

# Greedy controllability of finite dimensional linear systems \*

Martin Lazar<sup>a</sup> Enrique Zuazua<sup>b</sup>

<sup>a</sup>University of Dubrovnik, Čira Carića 4, 20 000 Dubrovnik, Croatia

<sup>b</sup>Departamento de Matemáticas, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

---

## Abstract

We analyse the problem of controllability for parameter dependent linear finite-dimensional systems. The goal is to identify the most distinguished realisations of those parameters so to better describe or approximate the whole range of controls. We adapt recent results on greedy and weak greedy algorithms for parameter dependent PDEs or, more generally, abstract equations in Banach spaces. Our results lead to optimal approximation procedures that, in particular, perform better than simply sampling the parameter-space to compute the controls for each of the parameter values. We apply these results for the approximate control of finite-difference approximations of the heat and the wave equation. The numerical experiments confirm the efficiency of the methods and show that the number of weak-greedy samplings that are required is particularly low when dealing with heat-like equations, because of the intrinsic dissipativity that the model introduces for high frequencies.

*Key words:* parametrised ODEs and PDEs, greedy control, weak-greedy, heat equation, wave equation, finite-differences.

---

## 1 Introduction and problem formulation

We analyse the problem of controllability for linear finite-dimensional systems submitted to parametrised perturbations, depending on unknown parameters in a deterministic manner.

In previous works we have analysed the property of averaged control looking for a control, independent of the values of these parameters, designed to perform well, in an averaged sense ([21], [13]).

Here we analyse the complementary issue of determining the most relevant values of the unknown parameters so

to provide the best possible approximation of the set of parameter dependent controls. Our analysis is based on previous work on reduced modelling and (weak) greedy algorithms for parameter dependent PDEs and abstract equations in Banach spaces ([4], [5]), which we adapt to the present context.

The problem is relevant in applications, as in practice the models under consideration are often not completely determined, submitted to unknown or uncertain parameters, either of deterministic or of stochastic nature. It is therefore essential to develop robust analytical and computational methods, not only allowing to control a given model, but also to deal with parameter-dependent families of systems in a stable and computationally efficient way.

Both reduced modelling and the control theory have experienced successful real-life implementations (we refer to the book [14] for a series of such interactions with the European industry). The merge of this two theories will allow of variety of applications in all fields involving problems modelled by parameter dependent systems (fluid dynamics, aeronautics, meteorology, economics, ...).

Although the greedy control is applicable to more general control problems and systems, here we concentrate on controllability issues and, to better illustrate

---

\* This paper has been partially developed while the first author was visiting the Basque Center for Applied Mathematics (Bilbao, Spain) within the MTM2011-29306 Grant of the MINECO. The first author was also supported by Croatian Science Foundation under the project WeConMApp/HRZZ-9780, as well as by University of Dubrovnik through the grant 11-18-10-2 and through the Erasmus+ programme.

The second author was partially supported by the Advanced Grant NUMERIWAVES/FP7-246775 of the European Research Council Executive Agency, the FA9550-15-1-0027 of AFOSR, the MTM2011-29306 and MTM2014-52347 Grants of the MINECO, and a Humboldt Award at the University of Erlangen-Nürnberg.

Email addresses: [martin.lazar@unidu.hr](mailto:martin.lazar@unidu.hr), [enrique.zuazua@uam.es](mailto:enrique.zuazua@uam.es) (Enrique Zuazua).

the main ideas of the new approach, we focus on linear finite-dimensional systems of parameter dependent ODEs. Infinite-dimensional systems, as a first attempt to later consider PDE models, are discussed separately in Section 6, as well as in the Conclusion section.

Consider the finite dimensional linear control system

$$\begin{cases} \frac{d}{dt}x(t, \nu) = \mathbf{A}(\nu)x(t, \nu) + \mathbf{B}(\nu)u(t, \nu), & 0 < t < T, \\ x(0) = x^0. \end{cases} \quad (1)$$

In (1) the (column) vector valued function  $x(t, \nu) = (x_1(t, \nu), \dots, x_N(t, \nu))^T \in \mathbf{R}^N$  is the state of the system at time  $t$  governed by dynamics determined by the parameter  $\nu \in \mathcal{N} \subset \mathbf{R}^d$ ,  $d \geq 1$ ,  $\mathcal{N}$  being a compact set,  $\mathbf{A}(\nu)$  is a  $N \times N$ -matrix governing its free dynamics and  $u = u(t, \nu)$  is a  $M$ -component control vector in  $\mathbf{R}^M$ ,  $M \leq N$ , entering and acting on the system through the control operator  $\mathbf{B}(\nu)$ , a  $N \times M$  parameter dependent matrix. In the sequel, to simplify the notation,  $d/dt$  will be simply denoted by  $'$ .

The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are assumed to be Lipschitz continuous with respect to the parameter  $\nu$ . However, some of our analytical results (Section 6) will additionally require analytic dependence conditions on  $\nu$ .

Here, to simplify the presentation, we have assumed the initial datum  $x^0 \in \mathbf{R}^N$  to be controlled, to be independent of the parameter  $\nu$ . Despite of this, the matrices  $\mathbf{A}$  and  $\mathbf{B}$  being  $\nu$ -dependent, both the control and the solution will depend on  $\nu$ . Similar arguments allow to handle the case when  $x^0$  also depends on the parameter  $\nu$ , which will be discussed separately.

We address the controllability of this system whose initial datum  $x^0$  is given, known and fully determined. We assume that the system under consideration is controllable for all values of  $\nu$ . This can be ensured to hold, for instance, assuming that the controllability condition is satisfied for some specific realisation  $\nu_0$  and that the variations of  $\mathbf{A}(\nu)$  and  $\mathbf{B}(\nu)$  with respect to  $\nu$  are small enough.

In these circumstances, for each value of  $\nu$  there is a control of minimal  $[L^2(0, T)]^M$ -norm,  $u(t, \nu)$ . This defines a map,  $\nu \in \mathcal{N} \rightarrow [L^2(0, T)]^M$ , whose regularity is determined by that of the matrices entering in the system,  $\mathbf{A}(\nu)$  and  $\mathbf{B}(\nu)$ .

Here we are interested on the problem of determining the optimal selection of a finite number of realisations of the parameter  $\nu$  so that all controls, for all possible values of  $\nu$ , are optimally approximated.

More precisely, the problem can be formulated as follows.

**Problem 1** Given a control time  $T > 0$  and arbitrary initial data  $x^0$  and final target  $x^1 \in \mathbf{R}^N$ , we consider the set of controls of minimal  $[L^2(0, T)]^M$ -norm,  $u(t, \mathcal{N})$ , corresponding to all possible values  $\nu \in \mathcal{N}$  of the parameter satisfying the controllability condition:

$$x(T, \nu) = x^1. \quad (2)$$

This set of controls is compact in  $[L^2(0, T)]^M$ .

Given  $\varepsilon > 0$  we aim at determining a family of parameters  $\nu_1, \dots, \nu_n$  in  $\mathcal{N}$ , whose cardinal  $n$  depends on  $\varepsilon$ , so that the corresponding controls, denoted by  $u_1, \dots, u_n$ , are such that for every  $\nu \in \mathcal{N}$  there exists  $u_\nu^* \in \text{span}\{u_1, \dots, u_n\}$  steering the system (1) in time  $T$  within the  $\varepsilon$  distance from the target  $x^1$ , i. e. such that

$$\|x(T, \nu) - x^1\| \leq \varepsilon. \quad (3)$$

Here and in the sequel, in order to simplify the notation, we denote by  $u_\nu$  the control  $u(t, \nu)$ , and similarly we use the simplified notation  $\mathbf{A}_\nu, \mathbf{B}_\nu, x_\nu$ .

Note that, in practice, the controllability condition (2) is relaxed to the approximate one (3). This is so since, in practical applications, when performing numerical approximations, one is interested in achieving the targets within a given error. This fact is also intrinsic to the methods we employ and develop in this paper, and that can only yield optimal procedures to compute approximations of the exact control, which turn out to be approximate controls in the sense of (3).

This problem is motivated by the practical issue of avoiding the construction of a control function  $u_\nu$  for each new parameter value  $\nu$  which, for large systems, although theoretically feasible by the uniform controllability assumption, would be computationally expensive. By the contrary, the methods we develop try to exploit the advantages that a suitable choice of the most representative values of  $\nu$  provides when computing rapidly the approximation of the control for any other value of  $\nu$ , ensuring that the system is steered to the target within the given error (3).

Of course, the compactness of the parameter set  $\mathcal{N}$  and the Lipschitz-dependence assumption with respect to  $\nu$  make the goal to be feasible. It would suffice, for instance, to apply a *naive* approach, by taking a fine enough uniform mesh on  $\mathcal{N}$  to achieve the goal. However, our aim is to minimise the number of spanning controls  $n$  and to derive the most efficient approximation. The *naive* approach is not suitable in this respect.

To achieve this goal we adapt to the present frame of finite-dimensional control, the theory developed in recent years based on greedy and weak-greedy algorithms for parameter dependent PDEs or abstract equations in Banach spaces, which optimise the dimension of the approximating space, as well as the number of steps required for its construction.

The rest of this paper is organised as follows. In Section 2 we summarise the needed controllability results for finite-dimensional systems and reformulate Problem 1 in terms of the corresponding Gramian operator. Section 3 is devoted to the review of (weak) greedy algorithms, while their application to the control problem under consideration and its solution is provided in the subsequent section.

The computational cost of the greedy control approach is analysed in Section 5. Section 6 contains a generalisation of the approach to infinite dimensional problems followed by a convergence analysis of the greedy approximation errors with respect to the dimension of the approximating space. Section 7 contains numerical examples and experiments for finite-difference discretisations of 1-D wave and heat problems. The paper is closed pointing towards future development lines of the greedy control approach.

## 2 Preliminaries on finite dimensional control systems. Problem reformulation

In order to develop the analysis in this paper it is necessary to derive a convenient characterisation of the control of minimal norm  $u_\nu$ , as a function of the parameter  $\nu$ . This can be done in a straightforward manner in terms of the Gramian operator. In this section we briefly summarise the most basic material on finite-dimensional systems that will be used along this article (we refer to [15,20] for more details).

Consider the finite-dimensional system of dimension  $N$ :

$$x' = \mathbf{A}x + \mathbf{B}u, \quad 0 \leq t \leq T; \quad x(0) = x^0, \quad (4)$$

where  $x$  is the  $N$ -dimensional state and  $u$  is the  $M$ -dimensional control, with  $M \leq N$ .

This corresponds to a specific realisation of the system above for a given choice of the parameter  $\nu$ . We omit however the presence of  $\nu$  from the notation since we are now considering a generic linear finite-dimensional system.

Here  $\mathbf{A}$  is an  $N \times N$  matrix with constant real coefficients and  $\mathbf{B}$  is an  $N \times M$  matrix. The matrix  $\mathbf{A}$  determines the dynamics of the system and the matrix  $\mathbf{B}$  models the way  $M$  controls act on it.

In practice, it is desirable to control the  $N$  components of the system with a low number of controls, the best possible case being the one of scalar controls:  $M = 1$ .

Recall that system (4) is said to be *controllable* when every initial datum  $x^0 \in \mathbf{R}^N$  can be driven to any final datum  $x^1$  in  $\mathbf{R}^N$  in time  $T$ . This controllability property can be characterised by a necessary and sufficient condition, which is of purely algebraic nature, the so called

*Kalman condition:* System (4) is controllable if and only if

$$\text{rank}[\mathbf{B}, \mathbf{AB}, \dots, \mathbf{A}^{N-1}\mathbf{B}] = N. \quad (5)$$

When this rank condition is fulfilled the system is controllable for all  $T > 0$ .

There is a direct proof of this result which uses the representation of solutions of (4) by means of the variations of constants formula. But for our purpose it is more convenient to use the point of view based on the dual problem of observability of the adjoint system that we discuss now.

Consider the *adjoint system*

$$\begin{cases} -\varphi' = \mathbf{A}^*\varphi, & 0 \leq t \leq T, \\ \varphi(T) = \varphi^0. \end{cases} \quad (6)$$

**Proposition 2.1** *System (4) is controllable in time  $T$  if and only if the adjoint system (6) is observable in time  $T$ , i. e. if there exists a constant  $C > 0$  such that, for all solution  $\varphi$  of (6),*

$$|\varphi^0|^2 \leq C \int_0^T |\mathbf{B}^*\varphi|^2 dt. \quad (7)$$

*Both properties hold in all time  $T$  if and only if the Kalman rank condition (5) is satisfied.*

*Furthermore, the control of minimal  $[L^2(0, T)]^M$ -norm can be built as a minimiser of a quadratic functional  $J : \mathbf{R}^N \rightarrow \mathbf{R}$ :*

$$J(\varphi^0) = \frac{1}{2} \int_0^T |\mathbf{B}^*\varphi(t)|^2 dt - \langle x^1, \varphi^0 \rangle + \langle x^0, \varphi(0) \rangle. \quad (8)$$

*More precisely, if  $\tilde{\varphi}^0$  is a minimiser for  $J$ , then the control*

$$u = \mathbf{B}^*\tilde{\varphi}, \quad (9)$$

*where  $\tilde{\varphi}$  is the corresponding solution of (6), is such that the state  $x$  of (4) satisfies the control requirement  $x(T) = x^1$ .*

Note that the minimiser of  $J : \mathbf{R}^N \rightarrow \mathbf{R}$  exists since the functional  $J$  is continuous, quadratic and coercive, in view of the observability inequality.

This characterisation of controls ensuring controllability also yields explicit bounds on the controls. Indeed, since the functional  $J \leq 0$  at the minimiser, and in view of the observability inequality (7), it follows that

$$\|u\| \leq 2\sqrt{C}[|x^0|^2 + |x^1|^2]^{1/2},$$

$C$  being the same constant as in (7). Therefore, we see that the square root of the observability constant is, up to a multiplicative factor, the norm of the control map associating to the data the control of minimal norm.

Summarising, the control of minimal  $[L^2(0, T)]^M$ -norm is of the form  $u = \mathbf{B}^* \tilde{\varphi}$ , where  $\tilde{\varphi} = \exp(\mathbf{A}^*(T-t))\tilde{\varphi}^0$ ,  $\tilde{\varphi}^0$  being the minimiser of  $J$ .

Let  $\Lambda$  be the quadratic form, known as (controllability) Gramian, associated to the pair  $(\mathbf{A}, \mathbf{B})$ , i.e.

$$\langle \Lambda \varphi^0, \psi^0 \rangle = \int_0^T \langle \mathbf{B}^* \varphi, \mathbf{B}^* \psi \rangle dt,$$

with  $\varphi, \psi$  being solutions to (6) with the data  $\varphi^0$  and  $\psi^0$ , respectively. Then, under the rank condition, because of the observability inequality, this operator is coercive and symmetric and therefore invertible. Its corresponding matrix, which we denote the same, is given by the relation

$$\Lambda = \int_0^T e^{(T-t)\mathbf{A}} \mathbf{B} \mathbf{B}^* e^{(T-t)\mathbf{A}^*} dt. \quad (10)$$

The minimiser  $\tilde{\varphi}^0$  can be expressed as the solution to the linear system

$$\Lambda \tilde{\varphi}^0 = x^1 - e^{T\mathbf{A}} x^0.$$

Hereby, the left hand side, up to the free dynamics component, represents the solution of the control system (4) with the control given by (9). As the solution is steered to the target  $x^1$ , the last relation follows.

In our context the adjoint system depends also on the parameter  $\nu$ :

$$\begin{cases} -\varphi'_\nu(t) = \mathbf{A}_\nu^*(\nu) \varphi_\nu(t), & t \in (0, T), \\ \varphi_\nu(T) = \varphi^0. \end{cases} \quad (11)$$

We assume that the system under consideration is controllable for all values of  $\nu$ . This can be ensured to hold assuming the following uniform controllability condition

$$\Lambda_- \mathbf{I} \leq \Lambda_\nu \leq \Lambda_+ \mathbf{I}, \quad (12)$$

where  $\Lambda_\pm$  are positive constants, while  $\Lambda_\nu$  is the Gramian of the system determined by  $\mathbf{A}_\nu, \mathbf{B}_\nu$ .

As we mentioned above, this assumption is fulfilled, in particular, as soon as the system is controllable for some specific value of the parameter  $\nu = \nu_0$ ,  $\mathbf{A}$  and  $\mathbf{B}$  depend continuously on  $\nu$ , and  $\nu$  is close enough to  $\nu_0$ . But our discussion and presentation makes sense in the more general setting where (12) is fulfilled.

As we restrict the analysis to the set of minimal  $[L^2(0, T)]^M$ -norm controls, each such control can be uniquely determined by the relation

$$u_\nu = \mathbf{B}_\nu^* e^{(T-t)\mathbf{A}_\nu^*} \varphi_\nu^0, \quad (13)$$

where  $\varphi_\nu^0$  is the unique minimiser of a quadratic functional  $J_\nu : \mathbf{R}^N \rightarrow \mathbf{R}$  given by (8), with  $\mathbf{B}$  replaced by  $\mathbf{B}_\nu$ , and  $\varphi$  by  $\varphi_\nu$ , the solution to (11).

As explained above, the minimiser  $\varphi_\nu^0$  can be equiva-

lently determined as the solution to the system

$$\Lambda_\nu \varphi_\nu^0 = x^1 - e^{T\mathbf{A}_\nu} x^0, \quad (14)$$

where  $\Lambda_\nu$  is the Gramian associated to  $(\mathbf{A}_\nu, \mathbf{B}_\nu)$ .

According to the uniform controllability condition (12), this defines a mapping  $\varphi^0 : \nu \in \mathcal{N} \rightarrow \mathbf{R}^N$ , whose smoothness is transferred from the mappings of  $\mathbf{A}_\nu$  and  $\mathbf{B}_\nu$  at all levels: Lipschitz, analytic etc. Having assumed these maps are Lipschitz continuous and the parameter  $\nu \in \mathcal{N}$  varies on a compact set, the set of minimisers  $\varphi^0(\mathcal{N})$  constitutes also a compact set in  $\mathbf{R}^N$ .

By using the 1-1 correspondence between the controls  $u_\nu$  and the associated minimisers  $\varphi_\nu^0$ , the original problem formulation can be reduced to the following one.

**Problem 2** Given a control time  $T > 0$ , an arbitrary initial data  $x^0$ , final target  $x^1 \in \mathbf{R}^N$ , and  $\varepsilon > 0$ , and taking into account that the set of minimisers  $\varphi^0(\mathcal{N})$  corresponding to all possible values  $\nu \in \mathcal{N}$  of the parameter is compact in  $\mathbf{R}^N$ , we aim at determining a family of parameters  $\nu_1, \dots, \nu_n$  in  $\mathcal{N}$  so that the corresponding minimisers, that we denote by  $\varphi_1^0, \dots, \varphi_n^0$ , are such that for every  $\nu \in \mathcal{N}$  there exists  $\varphi_{\nu}^{0,*} \in \text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$  such that the control  $u_\nu^*$  given by (13), with  $\varphi_\nu^0$  replaced by  $\varphi_{\nu}^{0,*}$ , steers the system (1) to the state  $x_\nu^*(T)$  within the  $\varepsilon$  distance from the target  $x^1$ .

In this formulation the elements of the manifold we want to approximate are  $N$ -dimensional vectors (instead of  $[L^2(0, T)]^M$ -functions in the first formulation), uniquely determined as solutions to linear systems (14). This enables to adapt the (weak) greedy algorithms and reduced bases methods for parameter dependent problems, that we present in the following section.

The approximate controls we obtain in this manner do not really belong to the space  $\text{span}\{u_1, \dots, u_n\}$  spanned by controls associated to selected parameter values, since the selection is done at the level of  $\varphi_{\nu}^{0,*} \in \text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$ , the control being simply the natural one corresponding to the choice of  $\varphi_{\nu}^{0,*}$ .

### 3 Preliminaries on (weak) greedy algorithms

In this section we present a brief introduction and main results of the linear approximation theory of parametric problems based on the (weak) greedy algorithms, that we shall use in our application to controllability problems. A more exhaustive overview can be found in some recent papers, e.g. [6,8].

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  $n$ . By increasing  $n$  one improves the accuracy of the approximation.

Determining *offline* an approximation subspace within a given error normally implies a high computational effort. However, this calculation is performed only once, resulting in a good subspace from which one can easily and computationally cheaply construct *online* approximations to every vector from  $\mathcal{K}$ .

Vectors  $x_i, i = 1 \dots n$  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  $V_n \leq X$  that approximate the set  $\mathcal{K}$  in the best possible manner. The algorithm is structured as follows.

### Weak greedy algorithm

---

Fix a constant  $\gamma \in (0, 1]$ .

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

$$\|x_1\|_X \geq \gamma \max_{x \in \mathcal{K}} \|x\|_X. \quad (15)$$

At the general step, having found  $x_1 \dots x_n$ , denote  $V_n = \text{span}\{x_1, \dots, x_n\}$  and

$$\sigma_n(\mathcal{K}) := \max_{x \in \mathcal{K}} \text{dist}(x, V_n). \quad (16)$$

Choose the next element  $x_{n+1}$  such that

$$\text{dist}(x_{n+1}, V_n) \geq \gamma \sigma_n(\mathcal{K}). \quad (17)$$

The algorithm stops when  $\sigma_n(\mathcal{K})$  becomes less than the given tolerance  $\varepsilon$ .

---

The algorithm produces a finite dimensional space  $V_n$  that approximates the set  $\mathcal{K}$  within the given tolerance  $\varepsilon$ . The choice of a new element  $x_n$  in each step is not unique, neither is the sequence of approximation rates  $\sigma_n(\mathcal{K})$ . But every such a chosen sequence decays at the same rate, which under certain assumptions given below, is close to the optimal one. Thus the algorithm optimises the number of steps required in order to satisfy the given tolerance, as well as the dimension of the final space  $V_n$ .

The pure greedy algorithm corresponds to the case  $\gamma = 1$ . As we shall see below, 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 performing the (weak) greedy algorithm one has to chose the next element of the approximation space by exploring the distance (16) for all possible values  $x \in \mathcal{K}$ . Such approach is faced with two crucial obstacles:

- i) the set  $\mathcal{K}$  in general consists of infinitely many vectors.
- ii) 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 is bypassed by performing a search 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.

As to deal with the second one, instead of considering the exact distance appearing in (16), one uses some *surrogate*, 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 a 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 [3] in the Hilbert space setting, and subsequently improved and extended to the case of a general Banach space  $X$  in [9].

**Theorem 3.1 ([9], Corollary 3.3)** *For the weak greedy algorithm with constant  $\gamma$  in a Hilbert space  $X$  we have the following: If the compact set  $\mathcal{K}$  is such that, for some  $\alpha > 0$  and  $C_0 > 0$*

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

*then*

$$\sigma_n(\mathcal{K}) \leq C_1 n^{-\alpha}, \quad n \in \mathbb{N}, \quad (18)$$

*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, and the result will be used in the convergence analysis of our method in Section 6. A similar estimate also holds for exponential decays (cf. [9]). In addition, it is also remarkable that the constant  $\gamma$  effects the efficiency only up to a multiplicative constant, while leaving approximation rates unchanged.

## 4 Greedy control and the main result

In this section we solve Problem 2 implementing the (weak) greedy algorithm in the manifold  $\mathcal{K} = \varphi^0(\mathcal{N}) = \{\varphi_\nu^0, \nu \in \mathcal{N}\}$  consisting of minimisers determined by the relation (14). The *goal* is to choose  $n$  parameters such

that  $\Phi_n^0 = \text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$  approximates the whole manifold  $\varphi^0(\mathcal{N})$  within the given error  $\varepsilon$ .

To this effect, as already stated in the introduction, we assume that the matrices  $\mathbf{A}(\nu)$  and  $\mathbf{B}(\nu)$  are Lipschitz continuous with respect to the parameter. In turn, this implies that the mapping  $\nu \rightarrow \varphi_\nu^0$  possesses the same regularity as well, with the Lipschitz constant denoted by  $C_\varphi$ .

#### 4.1 Construction of an approximating space $\Phi_n^0$

The greedy selection of each new snapshot relies on the relation (16), which in this setting maximises the distance of elements of  $\varphi^0(\mathcal{N})$  from the space spanned by already chosen snapshots. Theoretically, this process requires that we solve (14) for each value of  $\nu$ . And this is exactly what we want to avoid. Actually, here we face the obstacle *ii)* from previous section, since one has to apply the greedy algorithm within a set whose elements are not given explicitly. The problem is managed by identifying an appropriate surrogate for the unknown distances  $\text{dist}(\varphi_\nu^0, \Phi_n^0)$ . To this effect note that

$$\begin{aligned} \text{dist}(\varphi_\nu^0, \Phi_n^0) &\sim \text{dist}(\Lambda_\nu \varphi_\nu^0, \Lambda_\nu \Phi_n^0) \\ &= \text{dist}(x^1 - e^{T\mathbf{A}_\nu} \mathbf{x}^0, \Lambda_\nu \Phi_n^0), \end{aligned} \quad (19)$$

where  $\Lambda_\nu \Phi_n^0 = \text{span}\{\Lambda_\nu \varphi_1^0, \dots, \Lambda_\nu \varphi_n^0\}$ , while  $\sim$  denotes equivalence of terms resulting from the uniform controllability assumption (12).

In such a way we replace the unknown  $\varphi_\nu^0$  by an easy computed term  $x^1 - e^{T\mathbf{A}_\nu} \mathbf{x}^0$ , combining the target  $x^1$  and the solution of the free dynamics at time  $T$ . As for the other term, note that  $\Lambda_\nu \varphi_i^0, i = 1 \dots n$  represents the value at time  $T$  of the solution to the system

$$\begin{cases} x' = \mathbf{A}_\nu x + \mathbf{B}_\nu u_{\nu,i} \\ x(0) = 0 \end{cases}$$

where  $u_{\nu,i}$  is the *control* obtained by solving the corresponding adjoint problem (for the parameter  $\nu$ ) with initial datum  $\varphi_i^0$ .

Thus instead of dealing with  $\text{dist}(\varphi_\nu^0, \Phi_n^0)$  we use the *surrogate*  $\text{dist}(x^1 - e^{T\mathbf{A}_\nu} \mathbf{x}^0, \Lambda_\nu \Phi_n^0)$ , obtained by projecting an easy obtainable vector  $x^1 - e^{T\mathbf{A}_\nu} \mathbf{x}^0$  to a linear space  $\Lambda_\nu \Phi_n^0$  whose basis is obtained by solving the adjoint system plus the state one  $n$  times.

The surrogate measures the control performance of the snapshots  $\varphi_i^0, i = 1..n$  when applied to the system associated to parameter  $\nu$  (Figure 1). Namely, by relation (13) the minimiser  $\varphi_\nu^0$  uniquely determines the control  $u_\nu$  steering the system (1) from the initial datum  $x^0$  to the target  $x^1$ . By replacing  $\varphi_\nu^0$  with  $\varphi_i^0$  in (13) the system is driven to the state  $\Lambda_\nu \varphi_i^0 + e^{T\mathbf{A}_\nu} \mathbf{x}^0$ , whose distance from the target represents the surrogate value.

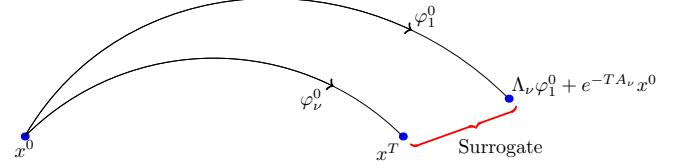


Fig. 1. The surrogate of  $\text{dist}(\varphi_\nu^0, \varphi_1^0)$

If for every  $\nu \in \mathcal{N}$  we can find a suitable linear combination of the above states close enough to the target, we deem that we have found a good approximation of the manifold  $\varphi^0(\mathcal{N})$ . Otherwise, we select as the next snapshot a value for which the already selected snapshots provide the worst performance. The precise description of the offline part of the algorithm is given below.

#### Greedy control algorithm - offline part

Fix the approximation error  $\varepsilon > 0$ .

##### STEP 1 (Discretisation)

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

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

where  $\delta > 0$  is a constant to be determined later (cf. (24), (33)), in dependence on the problem under consideration and the tolerance  $\varepsilon$ .

##### STEP 2 (Choosing $\nu_1$ )

Check the inequality

$$\max_{\tilde{\nu} \in \tilde{\mathcal{N}}} |x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} \mathbf{x}^0| < \frac{\varepsilon}{2} \quad (20)$$

and stop the algorithm if it holds.

Otherwise, determine the first distinguished parameter value as

$$\nu_1 = \operatorname{argmax}_{\tilde{\nu} \in \tilde{\mathcal{N}}} |x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} \mathbf{x}^0|. \quad (21)$$

and choose  $\varphi_1^0$  as the minimiser of  $J_\nu$  corresponding to  $\nu = \nu_1$ .

##### STEP 3 (Choosing $\nu_{j+1}$ )

Having chosen  $\nu_1, \dots, \nu_j$  calculate  $\Lambda_{\tilde{\nu}} \varphi_i^0, i = 1 \dots j$  for each  $\tilde{\nu} \in \tilde{\mathcal{N}}$ .

Check the approximation criteria

$$\max_{\tilde{\nu} \in \tilde{\mathcal{N}}} \text{dist}(x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} \mathbf{x}^0, \Lambda_{\tilde{\nu}} \Phi_j^0) < \frac{\varepsilon}{2}. \quad (22)$$

If the inequality is satisfied, stop the algorithm.

Otherwise, determine the next distinguished parameter value as

$$\nu_{j+1} = \operatorname{argmax}_{\tilde{\nu} \in \tilde{\mathcal{N}}} \text{dist}(x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} \mathbf{x}^0, \Lambda_{\tilde{\nu}} \Phi_j^0), \quad (23)$$

choose  $\varphi_{j+1}^0$  as the minimiser of  $J_\nu$  corresponding to  $\nu = \nu_{j+1}$  and repeat Step 3.

The algorithm results in the approximating space  $\Phi_n^0 =$

$\text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$ , where  $n$  is a number of chosen snapshots (specially  $\Phi_0^0 = \{0\}$  for  $n = 0$ ).

The value of the parameter  $\nu_1$  is chosen by testing the performance of the null control as an initial guess for all  $\tilde{\nu} \in \tilde{\mathcal{N}}$ . The selected value  $\nu_1$  is the one for which this performance provides the worst approximation. The algorithm is stopped at this initial level only if the null control ensures the uniform control of all system realisations within the given tolerance.

Note that the set  $\{\varphi_1^0, \dots, \varphi_n^0\}$  is linearly independent, as for vectors  $\varphi_\nu^0$  that linearly depend on already chosen ones, the corresponding surrogate distance (23), which is the criterion for the choice of new snapshots, vanishes. Thus the algorithm stops after, at most,  $n \leq N$  iterations, and it fulfils the requirements of the weak greedy theory. More precisely the following result holds.

**Theorem 4.1** *Let the  $\nu \rightarrow (\mathbf{A}(\nu), \mathbf{B}(\nu))$  be Lipschitz and such that the uniform controllability condition (12) holds, and let  $C_\varphi$  be the Lipschitz constant of the mapping  $\nu \rightarrow \varphi_\nu^0$  determined by (14).*

For a given  $\varepsilon > 0$  take the discretisation constant  $\delta$  such that

$$\delta \leq \varepsilon / (2C_\varphi \Lambda_-). \quad (24)$$

Then the above algorithm provides a weak greedy approximation of the manifold  $\varphi^0(\mathcal{N})$  with the constant

$$\gamma = \Lambda_- / (2\Lambda_+) \leq 1/2, \quad (25)$$

and the approximation error less than  $\varepsilon / \Lambda_-$ .

**Remark 1** The obtained approximation error  $\varepsilon / \Lambda_-$  of the family of minimisers  $\varphi^0(\mathcal{N})$  is a consequence of the required  $\varepsilon$  approximation of the set  $\{\Lambda_\nu \varphi_\nu^0, \nu \in \mathcal{N}\}$  (as  $|\Lambda \varphi_1^0 - \Lambda \varphi_2^0| < \varepsilon$  implies  $|\varphi_1^0 - \varphi_2^0| < \varepsilon / \Lambda_-$ ).

**Remark 2** The case  $n = 0$ , occurring when inequality (20) holds, is a trivial one resulting in a null approximating space that we exclude from the proof.

**Proof:** In order to prove the theorem we have to show:

a) that the selected minimisers  $\varphi_j^0$  associated to parameter values determined by (21) and (23) satisfy

$$\begin{aligned} |\varphi_1^0| &\geq \gamma \max_{\nu \in \mathcal{N}} |\varphi_\nu^0| \\ \text{dist}(\varphi_{j+1}^0, \Phi_j^0) &\geq \gamma \max_{\nu \in \mathcal{N}} \text{dist}(\varphi_\nu^0, \Phi_j^0), \quad j = 1..n-1 \end{aligned} \quad (26)$$

for the constant  $\gamma$  given by (25);

b) that the approximation error

$$\sigma_n(\varphi^0(\mathcal{N})) := \max_\nu \text{dist}(\varphi_\nu^0, \Phi_n^0)$$

obtained at the end of the algorithm is less than  $\varepsilon / \Lambda_-$ .

a) To this effect note that for the first snapshot the following estimates hold:

$$\begin{aligned} (\forall \tilde{\nu} \in \tilde{\mathcal{N}}) \quad |\varphi_1^0| &\geq \frac{1}{\Lambda_+} |x^1 - e^{T\mathbf{A}_1} \mathbf{x}^0| \\ &\geq \frac{1}{\Lambda_+} |x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} \mathbf{x}^0| \geq \frac{\Lambda_-}{\Lambda_+} |\varphi_{\tilde{\nu}}^0|, \end{aligned} \quad (27)$$

where we have used relation (14) (for  $\nu = \nu_1$ ) and the criterion (21).

In order to obtain an estimate including the whole set of parameters  $\mathcal{N}$ , we employ Lipschitz regularity of the mapping  $\nu \rightarrow \varphi_\nu^0$ . Thus for an arbitrary  $\nu \in \mathcal{N}$ , by taking  $\tilde{\nu} \in \tilde{\mathcal{N}}$  such that  $|\nu - \tilde{\nu}| < \delta$ , by means of (24) and (27) it follows

$$|\varphi_\nu^0| \leq |\varphi_{\tilde{\nu}}^0 - \varphi_\nu^0| + |\varphi_{\tilde{\nu}}^0| \leq \frac{\varepsilon}{2\Lambda_-} + \frac{\Lambda_+}{\Lambda_-} |\varphi_1^0|.$$

Having excluded the case (20), we have

$$\frac{\varepsilon}{2} \leq |\Lambda_1 \varphi_1^0| \leq \Lambda_+ |\varphi_1^0|.$$

implying

$$|\varphi_1^0| \geq \gamma \max_{\nu \in \mathcal{N}} |\varphi_\nu^0|, \quad (28)$$

with  $\gamma$  given by (25).

For a general  $j$ -th iteration, and an arbitrary  $\nu \in \mathcal{N}$ , by taking  $\tilde{\nu} \in B(\nu; \delta) \cap \tilde{\mathcal{N}}$  and using (24) we get

$$\begin{aligned} \text{dist}(\varphi_\nu^0, \Phi_j^0) &\leq |\varphi_\nu^0 - \varphi_{\tilde{\nu}}^0| + \text{dist}(\varphi_{\tilde{\nu}}^0, \Phi_j^0) \\ &\leq \frac{\varepsilon}{2\Lambda_-} + \frac{1}{\Lambda_-} \text{dist}(\Lambda_{\tilde{\nu}} \varphi_{\tilde{\nu}}^0, \Lambda_{\tilde{\nu}} \Phi_j^0) \\ &\leq \frac{1}{\Lambda_-} \left( \frac{\varepsilon}{2} + \text{dist}(\Lambda_{\tilde{\nu}} \varphi_{j+1}^0, \Lambda_{\tilde{\nu}} \Phi_j^0) \right), \end{aligned}$$

where in the last inequality follows from definition of the next snapshot (23). Since the stopping criteria (22) is not satisfied until  $j < n$ , the first term within the last bracket is less than the second one, implying

$$\begin{aligned} \text{dist}(\varphi_\nu^0, \Phi_j^0) &\leq \frac{2}{\Lambda_-} \text{dist}(\Lambda_{\tilde{\nu}} \varphi_{j+1}^0, \Lambda_{\tilde{\nu}} \Phi_j^0) \\ &\leq 2 \frac{\Lambda_+}{\Lambda_-} \text{dist}(\varphi_{j+1}^0, \Phi_j^0). \end{aligned}$$

Combining the last relation with (28), estimate (26) follows.

b) Finally, having achieved inequality (22) after  $n$  iterations, for an arbitrary  $\nu \in \mathcal{N}$ , taking  $\tilde{\nu}$  as above we get

$$\begin{aligned} \text{dist}(\varphi_\nu^0, \Phi_n^0) &\leq |\varphi_\nu^0 - \varphi_{\tilde{\nu}}^0| + \text{dist}(\varphi_{\tilde{\nu}}^0, \Phi_n^0) \\ &< \frac{\varepsilon}{2\Lambda_-} + \frac{1}{\Lambda_-} \text{dist}(\Lambda_{\tilde{\nu}} \varphi_{\tilde{\nu}}^0, \Lambda_{\tilde{\nu}} \Phi_n^0) < \frac{\varepsilon}{\Lambda_-}, \end{aligned} \quad (29)$$

thus obtaining the required approximation error of the set  $\varphi^0(\mathcal{N})$ .  $\square$

*Remark 3* Besides a choice of the discretisation constant  $\delta$  determined by (24), other choices are possible as well. Actually, taken any  $\delta > 0$ , define  $\tilde{C} := C_\varphi \delta / \varepsilon$ . The implementation of the algorithm in that case would lead to the greedy constant  $\gamma^{-1} = (2\tilde{C} + 1/\Lambda_-)\Lambda_+$  and the approximation error of the set  $\varphi^0(\mathcal{N})$  equal to  $\varepsilon(\tilde{C} + 1/(2\Lambda_-))$ . Note, however, that no choice of  $\delta$  can reduce the order of the error already obtained by (29).

#### 4.2 Construction of an approximate control for a given parameter value

Having constructed an approximating space  $\Phi_n^0$  of dimension  $n$ , we would like to exploit it for construction of an approximate control  $u_\nu^*$  associated to an arbitrary given value  $\nu \in \mathcal{N}$ .

Such a control is given by the relation

$$u_\nu^* = \mathbf{B}_\nu^* e^{(T-t)\mathbf{A}_\nu} \varphi_\nu^{0,*}, \quad (30)$$

where  $\varphi_\nu^{0,*}$  is appropriately chosen approximation of  $\varphi_\nu^0$  from  $\Phi_n^0$ . It steers the system to the state  $x_\nu^*(T) = \Lambda_\nu \varphi_\nu^{0,*} + e^{T\mathbf{A}_\nu} x^0$ . Comparing formula (30) with the one fulfilled by the exact control (13), we note that the only difference lies in the replacement of the unknown minimiser  $\varphi_\nu^0$  by its approximation  $\varphi_\nu^{0,*}$ .

Thus  $x_\nu^*(T) - e^{T\mathbf{A}_\nu} x^0$  is a suitable linear combination of vectors  $\Lambda_\nu \varphi_i^0, i = 1..n$ . As our goal is to steer the system to  $x^1$  as close as possible, the best performance is obtained if  $\Lambda_\nu \varphi_\nu^{0,*}$  is chosen as projection of  $\Lambda_\nu \varphi_\nu^0 = x^1 - e^{T\mathbf{A}_\nu} x^0$  to the space  $\Lambda_\nu \Phi_n^0 = \text{span}\{\Lambda_\nu \varphi_1^0, \dots, \Lambda_\nu \varphi_n^0\}$ .

For this reason we define approximation of  $\varphi_\nu^0$  as

$$\varphi_\nu^{0,*} = \sum_i^n \alpha_i \varphi_i^0, \quad (31)$$

where the coefficients  $\alpha_i$  are chosen such that  $\sum_i^n \alpha_i \Lambda_\nu \varphi_i^0$  represents projection of the vector  $x^1 - e^{T\mathbf{A}_\nu} x^0$  to the space  $\Lambda_\nu \Phi_n^0$ . This choice of  $\varphi_\nu^{0,*}$  corresponds to the minimisation of the functional  $J_\nu$ , determined by (8), over the space  $\Lambda_\nu \Phi_n^0$ .

In order to check performance estimate of the approximate control (30), note that for any value  $\nu \in \mathcal{N}$  we have

$$\begin{aligned} |x^1 - x_\nu^*(T)| &= |\Lambda_\nu(\varphi_\nu^0 - \varphi_\nu^{0,*})| \\ &\leq |\Lambda_\nu(\varphi_\nu^0 - \varphi_{\tilde{\nu}}^{0,*})| \\ &\leq |\Lambda_\nu \varphi_\nu^0 - \Lambda_{\tilde{\nu}} \varphi_{\tilde{\nu}}^0| + |\Lambda_{\tilde{\nu}}(\varphi_{\tilde{\nu}}^0 - \varphi_{\tilde{\nu}}^{0,*})| \\ &\quad + |(\Lambda_{\tilde{\nu}} - \Lambda_\nu)\varphi_{\tilde{\nu}}^{0,*}|, \end{aligned} \quad (32)$$

where we have used that  $\Lambda_\nu \varphi_\nu^{0,*}$  is the orthogonal projection of  $\Lambda_\nu \varphi_\nu^0$  to the space  $\Lambda_\nu \Phi_n^0$  (that also contains  $\Lambda_\nu \varphi_{\tilde{\nu}}^{0,*}$ ), while  $\tilde{\nu}$  is taken from the set  $B(\nu; \delta) \cap \mathcal{N}$ .

Taking into account the stopping criteria (22), the penultimate term in (32) is less than  $\varepsilon/2$ , while the preceding one equals

$$|(e^{T\mathbf{A}_\nu} - e^{T\mathbf{A}_{\tilde{\nu}}})x^0| \leq C_L |\nu - \tilde{\nu}|,$$

where  $C_L$  denotes the Lipschitz constant of the mapping  $\nu \rightarrow \Lambda_\nu \varphi_\nu^0 = x^1 - e^{T\mathbf{A}_\nu} x^0$ . Similarly, the last term in (32) can be estimated from above by  $C_\Lambda |\nu - \tilde{\nu}|$ , where  $C_\Lambda$  stands for the Lipschitz constant of the mapping  $\nu \rightarrow \Lambda_\nu$ . Thus in order to obtain a performance estimate less than  $\varepsilon$ , one possibly has to refine the discretisation used in Theorem 4.1 and to take

$$\delta \leq \frac{1}{2} \min\left\{\frac{\varepsilon}{C_L + C_\Lambda}, \frac{\varepsilon}{C_\varphi \Lambda_-}\right\}, \quad (33)$$

where, let it be repeated,  $C_\varphi$  denotes the Lipschitz constant of the mapping  $\nu \rightarrow \varphi_\nu^0$ .

This finalises solving of Problem 2 and leads us to the following result.

**Theorem 4.2** Let  $\nu \rightarrow (\mathbf{A}(\nu), \mathbf{B}(\nu))$  be a Lipschitz mapping such that the uniform controllability condition (12) holds.

Given  $\varepsilon > 0$ , let  $\Phi_n^0 = \text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$  be an approximating space constructed by the greedy control algorithm with the discretisation constant  $\delta$  given by (33). Then for any  $\nu \in \mathcal{N}$  the approximate control  $u_\nu^*$  given by (30) and (31) steers the control system (1) to the state  $x_\nu^*(T)$  within the  $\varepsilon$  distance from the target  $x^1$ .

*Remark 4* As already commented in Section 2, the approximate control  $u_\nu^*$  does not belong to the space  $\text{span}\{u_1, \dots, u_n\}$  spanned by controls associated to selected parameter values, as the control operator and system matrix entering (30) correspond to the given value  $\nu$ , while  $u_i = \mathbf{B}_i^* e^{(T-t)\mathbf{A}_i} \varphi_i^0$ .

*Remark 5* The greedy control also applies if we additionally assume that initial datum  $x^0$ , as well as target state  $x^1$  depend on the parameter  $\nu$  in a Lipschitz manner. In that case, the greedy search is performed in the same manner as above, i.e. by exploring elements of the set  $\{x_\nu^1 - e^{T\mathbf{A}_\nu} x_\nu^0 | \nu \in \mathcal{N}\}$  in calculation of the surrogate distance, and the obtained results remain valid with the same constants.

*Remark 6* All the above results on greedy control remain valid if instead of Lipschitz continuity we merely assume continuous dependence with respect to the parameter  $\nu \in \mathcal{K}$ . Namely, as  $\mathcal{K}$  is a compact set, the assumption directly implies uniform continuity, which suffices for the proof of Theorems 4.1 and 4.2 in which we need  $\varphi_\nu^0$  and  $\varphi_{\tilde{\nu}}^0$  to be close whenever  $\nu$  and  $\tilde{\nu}$  are. The only difference in that case is that the discretisation constant  $\delta$  can not be given explicitly in terms of  $\varepsilon$ , unlike expressions (24) and (33).

We finish the section describing the algorithm summarising the above procedure to construct the approximate

control  $u_\nu^*$ . Hereby we suppose that, for fixed  $\varepsilon > 0$ , the approximating space  $\Phi_n^0 = \text{span}\{\varphi_1^0, \dots, \varphi_n^0\}$  has been constructed by the offline part of the greedy control algorithm with the discretisation constant  $\delta$  given by (33).

### Greedy control algorithm - online part

A parameter value  $\nu \in \mathcal{N}$  is given.

**STEP 1** Calculate  $\Lambda_\nu \varphi_\nu^0 = x^1 - e^{T\mathbf{A}_\nu} x^0$ .

**STEP 2** Calculate  $\Lambda_\nu \varphi_i^0 = e^{(T-t)\mathbf{A}_\nu} B u_i^*, i = 1..n$ , where  $u_i^* = \mathbf{B}_\nu^* e^{(T-t)\mathbf{A}_\nu^*} \varphi_i^0$ .

**STEP 3** Project  $\Lambda_\nu \varphi_\nu^0$  to  $\Lambda_\nu \Phi_n^0 = \text{span}\{\Lambda_\nu \varphi_1^0, \dots, \Lambda_\nu \varphi_n^0\}$ . Denote the projection by  $P_n \Lambda_\nu \varphi_\nu^0$ .

**STEP 4** Solve the system  $P_n \Lambda_\nu \varphi_\nu^0 = \alpha_i \Lambda_\nu \varphi_i^0$  for  $\alpha_i, i = 1..n$ .

**STEP 5** The approximate control is given by

$$u_\nu^* = \sum_i \alpha_i u_i^*,$$

where  $u_i^*$  are already determined within Step 2.

For any  $\nu \in \mathcal{N}$  the obtained approximate control steers the system (1) within an  $\varepsilon$  distance from the target.

Note that for a parameter value  $\tilde{\nu}$  that belongs to the discretisation set  $\tilde{\mathcal{N}}$  used in the construction of an approximating space  $\Phi_n^0$  steps 1–3 of the last algorithm can be skipped as the corresponding terms have already been calculated within the offline part of the algorithm.

## 5 Computational cost

### 5.1 Offline part

The offline part of the greedy control algorithm consists of two main ingredients. First, the search for distinguished parameter values  $\nu_j$  by examining the surrogate value over the set  $\tilde{\mathcal{N}}$ , and, second, the calculation of the corresponding snapshots  $\varphi_j^0$ . This subsection is devoted to estimate its computational cost.

#### Choosing $\nu_1$

In order to identify the first parameter value, one has to maximise the expression  $|x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} x^0|$  over the set  $\tilde{\mathcal{N}}$ , which represents the distance of the target  $x^1$  from the free dynamics state. To this effect one has to solve the original system (1) with zero control  $k = \text{card}(\tilde{\mathcal{N}})$  times. Denoting the cost of solving the system by  $C$ , the cost of choosing  $\nu_1$  thus equals

$$k(C + 3N),$$

where the second term corresponds to calculation of the distance between two vectors in  $\mathbf{R}^N$ .

In general, the cost  $C$  differs depending on the type of matrices  $\mathbf{A}_\nu$  under consideration and the method chosen for solving the corresponding control system. For example, an implicit one step method consists on solving  $T/\Delta t$  linear systems, where  $\Delta t$  denotes the time discretisation step. However, as we consider time independent dynamics, all these systems have the same matrix, thus LU factorisation is required just ones. Consequently, the cost  $C$  can be estimated as

$$C = \frac{2}{3} N^3 + \frac{T}{\Delta t} 4N^2, \quad (34)$$

where the first part corresponds to the LU factorisation, while the latter one is the cost of building and solving a system of a type  $L U x_{k+1} = b_k$ .

#### Calculating $\varphi_1^0$

In order to determine the first snapshot one has to solve the system (14) for the chosen parameter value  $\nu_1$ . To this effect we construct (the unknown) Gramian matrix  $\Lambda_1$  by calculating  $\Lambda_1 e_i, i = 1..N$  for the vectors of the canonical basis  $e_i$ . According to the results presented in Section 2, for any  $\varphi^0 \in \mathbf{R}^N$  the corresponding vector  $\Lambda_1 \varphi^0$  can be determined as the state of the control system at time  $T$  with the control  $u = \mathbf{B}^* \varphi$ , with  $\varphi$  being the solution of the adjoint problem starting from  $\varphi^0$ .

In such a way, the cost of composing the system for  $\varphi_1^0$  equals  $2NC$  (corresponding to  $N$  control plus adjoint problems starting from  $N$  different data). As solving the system requires a number of operations of order  $2/3N^3$ , the cost of this part of the algorithm turns out as

$$2NC + \frac{2}{3} N^3.$$

#### Choosing $\nu_{j+1}$

Suppose we have determined the first  $j$  snapshots and have constructed the approximating space  $\Phi_j^0 = \text{span}\{\varphi_1^0, \dots, \varphi_j^0\}$ . The next parameter value is chosen by maximising the distance of the vector  $x^1 - e^{T\mathbf{A}_{\tilde{\nu}}} x^0$ , calculated already in the first iteration, from the space  $\Lambda_{\tilde{\nu}} \Phi_j^0$  over the set  $\tilde{\mathcal{N}} \setminus \{\nu_1, \dots, \nu_j\}$  (the already chosen values can obviously be excluded from the search). The basis of the above space has been determined gradually throughout the previous iterations up to the last vector  $\Lambda_{\tilde{\nu}} \varphi_j^0$ , whose calculation is performed at this level. As explained above, it requires  $2C$  operations, that have to be performed  $k - j$  times.

In addition, this basis is orthonormalised, thus enabling the efficient calculation of the distance in (22). This process is performed gradually throughout the algorithm as well, and in each iteration we just orthonormalise the last vector with respect to the rest of the, already orthonormalised set. The corresponding cost for a single

value  $\tilde{\nu} \in \tilde{\mathcal{N}}$  is of order  $4Nj$ . Similarly, the (orthogonal) projection of  $x^1 - e^T \mathbf{A}_{\tilde{\nu}} x^0$  to  $\Lambda_{\tilde{\nu}} \Phi_j^0$  takes into account its projection to  $\Lambda_{\tilde{\nu}} \Phi_{j-1}^0$  used in the previous iteration, and adds just a projection to the last introduced vector.

As its corresponding cost is of order  $4N$ , the total cost of this part of the algorithm equals

$$(k-j)(2C + 4Nj + 4N + 3N),$$

where, as in (34), the last term corresponds to calculating the distance between two vectors in  $\mathbf{R}^N$ .

### Calculating $\varphi_{j+1}^0$

In order to determine the next snapshot, we have to construct corresponding Gramian matrix  $\Lambda_{j+1}$ . As explained in the part related to the calculation of  $\varphi_1^0$ , this can be done by applying the Gramian to some basis of  $\mathbf{R}^N$ . As in the previous part of the algorithm we have already calculated  $\Lambda_{j+1} \varphi_i^0, i = 1 \dots j$ , it is enough to calculate  $\Lambda_{j+1} e_{j_i}, i = 1 \dots N-j$ , for vectors of canonical basis  $e_{j_i}$  complementing the set  $\Phi_j^0$  to the basis of  $\mathbf{R}^N$ . Thus building the matrix of the system (14) requires  $2(N-j)C$  operations.

In addition, the complementation of the set  $\Phi_j^0$  takes advantage of the basis obtained as complementation of the set  $\Phi_{j-1}^0$  in the previous iteration. Adding a new snapshot  $\varphi_j^0$  to that basis, one of vectors  $e_{j_i}$  has to be removed such that the new set results in a basis again. Identifying this vector requires solving a single  $N \times N$  system.

Finally, taking into account the cost of solving the system (14), the cost of calculating the next snapshot turns out to be

$$2(N-j)C + \frac{4}{3}N^3.$$

### Total cost

Summing the above costs for  $j = 1 \dots n$  the total cost of the algorithm results in

$$\begin{aligned} & k(C + 3N) + n(k - \frac{n}{2})(2C + 7N) + 2n^2N(k - \frac{2}{3}n) \\ & + 2Cn(N - \frac{n}{2}) + \frac{4}{3}N^3n \end{aligned} \quad (35)$$

As the cost  $C$  of solving the control system is  $O(N^3)$ , the most expensive part of the greedy control algorithm corresponds to the terms containing this cost.

It is interesting to notice that, as the number of chosen snapshots  $n$  approaches either the number of eligible parameter values  $k = |\tilde{\mathcal{N}}|$ , or the system dimension  $N$ , this part converges to

$$C(k + 2kN),$$

which corresponds, up to lower order terms, to the cost of calculating  $\varphi_{\tilde{\nu}}^0$  for all  $k$  values  $\tilde{\nu} \in \tilde{\mathcal{N}}$ .

This demonstrates that the application of the greedy control algorithm is always cheaper than a naive approach that consists of calculating controls for all values of the parameter from an uniform mesh on  $\mathcal{N}$ , taken fine enough so to achieve the approximative control (3).

Let us note that in the above analysis we have assumed that in the each iteration we calculate the surrogate distance (22) for all values of the set  $\tilde{\mathcal{N}}$ , except those already chosen by the greedy algorithm. And that we repeat the procedure until the residual is less than  $\varepsilon/2$  for all parameter values of  $\tilde{\mathcal{N}}$ . However, very likely, for some, or even many parameter values, the corresponding residual will be less than the given tolerance already before the last iteration  $n$ . And once it occurs, these values do not have to be explored any more, neither we have to calculate the corresponding surrogates for subsequently chosen snapshots. Therefore, the obtained cost estimates are rather conservative, and in practice we expect the real cost to be lower than the above one.

Further significant cost reductions are obtained under assumption of a system with parameter independent matrix  $\mathbf{A}$  and the control operator  $\mathbf{B}$  (with dependence remaining only in initial datum and/or in the target). In that case it turns that the corresponding Gramian matrix  $\Lambda$  is also parameter independent. For this reason, maximising the distance appearing in (22) requires solving of control plus adjoint problem just once in each iteration. As a result, the most expensive term of the total cost (35) is replaced by quite a moderate one

$$C(k + 2n + 2N).$$

## 5.2 Online part

### STEP 1

Calculating  $\Lambda_{\nu} \varphi_{\nu}^0$  corresponds to solving the control system once, whose computational cost equals  $C$ .

### STEP 2

Calculating  $\Lambda_{\nu} \varphi_i^0$  requires solving the loop of the adjoint plus control system  $n$  times, resulting in the cost of  $2nC$ .

### STEP 3

The most expensive part of this step consist in the Gram Schmidt orthonormalisation procedure of the set  $\{\Lambda_{\nu} \varphi_1^0, \dots, \Lambda_{\nu} \varphi_n^0\}$ , whose cost equals  $2Nn^2$ .

### STEP 4

Solving the system with QR decomposition requires  $2Nn^2 - 2n^2/3$  operations.

### STEP 5

The cost of this step is negligible compared to the previous ones.

### Total cost

Thus we obtain that the total cost of finding an approximative control  $u_\nu^*$  for a given parameter value equals

$$C(1 + 2n) + 4Nn^2 - 2n^2/3.$$

As  $n$  approaches the system dimension  $N$ , its most expensive part  $C(1 + 2n)$  converges, up to lower order terms, to the cost of calculating an exact control  $u_\nu$ . Consequently, the cost reduction obtained by choosing the approximative control  $u_\nu^*$  obtained by the greedy control algorithm instead of the exact one depends linearly on the ratio between the number of used snapshots  $n$  and the system dimension  $N$ .

## 6 Infinite dimensional problems

The theory and the (weak) greedy control algorithm developed in the preceding section for finite dimensional linear control systems extend to ODEs in infinite dimensional spaces. They can be written exactly as in the form (1) except for the fact that solutions  $x(t, \nu)$  for each value of the parameter  $\nu$  and each time  $t$  live in an infinite dimensional Hilbert space  $X$ . The key assumption that distinguishes these infinite-dimensional ODEs from Partial Differential Equations (PDE) is that the operators  $(\mathbf{A}(\nu), \mathbf{B}(\nu))$  entering in the system are assumed to be bounded.

The controllability of such systems has been extensively elaborated during last few decades, expressed in terms of semigroups generated by a bounded linear operator, see, for instance,[16–18]. In fact, the existing theory of controllability for linear Partial Differential Equations (PDE) can be applied in that context too. This allows to characterise controllability in terms of the observability of the corresponding adjoint systems. In this way, the uniform controllability conditions of parameter dependent control problems can be recast in terms of the uniform observability of the corresponding adjoint systems.

However, in here, we limit our analysis to the case of infinite-dimensional ODEs for which the evolution is generated by bounded linear operators, contrarily to the case of PDEs in which the generator is systematically an unbounded operator. The greedy theory developed here is easier to implement in the context of infinite-dimensional ODEs since, in particular, the analytic dependence property of controls with respect to the parameters entering in the system can be more easily established.

In fact, in the context of infinite-dimensional ODEs, most of the results in Section 2 on finite-dimensional systems apply as well. In particular, the characterisation of the controls as in (13), in terms of minimisers of quadratic functionals of the form  $J$  as in (8) holds in this case too, together with (10) for the Gramian operators. Obviously, the Kalman rank condition cannot

be extended to the infinite-dimensional case. But the open character of the property of controllability with respect to parametric perturbations remains true in the infinite-dimensional case too, i.e. if the system is controllable for a given value of  $\nu$ , under the assumption that  $(\mathbf{A}(\nu), \mathbf{B}(\nu))$  depends on  $\nu$  in a continuous manner, it is also controllable for neighbouring values of  $\nu$ .

In this section we shall analyse convergence rates of the constructed greedy control algorithm as the dimension of the approximating space tends to infinity, an issue that only makes sense for infinite-dimensional systems. In particular, we are interested in the dimension of the approximating space  $\varphi^0(\mathcal{N})$  required to provide an uniform control within the given tolerance  $\varepsilon$ . This problem can be stated in terms of the estimate (22): what is the number of the algorithm iterations  $n$  we have to repeat until the estimate is fulfilled.

In case of systems of a finite dimension  $N$ , the algorithm constructs an  $n$  dimensional approximating space of  $\varphi^0(\mathcal{N}) \subseteq \mathbf{R}^N$ , and it certainly stops after, at most,  $n \leq N$  iterations. For infinite-dimensional systems, however there is no such an obvious stopping criteria.

In general we analyse the performance of the algorithm by comparing the greedy approximation errors  $\sigma_n(\varphi^0(\mathcal{N})) = \max_\nu \text{dist}(\varphi_\nu^0, \Phi_n^0)$  with the Kolmogorov widths  $d_n(\varphi^0(\mathcal{N}))$ , which represent the best possible approximation of  $\varphi^0(\mathcal{N})$  by a  $n$  dimensional subspace in  $\mathbf{R}^N$ . To this effect one could try to employ Theorem 3.1 which connects sequences  $(\sigma_n)$  and  $(d_n)$ . However, we have to apply the theorem to the set  $\varphi^0(\mathcal{N})$  (instead of  $\mathcal{N}$ ). But while Kolmogorov width of a set of admissible parameters  $\mathcal{N}$  is usually easy to estimate, that is not the case for a corresponding set of solutions (to a parametric dependent equation), or minimisers (as studied in this manuscript). Fortunately, a result in that direction has been provided recently for holomorphic mappings ([5]) under the assumption of a polynomial decay of Kolmogorov widths.

**Theorem 6.1** *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  $C_0 > 0$*

$$d_n(\mathcal{K}) \leq C_0 n^{-\alpha} \implies d_n(u(\mathcal{K})) \leq C_1 n^{-\beta}, \quad n \in \mathbf{N}, \quad (36)$$

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

*Remark 7* *The proof of the theorem provides an explicit estimate of the constant  $C_1$  in dependence on  $C_0$ ,  $\alpha$  and the mapping  $u$ . However, due to its rather complicated form we do not expose it here.*

Going back to our problem, Theorem 6.1 can be applied under the assumption that the mapping

$\nu \rightarrow (\mathbf{A}(\nu), \mathbf{B}(\nu))$  is analytic (its image being embedded in the space of linear and bounded operators in  $X$ ), which implies, in view of the representation formula (13), that the mapping  $\mathcal{N}$  to  $\varphi^0(\mathcal{N})$  is analytic as well. Note that this issue is much more delicate in the PDE setting, with the generator of the semigroup  $\mathbf{A}(\nu)$  being an unbounded operator. In fact the property fails in the context of hyperbolic problems although it is essentially true for elliptic and parabolic equations.

As we consider a finite number of parameters, lying in the set  $\mathcal{N} \subset \mathbf{R}^d$ , the polynomial decay of  $d_n(\mathcal{N})$  can be achieved at any rate  $\alpha > 0$  just by adjusting the corresponding constant  $C_0 = C_0(\alpha)$  in (36) (note that  $d_n(\mathcal{N}) = 0$  for  $n \geq d$ ). Of course, the Kolmogorov widths of the set  $\varphi^0(\mathcal{N})$  do not have to vanish for  $n$  large, but Theorem 6.1 ensures their polynomial decay at any rate.

Combining Theorems 3.1, 4.1 and 6.1 we thus obtain the following result.

**Corollary 6.2** *Let the mapping  $\nu \rightarrow (\mathbf{A}(\nu), \mathbf{B}(\nu))$ , corresponding to the parameter dependent infinite-dimensional ODEs, be analytic and such that the uniform controllability condition (12) holds. Then the greedy control algorithm ensures a polynomial decay of arbitrary order of the approximation rates.*

More precisely, for all  $\alpha > 0$  there exists  $C_\alpha > 0$  such that for any  $\nu$  the minimiser  $\varphi_\nu^0$  determined by the relation (14) can be approximated by linear combinations of the weak-greedy ones as follows:

$$\text{dist}(\varphi_\nu^0; \text{span}\{\varphi_j^0 : j = 1, \dots, n\}) \leq C_\alpha n^{-\alpha},$$

where  $C_\alpha$  can be determined by exploring constants appearing in (18), (25) and (36).

The last result provides us with a stopping criteria of the greedy control algorithm. For a given tolerance  $\varepsilon > 0$ , the algorithm stops after choosing  $n > \sqrt[\alpha]{C_\alpha/\varepsilon}$  snapshots. It results with a  $n$  dimensional space  $\Phi_n^0$  approximating the family of minimisers  $\varphi^0(\mathcal{N})$  within the error  $\varepsilon/\Lambda_-$ , and providing, by means of formulæ (30) and (31), a uniform control of our system within an  $\varepsilon$  tolerance.

A similar result holds if  $\mathcal{N}$  is infinite-dimensional, provided its Kolmogorov width decays polynomially. A typical example of such a set is represented by the so called affine model in which the parameter dependence is given by

$$\mathbf{A}(\boldsymbol{\nu}) = \sum_{l=1}^{\infty} \nu_l \mathbf{A}_l, \quad (37)$$

and/or similarly for  $\mathbf{B}(\mathbf{x}, \boldsymbol{\nu})$ . Here it is assumed that  $\boldsymbol{\nu}$  belongs to the unit cube in  $l^\infty$ , i.e. that  $|\nu_l| \leq 1$  for any  $l \in \mathbf{N}$ , while the sequence of numbers  $a_l := \|\mathbf{A}_l\|$  belongs to  $l^p$  for some  $p < 1$ .

However, note that in this case the Kolmogorov width of the set  $\mathcal{N} = B(0; 1, l^\infty)$  does not decay polynomially (actually these are constants equal to 1), but the polynomial decay is obtained for the set  $\mathbf{A}(\boldsymbol{\nu}) = \{\mathbf{A}(\cdot, \boldsymbol{\nu}); \boldsymbol{\nu} \in \mathcal{N}\}$ . Indeed, rearranging the indices so that the sequence  $(a_l)$  is decreasing, it follows

$$\begin{aligned} \|\mathbf{A}(\boldsymbol{\nu}) - \sum_{l=1}^n \nu_l \mathbf{A}_l\| &\leq \sum_{l=n+1}^{\infty} \nu_l a_l \\ &\leq a_{n+1}^{1-p} \sum_{l=n+1}^{\infty} \nu_l a_l^p \leq C n^{1-1/p} \end{aligned}$$

where we have used that for a decreasing  $l^p$  sequence we must have  $a_n^p \leq C n^{-1}, n \geq 1$ .

Thus one can consider the mapping  $(\mathbf{A}(\boldsymbol{\nu}), \mathbf{B}(\boldsymbol{\nu})) \rightarrow \varphi_\boldsymbol{\nu}^0$  and apply Theorem 6.1 to  $K = (\mathbf{A}(\boldsymbol{\nu}), \mathbf{B}(\boldsymbol{\nu}))$ .

Furthermore, in the case of the affine model this theorem can be improved, implying the Kolmogorov  $n$ -widths of sets  $K$  and  $u(K)$  decay at the same rate  $n^{-\alpha}, \alpha = 1/p-1$  (e.g. [5]). Consequently, one obtains that the greedy approximation rates  $\sigma_n(\varphi^0(\mathcal{N}))$  decay at the same rate as well.

Finally, let us mention that the cost of the greedy control is significantly reduced if one considers an affine model in which the control operator has finite representation of the form (37), while the system matrix  $\mathbf{A}$  is taken as parameter independent.

In that case the corresponding Gramian is of the form:

$$\Lambda(\boldsymbol{\nu}) = \sum_{l,m=1}^L \nu_l \nu_m \Lambda_{lm} \quad (38)$$

where  $\Lambda_{lm} = \int_0^T e^{(T-t)\mathbf{A}} \mathbf{B}_l \mathbf{B}_m^* e^{(T-t)\mathbf{A}^*} dt$ , while  $\boldsymbol{\nu} = (\nu_1, \dots, \nu_L)$ . Here we consider a finite representation, but a more general one can be reduced to this one by truncation of the series. Consequently, computing  $\Lambda_{\tilde{\boldsymbol{\nu}}} \varphi_j^0$  for a chosen snapshot does not require solving the loop of the adjoint plus control system for each value  $\tilde{\boldsymbol{\nu}} \in \tilde{\mathcal{N}}$ . Instead, it is enough to perform the loop  $L^2$  times in order to obtain vectors  $\Lambda_{lm} \varphi_j^0, l, m = 1..L$ , and express  $\Lambda_{\tilde{\boldsymbol{\nu}}} \varphi_j^0$  as their linear combination by means of (38).

Such approach can result in a lower cost of the greedy control algorithm compared to the one obtained in the previous section, with a precise reduction rate depending on the relation between the series length  $L$  and the number of eligible parameters  $k$ .

## 7 Numerical examples

### 7.1 Wave equation

We consider the control system (1) whose governing matrix has the block form

$$\mathbf{A} = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{N/2} \\ \nu(N/2 + 1)^2 \tilde{\mathbf{A}} & \mathbf{0} \end{pmatrix},$$

where  $\mathbf{I}_{N/2}$  is the identity matrix of dimension  $N/2$ , while

$$\tilde{\mathbf{A}} = \begin{pmatrix} -2 & 1 & 0 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 & 0 \\ 0 & 1 & -2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -2 & 1 \\ 0 & 0 & 0 & \cdots & 1 & -2 \end{pmatrix}. \quad (39)$$

The control operator is assumed to be of the parameter-independent form

$$\mathbf{B} = (0, \dots, 0, (N/2 + 1)^2)^\top.$$

The system corresponds to the semi-discretisation of the wave equation problem with the control on the right boundary:

$$\begin{cases} \partial_{tt}v - \nu \partial_{xx}v = 0, & (t, x) \in \langle 0, T \rangle \times \langle 0, 1 \rangle, \\ v(t, 0) = 0, \quad v(t, 1) = u_\nu(t), \\ v(0, x) = v_0(x), \quad \partial_t v(x, 0) = v_1(x). \end{cases} \quad (40)$$

Parameter  $\nu$  represents the (square of ) velocity of propagation, while  $N/2$  corresponds to the number of inner points in the discretisation of the space domain.

For this example we specify the following ingredients:

$$T = 3, \quad N = 50, \quad v_0 = \sin(\pi x), \quad v_1 = 0.$$

The final target is set at  $x^1 = 0$ , and we assume

$$\nu \in [1, 10] = \mathcal{N}.$$

The system satisfies the Kalman's rank condition and accordingly the control exists for any value of  $\nu$ . Although the convergence of this direct approximation method, in which one computes the control for a finite-difference discretisation, hoping that it will lead to a good approximation of the continuous control, is false in general (see [19]), this is a natural way to proceed and it is interesting to test the efficiency of the greedy approximation for a given  $N$ , which in practice corresponds to fixing the space-mesh for the discretisation. In any case, the time-horizon  $T$  is chosen such that the Geometrical Control Condition (ensuring controllability of the continuous problem (40), see [2]) is satisfied for the continuous wave equation and all  $\nu \in \mathcal{N}$ .

The greedy control algorithm has been applied with

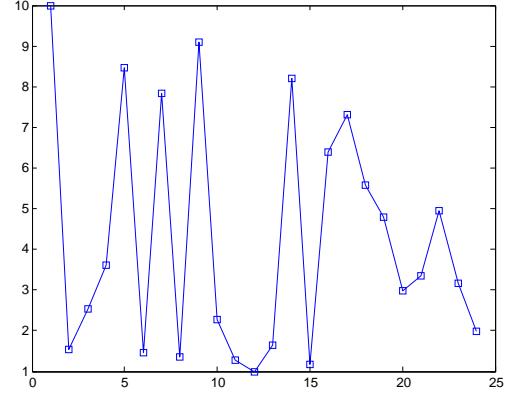


Fig. 2. Distribution of the selected parameter values.

$\varepsilon = 0.5$  and the uniform discretisation of  $\mathcal{N}$  in  $k = 100$  values.

The algorithm stopped after 24 iterations, choosing 24 (out of 100) parameter values depicted on Figure 2. The figure manifests the way by which the algorithm selects the values, taking them from different parts of the parameter set in a zigzag manner, in order to obtain the best possible approximation performance. The corresponding minimisers  $\varphi_i^0$  have been calculated and stored, completing the offline part of the process. The elapsed time of the algorithm is 2.312 seconds, performed on a personal notebook with a 2.7 GHz processor and DDR3 RAM with 8 GB and 1.6 GHz.

In the online part one explores the stored data in order to construct approximate controls for different parameter values ranging between 1 and 10 by means of formula (30). Here we present the results obtained for  $\nu = \pi$ .

The approximate control constructed by the greedy control algorithm is depicted on Figure 3.a. It is worth to mention that the constructed control turns out to be almost indistinguishable from the optimal exact control obtained by standard methods, based on minimisation of the cost functional  $J$  from (8). Although providing almost the same output as the classical methods, the advantage of the approach presented in this paper lies in the gain on computational cost. The elapsed time of the online part of the algorithm in this case is 7 seconds, compared to 51 second required to construct the optimal exact control by standard methods. This result confirms the computational efficiency of the greedy approach elaborated in Section 5.

Figure 3.b displays the evolution of the last 5 components of the system, corresponding to the time derivative of the solution to (40) at grid points, controlled by the approximate control  $u_\nu^*$ . As required, their trajectories are driven (close) to the zero target.

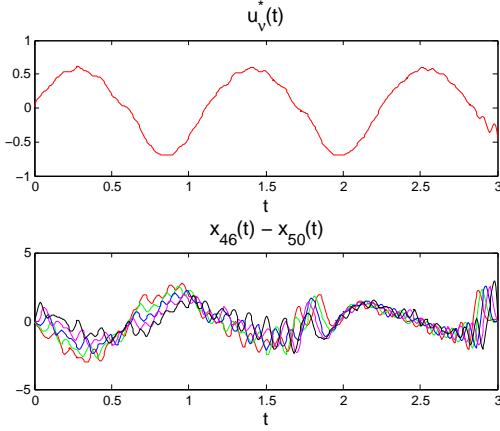


Fig. 3. Evolution of a) the approximate control and b) the last 5 system components for  $\nu = \pi$ .

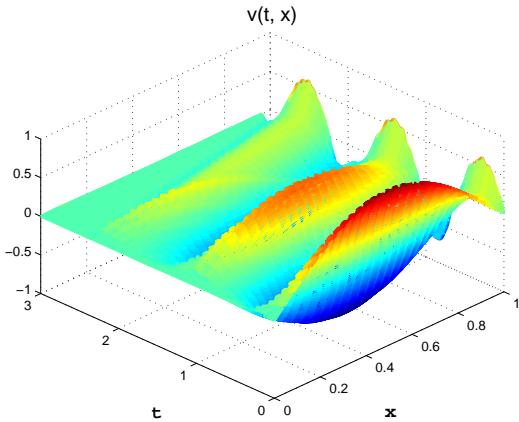


Fig. 4. Evolution of the solution to the semi-discretised problem (40) governed by the approximate control  $u_\nu^*$  for  $\nu = \pi$ .

The first 25 components of the system are plotted by a 3-D plot (Figure 4) depicting the evolution of the solution to a semi-discretised problem (40) governed by the approximate control  $u_\nu^*$ . The starting wave front, corresponding to the sinusoidal initial position, as well as the oscillations in time, gradually diminish and the solution is steered (close) to zero, as required. Furthermore, the total distance of the solution from the target equals  $|x(T)| = 0.05$ .

The efficiency of the greedy control approach is checked by exploring the corresponding approximation rates  $\sigma_n(\varphi^0(\tilde{\mathcal{N}})) = \max_{\tilde{\nu}} \text{dist}(\varphi_{\tilde{\nu}}^0, \Phi_n^0)$ , whose logarithmic values are depicted by the solid curve of Figure 5. The curve decreases and goes below  $\log(\varepsilon) = \log(0.5)$  for  $n = 24$ , stopping the algorithm after 24 iterations.

Almost linear shape of the curve suggests an exponential decay of the approximation rates  $\sigma_n(\varphi^0(\tilde{\mathcal{N}}))$ . Such appearance is in accordance with Corollary 6.2 providing their polynomial decay at any rate. The abrupt breakup

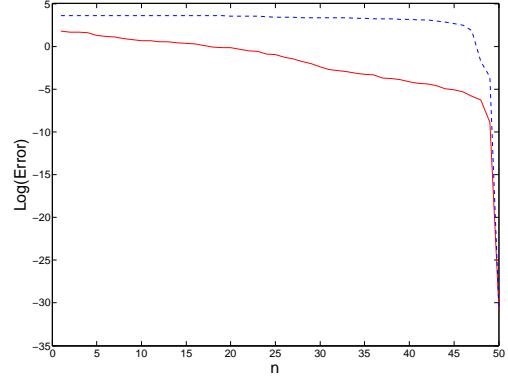


Fig. 5. Greedy (solid) vs canonical (dashed) approximation rates.

of the linear behaviour near  $n = 50$  is due to the fact that we consider a finite dimensional problem in which the approximation space is saturated when  $n$  reaches the size of the ambient space  $\mathbf{R}^N$ .

The bases of the approximation spaces  $\Phi_n^0$  are built iteratively in a *greedy* manner, by exploring the surrogate distance (19) over the set  $\tilde{\mathcal{N}}$ . As explained in the previous section, this search is costly, but should produce optimal approximation rates in the sense of the Kolmogorov widths. As opposed to that, one can explore some a cheaper approach, in which vectors spanning an approximation space are chosen arbitrarily, e.g. by taking vectors of the canonical basis.

The dashed curve of Figure 5 plots approximation errors of spaces  $E_n$  spanned by the first  $n$  vectors of the canonical basis in  $\mathbf{R}^N$ . Obviously the greedy approach wins over the latter one, in accordance with the theoretical results, ensuring optimal performance. Eventually, both curves exhibit sharp decay at the moment in which the size of the approximation space  $n$  reaches the size of the ambient space  $\mathbf{R}^N$ .

## 7.2 Heat equation

For  $\mathbf{A} = (N+1)^2 \tilde{\mathbf{A}}$ , with  $\tilde{\mathbf{A}}$  given by (39), and the control operator  $\mathbf{B} = (0, \dots, 0, (N+1)^2)^\top$ , the system (1) corresponds to the space-discretisation of the heat equation problem with  $N$  internal grid points and the control on the right boundary:

$$\begin{cases} \partial_t v - \nu \partial_{xx} v = 0, & (t, x) \in \langle 0, T \rangle \times \langle 0, 1 \rangle, \\ v(t, 0) = 0, & v(t, 1) = u_\nu(t), \\ v(0, x) = v_0. \end{cases} \quad (41)$$

The parameter  $\nu$  represents the diffusion coefficient and is supposed to range within the set  $\mathcal{N} = [1, 2]$ . The system satisfies the Kalman's rank condition for any  $\nu \in \mathcal{N}$  and any target time  $T$ .

We aim to control the system from the initial state  $v_0(x) = \sin(\pi x)$  to zero in time  $T = 0.1$ .

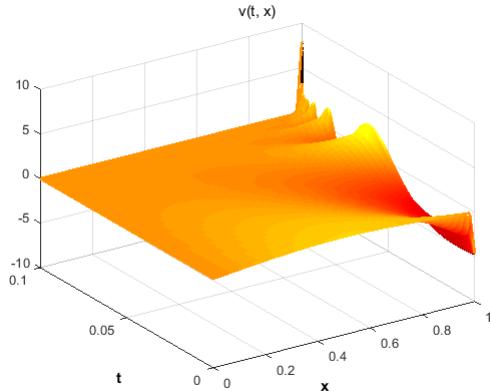


Fig. 6. Evolution of the solution to the semi-discretised problem (41) governed by the approximate control  $u_\nu^*$  for  $\nu = \sqrt{2}$ .

The greedy control algorithm has been applied for the system of dimension  $N = 50$  with  $\varepsilon = 0.0001$ , and the uniform discretisation of  $\mathcal{N}$  in  $k = 100$  values.

The algorithm stops after only three iterations, choosing parameter values (out of 100 eligibles) in the following order:

$$1.00, 1.18, 1.45.$$

The elapsed time is 213 seconds (obtained on the same machine used for example 7.1).

The corresponding three minimisers  $\varphi_i^0$  are used for constructing approximate controls for all parameter values by means of formulas (30) and (31). Here we present the results obtained for  $\nu = \sqrt{2}$ .

Evolution of the solution, presented by a 3-D plot is given by Figure 6. The system is driven to the zero state within the error  $|x(T)| = 1 \cdot 10^{-5}$ . The corresponding approximate control profile is depicted on Figure 7, exhibiting rather strong oscillations when approaching the target time – an intrinsic feature for the heat equation.

Let us mention that the elapsed time for constructing the approximate control  $u_\nu^*$  was 1.5 seconds only, compared to 37 seconds required to construct the optimal exact control. This huge difference is due to a rather small number of snapshots required to obtain approximation performance within the given accuracy, and it reconfirms computational efficiency of the greedy approach when compared to the standard one.

The choice of a rather tiny value of the target time  $T = 0.1$  is due to the strong dissipation effect of the heat equation, providing an exponential solution decay even in the absence of any control. For the same reason, the algorithm stopped after only three iterations although the precision was set rather high with  $\varepsilon = 0.0001$  (for the wave problem 24 iterations were required in order to produce uniform approximation within the error 0.5).

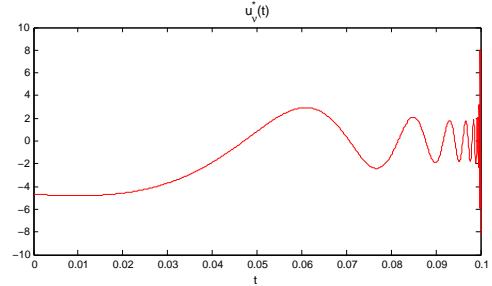


Fig. 7. Evolution of the approximate control for  $\nu = \sqrt{2}$ .

## 8 Conclusion and perspectives

- *Exact controllability.* In our analysis the exact control problem is relaxed to the approximate one by letting a tolerance  $\varepsilon$  on the target at time  $T$ , what is realistic in applications. Note however that this approximation of the final controllability condition is achieved by ensuring a sufficiently small error on the approximation of exact controls. Thus, in practice, the methods we have built are of direct application to the problem of exact controllability as well as the tolerance  $\varepsilon$  tends to zero.

- *Comparison of greedy and naive approach.* As we have seen the weak greedy algorithm we have developed has various advantages with respect to the naive one, the latter consisting of taking a fine enough uniform mesh on the set of eligible parameters  $\mathcal{N}$  and calculating all corresponding controls. Although the greedy approach requires an (expensive) search for distinguished parameter values, in the finite-dimensional case its computational cost is smaller than for the naive one. Meanwhile, in the infinite-dimensional context it leads to algorithms with an optimal convergence rate.

- *Model reduction:* The computational cost of the greedy methods, as developed here, remains high. This is so because on the search of the new distinguished parameter values, to compute the surrogate, we are obliged to solve fully the adjoint system first and then the controlled one. It would be interesting to build cheaper (from a computational viewpoint) surrogates. This would require implementing model reduction techniques. But this, as far as we know, has not been successfully done in the context of controllability. Model reduction methods have been implemented for elliptic and parabolic optimal control problems (see [10], [11], and [12]) but, as far as we know, they have not been developed in the context of controllability.

- *Parameter dependence.* Most existing practical realisations of (weak) greedy algorithms assume an affine parameter dependence presented at the end of Section 6 ([6,7,12]). The assumption provides a cheaper computation of a surrogate and reduces the cost of the search for a next snapshot.

By the contrary, the greedy control algorithm presented in this paper allows for very general param-

ter dependence, still providing an efficient algorithm beating the naive approach. Smooth dependence is needed in order to obtain sharp convergence rates. Still the algorithms can be applied for models depending on the parameters in a rough manner.

The greedy control algorithm and corresponding approximation results of Section 4, as well as solution to our Problem is derived under assumption of Lipschitz continuity of the control system entries with respect to the parameter. However, as explained in Remark 6, all these results remain valid if we merely assume continuous dependence with respect to the parameter. The Lipschitz regularity is chosen just to simplify the presentation and to obtain explicit expressions by means of the given approximation error  $\varepsilon$  and the Lipschitz constants.

A generalisation of the approach to problems involving discontinuous parameter dependence would require an approximation by piecewise smooth functions and separate applications of the algorithm on each subinterval fulfilling the continuity assumption.

The additional analyticity requirement is imposed in Section 6 in order to deduce convergence analysis of the algorithm. Namely, transfer of Kolmogorov width from the set of parameters into the set of solutions (or controls) is provided so far only by Theorem 6.1, which requires analyticity assumption, as well as polynomial decay of Kolmogorov widths. Once this result is expanded to a more general setting, it can be applied directly to the presented greedy control algorithm, and provide convergence estimates under weaker regularity assumptions, without modifying the very algorithm.

However, the present lack of an analogous result under more general assumptions does not prevent possible applications of the greedy control algorithm to non-analytic data. It can still be implemented, practically at no-risk. Namely, in the worst case when dimension of approximation space reaches number of eligible parameters, one will calculate controls for all these parameter values, which corresponds to the naive approach. And this will cost the same as applying naive method directly from the beginning. But there are many reasons to believe the algorithm will stop before this ultimate point.

- *Waves versus heat:* In our numerical experiments we have observed that the greedy algorithm is more efficient for the semi-discrete heat equation than for the wave one, in the sense that for the first one it stops after  $n = 3$  iterations while for the second one it goes up to  $n = 24$ . This is a natural result to be expected. Indeed, for the heat equation, because of the intrinsic dissipativity of the system, the high frequency components play a minor role when, as in our experiments, dealing with approximate controllability. Thus, even if the algebraic dimension of both systems, waves and heat, are the same, in practice the relevant dimensions for the heat equation is much smaller, thus explaining the faster convergence of the greedy algorithm for

heat like equations.

- *Extension of greedy control to PDE systems.* As we have seen the methods and ideas developed in this paper can be easily adapted to infinite-dimensional ODEs. The same ideas and methods can be implemented without major changes on the PDE setting. However, as we have seen above, in order to quantify the convergence rate of greedy algorithms, the analytic dependence of the controls with respect to parameters plays a key role. As far as we know, this issue has not been thoroughly addressed in the literature, i.e. that of whether or not controls for a given PDE controllability problem depend analytically on the parameters entering in the system. And at this level, the type of PDE under consideration can play a major role. Indeed, for heat equations, even when the unknown parameters enter in the principal part of the diffusion operator, the analytic dependence of the controls can be expected. This is not the case for wave equations (see [1]). Indeed, note that even for the simplest first order transport equation solutions fail to depend analytically on the velocity of propagation in a natural  $L^2$  or Sobolev energy setting. This issue requires significant further investigation.
- *Robust control.* The analysis and methods developed in this paper apply to control problems where the data to be controlled are prescribed a priori, either depending on the parameters or not. In that sense, our results are the analogue to what has been done for the greedy approximation of PDE solutions with given data (right hand side terms and boundary values) as in [3], [5] and [6], among others.

From the point of view of applications it would be interesting to develop greedy methods for control of potential application for all possible data to be controlled. This would require to establish a strategy for identifying the most relevant snapshots of the parameters  $\nu$  for the approximation of the Gramians  $\Lambda(\nu)$ , within the space of bounded linear operators. And this, in turn, requires identifying efficient surrogates. This is an interesting problem to be addressed.

Actually, as far as we know, this has not been done even in the context of the solvability of elliptic problems. The work done so far, as mentioned above, addresses approximation issues for given specific data. But the problem of applying greedy algorithms to approximate the resolvent operators, in a robust manner and independently of the data entering in the PDE, has not been addressed.

**Acknowledgements.** The authors acknowledges A. Cohen and R. DeVore for fruitful discussions, and J. Lohéac for valuable support with the numerical simulations.

## References

- [1] G. ALLAIRE AND A. KELLY, (2011). *Optimal Design of Low-Contrast Two Phase Structures for the Wave Equation*.

- Math. Models Methods Appl. Sci., 21(7), 1499–1538.
- [2] C. BARDOS, G. LEBEAU AND J. RAUCH, (1992). *Sharp sufficient conditions for the observation, control, and stabilization of waves from the boundary*. SIAM J. Control Optim., 30(5), 1024–1065.
- [3] P. BINEV, A. COHEN, W. DAHMEN, R. DEVORE, G. PETROVA AND P. WOJTASZCZYK, (2011). *Convergence rates for greedy algorithms in reduced basis methods*. SIAM J. Math. Anal., 43(3), 1457–1472.
- [4] A. BUFFA, Y. MADAY, A.T. PATERA, C. PRUDHOMME AND G. TURINICI, (2012). *A Priori convergence of the greedy algorithm for the parameterized reduced basis*. Math. Model. Numer. Anal., 46, 595–603.
- [5] A. COHEN AND R. DEVORE, *Kolmogorov widths under holomorphic mappings*. submitted
- [6] A. COHEN AND R. DEVORE, (2015). *Approximation of high-dimensional parametric PDEs*. Acta Numerica, 24, 1–159.
- [7] W. DAHMEN (2015). How to best sample a solution manifold? Arxiv preprint.
- [8] R. DEVORE, (2015). *The Theoretical Foundation of Reduced Basis Methods*. preprint.
- [9] R. DEVORE, G. PETROVA AND P. WOJTASZCZYK, (2013). *Greedy Algorithms for Reduced Bases in Banach Spaces*. Constr. Approx.. 37 455–466.
- [10] M.A. GREPL, N.C. NGUYEN, K. VEROY, A.T. PATERA AND G.R. LIU, *Certified Rapid Solution of Partial Differential Equations for Real-Time Parameter Estimation and Optimization*. in Real-Time PDE-Constrained Optimization, Lorenz T. Biegler et al. (eds), SIAM, 2007.
- [11] M.A. GREPL AND M. KÄRCHER, (2011). *Reduced basis a posteriori error bounds for parametrized linear-quadratic elliptic optimal control problems*. Comptes Rendus Mathematique, 349(15-16), 873–877.
- [12] M. KÄRCHER AND M.A. GREPL, (2014). *A certified reduced basis method for parametrized elliptic optimal control problems*. ESAIM: COCV, 20, 416 – 441.
- [13] M. LAZAR AND E. ZUAZUA, (2014). *Averaged control and observation of parameter-depending wave equations*. C. R. Acad. Sci. Paris, Ser. I, 352, 497–7502.
- [14] T. LERY ET AL. (EDS), (2011). *European Success Stories in Industrial Mathematics*, Springer.
- [15] S. MICU AND E. ZUAZUA, (2005). *An introduction to the controllability of linear PDE*. Contrôle non linéaire et applications, Sari, T., ed., Collection Travaux en Cours Hermann, 67–150.
- [16] R. TRIGGIANI, (1975). *Controllability and observability in Banach space with bounded operators*. SIAM J. Control, 13(2), 462–491
- [17] M. TUCSNAK, G. WEISS, (2009). *Observation and Control for Operator Semigroups*. Birkhäuser.
- [18] J. ZABCZYK, (2008). *Mathematical Control Theory*. Birkhäuser.
- [19] E. ZUAZUA, (2005). *Propagation, observation, and control of waves approximated by finite difference methods*. SIAM Review, 47 (2), 197–243.
- [20] E. ZUAZUA, (2006). *Controllability and Observability of Partial Differential Equations: Some results and open problems*. *Handbook of Differential Equations: Evolutionary Equations*, vol. 3, C. M. Dafermos and E. Feireisl (eds), Elsevier Science, Amsterdam, 527–621.
- [21] E. ZUAZUA, (2014). *Averaged control*. Automatica, 50, 3077–3087.