c \geq 0$. For the discretization of the elliptic problem, we choose $N_{x_1} = N_{x_2} = 400$ to have a uniform, equally sized mesh in both directions. Finally, for the optimal control problem, we set the parameter $\beta = 10^4$ while the desired target is the x_2-independent function

$$y^d(x) = \sin(2\pi x_1).$$

We take ω as the region $(0.2, 0.5) \times (0, 5, 0.9) \cup (0.5, 0.9) \times (0.1, 0.9)$ in the x_1x_2-plane (see Figure 2b).

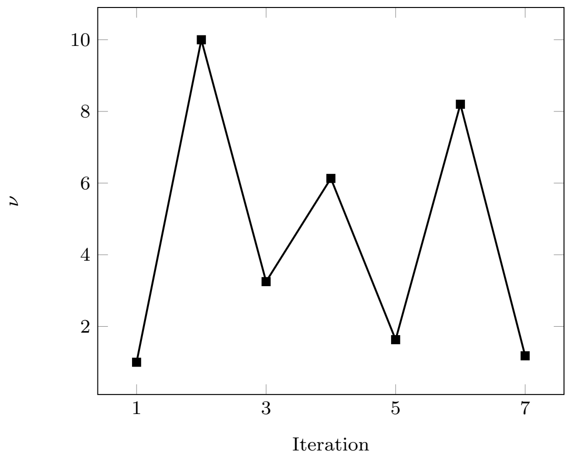
Since the functional J_ν is quadratic and coercive, a standard conjugate gradient (CG) algorithm is a quite natural and simple choice to solve the minimization problem (7). In fact, for a given $\nu \in \mathcal{K}$ one can directly solve the minimization problem (12). The average time for computing the corresponding control up to a given tolerance of 10^{-8} using the CG is around six seconds.

As mentioned in [28], Hypotheses **H1** and **H2** allow us to implement a *naive* approach for approximating the parameterized control set $\bar{\mathcal{U}}(\mathcal{K})$. This approach consists in discretizing the parameter set in a very fine mesh, that we denote it by $\tilde{\mathcal{K}}$, and then computing the corresponding control for each parameter in this finite-dimensional set. If the number of parameters in $\tilde{\mathcal{K}}$ is rather small, or the desired precision ε is large enough, one can directly use the naive approach.

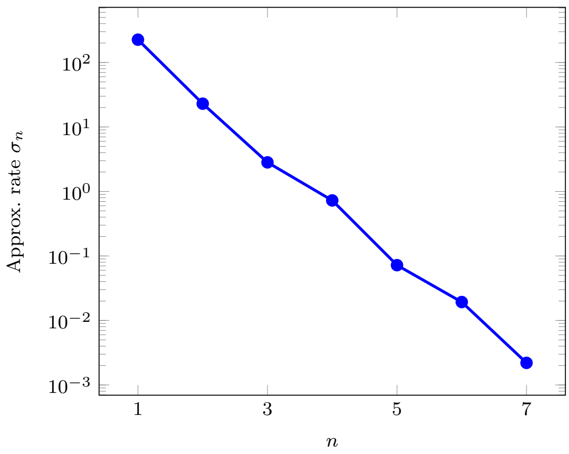
For instance, in this particular test, let us take $\tilde{\mathcal{K}}$ as the uniform discretization of the interval (38) in $k = 100$ values. Then, the *naive* approach amounts to solve 100 different times the minimization problem (12). The whole process takes around 600 seconds and requires to store all the computed controls. Therefore, one expects that the greedy approach will help us to reduce the computing effort.

We apply the greedy procedure described in Algorithm 2 over the set $\tilde{\mathcal{K}}$ and choosing a precision of $\varepsilon = 0.005$. The algorithm stops after seven iterations selecting 7 (out of 100) parameter values. The corresponding optimal states $(\bar{p}_{\nu_i}, \bar{y}_{\nu_i})$ associated to each selected parameter are stored in memory.

The time needed to finish the offline part of the greedy algorithm takes 476 seconds. Compared to the time that takes the *naive* approach, this might seem as a marginal gain. As shown in Section 6, the offline part is costly. However, this test being an academic example, does not exhibit all of the potential of the greedy theory. In fact, in real life applications where simulation time may take hours, it is indeed much



(a) Distribution of the selected parameters



(b) Approximation error

Fig. 1: Numerical results of the greedy test #1.

cheaper to compute a residual as (15) than solving a large number of control problems associated to that many parameters.

The way the parameters are chosen for this test is illustrated in Figure 1a. One can see that the selected values are taken from different parts of the parameter set in a zig-zag manner, leading to the best possible approximation. Indeed, in Figure 1b, we plot the approximation rate of the greedy algorithm corresponding to

$$\sigma_n(\tilde{\mathcal{K}}) = \max_{\nu \in \tilde{\mathcal{K}}} \operatorname{dist}\Big((\bar{p}_\nu, \bar{y}_\nu), (\bar{\mathcal{P}}_n, \bar{\mathcal{Y}}_n)\Big).$$

Such plot suggests an exponential decay of the approximation rate σ_n, which is in accordance with Corollary 1 that provides a polynomial decay at any rate.

Following the discussion in Section 3, once the offline part is completed, we can construct (approximate) optimal controls by choosing a suitable combination of the optimal states $\bar{p}_{\nu_i}$. The methodology to construct such approximations is described in Algorithm 3.

In Figure 2, we plot the approximation of the real control with the greedy algorithm for $\nu = \sqrt{2}$. The approximated control is nearly identical to the obtained, for instance, by minimizing directly functional (7) for the associated value $\nu = \sqrt{2}$. In fact, one can obtain for this particular experiment that

$$\|\bar{u}_{\sqrt{2}} - u^\star_{\sqrt{2}}\|_{L^2(\Omega)} \approx 4.61 \times 10^{-5}.$$

In spite of obtaining almost the same solution, the convergence of conjugate gradient method takes 5.4 s compared to 0.488 s in the online greedy part. This fact also shows the computational efficiency of the proposed algorithm.

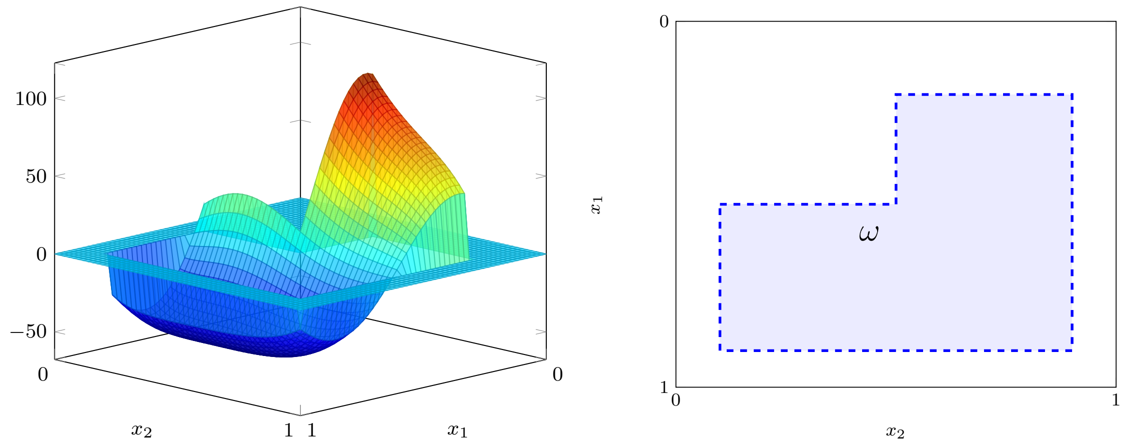
In Figure 3a, we plot the solution to (4) using the approximated control $u^\star_{\sqrt{2}}$. As for the control, we can compute the difference between the real optimal state against the approximate optimal state $\bar{y}^\star_{\sqrt{2}}$. In this particular test, we obtain the following estimate

$$\|\bar{y}_{\sqrt{2}} - y^\star_{\sqrt{2}}\|_{L^2(\Omega)} \approx 7.14 \times 10^{-7}.$$

Note, however, that the approximation of the state to the desired target (see figure 3b) is quite different from what one may expect. This fact is not related with the greedy procedure. Actually, it is closely related to the geometry of the domain ω and the penalization parameter β. One can obtain a better approximation of the desired target y^d by taking a larger value of β, but at the price of increasing the L^2-norm of the control, or by taking a bigger control set ω, leading to an increasing number of degrees of freedom.

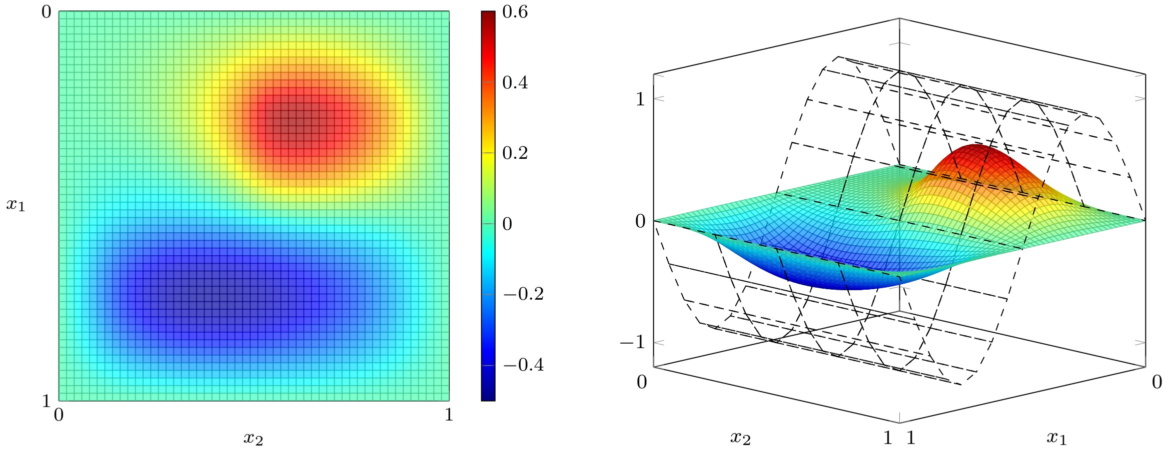
4.2 Greedy test # 2

Here, we present a series of experiments for the case that fit the setting of example 2 in Remark 2. As mentioned before, this will be of interest when discussing the applicability of the method for the turnpike problem.



(a) The approximated control $x \mapsto \bar{u}^{\star}_{\sqrt{2}}(x)$ (b) The control set ω

Fig. 2: Approximated control by greedy methods.



(a) The state $x \mapsto \bar{y}^{\star}_{\sqrt{2}}(x)$ (b) The state $\bar{y}^{\star}_{\sqrt{2}}$ and the target function y^d (dashed)

Fig. 3: Controlled solution.

To this end, let us consider the case where $a(x) \equiv 1$ for every parameter ν. It is well-known that the eigenvalues of the Laplacian on a square are given by

$$\lambda_{nm} = \pi^2(n^2 + m^2), \quad n, m \in \mathbb{N}.$$

Hence, a straightforward example for testing our greedy methodology is the case where

$$c(x, \nu) = \nu$$

with $\nu \in [-5\pi^2 + \epsilon, -2\pi^2 - \epsilon]$ and $\epsilon > 0$. We see that in this case, **H3** holds as long as we are sufficiently far from the eigenvalues.

For simulation purposes, let us take N_{x_1}, N_{x_2} and β as in the previous test. In addition, we consider

$$\nu \in [-45, -20] = \mathcal{K}$$

and the desired target

$$y^d(x) = \sin(\pi x_1)$$

We will use as a control region the shape depicted in Figure 4. For this particular test, the average time for computing the optimal control for different values of ν is around 10 seconds. Implementing the *naive* approach for the refined mesh $\tilde{\mathcal{K}}$ implies that at least 1000 seconds are needed to finish this process. We will see that by means of the greedy approach we can reduce the computational effort.

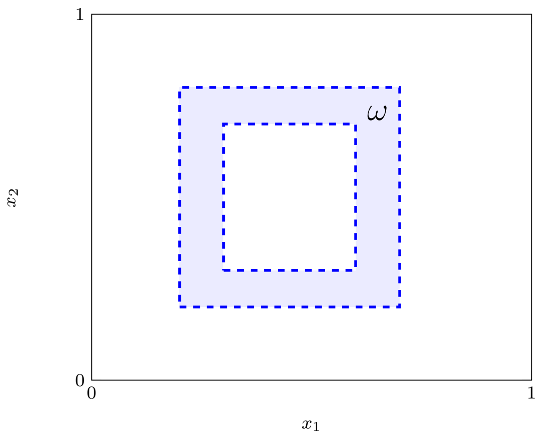
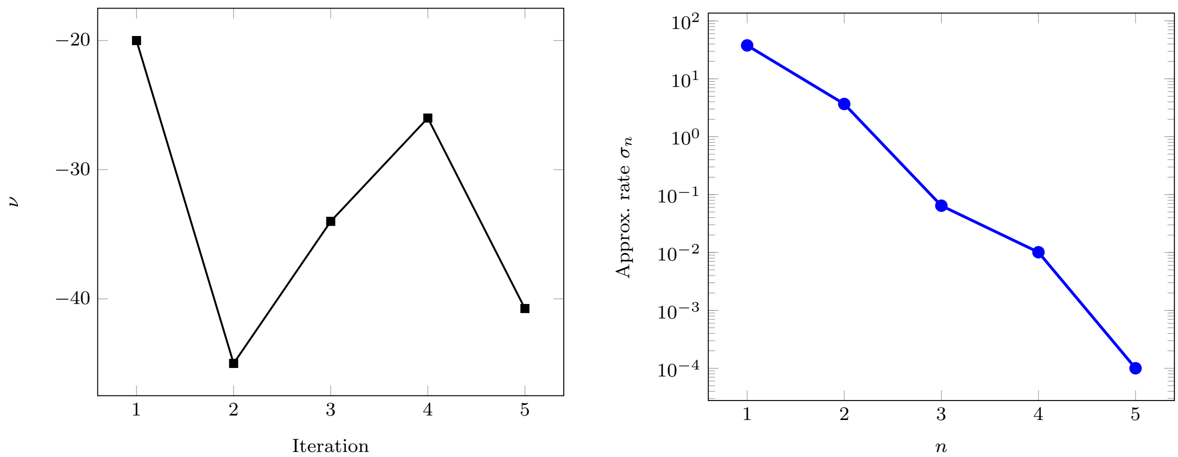


Fig. 4: Control region for the second test.



(a) Distribution of the selected parameters

(b) Approximation error

Fig. 5: Numerical results of the greedy test #2.

Using our computational tool, we choose again the approximation tolerance $\varepsilon = 0.005$. The greedy algorithm stops after five iterations. We present in figure 5a the selected parameters in the order they were chosen by the program. As in the previous test, the selection of the parameters leads to an exponential decay of the approximation rates σ_n (see figure 5b). The elapsed time for completing the offline process is 389 seconds. This shows a clear improvement of almost 60% less time with respect to the time consumed in *naive* approach.

In Figure 6, we plot the approximation of the optimal control obtained by the greedy method for a chosen value of $\nu = -5\pi^2/2$, while in Figure 7 we show corresponding controlled state and a comparison with the target y^d. For this particular test, the elapsed time to compute the online control is 0.406 seconds while computing the optimal control directly by means of the CG method takes 15.9 seconds. The approximation error with respect to the real control is

$$\|\bar{u}_{-5\pi^2/2} - u^\star_{-5\pi^2/2}\|_{L^2(\Omega)} \approx 5.81 \times 10^{-4},$$

while the approximation error for the state using the real and approximated control is

$$\|\bar{y}_{-5\pi^2/2} - y^\star_{-5\pi^2/2}\|_{L^2(\Omega)} \approx 6.69 \times 10^{-5}.$$

In Figure 7, we observe that the approximation to the desired target y^d is (to a certain extent) better than the one we obtain in greedy test #1. This is because the chosen y^d does not change sign in the domain Ω (cf. Figure 3b). This also translates into a less effort by the control (see Figure 6), even if the control set ω is smaller than in the previous case (cf. Figure 2).

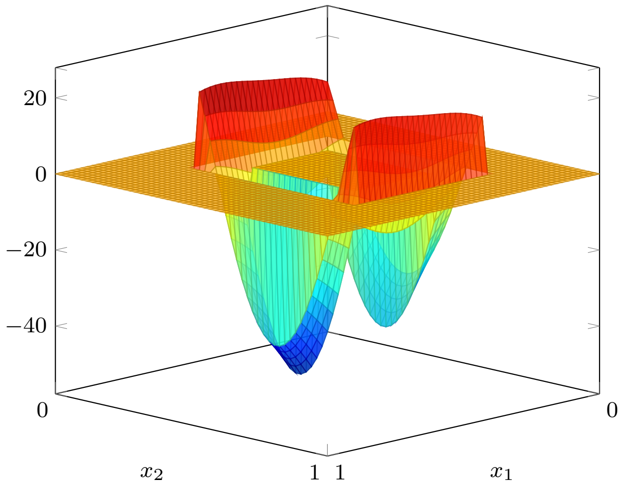
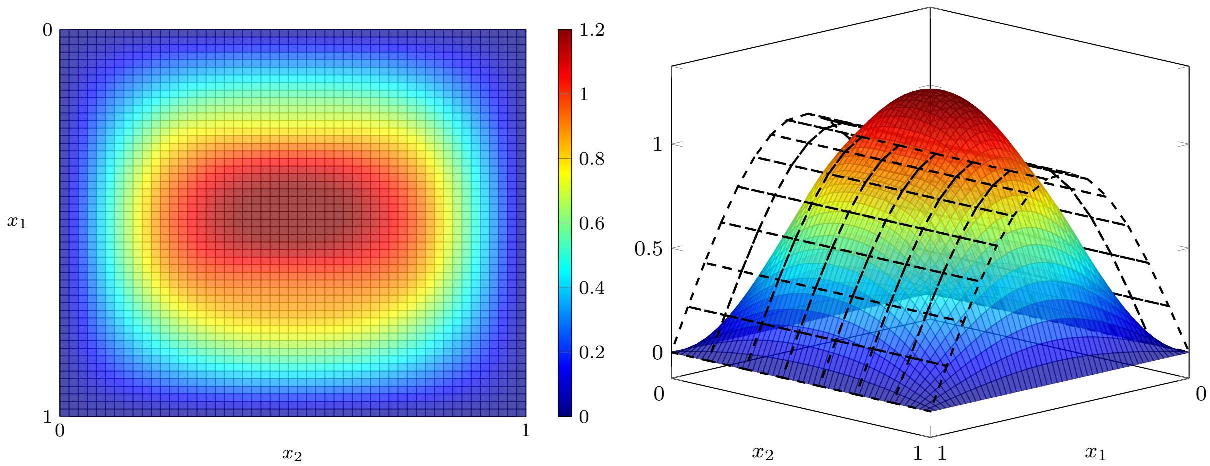


Fig. 6: The approximated control $x \mapsto \bar{u}^{\star}_{-5\pi^2/2}(x)$.



(a) The state $x \mapsto \bar{y}^{\star}_{-5\pi^2/2}(x)$

(b) The state $\bar{y}^{\star}_{-5\pi^2/2}$ and the target function y^d (dashed)

Fig. 7: Controlled solution.

5 Connection with turnpike problems

We devote this section to discuss the utility and limitations of the greedy approach for elliptic equations in the control of evolution problems.

For the moment, let us turn our attention to the time dependent optimization problem

$$\min_{u \in L^2(0,T;L^2(\omega))} J^T(u) = \frac{1}{2}\int_0^T |u(t)|^2_{L^2(\omega)} + \frac{\beta}{2}\int_0^T \|y(t) - y_d\|^2_{L^2(\Omega)}, \tag{39}$$

where y is solution to the parabolic problem

$$\begin{cases} y_t - \operatorname{div}(a_\nu \nabla y) + c_\nu y = 1_\omega u, & \text{in } \Omega \times (0,T), \\ y = 0, & \text{on } \partial\Omega \times (0,T), \\ y(x,0) = y^0(x), & \text{in } \Omega. \end{cases} \tag{40}$$

As stated in Section 2, the turnpike property through Theorem 1 ensures that the optimal state and control $(u^T(t), y^T(t))$ solution to (39) both simplify exponentially in time large to their steady counterpart $(\bar{u}, \bar{y})$. From this fact, a first intuitive experiment is to put the control $\bar{u}$ in (40) and test its capability to control the system.

Here, we present a series of experiments related to the (approximated) optimal steady controls in Sections 4.1 and 4.2. We use them to control the corresponding evolution equation and test their efficiency. We will differentiate two main cases to be studied.

5.1 The case $c(x) \geq 0$

In addition to the parameters already set in the greedy test #1 (see Section 4.1), let us take

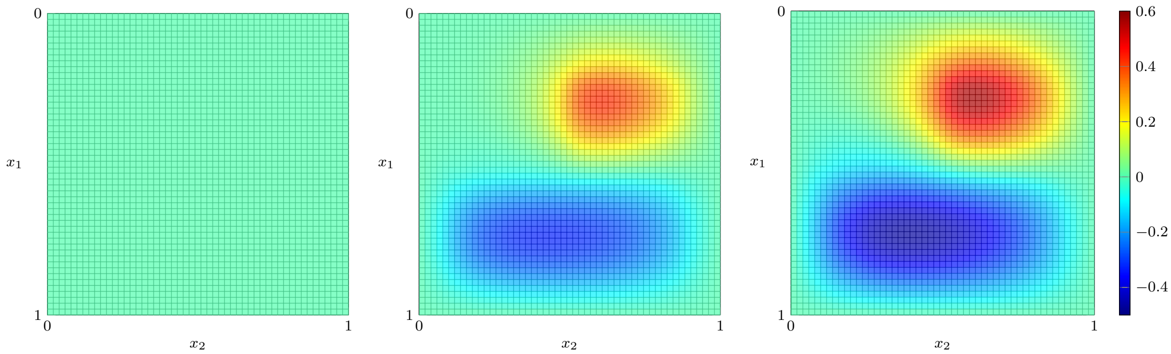
- $T = 3$,
- $y^0(x) \equiv 0$.

In this stage, we are going to use the optimal control $\bar{u}^\star_{\sqrt{2}}$ as a time-independent control for system (40). More precisely, we take

$$u(x,t) = \bar{u}^\star_{\sqrt{2}}(x) \quad \text{for } t \in (0,T). \tag{41}$$

We recall that this control is shown in Figure 2. Now, by plugging such control into equation (40), we obtain the controlled solution displayed in Figure 8. We see that this control steers y away from the initial condition at time $t = 0$ and reaches in a short amount of time a region near the optimal steady state $\bar{y}^\star_{\sqrt{2}}$ (cf. Figure 8c).

For this experiment, the efficiency of the time-independent control is closely related to the fact that $c(x) \geq 0$. In particular, the conditions on the coefficients a_ν and c allow to prove that the uncontrolled system (40) is exponentially stable regardless of the initial datum.



(a) $t = 0$ (b) $t = 0.025$ (c) $t \in (0.1, 3)$

Fig. 8: Evolution in time for a stable system controlled with the steady control approximated by greedy procedure.

In Figure 9, we illustrate the efficiency of the steady control by plotting different curves representing the time evolution of the L^2-norm of y solution to (40) with different initial datums and taking (41) as a control. To compare, we have computed the time-dependent solution $y^T(t)$ associated to the optimal control $u^T(t)$, obtained by minimizing (39), for the given parameter $\nu = \sqrt{2}$ and $y_0(x) = 0$. Note that we have split the time horizon in two different intervals for the sake of clarity.

The solid line corresponds to the trajectory obtained by using the turnpike control $u^T(t)$ and exhibits the prototypical behavior expected from the turnpike theory. For most of the time, except at the beginning and end, the solution remains exponentially close to corresponding steady state $\bar{y}$.

On the other hand, the discontinuous lines represent the trajectories of the system controlled with the steady control (41) and differential initial datums. We see that this control drives the solution y to the same region (up to an approximation error inherited from the greedy procedure) and maintains the solution there during the whole time interval independently of the initial data. For more that 95% percent of the time, both solutions are almost indistinguishable (up to the end of the interval, where they are superimposed).

To some extent, the turnpike property, together with the greedy procedure for elliptic optimal control problems, gives a partial answer on how to propose a greedy methodology that is independent of the initial datum for the system under consideration (cf. [28]).

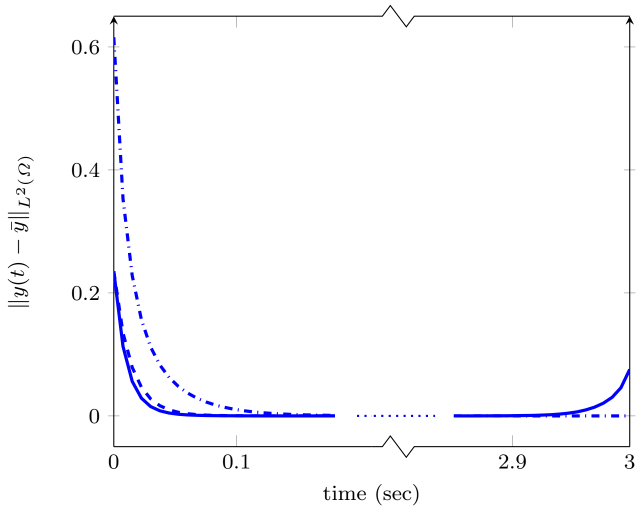


Fig. 9: L^2-norm of the difference between the solution to system (40) and the steady state for different controls: turnpike control (——), steady control with $y_0(x) = 0$ (- - -) and $\tilde{y}_0(x) = \chi_{(0.4,0.9)^2}(x)$ (-·-·-).

5.2 The case $c(x) < 0$

In spite of the above conclusions, it is important to mention that the same behavior cannot be expected for all coefficients $c(x) < 0$. In fact, if $c(x) < -\lambda_1$ where λ_1 if the first eigenvalue of the operator $-\mathrm{div}\,(a_\nu \nabla \cdot)$, then system (40) is shown to be unstable.

As an example, let us take the parameters and results in the greedy test #2 shown in Section 4.2 together with $y_0(x) = 0$ and $T = 3$. As before, we can put the (approximated) steady control $\bar{u}^\star_{-5\pi^2/2}(x)$ in system (40) and test its performance. Recall that we have chosen $c_\nu \equiv -5\pi^2/2$, which clearly satisfies $c_\nu < -\lambda_1$. Thus, in this case, the uncontrolled evolution equation is shown to be unstable.

We show in Figure 10 the solution to the evolution problem using the steady control $\bar{u}^\star_{-5\pi^2/2}$ for $t \in [0, 0, 3]$. We see that in this case, the steady control lacks to stabilize the system around the steady state shown in figure 7a and, moreover, continues growing exponentially during the remainder of the time interval.

In Figure 11, we present further experiments where we illustrate that for some initial data the controlled solution y grows in exponential manner. In fact, only for $y_0(x) = \bar{y}^\star_{5\pi^2/2}$ and a sufficiently small neighborhood of this initial datum, the steady control is effective to control the underlying system.

Such behaviour could suggest the lack of the turnpike property for unstable systems. However, the proof of Theorem 1 relies on the controllability of (40) and not of its the stability. Since (40) is known to be exactly controllable to trajectories regardless the sign of $c(x)$ (see, e.g. [18]), then Theorem 1 should hold.

Actually, for this particular case, to see the turnpike property it is necessary to compute the solution to the time-dependent minimization problem (40). We show in Figure 12, the time evolution of the optimal controlled state and control and, as expected, we can see the asymptotic simplification towards the state and control for most of the time horizon. However, we can also see that the control makes a large effort at the beginning of the temporal interval to move the system to this steady state.

Thus we conclude that unlike the case $c(x) \geq 0$, the steady turnpike controls, and in particular the approximated controls derived from the greedy approach, are in general not enough to control an unstable evolution system (40).

One possible solution to this issue is to exploit the so-called controllability to trajectories of (40) (see, for instance, [16],[17, Remark 6.1]). More precisely, we can take the steady solution $\bar{y}$ to (7) and define $z = y - \bar{y}$. Since $\bar{y}$ is in particular a reachable state for the solutions to the evolution problem, one can construct, for any given $T_0 > 0$, a null control v_0 steering the system such that z verifies $z(T_0) = 0$. This would imply that $y(T_0) = \bar{y}$.

Then, once we are near the steady state, one can switch to the steady control to remain there. In this way, the problem of computing an optimal control in a large time horizon can be reduced to an exact controllability to trajectories problem plus a steady optimal control one. The same idea can be applied if instead of the exact values of $(\bar{u}, \bar{y})$ we have an approximation by greedy methods.

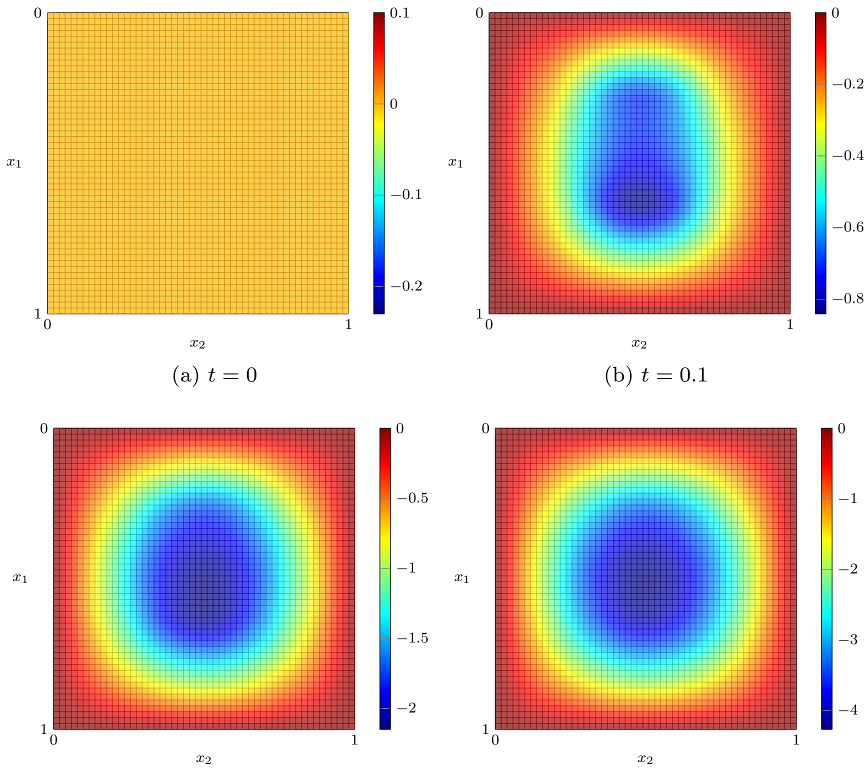


Fig. 10: Evolution in time for an unstable system controlled with the steady control approximated by means of the greedy approach.

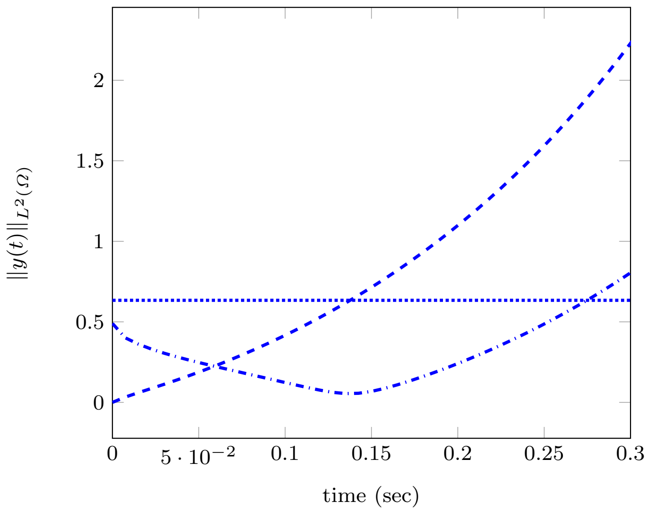
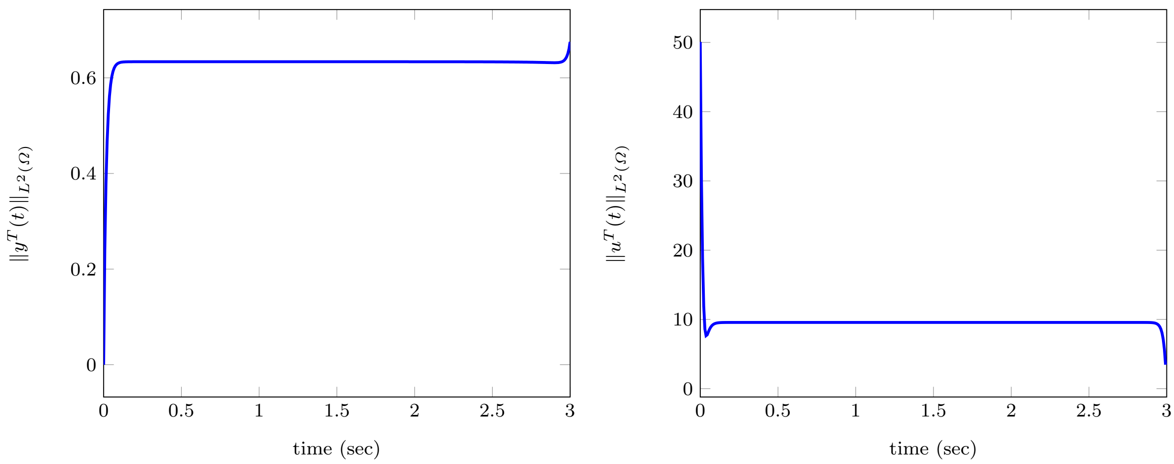


Fig. 11: Time evolution comparison for an unstable system controlled with steady control with different initial data: $y_0(x) = \bar{y}^{\star}_{-5\pi^2/2}(x)$ (dotted), $y_0(x) = 0$ (dashed) and $\tilde{y}_0(x) = \chi_{(0.4,0.9)^2}(x)$ (dash-dotted).

Note however that in this strategy, T_0 is a new design parameter. In fact, recall that null controllability for heat and parabolic problems hold for any time and that the cost of controllability can be analyzed precisely in terms of T_0 (see, for instance, [18]). The optimal way to choose T_0 such that this strategy emulates as close as posible the turnpike property is an open and interesting problem.



(a) The optimal state

(b) The optimal control

Fig. 12: Turnpike property for an unstable system.

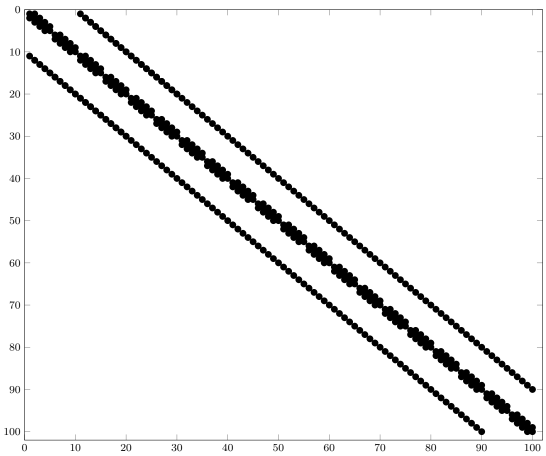


Fig. 13: Sparsity pattern of matrix $\mathcal{A}_h$ with $n = 10$. The total number of nonzero elements is 460 (out of 10000).

6 Computational cost

In this section, we make a more detailed analysis on the computational cost of the greedy algorithm versus the *naive approach*. We rely on classic numerical linear algebra results (see, for instance, [5,32]). The analysis is restricted to 2-D problems for which numerical results are presented.

During numerical implementation of Algorithm 2, two main basic problems arise systematically: matrix-vector multiplication and solving the set of linear equations

$$Ax = b, \quad A \in \mathbb{R}^{M \times M}. \tag{42}$$

For a given matrix A and any vector x, the cost of computing the product between them is $2M^2$ operations and standard methods for solving the linear system (42) have a computational cost that grows as $\mathcal{O}(M^3)$. As we mentioned before, the finite difference discretization of the operator $\mathcal{A} = -\mathrm{div}\,(a_\nu \nabla \cdot)$ on a $N \times N$ grid can be written as a system of $\mathcal{O}(N^2)$ equations. Arranging them in the form (42) yields a coefficient matrix with $\mathcal{O}(N^2 \times N^2)$ entries. In particular, since $M = N^2$ for our case, even for a small number N of grid points the standard methods become unsuitable.

By arranging equations (37) in a *natural* row order, one can see that the non zero entries of the resulting matrix are distributed along 5 diagonals of the matrix. In figure 13, we see an example of the distribution pattern for the nonzero entries of $\mathcal{A}_h$ for $N = 10$ grid points.

These kind of matrices are called sparse matrices and have the feature that their number of nonzero entries is of order N instead of N^2.

The matrix-vector multiplication for sparse matrices can be carried out in $2s$ operations where s is the number of nonzero entries of the matrix. For this particular problem, matrix $\mathcal{A}_h$ has a maximum number of $5N^2$ nonzero entries, thus the computational cost is $10N^2$.

Moreover, for our problem, $\mathcal{A}_h$ is an special case of sparse matrices with a defined structure called *banded* matrices. The bandwidth μ is the number such that $a_{i,j} = 0$ if $|i-j| > \mu$. It can be readily seen that for the general case, the FD matrix associated for a given problem, the bandwidth of $\mathcal{A}_h$ is $\mu = N$. This special structure will allow us to make a precise (but conservative) estimation of the computational cost for the greedy procedure presented here.

6.1 Cost of the offline part

The offline part of the greedy algorithm consists of two main ingredients. On one hand, the search for distinguised parameter values ν_j by examining the residual (15) over the set $\tilde{\mathcal{K}}$ and, on the other, the computation of the corresponding snapshots $(\bar{p}_{\nu_j}, \bar{y}_{\nu_j})$. We will estimate the computational cost by differentiating three main steps.

Step 1. Choosing ν_1. The first parameter is distinguished by maximizing over all of the possible targets y^d_ν. To do this, one has to the compute $k = \mathrm{card}(\tilde{\mathcal{K}})$ times the Euclidian norm in $\mathbb{R}^{N^2}$ and look for the maximum value. The implementation cost is

$$2kN^2.$$

Step 2. Computing $(\bar{p}_{\nu_1}, \bar{y}_{\nu_1})$. In order to determine the first snapshot one has to solve the minimization problem (12) with the selected ν_1 in the previous step. As we mentioned before, we are using a standard CG method to compute the optimal control.

The dominating computations during an iteration of the CG are matrix-vector products and the computation of the operator S^*S (see eq. (13)). The computation of S^*Sv for any $v \in L^2(\omega)$ amounts first to solve the system of cascade equations

$$\begin{cases} \mathcal{A}y + c\,y = \chi_\omega v, & \text{in } \Omega \\ \mathcal{A}^*p + c\,p = \beta(y - y^d), & \text{in } \Omega \\ y = p = 0 & \text{on } \partial\Omega. \end{cases}$$

and then we can take $S^*Sv = p|_\omega$.

Since A is self-adjoint both equations can be solved at the numerical level by taking $L_h = \mathcal{A}_h + cI_h$ and then finding for y and p. Since this has to be done during many iterations of the CG, we can exploit the nature of the matrix L_h (observe that the term cI only contributes to the diagonal and the sparsity pattern of A_h is preserved) and take the LU factorization of L_h once at the beginning of the process.

For the matrix L_h, which has a bandwidth $\mu = N$, the LU factorization gives matrices L and U with the same lower and upper bandwidth, respectively (see [32, Property 3.4]). The computational cost is

$$2N^4\,.$$

Once the LU factors are known, we can solve for either y or p by forward-backward substitution in $4N^2\mu = 4N^3$ operations.

Finally, the maximum number of iterations of the CG to achieve a given tolerance δ (see, for instance, [34]) for this particular case is

$$\ell \le \left\lceil \frac{1}{2}\sqrt{\kappa}\ln\left(\frac{2}{\delta}\right)\right\rceil$$

where κ is the condition number of S^*S.

Gathering all together, we see that the cost for computing an optimal control $\bar{u}$ for a given parameter ν is

$$C_{\mathrm{opt}} = 2N^4 + \ell\left(4N^3 + \mathcal{O}(N^2)\right) \tag{43}$$

where the last term involves all related matrix-vector operations of one iteration of the CG.

Remark 6 The cost provided here is rather conservative and far from being optimal. Observe that, even if A_h has a banded pattern, the band itself is sparse (see Figure 13). LU factorization in this case gives matrices L and U which are filled in the band and have more nonzero elements than A_h itself. Different order schemes or matrix compacting techniques can lead to a better performance than computing the corresponding LU matrices. In general, the computational cost for solving large sparse system depends on a complicated way of the $\mathcal{A}_h$, the number of nonzero elements, its sparsity pattern and the particular algorithm used. We refer the interested reader to [19] for a more detailed discussion.

Step 3. Choosing ν_{j+1}. Suppose that we have determined the first j snapshots and have constructed the approximating space $(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)$. The basis of the approximating space has been determined gradually through the previous iterations up to the last vector $(\bar{p}_{\nu_j}, \bar{y}_{\nu_j})$. As explained in the previous step, each vector requires C_{opt} operations.

According to Algorithm 2, the next parameter value is chosen by computing

$$|| \inf_{(p,y)\in(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)} R_{\tilde{\nu}}(p,y)||_{H^{-1}(\Omega)} \tag{44}$$

For any given parameter $\tilde{\nu}$, equations (27) and (28) enable the practical implementation of this infimum. First, the residual $G_{\tilde{\nu}}$ can be build just by vector and scalar multiplication of sparse matrices, see (27). The computational cost of this part can be estimated as $45N^2$.

As stated in Remark 2, the infimum appearing in (44) can be computed by projecting the vector $\begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}$ to the space $G_{\tilde{\nu}}(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)$. Indeed, one can compute by means of the Gram-Schmidt procedure an orthonormal basis for the space $(\overline{\mathcal{P}}_j, \overline{\mathcal{Y}}_j)$ and then finding the projection just by vector multiplications. To improve the efficiency, this basis is orthonormalized iteratively just by adding the last computed vector to the already orthonormal base. This has a cost of $4N^2j$. Finding the actual projection can be carried out by merely matrix-vector multiplication, which also has a total cost of $4N^2j$.

Adding up, the total cost of this part of the algorithm equals to

$$(k-j)(45N^2 + 8N^2j)$$

Remark 7 Observe that at the numerical level, the residual R_ν can be implemented at the very low cost of $\mathcal{O}(N^2)$ operations, since it is only composed by matrix-vector multiplications. In other works (see, e.g., [28]), where the surrogate is obtained by solving coupled forward-backward systems, the cost is dominated by LU-factorization which (for banded sparse matrices) has a cost of $2N^4$ (see *Step 2* above). The idea of obtaining cheaper surrogates is, of course, not new and has been successfully explored within the framework of optimal control for elliptic and parabolic equations, see, for instance, [20, 24, 25]. The difference between these works and the approach presented in this mansucript is in the way the greedy procedure is implemented. In the cited papers the authors employ Galerkin projections in order to determine the next snapshot, while we use orthogonal projection to the constructed space. However, the cost of both approaches is similar.

Step 4. Calculating $(\bar{p}_{j+1}, \bar{y}_{j+1})$. To compute the new snapshot it suffices to solve the optimal control problem for ν_{j+1}. As before, the cost is given in (43).

Total cost

Summing the above cost for $j = 1, \ldots, n$, the tocal cost of the algorithm results in

$$C_{total} = 2kN^2 + \frac{45}{2}N^2 n\,(2k-n-1) + \frac{4}{3}N^2 n(n+1)\,(3k-2n-1)) + nC_{\text{opt}}$$

As the cost C_{opt} of solving the optimal control problem is $\mathcal{O}(N^4)$, this part contributes with most of the computational cost of the greedy algorithm. As the number of chosen snapshots n approaches the number of eligible parameters $k = \text{card}(\tilde{\mathcal{K}})$, this part converges, up to lower order terms, to the cost of computing $\bar{u}_\nu$ for all the possible values k. This proves that the application of the greedy control algorithm is always cheaper than a naive approach that consists of computing controls for all values of the parameter from a very refined uniform mesh on $\mathcal{K}$.

6.2 Cost of the online part

Here, we will estimate the computational cost of the online part of the greedy procedure, which is described in Algorithm 3. Observe that only the first four steps of this algorithm contribute to the total cost of computing the approximated control $\bar{u}^\star$ with the greedy procedure.

- *Step 1.* In the first step, we have to project the vector

$$\begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix} \text{ to } G_\nu(\overline{\mathcal{P}}_n, \overline{\mathcal{Y}}_n)$$

  where $G_\nu(\overline{\mathcal{P}}_n, \overline{\mathcal{Y}}_n)$ denotes the space spanned by $G_\nu(\overline{p}_i, \overline{y}_i)$, $i = 1, \ldots, n$ and $\{(\overline{p}_i, \overline{y}_i), i = 1, \ldots, n\}$ are the snapshots selected in the offline part. To this effect we first build the linear part of the residual, G_ν, for each pair $(\bar{p}_i, \bar{y}_i)$, $i = 1, \ldots, n$, with a computational cost of $45N^2$ for each one. Then we orthonormalize the set $\{G_\nu(\overline{p}_i, \overline{y}_i), i = 1, \ldots, n\}$ with the Gram-Schmidt procedure, which can be achieved in $2N^2n^2$ operations. Finally, the projection can be carried out by by adding and multiplying vectors in this new orthonormal basis with a total cost of $4N^2n$.
- *Step 2.* Denoting the projection by $P_n \begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix}$, we solve for the system

$$P_n \begin{pmatrix} 0 \\ -\beta y^d \end{pmatrix} = \sum \alpha_i G_\nu(\overline{p}_i, \overline{y}_i).$$

  Solving this system with QR decomposition has a cots of $8N^2n^2 - \frac{16}{3}n^3$ operations.
- *Step 3.* This step can be achieved by scalar-vector multiplications with a total cost of N^2n.
- *Step 4.* The computational cost of this part is negligible with respect to the rest of the algorithm.

Adding up these contributions, the total cost for obtaining an approximative control $\bar{u}_\nu^\star$ for some parameter value equals to

$$10N^2n^2 + 50N^2n - \frac{16}{3}n^3$$

Recall that the the computational cost of obtaining the optimal control by the conjugate gradient method for some parameter ν is C_{opt} (see eq. (43)) is of order N^4. The cost for computing the $\bar{u}_\nu^\star$ only depends on n and, in particular, since no iterations are required the total cost of obtaining an approximation of the control by the greedy procedure is less than the cost of obtaining it, for instance, by the CG method.

Finally, *Step 5* of the online algorithm amounts to solve the corresponding system with control $\bar{u}_\nu^\star$ and, as mentioned before, we can solve it by computing the LU factors and then by forward-backward subsstitution in $2N^4 + 4N^3$ operations.

7 Additional comments and open problems

We devote this section to present some concluding remarks and state some open problems regarding the turnpike property as well as the greedy approach.

1. *Robust control of the evolution equation.* The application of the turnpike property and the reduction of the optimal control problem for parabolic equations to the steady state one allows us to employ greedy procedure and construct an approximating space for parameterized controls which is independent of the initial datum. This allows for a robust approach which is not datum sensitive. Thus we provide partial answers to the questions arising in [28] in the context of finite-dimensional systems, as well as in [9], where greedy approach for elliptic problems is explored in a robust manner independently of the given source terms.
2. *Stable vs unstable.* From the results presented in [31], it is clear that we can expect the asymptotic simplification of the time-dependent controls towards their steady version in time large. However, from the practical point of view, two main cases has to be differentiated. As discussed in Section 5, the possibility to use the steady control in the evolution problem largely depends on its stability. If the system is stable, one can effectively use the time independent control to steer the system near the turnpike and remain there indefinitely. When the system is unstable, one can use the approach discussed in Section 5.2 to first control the system to the steady state and then use the steady control to remain there. In this way, the original problem of computing an optimal control in a large time horizon is reduced to an exact controllability problem in a small time interval plus a steady optimal control one.

3. *On hypothesis* **H3**. In Section 2, we assumed that the mapping L_ν is uniformly elliptic with respect to ν. hold for all $\nu \in \mathcal{K}$. This condition allowed us to define the residual R_ν (and also to employ systematically (5)), which in turn is at the heart of the proof of Theorem 3 and the greedy procedure described in Algorithm 2.
   In the general case, when hypothesis **H3** is not fulfilled, there might be some $\nu \in \mathcal{K}$ such that $\ker(L_\nu) = \{0\}$. For that parameter value the solutions to system (4) are not uniquely determined and can be written as
   $$y = z + e \quad \text{with} \quad z \in \ker(L_\nu)^\perp, \quad e \in \ker(L_\nu). \tag{45}$$
   and the minimization problem (7) has to be replaced by the more general one
   $$\min \left\{ J(u) = \frac{1}{2}|u|^2_{L^2(\omega)} + \frac{\beta}{2}\|y - y_d\|^2_{L^2(\Omega)}, \quad u \in L^2(\omega), \quad y \in H^1_0(\Omega) \text{ verifying (45)} \right\}. \tag{46}$$
   In this case, the functional J appearing in (46) is again strictly convex, continuous and coercive and therefore the minimization problem is well posed and the optimal control $\bar{u}$ and the optimal state $\bar{y}$ can be identified. However, from the definition of the residual (15), we see that the operator L_ν is involved, and if **H3** is not fulfiled the residual is not longer a reliable way to measure the distance between two possiblly unknown controls. Thus for the general case in which **H3** is not satisfied a more careful analysis is needed and it remains as an open problem.
4. *Wave equations.* When the free dynamics of the system under consideration enjoys some stability property, it is expected for the turnpike property to hold. For wave-like models, where solutions of the free dynamics are of oscillatory nature and do not enjoy the property of asymptotic simplification, is less clear. However, as shown in [31, 39], the turnpike property still holds for wave-like models under suitable controllability assumptions. More precisely, except for an initial time layer $[0, \tau]$ and a final one $[0, T - \tau]$ during the rest of the time interval the solutions are exponentially close to the steady state ones. Therefore, the greedy methodology and the robust control approach discussed above applies for such models without major changes. In the same line as the unstable parabolic case, it remains to show what is the best strategy to control the system in the initial and final layer.
5. *Finite element methods.* The numerical simulations presented here were implemented by using finite differences methods. As usual, these methods are suitable for analyzing examples posed in simple cartesian geometries with regular enough meshes and are known to be easily programmable. One can use finite element methods to study more complex problems arising in applications and the same results presented in this work are expected. In particular, for problems where the cost of computing one solution to the optimal control problem is very large, the greedy approach and, in general, reduced bases techniques, are reliable tools to speed up the simulations, allowing to explore different parameter configuration within a reasonable amount of time.
6. *Greedy controllability for parabolic equations.* In this paper we generalize the greedy approach for controllability problems presented in [28] for ODEs to the PDE setting. Nevertheless, to quantify the convergence rate of greedy algorithms, the analytic dependence of the controls with respect to the parameters plays a key role. For heat and parabolic equations, when the unknown parameters enter in the principal part of the diffusion operator, the analytic dependence can be expected (this is not the case for wave equations, see, e.g. [1]). However, the functional setting in which the controls are analytically dependent of the parameters is not clear. Moreover, classical observability inequalities fail to provide good upper and lower bounds for a residual (surrogate) that allows to measure the distance between to possible unknown controls.
   By exploring the turnpike property, we can bypass the difficulties of analytical dependence for the controls of parabolic problems and reduce the problem to the elliptic setting where we have an appropriate functional setting (cf. subsection 3.4), which provides optimal approximation rates in terms on Kolmogorov widths. Furthermore, unlike other papers in which the affine parameter dependence is required for efficient implementation of a greedy sampling procedure ([11, 20, 24, 25]), here no special dependence of the coefficients with respect to the parameter is required.

References

1. G. Allaire and A. Kelly. Optimal design of low-contrast two-phase structures for the wave equation. *Mathematical Models and Methods in Applied Sciences*, **21**, 07 (2011), 1499–1538.
2. J. A. Atwell and B. B. King. Proper orthogonal decomposition for reduced basis feedback controllers for parabolic equations. *Mathematical and computer modelling*, **33**, 1-3 (2001), 1–19.

3. E. Bader, M. Kärcher, M. A. Grepl and K. Veroy. Certified reduced basis methods for parametrized distributed elliptic optimal control problems with control constraints. *SIAM Journal on Scientific Computing*, **38**, 6 (2016), A3921–A3946.
4. P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, **43**, 3 (2011), 1457–1472.
5. A. Boyd and L. Vandenberghe. Convex optimization. *Cambridge university press*, (2004).
6. H. Brezis. Functional Analysis, Sobolev Spaces and Partial Differential Equations. *Universitext, Springer-Verlag*, (2011).
7. A. Buffa, Y. Maday, A. T. Patera, C. Prudhomme and G. Turinici. A priori convergence of the greedy algorithm for the parameterized reduced basis. *Mathematical modelling and numerical analysis.* **46** (2012), 595-603.
8. D. A. Carlson, A. Haurie and A. Jabrane. Existence of overtaking solutions to infinite dimensional control problems on unbounded time intervals. *SIAM J. Control Optim.,* **25** (1987), 1517–1541.
9. M. Choulli , E. Zuazua Lipschitz dependence of the coefficients on the resolvent and greedy approximation for scalar elliptic problems *C. R. Acad. Sci. Paris, Ser I* **354** (2016), 1174-1187
10. A. Cohen and R. DeVore. Kolmogorov widths under holomorphic mappings. *IMA Journal of Numerical Analysis*, **36**, 1 (2015), 1–12.
11. A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, **24**, (2015), 1–159.
12. L. Dedè. Reduced basis method and a posteriori error estimation for parametrized linear-quadratic optimal control problems. *SIAM Journal on Scientific Computing*, **32**, 2 (2010), 997–1019.
13. R. DeVore, (2015). *The Theoretical Foundation of Reduced Basis Methods.* preprint.
14. R. DeVore, G. Petrova and P. Wojtaszczyk. Greedy Algorithms for Reduced Bases in Banach Spaces. *Constr. Approx.*, **37** (2013), 455–466.
15. L. C. Evans. Partial differential equations, volume 19 of Graduate Studies in Mathematics. *American Mathematical Society*, Providence, RI, second edition, (2010).
16. E. Fernández-Cara and E. Zuazua. Null and approximate controllability for weakly blowing up semilinear heat equations. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis,* **17**, 5 (2000), 583-616).
17. E. Fernández-Cara and E. Zuazua. The cost of approximate controllability for heat equations: the linear case. *Advances in Differential equations*, **5**, 4-6 (2000), 465–514
18. A. Fursikov and O. Yu. Imanuvilov. Controllability of evolution equations. *Lecture Notes, Research Institute of Mathematics*, Seoul National University, Korea, (1996).
19. A. George and J. W. Liu. Computer solution of large sparse positive definite. *Prentice Hall*, (1981).
20. M. A. Grepl and M. Kärcher. Reduced basis a posteriori error bounds for parametrized linear-quadratic elliptic optimal control problems. *Comptes rendus mathematique*, **349**, 15-16 (2011), 873–877.
21. R. Glowinski, J.-L. Lions and J. He. Exact and Approximate Controllability for Distributed Parameter Systems: A Numerical Approach. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2008.
22. K. Ito and S. Ravindran. A reduced basis method for control problems governed by PDEs, in *Control and Estimation of Distributed Parameter System*, W. Desch, F. Kappel, and K. Kunisch, eds., Internat. Ser. Numer. Math. 126, 1998, pp. 153–168.
23. K. Ito and S. S. Ravindran. Reduced basis method for optimal control of unsteady viscous flows. *International Journal of Computational Fluid Dynamics*, **15**, 2 (2001), 97–113.
24. M. Kärcher and M. A. Grepl. A certified reduced basis method for parametrized elliptic optimal control problems. *ESAIM: Control, Optimisation and Calculus of Variations*, **20**, 2 (2014), 416–441.
25. M. Kärcher and M. A. Grepl. A Posteriori Error Estimation for Reduced Order Solutions of Parametrized Parabolic Optimal Control Problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, **48**, 6 (2014), 1615–1638.
26. K. Kunisch and S. Volkwein. Control of the Burgers equation by a reduced-order approach using proper orthogonal decomposition. *Journal of optimization theory and applications*, **102**, 2 (1999), 345–371.
27. K. Kunisch, S. Volkwein and L. Xie. HJB-POD-based feedback design for the optimal control of evolution problems. *SIAM Journal on Applied Dynamical Systems*, **3**, 4 (2004), 701–722.
28. M. Lazar and E. Zuazua. Greedy controllability of finite dimensional linear systems. *Automatica*, **74** (2016), 327–340.
29. J.-L Lions. Optimal control of systems governed by partial differential equations. *Springer-Verlag*, 1971.
30. F. Negri, G. Rozza, A. Manzoni and A. Quarteroni. Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing*, **35**, 5 (2013), A2316–A2340.
31. A. Porretta and E. Zuazua. Long time versus steady state optimal control. *SIAM Journal on Control and Optimization*, **51**, 6 (2013), 4242–4273.
32. A. Quarteroni, R. Sacco, and F. Saleri. Numerical mathematics. Vol. 37. *Springer Science & Business Media*, (2010).
33. S. S. Ravindran. A reduced-order approach for optimal control of fluids using proper orthogonal decomposition. *Internat. J. Numer. Methods Fluids*, **34** (2000), 425–448.
34. J. R. Shewchuk. An introduction to the conjugate gradient method without the agonizing pain. *Manuscript*, (1994).
35. E. Trélat and E. Zuazua The turnpike property in finite-dimensional nonlinear optimal control. *J. Differ. Equ.,* **258** (2015), 81–114.
36. F. Tröltzsch and S. Volkwein. POD a-posteriori error estimates for linear-quadratic optimal control problems. *Computational Optimization and Applications*, **44**, 1 (2009), 83-115.
37. A. J. Zaslavski. Turnpike properties in the calculus of variations and optimal control. *Nonconvex. Optim. Appl. 80.* Springer-Verlag New York, 2006.
38. A. J. Zaslavski. Existence and structure of optimal solutions of infinite-dimensional control problems. *Appl. Math. Optim.,* **42** (2000), 291–313.
39. E. Zuazua. Large time control and turnpike properties for wave equations. *Annual Reviews in Control*, **44** (2017), 199–210.