Numerical aspect of uniform ensemble controllability for linear systems

Jérôme Lohéac, Martin Lazar

Numerical aspect of uniform ensemble controllability for linear systems

Martin Lazar[1] Jérôme Lohéac[2]

December 20, 2024

Abstract

We consider the ensemble controllability problem for a linear time-invariant system $\dot{x}(t,\theta) = A(\theta)x(t,\theta) + B(\theta)u(t)$, where A and B are continuous matrices with respect to the parameter θ, which belongs to some compact set $\Theta \subset \mathbb{R}$. Given any continuous initial state datum $\theta \mapsto x^0(\theta)$ and any continuous target state $\theta \mapsto x^1(\theta)$, we investigate the numerical computation of a *θ-independent* open loop control u such that x^0 is steered, in a given time $T > 0$, at a distance $\varepsilon > 0$ of x^1 in the uniform norm (with respect to the parameter).

We approach the problem both theoretically and numerically. Using the Fenchel-Rockafellar duality, we first prove the existence and uniqueness of the ensemble control of a minimal L^2 norm. The numerical recovery of the optimal control is obtained by solving the dual problem, which consists in the unconstrained minimization of a non-differentiable functional in the space of Radon measures.

Keywords: parameter dependent systems, ensemble controllability, linear control system, Fenchel-Rockafellar duality, numerical algorithms.

Contents

1 Introduction **2**

2 Preliminaries **3**

3 Minimal L^2-norm controls **5**
3.1 Optimality conditions 5
3.2 Continuity properties of the minimizers 10

4 Systems with a finite number of parameter values **14**

5 Discretized problem and convergence **18**

6 Numerical strategies **22**
6.1 Greedy coordinate descent 22
6.2 Reduced minimization problem 26
6.3 Proposed numerical strategies 26

7 Numerical examples **27**

A Proof of Proposition 6.1 **32**

B Some unreachable states **35**

[1]Faculty of Electrical Engineering and Applied Computing, University of Dubrovnik, Croatia
[2]Université de Lorraine, CNRS, CRAN UMR 7039 Nancy, France

1 Introduction

The concept of ensemble controllability [1, 8, 16, 21, 23] needs no more publicity nowadays with regard to its practical importance. Its aim is to control systems whose natural dynamic depends on some unknown or uncertain parameters in a robust and computationally efficient manner. In particular, the goal is to find a parameter invariant control that performs well for each particular realization of the parameters. This kind of problem arises for instance in the transport of quantum particles, where the goal is to develop external excitation that can simultaneously steer the ensemble of systems with variation in their internal parameters from an initial state to a desired final state [20]. One can also see [3] where it is used to study the controllability of the Bloch equation, for an ensemble of non-interacting half-spins, in a static magnetic field, with dispersion in the Larmor frequency. Other real life practical example can be found, as for instance in non-holonomic systems theory where ensemble controllability is used to derive an approximate steering algorithm for a non-holonomic unicycle in the presence of model perturbation [4]. Nowadays, however, the most widely application is related to machine learning, in particular to training processes in supervised learning. By using the dynamic perspective of neural networks, the problem of their training can be recast as a control problem in which the network coefficients (weights and biases) stand for data(parameter)-independent controls [13].

From the generic notion of ensemble controllability, many other notions emerged as the one of uniform ensemble controllability [16, 20, 21], uniform ensemble reachability, L^q-ensemble reachability [11]. Even if this topic is studied actively, it is far from being well understood. The necessary and sufficient condition of ensemble controllability can be obtained as a special form of the generalization of the Kalman rank theorem to infinite dimensional Banach spaces (e.g. [28, Theorem 3.1.1]). However, such a characterization is primarily theoretical and not suitable for checking the controllability properties of a particular system. In the above-mentioned non-exhaustive references, the authors have derived necessary and/or sufficient conditions for linear time invariant systems to be uniformly ensemble controllable in some special cases. From these references, especially [8, 10, 11], it could be seen that the topology of Θ plays a crucial role in the controllability issue. Roughly speaking, for regular enough parameter dependent systems, uniform ensemble controllability cannot hold if the compact set Θ is homeomorphic to some compact subset of $\mathbb{R}^d$ (with $d \geqslant 2$) with non-empty interior.

Most papers on the topic analyze conditions under which a system is ensemble controllable, but do not consider the associated optimization problem of finding a feasible control of minimal norm, nor do they deal with numerical recovery of the solution. An approximate construction of a feasible control is provided only in some special cases (e.g. [3, 4]). In [15] the authors consider an associated penalization problem in which the constraint of reaching the prescribed target with some a priori given precision is replaced by an additional term in the cost functional that penalizes deviation of the final state from the target. Consequently, such a problem consists in the unconstrained minimization of a smooth functional and is numerically easier to handle. In [25], the uniform ensemble control problem is replaced by a simultaneous one (consisting of hitting a reachable target). The author discusses the strategies by which the latter can be approximated by its discretized version (which is equivalent to an exact control problem for a large, finite-dimensional system). Let us also mention [26], where an ensemble control is designed using polynomial approximations. However, the results of [26] seem only applicable for single input systems.

In [18] we have characterized the optimal ensemble control in the discrete case (i.e., for a finite number of parameters). Then, assuming Lipschitz regularity, we showed how to obtain a suboptimal control for the continuous case. In the present article, we directly tackle the case of a continuous set Θ, and derive optimality conditions. Based on them we propose several numerical strategies for recovering the minimal norm control. More precisely, we prove the existence and uniqueness of a minimal norm ensemble control, and express it through the solution of the corre-

sponding dual problem, which consists in the unconstrained minimization of a non-differentiable functional in the space of Radon measures. The numerical strategies that we propose in this paper are obtained using the optimality condition on the adjoint problem.

The paper is organized as follows. In the next section, we recall the notion and some basic results on uniform ensemble controllability. In Section 3, we introduce the optimization problem of finding a uniform ensemble control of a minimal L^2 norm and derive the associated dual problem. We demonstrate the existence of a unique solution for both problems and list their properties. In particular, the continuity of solutions with respect to the problem entries is obtained. Section 4 deals with a special but important case of discretized parameters (obtaining only a finite number of values), where we also discuss the sparsity properties of the optimal measure. The convergence of solutions when passing from discrete to continuous parameter setting is demonstrated in Section 5. Numerical iterative algorithms are presented in Section 6, while their performances are checked and discussed on several particular examples in Section 7. Appendices A and B contain some technical and related results.

Notation. The following notations are used in the paper.
$\mathbb{N} = \{0, 1, 2, \dots\}$ stands for the set of natural numbers including zero, while $\mathbb{N}^*$ denotes the set of positive integers. $\mathbb{R}_+$ is the set of nonnegative real numbers, and we set $\mathbb{R}_+^* = \mathbb{R}_+ \setminus \{0\}$. For a matrix $M \in \mathbb{R}^{n\times m}$, $M^* \in \mathbb{R}^{m\times n}$ is the transpose matrix of M. $|\cdot|$ stands for the Euclidean norm in $\mathbb{R}^n$, and $\langle \cdot, \cdot \rangle$ is the scalar product on $\mathbb{R}^n$. The open (respectively closed) ball of $\mathbb{R}^n$ centered on some x and of radius $\varepsilon > 0$ is denoted by $B(x, \varepsilon)$ (respectively $\overline{B}(x, \varepsilon)$).
For X a Banach space and $f : X \to \mathbb{R} \cup \{\infty\}$, we define $\operatorname{dom} f = \{x \in X \ \mid \ f(x) < \infty\}$, and $\operatorname{cont} f$ is the set of points $x \in \operatorname{dom} f$ where f is continuous. For $S \subset X$, $\operatorname{int} S$ is the interior of S.
Given $n \in \mathbb{N}^*$, by $\mathcal{C}(\Theta)^n$ we denote the space of continuous complex vector functions on $\Theta \subseteq \mathbb{R}$ equipped with L^∞ norm, i.e., $\mathcal{C}(\Theta)^n = \mathcal{C}(\Theta; \mathbb{R}^n)$ and $\|f\|_{\mathcal{C}(\Theta)^n} = \sup_\theta |f(\theta)|$. $\mathcal{B}_\infty(y, \varepsilon)$ is the closed ball in the space $\mathcal{C}(\Theta)^n$ centered in y and of radius ε.
$\mathcal{M}(\Theta)$ stands for the set of Radon measures supported on Θ, while by $\mathcal{M}(\Theta)^n$ we denote the dual space to $\mathcal{C}(\Theta)^n$ with the dual product given by $\langle \mu, f \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} = \sum_{i=1}^n \int_\Theta f_i(\theta)\, \mathrm{d}\mu_i(\theta) = \int_\Theta f(\theta)^*\, \mathrm{d}\mu(\theta)$. In addition, for $F \in \mathcal{C}(\Theta)^{n\times m}$, we define $F^* \flat \mu \in \mathcal{M}(\Theta)^m$ by $\langle F^* \flat \mu, f \rangle_{\mathcal{M}(\Theta)^m, \mathcal{C}(\Theta)^m} = \langle \mu, Ff \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} = \int_\Theta \left(f(\theta)^* F(\theta)^* \right) \mathrm{d}\mu(\theta)$.

2 Preliminaries

Let Θ be a compact subset of $\mathbb{R}$ and consider for every constant $\theta \in \Theta$ the system described by

$$\dot{x}(t, \theta) = A(\theta)x(t, \theta) + B(\theta)u(t) \qquad (t > 0,\ \theta \in \Theta). \tag{2.1}$$

In (2.1), we assume that $A \in \mathcal{C}(\Theta)^{n\times n}$ and $B \in \mathcal{C}(\Theta)^{n\times m}$. The input u is independent of the parameter θ, and $\dot{x}$ is the derivative of x with respect to the time variable t. Given an initial state $x^0 \in \mathcal{C}(\Theta)^n$ and an input $u \in L^2(0, T)^m$, the solution of (2.1) at time t, is denoted by $x(t, \theta; x^0, u)$. More precisely, we have,

$$x(t, \theta; x^0, u) = e^{tA(\theta)}x^0(\theta) + \int_0^t e^{(t-\tau)A(\theta)}B(\theta)u(t)\, \mathrm{d}t \qquad (t \geqslant 0,\ \theta \in \Theta). \tag{2.2}$$

Definition 2.1 (Uniform Ensemble Controllability)**.** The system (2.1) is said to be *uniformly ensemble controllable* if for any $x^0, x^1 \in \mathcal{C}(\Theta)^n$ and any $\varepsilon > 0$, there exist a time $T > 0$ and a control $u \in L^2(0, T)^m$, such that

$$\sup_{\theta\in\Theta} \left| x(T, \theta; x^0, u) - x^1(\theta) \right| \leqslant \varepsilon.$$

The terminology *uniform* in the above definition is used to emphasize the fact that the set of output functions is the continuous one endowed with the uniform norm. This definition actually makes sense since for any continuous initial state datum x^0 and any θ-independent input $u \in L^2(0,T)^m$, the output trajectory is continuous with respect to the parameter θ.

Remark 2.2. The systems considered in the present paper focus on the case where $A(\theta)$ and $B(\theta)$ are real matrices. However, all the obtained results can be applied to complex matrix, having in mind that if $A(\theta) \in \mathbb{C}^{n\times n}$ and $B(\theta) \in \mathbb{C}^{n\times m}$ are complex matrices, then the *complex* system can be identified to a real system with matrices $\begin{pmatrix} \Re A & -\Im A \\ \Im A & \Re A \end{pmatrix} \in \mathbb{R}^{2n\times 2n}$ and $\begin{pmatrix} \Re B & -\Im B \\ \Im B & \Re B \end{pmatrix} \in \mathbb{R}^{2n\times 2m}$, where $\Re$ and $\Im$ denote the real and imaginary part respectively.
Let us also mention that it is classical that the set of parameter Θ is assumed to be a subset of $\mathbb{C}$. In this paper we only consider the case where $\Theta \subset \mathbb{R}$. This assumption is made, firstly, because it avoids technical difficulties and, secondly, because, as said in the introduction, from [8, 10, 11] typical sets $\Theta \in \mathbb{C}$ for which uniform ensemble controllability can be expected are one dimensional objects.

The following facts can be found in [11].

Proposition 2.3. *If (2.1) is uniformly ensemble controllable, then it is uniformly ensemble controllable in any time $T > 0$, i.e., for any $x^0, x^1 \in \mathcal{C}(\Theta)^n$, any $T > 0$ and any $\varepsilon > 0$, there exists a control $u \in L^2(0,T)^m$, such that $\sup_{\theta\in\Theta} \left|x(T,\theta;x^0,u) - x^1(\theta)\right| \leqslant \varepsilon$.*

Proposition 2.4. *If (2.1) is uniformly ensemble controllable, then for every $\theta \in \Theta$, the pair $(A(\theta), B(\theta))$ is controllable.*

Remark 2.5. The direct consequence of the last result is that the control operator can not vanish in any point, i.e. $B(\theta) \neq 0$ for every parameter θ.

More generally, we have the following exact controllability property.

Proposition 2.6. *Assume (2.1) is uniformly ensemble controllable. For every $K \in \mathbb{N}^*$ and every $\theta_1, \dots, \theta_K \in \Theta$, two by two distinct, the pair $(\mathbb{A}, \mathbb{B})$ is controllable, where*

$$\mathbb{A} = \begin{pmatrix} A(\theta_1) & & \\ & \ddots & \\ & & A(\theta_K) \end{pmatrix} \in \mathbb{R}^{Kn\times Kn}, \quad \text{and} \quad \mathbb{B} = \begin{pmatrix} B(\theta_1) \\ \vdots \\ B(\theta_K) \end{pmatrix} \in \mathbb{R}^{Kn\times m}.$$

Let us also mention the unique continuation property for the adjoint system (cf. [10]). To this end, we denote by $\mathcal{M}(\Theta)$ the set of Radon measures supported in Θ, and we define the *control to final state operator* $\Phi_T \in \mathcal{L}(L^2(0,T)^m, \mathcal{C}(\Theta)^n)$ by

$$(\Phi_T u)(\theta) = \int_0^T e^{(T-t)A(\theta)} B(\theta) u(t)\,\mathrm{d}t \qquad (\theta \in \Theta,\ u \in L^2(0,T)^m), \tag{2.3}$$

i.e., $(\Phi_T u)(\theta) = x(T,\theta,0,u)$ with x given by (2.2). When there is no fear of ambiguity, we shall use abbreviation $x^T = \Phi_T u$.

Note that the adjoint of Φ_T is $\Phi_T^* \in \mathcal{L}(\mathcal{M}(\Theta)^n, L^2(0,T)^m)$ defined by

$$(\Phi_T^* \mu)(t) = \int_\Theta B(\theta)^* e^{(T-t)A(\theta)^*}\,\mathrm{d}\mu(\theta) = \left\langle e^{(T-t)A^*} \flat \mu, B \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^{n\times m}}$$
$$(t \in (0,T),\ \mu \in \mathcal{M}(\Theta)^n). \tag{2.4}$$

Proposition 2.7. *The system (2.1) is uniformly ensemble controllable in time $T > 0$ if and only if*

$$\left\{\mu \in \mathcal{M}(\Theta)^n \;\middle|\; \forall t \in [0,T],\ \int_\Theta B(\theta)^* e^{(T-t)A(\theta)^*}\,\mathrm{d}\mu(\theta) = 0\right\} = \{0\}. \tag{2.5}$$

By the last proposition the uniform ensemble controllability implies the unique continuation property of the adjoint operator, i.e.,

$$\Phi_T^*\mu = 0 \Longrightarrow \mu = 0. \tag{2.6}$$

In particular, the Gramian operator defined as $\Lambda_T := \Phi_T\Phi_T^* \in \mathcal{L}(\mathcal{M}(\Theta)^n; \mathcal{C}(\Theta)^n)$ is injective.

3 Minimal L^2-norm controls

3.1 Optimality conditions

The aim of this section is to characterize the minimal L^2-norm control steering the system (2.1) from x^0 to a distance $\varepsilon > 0$ of x^1 in a given time $T > 0$. More precisely, given $T > 0$, $\varepsilon > 0$ and $x^0, x^1 \in \mathcal{C}(\Theta)^n$, we aim to find a minimizer of

$$\left|\begin{array}{ll} \min & \frac{1}{2}\|u\|^2_{L^2(0,T)^m}, \\ & u \in L^2(0,T)^m \quad \text{and} \quad \sup\limits_{\theta\in\Theta}\left|y(\theta) - \displaystyle\int_0^T e^{(T-t)A(\theta)}B(\theta)u(t)\,\mathrm{d}t\right| \leqslant \varepsilon. \end{array}\right. \tag{P}$$

where $y \in \mathcal{C}(\Theta)^n$ is the final target adjusted by the solution of the homogeneous equation, i.e.,

$$y(\theta) = x^1(\theta) - e^{TA(\theta)}x^0(\theta) \qquad (\theta \in \Theta).$$

When there is no fear of ambiguity, we shall also refer to y as (adjusted) target.

We rewrite the problem (P) by introducing functions $f \in \mathcal{C}(L^2(0,T)^m)$ and $g : \mathcal{C}(\Theta)^n \to \mathbb{R}_+ \cup \{\infty\}$ defined as,

$$\begin{aligned} f(u) &= \tfrac{1}{2}\|u\|^2_{L^2(0,T)^m} && (u \in L^2(0,T)^m), \\ g(\xi) &= \begin{cases} 0 & \text{if } \|\xi - y\|_{L^\infty(\Theta)^n} \leqslant \varepsilon, \\ \infty & \text{otherwise} \end{cases} && (\xi \in \mathcal{C}(\Theta)^n). \end{aligned}$$

In particular, g is the indicator function of the ball $\mathcal{B}_\infty(y,\varepsilon) \subset \mathcal{C}(\Theta)^n$.

Hence, (P) can be expressed as

$$\min_{u\in L^2(0,T)^m} f(u) + g(\Phi_T u).$$

Note that f and g are convex functions, and our aim is to approach the optimization problem (P) by the Fenchel-Rockafellar duality. To this end, we consider the Fenchel conjugates of f and g defined as

$$\begin{aligned} f^*(v) &= \sup_{u\in L^2(0,T)^m}\Big(\langle v,u\rangle_{L^2(0,T)^m} - f(u)\Big) = \tfrac{1}{2}\|v\|^2_{L^2(0,T)^m} && (v \in L^2(0,T)^m), \\ g^*(\mu) &= \sup_{\xi\in\mathcal{C}(\Theta)^n}\Big(\langle \mu,\xi\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} - g(\xi)\Big) = \sup_{\substack{\xi\in\mathcal{C}(\Theta)^n \\ \|\xi-y\|_{L^\infty(\Theta)^n}\leqslant\varepsilon}} \langle \mu,\xi\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} \\ &= \langle \mu, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} + \varepsilon\|\mu\|_{\mathcal{M}(\Theta)^n} && (\mu \in \mathcal{M}(\theta)^n). \end{aligned}$$

Here, and in the rest of this paper, for $\mu = (\mu_1, \dots, \mu_m) \in \mathcal{M}(\Theta)^n$, we set

$$\|\mu\|_{\mathcal{M}(\Theta)^n} = \sup_{\substack{\xi \in \mathcal{C}(\Theta)^n \\ \|\xi\|_{L^\infty(\Theta)^n} \leqslant 1}} \langle \mu, \xi \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} = \sup_{\substack{\xi_1, \dots, \xi_n \in \mathcal{C}(\Theta) \\ \sup_\Theta \sum_{i=1}^n |\xi_i|^2 \leqslant 1}} \sum_{i=1}^n \langle \mu_i, \xi_i \rangle_{\mathcal{M}(\Theta), \mathcal{C}(\Theta)} .$$

Specially, if the components of μ are written in the form $\mu_i = f_i \flat \nu$ for some i-independent scalar and positive measure ν, then $\|\mu\|_{\mathcal{M}(\Theta)^n} = \left\langle \left(\sqrt{\sum_i f_i^2} \right) \flat \nu, 1 \right\rangle_{\mathcal{M}(\Theta), \mathcal{C}(\Theta)} = \int_\Theta |f(\theta)| \, \mathrm{d}\nu(\theta)$. In the special case of μ being a combination of disjoint Dirac masses $\mu = \sum_k m_k \delta_{\theta_k}$ for $m_k \in \mathbb{R}^n$, its norm is given by $\sum_k |m_k|$.

Theorem 3.1. *Assume (2.1) is uniformly ensemble controllable. Then for every $y \in \mathcal{C}(\Theta)^n$ and every $\varepsilon > 0$, the minimization problem (P) admits a unique minimizer $u_{opt} \in L^2(0,T)^m$ given by the expression*

$$u_{opt} = \Phi_T^* \mu_{opt},$$

where $\mu_{opt} \in \mathcal{M}(\Theta)^n$ minimizes the dual functional

$$J(\mu) = \tfrac{1}{2} \|\Phi_T^* \mu\|^2_{L^2(0,T)^m} - \langle \mu, y \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} + \varepsilon \|\mu\|_{\mathcal{M}(\Theta)^n} \qquad (\mu \in \mathcal{M}(\Theta)^n). \tag{3.1}$$

Furthermore, the minimizer μ_{opt} is unique, and satisfies

$$\|u_{opt}\|^2_{L^2(0,T)^m} = \langle \mu_{opt}, y \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} - \varepsilon \|\mu_{opt}\|_{\mathcal{M}(\Theta)^n} , \tag{3.2}$$

$$\operatorname{supp} \mu_{opt} \subset \{\theta \in \Theta \;\mid\; |(\Phi_T u_{opt})(\theta) - y(\theta)| = \varepsilon\} \tag{3.3}$$

and, for every test function $v \in \mathcal{C}(\Theta)^n$ such that $\Phi_T u_{opt} + v \in \mathcal{B}_\infty(y, \varepsilon)$,

$$\langle \mu_{opt}, v \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \geqslant 0. \tag{3.4}$$

Finally, μ_{opt} is the zero measure if and only if $y \leqslant \varepsilon$.

Remark 3.2. As $|(\Phi_T u_{opt})(\theta) - y(\theta)|$ is the distance of the optimal final state from the target x^1, the characterization (3.3) restricts the support of the optimal measure to the parameters $\theta \in \Theta$ for which the optimal final state lies exactly on the boundary of the target ball $\mathcal{B}_\infty(x^1, \varepsilon)$.

Proof. We first observe that f is strictly convex, and $\operatorname{dom} g$ is not empty. Furthermore, we observe that $\left\{u \in L^2(0,T)^m \;\mid\; \Phi_T u \in \operatorname{dom} g\right\}$ is a convex set and g is lower semi-continuous. These facts ensure the existence and uniqueness of a minimizer of (P) (see e.g. [24, Theorem 2.19]). Note also that $\operatorname{cont} g = \operatorname{int} \operatorname{dom} g$ and $\operatorname{dom} f = L^2(0,T)^m$, hence, the uniform ensemble controllability assumption guaranties that $\Phi_T \operatorname{dom} f \cap \operatorname{cont} g \neq \emptyset$. Thus, according to Fenchel duality (see e.g. [5, Theorem 4.4.3], [24, Theorem 3.51] or [12]), we have

$$\begin{aligned} \min_{u \in L^2(0,T)^m} f(u) + g(\Phi_T u) &= \max_{\mu \in \mathcal{M}(\Theta)^n} -f^*(\Phi_T^* \mu) - g^*(-\mu) = - \min_{\mu \in \mathcal{M}(\Theta)^n} f^*(\Phi_T^* \mu) + g^*(-\mu) \\ &= - \min_{\mu \in \mathcal{M}(\Theta)^n} \tfrac{1}{2} \|\Phi_T^* \mu\|^2_{L^2(0,T)^m} - \langle \mu, y \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} + \varepsilon \|\mu\|_{\mathcal{M}(\Theta)^n} . \end{aligned}$$

In addition (3.1) admits a minimizer $\mu_{opt} \in \mathcal{M}(\Theta)^n$, and we have

$$f(u_{opt}) + g(\Phi_T u_{opt}) = -f^*(\Phi_T^* \mu_{opt}) - g^*(-\mu_{opt}) \tag{3.5}$$

and

$$\Phi_T^* \mu_{opt} \in \partial f(u_{opt}) \quad \text{and} \quad -\mu_{opt} \in \partial g(\Phi_T u_{opt}),$$

where ∂ stands for the sub-differential. As f is a smooth function, its sub-differential is a singleton, and from the first of the above inclusions we get that $u_{opt} = \Phi_T^* \mu_{opt}$.

Noticing that $f^* = f$, $g(\Phi_T u_{opt}) = 0$, we get from (3.5), $2f(u_{opt}) = -g^*(-\mu_{opt})$, which leads to (3.2).

We also have that $x_{opt}^T := \Phi_T u_{opt} \in \mathcal{C}(\Theta)^n$ is such that $\left\|x_{opt}^T - y\right\|_{L^\infty(\Theta)^n} \leqslant \varepsilon$, i.e., $g(x_{opt}^T) = 0$. Then, $-\mu_{opt} \in \partial g(x_{opt}^T)$ implies that for every $v \in \mathcal{C}(\Theta)^n$, $g(x_{opt}^T + v) = g(x_{opt}^T + v) - g(x_{opt}^T) \geqslant \langle -\mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$. In particular, if $\left\|x_{opt}^T + v - y\right\|_{L^\infty(\Theta)^n} \leqslant \varepsilon$, we have $0 \leqslant \langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$, thus proving (3.4).

From here we obtain the characterization (3.3) of the support of μ_{opt}. Indeed, for every $v \in \mathcal{C}(\Theta)^n$ such that $\operatorname{supp} v \subset \left\{\theta \in \Theta \mid \left|x_{opt}^T(\theta) - y(\theta)\right| < \varepsilon\right\}$, there exists $\alpha > 0$ such that $\left\|x_{opt}^T + \alpha v - y\right\|_{L^\infty(\Theta)^n} < \varepsilon$ and $\left\|x_{opt}^T - \alpha v - y\right\|_{L^\infty(\Theta)^n} < \varepsilon$. Hence, we shall have $0 \leqslant \langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$ and $0 \leqslant -\langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$, leading to $0 = \langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$.

Dealing with the uniqueness of μ_{opt}, assume by contradiction that μ_1 and μ_2 are minimizer of J. By the uniqueness of u_{opt}, we have $u_{opt} = \Phi_T^* \mu_1 = \Phi_T^* \mu_2$. According to Proposition 2.7, we conclude that $\mu_1 = \mu_2$.

Finally, if $\|y\|_{\mathcal{C}(\Theta)^n} \leqslant \varepsilon$ (meaning that the free dynamics brings the state within the target ball) by the definition (3.1) of the functional J and the Cauchy-Schwarz inequality, we have that for every measure $\mu \in \mathcal{M}(\Theta)^n$

$$J(\mu) \geqslant -\|\mu\|_{\mathcal{M}(\Theta)^n} \|y\|_{\mathcal{C}(\Theta)^n} + \varepsilon \|\mu\|_{\mathcal{M}(\Theta)^n} \geqslant 0 = J(0),$$

i.e., the functional J is nonnegative and $\mu_{opt} = 0$ is its unique minimizer.

The reverse implication holds trivially, as if $\|y\|_{\mathcal{C}(\Theta)^n} > \varepsilon$ then a nonzero control is required in order to satisfy the constraint (of reaching the target ball). Using classical optimization techniques arguments, it can be even shown that in that case $\|\mu_{opt}\|_{\mathcal{M}(\Theta)^n}$ is bounded from below by a positive constant depending linearly on the difference $\|y\|_{\mathcal{C}(\Theta)^n} - \varepsilon$ (e.g. [17, Proposition 2.3]). □

Proposition 3.3. *Using the assumptions and notations introduced in Theorem 3.1, the optimal measure μ_{opt} is of the form*

$$\mathrm{d}\mu_{opt}(\theta) = \alpha(\theta)\left(y(\theta) - x_{opt}^T(\theta)\right) \mathrm{d}\left|\mu_{opt}\right|(\theta), \tag{3.6}$$

where $|\mu_{opt}| = \sum_{i=1}^n |\mu_{opt,i}| \in \mathcal{M}(\Theta)$ is the sum of total variations of components of the vector measure μ_{opt}, while $\alpha : \Theta \to \mathbb{R}_+$ is an integrable (with respect to $|\mu_{opt}|$) nonnegative scalar function.

Proof. **Step 1.** We first show that

$$\langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \geqslant 0, \tag{3.7}$$

for every test function $v \in \mathcal{C}(\Theta)^n$ such that $\left\langle v(\theta), y(\theta) - x_{opt}^T(\theta)\right\rangle \geqslant 0$ whenever $\left|(x_{opt}^T)(\theta) - y(\theta)\right| = \varepsilon$, $\theta \in \Theta$.

Suppose first that v satisfies the last inequality strictly, i.e., that $\left\langle v(\theta), y(\theta) - x_{opt}^T(\theta)\right\rangle > 0$ for every $\theta \in \Theta$ satisfying $|x_{opt}(\theta) - y(\theta)| = \varepsilon$. Then we show there exists a θ independent scalar $\alpha > 0$ small enough such that $\|x_{opt} + \alpha v - y\|_{L^\infty(\Theta)^n} \leqslant \varepsilon$. Indeed, let us set $\Theta_{opt} = \{\theta \in \Theta \mid |x_{opt}(\theta) - y(\theta)| = \varepsilon\}$, since $\theta \mapsto |x_{opt}(\theta) - y(\theta)|$ is continuous, we have that Θ_{opt} is compact. On this set, we have $\langle v(\theta), y(\theta) - x_{opt}(\theta)\rangle > 0$, using the uniform continuity on Θ of this function, we get the existence of $\delta > 0$ such that $\langle v(\theta), y(\theta) - x_{opt}(\theta)\rangle > 2\delta$ for every $\theta \in \Theta_{opt}$, and there also exists $\eta > 0$ such that $\langle v(\theta), y(\theta) - x_{opt}(\theta)\rangle > \delta$ for every $\theta \in \Theta_{opt} + B(0, \eta)$. Using the fact that $\Theta \setminus (\Theta_{opt} + B(0, \eta))$ is compact, we also obtain the existence of $\widetilde{\varepsilon} < \varepsilon$ such that $|x_{opt}(\theta) - y(\theta)| \leqslant \widetilde{\varepsilon}$ for every $\theta \in \Theta \setminus (\Theta_{opt} + B(0, \eta))$. All in all, we have for every $\alpha > 0$,

- for $\theta \in \Theta_{opt} + B(0,\eta)$,

$$|x_{opt}(\theta) + \alpha v(\theta) - y(\theta)|^2 = |x_{opt}(\theta) - y(\theta)|^2 - 2\alpha \langle v(\theta), y(\theta) - x_{opt}(\theta)\rangle + \alpha^2 |v(\theta)|^2 \\ \leqslant \varepsilon^2 - 2\alpha\delta + \alpha^2 \|v\|^2_{L^\infty(\Theta)^n} .$$

- for $\theta \in \Theta \setminus (\Theta_{opt} + B(0,\eta))$,

$$\left|x^T_{opt}(\theta) + \alpha v(\theta) - y(\theta)\right|^2 \leqslant \widetilde{\varepsilon}^2 + 2\alpha\widetilde{\varepsilon}\,\|v\|_{L^\infty(\Theta)^n} + \alpha^2 \|v\|^2_{L^\infty(\Theta)^n} .$$

Hence, taking $\alpha > 0$ small enough, we have $\left|x^T_{opt}(\theta) + \alpha v(\theta) - y(\theta)\right| \leqslant \varepsilon$ for every $\theta \in \Theta$. By (3.4) this ensures that $\langle \mu_{opt}, v\rangle_{\mathcal{M}(\theta)^n, \mathcal{C}(\Theta)^n} \geqslant 0$ for every $v \in \mathcal{C}(\Theta)^n$ such that $\left\langle v(\theta), y(\theta) - x^T_{opt}(\theta)\right\rangle > 0$ for every $\theta \in \Theta_{opt}$.

It remains to prove that (3.7) holds true, even if $\left\langle v(\theta), y(\theta) - x^T_{opt}(\theta)\right\rangle \geqslant 0$ for every $\theta \in \Theta_{opt}$. In fact, if v satisfies the last condition, then $\left\langle v(\theta) + \alpha\left(y(\theta) - x^T_{opt}(\theta)\right), y(\theta) - x^T_{opt}(\theta)\right\rangle > 0$ for every $\theta \in \Theta_{opt}$ and every $\alpha > 0$. This ensures that $0 \leqslant \left\langle \mu_{opt}, v(\theta) + \alpha\left(y(\theta) - x^T_{opt}(\theta)\right)\right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$, and hence, taking the limit $\alpha \to 0$, we obtain the desired result.

Step 2. Now we prove the relation (3.6)

First, let us note that μ_{opt} is absolutely continuous with respect to the measure $|\mu_{opt}|$. By the Lebesgue-Radon-Nikodym theorem (see e.g. [14]) there exists an integrable (with respect to $|\mu_{opt}|$) vector function φ such that $\mathrm{d}\mu_{opt}(\theta) = \varphi(\theta)\,\mathrm{d}|\mu_{opt}|(\theta)$.

It remains to prove that $\varphi = \alpha(y - x^T_{opt})$ for some $\alpha : \Theta \to \mathbb{R}_+$ which is integrable (with respect to $|\mu_{opt}|$). To this end, without loosing generality, let us write φ as

$$\varphi(\theta) = \alpha(\theta)(y(\theta) - x^T_{opt}(\theta)) + w(\theta), \tag{3.8}$$

where $w(\theta) \in \left\{y(\theta) - x^T_{opt}(\theta)\right\}^\perp$, while α is an arbitrary (not necessarily non-negative) scalar function. Without loss of generality, we can assume that $\operatorname{supp}\varphi \subset \operatorname{supp}\mu_{opt}$, i.e., we can assume that α and w vanish on $\Theta\setminus\operatorname{supp}\mu_{opt}$. Taking into account that $y - x^T_{opt}$ is continuous on Θ and that $\left|y - x^T_{opt}\right| = \varepsilon$ on $\operatorname{supp}\mu_{opt}$, we get that $\alpha(\theta) = \left\langle y(\theta) - x^T_{opt}(\theta), \varphi(\theta)\right\rangle/\varepsilon^2$ for every $\theta \in \operatorname{supp}\mu_{opt}$. This ensures that α is $|\mu_{opt}|$-measurable, and then $w = \varphi - \alpha\left(y - x^T_{opt}\right)$ is also $|\mu_{opt}|$-measurable.

Define $v = \psi\left(y - x^T_{opt}\right)$, where $\psi \in \mathcal{C}(\Theta)$ is an arbitrary non-negative, scalar test function. v is therefore an eligible test function in (3.7), implying

$$0 \leqslant \langle \mu_{opt}, v\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} = \int_\Theta \left\langle \left(\alpha(\theta)(y(\theta) - x^T_{opt}(\theta)) + w(\theta)\right), \psi(\theta)\left(y(\theta) - x^T_{opt}(\theta)\right)\right\rangle \mathrm{d}\,|\mu_{opt}|\,(\theta) \\ = \int_\Theta \psi(\theta)\left|y(\theta) - x^T_{opt}(\theta)\right|^2\ \alpha(\theta)\,\mathrm{d}\,|\mu_{opt}|\,(\theta) = \left\langle \alpha_\flat\,|\mu_{opt}|, \psi\left|y - x^T_{opt}\right|^2\right\rangle_{\mathcal{M}(\Theta), \mathcal{C}(\Theta)} .$$

As $|\mu_{opt}|$ is a non-negative measure, by varying ψ and taking again into account that $\left|y - x^T_{opt}\right| = \varepsilon$ on $\operatorname{supp}\mu_{opt}$, it follows $\alpha \geqslant 0$ for $|\mu_{opt}|$-a.e. θ. Taking $\psi = 1$, we also get that α is $|\mu_{opt}|$-integrable.

Furthermore, (3.2) together with $\|u_{opt}\|^2_{L^2(0,T)^m} = \|\Phi^*_T\mu_{opt}\|^2_{L^2(0,T)^m} = \left\langle \mu_{opt}, x^T_{opt}\right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$ leads to

$$\varepsilon\,\|\mu_{opt}\|_{\mathcal{M}(\Theta)^n} = \left\langle \mu_{opt}, y - x^T_{opt}\right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} .$$

At this stage we exploit the form (3.8) and put $\mu_{opt} = \left(\alpha(y - x^T_{opt}) + w\right)_\flat\,|\mu_{opt}|$ in the last expression. By taking into account that $w(\theta)$ is orthogonal to $y(\theta) - x^T_{opt}(\theta)$, and recalling that

$\left|y - x_{opt}^T\right| = \varepsilon$ on $\operatorname{supp}\mu_{opt}$, we get

$$\varepsilon \int_\Theta \left|\alpha(\theta)(y(\theta) - x_{opt}^T(\theta)) + w(\theta)\right| \mathrm{d}\left|\mu_{opt}\right|(\theta) = \int_\Theta \alpha(\theta)\left|(y(\theta) - x_{opt}^T(\theta))\right|^2 \mathrm{d}\left|\mu_{opt}\right|(\theta)$$
$$= \varepsilon \int_\Theta \alpha(\theta)\left|(y(\theta) - x_{opt}^T(\theta))\right| \mathrm{d}\left|\mu_{opt}\right|(\theta) \quad (3.9)$$

As α is a non-negative function and $\left|\alpha(y - x_{opt}^T) + w\right| \geqslant \alpha\left|y - x_{opt}^T\right|$, the equality (3.9) implies $\left|\alpha(y - x_{opt}^T) + w\right| = \alpha\left|y - x_{opt}^T\right|$ for $|\mu_{opt}|$-a.e. θ, that is to say that $w = 0$ for $|\mu_{opt}|$ (and consequently for μ_{opt}) a.e. θ. □

Remark 3.4. The last result covers the well-known behavior of the optimal control in the deterministic case, i.e., when Θ is a singleton. In that case, unless the optimal control is trivial, the solution of the dual problem is a vector having the same direction as $y - x_{opt}^T$ (see e.g. [19]).
In general, when a functional of the form (3.1) is defined on a Hilbert space, its minimizer h_{opt} is either trivial or a solution to the Euler-Lagrange equation which formally reads as:

$$y - x_{opt}^T = \varepsilon \frac{h_{opt}}{\|h_{opt}\|}.$$

In that case it follows directly that the minimizer has the same direction as $y - x_{opt}^T$. This kind of result is stronger than the one obtained in the last proposition, and it corresponds to a constant function α in (3.6). The reason why such approach is not possible in our setting is that the norm term $\|\mu\|_{\mathcal{M}(\Theta)^n}$ entering the functional J is not differentiable.

As a consequence of the last proposition, we obtain the following characterization of the optimal measure in the scalar case.

Corollary 3.5. *Using the assumptions and notations introduced in Theorem 3.1, let us suppose* $n = m = 1$.

Then the control operator does not change sign, i.e., there exists $\epsilon \in \{-1, 1\}$ *and* $\beta \in \mathcal{C}(\Theta)$ *such that* $\beta(\theta) > 0$ *and* $B(\theta) = \epsilon\beta(\theta)$ *for every* $\theta \in \Theta$.

Set $\Theta_+ = \left\{\theta \in \Theta \mid y(\theta) - x_{opt}^T(\theta) = \varepsilon\right\}$ *and* $\Theta_- = \left\{\theta \in \Theta \mid y(\theta) - x_{opt}^T(\theta) = -\varepsilon\right\}$. *Then* $\operatorname{supp}\mu_{opt} \subset \Theta_- \cup \Theta_+$ *and* μ_{opt} *is nonnegative (respectively, nonpositive) on* Θ_+ *(respectively* Θ_-*).*

In addition, if $\Theta_- = \emptyset$ *(respectively* $\Theta_+ = \emptyset$*), then either* $u_{opt} \equiv 0$ *or* $\operatorname{sign}(u_{opt}(t)) = \epsilon$ *(respectively* $\operatorname{sign}(u_{opt}(t)) = -\epsilon$*) for every* $t \in [0, T]$.

Proof. The fact that $B \in \mathcal{C}(\Theta)$ does not vanish directly follows from Proposition 2.4. Being a continuous function, its sign is constant.
The support and sign conditions of μ_{opt} directly follow from (3.3) and Proposition 3.3, respectively. For the sign of u_{opt}, let us assume that $\Theta_- = \emptyset$ (the proof for other case is identical). We have for every $t \in [0, T]$,

$$u_{opt}(t) = \int_\Theta e^{(T-t)A^*(\Theta)} B(\theta)\, \mathrm{d}\mu_{opt}(\theta) = \epsilon \int_{\Theta_+} e^{(T-t)A^*(\Theta)} \beta(\theta)\, \mathrm{d}\mu_{opt}(\theta),$$

We conclude by noticing that, for every $\theta \in \Theta$, $e^{(T-t)A^*(\Theta)}\beta(\theta) > 0$ and μ_{opt} is nonnegative. In particular, $u_{opt} \equiv 0$ if and only if $\mu_{opt} = 0$ and, according to Theorem 3.1, this holds if and only if $\|y\|_{\mathcal{C}(\Theta)^n} \leqslant \varepsilon$. Otherwise, the optimal control has a constant sign equal to ϵ for every $t \in [0, T]$. □

The rest of this subsection provides results by which one can construct a family of optimization problem (P) for which the minimizers of (3.1) can be explicitly calculated. This will allow us to

compare the results obtained by the numerical procedure to be presented in the Section 6 with the exact ones and to verify efficiency and correctness of the proposed algorithms (see Section 7).

The idea is to invert the problem: we specify a positive measure $\widetilde{\mu}$ and calculate $\widetilde{x}^T = \Phi_T \Phi_T^* \widetilde{\mu}$, the final state of the system run by the control $\widetilde{u} = \Phi_T^* \widetilde{\mu}$. The adjusted target y is then carefully chosen as a center of a square whose one vertex is $\widetilde{x}^T$ (note that the geometrical representation of a ball in $\mathcal{C}(\Theta)^n$ is actually a square). In particular, in accordance to Proposition 3.3, the difference $y - \widetilde{x}^T$ has the same direction as $\widetilde{\mu}$ for every parameter value θ.

Proposition 3.6. *Using the assumptions and notations introduced in Theorem 3.1 take $\varepsilon > 0$, $\eta \in \mathcal{M}(\Theta)$ a nonnegative scalar measure, and $f \in \mathcal{C}(\Theta)^n$ such that $|f|$ does not vanish. Define $\widetilde{\mu} = f_\flat \eta$, $\widetilde{x}^T = \Phi_T \Phi_T^* \widetilde{\mu}$ and $y = \widetilde{x}^T + \varepsilon \frac{f}{|f|}$. Then $\widetilde{\mu}$ is the minimizer of J defined by (3.1).*

Proof. Let us first observe that for every $\mu \in \mathcal{M}(\Theta)^n$, we have

$$\begin{aligned} J(\mu) &= \tfrac{1}{2} \left\| \Phi_T^* \mu \right\|^2_{L^2(0,T)^m} - \langle \mu, y \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} + \varepsilon \left\| \mu \right\|_{\mathcal{M}(\Theta)^n} \\ &= \tfrac{1}{2} \left\| \Phi_T^* \mu \right\|^2_{L^2(0,T)^m} - \langle \Phi_T^* \mu, \Phi_T^* \widetilde{\mu} \rangle_{L^2(0,T)^m} + \varepsilon \left(\left\| \mu \right\|_{\mathcal{M}(\Theta)^n} - \left\langle \mu, \tfrac{f}{|f|} \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \right) \end{aligned}$$

and

$$J(\widetilde{\mu}) = -\tfrac{1}{2} \left\| \Phi_T^* \widetilde{\mu} \right\|^2_{L^2(0,T)^m} + \varepsilon \left(\int_\Theta |f| \, \mathrm{d}\eta - \int_\Theta \frac{\langle f, f \rangle}{|f|} \, \mathrm{d}\eta \right) = -\tfrac{1}{2} \left\| \Phi_T^* \widetilde{\mu} \right\|^2_{L^2(0,T)^m}.$$

Let $\mu_{opt} \in \mathcal{M}(\Theta)^n$ be the minimizer of J, we have $J(\mu_{opt}) \leqslant J(\widetilde{\mu})$, that is to say that

$$\tfrac{1}{2} \left\| \Phi_T^* (\mu_{opt} - \widetilde{\mu}) \right\|^2_{L^2(0,T)^m} \leqslant -\varepsilon \left(\left\| \mu_{opt} \right\|_{\mathcal{M}(\Theta)^n} - \left\langle \mu_{opt}, \tfrac{f}{|f|} \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \right).$$

Obviously, we have $\left\| \mu_{opt} \right\|_{\mathcal{M}(\Theta)^n} \geqslant \left\langle \mu_{opt}, \tfrac{f}{|f|} \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$, from which we can conclude that we necessarily have $\Phi_T^* (\mu_{opt} - \widetilde{\mu}) = 0$. We conclude that $\mu_{opt} = \widetilde{\mu}$ using the unique continuation property (2.6) of the adjoint. □

As a direct corollary, we have the following result.

Corollary 3.7. *Using the assumptions and notations introduced in Theorem 3.1. Suppose $n = 1$, let $\varepsilon > 0$, $\epsilon \in \{-1, 1\}$, $\eta \in \mathcal{M}(\Theta)$ a nonnegative measure, $y = \epsilon(\Phi_T \Phi_T^* \eta + \varepsilon)$. The minimizer of (3.1) is $\epsilon\eta$.*

Proof. This is Proposition 3.6 with $f = \epsilon$. □

From Proposition 3.6 and Corollary 3.7, we see that there is no reason that the optimal measure μ_{opt} is sparse, i.e., the Lebesgue measure of its support is in general not zero. This might look surprisingly at the first glance, as μ_{opt} is a minimizer of the functional containing the $\|\cdot\|_{\mathcal{M}(\Theta)^n}$ term. This non-sparsity pitfall will be discussed in Section 4 in the case where the parameter set is of a finite cardinality.

3.2 Continuity properties of the minimizers

The characterization of the solution through the adjoint problem, that is obtained in the previous subsection, allows us also to demonstrate the continuity of solutions to optimization problem (P) with respect to the tolerance ε, the time T and the adjusted target y. More precisely, we consider a sequence of functionals

$$J_k(\mu) = \tfrac{1}{2} \left\| \Phi_{T_k}^* \mu \right\|^2_{L^2(0,T)^m} - \langle \mu, y_k \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} + \varepsilon_k \left\| \mu \right\|_{\mathcal{M}(\Theta)^n} \qquad (\mu \in \mathcal{M}(\Theta)^n). \tag{3.10}$$

where:

(i) $y_k \to y$ strongly in $\mathcal{C}(\Theta)^n$,
(ii) $T_k \to T$ in $\mathbb{R}_+^*$
(iii) $\varepsilon_k \to \varepsilon$ in $\mathbb{R}_+^*$.

By Theorem 3.1, for each $k \in \mathbb{N}$, the functional J_k admits the unique minimizer μ_k satisfying the relation

$$\left\|\Phi_{T_k}^* \mu_k\right\|_{L^2(0,T_k)^m}^2 = \langle \mu_k, y_k \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} - \varepsilon_k \left\|\mu_k\right\|_{\mathcal{M}(\Theta)^n} \tag{3.11}$$

and the control $u_k = \Phi_{T_k}^* \mu_k$ is the solution of the optimal ensemble control problem

$$\begin{array}{l} \min \quad \frac{1}{2}\left\|u\right\|_{L^2(0,T)^m}^2, \\ \quad \big| \quad u \in L^2(0,T)^m \quad \text{and} \quad \left\|y_k - \Phi_{T_k} u\right\|_{L^\infty(\Theta)^n} \leqslant \varepsilon_k. \end{array} \tag{P_k}$$

We are going to prove that $(\mu_k)_k$ and $(u_k)_k$ converge respectively vaguely and strongly (in the appropriate spaces) to the limit points μ and u, respectively, where μ is the minimizer of the limit functional (3.1) and u is the solution of the ensemble optimal control problem (P).

Note that the convergence of time horizons T_k implies strong convergence of the operators Φ_{T_k} and $\Phi_{T_k}^*$. The following technical result characterizes strong convergence of arbitrary operators on Banach spaces.

Lemma 3.8. *Let $(L_k)_k$ be a sequence of linear operators in $\mathcal{L}(X,Y)$, with X and Y being Banach spaces. The following statements are equivalent.*

(a) $L_k \longrightarrow L$ strongly in $\mathcal{L}(X,Y)$, i.e.,

$$(\forall\, u \in X) \quad L_k u \longrightarrow Lu \quad \text{in } Y.$$

(b) For every weakly $*$ converging sequence $(v_k)_k$ in Y' it holds

$$v_k \overset{*}{\rightharpoonup} v \quad \Longrightarrow \quad L_k^* v_k \overset{*}{\rightharpoonup} L^* v \quad \text{in } X'.$$

(c) For every strongly converging sequence $(u_k)_k$ in X it holds

$$u_k \longrightarrow u \quad \Longrightarrow \quad L_k u_k \longrightarrow Lu \quad \text{in } Y.$$

Proof. (a) $\Rightarrow$ (b): Let $(v_k)_k$ converge weakly $*$ to $v \in Y'$. By employing the strong convergence of the operators L_k, for every $u \in X$ we get

$$\langle L_k^* v_k, u \rangle_{X',X} = \langle v_k, L_k u \rangle_{Y',Y} \longrightarrow \langle v, Lu \rangle_{Y',Y} = \langle L^* v, u \rangle_{X',X}\,,$$

i.e., $L_k^* v_k \overset{*}{\rightharpoonup} L^* v$ in X'.

(b) $\Rightarrow$ (c): We first prove the weak convergence of the sequence $(L_k u_k)_k$. To this end, note that for an arbitrary $v \in Y'$ we have

$$\langle v, L_k u_k \rangle_{Y',Y} = \langle L_k^* v, u_k \rangle_{X',X} \longrightarrow \langle L^* v, u \rangle_{X',X} = \langle v, Lu \rangle_{Y',Y}\,.$$

Here we have employed the strong convergence of (u_k) and the assumption (b). In particular, we have

$$\left\|Lu\right\|_Y \leqslant \liminf_{k\to\infty} \left\|L_k u_k\right\|_Y\,. \tag{3.12}$$

On the other hand, for each k there exists $v_k \in Y'$ such that

$$\left\|L_k u_k\right\|_Y^2 = \langle v_k, L_k u_k \rangle_{Y',Y}\,, \tag{3.13}$$

and $\|v_k\|_{Y'} = \|L_k u_k\|_Y$ (see e.g. [6, Corollary 1.3.]).

Due to the weak convergence of $(L_k u_k)_k$, the sequence of norms $\|L_k u_k\|_Y$ is bounded, in sequel implying boundedness of $(v_k)_k$ in Y'. In particular, up to a subsequence, $(v_k)_k$ converge weakly $*$ to some $v \in Y'$, and it holds

$$\|v\|_{Y'} \leqslant \liminf_{k\to\infty} \|v_k\|_{Y'} = \liminf_{k\to\infty} \|L_k u_k\|_Y \,. \tag{3.14}$$

From here it follows

$$\|L_k u_k\|_Y^2 = \langle v_k, L_k u_k\rangle_{Y',Y} = \langle L_k^* v_k, u_k\rangle_{X',X} \longrightarrow \langle L^* v, u\rangle_{X',X} = \langle v, Lu\rangle_{Y',Y} \,, \tag{3.15}$$

where we passed to the limit by using the assumption (b) again.

In particular, by employing (3.15) and (3.14), we obtain

$$\lim_{k\to\infty} \|L_k u_k\|_Y^2 = {}_{Y'}\langle\, v, Lu\,\rangle_Y \leqslant \|v\|_{Y'} \|Lu\|_Y \leqslant \lim_{k\to\infty} \|L_k u_k\|_Y \|Lu\|_Y \,,$$

implying $\lim_{k\to\infty} \|L_k u_k\|_Y \leqslant \|Lu\|_Y$. Together with (3.12) this finishes the proof.

(c) $\Rightarrow$ (a): Trivial (take a constant sequence $u_k = u$). □

Before proving the convergence, we first demonstrate the boundedness properties of the minimizers.

Lemma 3.9. *The sequence $(\mu_k)_k$ of minimizers of functionals J_k given by (3.10) is bounded in $\mathcal{M}(\Theta)^n$.*

Proof. We argue by contradiction and assume that

$$\|\mu_k\|_{\mathcal{M}(\Theta)^n} \longrightarrow \infty \tag{3.16}$$

(up to a non-relabelled subsequence).

Dividing the relation (3.11) by $\|\mu_k\|_{\mathcal{M}(\Theta)^n}$ and denoting $\nu_k = \mu_k / \|\mu_k\|_{\mathcal{M}(\Theta)^n}$ we get

$$\|\mu_k\|_{\mathcal{M}(\Theta)^n} \left\|\Phi_{T_k}^* \nu_k\right\|_{L^2(0,T)^m}^2 = \langle \nu_k, y_k\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} - \varepsilon_k. \tag{3.17}$$

As all the terms on the right-hand side are bounded (with respect to k), the contradictory assumption (3.16) implies

$$\Phi_{T_k}^* \nu_k \to 0 \quad \text{strongly in } L^2(0,T)^m. \tag{3.18}$$

On the other hand, denoting by ν the vague limit of (ν_k) (up to a subsequence), the strong convergence of operators Φ_{T_k} and Lemma 3.8 imply $\Phi_{T_k}^* \nu_k \longrightarrow \Phi_T^* \nu$ weakly in $L^2(0,T)^m$. Together with (3.18), we get $\Phi_T^* \nu = 0$, while the ensemble controllability assumption (2.5) implies $\nu = 0$.

Going back to (3.17) and passing to the limit we get

$$0 > \varepsilon_k - \langle \nu_k, y_k\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \longrightarrow \varepsilon.$$

Here we use that $\langle \nu_k, y_k\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$ is the dual product of a strongly convergent functions with measures converging vaguely to zero, thus vanishing on the limit.

The last relation provides the desired contradiction which, in turn, implies the result. □

Based on the last two lemmas, we are going to obtain the main result of this paragraph.

Proposition 3.10. *The sequence of minimizers μ_k of the functionals J_k given by (3.10) converges vaguely in $\mathcal{M}(\Theta)^n$ to μ_{opt} (the minimizer of the limit functional (3.1)), and $\|\mu_k\|_{\mathcal{M}(\Theta)^n}$ converges to $\|\mu_{opt}\|_{\mathcal{M}(\Theta)^n}$. In addition, the sequence of optimal ensemble controls $u_k = \Phi^*_{T_k}\mu_k$ converges strongly in $L^2(0,T)^m$ to the solution of the limit problem (P).*

Proof. By Lemma 3.9, there exists a (non-relabelled) subsequence of the minimizers μ_k converging vaguely to some measure $\mu_L \in \mathcal{M}(\Theta)^n$.

In the first part of the proof, we show that the limit μ_L is the unique minimizer of the limit functional (3.1). Thus, it suffices to show

$$J(\mu_L) \leqslant J(\mu), \quad \mu \in \mathcal{M}(\Theta)^n. \tag{3.19}$$

To this end, we explore the convergence properties of the sequence of minimal functional values

$$\liminf_{k\to\infty} J_k(\mu_k) = \liminf_{k\to\infty} \Big(\tfrac{1}{2} \left\|\Phi^*_{T_k}\mu_k\right\|^2_{L^2(0,T)^m} - \langle \mu_k, y_k\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} + \varepsilon_k \|\mu_k\|_{\mathcal{M}(\Theta)^n} \Big). \tag{3.20}$$

Due to the weak lower semi-continuity of norms in Banach spaces, the last term is bounded from below by $\varepsilon \|\mu_L\|_{\mathcal{M}(\Theta)^n}$.

The second term on the right-hand side of (3.20) is the product of a strongly and vaguely convergent sequence, thus $\langle \mu_k, y_k\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} \to \langle \mu_L, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n}$.

As for the first term entering functional J_k by Lemma 3.8 and the strong convergence of the operators Φ_{T_k}, we have $\Phi^*_{T_k}\mu_k \longrightarrow \Phi^*_T\mu_L$ weakly in $L^2(0,T)^n$, implying

$$\liminf_{k\to\infty} \left\|\Phi^*_{T_k}\mu_k\right\|^2_{L^2(0,T)^m} \geqslant \|\Phi^*_T\mu_L\|^2_{L^2(0,T)^m}. \tag{3.21}$$

Consequently, from (3.20) we get

$$\liminf_{k\to\infty} J_k(\mu_k) \geqslant \tfrac{1}{2} \|\Phi^*_T\mu_L\|^2_{L^2(0,T)^m} - \langle \mu_L, y\rangle + \varepsilon \|\mu_L\|_{\mathcal{M}(\Theta)^n} = J(\mu_L)$$

As μ_k minimizes J_k, for an arbitrary $\mu \in \mathcal{M}(\Theta)^n$ we have

$$\liminf_{k\to\infty} J_k(\mu) \geqslant J(\mu_L). \tag{3.22}$$

On the other hand, employing the assumed convergences, (i), (ii) and (iii), for each fixed μ we have

$$\begin{aligned} J_k(\mu) &= \tfrac{1}{2} \left\|\Phi^*_{T_k}\mu\right\|^2_{L^2(0,T)^m} - \langle \mu, y_k\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} + \varepsilon_k \|\mu\|_{\mathcal{M}(\Theta)^n} \\ &\to \tfrac{1}{2} \|\Phi^*_T\mu\|^2_{L^2(0,T)^m} - \langle \mu, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} + \varepsilon \|\mu\|_{\mathcal{M}(\Theta)^n} = J(\mu). \end{aligned} \tag{3.23}$$

Together with (3.22) this implies (3.19), and in particular $\mu_k \xrightarrow{*} \mu_{opt}$.

In the second part of the proof, we will show that the sequence $\|\mu_k\|_{\mathcal{M}(\Theta)^n}$ converges to $\|\mu_{opt}\|_{\mathcal{M}(\Theta)^n}$.

Taking $\mu = \mu_{opt}$ in (3.23) we have,

$$J(\mu_{opt}) = \lim_{k\to\infty} J_k(\mu_{opt}) \geqslant \lim_{k\to\infty} J_k(\mu_k).$$

Combining the obtained inequality with (3.22) (remember that $\mu_{opt} = \mu_L$) we get the convergence of the optimal values $J_k(\mu_k) \to J(\mu_{opt})$. Together with (3.21) this convergence implies,

$$\liminf_{k\to\infty} \|\mu_k\|_{\mathcal{M}(\Theta)^n} \leqslant \|\mu_{opt}\|_{\mathcal{M}(\Theta)^n}.$$

which together with the vague convergence of the sequence (μ_k) implies the desired result.

The strong convergence of optimal ensemble controls $u_k = \Phi^*_{T_k}\mu_k$ follows now directly from the strong convergence of the operators Φ_{T_k} and the vague convergence of the measures μ_k (cf. Lemma 3.8 (c)). □

4 Systems with a finite number of parameter values

In this section, we restrict the analysis to the case of a finite number of parameters, i.e., $\operatorname{card}\Theta < \infty$. Such situation deserves a special attention, not only because in some applications the parameter assumes only discrete and bounded values, but also because the most general situations (i.e., those in which the parameter ranges over a continuous set of values) are numerically treated and analyzed as the limit case of discretized problems as the number of elements in the parameter set goes to infinity.

The aim is to provide a geometrical description of the problem and its solution, which will allow us to discuss the conditions under which the optimal measure (i.e., the minimizer of the functional J given by (3.1)) is sparse.

To this end, we assume that the parameter dependent system (2.1) is uniformly ensemble controllable, and that we deal with a parameter set Θ of finite cardinality. As we consider finite dimensional systems only, these assumptions imply that the Gramian operator $\Lambda_T = \Phi_T\Phi_T^* \in \mathcal{L}(\mathcal{M}(\Theta)^n;\mathcal{C}(\Theta)^n)$ is not just injective (due to Proposition 2.7), but also of full range. Indeed, due to the assumed finite cardinality of the parameter set, both spaces $\mathcal{M}(\Theta)^n$ and $\mathcal{C}(\Theta)^n$ are finite dimensional and are isomorphic to $\mathbb{R}^{nK}$ (where $K = \operatorname{card}\Theta \in \mathbb{N}^*$ denotes the number of parameters), thus the statement follows from the rank-nullity theorem.

For $c \geqslant 0$, we introduce the subset $E_c \subset \mathcal{C}(\Theta)^n$ consisting of all the final states x^T that can be reached with a control of norm less or equal to c, i.e.,

$$E_c := \left\{ x^T \in \mathcal{C}(\Theta)^n \;\middle|\; \exists u \in L^2(0,T)^m \text{ such that } \|u\|_{L^2(0,T)^m} \leqslant c \text{ and } x^T = \Phi_T u \right\}.$$

As the minimal norm control reaching any state x^T is of the HUM form $\Phi_T^*\mu$ for some $\mu \in \mathcal{M}(\Theta)$ (see e.g. [29]), the introduced set can be characterized as follows

$$E_c = \left\{ x^T \in \mathcal{C}(\Theta)^n \;\middle|\; \exists \mu \in \mathcal{M}(\Theta)^n \text{ such that } x^T = \Lambda_T\mu \text{ and } \|\Phi_T^*\mu\|_{L^2(0,T)^m} \leqslant c \right\},$$

where Λ_T stands for the Gramian $\Phi_T\Phi_T^*$. As the ensemble controllability assumption ensures the injectivity of the Gramian, the optimal control norm can be rewritten as

$$\|\Phi_T^*\mu\|^2_{L^2(0,T)^m} = \langle \mu, \Phi_T\Phi_T^*\mu\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} = \left\langle \Lambda_T^{-1}x^T, x^T\right\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n},$$

leading to the following characterization of the set E_c:

$$E_c = \left\{ x^T \in \mathcal{C}(\Theta)^n \;\middle|\; \left\langle \Lambda_T^{-1}x^T, x^T\right\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} \leqslant c^2 \right\}. \tag{4.1}$$

In other words, E_c is a sublevel set of a smooth function $F(x^T) = \left\langle \Lambda_T^{-1}x^T, x^T\right\rangle$. In particular, $\{E_c,\ c \geqslant 0\}$ is a nested family of closed, convex sets with smooth boundary $\partial E_c = \{x^T \mid F(x^T) = c^2\}$. The normal functional at the point $x^T \in \partial E_c$ is $n_{E_c}(x^T) = \nabla F(x^T) = 2\Lambda_T^{-1}x^T$.

Theorem 4.1. *Let $x^T, x^1 \in \mathcal{C}(\Theta)^n$ be such that $\left|x^1(\theta) - x^T(\theta)\right| \leqslant \varepsilon$ for every $\theta \in \Theta$. Then x^T is the optimal final state of the problem (P) with the target x^1 and initial datum $x^0 = 0$ if and only if the normal functional $n_{E_c}(x^T)$ is of the form*

$$n_{E_c}\left(x^T\right) = \left(\alpha(x^1 - x^T)\right)\flat\nu, \tag{4.2}$$

where $\alpha \in \mathcal{C}(\Theta)$ and $\nu \in \mathcal{M}(\Theta)$ are respectively a non-negative scalar function and a non-negative measure, and $\alpha(\theta) = 0$ whenever $\left|x^1(\theta) - x^T(\theta)\right| < \varepsilon$.

Here E_c is the sublevel set defined by (4.1) with the constant $c = \left\|\Phi_T^*\Lambda_T^{-1}x^T\right\|_{L^2(0,T)^m}$ being the minimal norm required to reach x^T.

Remark 4.2. By the choice of the level c, it follows that $x^T \in \partial E_c$. If additionally the relation (4.2) holds, i.e., if x^T is the optimal final state of the problem (P), then $c = \left\|\Phi_T^* \Lambda_T^{-1} x^T\right\|_{L^2(0,T)^m}$ is the smallest level such that the intersection of the sublevel set E_c with the target box is non-empty (see Figure 1).

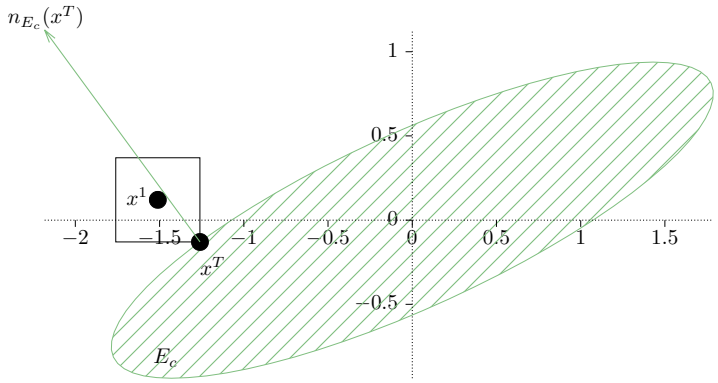


Figure 1: Set E_c for $n = 1$ and $K = 2$, together with the target ball, the optimal final state x^T and the normal functional $n_{E_c}(x^T)$. Each $x \in \mathcal{C}(\Theta)$ is presented by a point $(x(\theta_1), x(\theta_2)) \in \mathbb{R}^2$. For this plot, we have set $A(\theta_1) = 1$, $A(\theta_2) = 5$, $B(\theta_1) = 1$, $B(\theta_2) = 1/50$, $T = 1$, $\varepsilon = 1/4$, and the plotted set E_c is for $c = 1$.

Proof of Theorem 4.1. $\Longrightarrow$ Let us assume that x^T is the optimal final state of the problem (P). In that case $\frac{1}{2} n_{E_c}(x^T) = \Lambda_T^{-1} x^T$ is the optimal measure and the result follows directly from Proposition 3.3 (ensuring that it is of the form $\alpha(x^1 - x^T)|\mu_{opt}|$, with α and $|\mu_{opt}|$ being a non-negative scalar function and measure, respectively) and the description of the support of the optimal measure (3.3).

$\Longleftarrow$ Assume that (4.2) holds. In order to prove the statement, it is enough to show that x^T is the only element in the intersection of the sublevel set E_c and the ε ball around the target $x^1 \in \mathcal{C}(\Theta)^n$. We assume the contrary, i.e., that there exists some $\widetilde{x}^T \neq x^T$ lying in the above intersection.

We split the parameter set into two disjoint subsets, defined as

$$\Theta_b = \left\{\theta \in \Theta \;\middle|\; \left|x^1(\theta) - x^T(\theta)\right| = \varepsilon\right\} \quad \text{and} \quad \Theta_i = \left\{\theta \in \Theta \;\middle|\; \left|x^1(\theta) - x^T(\theta)\right| < \varepsilon\right\}.$$

By the assumption made on α, Θ_i is a zero measure set for $n_{E_c}(x^T)$. On the other hand, for any $\theta \in \Theta_b$, the vector $x^T(\theta)$ lies on the boundary of the ε ball (in $\mathbb{R}^n$) around the target $x^1(\theta)$, which, together with the fact that $\widetilde{x}^T(\theta)$ belongs to the same ball, implies that

$$\left\langle \left(x^1 - x^T\right)(\theta), \left(\widetilde{x}^T - x^T\right)(\theta) \right\rangle \geqslant 0 \qquad (\theta \in \Theta_b). \tag{4.3}$$

Consequently, we get

$$\begin{aligned} \left\langle n_{E_c}\left(x^T\right), \widetilde{x}^T - x^T \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} &= \left\langle n_{E_c}\left(x^T\right), \widetilde{x}^T - x^T \right\rangle_{\mathcal{M}(\Theta^1)^n, \mathcal{C}(\Theta^1)^n} \\ &= \left\langle \left(\alpha\left(x^1 - x^T\right)\right)_\flat \nu, \widetilde{x}^T - x^T \right\rangle_{\mathcal{M}(\Theta^1)^n, \mathcal{C}(\Theta^1)^n} \\ &= \left\langle \alpha_\flat \nu, \left(x^1 - x^T\right) \cdot \left(\widetilde{x}^T - x^T\right) \right\rangle_{\mathcal{M}(\Theta^1)^n, \mathcal{C}(\Theta^1)^n} \geqslant 0, \end{aligned} \tag{4.4}$$

where the final inequality follows by (4.3) and the assumed non-negativity of the measure $\alpha_\flat \nu$.

As $n_{E_c}(x^T)$ is the normal functional, the only way for $\widetilde{x}^T \in E_c \cap \mathcal{B}_\infty(x^1, \varepsilon)$ to satisfy the inequality (4.4) is that $\left\langle n_{E_c}\left(x^T\right), \widetilde{x}^T - x^T \right\rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} = 0$, i.e., that $\widetilde{x}^T$ belongs to the tangential space $T_{E_c}(x^T)$. As the convex set E_c intersects the tangential space $T_{E_c}(x^T)$ only in its boundary points, it follows that $\widetilde{x}^T \in \partial E_c$. In such way we obtain two distinct points $\widetilde{x}^T$ and x^T lying in $\partial E_c \cap \mathcal{B}_\infty(x^1, \varepsilon)$. However, that would imply existence of two different feasible controls of the same norm, c, which contradicts the uniqueness of the solution to problem (P). □

Remark 4.3. Note that the relation (4.2) implies that the optimal measure

$$\mu_{opt} = \Lambda_T^{-1} x^T = \tfrac{1}{2} n_{E_c}\left(x^T\right) \tag{4.5}$$

has, in every point $\theta \in \Theta$, the same direction as $(x^1 - x^T)(\theta)$, which is in accordance with the results of Proposition 3.3.

The last result allows an alternative, more direct proof of Proposition 3.6, in the case where the parameter set is of finite cardinal, by which one can construct a series of problems for which the exact solution is known.

Alternative proof of Proposition 3.6. With the notations and assumptions of Proposition 3.6, let $c = \left\| \Phi_T^* \Lambda_T^{-1} \widetilde{x}^T \right\|_{L^2(0,T)^m}$ be the minimal norm required to reach the state $\widetilde{x}^T$. Then for the normal functional $n_{E_c}(\widetilde{x}^T)$ we have

$$n_{E_c}\left(\widetilde{x}^T\right) = 2\Lambda_T^{-1}\widetilde{x}^T = 2 f \flat \eta = 2\left(\frac{|f|}{\varepsilon}\left(y - \widetilde{x}^T\right)\right) \flat \eta,$$

where $\eta \in \mathcal{M}(\Theta)$ and $f \in \mathcal{C}(\Theta)^n$ are the nonnegative scalar measure and the continuous vector function from the statement of Proposition 3.6 such that $\Lambda_T^{-1}\widetilde{x}^T = f \flat \eta$.

From here we see that the normal $n_{E_c}(\widetilde{x}^T)$ is of the form (4.2) (note that $y = x^1$ for the zero initial datum). The statement now follows from Theorem 4.1. □

The introduced sets E_c allow us to describe the set of targets for which the associated optimal control vanishes for some parameter $\theta \in \Theta$. This allows us to construct examples for which the exact solution is known. This is important in numerical implementations, where the detection of active parameters plays a crucial role in constructing the optimal measure.

For simplicity, the discussion is provided for scalar systems, i.e., we assume $n = 1$.

Given some optimal final state x^T, we know that the corresponding optimal measure is of the form (4.5). In particular, its i-th component (with $i \in \{1, \dots, K\}$ and $K = \operatorname{card}\Theta$) is going to be zero measure if and only if $x^T = \Lambda_T \mu_{opt}$ lies in the hyperplane $H_i = [\Lambda_1, \dots, \Lambda_{i-1}, \Lambda_{i+1} \dots \Lambda_K]$ spanned by all the columns of the Gramian matrix except the i-th one. This happens exactly for the states x^T for which the normal derivative $n_{E_c}(x^T) = 2\mu_{opt}$ is orthogonal to the i-th coordinate axes (Figure 2).

As any target x^1 lies at most $\sqrt[K]{2}\,\varepsilon/2$ from the optimal final state, this implies that the set of targets N_i for which the i-th component of the optimal measure equals zero is contained within the hyperstrip $H_i^\varepsilon := H_i + B(0, \sqrt[K]{2}\,\varepsilon/2)$. More precise characterization of this set can be obtained in the case $K = 2$, i.e., when the parameter obtains only two values. In that case, the set of targets for which the i-th component of the optimal measure is zero is given by

$$N_i = \left\{ x^T + \varepsilon\left(\frac{n_{E_c}(x^T)}{|n_{E_c}(x^T)|} + \beta e_i\right) \;\middle|\; x^T \in H_i,\ |\beta| \leqslant 1 \right\} \cup \mathcal{B}_\infty(0, \varepsilon),$$

and depicted on Figure 3.

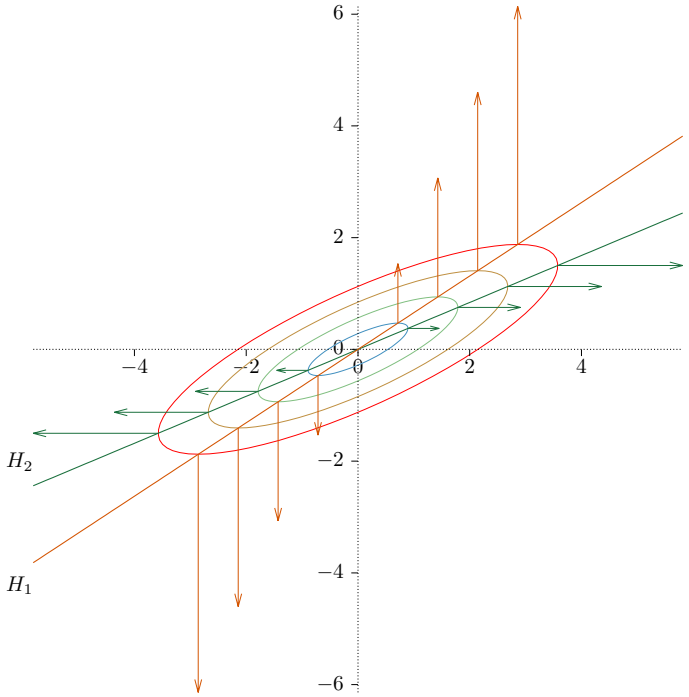


Figure 2: The hyperplanes H_1 and H_2 for $n = 1$ and $K = 2$. Arrows represent $n_{E_c}(x^T)$ for different values of c and well-chosen values of x^T. The parameters used for this plot are the ones of Figure 1.

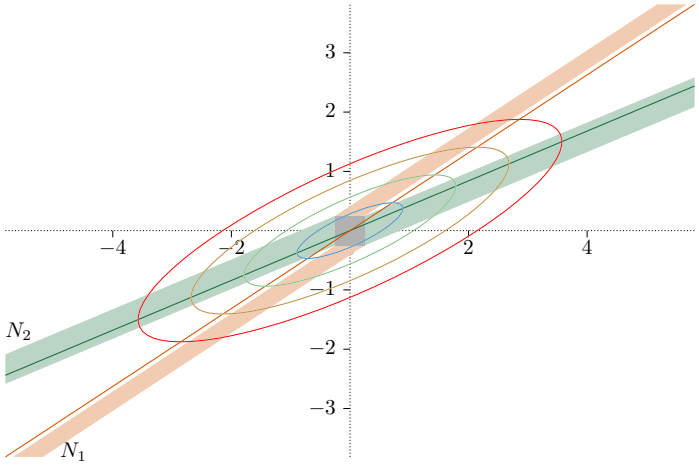


Figure 3: Sets N_1 and N_2 for $n = 1$ and $K = 2$. The parameters used for this plot are the ones of Figure 1.

For numerical implementations, it would be very useful to know whether the given target belongs to some set N_i. This is because revealing the support of the optimal measure is often the most challenging part of the numerical process, while, as we shall see in Section 7 (see in particular Examples 7.1 and 7.2), tuning the weights of active Dirac distributions turns out to be easier. However, the description of the sets N_i requires computation of the Gramian operator, which is a computationally very demanding task (therefore construction of the Gramian is generally avoided and instead various iterative procedures are used for solving control optimization problems). For this reason, the above description of targets for which the optimal measure vanishes at some points cannot directly accelerate numerical procedures. However, it contributes to a better understanding of the problem and its solution, and allows us to describe some characteristics of the optimal measures, in particular the sparsity.

In general, we can say that for the targets that are not contained in the union of hyperstrips $\cup_i H_i^\varepsilon$ the optimal measure is going to be fully supported. Of course, as hyperstrips H_i^ε, $i \in \{1, \dots, K\}$ cover just a smaller portion of the whole output space $\mathbb{R}^K$, this will be the case for the most of the targets. This might look surprisingly, having in mind that the term $\varepsilon \|\mu\|_{\mathcal{M}(\Theta)^n}$ entering the dual functional (3.1) is supposed to imply the sparsity in the support of the optimal measure. However, in the discrete case, the norm $\|\cdot\|_{\mathcal{M}(\Theta)^n}$ is equivalent to the ℓ^1 norm (see comments preceding Theorem 3.1), while the size of the measure support (i.e., number of active Dirac distributions) is given by the ℓ^0 norm. Although the ℓ^1 norm might be used to recover sparse solutions under suitable assumptions (see e.g. [7]), in general the ℓ^0 term is the one that induces sparsity.

5 Discretized problem and convergence

In this section, we show that the solution to the problem (P) can be arbitrarily well approximated by a solution to a discretized problem taking into account only a finite number of parameter values. This is of crucial importance from the numerical point of view, as the minimization of the dual functional in the latter case (originally defined on $\mathcal{M}(\Theta)^n$) reduces to the minimization of a function on $\mathbb{R}^{Kn}$ for some large enough $K \in \mathbb{N}^*$.

To this end, let $h > 0$ and consider Θ_h a compact subset of Θ such that $\operatorname{dist}(\theta, \Theta_h) \leqslant h$ for every $\theta \in \Theta$. The discretized version of (P) is

$$\left| \begin{array}{ll} \min & \frac{1}{2} \|u\|^2_{L^2(0,T)^m}, \\ & u \in L^2(0,T)^m \quad \text{and} \quad \sup\limits_{\theta \in \Theta_h} \left| x^1(\theta) - e^{TA(\theta)} x^0(\theta) - \displaystyle\int_0^T e^{(T-t)A(\theta)} B(\theta) u(t) \, \mathrm{d}t \right| \leqslant \varepsilon. \end{array} \right. \tag{P_h}$$

Following [18], we have the following results.

Theorem 5.1. *Assume that (2.1) is uniformly ensemble controllable. Given $y \in \mathcal{C}(\Theta)^n$ and $\varepsilon > 0$, we set $u_{opt} \in L^2(0,T)^m$ (respectively $u_h \in L^2(0,T)$) the unique minimizer of (P) (respectively (P_h)). Let us also set $\mu_{opt} \in \mathcal{M}(\Theta)^n$ and $\mu_h \in \mathcal{M}(\Theta_h)^n$ the unique minimizers of J defined by (3.1) over $\mathcal{M}(\Theta)^n$ and $\mathcal{M}(\Theta_h)^n$ respectively.*
We have $\|u_h\|_{L^2(0,T)^m} \leqslant \|u_{opt}\|_{L^2(0,T)^m}$, for every $h > 0$, and in addition, $u_h \to u_{opt}$ in $L^2(0,T)^m$ and μ_h vaguely converges to μ_{opt} as $h \to 0$.

Before giving the proof of this result, let us first recall that according to Proposition 2.4, if (2.1) is uniformly ensemble controllable with parameter set Θ, then (2.1) is uniformly ensemble controllable with parameter set Θ_h. Hence, the existence and uniqueness of u_h and μ_h (and of u_{opt} and μ_{opt}) directly follow from Theorem 3.1. Let us also mention that the strong convergence

of $(u_h)_h$ has already been given in [18, §2.2], under the additional assumption that A, B and y are Lipschitz function with respect to θ. It seems that this regularity assumption is not required, and we give this proof below.

Proof of Theorem 5.1. We identify μ_h to a measure in $\mathcal{M}(\Theta)$ supported on Θ_h. Since $\Theta_h \subset \Theta$, it is obvious that $\|u_h\|_{L^2(0,T)^m} \leqslant \|u_{opt}\|_{L^2(0,T)^m}$.

Step 1. Let us show that for every sequence $(h_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$ converging to 0, $\|\mu_{h_k}\|_{\mathcal{M}(\Theta)^n}$ is uniformly bounded with respect to k.
Assume by contradiction that there exist a sequence $(h_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$ converging to 0 such that $\|\mu_{h_k}\|_{\mathcal{M}(\Theta)^n} \to \infty$ as $k \to \infty$. We then set $\lambda_k = \|\mu_{h_k}\|_{\mathcal{M}(\Theta)^n}$, and we assume without loss of generality that $\lambda_k > 0$ for every $k \in \mathbb{N}$, and we set $\varsigma_k = \mu_{h_k}/\lambda_k$. By vague compactness (see e.g. [2, Chapter IV]), there exist a subsequence (still denoted by $(\varsigma_k)_{k\in\mathbb{N}}$) such that $(\varsigma_k)_k$ is vaguely convergent to $\varsigma_0 \in \mathcal{M}(\Theta)^n$. According to eq. (3.2), we have

$$\lambda_k \|\Phi_T^*\varsigma_k\|^2_{L^2(0,T)} = \langle \varsigma_k, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} - \varepsilon \|\varsigma_k\|_{\mathcal{M}(\Theta)^n} = \langle \varsigma_k, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} - \varepsilon.$$

This ensures that $\|\Phi_T^*\varsigma_k\|_{L^2(0,T)} \to 0$ as $k \to \infty$, i.e., $\Phi_T^*\varsigma_0 = 0$, and hence, $\varsigma_0 = 0$ by Proposition 2.7. Hence, for k large enough, we have, $\lambda_k \|\Phi_T^*\varsigma_k\|^2_{L^2(0,T)} = \langle \varsigma_k, y\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} - \varepsilon < 0$, which leads to a contradiction with $\lambda_k \|\Phi_T^*\varsigma_k\|^2_{L^2(0,T)} \geqslant 0$.

Step 2. Let us consider a sequence $(h_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$ such that $h_k \to 0$ as $k \to \infty$, and assume that $(\mu_{h_k})_{k\in\mathbb{N}}$ converges vaguely in $\mathcal{M}(\Theta)^n$. We aim to prove that $(\mu_{h_k})_{k\in\mathbb{N}}$ converges vaguely to μ_{opt}, and that the corresponding sequence of controls $(u_{h_k})_{k\in\mathbb{N}}$ converges strongly to u_{opt} in $L^2(0,T)^m$.
Let us denote by μ_0 the vague limit of $(\mu_{h_k})_k$ and we set $u_0 = \Phi_T^*\mu_0 \in L^2(0,T)^m$. For every $\varphi \in L^2(0,T)^m$, we have as $k \to \infty$,

$$\langle u_{h_k}, \varphi\rangle_{L^2(0,T)^m} = \langle \mu_{h_k}, \Phi_T\varphi\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} \longrightarrow \langle \mu_0, \Phi_T\varphi\rangle_{\mathcal{M}(\Theta)^n,\mathcal{C}(\Theta)^n} = \langle u_0, \varphi\rangle_{L^2(0,T)^m}.$$

This ensures the weak convergence of $(u_{h_k})_{k\in\mathbb{N}}$ to u_0, in particular, we have $\|u_0\|_{L^2(0,T)^m} \leqslant \liminf_{k\to\infty} \|u_{h_k}\|_{L^2(0,T)^m} \leqslant \|u_{opt}\|_{L^2(0,T)^m}$.

Since A and B are continuous on the compact set Θ, there are uniformly continuous and bounded. In addition, it is easy to see that there exists a constant $C > 0$, depending only on T and on the bounds of A and B, such that

$$\left|e^{(T-t)A(\theta_1)}B(\theta_1) - e^{(T-t)A(\theta_0)}B(\theta_0)\right| \leqslant C\left(|A(\theta_1) - A(\theta_0)| + |B(\theta_1) - B(\theta_0)|\right) \qquad (t \in [0,T],\ \theta_0,\theta_1 \in \Theta).$$

This ensures that for every $v \in L^2(0,T)^m$ and every $\theta_0, \theta_1 \in \Theta$ we have

$$|(\Phi_T v)(\theta_1) - (\Phi_T v)(\theta_0)| \leqslant \sqrt{T}C\|v\|_{L^2(0,T)^m}\left(|A(\theta_1) - A(\theta_0)| + |B(\theta_1) - B(\theta_0)|\right).$$

The last inequality clearly implies the uniform continuity of $\Phi_T u_{h_k}$, more precisely, for every $\delta > 0$, there exists $\eta(\delta) > 0$ (with $\eta(\delta) \to 0$ as $\delta \to 0$) such that for every $k \in \mathbb{N}$ and every $\theta_0, \theta_1 \in \Theta$ satisfying $|\theta_1 - \theta_0| \leqslant \delta$, we have

$$|(\Phi_T u_{h_k})(\theta_1) - (\Phi_T u_{h_k})(\theta_0)| \leqslant \eta(\delta).$$

By taking larger $\eta(\delta)$ if needed, we can also assume that

$$|y(\theta_1) - y(\theta_0)| \leqslant \eta(\delta).$$

Indeed, y is also uniformly continuous on Θ, since y is continuous on the compact set Θ.

For every $\theta \in \Theta$ and every $k \in \mathbb{N}$, let us set $\theta_{h_k} \in \Theta_{h_k}$ such that $|\theta - \theta_{h_k}| \leqslant h_k$. We recall that since $\theta_{h_k} \in \Theta_{h_k}$, we have $|(\Phi_T u_{h_k})(\theta_{h_k}) - y(\theta_{h_k})| \leqslant \varepsilon$. We then have, for every $\theta \in \Theta$ and every $k \in \mathbb{N}$,

$$\begin{aligned}|(\Phi_T u_0)(\theta) - y(\theta)| &\leqslant |(\Phi_T u_0)(\theta) - (\Phi_T u_{h_k})(\theta)| + |(\Phi_T u_{h_k})(\theta) - (\Phi_T u_{h_k})(\theta_{h_k})| \\ &\quad + |(\Phi_T u_{h_k})(\theta_{h_k}) - y(\theta_{h_k})| + |y(\theta_{h_k}) - y(\theta)| \\ &\leqslant |(\Phi_T u_0)(\theta) - (\Phi_T u_{h_k})(\theta)| + 2\eta(h_k) + \varepsilon.\end{aligned}$$

Taking the limit $k \to \infty$, we conclude that $|(\Phi_T u_0)(\theta) - y(\theta)| \leqslant \varepsilon$. This ensures that u_0 solves the uniform ensemble controllability problem with parameter set Θ. Thus, we have $\|u_0\|_{L^2(0,T)^m} \geqslant \|u_{opt}\|_{L^2(0,T)^m}$. But, we already know that $\|u_0\|_{L^2(0,T)^m} \leqslant \|u_{opt}\|_{L^2(0,T)^m}$ from the weak convergence of the sequence $(u_{h_k})_k$. This ensures that $\|u_0\|_{L^2(0,T)^m} = \|u_{opt}\|_{L^2(0,T)^m}$, and by uniqueness of the minimal L^2-norm control, $u_0 = u_{opt}$. Finally, from $\|u_{opt}\|_{L^2(0,T)^m} \geqslant \|u_{h_k}\|_{L^2(0,T)^m}$ and $\|u_0\|_{L^2(0,T)^m} \leqslant \liminf_{k\to\infty} \|u_{h_k}\|_{L^2(0,T)^m}$, we conclude that $\lim_{k\to\infty} \|u_{h_k}\|_{L^2(0,T)^m} = \|u_{opt}\|_{L^2(0,T)^m}$. This ensures that $u_{h_k} \to u_{opt}$ strongly in $L^2(0,T)^m$ as $k \to \infty$.

The fact that $u_0 = u_{opt}$ also ensures that $\Phi_T^* \mu_0 = \Phi_T^* \mu_{opt}$ and hence, $\mu_0 = \mu_{opt}$ by Proposition 2.7.

Conclusion. Assume by contradiction that μ_h is not vaguely convergent to μ_{opt} as $h \to 0$. Then, there exists $f \in \mathcal{C}(\Theta)^n$ such that $\langle \mu_h, f \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$ does not converge to $\langle \mu_{opt}, f \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n}$ as $h \to 0$. In particular, there exists $r > 0$ and a sequence $(h_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+^*)^{\mathbb{N}}$ converging to 0 such that $\left| \langle \mu_{h_k} - \mu_{opt}, f \rangle_{\mathcal{M}(\Theta)^n, \mathcal{C}(\Theta)^n} \right| > r$. But according to step 1, $(\mu_{h_k})_{k\in\mathbb{N}}$ is uniformly bonded. Using vague compactness, up to the extraction of a subsequence $(\mu_{h_k})_k$ is vaguely convergent. Finally, the step 2 ensures that the only possible vague limit is μ_{opt}. This leads to a contradiction, and concludes the proof. □

Remark 5.2. Let us mention that the arguments used in the first step of the proof of Theorem 5.1 are similar to the ones used in Lemma 3.9. This type of similarity is expected since in Theorem 5.1, we prove continuity of the minimizers of J with respect to the parameter set Θ, while Lemma 3.9 is an ingredient to prove continuity with respect to T, ε and y (see Proposition 3.10).

Let us now consider a discretized version of the problem. To this end, we consider $K \in \mathbb{N}^*$, $\theta_1, \dots, \theta_K \in \Theta$ be two by two distinct, $\Theta_{h_K} = \{\theta_1, \dots, \theta_K\}$, with $h_K = \max_{\theta\in\Theta} \operatorname{dist}(\theta, \Theta_{h_K})$. Let μ_{h_K} be the minimizer of J defined by (3.1) over Θ_{h_K}, then $\mu_{h_K} = \sum_{k=1}^K w_k \delta_{\theta_k}$, where δ_{θ_k} is the atomic mass located at θ_k, and $w = (w_k)_{k\in\{1,\dots,K\}} \in (\mathbb{R}^n)^K$ is the unique minimizer of

$$J_{\Theta_{h_K}}(w) = \frac{1}{2} \int_0^T \left| \sum_{k=1}^K B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \right|^2 \mathrm{d}t - \sum_{k=1}^K \langle w_k, y(\theta_k) \rangle + \varepsilon \sum_{k=1}^K |w_k| \tag{5.1}$$

and we have,

$$u_{h_K}(t) = \Phi_T^* \mu_{h_K} = \sum_{k=1}^K B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \qquad (t \in [0,T]).$$

In order to prepare § 6.1, let us give the following results.

Proposition 5.3. *$w = (w_k)_{k\in\{1,\dots,K\}} \in (\mathbb{R}^n)^K$ is a minimizer of $J_{\Theta_{h_K}}$ if and only if for every $l \in \{1, \dots, K\}$, $w_l \in \mathbb{R}^n$ is a minimizer of*

$$J_l : \widehat{w} \in \mathbb{R}^n \mapsto \frac{1}{2} \widehat{w}^* \Lambda_{\theta_l} \widehat{w} - \left\langle \widehat{w}, y(\theta_l) - \int_0^T e^{(T-t)A(\theta_l)} B(\theta_l) u_l(t) \, \mathrm{d}t \right\rangle + \varepsilon |\widehat{w}|,$$

with

$$u_l(t) = \sum_{\substack{k=1\\k\neq l}}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \quad \textit{and} \quad \Lambda_{\theta_l} = \int_0^T e^{(T-t)A(\theta_l)} B(\theta_l) B(\theta_l)^* e^{(T-t)A(\theta_l)^*} \,\mathrm{d}t.$$

Proof. According to Theorem A.1, $w = (w_k)_{k\in\{1,\dots,K\}}$ is a minimizer of $J_{\Theta_{h_K}}$ if and only if for every $l \in \{1,\dots,K\}$, w_l is a minimizer of $\widehat{w} \in \mathbb{R}^n \mapsto J_{\Theta_{h_K}}(\widetilde{w})$ with $\widetilde{w}_k = \begin{cases} w_k, & \text{if } k \neq l, \\ \widehat{w}, & \text{otherwise} \end{cases}$.
But, for every $l \in \{1,\dots,K\}$, we have,

$$\begin{aligned}
J_{\Theta_{h_K}}(w) &= \frac{1}{2}\int_0^T \left| \sum_{k=1}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \right|^2 \mathrm{d}t - \sum_{k=1}^{K} \langle w_k, y(\theta_k)\rangle + \varepsilon \sum_{k=1}^{K} |w_k| \\
&= \frac{1}{2}\int_0^T \left| B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l \right|^2 \mathrm{d}t \\
&\quad + \int_0^T \left\langle B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l, \sum_{\substack{k=1\\k\neq l}}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \right\rangle \mathrm{d}t - \langle w_l, y(\theta_l)\rangle + \varepsilon |w_l| \\
&\quad + \frac{1}{2}\int_0^T \left| \sum_{\substack{k=1\\k\neq l}}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \right|^2 \mathrm{d}t - \sum_{\substack{k=1\\k\neq l}}^{K} \langle w_k, y(\theta_k)\rangle + \varepsilon \sum_{\substack{k=1\\k\neq l}}^{K} |w_k| .
\end{aligned}$$

It is then clear that minimizers of $\widehat{w} \mapsto J_{\Theta_{h_K}}(\widetilde{w})$ are minimizers of J_l. □

Let us now give condition on the minimizers of J_l defined in Proposition 5.3.

Proposition 5.4. *Let $\Lambda \in \mathbb{R}^{n\times n}$, be a positive definite matrix, $z \in \mathbb{R}^n$ and $\varepsilon > 0$, the minimizer $p_{opt} \in \mathbb{R}^n$ of $\mathcal{J} : p \in \mathbb{R}^n \mapsto \frac{1}{2}p^*\Lambda p - \langle p, z\rangle + \varepsilon |p|$ over $\mathbb{R}^n$ is unique and satisfies $p_{opt} = 0$ if $|z| < \varepsilon$, and otherwise $p_{opt} = \rho q$ where,*

- $|q| = 1$;
- $\mathrm{shrink}(|z|, \varepsilon)/C_\Lambda \leqslant \rho \leqslant \mathrm{shrink}(|z|, \varepsilon)/c_\Lambda$;
- $(\varepsilon I + \rho\Lambda)q = z$,

where $0 < c_\Lambda \leqslant C_\Lambda$ are such that

$$c_\Lambda |p|^2 \leqslant p^*\Lambda p \leqslant C_\Lambda |p|^2 \qquad (p \in \mathbb{R}^n)$$

and

$$\mathrm{shrink}(s,\mu) = \begin{cases} s+\mu & \text{if } s < -\mu, \\ 0 & \text{if } -\mu \leqslant s \leqslant \mu, \\ s-\mu & \text{if } \mu < s \end{cases} \qquad (s \in \mathbb{R},\ \mu \in \mathbb{R}_+).$$

Remark 5.5. When $n = 1$, we have $p_{opt} = \mathrm{shrink}(z, \varepsilon)/\Lambda$.

Proof. First, $\mathcal{J}$ is strictly convex and coercive, hence admits a unique minimizer p_{opt} on $\mathbb{R}^n$. Furthermore, p_{opt} is characterized by $0 \in \partial\mathcal{J}(p_{opt})$. But, we have

$$\partial\mathcal{J}(p) = \begin{cases} \Lambda p - z + \varepsilon \frac{p}{|p|} & \text{if } p \neq 0, \\ \overline{B}(0,\varepsilon) - z & \text{if } p = 0, \end{cases}$$

recall that $\overline{B}(0,\varepsilon) \subset \mathbb{R}^n$ is the closed ball of $\mathbb{R}^n$ centered on 0 and of radius ε. In particular, we have $0 \in \partial\mathcal{J}(0)$ if $|z| \leqslant \varepsilon$. Let us then assume that $|z| > \varepsilon$, we then have $p_{opt} \neq 0$, and we set $\rho = |p_{opt}|$ and $q = p_{opt}/\rho$. We thus have $(\varepsilon I + \rho\Lambda)q = z$, and hence $\rho > 0$ shall be such that $1 = |q|^2 = \left|(\varepsilon I + \rho\Lambda)^{-1} z\right|^2$. This leads to the bounds on ρ. □

Before entering the algorithmic computation of a minimizer of $J_{\Theta_{h_K}}$, let us give the sub-differential of $J_{\Theta_{h_K}}$. For every $k \in \{1,\dots,K\}$, we have,

$$\partial_{w_k} J_{\Theta_{h_K}}(w) = \begin{cases} y(\theta_k) - \displaystyle\int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) u(t)\,\mathrm{d}t + \varepsilon \frac{w_k}{|w_k|}, & \text{if } w_k \neq 0, \\ y(\theta_k) - \displaystyle\int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) u(t)\,\mathrm{d}t + \mathcal{B}(0,\varepsilon), & \text{if } w_k = 0, \end{cases}$$

with $u(t) = \sum_{k=1}^K B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k$. In particular, if $0 \in \partial_{w_k} J_{\Theta_{h_K}}(w)$, we have

$$\operatorname{shrink}\left(\left| y(\theta_k) - \int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) u(t)\,\mathrm{d}t \right|, \varepsilon\right) = 0.$$

In addition, when $w_k = 0$, we have

$$\operatorname{shrink}\left(\left| y(\theta_k) - \int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) u(t)\,\mathrm{d}t \right|, \varepsilon\right) = \min\left\{|z| \;\middle|\; z \in \partial_{w_k} J(w)\right\}.$$

Alternatively to the algorithm proposed in § 6.1, one can use a direct optimization based on the optimality conditions given in Proposition 3.3. We have the following direct corollary of Proposition 3.3.

Corollary 5.6. *$w = (w_k)_{k\in\{1,\dots,K\}} \in (\mathbb{R}^n)^K$ is a minimizer of $J_{\Theta_{h_K}}$ if there exists $(\alpha_k)_{k\in\{1,\dots,K\}} \in \mathbb{R}^K$ such that for every $k \in \{1,\dots,K\}$,*

- *$\alpha_k \geqslant 0$, and $\alpha_k = 0$ if $|y(\theta_k) - (\Phi_T \Phi_T^* \mu_{h_K})(\theta_k)| < \varepsilon$;*
- *$w_k = \alpha_k \left(y(\theta_k) - (\Phi_T \Phi_T^* \mu_{h_K})(\theta_k)\right)$,*

where we have set $\mu_{k_K} = \sum_{k=1}^K w_k \delta_{\theta_k}$.

These condition will lead to the minimization problem given in § 6.2.

6 Numerical strategies

In this section, we propose three ways to numerically compute a minimizer of J_{Θ_h}.

The first one, S.0 is basically using existing solvers without taking in consideration differentiability issues that emerge from non-smoothness of the dual functional. The second one, S.1 is based on the optimality conditions given in Corollary 5.6, see § 6.2. The last ones, S.2.0 and S.2.1, use the *greedy coordinate descent method* proposed in [22], see § 6.1.
Finally, the strategies S.2.0 and S.2.1 can also be combined with the strategies S.0 and S.1. All the possibility are given in § 6.3, and compared on some numerical examples in Section 7.

6.1 Greedy coordinate descent

Inspired from [22], we propose the following *greedy coordinate descent* algorithm (Algorithm 1). This algorithm is a coordinate descent algorithm, but instead of considering each coordinate sequentially, we select at each iteration the coordinate for which the decrease of the cost function is the most important.

The convergence of Algorithm 1 is ensured by the following result.

Proposition 6.1. *Let $f_1, \dots, f_K \in \mathcal{C}(\mathbb{R}^n)$ and $F \in \mathcal{C}^1((\mathbb{R}^n)^K)$ be all convex. We consider the cost function*

$$\mathfrak{J}(w_1, \dots, w_K) = F(w_1, \dots, w_K) + \sum_{k=1}^{K} f_k(w_k) \qquad (w_1, \dots, w_K \in \mathbb{R}^n),$$

and we assume that $\mathfrak{J}$ is coercive and strictly convex.
Given $w^0 = (w_1^0, \dots, w_K^0) \in (\mathbb{R}^n)^K$, we define the sequences $(k_i)_{i\in\mathbb{N}} \in \{1, \cdots, K\}^{\mathbb{N}}$ and $(w^i)_{i\in\mathbb{N}} \in \left((\mathbb{R}^n)^K\right)^{\mathbb{N}}$ by:

$$k_i \in \operatorname*{argmin}_{k\in\{1,\dots,K\}} \left(\min_{\widehat{w}\in\mathbb{R}^n} J(w_1^i, \dots, w_{k-1}^i, \widehat{w}, w_{k+1}^i, \dots, w_K^i) \right), \tag{6.1}$$

and

$$\begin{aligned} w_{k_i}^{i+1} &\in \operatorname*{argmin}_{\widehat{w}\in\mathbb{R}^n} J(w_1^i, \dots, w_{k_i-1}^i, \widehat{w}, w_{k_i+1}^i, \dots, w_K^i), \\ w_k^{i+1} &= w_k^i \quad (\forall k \in \{1, \dots, K\} \setminus \{k_i\}), \end{aligned}$$

Then $\lim_{i\to\infty} w^i = w^{opt} \in (\mathbb{R}^n)^K$, where w^{opt} is the minimizer of $\mathfrak{J}$.

The proof of this result follows the arguments given in [22, Section 2.4] and is given in Appendix A for the sake of completeness. Note that in [22] the proof is given for $n = 1$.

Remark 6.2. We expected that the convergence result would also hold by selecting

$$k_i \in \operatorname*{argmax}_{k\in\{1,\dots,K\}} \operatorname{dist}\left(0, \partial_{w_k}\mathfrak{J}(w^i)\right). \tag{6.2}$$

But, we did not succeed to prove this result without additional regularity assumptions on the functions f_k. However, in practice, it seems that using the selection rule (6.2), give similar results compared to the selection rule (6.1), but is much faster (see examples in Section 7).

Algorithm 1 Greedy coordinate descent for uniform ensemble control.

Require: $T > 0$, $\varepsilon > 0$, $(A, B, x^0, x^1) \in \mathcal{C}(\Theta)^{n\times n} \times \mathcal{C}(\Theta)^{n\times m} \times \mathcal{C}(\Theta)^n \times \mathcal{C}(\Theta)^n$, with $n, m \in \mathbb{N}^*$
Require: $\Theta_h \subset \Theta$ with $K := \operatorname{card}(\Theta_h) \in \mathbb{N}^*$ and $\Theta_h = \{\theta_1, \dots, \theta_K\}$
Require: $\eta > 0$ ▷ A tolerance parameter.
Ensure: $u \in L^2(0,T)^m$ is such that $\sup_{\theta\in\Theta_h} \left| x^1(\theta) - e^{TA(\theta)}x^0(\theta) - \int_0^T e^{(T-t)A(\theta)}B(\theta)u(t)\,\mathrm{d}t \right| \leqslant \varepsilon + \delta$

Algorithm 1 (continued)

for all $k \in \{1, \dots, K\}$ **do**
  $w_k \leftarrow 0$
  $y_k \leftarrow x^1(\theta_k) - e^{TA(\theta_k)} x^0(\theta_k)$
  $G_k \leftarrow \text{shrink}(|y_k|, \varepsilon)$ ▷ $G_k = \text{dist}\{0, \partial_{w_k} J_{\Theta_h}(w)\}$.
end for
$u(t) \leftarrow 0$ ▷ $u(t) = \sum_{k=1}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k$.
$c \leftarrow 0$ ▷ $c = J_{\Theta_h}(w)$.
while $\sup_{k \in \{1,\dots,K\}} |y_k| > \varepsilon + \eta$ **do**
  for all $l \in \{1, \dots, K\}$ **do** ▷ Look for the best descent coordinate.
    if $G_l \neq 0$ **then**
      for all $k \in \{1, \dots, K\}$ **do** $\widehat{w}_k \leftarrow w_k$
      end for
      $\widehat{u}(t) \leftarrow B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l$
      $\widehat{y}_l \leftarrow y_l + \int_0^T e^{(T-t)A(\theta_l)} B(\theta_l) \widehat{u}(t)\, \mathrm{d}t$
      ▷ We have $\widehat{y}_l = x^1(\theta_l) - e^{TA(\theta_l)} x^0(\theta_l) - \int_0^T e^{(T-t)A(\theta_l)} B(\theta_l) \sum_{\substack{k=1 \\ k \neq l}}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k \, \mathrm{d}t$.
      $\widehat{w}_l \leftarrow$ Approximate Control$(\varepsilon, T, A(\theta_l), B(\theta_l), \widehat{y}_l)$ ▷ See Algorithm 2.
      $\widehat{c}_l \leftarrow J_{\Theta_h}(\widehat{w})$
    else $\widehat{c}_l \leftarrow c$
    end if
  end for
  $l \leftarrow$ an element of $\operatorname{argmax}_{k \in \{1,\dots,K\}} \widehat{c}_k$ ▷ l is the best coordinate to update.
  $\widehat{u}(t) \leftarrow B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l$
  $\widehat{y}_l \leftarrow y_l + \int_0^T e^{(T-t)A(\theta_l)} B(\theta_l) \widehat{u}(t)\, \mathrm{d}t$
  $w_l \leftarrow$ Approximate Control$(\varepsilon, T, A(\theta_l), B(\theta_l), \widehat{y}_l)$
  $u(t) \leftarrow u(t) - \widehat{u}(t) + B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l$ ▷ We have $u(t) = \sum_{k=1}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k$.
  $c \leftarrow J_{\Theta_h}(w)$
  for all $k \in \{1, \dots, K\}$ **do**
    $y_k \leftarrow x^1(\theta_k) - e^{TA(\theta_k)} x^0(\theta_k) - \int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) u(t)\, \mathrm{d}t$
    if $w_k = 0$ **then** $G_k \leftarrow \text{shrink}(|y_k|, \varepsilon)$
    else $G_k \leftarrow \left| y_k - \varepsilon \frac{w_k}{|w_k|} \right|$
    end if
  end for
end while
return $u(t)$

Remark 6.3. Here are some comments on Algorithm 1.

- The loop, lines 9 to 19, can be done in parallel.

- What is computed in the loop, lines 9 to 19, can be stored and reused lines 21 to 23 and 25.
- Algorithm 2 will be called many times, and hence has to be fast.
- This algorithm is adapted from [22], where the aim is to find sparse solutions. Thus, as we will see on some examples in Section 7, will lead to sparse minimizers. Consequently, it might not be optimal when we are in the situation described in Proposition 3.6 and Corollary 3.7, i.e., when the optimal measure is fully supported. However, we shall see that the controls obtained in this way still provide a fairy good approximation of the optimal one.

Algorithm 2 Computation of an approximate control for the system $\dot{x} = Ax + Bu$.

Require: $T > 0$, $\varepsilon > 0$ and $(A, B, y) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m} \times \mathbb{R}^n$ with $n, m \in \mathbb{N}^*$

Ensure: $w \in \mathbb{R}^n$ is such that $\left| y - \int_0^T e^{(T-t)A} B u(t)\,\mathrm{d}t \right| \leqslant \varepsilon$ with $u(t) = B^* e^{(T-t)A^*} w$

1: **function** APPROXIMATE CONTROL(ε, T, A, B, y)
2: **if** $|y| \leqslant \varepsilon$ **then** $w \leftarrow 0$ ▷ See Remark 5.5.
3: **else** ▷ In that case, we have $w \neq 0$.
4: **if** $n = 1$ **then**
5: $\Lambda \leftarrow \int_0^T e^{(T-t)A} B B^* e^{(T-t)A^*}\,\mathrm{d}t$
6: $w \leftarrow \mathrm{shrink}(y, \varepsilon)/\Lambda$ ▷ See Remark 5.5.
7: **else if** the control Gramian is "computable" **then** ▷ e.g. for n "small".
8: $\Lambda \leftarrow \int_0^T e^{(T-t)A} B B^* e^{(T-t)A^*}\,\mathrm{d}t$
9: $(r, q) \leftarrow$ solution of $r > 0$, $q \in \mathbb{R}^n$, $|q| = 1$ and $(\varepsilon I + r\Lambda)q(r) = y$
10: $w \leftarrow rq(r)$ ▷ See Proposition 5.4.
11: **else**
12: $w \leftarrow \underset{\nu\in\mathbb{R}^n}{\mathrm{argmin}} \left(\frac{1}{2} \int_0^T \left| B^* e^{(T-t)A^*} \nu \right|^2 \mathrm{d}t - \langle \nu, y \rangle + \varepsilon\,|\nu| \right)$
13: **end if**
14: **end if**
15: **return** w
16: **end function**

Remark 6.4.
- In Algorithm 2 (line 9), finding a solution of $r > 0$, $q \in \mathbb{R}^n$, $|q| = 1$ and $(\varepsilon I + r\Lambda)q(r) = y$ can be done using a dichotomy search.
- To find a minimizer of J (line 12 of Algorithm 2), one can use a gradient descent method. In fact, note that J is only not differentiable at 0, and note that 0 cannot be the minimizer in the situation of line 12 this is because $|y| > \varepsilon$. A way to avoid the evaluation of $\mathcal{J}(\nu) = \frac{1}{2}\int_0^T \left|B^* e^{(T-t)A^*}\nu\right|^2 \mathrm{d}t - \langle \nu, y\rangle + \varepsilon\,|\nu|$ at 0 is to initialize the gradient descent method with $\nu \in \mathbb{R}^n$ such that $\mathcal{J}(\nu) < \mathcal{J}(0) = 0$. This is for instance possible by setting $\nu = \alpha y$, with $\alpha = \mathrm{shrink}(|y|^2, \varepsilon\,|y|)/C$, with $C = \int_0^T \left|B^* e^{(T-t)A^*} y\right|^2 \mathrm{d}t$. We thus have $\nu \neq 0$ and $\mathcal{J}(\nu) < 0$, see Remark 5.5.
- Computing the Gramian matrix might be resource consuming. Hence, instead of computing it in Algorithm 2, it can be computed and stored in Algorithm 1.

6.2 Reduced minimization problem

For this strategy, we will use the optimality condition given in Corollary 5.6. More precisely, we consider the following minimization problem (recall that we have set $y(\theta) = x^1(\theta) - e^{TA(\theta)}x^0(\theta)$),

$$\min \frac{1}{2} \left\| \sum_{k=1}^{K} B(\theta_k)^* e^{(T-\cdot)A(\theta_k)^*} w_k \right\|^2_{L^2(0,T)^m} \tag{6.3a}$$

with respect to $w_1, \dots, w_K \in \mathbb{R}^n$, subject to the following constraints

$$\forall k \in \{1, \dots, K\},\ \exists \alpha_k \geqslant 0 \quad \text{such that:} \quad \begin{cases} \left|y(\theta_k) - y^T(\theta_k)\right|^2 \leqslant \varepsilon^2, \\ \alpha_k \left(\left|y(\theta_k) - y^T(\theta_k)\right|^2 - \varepsilon^2 \right) \geqslant 0, \\ w_k = \alpha_k \left(y(\theta_k) - y^T(\theta_k) \right), \end{cases} \tag{6.3b}$$

where we have set

$$y^T(\theta_k) = \int_0^T e^{(T-t)A(\theta_k)} B(\theta_k) \left(\sum_{l=1}^{K} B(\theta_l)^* e^{(T-t)A(\theta_l)^*} w_l \right) \mathrm{d}t. \tag{6.3c}$$

Observe that the last three constraints coincide with the optimality condition given in Corollary 5.6, and the control is given by $u(t) = \sum_{k=1}^{K} B(\theta_k)^* e^{(T-t)A(\theta_k)^*} w_k$. Using Corollary 5.6, it is obvious that minimizers of $J_{\Theta_{h_K}}$ coincide with minimizers of (6.3).

Remark 6.5. In practice the minimization problem (6.3) can be expressed as a minimization only on $(\alpha_k)_{k\in\{1,\dots,K\}}$, however, from a numerical point of view, it is easier to consider the minimization problem (6.3) as a minimization with respect to $(w_k, \alpha_k, y(\theta_k))_{k\in\{1,\dots,K\}}$ since gradient and Hessians can be easily computed.

6.3 Proposed numerical strategies

Based on the previous consideration, we can propose the following numerical strategies.

- S.0. Use existing nonlinear solvers, and try to directly minimize $J_{\Theta_{h_K}}$, without considering potential issues related to the non-differentiability of the cost functional. In the examples of Section 7, we use the `fminunc` function of `Matlab` with the quasi-Newton algorithm.
- S.1. Use existing nonlinear solvers to find a minimizer of (6.3). In the examples of Section 7, we use the `fmincon` function of `Matlab` with the interior-point algorithm.

In addition, one can use algorithms based on Algorithm 1. This leads to two strategies:

- S.2.0. Use directly Algorithm 1.
- S.2.1. Use Algorithm 1 but instead of using the selection rule given by (6.1), we use the one given by (6.2) (see Remark 6.2).

As we will see in Section 7, with strategies S.2.0 and S.2.1, it appears that many steps are between non-zeros values. Hence, we propose a refinement of these two algorithms by running S.0 or S.1 after some component of $(w_k)_k$ switches from null to a non-null value, or vice versa. More precisely, using notations of Algorithm 1, if w_l was 0 and is updated to a non-null value or if w_l was non-null and is updated to 0, we run S.0 or S.1 for the same minimization problem, but with $\widetilde{\Theta} = \{\theta_k \in \Theta_h,\ \forall k \in \{1, \dots, K\} \text{ s.t. } w_k \neq 0\}$ in place of Θ_h. In other words, we try to update all the non-zero values of $(w_k)_k$ in one shoot. This leads to the four additional strategies, S.2.i–S.j (with $i, j \in \{0, 1\}$).

Let us recall that we have analytical guaranties of convergence only for the strategy S.2.0.

7 Numerical examples

Unless explicitly said otherwise, for all the following examples

- the final time is $T = 1$;
- $\Theta = [1, 2]$, and Θ_h is a uniform discretization of Θ with step $h = 10^{-2}$, leading to $K = \operatorname{card} \Theta_h = 101$;
- $\varepsilon = 5 \times 10^{-2}$;
- the time integrals are computed using the trapezoidal rule with a uniform time step of 10^{-4}. In particular, to compute Φ_T, we use the trapezoidal rule, and the matrix exponentials are computed numerically. Another alternative could have been to solve the ordinary differential equation using a numerical scheme;
- the stopping criterion on the gradient (i.e., δ in Algorithm 1) is set to 10^{-12}, we also set a step tolerance of 10^{-20} (i.e., step sizes shall be greater than this number) and a maximal number of iterations of 10^5.

Examples 7.1 and 7.2 correspond to situation where the optimal control and the adjoint measure are known (see Proposition 3.6 and Corollary 3.7). Example 7.3 corresponds to a situation where the target cannot be reached (see Appendix B). Finally, Example 7.4 tackled the uniform ensemble controllability for a discretized heat equation.

In the tables used for the comparison of the different numerical strategies, $G = (G_1, \dots, G_K) \in (\mathbb{R}_+)^K$ is given by lines 27 to 29 of Algorithm 1 that is $G_k = \operatorname{dist}(0, \partial_{w_k} J_{\Theta_h}(w))$.

The codes are run using `Matlab` version `R2020b` on a 12 core processor of 64 bits and 3.20 GHz.

Example 7.1. We consider a first basic example, where $n = m = 1$, $A(\theta) = -\theta$ and $B(\theta) = 1$. It is well-known that this system is ensemble controllable (see [21, 27]). We consider $x^0(\theta) = 0$, $\mathrm{d}\mu_{opt}(\theta) = \chi_\omega(\theta)\,\mathrm{d}\theta$, where ω is an open subset of Θ. According to Corollary 3.7, we know that $u_{opt} = \Phi_T^* \mu_{opt}$ is the minimizer of (P) when the target is set to $x^1(\theta) = (\Phi_T \Phi_T^* \mu_{opt})(\theta) + \varepsilon$.

In practice, since Θ_h (introduced in Section 5) is chosen as a uniform discretization of Θ, we set $\mu_{opt,h} = h \sum_{\theta \in \Theta_h \cap \omega} \delta_\theta$, and it is clear that $\mu_{opt,h}$ is vaguely convergent to μ_{opt}. Also, the target x_h^1 is set accordingly, i.e., $x_h^1(\theta) = (\Phi_T \Phi_T^* \mu_{opt,h})(\theta) + \varepsilon$, and the discretized optimal control is $u_{opt,h} = \Phi_T^* \mu_{opt,h}$.

The numerical simulations are run with $\omega = (4/3, 7/4)$ (we recall that the other parameters are given in the preamble of Section 7). A comparison of the different strategies listed in § 6.3 is given in Table 1.

Strategy	CPU time (s)	running time (s)	nb. iter.	$\max G$	$\frac{\|u_h - u_{opt,h}\|_{L^2}}{\|u_{opt,h}\|_{L^2}}$	$\frac{\|\mu_h - \mu_{opt,h}\|_{\mathcal{M}}}{\|\mu_{opt,h}\|_{\mathcal{M}}}$
S.0	6.07	0.63	7	2.553×10^{-5}	1.561×10^{-3}	**0.959**
S.1	4.59	2.82	83	1.293×10^{-13}	4.384×10^{-8}	1.271
S.2.0	1409.58	201.93	100000	2.387×10^{-8}	2.467×10^{-5}	1.952
S.2.0–S.0	39.29	8.49	11	8.161×10^{-10}	4.432×10^{-6}	1.952
S.2.0–S.1	11.19	2.47	14	$\mathbf{7.452 \times 10^{-15}}$	$\mathbf{6.833 \times 10^{-9}}$	1.902
S.2.1	198.18	20.01	100000	3.389×10^{-7}	1.140×10^{-3}	1.800
S.2.1–S.0	**1.06**	**0.19**	14	3.043×10^{-9}	1.617×10^{-5}	1.905
S.2.1–S.1	4.28	1.19	12	3.757×10^{-14}	2.951×10^{-8}	1.857

(a) Errors, time of computation and number of iterations.

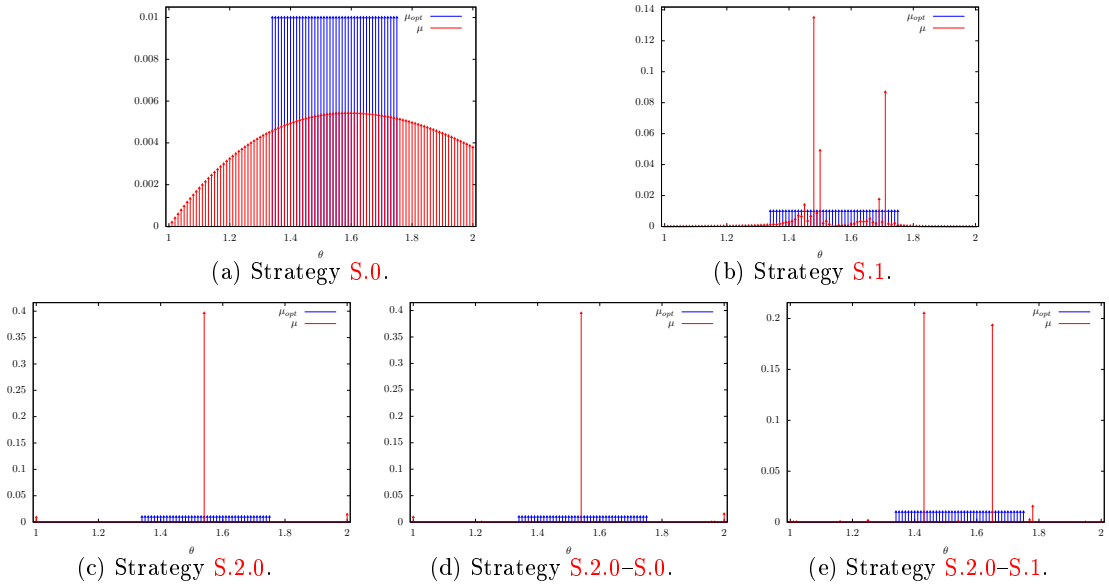
Table 1: Comparison of the strategies listed in § 6.3 for Example 7.1.

Strategy	Stopping reason
S.0	Cannot decrease the objective function along the current search direction.
S.1	Local minimum possible. Constraints satisfied.
S.2.0	Maximal number of iteration reached.
S.2.0–S.0	Cost is not decreasing.
S.2.0–S.1	Cost is not decreasing.
S.2.1	Maximal number of iteration reached.
S.2.1–S.0	Cost is not decreasing.
S.2.1–S.1	Cost is not decreasing.

(b) Stopping reasons.

Table 1: (continued)

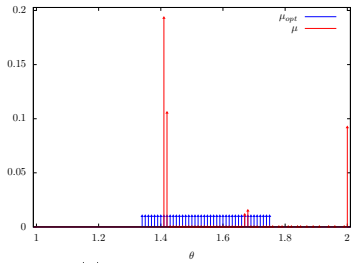
In Table 1b, for strategies S.2.i–S.j ($i, j \in \{0, 1\}$) the last cost modifications is of order 10^{-17}, which is close to the numerical 0 (indeed, by denoting c_{opt} the minimal cost, we have numerically $c_{opt} + 10^{-18} = c_{opt}$). Let us also mention that for all simulations, we have $\left|x^1(\theta) - x(T, \theta)\right| = \varepsilon$ up to an error of order 10^{-8}. We also see that the strategy S.2.0–S.1 gives the best results and strategy S.2.1–S.0 is the fastest one. As claimed in § 6.3, for strategy S.2.0 and S.2.1, most of the time is spent in updating non-zero values in fact for S.2.0, for all 10^5 steps the selected index is in $\{1, 55, 101\}$, and for S.2.1, after some iterations the selected index loops between 1, 42 and 101 and then between 42, 69 and 101. In Table 1a, we see that even if the control is closed to the optimal ones, the adjoint is rather far from the optimal one, this is also illustrated on Figure 4. As we can see from Figure 4 and Table 1a, wide variation on the adjoint does not necessarily produce large variations on the control. One can see from Figure 4, that the strategies S.0 and S.1 lead to adjoints that are fully supported on Θ_h, while the other strategies lead to adjoints for which the support consist in only few points of Θ_h. These facts will be discussed later on.



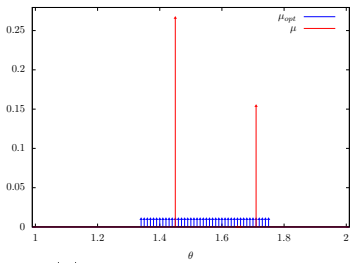
(a) Strategy S.0. (b) Strategy S.1.

(c) Strategy S.2.0. (d) Strategy S.2.0–S.0. (e) Strategy S.2.0–S.1.

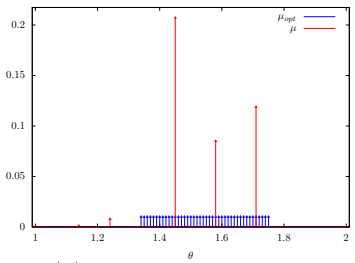
Figure 4: Adjoints obtained with strategies listed in § 6.3 for Example 7.1.



(f) Strategy S.2.1.



(g) Strategy S.2.1–S.0.



(h) Strategy S.2.1–S.1.

Figure 4: (continued)

Example 7.2. We consider the system and data given in Example 7.1, except that we set $\mu_{opt} = \delta_{\widetilde{\theta}_1} + 2\delta_{\widetilde{\theta}_2}$. In that case, for $i \in \{1,2\}$, we set $\widetilde{\theta}_{i,h} \in \operatorname{argmin}\left\{\left|\widetilde{\theta}_i - \theta\right|,\ \theta \in \Theta_h\right\}$ and $\mu_{opt,h} = \delta_{\widetilde{\theta}_{1,h}} + 2\delta_{\widetilde{\theta}_{2,h}}$.

The numerical simulation is run with $\widetilde{\theta}_1 = 7/6$ and $\widetilde{\theta}_2 = 17/10$ (we recall that the other parameters are given in the preamble of Section 7). A comparison of the different strategies listed in § 6.3 is given in Table 2.

Strategy	CPU time (s)	running time (s)	nb. iter.	$\max G$	$\frac{\|u_h - u_{opt,h}\|_{L^2}}{\|u_{opt,h}\|_{L^2}}$	$\frac{\|\mu_h - \mu_{opt,h}\|_{\mathcal{M}}}{\|\mu_{opt,h}\|_{\mathcal{M}}}$
S.0	4.11	0.44	19	2.431×10^{-8}	2.421×10^{-5}	**1.960**
S.1	2.92	1.78	76	$\mathbf{2.058\times10^{-13}}$	$\mathbf{2.675\times10^{-8}}$	1.981
S.2.0	1478.55	213.11	100000	7.575×10^{-7}	1.065×10^{-4}	2.000
S.2.0–S.0	0.55	0.10	10	4.783×10^{-9}	1.999×10^{-5}	2.000
S.2.0–S.1	5.60	1.38	11	2.553×10^{-12}	3.665×10^{-7}	2.000
S.2.1	194.10	19.49	100000	2.495×10^{-6}	3.447×10^{-4}	2.000
S.2.1–S.0	**0.28**	**0.04**	5	2.188×10^{-8}	4.409×10^{-5}	2.000
S.2.1–S.1	3.01	0.78	8	9.378×10^{-12}	5.331×10^{-7}	2.000

(a) Errors, time of computation and number of iterations

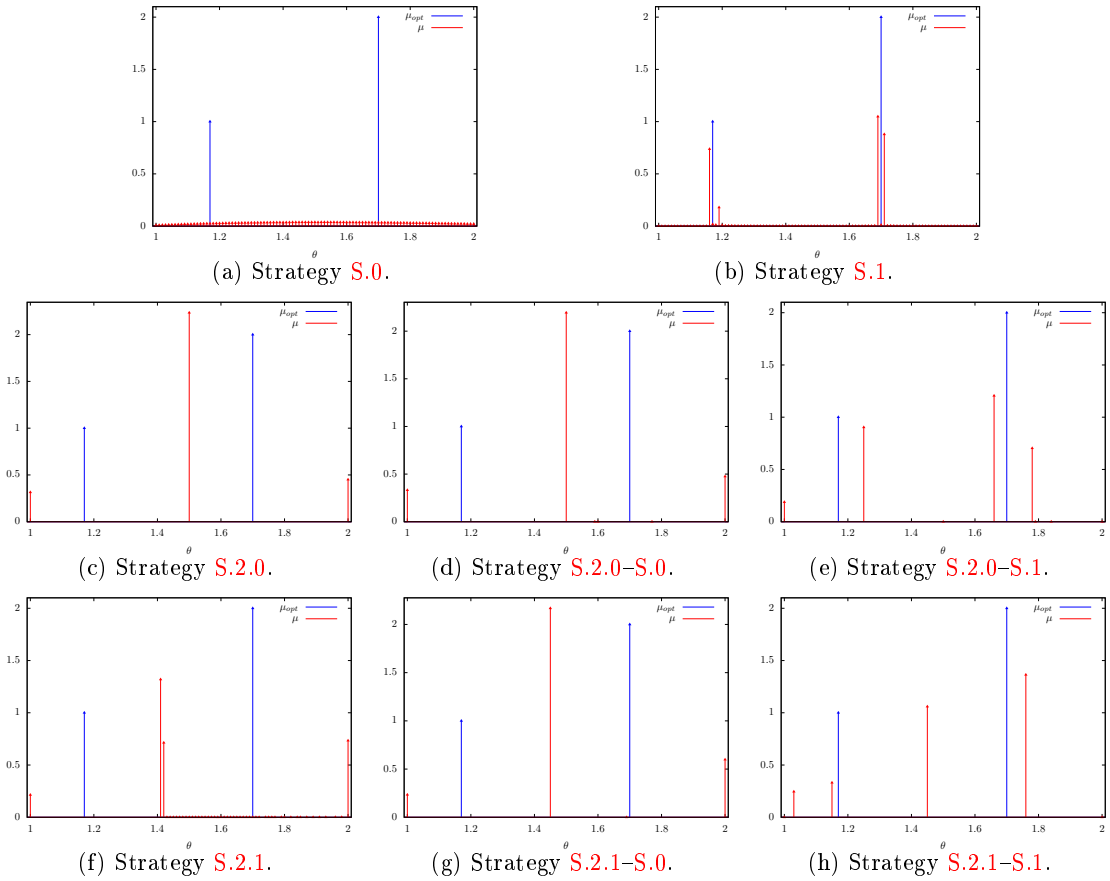
Strategy	Stopping reason
S.0	Cannot decrease the objective function along the current search direction.
S.1	Local minimum possible. Constraints satisfied.
S.2.0	Maximal number of iteration reached.
S.2.0–S.0	Step size is less than the step tolerance.
S.2.0–S.1	Step size is less than the step tolerance.
S.2.1	Maximal number of iteration reached.
S.2.1–S.0	Cost is not decreasing.
S.2.1–S.1	Cost is not decreasing.

(b) Stopping reasons.

Table 2: Comparison of the strategies listed in § 6.3 for Example 7.2.

We see form Table 2 that almost all the conclusion made for Example 7.1 also holds for Example 7.1, except that here the strategy S.1 give slightly better results than the strategy S.2.0–S.1. As for Example 7.1, we display on Figure 5 the adjoints numerically computed. The comments made in Example 7.1 also applies here, wide variation on the adjoint does not necessarily produce large

variations on the control, and the strategies S.0 and S.1 lead to adjoints that are fully supported on Θ_h, while the other strategies lead to adjoints for which the support consist in only few points of Θ_h.



(a) Strategy S.0. (b) Strategy S.1.

(c) Strategy S.2.0. (d) Strategy S.2.0–S.0. (e) Strategy S.2.0–S.1.

(f) Strategy S.2.1. (g) Strategy S.2.1–S.0. (h) Strategy S.2.1–S.1.

Figure 5: Adjoints obtained with strategies listed in § 6.3 for Example 7.1.

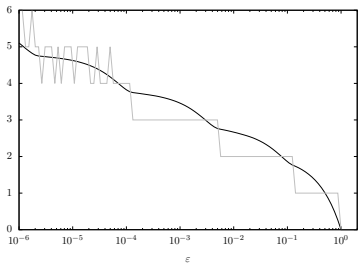
Based on the results given in Examples 7.1 and 7.2, we can make the following comments.

- All strategy lead to some acceptable control.
- Strategies S.0 and S.1 lead to adjoint measure, which are everywhere non-zero. In fact, they are not adapted for sparsity, contrary to strategies S.2.0 and S.2.1 and their derivatives. In addition, for strategy S.1 since an interior-point algorithm is used, the fact that the obtained solution is non-zero everywhere is expected.
- Even if the strategy S.2.1–S.1 did not give the best results, it is relatively fast and provide good results for the control. We will then use this strategy for the following examples.

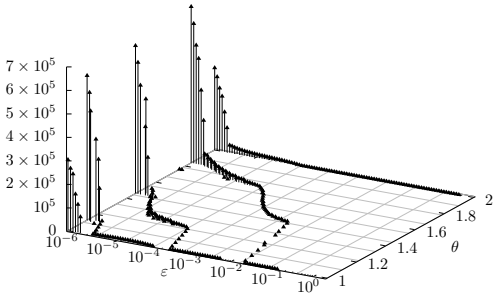
Example 7.3. We consider the system and data given in Example 7.1, but we set $x^1(\theta) = 1$. In that case, it is known from Proposition B.1 that the target state in not reachable. Hence, it is expected that the norm of the minimal L^2-norm control goes to ∞ as ε goes to 0.

This fact is illustrated on Figure 6. On Figure 7, we also display the obtained results for $\varepsilon = 2 \times 10^{-4}$ (we recall that the other parameters are given in the preamble of Section 7). In particular, Figures 7b and 7d confirm the claim of Corollary 3.5, i.e., the sign of the optimal measure

coincide with the sign of $x^1 - x^T$ and the measure is only supported on the set of parameters θ such that $\left|x^1(\theta) - x^T(\theta)\right| = \varepsilon$. The results for this example have been obtained using strategy S.2.1–S.1.

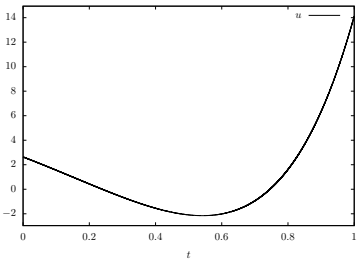


(a) Norm of the control (black) and number of Dirac masses in the adjoint state (gray), with respect to ε.

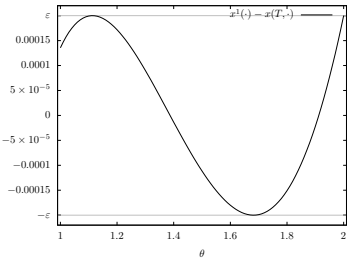


(b) $|\mu_h|$ with respect to ε.

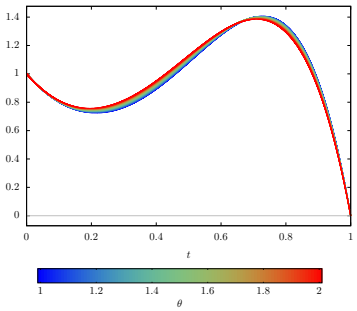
Figure 6: Norm of the control, number of Dirac masses in the adjoint state and absolute value of the adjoint state, with respect to ε, for Example 7.3.



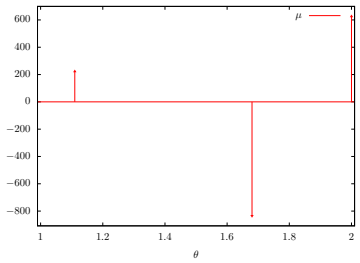
(a) Obtained control ($\|u\|_{L^2(0,T)} \simeq 3.713092$).



(b) Error to the target.



(c) Time dependent solutions for some values of θ. On this graph, we plot $x^1(\theta) - x(\cdot, \theta)$.



(d) Obtained adjoint measure.

Figure 7: Results obtained with strategy S.2.1–S.1 for the system given in Example 7.3, with $\varepsilon = 2 \times 10^{-4}$.

Example 7.4. For this example, we consider a discretized version of the family indexed by $\theta \in \Theta$ of systems of 1D heat equation:

$$\dot{y}(t,\theta,x) = \partial_x^2 y(t,\theta,x) - (\theta-1)y(t,\tau,x) + u(t,x) \qquad (t>0,\ x\in(0,1)), \tag{7.1a}$$
$$y(t,\theta,0) = y(t,\theta,1) = 0 \qquad (t>0), \tag{7.1b}$$
$$y(0,\theta,x) = y^0(\theta,x) \qquad (x\in(0,1)). \tag{7.1c}$$

One can refer to [9] for some ensemble controllability results related to parabolc systems.

Given $n \in \mathbb{N}^*$, using centered finite differences, the discretized version of (7.1) is

$$\dot{Y}(t,\theta) = A(\theta)Y(t,\theta) + B(\theta)U(t), \qquad Y(0,\theta) = Y^0(\theta), \tag{7.2}$$

with

$$A(\theta) = \frac{1}{(n+1)^2}\begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 1 & -2 \end{pmatrix} - (\theta-1)I_n \in \mathbb{R}^{n\times n} \quad \text{and} \quad B(\theta) = I_n \in \mathbb{R}^{n\times n}$$

and where $Y_i(t,\theta) \simeq y(t,\theta,\frac{i}{n+2})$, $U_i(t) \simeq u(t,\frac{i}{n+2})$ and $Y_i^0(\theta) = y^0(\theta,\frac{i}{n+2})$.

Let us mention that using the trapezoidal rule, $\|y(t,\theta,\cdot)\|_{L^2(0,1)}$ is approximated by $\sqrt{\frac{1}{n+1}}\,|Y(t,\theta)|$ hence, the value of ε will be adapted accordingly. For the control problem, we set $y^0(\theta,x) = \sin(x+\theta)$ and null target, i.e., $y^1(\theta) = 0$. The goal is to find a control $u \in L^2(0,T;L^2(0,1))$ such that $\sup_{\theta\in\Theta}\|y(T,\theta,\cdot)\|_{L^2(0,1)} \leqslant \varepsilon$, ε, T, Θ and discretization parameters are defined in the preamble of Section 7. For the numerical illustration, we use the strategy S.2.1–S.1, with $n = 19$. The obtained results are displayed on Figure 8.

A Proof of Proposition 6.1

For this proof, we follow [22, Section 2.4]. We consider a cost function of the form

$$\mathfrak{J}(w_1,\dots,w_K) = F(w_1,\dots,w_K) + \sum_{k=1}^{K} f_k(w_k) \qquad (w_1,\dots,w_K\in\mathbb{R}^n),$$

where $f_1,\dots,f_K$ are continuous convex function and F is a convex function of class $\mathcal{C}^1$. For every $k\in\{1,\dots,K\}$, we define $\breve{w}_k = (w_1,\dots,w_{k-1},w_{k+1},\dots,w_K) \in (\mathbb{R}^n)^{K-1}$, and $\mathfrak{J}_{k,\breve{w}_k}(w_k) = \mathfrak{J}(w)$ and $F_{k,\breve{w}_k}(w_k) = F(w)$.

Theorem A.1. *Using the notations and assumptions introduced above, $w = (w_1,\dots,w_k)$ is a minimizer of $\mathfrak{J}$ if and only if for every $k\in\{1,\dots,K\}$, w_k is a minimizer of $\mathfrak{J}_{k,\breve{w}_k}$.*

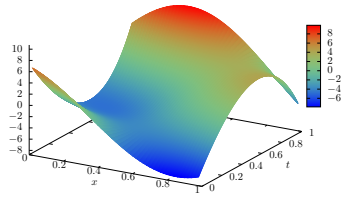
Proof. If w is a minimizer of $\mathfrak{J}$, then for every $\delta\in\mathbb{R}^n$ and every k, we have

$$\mathfrak{J}(w_1,\dots,w_{k-1},w_k+\delta,w_{k+1},\dots,w_k) \geqslant \mathfrak{J}(w_1,\dots,w_{k-1},w_k,w_{k+1},\dots,w_K),$$

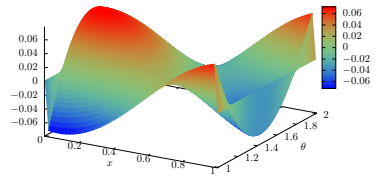
That is to say that $\mathfrak{J}_{k,\breve{w}_k}(w_k+\delta) \geqslant \mathfrak{J}_{k,\breve{w}_k}(w_k)$, i.e., w_k minimizes $\mathfrak{J}_{k,\breve{w}_k}$.

Reciprocally, if w_k minimizes $\mathfrak{J}_{k,\breve{w}_k}$, for every $\delta\in\mathbb{R}^n$ and every $t\in[0,1]$, we have,

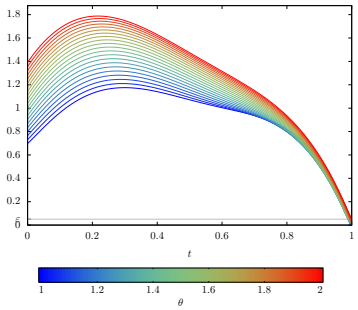
$$0 \leqslant \mathfrak{J}_{k,\breve{w}_k}(w_k+t\delta) - \mathfrak{J}_{k,\breve{w}_k}(w_k) = F_{k,\breve{w}_k}(w_k+t\delta) - F_{k,\breve{w}_k}(w_k) + f_k(w_k+t\delta) - f_k(w_k).$$



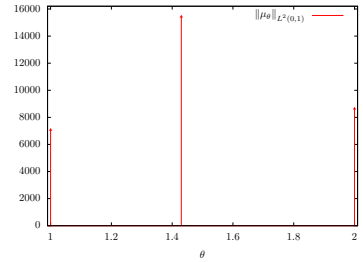
(a) Obtained control ($\|u\|_{L^2((0,T)\times(0,1))} \simeq 3.4857$).



(b) State at final time.



(c) L^2-norm if the time dependent solutions for some values of θ. On this graph, we plot $\|y(t,\theta,\cdot)\|_{L^2(0,1)}$.



(d) L^2-norm of the obtained adjoint measure.

Figure 8: Results obtained with strategy S.2.1–S.1 for the system given in Example 7.4.

But, since f_k is convex, we have $f_k(w_k+t\delta) = f_k\left((1-t)w_k + t(w_k+\delta)\right) \leqslant (1-t)f_k(w_k)+tf_k(w_k+\delta)$, and we get

$$f_k(w_k + t\delta) - f_k(w_k) \leqslant t\left(f_k(w_k+\delta) - f_k(w_k)\right) \qquad (t \in [0,1]).$$

This ensures

$$0 \leqslant \frac{1}{t}\left(F_{k,\breve{w}_k}(w_k + t\delta) - F_{k,\breve{w}_k}(w_k)\right) + f_k(w_k+\delta) - f_k(w_k) \qquad (t \in (0,1)).$$

Since F is of class $\mathcal{C}^1$, $F_{k,\breve{w}_k}$ is also of class $\mathcal{C}^1$ and taking the limit $t \to 0$ in the above equation, we get

$$0 \leqslant \langle \nabla F_{k,\breve{w}_k}(w_k), \delta\rangle + f_k(w_k+\delta) - f_k(w_k).$$

Now, assume that $w = (w_1, \dots, w_K)$ is such that w_k minimizes $F_{k,\breve{w}_k}$ for every $k \in \{1,\dots,K\}$.

For every $\delta = (\delta_1, \dots, \delta_K) \in (\mathbb{R}^n)^K$, we have, using the convexity (and $\mathcal{C}^1$ regularity) of F,

$$\begin{aligned}
\mathfrak{J}(w+\delta) - \mathfrak{J}(w) &= F(w+\delta) - F(w) + \sum_{k=1}^{K} \left(f_k(w_k+\delta_k) - f_k(w_k)\right) \\
&\geqslant \langle \nabla F(w), \delta \rangle + \sum_{k=1}^{K} \left(f_k(w_k+\delta_k) - f_k(w_k)\right) \\
&= \sum_{k=1}^{K} \left(\langle \nabla_{w_k} F(w), \delta_k \rangle + f_k(w_k+\delta_k) - f_k(w_k)\right) \\
&= \sum_{k=1}^{K} \left(\langle \nabla F_{k,\breve{w}_k}(w_k), \delta_k \rangle + f_k(w_k+\delta_k) - f_k(w_k)\right) \\
&\geqslant 0.
\end{aligned}$$

That is to say that w is a minimizer of $\mathfrak{J}$. □

Lemma A.2. *Assume $\mathfrak{J}$ is strictly convex and coercive. Then, for every $k \in \{1, \dots, K\}$, the function $g_k : (\mathbb{R}^n)^{K-1} \to \mathbb{R}^n$ defined by $g_k(\breve{w}) = \operatorname{argmin} \mathfrak{J}_{k,\breve{w}}$ is continuous.*

Proof. Firstly, since we assumed that $\mathfrak{J}$ is strictly convex and coercive, we have that $\mathfrak{J}_{k,\breve{w}}$ is also strictly convex and coercive, this holds for every $\breve{w} \in (\mathbb{R}^n)^{K-1}$ and every $k \in \{1, \dots, K\}$. This ensures the existence and uniqueness of minimizers of $\mathfrak{J}_{k,\breve{w}}$, i.e., the function g_k is well-defined.

Let $\breve{w} \in (\mathbb{R}^n)^{K-1}$ and $(\breve{w}^i)_{i\in\mathbb{N}} \in \left((\mathbb{R}^n)^{K-1}\right)^{\mathbb{N}}$ be a sequence converging to $\breve{w}$, for every $i \in \mathbb{N}$, we define $w_k^i = g_k(\breve{w}^i)$ and $w_k = g_k(\breve{w})$. We then have, $\mathfrak{J}_{k,\breve{w}^i}(w_k^i) \leqslant \mathfrak{J}_{k,\breve{w}^i}(w_k)$, and $\lim_{i\to\infty} \mathfrak{J}_{k,\breve{w}^i}(w_k) = \mathfrak{J}_{k,\breve{w}}(w_k)$. This ensures that the sequence $\left(\mathfrak{J}_{k,\breve{w}^i}(w_k^i)\right)_i$ is bounded. But since $\mathfrak{J}$ is coercive (and recalling that $\mathfrak{J}_{k,\breve{w}^i}(w_k^i) = \mathfrak{J}(w^i)$), this implies that the sequence $(w_k^i)_i$ is bounded, and hence admits a convergent subsequence (still denote by $(w_k^i)_i$), and we denote by w_k^∞ its limit. We then have $\mathfrak{J}_{k,\breve{w}}(w_k^\infty) = \lim_{i\to\infty} \mathfrak{J}_{k,\breve{w}^i}(w_k^i) \leqslant \lim_{i\to\infty} \mathfrak{J}_{k,\breve{w}^i}(w_k) = \mathfrak{J}_{k,\breve{w}}(w_k)$, this implies $w_k = w_k^\infty$, by uniqueness of the minimizer of $\mathfrak{J}_{k,\breve{w}}$. We have thus proved that $\lim_{i\to\infty} g_k(\breve{w}^i) = g_k(\breve{w})$. □

We are now in position to prove the convergence of the algorithm.

Proof of Proposition 6.1. It is trivial that

$$\min_{(\mathbb{R}^n)^K} \mathfrak{J} \leqslant \mathfrak{J}(w^{i+1}) \leqslant \mathfrak{J}(w^i) \qquad (i \in \mathbb{N}). \tag{A.1}$$

This ensures that $\left(\mathfrak{J}(w^i)\right)_i$ is bounded and convergent. Due to the coercivity of $\mathfrak{J}$, we also have, by compactness, the existence of $\phi : \mathbb{N} \to \mathbb{N}$ increasing and $w^\infty \in (\mathbb{R}^n)^K$ such that $\lim_{i\to\infty} w^{\phi(i)} = w^\infty$. The continuity of $\mathfrak{J}$ ensures that $\mathfrak{J}(w^\infty) = \lim_{i\to\infty} \mathfrak{J}(w^{\phi(i)}) = \lim_{i\to\infty} \mathfrak{J}(w^i)$ (the last equality follows from the uniqueness of limits).

For every $k \in \{1, \dots, K\}$ and every $w \in (\mathbb{R}^n)^K$, we define $\Delta\mathfrak{J}_k(w) = \mathfrak{J}(w_1, \dots, w_K) - \mathfrak{J}(w_1, \dots, w_{k-1}, g_k(\breve{w}_k), w_{k+1}, \dots, w_K) \geqslant 0$, where g_k is defined in Lemma A.2. From Lemma A.2, g_k is continuous, hence $\Delta\mathfrak{J}_k$ is also continuous. We then have $\Delta\mathfrak{J}_k(w^\infty) = \lim_{i\to\infty} \Delta\mathfrak{J}_k(w^{\phi(i)})$. We also have, by definition of the sequence,

$$0 \leqslant \Delta\mathfrak{J}_k(w^{\phi(i)}) \leqslant \Delta\mathfrak{J}_{k_{\phi(i)}}(w^{\phi(i)}) = \mathfrak{J}(w^{\phi(i)}) - \mathfrak{J}(w^{\phi(i)+1}) \qquad (i \in \mathbb{N},\ k \in \{1, \cdots, K\}).$$

which goes to 0 as i goes to ∞. Hence, taking the limit $i \to \infty$, we conclude that

$$0 = \Delta\mathfrak{J}_k(w^\infty) \qquad (k \in \{1, \cdots, K\}).$$

This ensures, using the definition of $\Delta\mathfrak{J}_k$ and Theorem A.1, that w^∞ is the minimizer of $\mathfrak{J}$. Note that the existence and uniqueness of the minimizer of $\mathfrak{J}$ is ensured by the coercivity and strict convexity of $\mathfrak{J}$.

Finally, the uniqueness of minimizers, the coercivity of $\mathfrak{J}$, and inequality (A.1) lead to $\lim_{i\to\infty} w^i = w^\infty$. □

B Some unreachable states

In this paragraph, we will show, in the case $m = 1$, that for every $k \in \mathbb{N}$, $\theta \in \Theta \mapsto A(\theta)^k B(\theta) \in \mathbb{R}^n$ is unreachable for the system (2.1). To get this result, we will assume that the system (2.1) is uniformly ensemble controllable, and that Θ has a nonempty interior. Let us first define the set of reachable points in time $T > 0$,

$$\mathcal{R}_{A,B}(T) = \left\{ \theta \in \Theta \mapsto \int_0^T e^{(T-t)A(\theta)} B(\theta) u(t)\, \mathrm{d}t \;\middle|\; u \in L^2(0,T)^m \right\}.$$

Before entering the proof of this result, let us mention that (2.1) is uniformly ensemble controllable if and only if $\mathrm{Span}\left\{\theta \in \Theta \mapsto A(\theta)^k B(\theta) v \;\middle|\; k \in \mathbb{N},\ v \in \mathbb{R}^m\right\}$ is dense in $\mathcal{C}(\Theta)$. This fact can be found in [28] (see also [11]).

Let us first prove our result in the case $n = 1$.

Proposition B.1. *Let $\Theta \subset \mathbb{R}$ be a compact set and assume that there exists $\theta_0 \in \Theta$ such that θ_0 is an accumulation point of Θ. Let $a, b \in \mathcal{C}(\Theta)$ and assume that (a, b) is uniformly ensemble controllable. Then for every $T > 0$, and every $k \in \mathbb{N}$, $a^k b \notin \mathcal{R}_{a,b}(T)$.*

Proof. Let us set $\widetilde{\Theta} = a(\Theta)$, we have $\widetilde{\Theta} \subset \mathbb{R}$ is a compact set, $\widetilde{\theta}_0 = a(\theta_0) \in \widetilde{\Theta}$ is an accumulation point of $\widetilde{\Theta}$, and we set $\imath_{\widetilde{\Theta}}$ the identity map on $\widetilde{\Theta}$. Since the pair (a, b) is ensemble controllable, we necessarily have that a is injective (see [11, Proposition 4]) and $b(\theta) \neq 0$ for every $\theta \in \Theta$ (see Remark 2.5). This ensures that if $y \in \mathcal{R}_{a,b}(T)$ then $y \circ a^{-1} \in \mathcal{R}_{\imath_{\widetilde{\Theta}}, b\circ a^{-1}}(T) = \left(b \circ a^{-1}\right) \mathcal{R}_{\imath_{\widetilde{\Theta}},1}(T)$. Consequently, it is enough to prove this result for $a = \imath_\Theta$ and $b = 1$.

Let us assume by contradiction that there exists $k \in \mathbb{N}$ and $u \in L^2(0,T)$ such that $\int_0^T e^{(T-t)\theta} u(t)\, \mathrm{d}t = \theta^k$ for every $\theta \in \Theta$. Since Θ admits an accumulation point, by analytic continuation, we have

$$\int_0^T e^{(T-t)\theta} u(t)\, \mathrm{d}t = \theta^k \qquad (\theta \in \mathbb{R}).$$

For every $\ell \in \mathbb{N}$, we also have by derivation with respect to θ,

$$\int_0^T (T-t)^{k+1+\ell} e^{(T-t)\theta} u(t)\, \mathrm{d}t = 0 \qquad (\theta \in \mathbb{R}).$$

Finally, using Stone-Weierstrass Theorem, we conclude that

$$(T-t)^{k+1} e^{(T-t)\theta} u(t) = 0 \qquad (\theta \in \mathbb{R},\ t \in [0,T] \text{ a.e.})$$

and using the fact that $(T-t)^{k+1} e^{(T-t)\theta}$ does not vanish on $(0,T)$, we conclude that $u = 0$. This leads to a contradiction with $\int_0^T e^{(T-t)\theta} u(t)\, \mathrm{d}t = \theta^k$. □

Even if the above result is given in the real context, it can be easily extended to the complex case (i.e., with $a(\theta), b(\theta) \in \mathbb{C}$, complex controls and $\Theta \subset \mathbb{C}$). Its complex version will be used in the next result.
We are now in position to give the general result.

Proposition B.2. *Let $\Theta \subset \mathbb{R}$ be a compact set and assume that $\operatorname{int}\Theta \neq \emptyset$. Let $A \in \mathcal{C}(\Theta)^{n\times n}$ and $B \in \mathcal{C}(\Theta)^n$, and assume that (A, B) is uniformly ensemble controllable. Then for every $T > 0$ and every $k \in \mathbb{N}$, we have*

$$A^k B \notin \mathcal{R}_{A,B}(T).$$

Proof. Following [10, Lemma A.2], there exists $\theta_0 \in \operatorname{int}\Theta$, $\delta > 0$, $\lambda \in \mathcal{C}([\theta_0 - \delta, \theta_0 + \delta], \mathbb{C})$ and $v \in \mathcal{C}([\theta_0 - \delta, \theta_0 + \delta], \mathbb{C})^n$ such that $v(\theta) \neq 0$ and $A(\theta)^* v(\theta) = \lambda(\theta) v(\theta)$ for every $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$. We now notice that if the system (2.1) is ensemble controllable on Θ, then the system given by the pair $(\lambda, v^* B)$ is also uniformly ensemble controllable on $[\theta_0 - \delta, \theta_0 + \delta]$. According to Proposition B.1, we have that $\lambda^k v^* B \notin \mathcal{R}_{\lambda, v^* B}(T)$ for every $T > 0$. This clearly ensures that $A^k B \notin \mathcal{R}_{A,B}(T)$ for every $T > 0$. □

Remark B.3. Let us mention that in the case $m > 1$, the situation is not so trivial.
For instance, let $T > 0$, $v \in \mathcal{C}^\infty([0, T])$ such that $v^{(k)}(0) = v^{(k)}(T) = 0$ for every $k \in \mathbb{N}$, and such that $\int_0^T e^{(T-t)\theta} v(t)\,\mathrm{d}t \neq 0$, and define $\beta(\theta) = \left(\int_0^T e^{(T-t)\theta} v(t)\,\mathrm{d}t\right)^{-1}$. For every $k \in \mathbb{N}$, we have $\theta^k \in \mathcal{R}_{\theta,(1\ \beta)}(T)$. In fact, we clearly have $1 = \int_0^T e^{(T-t)\theta}\beta(\theta) v(t)\,\mathrm{d}t$, from this relation and by integration by parts, one can deduce that $\theta^k = \int_0^T e^{(T-t)\theta}\beta(\theta) v^{(k)}(t)\,\mathrm{d}t$ for every $k \in \mathbb{N}$.

References

[1] A. Agrachev, Y. Baryshnikov, and A. Sarychev. Ensemble controllability by Lie algebraic methods. *ESAIM, Control Optim. Calc. Var.*, 22(4):921–938, 2016.

[2] H. Bauer. *Measure and integration theory. Transl. from the German by Robert B. Burckel*, volume 26 of *De Gruyter Stud. Math.* Berlin: de Gruyter, 2001.

[3] K. Beauchard, J.-M. Coron, and P. Rouchon. Controllability issues for continuous-spectrum systems and ensemble controllability of Bloch equations. *Commun. Math. Phys.*, 296(2):525–557, 2010.

[4] A. Becker and T. Bretl. Approximate steering of a unicycle under bounded model perturbation using ensemble control. *IEEE Transactions on Robotics*, 28(3):580–591, 2012.

[5] J. M. Borwein and Q. J. Zhu. *Techniques of variational analysis*, volume 20 of *CMS Books Math./Ouvrages Math. SMC*. New York, NY: Springer, 2005.

[6] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[7] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[8] X. Chen. Controllability issues of linear ensemble systems over multi-dimensional parameterization spaces, 2020. preprint.

[9] B. Danhane and J. Lohéac. Ensemble controllability of parabolic type equations. *Syst. Control Lett.*, 183:10, 2024. Id/No 105683.

[10] B. Danhane, J. Lohéac, and M. Jungers. Conditions for uniform ensemble output controllability, and obstruction to uniform ensemble controllability. *Math. Control Relat. Fields*, 14(3):1128–1175, 2024.

[11] G. Dirr and M. Schönlein. Uniform and L^q-ensemble reachability of parameter-dependent linear systems. *J. Differ. Equations*, 283:216–262, 2021.

[12] I. Ekeland and R. Témam. *Convex analysis and variational problems.*, volume 28 of *Classics Appl. Math.* Philadelphia, PA: Society for Industrial and Applied Mathematics, unabridged, corrected republication of the 1976 English original edition, 1999.

[13] C. Esteve, B. Geshkovski, D. Pighin, and E. Zuazua. Large-time asymptotics in deep learning, 2021.

[14] G. B. Folland. *Real analysis.* Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[15] P. A. Guth, K. Kunisch, and S. S. Rodrigues. Tracking optimal feedback control under uncertain parameters. *Phys. D*, 467:Paper No. 134245, 13, 2024.

[16] U. Helmke and M. Schönlein. Uniform ensemble controllability for one-parameter families of time-invariant linear systems. *Syst. Control Lett.*, 71:69–77, 2014.

[17] E. Kreindler and P. Sarachik. Constrained linear-quadratic optimization problems with parameter-dependent entries. *Journal of Optimization Theory and Applicationsl*, 198(2):781–804, 2023.

[18] M. Lazar and J. Lohéac. Chapter 8 - control of parameter dependent systems. In E. Trélat and E. Zuazua, editors, *Numerical Control: Part A*, volume 23 of *Handbook of Numerical Analysis*, pages 265–306. Elsevier, 2022.

[19] M. Lazar and C. Molinari. Optimal distributed control of linear parabolic equations by spectral decomposition. *Optimal Control Appl. Methods*, 42(4):891–926, 2021.

[20] J.-S. Li and N. Khaneja. Control of inhomogeneous quantum ensembles. *Phys. Rev. A*, 73:030302, Mar 2006.

[21] J.-S. Li and J. Qi. Ensemble control of time-invariant linear systems with linear parameter variation. *IEEE Trans. Autom. Control*, 61(10):2808–2820, 2016.

[22] Y. Li and S. Osher. Coordinate descent optimization for ℓ^1 minimization with application to compressed sensing; a greedy algorithm. *Inverse Probl. Imaging*, 3(3):487–503, 2009.

[23] W. Miao, G. Cheng, and J.-S. Li. On uniform ensemble controllability of diagonalizable linear ensemble systems. preprint, 2021.

[24] J. Peypouquet. *Convex optimization in normed spaces. Theory, methods and examples.* SpringerBriefs Optim. Cham: Springer, 2015.

[25] M. Schönlein. Computation of open-loop inputs for uniformly ensemble controllable systems. *Math. Control Relat. Fields*, 12(3):813–829, 2022.

[26] M. Schönlein. Polynomial methods to construct inputs for uniformly ensemble reachable linear systems. *Math. Control Signals Syst.*, 36(2):251–296, 2024.

[27] M. Schönlein and U. Helmke. Controllability of ensembles of linear dynamical systems. *Math. Comput. Simul.*, 125:3–14, 2016.

[28] R. Triggiani. Controllability and observability in Banach space with bounded operators. SIAM J. Control 13, 462–491 (1975)., 1974.

[29] E. Zuazua. Controllability and observability of partial differential equations: some results and open problems. In *Handbook of differential equations: evolutionary equations. Vol. III*, Handb. Differ. Equ., pages 527–621. Elsevier/North-Holland, Amsterdam, 2007.