OPTIMAL ACTUATOR DESIGN VIA BRUNOVSKY'S NORMAL FORM

BORJAN GESHKOVSKI AND ENRIQUE ZUAZUA

Abstract. In this paper, by using the Brunovsky normal form, we provide a reformulation of the problem consisting in finding the actuator design which minimizes the controllability cost for finite-dimensional linear systems with scalar controls. Such systems may be seen as spatially discretized linear partial differential equations with lumped controls. The change of coordinates induced by Brunovsky's normal form allows us to remove the restriction of having to work with diagonalizable system dynamics, and does not entail a randomization procedure as done in past literature on diffusion equations or waves. Instead, the optimization problem reduces to a minimization of the norm of the inverse of a change of basis matrix, and allows for an easy deduction of existence of solutions, and for a clearer picture of some of the problem's intrinsic symmetries. Numerical experiments help to visualize these artifacts, indicate further open problems, and also show a possible obstruction of using gradient-based algorithms – this is alleviated by using an evolutionary algorithm.

Contents

Keywords. Brunovsky normal form, controllability, finite-dimensional systems, Kalman rank condition, lumped control, optimal actuator design.

AMS Subject Classification. 93B05, 93B60, 90C26, 34H05.

1. Introduction

Due to their importance in many engineering applications, optimal design problems consisting in finding the location wherein a control of least amplitude actuates and ensures the controllability of the underlying system have been investigated in several works over the past decades, in of both the finite and infinite dimensional dynamical systems context. The simplest setting in which one can formulate the fundamental

Date: August 12, 2021.

problem is that of finite-dimensional linear systems with scalar controls:

$$\boxed{\begin{cases} y'(t) - Ay(t) = bu(t) & \text{in } (0,T), \\ y(0) = y_0, \end{cases}} \tag{1.1}$$

where $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $b \in \mathbb{R}^n$. Let us assume that (A,b) is controllable, namely, that the Kalman rank condition is satisfied:

$$\operatorname{span}\{b, Ab, \dots, A^{n-1}b\} = \mathbb{R}^n. \tag{1.2}$$

Now, it is well-known (see [Zuazua, 2007]) that the control $u(\cdot)$ of minimal $L^2(0,T)$–norm steering (1.1) to 0 in any given time $T > 0$ satisfies

$$\|u\|_{L^2(0,T)} \leqslant \mathfrak{C}(b,T)\,\|y_0\| \tag{1.3}$$

for some constant $\mathfrak{C}(b,T) > 0$ (which also depends on the dynamics A) and for all $y_0 \in \mathbb{R}^n$. So, for fixed $T > 0$, by denoting

$$\mathfrak{C}^*(b,T) := \inf\big\{\mathfrak{C}(b,T) > 0\colon (1.3) \text{ holds}\big\} = \inf_{\|y_0\|=1} \|\Gamma_b(y_0)\|_{L^2(0,T)},$$

where $\mathbb{R}^n \ni y_0 \mapsto \Gamma_b(y_0) = u \in L^2(0,T)$ is the "datum to minimal L^2–norm control" operator, the problem consisting of finding an actuator b which minimizes the cost of control may be formulated as

$$\boxed{\min_{b\in\mathbb{S}^{n-1}} \mathfrak{C}^*(b,T).} \tag{1.4}$$

As it is often done in control theory, looking at problems from the perspective of the *adjoint* may be more illustrative. We recall that (1.1) is controllable if and only if the adjoint system

$$\begin{cases} p'(t) + A^\top p(t) = 0 & \text{in } (0,T), \\ p(T) = p_T, \end{cases} \tag{1.5}$$

is observable in any time $T > 0$, in the sense that there exists a constant $\mathfrak{C}_T(b) > 0$ such that

$$\mathfrak{C}_T(b)\,\|p(0)\|^2 \leqslant \int_0^T \big|\langle b, p(t)\rangle\big|^2\,\mathrm{d}t \tag{1.6}$$

holds for all $p_T \in \mathbb{R}^n$. If we assume assume that $A^\top$ is diagonalizable, i.e., it admits a sequence of eigenvalues $\{\lambda_1, \dots, \lambda_n\}$ and an associated sequence of eigenvectors $\{\Psi_1, \dots, \Psi_n\}$ forming an orthonormal basis of $\mathbb{R}^n$, we may rewrite the smallest observability constant $\mathfrak{C}_T^*(b)$ by using separation of variables. Indeed, since

$$p(t) = \sum_{j=1}^n a_j e^{-\lambda_j(T-t)}\Psi_j \qquad \text{for } \ t \in [0,T],$$

where $a_j := \langle p_T, \Psi_j\rangle$, and setting $c_j := a_j e^{-\lambda_j T}$, it may readily be seen that the smallest constant $\mathfrak{C}_T^*(b) > 0$ such that (1.6) holds can be written as

$$\begin{aligned} \mathfrak{C}_T^*(b) &= \inf_{\sum_{j=1}^n |c_j|^2=1} \int_0^T \left|\sum_{j=1}^n c_j e^{\lambda_j t}\langle b, \Psi_j\rangle\right|^2 \mathrm{d}t \\ &= \inf_{\sum_{j=1}^n |c_j|^2=1} \left(\sum_{j=1}^n c_j^2 \big|\langle b, \Psi_j\rangle\big|^2 \frac{e^{2\lambda_j T}-1}{e^{2\lambda_j}} + 2\sum_{j=1}^n \sum_{k=1}^{j-1} c_k c_j \langle b, \Psi_j\rangle\langle b, \Psi_k\rangle \frac{e^{2(\lambda_j+\lambda_k)T}-1}{2(\lambda_j+\lambda_k)}\right). \end{aligned}$$

However, there is no direct way to simplify the above identity – due to the appearance of cross terms when expanding the square – without making specific assumptions on the coefficients c_j of the initial data (e.g., by randomizing them as done in previous literature, as discussed in a subsequent section).

1.1. **Our contributions.** The goal of this work is to rewrite (1.4) in a problem which is more tractable from both an analytical and computational perspective, and does not require 1). the system to be diagonalizable, or 2). a randomization procedure of the Fourier coefficients of the initial data. We do so by leveraging the finite-dimensional and scalar control structure. Namely,

- By using the Brunovsky normal form ([Brunovskỳ, 1970], see Lemma 2.1), we discover that we can rewrite (1.4) as a minimization problem for the norm of the inverse of a change of basis matrix. In particular, the cost $\mathfrak{C}^*(T, b)$ can be written as the tensor product of a function of T and another function of b. Hence, any optimal actuator b^* is independent of the time horizon T. See Proposition 2.1.
- We further rewrite the reformulated minimization problem as a maximization of the smallest eigenvalue of a related, symmetric and positive definite matrix. (See Lemma 2.2.) This variational formulation allows us to ensure the existence of solutions (see Proposition 2.2) and also an invariance of the cost with respect to orthogonal transformations which commute with the system dynamics A (see Proposition 3.1). The latter, in turn, entails non-uniqueness in some cases.
- Finally, in Section 4, we present numerical experiments on three different examples (in low dimensions) to illustrate the insinuated artifacts and stimulate prospective directions and open problems.

Remark 1. *Note that since the Kalman rank condition is equivalent to having*[1] $\langle b, \Psi\rangle \neq 0$ *for all* $\Psi : A\Psi = \lambda\Psi$*, the functional is nontrivial.*

1.2. **Background.** Actuator optimization problems such as the one studied in this work can be formulated easily for a wide variety of finite and infinite dimensional control systems. In particular, such problems are the motivation of a series of works by Privat, Trélat, and Zuazua [Privat et al., 2013a,b, 2015, 2016, 2017, 2019]. (See also [Gimperlein and Waters, 2017; Bergounioux et al., 2019] for subsequent studies, [Trélat, 2018] for a concise presentation, and [Morris, 2010; Kalise et al., 2018] for problems with fixed initial data.) In these works, Privat, Trélat, and Zuazua consider the setting of linear partial differential equations (typically diffusion equations or waves) – to illustrate their approach, let us consider the adjoint heat equation

$$\begin{cases} -p_t - \Delta p = 0 & \text{in } \Omega \times (0,T), \\ p = 0 & \text{in } \partial\Omega \times (0,T), \\ p|_{t=T} = p_T & \text{in } \Omega, \end{cases} \tag{1.7}$$

[1]This fact follows by a unique continuation argument, see e.g. [Tucsnak and Weiss, 2009, Section 1.5] for more detail (where this property is referred to as the Hautus test); see also [Beauchard and Zuazua, 2011, Lem. 1] where this test referred to as the Shizuta-Kawashima (SK) condition is used in the context of hypocoercivity.

where $\Omega \subset \mathbb{R}^d$. Equation (1.7) is observable in any time $T > 0$ and from any open and non-empty subset $\omega \subset \Omega$ in the sense that the observability inequality

$$\mathfrak{C}_T(\omega)\|p(0)\|^2_{L^2(\Omega)} \leqslant \int_0^T \int_\omega |p(t,x)|^2 \,\mathrm{d}x\,\mathrm{d}t \tag{1.8}$$

for some constant $\mathfrak{C}_T(\omega) > 0$ and for all $p_T \in L^2(\Omega)$. In this setting, the dual and equivalent problem to optimal actuator design, is that of optimal sensor placement, which consisting in answering: *What is the domain $\omega^* \subset \Omega$ with $|\omega^*| = \gamma$ such that the smallest constant $\mathfrak{C}_T(\omega^*) > 0$ for which (1.8) holds, is minimized?*

The approach of these works involves separation of variables using a basis of eigenfunctions $-\Delta\Psi_j = \lambda_j\Psi_j$. Sticking to the design problem for (1.7) – (1.8) for ease of presentation, one would decompose the solution of (1.7) into this basis as $p(t,x) = \sum_{j=1}^\infty a_j e^{-\lambda_j(T-t)}\Psi_j(x)$. If one defines $b_j := a_j e^{-\lambda_j T}$, the shape optimization problem can be addressed by examining

$$\begin{aligned}\mathfrak{C}_T(\omega) &= \inf_{\sum_{j=1}^\infty |b_j|^2=1} \int_0^T \int_\omega \left|\sum_{j=1}^\infty b_j e^{\lambda_j t}\ \Psi_j(x)\right|^2 \,\mathrm{d}x\,\mathrm{d}t \\ &= \inf \sigma\left(\left\{\frac{e^{(\lambda_j+\lambda_k)T}-1}{\lambda_j+\lambda_k}\int_\omega \Psi_j(x)\Psi_k(x)\,\mathrm{d}x\right\}_{j,k}\right),\end{aligned}$$

where σ denotes the spectrum of the intervening infinite-dimensional, symmetric, and nonnegative matrix. This is a challenging spectral optimization problem since little is known about the mixed terms $\int_\omega \Psi_j(x)\Psi_k(x)\,\mathrm{d}x$. Indeed, even in the case of the disk, the restriction of inner products of arbitrary Bessel functions to subsets $\omega \subset \Omega$ cannot be computed explicitly.

In order to avoid computing these mixed terms, Privat, Trélat and Zuazua replace $\{a_j\}_{j\in\mathbb{N}}$ by a sequence of real-valued random variables $\{\beta_j^\nu a_j\}_{j\in\mathbb{N},\nu\in\mathcal{X}}$; the random variables $\{\beta_j^\nu\}_{j\in\mathbb{N},\nu\in\mathcal{X}}$ are independent and identically distributed, of mean 0 and variance 1, and have fast decay (e.g., following a Bernouilli distribution). The authors then study the case of an averaged observability constant, in which the mixed terms vanish when expanding the quadratic term:

$$\begin{aligned}\mathfrak{C}_T^{\mathrm{rand}}(\omega) &= \inf \sigma\left(\left\{\frac{e^{(\lambda_j+\lambda_k)T}-1}{\lambda_j+\lambda_k}\mathbb{E}(\beta_j^\nu\beta_k^\nu)\int_\omega \Psi_j(x)\Psi_k(x)\,\mathrm{d}x\right\}_{j,k}\right) \\ &= \inf \sigma\left(\left\{\frac{e^{2\lambda_j T}-1}{2\lambda_j}\ \int_\omega \Psi_j(x)^2\,\mathrm{d}x\right\}_j\right) \\ &= \inf_{j\in\mathbb{N}} \frac{e^{2\lambda_j T}-1}{2\lambda_j}\int_\omega \Psi_j(x)^2\,\mathrm{d}x.\end{aligned}$$

It is to be noted herein that the randomization hypothesis renders the shape optimization problem significantly more tractable, but of course, with the price that there might be a gap between the deterministic and the randomized problem. Going back to the deterministic problem is thus very challenging, which motivates our approach of reformulating the deterministic control (or observation) cost in a different coordinate system.

Remark 2. *Note that, formulated as such for linear finite-dimensional systems,* (1.4) *does not strictly represent a finite-dimensional, discretized version of localized actuator or sensor problems for partial differential equations (such as* (1.7)*). Rather, whenever A is a numerical discretization of some differential operator in one space dimension (e.g., by finite-differences),* (1.4) *can be seen as finding the optimal controller location for a corresponding* lumped control system. *In the context of the heat equation for instance, this would be*

$$\begin{cases} y_t(t,x) - y_{xx}(t,x) &= b(x)u(t) \quad \text{in } (0,T)\times(0,1), \\ y(t,0) = y(t,1) &= 0 \quad \text{in } (0,T), \end{cases} \tag{1.9}$$

and A could thus represent the finite-difference Laplacian.

Notation. For $n \geqslant 2$, we denote $\mathbb{S}^{n-1} := \{x \in \mathbb{R}^n \,:\, \|x\| = 1\}$, and by $\mathrm{GL}_n(\mathbb{R})$ the group of invertible matrices. Unless otherwise stated, we denote by $\|\cdot\|$ the standard euclidean (ℓ^2) norm.

2. Reformulation via Brunovsky's normal form

We begin our study by motivating and recalling the Brunovsky normal form, as to enhance the clarity of the subsequent results. Consider the n-th order linear equation

$$\zeta^{(n)}(t) + k_1\zeta^{(n-1)}(t) + \ldots + k_n\zeta(t) = u(t), \tag{2.1}$$

with real constant coefficients $\{k_i\}_{i=1}^n$. By setting $z := \left[\zeta \;\; \zeta' \;\ldots\; \zeta^{(n-1)}\right]^\top$, one sees that the above equation is equivalent to the linear system

$$z'(t) = \mathfrak{A}z(t) + \mathbf{e}_n u(t), \tag{2.2}$$

where $\mathbf{e}_n := [0,\ldots,0,1]^\top$ denotes the last vector of the canonical basis of $\mathbb{R}^n$, and

$$\mathfrak{A} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 0 & 1 \\ -k_n & \ldots & \ldots & \ldots & -k_1 \end{bmatrix} \tag{2.3}$$

is a *companion* matrix. A natural question that arises is the converse: *When can a constant coefficient linear system*

$$y'(t) = Ay(t) + bu(t) \tag{2.4}$$

where $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $b \in \mathbb{R}^n$ be transformed to (2.2) *via $y = Pz$ for some invertible matrix $P \in \mathcal{M}_{n\times n}(\mathbb{R})$?*

Note that, should such a relation hold, then

$$\begin{aligned} z'(t) = \left(P^{-1}y\right)'(t) &= \left(P^{-1}AP\right)\left(P^{-1}y\right)(t) + \left(P^{-1}b\right)u(t) \\ &= \left(P^{-1}AP\right)z(t) + \left(P^{-1}b\right)u(t), \end{aligned}$$

and so we are led to ask if there exists an invertible matrix $P \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that $P^{-1}AP$ is a companion matrix, and $P^{-1}b = \mathbf{e}_n$.

To answer such a question, the Brunovsky's normal form comes into play.

Lemma 2.1 (Brunovsky normal form, [Brunovskỳ, 1970])**.** *Let $A \in \mathcal{M}_{n\times n}$ with $n \geqslant 2$ be given. If there exists a vector $b \in \mathbb{R}^n$ such that (A, b) satisfy the Kalman rank condition (1.2), then there exists an invertible matrix $P = P(b) \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that*

$$A = P\mathfrak{A}P^{-1} \qquad \text{and} \qquad b = P\mathbf{e}_n, \tag{2.5}$$

where $\mathfrak{A}$ is the companion matrix of A defined as

$$\mathfrak{A} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 1 \\ -a_n & \dots & \dots & \dots & -a_1 \end{bmatrix}, \tag{2.6}$$

where $\{a_1, \dots, a_n\}$ are the coefficients of the characteristic polynomial of A

$$\det(A - x\mathrm{Id}) := \prod_{j=1}^{n}(x - \lambda_j)^{r_j} = x^n + a_1 x^{n-1} + \dots + a_{n-1}x + a_n = 0.$$

Moreover, the matrix $P(b)$ ensuring (2.5) is unique, its columns $\{f_1, \dots, f_n\}$ being given by

$$f_k = \begin{cases} b & k = n \\ \left(A^{n-k} + \displaystyle\sum_{j=1}^{n-k} a_j A^{n-k-j}\right) b, & 1 \leqslant k \leqslant n-1. \end{cases} \tag{2.7}$$

Conversely, if there exists an invertible matrix $P \in \mathcal{M}_{n\times n}(\mathbb{R})$ such that $A = P\mathfrak{A}P^{-1}$, then (A, b), with $b := P\mathbf{e}_n$, satisfies the Kalman rank condition (1.2).

For the sake of completeness and clarity, we provide a proof in the appendix (see also [Brunovskỳ, 1970; Trélat, 2005]). The Brunovsky normal form has found great success in a variety of contexts, going as far as gradient descent convergence for machine learning applications ([Hardt et al., 2016]). Before proceeding, let us provide some comments.

Remark 3. *A well-known result in linear algebra states that a matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is similar to its companion matrix $\mathfrak{A}$ (i.e., there exists a $P \in \mathrm{GL}_n(\mathbb{R})$ such that $A = P\mathfrak{A}P^{-1}$) if and only if A has a cyclic vector (i.e., there exists some $b \in \mathbb{R}^n$ such that (1.2) holds) – see for instance [Horn and Johnson, 2012, Theorem 3.3.15]. In fact, one sees that Lemma 2.1 is nothing but a rewriting of this fact. Furthermore, both conditions are equivalent to A having all of its eigenspaces with dimension $\leqslant 1$. Hence, a sufficient condition for a square matrix A to be similar to its companion matrix (or equivalently, to have a cyclic vector) is that it has n distinct eigenvalues. This will be the case for the examples we shall consider; a notable one being the finite-difference discretization of the Dirichlet Laplacian in 1d, whose eigenvalues are precisely $\lambda_j = -\frac{4}{h^2}\sin^2\left(\frac{\pi j}{2(n+1)}\right)$ for $j = 1, \dots, n$ (see [Vichnevetsky and Bowles, 1982]).*

By virtue of the change of coordinates provided by Brunovsky canonical form, we can obtain the following result which allows us to consider an equivalent, but more explicit representation of the cost to be minimized.

Proposition 2.1 ((1.4) in Brunovsky coordinates)**.** *Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ with $n \geqslant 2$ be given, and suppose that $b \in \mathbb{R}^n$ is such that (A, b) satisfies the Kalman rank condition (1.2). Then,*

$$\mathfrak{C}(b,T) = \Big(\big\|P^{-1}(\cdot)\big\| \otimes \kappa(\cdot)\Big)(b,T) := \big\|P^{-1}(b)\big\|\, \kappa(T),$$

where $\kappa(T) > 0$ denotes the cost of controllability for $(\mathfrak{A}, \mathbf{e}_n)$ (and thus depends solely on A).

Consequently, whenever A is similar to its companion matrix $\mathfrak{A}$, problem (1.4) is equivalent to

$$\boxed{\inf_{b\in\mathbb{S}^{n-1}} \big\|P^{-1}(b)\big\|\,.} \tag{2.8}$$

Remark 4. *In other words, one sees that now the cost function is independent of T, and hence an optimal design $b^* \in \mathbb{S}^{n-1}$ will be as well. One should avoid confusion in this insight, as clearly a minimal $L^2(0,T)$–norm control will depend on the time horizon since the controllability cost $\mathfrak{C}^*(b,T)$ will too – the splitting of time and controller variables does not contradict existing results which ensure that $\kappa(T)$ decays as $T \nearrow \infty$, and explodes like $\gamma T^{-\frac{n+1}{2}}$ with $\gamma = (n-1)!\left(\mathfrak{A}^{n-1}\cdot \mathbf{e}_{n-1}\right)^{-1}$ as $T \searrow 0$ (see [Seidman, 1988]).*

Proof of Proposition 2.1. Let us suppose that there exists a vector $b \in \mathbb{R}^n$ such that (A, b) is controllable, i.e., (A, b) satisfies (1.2). We consider the system

$$\begin{cases} z'(t) - \mathfrak{A}z(t) = \mathbf{e}_n u(t) & \text{in } \ (0,T), \\ z(0) = z_0, \end{cases} \tag{2.9}$$

which is also controllable. Moreover, given any $T > 0$, there exists a constant $\kappa(T) > 0$ depending only on T and $\mathfrak{A}$ (and thus A) such that the minimal L^2–norm function $u(\cdot)$ ensuring controllability for (1.2) satisfies

$$\|u\|_{L^2(0,T)} \leqslant \kappa(T)\, \|z_0\| \tag{2.10}$$

for all $z_0 \in \mathbb{R}^n$. Note that the cost of control $\kappa(T) > 0$, defined as the smallest constant appearing in (2.10), is a priori independent of $b \in \mathbb{R}^n$, since the companion matrix $\mathfrak{A}$ is itself independent of b and depends only on A via its characteristic polynomial. Since $\mathfrak{A} = P^{-1}AP$ and $\mathbf{e}_n = P^{-1}b$, we see that

$$z'(t) - P^{-1}APz(t) = P^{-1}b\, u(t) \qquad \text{for } \ t \in (0,T), \tag{2.11}$$

and multiplying by P to the left, we obtain

$$(Pz)'(t) - A(Pz)(t) = bu(t) \qquad \text{for } \ t \in (0,T). \tag{2.12}$$

Therefore, with $y = Pz$, we recover the system (1.1) from (2.9) – (2.6). By virtue of the above computations, and (2.10), we deduce

$$\begin{aligned} \|u\|_{L^2(0,T)} \leqslant \kappa(T)\, \|z_0\| &= \kappa(T)\, \big\|P^{-1}y_0\big\| \\ &\leqslant \kappa(T)\, \big\|P^{-1}\big\|\, \|y_0\|\,. \end{aligned}$$

This bound is sharp, as the cost of control of the original system (1.1) is precisely

$$\mathfrak{C}(b,T) := \kappa(T)\, \big\|P^{-1}(b)\big\|\,.$$

This concludes the proof. □

In other words, the transformation induced by writing the Brunovsky normal form of the original system (1.1) has allowed to perform a separation of variables of the control cost. Hence, the problem of choosing the controller $b \in \mathbb{S}^{n-1}$ so that the cost of control of (1.1) is optimized, i.e. (1.4), can be reformulated to the problem of optimizing the norm of the inverse of change-of-basis matrix $P(b)$.

2.1. **Computing the norm of $P^{-1}(b)$.** As we have seen in what precedes, provided there exists $b \in \mathbb{R}^n$ such that (A, b) satisfies the Kalman rank condition, the change-of-basis matrix $P(b) \in \mathrm{GL}_n(\mathbb{R})$ is fully determined out of the coefficients of the characteristic polynomial of A, and the value of b. It would however be convenient to have a simplified description of the norm of $P^{-1}(b)$. The norm which canonically appears in (2.8) is the standard operator norm, namely $\left\|P^{-1}(b)\right\| := \sup_{\|x\|=1} \left\|P^{-1}(b)x\right\|$ (where the underlying norm is the euclidean one), which could be defined as the largest eigenvalue of an associated symmetric and positive definite matrix, and hence avoids computing the inverse.

In fact, one has the following characterization.

Lemma 2.2 (Variational form)**.** *Suppose that $A \in \mathcal{M}_{n\times n}$ with $n \geqslant 2$ is similar to its companion matrix. Problem (2.8) is then equivalent to*

$$\boxed{\max_{b\in\mathbb{S}^{n-1}} \lambda_1\left(P(b)P(b)^\top\right).} \tag{2.13}$$

Here $\lambda_1(M)$ denotes the smallest eigenvalue of a matrix $M \in \mathcal{M}_{n\times n}(\mathbb{R})$.

Proof of Lemma 2.2. Noting that $(P(b)^{-1})^\top P(b)^{-1}$ is a symmetric and positive definite matrix (by virtue of the Kalman rank condition, which holds due to the equivalence with A being similar to its companion matrix), it thus admits a sequence of n real eigenvalues $0 < \lambda_1 \leqslant \ldots \leqslant \lambda_n$. Moreover using classical results from linear algebra, we have

$$\left\|P(b)^{-1}\right\| = \sqrt{\lambda_n\Big((P(b)^{-1})^\top P(b)^{-1}\Big)}, \tag{2.14}$$

and, noting that $\left(P(b)^{-1}\right)^\top = \left(P(b)^\top\right)^{-1}$, we see that

$$\left(P(b)^{-1}\right)^\top P(b)^{-1} = \left(P(b)^\top\right)^{-1} P(b)^{-1} = \left(P(b)P(b)^\top\right)^{-1}.$$

Using once again the symmetry of $P(b)P(b)^\top$, we see that

$$\lambda_n\Big((P(b)P(b)^\top)^{-1}\Big) = \frac{1}{\lambda_1(P(b)P(b)^\top)}. \tag{2.15}$$

Accordingly, by positivity and the convexity of the square root, the optimisation problem (2.8) is equivalent to (2.13). □

Remark 5. *We may, for instance, also consider an explicit representation of the inverse of $P^{-1}(b)$ by the Cayley-Hamilton formula*

$$P^{-1}(b) = \frac{1}{\det(P(b))} \sum_{s=0}^{n-1} P(b)^s \sum_{k_1,k_2,\ldots,k_{n-1}} \prod_{l=1}^{n-1} \frac{(-1)^{k_l+1}}{l^{k_l} k_l!} \mathrm{trace}(P^l(b))^{k_l},$$

where $k_l \geqslant 0$ solve the linear Diophantine equation $s + \sum_{l=1}^{n-1} l k_l = n-1$, and consider the Frobenius norm instead of the standard operator norm in (2.8). Such a formulation is

however not all too appealing for numerical purposes due to the implicit need to solve a Diophantine equation in each iteration of the minimization algorithm.

Another way to characterize the inverse could be by using the Cramer formula, but this becomes difficult to track when $n \geqslant 3$ due to the involved form of the minors composing the adjunct matrix. In any case, such explicit formulas for the inverse of $P(b)$ appear quite convoluted and difficult to use for a further analysis.

In view of the equivalent characterization of (2.8) given by (2.13), and the well-known continuity results for eigenvalues with respect to parameters whenever the underlying matrix possesses such continuity[2], we may deduce the following result.

Proposition 2.2. *Suppose that $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ with $n \geqslant 2$ is similar to its companion matrix. Then, both problems (2.8) and (2.13) admit a solution $b^* \in \mathbb{S}^{n-1}$.*

This result is a priori not evident when looking at the equivalent problem of minimizing the norm of the inverse of $P(b)$, but follows as a direct corollary.

3. Symmetries

A question which merits asking however, and which does not seem that obvious at first glance since it is not quite clear how one may study the convexity of $b \longmapsto P^{-1}(b)$ (or concavity of $b \longmapsto \lambda_1\left(P(b)P(b)^\top\right)$), is that of uniqueness of minimizers (or the lack thereof). There is no reason as to why one may expect uniqueness. In fact, we prove the following result, which stipulates an invariance of the functional with respect to orthogonal transformations which commute with the system dynamics A.

Proposition 3.1 (Invariants)**.** *Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ with $n \geqslant 2$ be similar to its companion matrix, and let $\mathbf{R} \in \mathcal{M}_{n\times n}(\mathbb{R})$ be such that*

(i) $[A, \mathbf{R}] = A\mathbf{R} - \mathbf{R}A = 0$ (i.e. A and $\mathbf{R}$ commute);
(ii) $\mathbf{R} \in \mathcal{M}_{n\times n}$ is orthogonal, meaning that $\mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top\mathbf{R} = \mathrm{Id}_n$.

Then we have that

$$\min_{b\in\mathbb{S}^{n-1}} \left\|P^{-1}(\mathbf{R}b)\right\|^2 = \min_{b\in\mathbb{S}^{n-1}} \left\|P^{-1}(b)\right\|^2. \tag{3.1}$$

In other words, provided a minimizer b^*, one may, provided commutativity with A, rotate b^* to obtain another minimizer $\mathbf{R}b^*$.

For example, as seen in the numerical experiments in the following section, the finite-difference Dirichlet Laplacian in $n = 2$:

$$\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix},$$

commutes with the orthogonal matrices

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

[2]All eigenvalues of a matrix $M(t)$ are continuous functions of t whenever the entries of $M(t)$ are continuous functions of t. This fact holds whether or not $M(\cdot)$ is invertible and/or positive definite (see e.g., [Kato, 2013, pp. 116]).

Proof of Proposition 3.1. We will make use of the characterization (2.14) – (2.15) of the spectral norm of $P^{-1}(\cdot)$. In other words, we recall that since $P(\cdot)P(\cdot)^\top$ is a symmetric and positive definite matrix, we have that

$$\left\|P^{-1}(\cdot)\right\|^2 = \frac{1}{\lambda_1\left(P(\cdot)P(\cdot)^\top\right)}, \tag{3.2}$$

where $\lambda_1\left(P(\cdot)P(\cdot)^\top\right)$ denotes the smallest eigenvalue of $P(\cdot)P(\cdot)^\top$. Let us thus concentrate on investigating the invariance properties of λ_1.

Let $b \in \mathbb{R}^n$ be fixed. We recall that by the Rayleigh's min-max theorem, we have

$$\lambda_1\left(P(b)P(b)^\top\right) := \min_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\langle P(b)P(b)^\top x, x\rangle}{\|x\|^2}.$$

On another hand, making use of (2.7), we may see that

$$P(b) = \underbrace{\begin{bmatrix} p_1(A) & \dots & p_n(A)\end{bmatrix}}_{\in\,\mathcal{M}_{n\times n^2}(\mathbb{R})} \underbrace{\begin{bmatrix} b & & \\ & \ddots & \\ & & b\end{bmatrix}}_{\in\mathcal{M}_{n^2\times n}(\mathbb{R})}, \tag{3.3}$$

where

$$p_k(A) := \begin{cases} A^{n-k} + \displaystyle\sum_{j=1}^{n-k} a_j A^{n-k-j} & \text{for } k \leqslant n-1, \\ \mathrm{Id} & \text{for } \ k = n. \end{cases} \tag{3.4}$$

After some computations using (3.3), we can deduce that

$$P(b)P(b)^\top = \sum_{k=1}^{n} p_k(A) b b^\top p_k(A)^\top. \tag{3.5}$$

The above representation combined with the Rayleigh quotient characterization yield

$$\begin{aligned}\lambda_1\left(P(b)P(b)^\top\right) &:= \min_{x\in\mathbb{R}^n\setminus\{0\}} \sum_{k=1}^{n} \frac{\left\langle p_k(A) b b^\top p_k(A)^\top x, x\right\rangle}{\|x\|^2} \\ &= \min_{x\in\mathbb{R}^n\setminus\{0\}} \sum_{k=1}^{n} \frac{\left\langle b^\top p_k(A)^\top x, b^\top p_k(A)^\top x\right\rangle}{\|x\|^2} \\ &= \min_{x\in\mathbb{R}^n\setminus\{0\}} \sum_{k=1}^{n} \frac{\left\|(p_k(A)b)^\top x\right\|^2}{\|x\|^2}.\end{aligned}$$

Now since $[A, \mathbf{R}] = 0$ we clearly also have $[p_k(A), \mathbf{R}] = 0$ for $k \leqslant n$. Whence for $x \in \mathbb{R}^n$,

$$\left\|(p_k(A)\mathbf{R}b)^\top x\right\|^2 = \left\|(\mathbf{R}p_k(A)b)^\top x\right\|^2 = \left\|(p_k(A)b)^\top \mathbf{R}^\top x\right\|^2$$

holds. Since $\mathbf{R}^\top$ is orthogonal,

$$\frac{\left\|(p_k(A)\mathbf{R}b)^\top x\right\|^2}{\|x\|^2} = \frac{\left\|(p_k(A)b)^\top \mathbf{R}^\top x\right\|^2}{\|\mathbf{R}^\top x\|^2}.$$

Clearly, since $\mathbf{R}$ is invertible,

$$\min_{y\in\mathbb{R}^n\setminus\{0\}}\sum_{k=1}^{n}\frac{\left\|(p_k(A)b)^\top y\right\|^2}{\|y\|^2}=\min_{x\in\mathbb{R}^n\setminus\{0\}}\sum_{k=1}^{n}\frac{\left\|(p_k(A)b)^\top\mathbf{R}^\top x\right\|^2}{\|\mathbf{R}^\top x\|^2},$$

whence we may conclude the proof. □

4. Numerical experiments

We henceforth provide a brief numerical study of the optimization problem. We focus on the reformulation provided by (2.13), which we recall consists in solving

$$\max_{b\in\mathbb{S}^{n-1}}\lambda_1\left(P(b)P(b)^\top\right)=\max_{b\in\mathbb{S}^{n-1}}\min_{x\in\mathbb{R}^n\setminus\{0\}}\frac{\left\langle P(b)P(b)^\top x,x\right\rangle}{\|x\|^2}. \tag{4.1}$$

We recall the synthetic definition of $P(b)$ and characterization of $P(b)P(b)^\top$ in (3.3) and (3.5), respectively. Given a matrix $A\in\mathcal{M}_{n\times n}(\mathbb{R})$ which is similar to its companion matrix, we shall solve numerically the above optimization problem (i.e. find some maximizer $b^*\in\mathbb{R}^n$) by using

- **Case $n=2$:** The `IPOPT` method via `CasADi` ([Andersson et al., 2019]) in `Matlab`.[3] We make use of the power iteration algorithm to find the smallest eigenvalue of the symmetric, positive-definite matrix $P(b)P(b)^\top$ by a simple spectral shift: we first find the largest eigenvalue $\lambda_{\max}$, and then find the largest eigenvalue of $P(b)P(b)^\top-\lambda_{\max}$; the sum of both resulting eigenvalues yields the desired smallest eigenvalue. We emphasize the necessity of not using a pre-defined routine for computing the eigenvalue, due to the fact that automatic differentiation requires a graph-like object to be able to differentiate and obtain gradients, and traceability with respect to the optimization variable is in general not provided in a pre-defined routine.
- **Case $n\geqslant 3$:** Due to a lack of convergence of `IPOPT` for $n\geqslant 3$, which could be due to non-concavity, we make use of an evolutionary algorithm[4]. Namely, we use the *differential evolution* algorithm implemented in `SciPy` ([Storn and Price, 1997]). (Such obstacles have been encountered – and bypassed – by use of a genetic in related works, see [Hébrard and Henrott, 2003; Freitas, 1999].)

The algorithms suffer from a curse of dimensionality and are, at least for the examples presented below, providing answers up to $n\leqslant 10$ (an optimization run for $n=10$ took around $8h$ on a personal computer). We provide three basic experiments to motivate possible characterizations of optimal solutions depending on the symmetry properties of the system dynamics A.

Remark 6. *The likely cause of the lack of convergence of gradient-based methods is the lack of concavity of the functional $b\mapsto\lambda_1\left(P(b)P(b)^\top\right)$. Let us briefly comment on this artifact. By using the Rayleigh characterization of λ_1, we see that to differentiate one needs to inject derivatives inside the* min. *Formally applying Danskin's theorem ([Danskin, 1966]), to differentiate $b\mapsto\lambda_1\left(P(b)P(b)^\top\right)$ it would roughly suffice to differentiate the map $\Psi:b\mapsto\langle Mbb^\top M^\top x,x\rangle$ for fixed $x\in\mathbb{R}^n$, where $M\in\mathcal{M}_{n\times n}(\mathbb{R})$ is fixed. In*

[3] see `https://github.com/borjanG/optimal.controller`. Experiments were conducted on a personal MacBook Pro laptop (2.4 GHz Quad-Core Intel Core i5, 16GB RAM, Intel Iris Plus Graphics 1536 MB).

[4] We thank Emmanuel Trélat for this insight and suggestion.

essence, this reduces to differentiating the square matrix $bb^\top \in \mathcal{M}_{n\times n}(\mathbb{R})$ *with respect to* b *– a first differentiation yields a 3-tensor* $\mathbf{D}^1 \in \mathbb{R}^{n\times n\times n}$ *where* $\mathbf{D}^1_{k,j,\ell} = \partial_{b_\ell}(bb^\top)_{j,k} = b_j\delta_{\ell,k} + b_k\delta_{\ell,j}$*, where* $\delta_{j,k}$ *denotes the Kronecker delta. A second differentiation would yield a 4-tensor* $\mathbf{D}^2 \in \mathbb{R}^{n\times n\times n\times n}$*, where* $\mathbf{D}^2_{j,k,\ell,r} = \partial_{b_r}\left(\mathbf{D}^1_{k,j,\ell}\right) = \delta_{r,j}\delta_{\ell,k} + \delta_{r,k}\delta_{j,\ell}$*. This would mean that the Hessian of* Ψ *is very sparse and possibly not negative-definite.*

Example 4.1 (Heat equation with lumped control)**.** *We begin this section by considering a finite difference discretization of the one-dimensional heat equation*

$$\begin{cases} y_t(t,x) - y_{xx}(t,x) = b(x)u(t) & \text{in } (0,T)\times(0,1), \\ y(t,0) = y(t,1) = 0 & \text{in } (0,T). \end{cases}$$

Here $b(x) \in \mathbb{R}$ *is a scalar function designating the location wherein the controller actuates with amplitude* $u(t)$ *in each time* t*. By using the classical two-point difference scheme for approximating the second derivative, we obtain the system*

$$y_h'(t) - A_{\Delta,h}y_h(t) = b_h u(t) \qquad \text{in } \ (0,T). \tag{4.2}$$

Here $h = \frac{1}{n-1}$ *where* $n \geqslant 2$ *represents the number of spatial grid points, with* $b_h \in \mathbb{R}^n$ *representing the optimization variable, and*

$$A_{\Delta,h} := \frac{1}{h^2}\begin{bmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{bmatrix}$$

being the standard finite-difference discretization of the Dirichlet Laplacian.

Let us henceforth address a couple of illustrative cases. We provide illustrations of the results in Figure 1 and Figure 2.

Case 1): $(n = 2)$*. We shall begin by focusing our attention on the case* $n = 2$*, and thus consider*

$$A_\Delta = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

In this case, several computations can be done explicitly. Indeed, first note that

$$P(b)P(b)^\top = \begin{bmatrix} (2b_1+b_2)^2 + b_1^2 & (2b_1+b_2)(b_1+2b_2) + b_1b_2 \\ (2b_1+b_2)(b_1+2b_2) + b_1b_2 & (b_1+2b_2)^2 + b_2^2 \end{bmatrix},$$

whence

$$\lambda_1\Big(P(b)P(b)^\top\Big) = 4b_1b_2 + 3\big(b_1^2+b_2^2\big) - 2\Big(\big(2b_1^2+2b_1b_2+b_2^2\big)\big(b_1^2+2b_1b_2+2b_2^2\big)\Big)^{\frac{1}{2}}.$$

Making use of Lagrange multipliers and symbolic computation, one can find that the above function has 4 *maximizers. Numerically, we find the following* 4 *maximizers:*

$$b^* = \begin{bmatrix} b_1^* \\ b_2^* \end{bmatrix} \in \left\{ \begin{bmatrix} -0.257983 \\ 0.96614944 \end{bmatrix}, \begin{bmatrix} 0.257983 \\ -0.96614944 \end{bmatrix}, \begin{bmatrix} 0.96614944 \\ -0.257983 \end{bmatrix}, \begin{bmatrix} -0.96614944 \\ 0.257983 \end{bmatrix} \right\}. \tag{4.3}$$

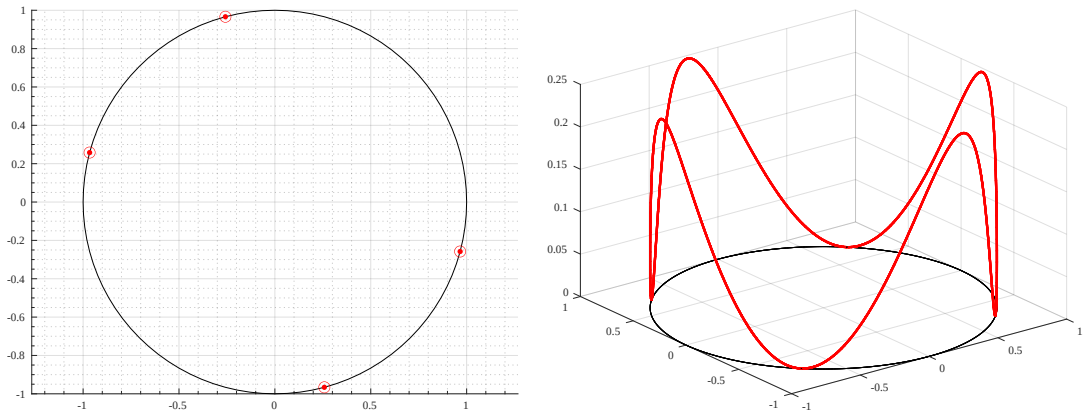

FIGURE 1. **Example 4.1** ($n = 2$). *Left*: The 4 maximizers on $\mathbb{S}^1$ found by the `IPOPT` algorithm, as indicated in (4.3). *Right*: the graph of the function $\mathbb{S}^1 \ni b \mapsto \lambda_1(P(b)P(b)^\top)$, wherein we see 1). the maximum equal to 0.24913 attained at the computed maximizers located on the left plot; 2). the zeros are attained at points which do not satisfy the Kalman rank condition, which are precisely the 4 points with $|b_1| = |b_2| = \frac{\sqrt{2}}{2}$; 3). the rotational symmetry of the cost functional.

We depict these maximizers on $\mathbb{S}^1$ *in Figure 1. Interestingly enough, we see that*

$$\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0.96614944 \\ -0.257983 \end{bmatrix} = \begin{bmatrix} -0.96614944 \\ 0.257983 \end{bmatrix}$$
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0.96614944 \\ -0.257983 \end{bmatrix} = \begin{bmatrix} -0.257983 \\ 0.96614944 \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0.96614944 \\ -0.257983 \end{bmatrix} = \begin{bmatrix} 0.257983 \\ -0.96614944 \end{bmatrix},$$

whence one may generate all the maximizers from $[0.96614944, -0.257983]^\top$ *and applying the orthogonal (rotation) matrices appearing in the identities just above, all of which commute with* A_Δ. *This may also be seen in Figure 1.*

Case 2): ($n = 3$). *We also provide the numerical results in the case* $n = 3$, *and depict the functional to be maximized in Figure 2. We numerically find the following* 8 *maximizers:*

$$b^* \in \left\{ \begin{bmatrix} -0.7633 \\ 0.6325 \\ 0.1311 \end{bmatrix}, \begin{bmatrix} -0.1311 \\ 0.6325 \\ -0.7633 \end{bmatrix}, \begin{bmatrix} -0.1311 \\ -0.6325 \\ 0.7633 \end{bmatrix}, \begin{bmatrix} 0.7633 \\ -0.6325 \\ -0.1311 \end{bmatrix}, \right. \tag{4.4}$$
$$\left. \begin{bmatrix} -1.346 * 10^{-7} \\ 0.44707 \\ -0.8944 \end{bmatrix}, \begin{bmatrix} 4.975 * 10^{-7} \\ -4.44707 \\ 0.8944 \end{bmatrix}, \begin{bmatrix} -9.089 * 10^{-8} \\ 0.44707 \\ -0.8944 \end{bmatrix}, \begin{bmatrix} -4.8519 * 10^{-8} \\ 0.44707 \\ -0.8944 \end{bmatrix} \right\}.$$

We again note a similar rotational symmetry among the obtained maximizers. The latter can be visualized as the peaks in brightly colored patches in Figure 2. We do not conjecture that these maximizers are the sole ones that the functional possesses, as the yellow patches appearing in Figure 2 could contain multiple peaks.

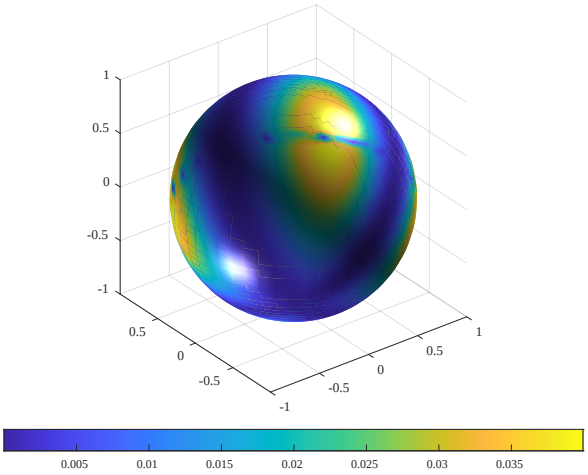

FIGURE 2. **Example 4.1** ($n = 3$). The functional $b \mapsto \lambda_1(P(b)P(b)^\top)$ on $\mathbb{S}^2$; the opposite side of the sphere manifests the same pattern. We dispose of 8 maximizers at which the maximum value equal to ~ 0.0399 is attained. Rotational symmetry is also apparent.

Example 4.2 (Wave equation with lumped control)**.** *We now consider a finite-difference discretization of the one-dimensional wave equation with lumped control:*

$$\begin{cases} z_{tt}(t,x) - z_{xx}(t,x) &= b(x)u(t) \quad \text{in } (0,T)\times(0,1), \\ z(t,0) = z(t,1) &= 0 \quad \text{in } (0,T). \end{cases}$$

By setting $y := [z, z_t]^\top$, we rewrite the equation in the above system in the canonical first-order form as

$$y_t(t,x) - \begin{bmatrix} 0 & \mathrm{Id} \\ \partial_x^2 & 0 \end{bmatrix} y(t,x) = \begin{bmatrix} 0 \\ b(x) \end{bmatrix} u(t) \quad \text{in } (0,T)\times(0,1).$$

When the Dirichlet Laplacian is discretized as in the previous examples, we find ourselves with a linear control system in $\mathbb{R}^{2n}$, with system dynamics

$$A_{\square,h} := \begin{bmatrix} 0 & \mathrm{Id}_n \\ A_{\Delta,h} & 0 \end{bmatrix}$$

with $A_{\Delta,h}$ as in Example 4.1. We depict the shape of the functional $b \mapsto \lambda_1(P(b)P(b)^\top)$ in Figure 4 ($n = 2$) and Figure 5 ($n = 3$). We in fact see that the functional is identical to that of the heat case, thus the found maximizers are as well. This is due to the following result.

Proposition 4.1. *Let $P_\square(b) \in \mathrm{GL}_{2n}(\mathbb{R})$ denote the change-of-basis matrix for $A_{\square,h} \in \mathcal{M}_{2n\times 2n}$, and $P_\Delta(b) \in \mathrm{GL}_n(\mathbb{R})$ that for $A_{\Delta,h} \in \mathcal{M}_{n\times n}(\mathbb{R})$. Then*

$$P_\square(b)P_\square(b)^\top = \begin{bmatrix} P_\Delta(b)P_\Delta(b)^\top & 0 \\ 0 & P_\Delta(b)P_\Delta(b)^\top \end{bmatrix}. \tag{4.5}$$

Consequently, $\lambda_1\left(P_\square(b)P_\square(b)^\top\right) = \lambda_1\left(P_\Delta(b)P_\Delta(b)^\top\right)$.

Proof of Proposition 4.1. We begin by recalling that (we drop the indexes h)

$$P_\square(b)P_\square(b)^\top = \sum_{k=1}^{2n} p_k(A_\square) \begin{bmatrix} 0 & 0 \\ 0 & bb^\top \end{bmatrix} p_k(A_\square)^\top,$$

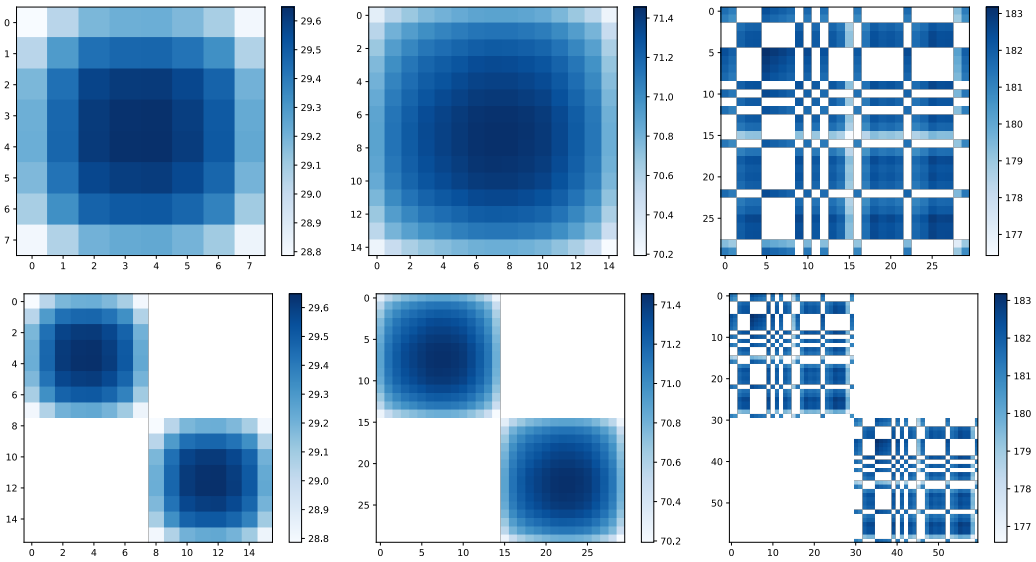

FIGURE 3. Graphical depiction of Proposition 4.1: we display $P_\Delta(b)P_\Delta(b)^\top$ (top) and $P_\square(b)P_\square(b)^\top$ (bottom) for $n \in \{8, 15, 30\}$, with b picked at random per each selected n. More precisely, we display the $\log_{10}$ of these matrices to enhance visibility. An interesting pattern starts to appear for $n \geqslant 26$ as seen on the rightmost figures, likely due to dimensionality.

with

$$p_k(A_\square) := \begin{cases} A_\square^{2n-k} + \sum_{j=1}^{2n-k} a_j^\square A_\square^{2n-k-j} & k \leqslant 2n-1, \\ \mathrm{Id}_{2n} & k = 2n. \end{cases}$$

We distinguish two cases.

Case 1): k is even. One can easily show by induction that

$$A_\square^k = \begin{bmatrix} A_\Delta^{\frac{k}{2}} & 0 \\ 0 & A_\Delta^{\frac{k}{2}} \end{bmatrix}, \tag{4.6}$$

and, moreover, $a_j^\square = 0$ for j odd and $a_{2j}^\square = a_j^\Delta$ for j even. Hence,

$$p_k(A_\square) = \begin{bmatrix} A_\Delta^{\frac{2n-k}{2}} & 0 \\ 0 & A_\Delta^{\frac{2n-k}{2}} \end{bmatrix} + \sum_{j=2}^{2n-k} a_{\frac{j}{2}}^\Delta \begin{bmatrix} A_\Delta^{\frac{2n-k-j}{2}} & 0 \\ 0 & A_\Delta^{\frac{2n-k-j}{2}} \end{bmatrix}.$$

Setting $k = 2\kappa$ and $j = 2r$, we see that

$$p_{2\kappa}(A_\square) = \begin{bmatrix} A_\Delta^{n-\kappa} & 0 \\ 0 & A_\Delta^{n-\kappa} \end{bmatrix} + \sum_{r=1}^{n-\kappa} a_r^\Delta \begin{bmatrix} A_\Delta^{n-\kappa-r} & 0 \\ 0 & A_\Delta^{n-\kappa-r} \end{bmatrix} = \begin{bmatrix} p_\kappa(A_\Delta) & 0 \\ 0 & p_\kappa(A_\Delta) \end{bmatrix}.$$

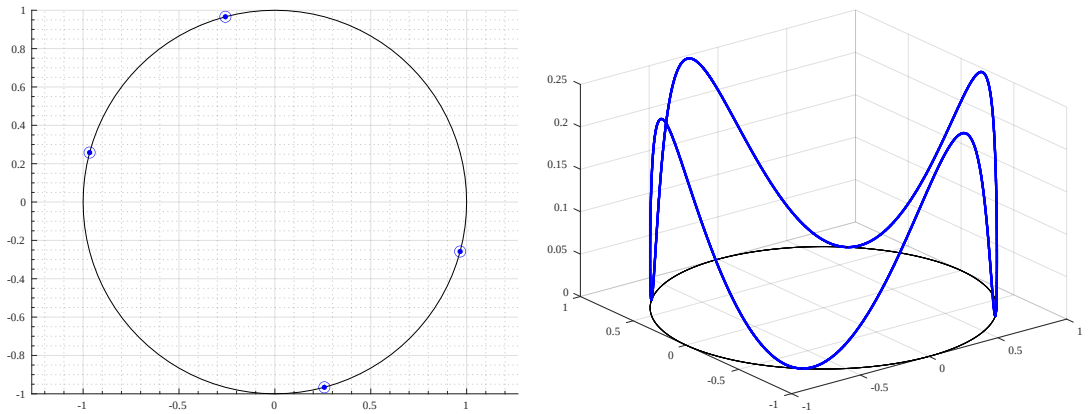

FIGURE 4. **Example 4.2** ($n = 2$). The maximizers and the functional are identical to the heat system in Example 4.1.

Consequently, for $k = 2\kappa$, $\kappa \geqslant 1$,

$$p_{2\kappa}(A_\square)\begin{bmatrix} 0 & 0 \\ 0 & bb^\top \end{bmatrix} p_{2\kappa}(A_\square)^\top = \begin{bmatrix} 0 & 0 \\ 0 & p_\kappa(A_\Delta)bb^\top p_\kappa(A_\Delta)^\top \end{bmatrix}. \tag{4.7}$$

Case 2): k is odd. One can, once again, easily show by induction that

$$A_\square^k = \begin{bmatrix} 0 & A_\Delta^{\frac{k-1}{2}} \\ A_\Delta^{\frac{k+1}{2}} & 0 \end{bmatrix}.$$

Hence,

$$p_k(A_\square) = \begin{bmatrix} 0 & A_\Delta^{\frac{2n-k-1}{2}} \\ A_\Delta^{\frac{2n-k+1}{2}} & 0 \end{bmatrix} + \sum_{j=2}^{2n-k-1} a_{\frac{j}{2}}^\Delta \begin{bmatrix} 0 & A_\Delta^{\frac{2n-k-j-1}{2}} \\ A_\Delta^{\frac{2n-k-j+1}{2}} & 0 \end{bmatrix}.$$

By setting $k = 2\kappa - 1$ with $\kappa \geqslant 1$, and $j = 2r$, we find

$$p_{2\kappa-1}(A_\square) = \begin{bmatrix} 0 & A_\Delta^{n-\kappa} \\ A_\Delta^{n-\kappa+1} & 0 \end{bmatrix} + \sum_{r=1}^{n-\kappa} a_r^\Delta \begin{bmatrix} 0 & A_\Delta^{n-\kappa-r} \\ A_\Delta^{n-\kappa+r+1} & 0 \end{bmatrix}.$$

It then follows that for $\kappa \geqslant 1$,

$$p_{2\kappa-1}(A_\square)\begin{bmatrix} 0 & 0 \\ 0 & bb^\top \end{bmatrix} p_{2\kappa-1}(A_\square)^\top = \begin{bmatrix} p_\kappa(A_\Delta)bb^\top p_\kappa(A_\Delta)^\top & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.8}$$

Combining (4.7) and (4.8), we may conclude. □

Example 4.3 (Advection-diffusion equation with lumped control)**.** *We now consider a system which is non-diagonalizable, hence existing methods based on randomization are not applicable. Namely, we consider the finite difference discretization of the one-dimensional advection-diffusion equation*

$$\begin{cases} y_t(t,x) - y_{xx}(t,x) + y_x(t,x) = b(x)u(t) & (0,T)\times(0,1), \\ y(t,0) = y(t,1) = 0 & (0,T), \end{cases}$$

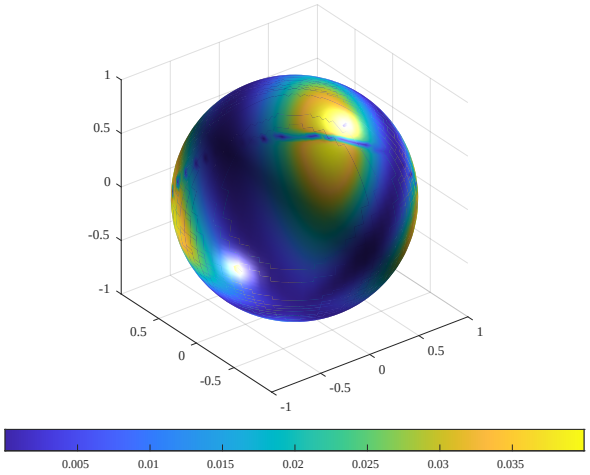

FIGURE 5. **Example 4.2** ($n = 3$). The functional $b \mapsto \lambda_1(P(b)P(b)^\top)$ on $\mathbb{S}^2$ (and thus the maximizers) are the same as for the heat system in Example 4.1.

as well as

$$\begin{cases} y_t(t,x) - y_{xx}(t,x) - y_x(t,x) = b(x)u(t) & (0,T)\times(0,1), \\ y(t,0) = y(t,1) = 0 & (0,T). \end{cases}$$

Using a finite difference approximation as for Example 4.1 and in particular a centered difference scheme for the advection term, we obtain a couple of finite-dimensional control systems with system dynamics of the form

$$A_{\pm\partial_x} := \frac{1}{h^2}\begin{bmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{bmatrix} + \frac{1}{2h}\begin{bmatrix} 0 & \pm 1 & 0 & \dots & 0 \\ \mp 1 & 0 & \pm 1 & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \dots & \mp 1 & 0 & \pm 1 \\ 0 & \dots & \dots & \mp 1 & 0 \end{bmatrix}.$$

We provide illustrations of the results in Figure 6 ($n = 2$) and Figure 7 ($n = 3$).

In the case $n = 2$, the (approximate) maximal value of 0.32236 of the functional (same for both A_{∂_x} and $A_{-\partial_x}$) is attained at the points

$$b^*_{\partial_x} \in \left\{ \begin{bmatrix} -0.9548099 \\ 0.296895 \end{bmatrix}, \begin{bmatrix} 0.9548099 \\ -0.296895 \end{bmatrix} \right\},$$
$$b^*_{-\partial_x} \in \left\{ \begin{bmatrix} -0.296895 \\ 0.9548099 \end{bmatrix}, \begin{bmatrix} 0.296895 \\ -0.9548099 \end{bmatrix} \right\}. \tag{4.9}$$

Note that the maximizers $b^*_{\partial_x}$ and $b^*_{-\partial_x}$ are themselves an axial symmetry of one another.

Similarly, for $n = 3$, we find

$$b^*_{\partial_x} \in \left\{ \begin{bmatrix} -0.8716 \\ 0.4901 \\ -9.34 * 10^{-9} \end{bmatrix}, \begin{bmatrix} -0.8716 \\ 0.4901 \\ 1.246 * 10^{-6} \end{bmatrix} \begin{bmatrix} 0.8716 \\ -0.4901 \\ -7.297 * 10^{-8} \end{bmatrix}, \begin{bmatrix} 0.8716 \\ -0.4901 \\ 1.541 * 10^{-7} \end{bmatrix} \right\}, \tag{4.10}$$

as well as

$$b^*_{-\partial_x} \in \left\{ \begin{bmatrix} -9.229 * 10^{-8} \\ 0.4901 \\ -0.8716 \end{bmatrix}, \begin{bmatrix} -3.581 * 10^{-8} \\ 0.4901 \\ -0.8716 \end{bmatrix} \begin{bmatrix} -2.223 * 10^{-7} \\ -0.4901 \\ 0.8716 \end{bmatrix}, \begin{bmatrix} 1.787 * 10^{-7} \\ -0.4901 \\ 0.8716 \end{bmatrix} \right\}. \tag{4.11}$$

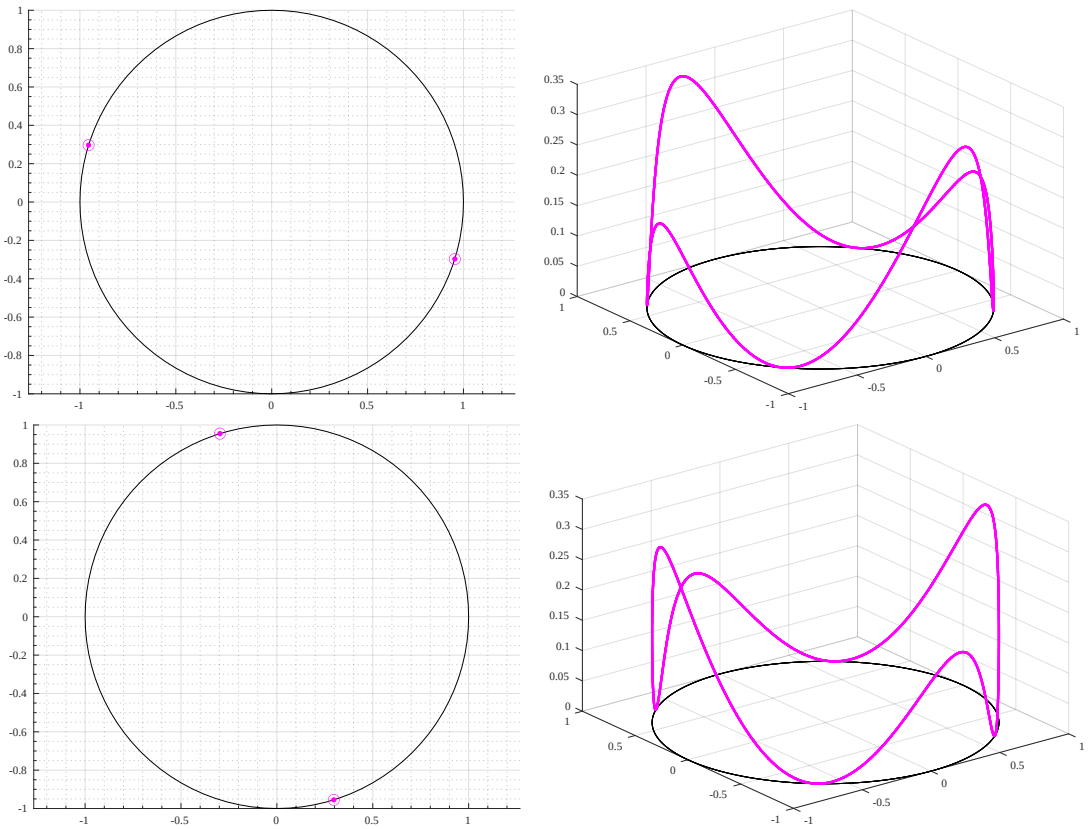

FIGURE 6. **Example 4.3** ($n = 2$). *Left*: The 2 maximizers on $\mathbb{S}^1$ for both $A_{-\partial_x}$ (*top*) and A_{∂_x} (*bottom*), as indicated in (4.9). *Right*: the graph of the function $\mathbb{S}^1 \ni b \mapsto \lambda_1(P(b)P(b)^\top)$ for both $A_{-\partial_x}$ (*top*) and A_{∂_x} (*bottom*), wherein we see that the maximum ~ 0.32236 is attained at the computed maxima located on the left plots; axial symmetry of the maximizers, as well as the rotational symmetry between both functionals is also apparent.

5. Concluding remarks and outlook

By using the Brunovsky normal form, we discovered a reformulation of the problem consisting in finding the actuator which minimizes the controllability cost for finite dimensional linear systems with scalar controls. Such problems can be seen as, for instance, discretizations of one-dimensional lumped control problems for linear partial differential equations. We emphasize the fact that our study does not require the matrix generating the dynamics to be diagonalizable or rely on a randomization procedure of the initial data (as done in past literature in the infinite-dimensional setting).

The Brunovsky reformulation provides a formulation of the control cost as a tensor product as it separates the time horizon and the controller. The resulting optimization problem reduces to the optimization of the norm of the inverse of a change of basis matrix, and allows us to stipulate the existence of minimizers (or maximizers for an equivalent variational problem), as well as non-uniqueness due to an invariance of the cost with respect to orthogonal transformations.

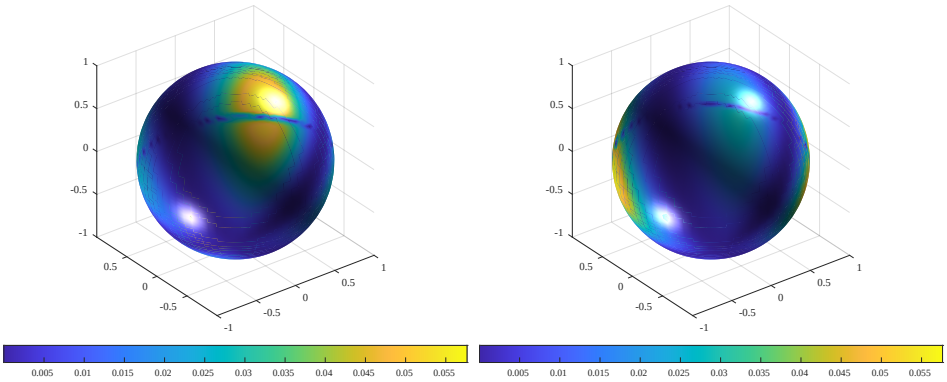

FIGURE 7. **Example 4.3** ($n = 3$). The functional $b \mapsto \lambda_1(P(b)P(b)^\top)$ on $\mathbb{S}^2$. The maximizers (found in (4.10) and (4.11)) for both $A_{-\partial_x}$ (*left*) and A_{∂_x} (*right*) may be found in the bright yellow patches, which replicate on the opposite sides of the sphere.

Let us emphasize several caveats and obstacles regarding our study, which we hope would shed some light on the possible directions of research, in view of providing a complete resolution of the optimal design problem in the deterministic case.

- The optimization of a functional which includes the inverse of a matrix is expected to not scale well with the dimension and thus possibly suffer from a curse of dimensionality. Whence, one should be wary regarding the transfer of the insights of the finite dimensional to the infinite dimensional setting.
- Even after considering the variational reformulation of the problem, which consists in maximizing the first eigenvalue of a positive-definite symmetric matrix, there are no obvious ways (to our knowledge) to solve such a mixed max–min problem over a manifold such as $\mathbb{S}^{n-1}$. In fact, we saw that gradient-based methods seem to fail to converge in dimensions $n \geqslant 3$ – we hence used a global optimization method based on an evolutionary algorithm, which, nonetheless, requires $\sim 8h$ to run when $n = 10$ on a personal machine. We believe that a full clarification of the underlying difficulty of a numerical resolution of this problem in higher dimension, as well as the proposal of novel methods for its resolution are required.

In addition, we believe that there are a multitude of problems regarding the analysis of this problems which ought to be conducted. These include the following.

5.1. **Time-dependent coefficients, neural networks.** Once all of the aforementioned problems are solved, one could look to time-dependent coefficient problems, namely for systems of the form

$$x'(t) - A(t)x(t) = b(t)u(t) \qquad \text{in } (0,T). \tag{5.1}$$

Note that the sparsity of $b(t)$ could also be enhanced imposing other restrictions of the form $\|b(\cdot)\|_{L^1(0,T;\mathbb{R}^n)} = 1$.

Considering systems of the form (5.1) is particularly important in the context of *deep learning* via *continuous-time residual neural networks* (ResNets) (see [Weinan, 2017; Esteve et al., 2020; Ruiz-Balet and Zuazua, 2021; Geshkovski, 2021]), which are

systems taking the form

$$x'(t) = \mathbf{w}(t)\sigma(x(t)) + \mathbf{b}(t) \qquad \text{in} \quad (0,T). \tag{5.2}$$

Here $\mathbf{w}(t) \in \mathcal{M}_{n\times n}(\mathbb{R})$ and $\mathbf{b}(t) \in \mathbb{R}^n$ play the role of the controls, and $\sigma \in \text{Lip}(\mathbb{R})$. Simplifying by assuming that $\sigma = \text{Id}$, fixing $\mathbf{w}(t)$, and writing $\mathbf{b}(t) = bu(t)$ for $b \in \mathbb{R}^n$, we deduce a system of the form (5.1).

For neural networks such as (5.2), minimizing the cost of control by means of controls which are as sparse as possible is clearly relevant for computational purposes due to the high dimensional data involved, and a linear study along with perturbation arguments could yield important insights (see [Yagüe and Geshkovski, 2021] for an optimal control approach to the sparsity issue). There is, of course, a huge gap between the linear constant coefficient case presented above and the study of optimal controllers for ResNets. But, the problems discussed above are deemed necessary in the bigger picture.

5.2. **Uniqueness modulo rotations.** We have seen that optimal actuators are in general not unique due to the invariance of the minimization (or maximization) problem with respect to orthogonal matrices which commute with the dynamics A. It would be of interest to see, at least in very particular test cases, whether a general result can be obtained characterizing the sets of optimal controllers depending on the symmetry properties of the matrix A. In such a case, one could perhaps deduce a uniqueness result modulo the rotated solutions. This insight is reinforced by our numerical simulations in dimensions $n = 2, 3$.

5.3. **Non-scalar controls and PDEs.** The Brunovsky normal form can also be extended to the case $m > 1$, and thus $b \in \mathcal{M}_{n\times m}(\mathbb{R})$. It would be of interest to see how the original problem of finding an optimal b may be reformulated by means of the Brunovsky coordinates in the case $m > 1$. This naturally raises the question of PDE shape design, which seems out of the scope of this particular method.

5.4. **Optimization methods on manifolds.** The algorithms we used need not always converge to a global maximizer lying on $\mathbb{S}^{n-1}$. The algorithm could be enforced by considering optimization methods (including gradient descent) specifically designed to variables lying on manifolds (see e.g., [Boumal, 2020][5]). We leave this open to further investigation.

Acknowledgments. We thank Yannick Privat for generally helpful comments.

Funding. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No.765579-ConFlex. E.Z. has received funding from the Alexander von Humboldt-Professorship program, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement NO. 694126-DyCon), the Transregio 154 Project "Mathematical Modeling, Simulation and Optimization Using the Example of Gas Networks" of the German DFG, grant MTM2017-92996-C2-1-R COSNET of MINECO (Spain), by the Elkartek grant KK-2020/00091 CONVADP of the Basque government and by the Air Force Office of Scientific Research (AFOSR) under Award NO: FA9550-18-1-0242.

[5]We thank Arieh Iserles for this reference.

Appendix A. Auxiliary proofs

Proof of Lemma 2.1. We only prove the first direction of the statement. We split the proof in two steps.

Step 1). Let us first assume that (2.5) is fulfilled for some invertible matrix $\overline{P} \in \mathcal{M}_{n\times n}(\mathbb{R})$, whose columns we denote $\{\overline{f}_k\}_{k=1}^n$. From $b = \overline{P}\mathbf{e}_n$, we immediately deduce that $b = \overline{f}_n$, while each columns of the system $A\overline{P} = \overline{P}\mathfrak{A}$ yields

$$\begin{cases} A\overline{f}_n = \overline{f}_{n-1} - a_1\overline{f}_n \\ A\overline{f}_{n-1} = \overline{f}_{n-2} - a_2\overline{f}_n \\ \quad\vdots \\ A\overline{f}_3 = \overline{f}_2 - a_{n-2}\overline{f}_n \\ A\overline{f}_2 = \overline{f}_1 - a_{n-1}\overline{f}_n \\ A\overline{f}_1 = -a_n\overline{f}_n. \end{cases} \tag{A.1}$$

Here, we recall that $a_1, \dots, a_n$ denote the coefficients of the characteristic polynomial of A. The above relation can readily be rewritten to read as

$$\begin{cases} \overline{f}_n = b, \\ A\overline{f}_k = \overline{f}_{k-1} - a_{n-k}\overline{f}_n, \quad \text{for all } k \in \{2, \dots, n\}, \\ A\overline{f}_1 = -a_n\overline{f}_n. \end{cases} \tag{A.2}$$

Using the fact that (A.2) entails $\overline{f}_{k-1} = A\overline{f}_k + a_{n-k}b$ for $k \geqslant 2$, by a brief induction argument we may further rewrite (A.2) to see that

$$\overline{f}_k = \begin{cases} b, & k = n \\ \left(A^{n-k} + \displaystyle\sum_{j=1}^{n-k} a_j A^{n-k-j}\right) b & 1 \leqslant k \leqslant n-1. \end{cases} \tag{A.3}$$

Step 2). Let us now define

$$P(b) := \Big[f_1 \mid \dots \mid f_n \Big], \tag{A.4}$$

with the columns $\{f_k\}_{k=1}^n$ of $P(b)$ being defined as in (A.3). We shall prove that this $P(b)$ is invertible, and is the unique matrix such that (2.5) holds.

We begin by noting that

$$\begin{aligned} P(b) = \quad & \Big[A^{n-1}b \;\; A^{n-2}b \;\; A^{n-3}b \;\; \dots \;\; A^3b \;\; A^2b \;\; Ab \;\; b\Big] \\ & + a_1\Big[A^{n-2}b \;\; A^{n-3}b \;\; A^{n-4}b \;\; \dots \;\; A^2b \;\; Ab \;\; b \;\; 0\Big] \\ & + a_2\Big[A^{n-3}b \;\; A^{n-4}b \;\; A^{n-5}b \;\; \dots \;\; Ab \;\; b \;\; 0 \;\; 0\Big] \\ & + \dots \\ & + a_{n-3}\Big[\; A^2b \;\; Ab \;\; b \;\; \dots \;\; 0 \;\; 0 \;\; 0 \;\; 0\Big] \\ & + a_{n-2}\Big[\; Ab \;\; b \;\; 0 \;\; \dots \;\; 0 \;\; 0 \;\; 0 \;\; 0\Big] \\ & + a_{n-1}\Big[\; b \;\; 0 \;\; 0 \;\; \dots \;\; 0 \;\; 0 \;\; 0 \;\; 0\Big]. \end{aligned} \tag{A.5}$$

Whence, by the Kalman rank condition, P has full rank and is thus invertible. Left-multiplying the first column in (A.5) by A, one obtains

$$Af_1 = \big(A^n + a_1 A^{n-1} + \ldots + a_{n-2}A^2 + a_{n-1}A\big)b = -a_n b, \tag{A.6}$$

where the rightmost equality is a consequence of the Cayley–Hamilton theorem. Now, the definition of the columns in (A.3) combined with (A.6) leads us to deduce that (A.2) holds for the columns $\{f_k\}_{k=1}^n$. Hence $AP = P\mathfrak{A}$, and one clearly also has $P\mathbf{e}_n = b$. Thus, P defined in (A.4) is invertible and is the unique matrix such that (2.5) holds. This concludes the proof. □

Remark 7 (On the uniqueness of P)**.** *Another way to see that P is the unique invertible matrix such that (2.5) holds is the following. Let P_0 be another matrix such that*

$$A = P_0 \mathfrak{A} P_0^{-1} \qquad \text{and} \qquad b = P_0 \mathbf{e}_n. \tag{A.7}$$

Then, since P is invertible, we may write

$$P_0 = QP \tag{A.8}$$

for some matrix $Q \in \mathcal{M}_{n\times n}(\mathbb{R})$. Thus

$$AQP = QP\mathfrak{A} = QAP, \tag{A.9}$$

so Q commutes with A. Moreover, $QP\mathbf{e}_n = b$. But then

$$A^k b = A^k QP\mathbf{e}_n = QA^k P\mathbf{e}_n = QA^k b \qquad \text{for } k \geqslant 1. \tag{A.10}$$

Since the vectors $A^k b$ span $\mathbb{R}^n$ (by virtue of the Kalman rank condition), we conclude that $Q \equiv \mathrm{Id}$.

References

Andersson, J. A., Gillis, J., Horn, G., Rawlings, J. B., and Diehl, M. (2019). CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1):1–36.

Beauchard, K. and Zuazua, E. (2011). Large time asymptotics for partially dissipative hyperbolic systems. *Archive for rational mechanics and analysis*, 199(1):177–227.

Bergounioux, M., Bretin, É., and Privat, Y. (2019). How to position sensors in thermo-acoustic tomography. *Inverse Problems*, 35(7):074003.

Boumal, N. (2020). An introduction to optimization on smooth manifolds. *Available online, May*.

Brunovskỳ, P. (1970). A classification of linear controllable systems. *Kybernetika*, 6(3):173–188.

Danskin, J. M. (1966). The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, 14(4):641–664.

Esteve, C., Geshkovski, B., Pighin, D., and Zuazua, E. (2020). Large-time asymptotics in deep learning. *arXiv preprint arXiv:2008.02491*.

Freitas, P. (1999). Optimizing the rate of decay of solutions of the wave equation using genetic algorithms: a counterexample to the constant damping conjecture. *SIAM journal on control and optimization*, 37(2):376–387.

Geshkovski, B. (2021). Control in moving interfaces and deep learning.

Gimperlein, H. and Waters, A. (2017). A deterministic optimal design problem for the heat equation. *SIAM Journal on Control and Optimization*, 55(1):51–69.

Hardt, M., Ma, T., and Recht, B. (2016). Gradient descent learns linear dynamical systems. *arXiv preprint arXiv:1609.05191*.

Hébrard, P. and Henrott, A. (2003). Optimal shape and position of the actuators for the stabilization of a string. *Systems & control letters*, 48(3-4):199–209.

Horn, R. A. and Johnson, C. R. (2012). *Matrix analysis*. Cambridge university press.

Kalise, D., Kunisch, K., and Sturm, K. (2018). Optimal actuator design based on shape calculus. *Mathematical Models and Methods in Applied Sciences*, 28(13):2667–2717.

Kato, T. (2013). *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media.

Morris, K. (2010). Linear-quadratic optimal actuator location. *IEEE Transactions on Automatic Control*, 56(1):113–124.

Privat, Y., Trélat, E., and Zuazua, E. (2013a). Optimal location of controllers for the one-dimensional wave equation. In *Annales de l'IHP Analyse non linéaire*, volume 30, pages 1097–1126.

Privat, Y., Trélat, E., and Zuazua, E. (2013b). Optimal observation of the one-dimensional wave equation. *Journal of Fourier Analysis and Applications*, 19(3):514–544.

Privat, Y., Trélat, E., and Zuazua, E. (2015). Optimal shape and location of sensors for parabolic equations with random initial data. *Archive for Rational Mechanics and Analysis*, 216(3):921–981.

Privat, Y., Trélat, E., and Zuazua, E. (2016). Optimal observability of the multi-dimensional wave and Schrödinger equations in quantum ergodic domains. *Journal of the European Mathematical Society*, 18(5):1043–1111.

Privat, Y., Trélat, E., and Zuazua, E. (2017). Actuator design for parabolic distributed parameter systems with the moment method. *SIAM Journal on Control and Optimization*, 55(2):1128–1152.

Privat, Y., Trélat, E., and Zuazua, E. (2019). Spectral shape optimization for the Neumann traces of the Dirichlet-Laplacian eigenfunctions. *Calculus of Variations and Partial Differential Equations*, 58(2):1–45.

Ruiz-Balet, D. and Zuazua, E. (2021). Neural ODE control for classification, approximation and transport. *arXiv preprint arXiv:2104.05278*.

Seidman, T. I. (1988). How violent are fast controls. *Mathematics of Control, Signals and Systems*, 1(1):89–95.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

Trélat, E. (2005). *Contrôle optimal: théorie & applications*. Vuibert.

Trélat, E. (2018). Optimal shape and location of sensors or actuators in PDE models. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3843–3863. World Scientific.

Tucsnak, M. and Weiss, G. (2009). *Observation and control for operator semigroups*. Springer Science & Business Media.

Vichnevetsky, R. and Bowles, J. B. (1982). *Fourier analysis of numerical approximations of hyperbolic equations*. SIAM.

Weinan, E. (2017). A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5(1):1–11.

Yagüe, C. E. and Geshkovski, B. (2021). Sparse approximation in learning via neural ODEs. *arXiv preprint arXiv:2102.13566*.

Zuazua, E. (2007). Controllability and observability of partial differential equations: some results and open problems. In *Handbook of differential equations: evolutionary*

equations, volume 3, pages 527–621. Elsevier.

Borjan Geshkovski
Departamento de Matemáticas
Universidad Autónoma de Madrid
28049 Madrid, Spain
and
Chair of Computational Mathematics
Fundación Deusto
Av. de las Universidades, 24
48007 Bilbao, Basque Country, Spain
Email address: borjan.geshkovski@uam.es

Enrique Zuazua
Chair in Applied Analysis, Alexander von Humboldt-Professorship
Department of Mathematics
Friedrich-Alexander-Universität Erlangen-Nürnberg
91058 Erlangen, Germany
and
Chair of Computational Mathematics
Fundación Deusto
Av. de las Universidades, 24
48007 Bilbao, Basque Country, Spain
and
Departamento de Matemáticas
Universidad Autónoma de Madrid
28049 Madrid, Spain
Email address: enrique.zuazua@fau.de