Optimal Control of the Fokker-Planck Equation with State-Dependent Controls

A. Fleig* and R. Guglielmi†

April 8, 2016

Abstract

For a large class of stochastic processes, the evolution of the underlying probability density function is prescribed by a forward Kolmogorov equation, also called Fokker-Planck equation, a second-order parabolic partial differential equation. In this manner, an optimal control problem subject to an Itô stochastic differential equation can be rendered deterministic by recasting it as an optimal control problem for the Fokker-Planck equation, which we study in this work.

In this setting, the control acts as a coefficient of the state variable in the advection term, i.e., it is of bilinear type. This optimal control problem has been firstly studied by Annunziato and Borzì (2010, 2013) for constant or time dependent controls. We extend the analysis to the case of time and space dependent controls, which allows to consider a wider variety of possible objectives. In order to deduce existence of nonnegative solutions for the state equation we require suitable integrability assumptions on the coefficients of the Fokker-Planck equation and thus on the control function. Therefore, the optimization takes place in a Banach space. We develop a systematic analysis of the existence of optimal controls and derive the system of first order necessary optimality conditions.

Keywords Bilinear control · Fokker-Planck equation · Optimal control theory · Optimization in Banach spaces · Probability density function · Stochastic process

Mathematics Subject Classification (2010) 35Q84 · 35Q93 · 49J20 · 49K20

1 Introduction

Initiated by Kolmogorov's work [20], the study of the Fokker-Planck (FP) equation, also known as Kolmogorov forward equation, has received great and increasing attention, since

*Chair of Applied Mathematics, Department of Mathematics, University of Bayreuth, 95440 Bayreuth, Germany, email: arthur.fleig@uni-bayreuth.de

†Radon Institut for Computational and Applied Mathematics - RICAM, 4040 Linz, Austria, email: roberto.guglielmi@ricam.oeaw.ac.at

it describes the time evolution of the probability density function (PDF) of the velocity of a particle. After that, the FP equation has been applied to several problems in physics, chemistry, and biology and a large amount of literature concerning the FP equation in connection with the evolution of the transition PDF associated with stochastic processes has been developed, see for example [15, 17]. In recent years, the well-posedness of the FP under low regularity assumptions on the coefficients has been studied in connection with existence, uniqueness and stability of martingale solutions to the related stochastic differential equation [23, 11], and control properties of the FP has become of main interest in mean field game theory, see [27] for further insight on this connection.

In a similar way, our main interest in the optimal control of the FP equation stems from a statistical approach that allows to recast a stochastic optimal control problem into a deterministic minimization problem, where the state is constrained to a FP equation. The idea behind this approach is that the state of a stochastic process can be characterized by the shape of its statistical distribution, which is represented by the PDF. Therefore, controlling the PDF yields an accurate and flexible control strategy that can accommodate a wide class of objectives, cf. also [8, Section 4]. In this direction, in [14, 18, 19, 34] PDF-control schemes were proposed, where the cost functional depends on the PDF of the stochastic state variable. In this way, a deterministic objective results and no averaging, which is usually considered for stochastic optimal control problems, see, e.g., [13], is needed.

However, in these references, stochastic methods were still adopted in order to approximate the state variable of the random process. Conversely, in [2, 3] the authors approach the problem of tracking the PDF associated with the stochastic process directly. If the control acts through the drift term, the evolution of the PDF is controlled through the advection term of the FP equation. This is a rather weak action of the controller on the system, usually called of bilinear type, since the control takes action as a coefficient of the state variable. Indeed, only few controllability results are known for such kind of control systems, for instance in connection with quantum control system and stochastic control [7] or in relation to the planning problem for the mean field games system [26]. Concerning the existence of bilinear optimal controls, a first result was given by [1] for a control function that only depends on time. As we will explain in more details below, this choice significantly simplifies the functional framework needed to ensure well-posedness of the FP equation and the related optimal control problem. However, in certain situations it could be handier or even required to act on the space variable as well. In general, the richer structure of a control $u = u(x, t)$ allows to substantially improve the tracking performance of a PDF, as shown in [12].

Therefore, the aim of this work is to extend the theoretical study on the existence of bilinear optimal controls of the FP equation by [1] to the case of more general control functions, i.e., to the case of a bilinear control that depends on both time and space. This feature requires a careful analysis of the well-posedness of the FP equation. Indeed, suitable assumptions are needed on the coefficient of the advection term in order to give meaning to the weak formulation of the equation. For this reason, we will assume the functional framework proposed in the work of Aronson [4] and Aronson-Serrin [5]. In this setting, the advection coefficient belongs to the space $L^q(0, T; L^p(\Omega))$ for some $p, q \geq 1$ such that $1/q + d/(2p) < 1/2$.

This implies, in particular, that the space of controls cannot be chosen as $L^2(Q)$, but only as an appropriate Lebesgue space that is not bounded and closed in $L^2(Q)$. As a result, the optimization problem is defined on a Banach space. This is in contrast with the usual setting for most of the optimization problems, where one exploits the richer Hilbert structure of the space of admissible controls. For example, the control function space always fulfills a Hilbert structure in the linear-quadratic case [22, 25], e.g., the cost functional penalizes the square of the cost effort and the state is constrained to a linear partial differential equation (PDE) with additive control through a source term acting on the boundary or in the interior of the domain. On the other hand, optimization on Banach spaces is often considered whenever the state variable is subject to a nonlinear PDE, see for example [9, 30]. Let us observe that, in recent works [23, 27], the well-posedness of the FP equation has been established even for drift coefficients $b \in L^2(Q)$, in the context of renormalized solutions. These papers could describe the right framework for studying the optimal control problem of the FP equation in a Hilbert setting. A desirable feature of the functional framework adopted in this paper is the fact that we do not require any differentiability property of the control function, which is in accordance with the numerical simulations shown in [12].
In the sequel, after describing the setting in Section 2, in Section 3 we introduce the proper assumptions from [4] on the functional framework to ensure existence of (nonnegative) solutions to the state equation. Section 4 is devoted to recast the FP equation in an abstract setting and to deduce useful a-priori estimates of its solution, which we use to prove our main result on existence of solutions to the optimal control problem in Section 5. In Section 6, we deduce the system of first order necessary optimality conditions that characterizes the optimal solutions. Section 7 concludes.

2 Preliminaries

Given $T > 0$, let us consider a controlled continuous-time stochastic process described by the (Itô) stochastic differential equation

$$\begin{aligned} &dX_t = b(X_t, t; u)\,\mathrm{d}t + \sigma(X_t, t)\,dW_t\,, \qquad t \in (0, T)\,, \\ &X(t=0) = X_0\,, \end{aligned} \tag{1}$$

where $X_0 \in \mathbb{R}^d$ is the initial condition, $d \geq 1$, $W_t \in \mathbb{R}^m$ is an $m-$dimensional Wiener process, $m \geq 1$, $b = (b_1, \ldots, b_d)$ is a vector valued drift function, and the dispersion matrix $\sigma(X_t, t) \in \mathbb{R}^{d\times m}$ is assumed to have full rank. We postpone the analysis of the degenerate diffusion case to a next paper. The control u acting on (1) through the drift term b has to be chosen from a suitable class of admissible functions in a way to minimize a certain cost functional.
Assuming for simplicity that the state variable X_t evolves in a bounded domain Ω of $\mathbb{R}^d$ with smooth boundary $\partial\Omega$, we set the notations $Q := \Omega \times (0, T)$, $\Sigma := \partial\Omega \times (0, T)$, and $a_{ij} = \sum_{k=1}^{d} \sigma_{ik}\sigma_{kj}/2$, $i, j = 1, \ldots d$. Notice that the matrix $\left(a_{ij}\right)_{i,j}$ is symmetric positive semi-definite. We will denote by ∂_i and ∂_t the partial derivative with respect to x_i and t, respectively, where $i = 1, \ldots, d$.

Under suitable assumptions on the coefficients b and σ, it is well known [28, p. 227], [29, p. 297] that, given an initial distribution y_0, the evolution of the PDF associated with the stochastic process (1) satisfies the following FP equation

$$\partial_t y - \sum_{i,j=1}^{d} \partial_{ij}^2 (a_{ij} y) + \sum_{i=1}^{d} \partial_i (b_i(u) y) = 0 \,, \qquad \text{in } Q \,, \tag{2}$$

$$y(x,0) = y_0(x) \,, \qquad \text{in } \Omega \,, \tag{3}$$

where the arguments (x,t) are omitted here and in the following, whenever clear from the context. We refer to [31] for an exhaustive theory and numerical methods on the FP equation. A solution y to (2)-(3) shall furthermore satisfy the standard properties of a PDF, i.e.,

$$y(x,t) \geq 0 \,, \quad (x,t) \in Q \,, \qquad (\textit{positivity})$$

$$\int_\Omega y(x,t)\, \mathrm{d}x = 1 \,, \quad t \in (0,T) \,. \qquad (\textit{unitary mass})$$

Notice that, in general, equation (2) evolves in the space domain $\mathbb{R}^d$ rather than in Ω. However, if localized SDEs are under consideration or if the objective is to keep the PDF within a given compact set of Ω and the probability to find X_t outside of Ω is negligible, we might focus on the description of the evolution of the PDF in the bounded domain $\Omega \subset \mathbb{R}^d$. Thus, assuming that the physical structure of the problem ensures the confinement of the stochastic process within Ω, it is reasonable to employ homogeneous Dirichlet boundary conditions

$$y(x,t) = 0 \quad \text{on } \Sigma \,,$$

also known as absorbing boundary conditions [28, p. 231] (see the work of Feller [10] for a complete characterization of possible boundary conditions in one space dimension; in the multidimensional case, usually either absorbing boundary conditions $y(x,t) = 0$ on Σ, or reflecting boundary conditions $n \cdot J(x,t) = 0$ on Σ are adopted, where J denotes the probability current and n the unit normal vector to the surface $\partial\Omega$. See also [17, Chapter 5] for a comparison between the Gihman-Skorohod [16] and the Feller classification of boundary conditions).

3 Well-posedness of the Fokker–Planck Equation

In this section, we describe the functional framework that we will adopt to ensure the existence of solutions to the FP equation including a source term $f\colon Q \to \mathbb{R}$.

$$\partial_t y - \sum_{i,j=1}^{d} \partial_{ij}^2 (a_{ij} y) + \sum_{i=1}^{d} \partial_i \big(b_i(u) y\big) = f \quad \text{in } Q \,. \tag{4}$$

Assuming that $a_{ij} \in C^1(\overline{Q})$ for all $i, j = 1, \dots, d$, and setting $\tilde{b}_j(u) := \sum_{i=1}^{d} \partial_i a_{ij} - b_j(u)$, equation (4) can be recast in flux formulation

$$\partial_t y - \sum_{j=1}^{d} \partial_j \Big(\sum_{i=1}^{d} a_{ij} \partial_i y + \tilde{b}_j(u) y \Big) = f \,, \quad \text{in } Q\,. \tag{5}$$

Furthermore, we have initial and boundary conditions

$$y(x,t) = 0\,, \qquad (x,t) \in \Sigma\,, \tag{6}$$
$$y(x,0) = y_0(x) \in L^2(\Omega)\,, \qquad x \in \Omega\,, \tag{7}$$

with the associated weak formulation

$$\begin{aligned}\iint\limits_Q fv \,\mathrm{d}x\mathrm{d}t &= \iint\limits_Q \partial_t y v \,\mathrm{d}x\mathrm{d}t - \iint\limits_Q \Big(\sum_{j=1}^{d} \partial_j \Big(\sum_{i=1}^{d} a_{ij} \partial_i y + \tilde{b}_j(u) y \Big)\Big) v \,\mathrm{d}x\mathrm{d}t \\ &= - \iint\limits_Q y \partial_t v \,\mathrm{d}x\mathrm{d}t - \int\limits_\Omega y(\cdot,0) v(\cdot,0) \,\mathrm{d}x + \iint\limits_Q \sum_{j=1}^{d} \Big(\sum_{i=1}^{d} a_{ij} \partial_i y + \tilde{b}_j(u) y \Big) \partial_j v \,\mathrm{d}x\mathrm{d}t\end{aligned}$$

for test functions $v \in W_2^{1,1}(Q)$ with $v_{|\partial\Omega} = 0$ and $v(\cdot,T) = 0$. Here and in the following sections we assume the following hypotheses on the data of equation (5), derived from [4]:

Assumption 1

1. *$a_{ij} \in C^1(\overline{Q})$ for all $i, j = 1, \dots, d$.*
2. *$\forall \xi \in \mathbb{R}^d$ and almost all $(x,t) \in Q$:*
   (a) *$\sum_{i,j=1}^{d} a_{ij}(x,t)\xi_i\xi_j \geq \theta|\xi|^2$ for some constant $0 < \theta < \infty$,*
   (b) *$|a_{ij}(x,t)| \leq M, i, j = 1, ..., d$ for some constant $0 < M < \infty$.*
3. *$f, \tilde{b}_j(u) \in L^q(0,T; L^\infty(\Omega)), j = 1, ..., d$ with $2 < q \leq \infty$.*

Remark 2

1. *As far as well-posedness is concerned, Assumption 1(3) can be weakened to the less demanding requirement [4]*

   $$f, \tilde{b}_j(u) \in L^q(0,T; L^p(\Omega)), j = 1, ..., d \text{ with } 2 < p, q \leq \infty \text{ and } \tfrac{d}{2p} + \tfrac{1}{q} < \tfrac{1}{2}.$$

   However, we directly assume the additional regularity of the coefficients $\tilde{b}_j(u)$, which is required in the following sections.

2. *If the right-hand-side is of the form* $f = \operatorname{div}(F)$ *with* $F\colon Q \to \mathbb{R}^d$*, it is enough to assume* $F_i \in L^2(Q), i = 1, ..., d$ *instead of requiring* f *as in Assumption 1(3), see [4].*

The following theorem, a special case of [4, Thm. 1, p. 634], guarantees the existence and uniqueness of (nonnegative) solutions.

Theorem 3
Suppose that Assumption 1 holds and let $y_0 \in L^2(\Omega)$*. Then there exists a unique* $y \in L^2(0,T; H_0^1(\Omega)) \cap L^\infty(0,T; L^2(\Omega))$ *satisfying*

$$\iint\limits_Q -y\partial_t v + \sum_{j=1}^d \Big(\sum_{i=1}^d a_{ij}\partial_i y + \tilde{b}_j(u)y\Big)\partial_j v - fv \,\mathrm{d}x\mathrm{d}t = \int\limits_\Omega y_0 v(\cdot,0)\,\mathrm{d}x \tag{8}$$

for every $v \in W_2^{1,1}(Q)$ *with* $v_{|\partial\Omega} = 0$ *and* $v(\cdot,T) = 0$*, i.e.,* y *is the unique weak solution of the Fokker-Planck initial boundary value problem defined by* (4), (6)*, and* (7)*.*
Moreover, if $f \equiv 0$ *and* $0 \le y_0 \le m$ *almost everywhere in* Ω*, then*

$$0 \le y(x,t) \le m(1 + C_{FP}k) \quad \textit{almost everywhere in } Q\,,$$

where

$$k := \sum_{j=1}^d \left\|\tilde{b}_j(u)\right\|_{L^q(0,T;L^\infty(\Omega))}$$

and the constant $C_{FP} > 0$ *depends only on* T, Ω*, and the structure of the FP equation.*

The solution obtained by Theorem 3 is more regular; indeed, it belongs to the $W(0,T)$ space. To this end, consider the Gelfand triple $V \hookrightarrow H \hookrightarrow V'$, with $H := L^2(\Omega)$, $V := H_0^1(\Omega)$, and $V' = H^{-1}(\Omega)$ the dual space of V, endowed with norms

$$\|y\|_H^2 := \int\limits_\Omega y^2\,\mathrm{d}x\,, \quad \|y\|_V^2 := \int\limits_\Omega |\nabla y|^2\,\mathrm{d}x\,, \quad \|L\|_{V'} := \sup_{y\in V, \|y\|_V = 1} |\langle L, y\rangle_{V',V}|\,,$$

respectively, where $|\cdot|$ is the Euclidean norm and $\langle\cdot,\cdot\rangle_{V',V}$ represents the duality map between V and V'. This notation and these spaces are used throughout the paper. We remind that

$$W(0,T) := \{y \in L^2(0,T;V) : \dot{y} \in L^2(0,T;V')\} \subset C([0,T];H)\,,$$

$\dot{y}$ denoting the time derivative of y.

Proposition 4
Under the assumptions of Theorem 3 the solution y *in Theorem 3 belongs to* $W(0,T)$*, possibly after a modification on a set of measure zero.*

Proof. The proof is analogous to the proof of [33, Theorem 3.12], the only change being a different functional. In our case, for any fixed t, the linear functional is given by $F(t)\colon H_0^1(\Omega) \to \mathbb{R}$,

$$v \mapsto -\sum_{j=1}^{d} \Big(\sum_{i=1}^{d} a_{ij}(t)\partial_i y(t) + \tilde{b}_j(t; u(t))y(t), \partial_j v \Big)_H + (f(t), v)_H \,.$$

F(t) is bounded and thus continuous for all $t \in (0, T)$:

$$\begin{aligned}
|F(t)v| &= \left| -\sum_{j=1}^{d} \int_\Omega \Big(\sum_{i=1}^{d} a_{ij}(t)\partial_i y(t) + \tilde{b}_j(t; u(t))y(t) \Big) \partial_j v \,\mathrm{d}x + \int_\Omega f(t)v \,\mathrm{d}x \right| \\
&\leq \sum_{j=1}^{d} \int_\Omega \sum_{i=1}^{d} |a_{ij}(t)| \, |\partial_i y(t)| \, |\partial_j v| \,\mathrm{d}x + \int_\Omega |f(t)||v| \,\mathrm{d}x \\
&\quad + \sum_{j=1}^{d} \int_\Omega |\tilde{b}_j(t; u(t))| \, |y(t)| \, |\partial_j v| \,\mathrm{d}x \\
&\leq \underbrace{\sum_{i,j=1}^{d} \|a_{ij}(t)\|_{L^\infty(\Omega)}}_{=:C_1(t)} \|y(t)\|_{H_0^1(\Omega)} \|v\|_{H_0^1(\Omega)} + c_\Omega \|f(t)\|_{L^2(\Omega)} \|v\|_{H_0^1(\Omega)} \\
&\quad + \sum_{j=1}^{d} \left\|\tilde{b}_j(t; u(t))\right\|_{L^\infty(\Omega)} \|y(t)\|_{L^2(\Omega)} \|v\|_{H_0^1(\Omega)} \,,
\end{aligned}$$

where c_Ω is such that $\|v\|_{L^2(\Omega)} \leq c_\Omega \|v\|_{H_0^1(\Omega)}$ for any $v \in H_0^1(\Omega)$. Therefore,

$$\begin{aligned}
\|F(t)\|_{V'} \leq C_1(t) \|y(t)\|_{H_0^1(\Omega)} &+ \sum_{j=1}^{d} \left\|\tilde{b}_j(t; u(t))\right\|_{L^\infty(\Omega)} \|y(t)\|_{L^2(\Omega)} \\
&+ c_\Omega \|f(t)\|_{L^2(\Omega)} \,.
\end{aligned} \tag{9}$$

Since $C_1(t) \in L^\infty(0, T)$, $\|y(t)\|_{H_0^1(\Omega)} \in L^2(0, T)$, $\left\|\tilde{b}_j(t; u(t))\right\|_{L^\infty(\Omega)} \in L^q(0, T)$, $\|y(t)\|_{L^2(\Omega)} \in L^\infty(0, T)$, and $\|f(t)\|_{L^2(\Omega)} \in L^q(0, T)$, $q > 2$, the right-hand-side of (9) belongs to $L^2(0, T)$, i.e., $F \in L^2(0, T; V')$. Note that this result also holds if f is of the form mentioned in Remark 2(2) due to the spatial derivatives being transferred to v. The remaining steps in this case are again the same as in the proof of [33, Theorem 3.12]. $\square$

Note that for all $v \in H_0^1(\Omega)$ we have $\int_\Omega y_0 v \,\mathrm{d}x = \lim_{t\to 0} \int_\Omega y(\cdot, t)v \,\mathrm{d}x = \int_\Omega y(\cdot, 0)v \,\mathrm{d}x$, where the first equality follows from equation (8) and the second holds because $W(0, T) \subset C([0, T]; H)$. Consequently, $y(\cdot, 0) = y_0(\cdot)$ in Ω.

4 A-priori estimates

In this section we deduce a-priori estimates of solutions to the Fokker-Planck initial boundary value problem defined by (4), (6), and (7) with $f \in L^2(0,T;V')$. For the sake of clarity, we recast it in abstract form

$$\begin{cases} \dot{y}(t) + Ay(t) + B(u(t), y(t)) = f(t) \quad \text{in } V' \,, \; t \in (0,T) \\ y(0) = y_0 \,, \end{cases} \tag{10}$$

where $y_0 \in H$, $A\colon V \to V'$ is a linear and continuous operator such that

$$\langle Az, \varphi \rangle_{V',V} = \int_\Omega \sum_{i,j=1}^d \partial_i (a_{ij} z) \, \partial_j \varphi \, \mathrm{d}x \qquad \forall z, \varphi \in V \,,$$

and the operator $B\colon L^\infty(\Omega; \mathbb{R}^d) \times H \to V'$ is defined by

$$\langle B(u,y), \varphi \rangle_{V',V} = - \int_\Omega \sum_{i=1}^d b_i(u) y \, \partial_i \varphi \, \mathrm{d}x = - \int_\Omega y b(u). \nabla \varphi \, \mathrm{d}x$$

for all $u \in L^\infty(\Omega; \mathbb{R}^d)$, $y \in H$, $\varphi \in V$. In the following, $\mathcal{E}(y_0, u, f)$ refers to (10), whenever we want to point out the data (y_0, u, f). From this section on, we denote by M and C generic positive constants that might change from line to line. Furthermore, we assume the following properties.

Assumption 5

1. *Besides Assumption 1(1)-(2), the coefficient functions a_{ij} depend only on space, i.e., $a_{ij} \in C^1(\overline{\Omega})$ with $|a_{ij}(x)|, |\partial_i a_{ij}(x)| \le M$, $i,j = 1, ..., d$ for all $x \in \Omega$ and some constant $0 < M < \infty$.*
2. *The function $b\colon \mathbb{R}^{d+1} \times \mathcal{U} \to \mathbb{R}^d, (x,t;u) \mapsto b(x,t;u(x,t))$ satisfies the growth condition*

$$\sum_{i=1}^d |b_i(x,t;u)|^2 \le M(1 + |x|^2 + |u(x,t)|^2) \quad \forall x \in \mathbb{R}^d \,, \tag{11}$$

for every $i = 1, \dots, d$, $t \in [0,T]$, and u in a suitable space $\mathcal{U}$.

We assume for simplicity the coefficients a_{ij} to be independent of time in order to cope with an autonomous operator A. In this setting, $u(t) \in L^\infty(\Omega; \mathbb{R}^d)$ implies $b(t; u(t)) \in L^\infty(\Omega; \mathbb{R}^d)$, which occurs, in particular, in the case

$$b_i(x,t;u) = \gamma_i(x) + u_i(x,t) \quad \text{for some } \gamma_i \in C^1(\Omega) \,, \; u_i(\cdot, t) \in L^\infty(\Omega) \tag{12}$$

for any $t \in (0,T)$ and $i = 1, \dots, d$. Furthermore, because of (11),

$$\|B(u,y)\|_{V'} \le M(1 + \|u\|_{L^\infty(\Omega;\mathbb{R}^d)}) \, \|y\|_H \qquad \forall u \in L^\infty(\Omega; \mathbb{R}^d) \,, y \in H \,.$$

Given $q > 2$, admissible controls are functions

$$u \in \mathcal{U} := L^q(0,T; L^\infty(\Omega;\mathbb{R}^d)) \subset L^2(0,T; L^\infty(\Omega;\mathbb{R}^d)),$$

for which holds

$$\|u\|_{L^2(0,T;L^\infty(\Omega;\mathbb{R}^d))} \leq T^{\frac{q-2}{2q}} \|u\|_{\mathcal{U}}. \tag{13}$$

To ease the notation, we will still denote by A and B the two operators $A\colon L^2(0,T;V) \to L^2(0,T;V')$ and $B\colon \mathcal{U} \times L^\infty(0,T;H) \to L^q(0,T;V')$ such that for all $z, \varphi \in L^2(0,T;V)$, we have $Az = -\sum_{i,j=1}^d \partial^2_{ij}(a_{ij}z)$ and

$$\int_0^T \langle Az(t), \varphi(t)\rangle_{V',V}\, \mathrm{d}t = \iint_Q \sum_{i,j=1}^d \partial_i(a_{ij}z)\, \partial_j\varphi \, \mathrm{d}x\mathrm{d}t, \tag{14}$$

and $B(u,y) = \sum_{i=1}^d \partial_i(b_i(u)y) = \operatorname{div}(b(u)y)$ for all $u \in \mathcal{U}$ and $y \in L^\infty(0,T;H)$, such that

$$\int_0^T \langle B(u(t),y(t)), \varphi(t)\rangle_{V',V}\, \mathrm{d}t = -\iint_Q \sum_{i=1}^d b_i(u)y\, \partial_i\varphi \, \mathrm{d}x\mathrm{d}t \tag{15}$$

for all $\varphi \in L^{q'}(0,T;V)$ with $1/q + 1/q' = 1$. Indeed, for every $u \in \mathcal{U}$ and $y \in L^\infty(0,T;H)$ we have that $\operatorname{div}(b(u)y) \in L^q(0,T;V')$ and

$$\|B(u,y)\|_{L^q(0,T;V')} = \|\operatorname{div}(b(u)y)\|_{L^q(0,T;V')} \leq M(1 + \|u\|_{\mathcal{U}})\, \|y\|_{L^\infty(0,T;H)}.$$

Note that the integral on the r.h.s. in (14) is not symmetric in z and φ, owing to the fact that A is not self-adjoint.
Let us now consider the bilinear form $a\colon (0,T) \times V \times V \to \mathbb{R}$ such that

$$\begin{aligned} a(t,\psi,\varphi) &:= \int_\Omega \left(\sum_{i,j=1}^d \partial_i(a_{ij}\psi)\, \partial_j\varphi - \sum_{i=1}^d b_i(t,u(t))\psi\partial_i\varphi \right) \mathrm{d}x \\ &= \int_\Omega \left(\sum_{i,j=1}^d a_{ij}\partial_i\psi\, \partial_j\varphi + \sum_{j=1}^d \tilde{b}_j(t,u(t))\psi\partial_j\varphi \right) \mathrm{d}x \end{aligned}$$

for all $t \in (0,T)$ and $\psi, \varphi \in V$. Thanks to the uniform ellipticity of A and Young's inequality, for every $\varepsilon > 0$, $t \in (0,T)$, and every $\varphi \in V$ we have that

$$\begin{aligned} \theta \int_\Omega |\nabla\varphi|^2\, \mathrm{d}x \leq \int_\Omega \sum_{i,j=1}^d a_{ij}\partial_i\varphi\, \partial_j\varphi \, \mathrm{d}x &= a(t,\varphi,\varphi) - \int_\Omega \sum_{j=1}^d \tilde{b}_j(t;u(t))\varphi\partial_j\varphi\, \mathrm{d}x \\ &\leq a(t,\varphi,\varphi) + \|\tilde{b}(t;u(t))\|_{L^\infty(\Omega;\mathbb{R}^d)} \left(\varepsilon \int_\Omega |\nabla\varphi|^2\, \mathrm{d}x + \frac{1}{4\varepsilon}\int_\Omega |\varphi|^2\, \mathrm{d}x \right). \end{aligned}$$

Thus, choosing $\varepsilon = 3\theta/(4\|\tilde{b}(t;u(t))\|_\infty) = 3\theta/(4\|\tilde{b}(t;u(t))\|_{L^\infty(\Omega;\mathbb{R}^d)})$, we conclude that

$$\frac{\theta}{4}\left\|\varphi\right\|_V^2 \le a(t,\varphi,\varphi) + C_1(t)\left\|\varphi\right\|_H^2 , \tag{16}$$

where $C_1(t) = \frac{\left\|\tilde{b}(t;u(t))\right\|_\infty^2}{3\theta}$. We now derive some useful a-priori estimates on the solution of (10).

Lemma 6
Let $y_0 \in H$, $f \in L^2(0,T;V')$ and $u \in \mathcal{U}$. Then a solution y of (10) satisfies the estimates

$$\|y\|^2_{L^\infty(0,T;H)} \le M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')}\right) , \tag{17}$$

$$\|y\|^2_{L^2(0,T;V)} \le (1+\|u\|_{\mathcal{U}}^2)M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')}\right) , \tag{18}$$

$$\|\dot{y}\|^2_{L^2(0,T;V')} \le (1+\|u\|_{\mathcal{U}}^2)M(u)\left(\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')}\right) + 2\left\|f\right\|^2_{L^2(0,T;V')} , \tag{19}$$

where $M(u) := Ce^{c(1+\|u\|_{\mathcal{U}}^2)}$ for some positive constants c, C.

Proof. Let y a solution of (10) and $t \in (0,T)$. Multiplying equation (10) by $y(t)$, we deduce that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right) + a(t,y(t),y(t)) = \langle f(t), y(t)\rangle_{V',V} , \quad t \in (0,T) ,$$

and thus

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right) + \frac{\theta}{2}\left\|y(t)\right\|_V^2 &\le \frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right) + 2a(t,y(t),y(t)) + 2C_1(t)\left\|y(t)\right\|_H^2 \\ &= 2\langle f(t), y(t)\rangle_{V',V} + 2C_1(t)\left\|y(t)\right\|_H^2 \le 2\varepsilon\left\|y(t)\right\|_V^2 + \frac{1}{2\varepsilon}\left\|f(t)\right\|_{V'}^2 + 2C_1(t)\left\|y(t)\right\|_H^2 .\end{aligned}$$

Fixing $\varepsilon = \theta/8$, we can apply Gronwall's inequality and have that

$$\|y(t)\|_H^2 \le e^{\int_0^t 2C_1(\tau)\mathrm{d}\tau}\left[\|y(0)\|_H^2 + \frac{4}{\theta}\int_0^t \|f(\tau)\|_{V'}^2 \,\mathrm{d}\tau\right] .$$

By the definition of C_1 and the assumptions (11) and (13) on b, we deduce that $\int_0^T 2C_1(t)\mathrm{d}t \le M(1+\|u\|_{\mathcal{U}}^2)$, and thus

$$\|y\|^2_{L^\infty(0,T;H)} \le Ce^{c(1+\|u\|_{\mathcal{U}}^2)}\left[\|y(0)\|_H^2 + \|f\|^2_{L^2(0,T;V')}\right] .$$

Integrating in $(0,T)$ the relation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\|y(t)\|_H^2\right) + \frac{\theta}{4}\left\|y(t)\right\|_V^2 \le \frac{4}{\theta}\left\|f(t)\right\|_{V'}^2 + 2C_1(t)\left\|y(t)\right\|_H^2 ,$$

we conclude that

$$\|y\|^2_{L^2(0,T;V)} \le C\left(\|y(0)\|^2_H + \|f\|^2_{L^2(0,T;V')}\right) + C(1+\|u\|^2_{\mathcal{U}})\,\|y\|^2_{L^\infty(0,T;H)}$$
$$\le C(1+\|u\|^2_{\mathcal{U}})e^{c(1+\|u\|^2_{\mathcal{U}})}\left(\|y(0)\|^2_H + \|f\|^2_{L^2(0,T;V')}\right).$$

We remind that C might change from line to line. Finally, multiplying (10) by $\varphi \in L^2(0,T;V)$ and integrating over $(0,T)$,

$$\left|\int_0^T \langle \dot{y}(t), \varphi(t)\rangle_{V',V}\,\mathrm{d}t\right| \le C_\alpha \|y\|_{L^2(0,T;V)}\|\varphi\|_{L^2(0,T;V)}$$
$$+ \|y\|_{L^\infty(0,T;H)}\,\|u\|_{L^2(0,T;L^\infty(\Omega;\mathbb{R}^d))}\,\|\varphi\|_{L^2(0,T;V)} + \|f\|_{L^2(0,T;V')}\,\|\varphi\|_{L^2(0,T;V)},$$

where $C_\alpha > 0$ such that $\|A\xi\|_{V'} \le C_\alpha \|\xi\|_V$ for all $\xi \in V$. Thanks to relation (13),

$$\|\dot{y}\|_{L^2(0,T;V')} \le C_\alpha \|y\|_{L^2(0,T;V)} + C\,\|y\|_{L^\infty(0,T;H)}\,\|u\|_{\mathcal{U}} + \|f\|_{L^2(0,T;V')}.$$

Using twice the estimate $(a+b)^2 \le 2a^2 + 2b^2$, we derive the relation (19) by the estimates on $\|y\|_{L^\infty(0,T;H)}$ and $\|y\|_{L^2(0,T;V)}$. □

5 Existence of optimal control

In this section we consider the minimization problem of a cost functional $\tilde{J}(y,u)$, where the state y is subject to equation (10) with control u and source $f \equiv 0$. We require Assumptions 1 and 5 to hold in this and the following sections.
Fixing $y_0 \in H$, we introduce the control-to-state operator $\Theta\colon \mathcal{U} \to C([0,T];H)$ such that $u \mapsto y \in C([0,T];H)$ solution of $\mathcal{E}(y_0,u,0)$. Thus, the optimization problem turns into minimizing the so-called reduced cost functional $J\colon \mathcal{U} \to \mathbb{R}$ such that $J(u) := \tilde{J}(\Theta(u),u)$, which we assume to be bounded from below, over a suitable non-empty subset of admissible controls $\mathcal{U}_{ad}$. Without loss of generality, we assume the existence of a control $\tilde{u} \in \mathcal{U}_{ad}$ such that $J(\tilde{u}) < \infty$. In the following, we consider box constraints for the space of admissible controls, i.e.,

$$\mathcal{U}_{ad} := \{u \in \mathcal{U} : u_a \le u(x,t) \le u_b \quad \text{for almost all } (x,t) \in Q\} \tag{20}$$

where $u_a, u_b \in \mathbb{R}^d$ and $u_a \le u_b$ is to be understood component-wise. In order to prove the main theorem we will need the following compactness result (see [6], [24, Théorème 5.1, p. 58] or [32]).

Theorem 7
Let I be an open and bounded interval of $\mathbb{R}$, and let X, Y, Z be three Banach spaces, with dense and continuous inclusions

$$Y \hookrightarrow X \hookrightarrow Z,$$

the first one being compact. Then, for every $p \in [1, +\infty)$ and $r > 1$ we have the compact inclusions

$$L^p(I; Y) \cap W^{1,1}(I; Z) \hookrightarrow L^p(I; X)$$

and

$$L^\infty(I; Y) \cap W^{1,r}(I; Z) \hookrightarrow C(\overline{I}; X).$$

Theorem 8

Let $y_0 \in H$ and assume $b(x; u) = (\gamma_i(x) + u_i(x, t))_i$ for some $\gamma_i \in C^1(\Omega)$, $i = 1, \dots, d$. Consider the reduced cost functional $J(u) = \tilde{J}(\Theta(u), u)$, where the function $\Theta(u)(t)$ is the unique solution of the equation

$$\begin{cases} \dot{y}(t) + Ay(t) + B(u(t), y(t)) = 0 \quad \text{in } V', \ t \in (0, T), \\ y(0) = y_0, \end{cases} \tag{21}$$

to be minimized over the controls $u \in \mathcal{U}_{ad}$. Assume that J is bounded from below and (sequentially) weakly-star lower semicontinuous.

Then there exists a pair $(\bar{y}, \bar{u}) \in C([0, T]; H) \times \mathcal{U}_{ad}$ such that $\bar{y}$ solves $\mathcal{E}(y_0, \bar{u}, 0)$ and $\bar{u}$ minimizes J in $\mathcal{U}_{ad}$.

Proof. Let $(u_n)_{n\geq 1}$ be a minimizing sequence converging to $I := \inf_{u\in\mathcal{U}_{ad}} J(u)$. Since $(u_n)_{n\geq 1} \subset \mathcal{U}_{ad}$, we have that $\|u_n\|_{\mathcal{U}} \leq c\|u_n\|_{L^\infty(Q)} \leq C$ for some positive constants c and C, for any $n \geq 1$. Moreover, the pair (u_n, y_n) satisfies the state equation

$$\dot{y}_n(t) + Ay_n(t) + B(u_n(t), y_n(t)) = 0, \quad y_n(0) = y_0. \tag{22}$$

The a-priori estimates of Lemma 6 ensure that there exists a positive constant, still denoted by C, such that

$$\|y_n\|_{L^\infty(0,T;H)}, \quad \|y_n\|_{L^2(0,T;V)}, \quad \|\dot{y}_n\|_{L^2(0,T;V')} \leq C,$$

and so we deduce that

$$\begin{aligned} \|Ay_n\|_{L^2(0,T;V')} &\leq C_\alpha \|y_n\|_{L^2(0,T;V)} \leq C, \\ \|B(u_n, y_n)\|_{L^2(0,T;V')} &\leq c \|B(u_n, y_n)\|_{L^q(0,T;V')} \\ &\leq M(1 + \|u_n\|_{\mathcal{U}}) \|y_n\|_{L^\infty(0,T;H)} \leq C. \end{aligned}$$

Thus, there exist subsequences (still indexed with the subscript n) such that

$$\begin{array}{ll} u_n \overset{*}{\rightharpoonup} \bar{u} & \text{weakly-star in } \mathcal{U}, \\ y_n \overset{*}{\rightharpoonup} \bar{y} & \text{weakly-star in } L^\infty(0, T; H), \\ y_n \rightharpoonup \bar{y} & \text{weakly in } L^2(0, T; V), \\ \dot{y}_n \rightharpoonup \psi & \text{weakly in } L^2(0, T; V'), \\ Ay_n \rightharpoonup \chi & \text{weakly in } L^2(0, T; V'), \\ B(u_n, y_n) \rightharpoonup \Lambda & \text{weakly in } L^2(0, T; V'). \end{array}$$

Since the Banach-Alaoglu theorem ensures that $\mathcal{U}_{ad}$ is weakly-star closed, we deduce that $\bar{u} \in \mathcal{U}_{ad}$. We now want to pass to the limit in the state equation (22). First of all, we observe that $\psi = \dot{\bar{y}}$, thanks to the convergence in the $\sigma(\mathcal{D}(0,T;V), \mathcal{D}'(0,T;V'))$ topology, thus $\bar{y}$ belongs to $W(0,T) \subset C([0,T];H)$. Moreover, since the operator $A\colon L^2(0,T;V) \to L^2(0,T;V')$ is strongly continuous, and therefore weakly continuous too, we deduce that $A\bar{y} = \chi$. Finally, we claim that $B(\bar{u},\bar{y}) = \Lambda$, which, because of the bilinear action of the control, is the most difficult part of the proof. Note that, thanks to the first relation in Theorem 7 with $Y := V$, $X := H$, and $Z := V'$, the embedding $W(0,T) \subset L^2(0,T;H)$ is compact, thus $(y_n)_n$ admits a subsequence strongly convergent to $\bar{y}$ in $L^2(0,T;H)$. Therefore, for every $\varphi \in L^2(0,T;V)$,

$$
\begin{aligned}
&\int_0^T \langle B(\bar{u},\bar{y}) - \Lambda, \varphi\rangle_{V',V}\, \mathrm{d}t \\
&\qquad = -\iint_Q \bar{y}b(\bar{u}).\,\nabla\varphi\, \mathrm{d}x\mathrm{d}t - \lim_{n\to\infty}\int_0^T \langle B(u_n,y_n),\varphi\rangle_{V',V}\, \mathrm{d}t \\
&\qquad = -\iint_Q \bar{y}b(\bar{u}).\,\nabla\varphi\, \mathrm{d}x\mathrm{d}t + \lim_{n\to\infty}\iint_Q y_n b(u_n).\,\nabla\varphi\, \mathrm{d}x\mathrm{d}t \\
&\qquad = -\lim_{n\to\infty}\iint_Q \left(\bar{y}b(\bar{u}) - y_n b(u_n)\right).\nabla\varphi\, \mathrm{d}x\mathrm{d}t \\
&\qquad = -\lim_{n\to\infty}\iint_Q \bar{y}\left(b(\bar{u}) - b(u_n)\right).\nabla\varphi\, \mathrm{d}x\mathrm{d}t - \lim_{n\to\infty}\iint_Q \left(\bar{y} - y_n\right) b(u_n).\,\nabla\varphi\, \mathrm{d}x\mathrm{d}t\,.
\end{aligned}
$$

We observe that $\bar{y} \in L^\infty(0,T;H)$ and $\partial_i\varphi \in L^2(0,T;H)$ for all $i = 1,\dots,d$, thus $\bar{y}\ \partial_i\varphi \in L^2(0,T;L^1(\Omega)) \subset L^{q'}(0,T;L^1(\Omega))$ with q' such that $1/q + 1/q' = 1$ and $L^q(0,T;L^\infty(\Omega)) = \left[L^{q'}(0,T;L^1(\Omega))\right]^*$, since the Lebesgue measure is $\sigma-$finite. Moreover, the expression (12) of b gives that $b(\bar{u}) - b(u_n) = (\bar{u}_i - u_{n,i})_{i=1,\dots,d}$, and $u_n \overset{*}{\rightharpoonup} \bar{u}$ weakly-star in $\mathcal{U}$ ensures that the first integral goes to 0 as $n \to +\infty$. Furthermore, using the fact that the sequence $(b(u_n))_n$ is uniformly bounded and $y_n \to \bar{y}$ strongly in $L^2(0,T;H)$,

$$
\Big|\iint_Q \left(\bar{y} - y_n\right) b(u_n).\,\nabla\varphi\, \mathrm{d}x\mathrm{d}t\Big| \le C\,\|\bar{y} - y_n\|_{L^2(0,T;H)}\,\|\varphi\|_{L^2(0,T;V)} \to 0
$$

as $n \to +\infty$. Additionally, we observe that $\bar{y}(0) = y_0$, hence

$$
\dot{\bar{y}}(t) + A\bar{y}(t) + B(\bar{u}(t),\bar{y}(t)) = 0\,, \quad \bar{y}(0) = y_0\,.
$$

Finally, owing to the weakly-star lower semicontinuity of J, we conclude that

$$
J(\bar{u}) \le \liminf_{n\to\infty} J(u_n) = I.
$$

Thus, $(\bar{y},\bar{u})$ is an optimal pair for the optimal control problem under consideration. $\square$

Requiring box constraints as in (20) might seem a too restrictive choice. However, we note that in case of bilinear action of the control into the system, even box constraints might not suffice to ensure the existence of optimal controls in general, see for example [25, Section 15.3, p. 237].
Theorem 8 clearly also holds for any $\mathcal{U}_{ad}$ that is a bounded weakly-star closed subset of $\mathcal{U}$. However, observe that in the unconstrained case $\mathcal{U}_{ad} \equiv \mathcal{U}$, asking only $J(u) \geq \lambda\|u\|_{\mathcal{U}}$ for some $\lambda > 0$ is not enough. Instead, for the arguments in the proof to hold, one would need $J(u) \geq \lambda\|u\|_{L^\infty(Q;\mathbb{R}^d)}$, which is not very practical. Therefore, we focus on the former case in this paper.

Corollary 9
As a consequence of the previous proof, we have also proved that the control-to-state map $\Theta\colon \mathcal{U}_{ad} \subset \mathcal{U} \to C([0,T];H) \subset L^2(0,T;H)$ *such that* $u \mapsto \Theta(u) = y \in L^2(0,T;H)$ *solution of* $\mathcal{E}(y_0, u, 0)$ *is sequentially continuous from* $\mathcal{U}_{ad}$ *(with the weak-star topology induced by* $\mathcal{U}$*) to* $L^2(0,T;H)$ *(with the strong topology).*

Corollary 10
Assume that $b(x,t;u) = (\gamma_i(x) + u_i(x,t))_i$ *for some* $\gamma_i \in C^1(\Omega)$, $i = 1, \dots, d$, *with* $u \in \mathcal{U}_{ad}$ *as in* (20)*, and let* $y_d \in L^2(0,T;H)$, $y_\Omega \in H$, $\alpha, \beta, \lambda \geq 0$ *with* $\max\{\alpha, \beta\} > 0$. *Then an optimal pair* $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$ *exists for the reduced cost functional*

$$J(u) := \frac{\alpha}{2} \|\Theta(u) - y_d\|^2_{L^2(Q)} + \frac{\beta}{2} \|\Theta(u)(T) - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \|u\|^2_{L^2(Q;\mathbb{R}^d)} . \tag{23}$$

Proof. The cost functional (23) is bounded from below by zero and it is weakly lower semicontinuous in $L^2(Q;\mathbb{R}^d)$. Moreover, a minimizing sequence $(u_n)_{n\geq 1}$ in $\mathcal{U}_{ad}$ converging to I is uniformly bounded both in $\mathcal{U}$ and in $L^2(Q;\mathbb{R}^d)$. Since the weak-star convergence in $\mathcal{U}$ implies the weak convergence in $L^2(Q;\mathbb{R}^d)$, we do not need to require weakly-star lower semicontinuity of J and therefore we can conclude the existence of an optimal pair $(\bar{y}, \bar{u}) \in C([0,T];H) \times \mathcal{U}_{ad}$. □

Remark 11
If one wants to use the cost functional (23) *without imposing box constraints on the control, e.g.,* $\mathcal{U}_{ad} \equiv \mathcal{U}$, *one shall require more regularity on the state* y *and on the control* u, *in order to gain the same level of compactness required in the proof of Theorem 8 to deduce that* $B(\bar{u}, \bar{y}) = \Lambda$. *Indeed, further regularity of* y *can be ensured by standard improved regularity results, see for example [35, Theorems 27.2 and 27.5] and [21, Theorem 6.1 and Remark 6.3]. However, these results come at the price of requiring more regularity of the coefficients in the PDE, which, in our case, translates to more regularity of the control. In particular, one would need to require differentiability of* u *both in time and space.*

Remark 12
Corollary 10 applies analogously to the case of time-independent controls

$$u \in \tilde{\mathcal{U}}_{ad} := \{u \in L^\infty(\Omega;\mathbb{R}^d) : u_a \leq u(x) \leq u_b \quad \textit{for almost every } x \in \Omega\} \tag{24}$$

for some $u_a, u_b \in \mathbb{R}^d$ such that $u_a \leq u_b$ (component-wise) and the reduced cost functional

$$J_2(u) := \frac{\alpha}{2} \|\Theta(u) - y_d\|^2_{L^2(Q)} + \frac{\beta}{2} \|\Theta(u)(T) - y_\Omega\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \|u\|^2_{L^2(\Omega;\mathbb{R}^d)} .$$

6 Adjoint state and optimality conditions

From now on we consider b and B such that $b(u) = u$ and

$$B(u, y) = \operatorname{div}(uy) \quad \forall u \in \mathcal{U}\,,\ y \in L^\infty(0, T; H)\,,$$

respectively. This choice does not affect the generality of the problem. Indeed, for b as in Theorem 8, assuming $\max_i\{\gamma_i, \gamma_i'\}$ sufficiently small, we can include the contribution $\operatorname{div}(\gamma y)$ in the operator A, which becomes

$$A_\gamma z := Az + \operatorname{div}(\gamma z)\,,$$

that still satisfies the assumptions required on A. We start with the following lemma.

Lemma 13
Let $y_0 \in H$ and consider the map $\Theta \colon \mathcal{U} \to C([0, T]; H)$ such that $u \mapsto y \in C([0, T]; H)$ solution of $\mathcal{E}(y_0, u, 0)$. Then Θ is differentiable in the Fréchet sense and, for every $\bar{u}$, $h \in \mathcal{U}$, the function $\Theta'(\bar{u})h$ satisfies

$$\begin{cases} \dot{z}(t) + Az(t) + B(\bar{u}(t), z(t)) = -B(h(t), \bar{y}(t)) \quad \text{in } V'\,,\ t \in (0, T)\,, \\ z(0) = 0\,, \end{cases} \tag{25}$$

where $\bar{y} = \Theta(\bar{u})$.

We observe that, thanks to Remark 2(ii), Theorem 3 ensures the existence of a unique weak solution of equation (25).

Proof. Thanks to the assumptions on b, the map $L \colon \mathcal{U} \to C([0, T]; H)$ such that $h \mapsto z \in C([0, T]; H)$ solution of (25) is linear. Moreover, L is continuous; indeed, the estimate (17) yields

$$\begin{aligned} \|z\|^2_{L^\infty(0,T;H)} &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|B(h, \bar{y})\|^2_{L^2(0,T;V')} \\ &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|\bar{y}\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}} \leq C \|h\|^2_{\mathcal{U}}\,. \end{aligned}$$

Let us now introduce $y_h := \Theta(\bar{u} + h)$ solution of $\mathcal{E}(y_0, \bar{u} + h, 0)$, and set $y := y_h - \bar{y}$. Thus, y satisfies

$$\begin{cases} \dot{y}(t) + Ay(t) + B(\bar{u}(t), y(t)) = -B(h(t), y_h(t)) \quad \text{in } V'\,,\ t \in (0, T)\,, \\ y(0) = 0\,. \end{cases}$$

Moreover, $B(h, y_h) \in L^q(0,T;V') \subset L^2(0,T;V')$, and relation (17) ensures

$$\begin{aligned}\|y\|^2_{L^\infty(0,T;H)} &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|B(h,y_h)\|^2_{L^2(0,T;V')}\\ &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|y_h\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}} ,\end{aligned}$$

with $\|y_h\|^2_{L^\infty(0,T;H)} \leq Ce^{c(1+\|\bar{u}+h\|^2_{\mathcal{U}})} \|y_0\|^2_H$, which is locally bounded in h. Finally, set $w := y - z$. Then w is solution of $\mathcal{E}(0, \bar{u}, -B(h(t), y(t)))$ and satisfies

$$\begin{aligned}\|w\|^2_{L^\infty(0,T;H)} &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|B(h,y)\|^2_{L^2(0,T;V')}\\ &\leq Ce^{c(1+\|\bar{u}\|^2_{\mathcal{U}})} \|y\|^2_{L^\infty(0,T;H)} \|h\|^2_{\mathcal{U}} ,\end{aligned}$$

that is,

$$\|\Theta(\bar{u}+h) - \Theta(\bar{u}) - z\|^2_{L^\infty(0,T;H)} \leq C \|y_0\|^2_H e^{c(1+\|\bar{u}+h\|^2_{\mathcal{U}})} \|h\|^4_{\mathcal{U}} .$$

Therefore, Θ is Fréchet differentiable, $\Theta' \in \mathcal{L}(\mathcal{U}, \mathcal{L}(\mathcal{U}, C([0,T];H)))$ and, for all $\bar{u} \in \mathcal{U}$, $\Theta'(\bar{u}) \colon \mathcal{U} \to C([0,T];H)$ is defined by $\Theta'(\bar{u})h = z$ for all $h \in \mathcal{U}$. □

We introduce the operators $A^* : L^2(0,T;V) \to L^2(0,T;V')$ such that

$$A^* z = -\sum_{i,j=1}^{d} a_{ij} \partial^2_{ij} z \qquad \forall z \in L^2(0,T;V) ,$$

and $\tilde{B} \colon L^2(0,T;V) \to L^2(0,T;L^2(\Omega;\mathbb{R}^d))$ such that $\tilde{B}(v) = \nabla_x v$ for all $v \in L^2(0,T;V)$, where ∇_x denotes the gradient with respect to the space variable $x \in \mathbb{R}^d$. Observe that, for every $v, \varphi \in L^2(0,T;V)$,

$$\int_0^T \langle A^* v(t), \varphi(t)\rangle_{V',V} \, \mathrm{d}t = \int_0^T \langle A\varphi(t), v(t)\rangle_{V',V} \, \mathrm{d}t \tag{26}$$

and for every $u \in \mathcal{U}, v \in L^2(0,T;V)$ and $w \in L^\infty(0,T;H)$,

$$\begin{aligned}\int_0^T (b(u).\, \tilde{B}(v), w)_H \, \mathrm{d}t &= \iint_Q \sum_{i=1}^{d} b_i(u) w \, \partial_i v \, \mathrm{d}x\mathrm{d}t\\ &= -\int_0^T \langle B(u(t), w(t)), v\rangle_{V',V} \, \mathrm{d}t \end{aligned}\tag{27}$$

and the above integrals are well-defined.

In the following, we derive the first order necessary optimality conditions for the cost functional J as in (23). We start by proving an explicit representation formula for the derivative of J, as stated in the following result. Incidentally, let us point out that J is one of the objective functionals most commonly used in the numerical simulations, see, for example, [3, 12].

Proposition 14
Let us consider the functional J of the form (23), with $y_d \in L^q(0,T;L^\infty(\Omega)), y_\Omega \in L^2(\Omega)$, and $y_0 \in L^\infty(\Omega)$. Then J is differentiable in $\mathcal{U}$ and, for all $u, h \in \mathcal{U}$,

$$\mathrm{d}J(u)h = \sum_{i=1}^{d} \iint_Q h_i(t)\,[y(t)\partial_i p(t) + \lambda u_i(t)]\,\mathrm{d}x\mathrm{d}t\,, \tag{28}$$

holds, where $y \in W(0,T)\cap L^\infty(Q)$ is a solution of $\mathcal{E}(y_0,u,0)$ and $p \in W(0,T)$ is the solution of the adjoint equation

$$\begin{cases} -\dot{p}(t) + A^* p(t) - b(u(t)).\,\tilde{B}p(t) = \alpha\,[y(t) - y_d(t)] & \text{in } V'\,,\ t \in (0,T)\,, \\ p(T) = \beta\,[y(T) - y_\Omega]\ . \end{cases} \tag{29}$$

Let us observe that, for all $i = 1, \dots, d$, the function $h_i \langle \partial_i p, y\rangle_{V',V} \colon (0,T) \to \mathbb{R}$ belongs to $L^1(0,T)$, owing to $h_i \in L^q(0,T;L^\infty(\Omega))$ with $q > 2$, $y \in L^2(0,T;V)$ and $\partial_i p \in L^\infty(0,T;V')$. Moreover, the existence and uniqueness of solutions for equation (29) is ensured as in Theorem 3. Indeed, $y_0 \in L^\infty(\Omega)$ implies $y \in L^\infty(Q)$, thus $y - y_d \in L^q(0,T;L^\infty(\Omega))$ as required by Assumption 1, and $y(T) - y_\Omega \in L^2(\Omega)$. By the change of variable $q(t) = p(T-t)$, $v(t) = u(T-t)$ and $f(t) = \alpha[y(T-t) - y_d(T-t)]$, equation (29) is recast in a form similar to equation (10) such that Theorem 3 and Proposition 4 can be applied. In addition, if $y_\Omega \in L^\infty(\Omega)$, then we conclude that $p \in W(0,T)\cap L^\infty(Q)$, see [4, Thm. 1, p. 634].

Proof. Thanks to Lemma 13, the functional J is differentiable in $\mathcal{U}$, and moreover, for all $u, h \in \mathcal{U}$, setting $z = \Theta'(u)h \in C([0,T];H)$ solution of (25), we derive that

$$\mathrm{d}J(u)h = \langle z, \alpha[y - y_d]\rangle_{L^2(0,T;H)} + \langle z(T), \beta[y(T) - y_\Omega]\rangle_H + \lambda\langle h, u\rangle_{L^2(0,T;L^2(\Omega;\mathbb{R}^d))}\,,$$

where y is the solution of the state equation (21). We now exploit the adjoint state p in order to figure out the dependence of $\mathrm{d}J(u)h$ on h. Indeed, owing to relations (27) and (29), we have that

$$\begin{aligned}
&\int_0^T \langle z(t), \alpha[y(t) - y_d(t)]\rangle_H \,\mathrm{d}t \\
&\quad= \int_0^T \langle -\dot{p}(t) + A^* p(t) - b(u(t)).\,\tilde{B}p(t), z(t)\rangle_{V',V}\,\mathrm{d}t \\
&\quad= -\langle z(T), p(T)\rangle_H + \langle z(0), p(0)\rangle_H + \int_0^T \langle \dot{z}(t) + Az(t) + B(u(t), z(t)), p(t)\rangle_{V',V}\,\mathrm{d}t \\
&\quad= -\langle z(T), p(T)\rangle_H - \int_0^T \langle B(h(t), y(t)), p(t)\rangle_{V',V}\,\mathrm{d}t \\
&\quad= -\langle z(T), p(T)\rangle_H + \iint_Q y(t)h(t).\,\nabla p(t)\,\mathrm{d}x\mathrm{d}t\,.
\end{aligned}$$

Since $\langle z(T), \beta[y(T) - y_\Omega]\rangle_H = \langle z(T), p(T)\rangle_H$, we conclude that

$$\begin{aligned} \mathrm{d}J(u)h &= \sum_{i=1}^{d} \iint_Q h_i(t) y(t) \partial_i p(t) \, \mathrm{d}x\mathrm{d}t + \lambda \sum_{i=1}^{d} \langle h_i, u_i\rangle_{L^2(Q)} \\ &= \sum_{i=1}^{d} \iint_Q h_i(t) \left[y(t)\partial_i p(t) + \lambda u_i(t)\right] \mathrm{d}x\mathrm{d}t \,. \end{aligned}$$

□

We observe that, a priori, $\mathrm{d}J(u)$ is defined in $\mathcal{U}$ for every $u \in \mathcal{U}$. However, thanks to the representation formula (28), it admits an extension operator which is well defined on $L^2(0, T; L^2(\Omega; \mathbb{R}^d))$.
As a consequence of Proposition 14 and the variational inequality $dJ(\bar{u})(u - \bar{u}) \geq 0$ for any $u \in \mathcal{U}_{ad}$ and locally optimal solution $\bar{u}$, we deduce the first order necessary optimality conditions, our second main result. The local optimality comes from the control-to-state operator being nonlinear, i.e., the reduced cost functional is non-convex even for standard quadratic costs like (23).

Corollary 15
Let $y_0 \in L^\infty(\Omega)$, $y_d \in L^q(0, T; L^\infty(\Omega))$, and $y_\Omega \in H$. Consider the cost functional J defined by (23) with $\alpha, \beta, \gamma \geq 0$ and $\max\{\alpha, \beta\} > 0$. An optimal pair $(\bar{y}, \bar{u}) \in C([0, T]; H) \times \mathcal{U}_{ad}$ for J with corresponding adjoint state $\bar{p}$ satisfies the following necessary conditions:

$$\begin{aligned} &\partial_t \bar{y} - \sum_{i,j=1}^{d} \partial^2_{ij}(a_{ij}\bar{y}) + \sum_{i=1}^{d} \partial_i\big(\bar{u}_i \bar{y}\big) = 0\,, && \textit{in } Q\,, \\ &-\partial_t \bar{p} - \sum_{i,j=1}^{d} a_{ij}\partial^2_{ij}\bar{p} - \sum_{i=1}^{d} \bar{u}_i \partial_i \bar{p} = \alpha[\bar{y} - y_d]\,, && \textit{in } Q\,, \\ &\bar{y} = \bar{p} = 0 && \textit{on } \Sigma\,, \\ &\bar{y}(0) = y_0\,, \quad \bar{p}(T) = \beta[\bar{y}(T) - y_\Omega]\,, && \textit{in } \Omega\,, \\ &\iint_Q \left[\bar{y}\partial_i \bar{p} + \lambda \bar{u}_i\right](u_i - \bar{u}_i) \, \mathrm{d}x\mathrm{d}t \geq 0 && \forall u \in \mathcal{U}_{ad}\,. \end{aligned} \tag{30}$$

Proof. The necessary optimality conditions (30) are derived by combining equation (21) for the state $\bar{y}$, equation (29) for the adjoint $\bar{p}$ and the variational inequality $\mathrm{d}J(\bar{u})(u - \bar{u}) \geq 0$ for all $u \in \mathcal{U}_{ad}$ for the optimal $\bar{u}$. Thanks to relations (14), (15), and (27), which define the operators A, B, and $\tilde{B}$, respectively, we deduce the desired system. □

In case of time-independent control considered in Remark 12, the only modification needed in the optimality system (30) is the variational inequality, which changes to

$$\int_\Omega \left[\int_0^T \bar{y}\partial_i \bar{p} \, \mathrm{d}t + \lambda \bar{u}_i \right] (u_i - \bar{u}_i) \, \mathrm{d}x \geq 0 \quad \forall u \in \tilde{\mathcal{U}}_{ad}\,,$$

where $\tilde{\mathcal{U}}_{ad}$ is given by (24).

7 Conclusions

In this paper, we considered an optimal control problem subject to the Fokker-Planck equation. In our setting, the control $u(x,t)$ takes action as a coefficient of the state variable, resulting in a control of bilinear type. We proved existence of optimal controls associated with a nonnegative state solution and derived the first order necessary optimality conditions rigorously, thereby extending the results of [1]. Since the control-to-state operator is nonlinear, the reduced cost functional is non-convex even for standard quadratic costs. Therefore, only local optimality can be expected. Furthermore, the uniqueness of the optimal control for a space-dependent controller is still an open question.

Acknowledgements

The authors wish to express their gratitude to Lars Grüne for suggesting them this very interesting subject and for many helpful comments. They would also like to thank Alfio Borzì for very helpful discussions. This work was partially supported by the EU under the 7th Framework Program, Marie Curie Initial Training Network FP7-PEOPLE-2010-ITN SADCO, GA 264735-SADCO, by the DFG project Model Predictive Control for the Fokker-Planck Equation, GR 1569/15-1, and by the INdAM through the GNAMPA Research Project 2015 "Analisi e controllo di equazioni a derivate parziali nonlineari".

References

[1] A. Addou and A. Benbrik. Existence and uniqueness of optimal control for a distributed-parameter bilinear system. *J. Dynam. Control Systems*, 8(2):141–152, 2002.

[2] M. Annunziato and A. Borzì. Optimal control of probability density functions of stochastic processes. *Math. Model. Anal.*, 15(4):393–407, 2010.

[3] M. Annunziato and A. Borzì. A Fokker-Planck control framework for multidimensional stochastic processes. *J. Comput. Appl. Math.*, 237(1):487–507, 2013.

[4] D. G. Aronson. Non-negative solutions of linear parabolic equations. *Ann. Scuola Norm. Sup. Pisa (3)*, 22:607–694, 1968.

[5] D. G. Aronson and James Serrin. Local behavior of solutions of quasilinear parabolic equations. *Arch. Rational Mech. Anal.*, 25:81–122, 1967.

[6] Jean-Pierre Aubin. Un théorème de compacité. *C. R. Acad. Sci. Paris*, 256:5042–5044, 1963.

[7] A. Blaquière. Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version). *Dynam. Control*, 2(3):235–253, 1992.

[8] Roger Brockett. New issues in the mathematics of control. In *Mathematics unlimited—2001 and beyond*, pages 189–219. Springer, Berlin, 2001.

[9] Eduardo Casas. Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.*, 35(4):1297–1327, 1997.

[10] William Feller. Diffusion processes in one dimension. *Trans. Amer. Math. Soc.*, 77:1–31, 1954.

[11] Alessio Figalli. Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.*, 254(1):109–153, 2008.

[12] Arthur Fleig, Lars Grüne, and Roberto Guglielmi. Some results on model predictive control for the Fokker-Planck equation. In *MTNS 2014 : 21st International Symposium on Mathematical Theory of Networks and Systems, July 7-11, 2014*, University of Groningen, The Netherlands, pages 1203–1206, 2014.

[13] Wendell H. Fleming and Raymond W. Rishel. *Deterministic and stochastic optimal control*. Springer-Verlag, Berlin-New York, 1975. Applications of Mathematics, No. 1.

[14] Michael G. Forbes, J. Fraser Forbes, and Martin Guay. Regulating discrete-time stochastic systems: focusing on the probability density function. *Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms*, 11(1-2):81–100, 2004.

[15] Crispin Gardiner. *Stochastic methods*. Springer Series in Synergetics. Springer-Verlag, Berlin, fourth edition, 2009. A handbook for the natural and social sciences.

[16] Ĭ. Ī. Gīhman and A. V. Skorohod. *Stochastic differential equations*. Springer-Verlag, New York-Heidelberg, 1972. Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72.

[17] Werner Horsthemke and René Lefever. *Noise-induced transitions*, volume 15 of *Springer Series in Synergetics*. Springer-Verlag, Berlin, 1984. Theory and applications in physics, chemistry, and biology.

[18] Guy Jumarie. Tracking control of nonlinear stochastic systems by using path cross-entropy and Fokker-Planck equation. *Internat. J. Systems Sci.*, 23(7):1101–1114, 1992.

[19] Miroslav Kárný. Towards fully probabilistic control design. *Automatica J. IFAC*, 32(12):1719–1722, 1996.

[20] A. Kolmogoroff. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 104(1):415–458, 1931.

[21] O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. Ural'tseva. *Linear and quasilinear equations of parabolic type.* Moskva: Izdat. 'Nauka'. 736 pp., 1967.

[22] Irena Lasiecka and Roberto Triggiani. *Control theory for partial differential equations: continuous and approximation theories. I*, volume 74 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2000. Abstract parabolic systems.

[23] C. Le Bris and P.-L. Lions. Existence and uniqueness of solutions to Fokker-Planck type equations with irregular coefficients. *Comm. Partial Differential Equations*, 33(7-9):1272–1317, 2008.

[24] J.-L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires.* Dunod; Gauthier-Villars, Paris, 1969.

[25] J.-L. Lions. *Optimal control of systems governed by partial differential equations.* Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, Band 170. Springer-Verlag, New York-Berlin, 1971.

[26] Alessio Porretta. On the planning problem for the mean field games system. *Dyn. Games Appl.*, 4(2):231–256, 2014.

[27] Alessio Porretta. Weak solutions to Fokker-Planck equations and mean field games. *Arch. Ration. Mech. Anal.*, 216(1):1–62, 2015.

[28] Serguei Primak, Valeri Kontorovich, and Vladimir Lyandres. *Stochastic methods and their applications to communications.* John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[29] Philip E. Protter. *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2005.

[30] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Appl. Math. Optim.*, 39(2):143–177, 1999.

[31] H. Risken. *The Fokker-Planck equation*, volume 18 of *Springer Series in Synergetics.* Springer-Verlag, Berlin, second edition, 1989.

[32] Jacques Simon. Compact sets in the space $L^p(0,T;B)$. *Ann. Mat. Pura Appl. (4)*, 146:65–96, 1987.

[33] Fredi Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI, 2010.

[34] Hong Wang. Robust control of the output probability density functions for multivariable stochastic systems with guaranteed stability. *IEEE Trans. Automat. Control*, 44(11):2103–2107, 1999.

[35] J. Wloka. *Partial differential equations.* Cambridge University Press, Cambridge, 1987.