Applied Mathematics & Optimization (2024) 90:5
https://doi.org/10.1007/s00245-024-10151-4

Pontryagin's Principle for Some Probabilistic Control Problems

Wim van Ackooij[1] · **René Henrion**[2] · **Hasnaa Zidani**[3]

Accepted: 23 May 2024

Abstract
In this paper we investigate optimal control problems perturbed by random events. We assume that the control has to be decided prior to observing the outcome of the perturbed state equations. We investigate the use of probability functions in the objective function or constraints to define optimal or feasible controls. We provide an extension of differentiability results for probability functions in infinite dimensions usable in this context. These results are subsequently combined with the optimal control setting to derive a novel Pontryagin's optimality principle.

Keywords Optimal control problems · Pontryagin maximum principle · Probabilistic cost · Probust control · Chance constraints

Mathematics Subject Classification 49J15 · 93E03 · 90C15

1 Introduction

Many physical or engineering systems are usually described by complex models including inherent uncertainties related to the evolution of the system or to the environment in which this evolution takes place. This is the case for example in finance or

Wim van Ackooij, René Henrion and Hasnaa Zidani have contributed equally to this work.

✉ Hasnaa Zidani
Hasnaa.Zidani@insa-rouen.fr

Wim van Ackooij
wim.van-ackooij@edf.fr

René Henrion
rene.henrion@wias-berlin.de

1 EDF R&D, Saclay, Palaiseau, France

2 Weierstrass Institute, Berlin, Germany

3 LMI, INSA Rouen Normandie, Rouen, France

Published online: 05 June 2024

in energy management where uncertainties about the price of commodities, demand or supply must be taken into account in the mathematical formulation.

Stochastic optimization provides a general and convenient framework for the optimization of uncertain systems. In this context, it is relevant to consider optimal solutions that are risk-averse in the sense of probability with respect to the model's uncertainties. This notion distinguishes from the "worst case" approach by the fact that it aims to define a robust solution against the uncertainties for a reasonable level of probability set by the decision-maker, while the worst-case problem, which aims for a robust solution against *all* uncertainties, may admit no feasible solution (and even when an admissible strategy exists, it is generally too pessimistic).

Optimization under probability constraints, introduced in the 1950 s, is currently featured as a predominant model of stochastic optimization. The traditional framework is set up by finite dimensional decisions and systems of finite random inequalities. We refer to the classic monograph [39] and the more recent presentation [44] for an overview of the theory, algorithms and applications of this model. The traditional framework is already seen as both a theoretical and a numerical challenge. Driven by the need to handle probabilistic constraints in important engineering applications, substantial algorithmic advances have been made over the past two decades, e.g., [8, 17, 35]. These and various other popular approaches rely either on sampling approaches that replace the underlying random vector with a discretized version or on some reformulation of the probability function. A very popular trend is the replacement of the probability function by a substitute, often derived from some approximation of the indicator function inside the expectation, e.g., [25, 29]. Recent investigations along such lines consist in replacing the probability function with it's inverse: quantile approximations, e.g., [37]. Traditionally however resolution methods have relied on the observation that probability functions are a special kind of nonlinear mapping. In principle therefore classic nonlinear programming solvers would be appropriate. This line of investigation has also shown great potential, e.g., [14]. Although a thorough investigation of various alternatives - trying to investigate pros and cons honestly - has of yet still to be carried out, it would seem that classic approaches are not only competitive, but offer advantages in obtaining feasible solutions. Now in order to put the classic approaches to work, first-order information of probability functions is usually required. This was recognized a long time ago and different strategies pursued: a generalistic one, of which, e.g., [45] offers a description and a more practical one starting from specific structures, such as those arising in concrete engineering applications, e.g., [27]. Some examples of the latter investigations are [2, 3, 6] and we refer the reader to [1] for a recent survey and overview.

The infinite-dimensional setting poses a lot of new theoretical questions on the structure of probabilistic constraints. Some fundamental structural analysis (weak semi-continuity, convexity, existence of solutions, stability of solutions with respect to perturbations of the probability measure) have been carried out for infinitely many probabilistic constraints in a Banach space and applied to PDE-constrained optimization problems [19, 24]. By using generalized differential calculus, sub-differential or differential formulas have been derived in the case of a single Lipschitzian random inequality with infinite-dimensional decisions and in the case of a finite-dimensional setting with infinite random inequality systems [8, 28].

Recently, the interest in applying probabilistic constraints (or Value-at-Risk models) in the context of optimal control setting has increased considerably (for example, [10, 15, 18, 20–22, 26, 30, 38, 43]). Optimal control problems consist in analysing the evolution of complex systems which, under the effect of a control input, can give the best performance while respecting the constraints of the system. This class of problems arise in many technological fields (aeronautics, mobile robotics, power management, gas transport, ...).

The present work is devoted to a class of control problems with uncertainties. Let $\mathbb{H}$ and $\mathbb{U}$ be two Hilbert spaces, let T be a given finite final time, and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where Ω represents the sample space, $\mathcal{F}$ is the σ-algebra of events, and $\mathbb{P} : \Omega \to [0, 1]$ is a probability measure. Consider an ensemble of controlled state equations parametrized by the random event ω on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

$$\begin{cases} \dot{\mathbf{x}}_\omega(t) = A\mathbf{x}_\omega(t) + B(\omega)\boldsymbol{u}(t) + E(\omega) & \text{for a.e. } t \in [0, T], \\ \mathbf{x}_\omega(0) = x_0, & \\ u(t) \in U, & \text{for a.e. } t \in [0, T], \end{cases} \tag{1}$$

where U is a closed metric space, x_0 is an initial data, A, and B are linear operators (the assumptions on these operators will be made precise later). An admissible control input $u : [0, T] \to \mathbb{U}$ is a measurable function assumed to be ω-independent, which means that the states of the ensemble (the parametrized family) are driven by the same control. The optimal control problems consists of an optimization problem governed by the uncertain system (1) with a cost function composed of an expectation term and a probability of success for the terminal state. The latter has the following form

$$\mathbb{P}\big(\Psi(\mathbf{x}_\omega(T)) \leq 0\big), \tag{2}$$

where $\Psi : \mathbb{H} \to \mathbb{R}$ is a given function. This cost function evaluates the probability that the ensemble of controlled states verify a constraint at the final time.

The control problem may alternatively include such a probability of success as a "target constraint" at the final time.

In the state equation (1), the uncertainty appears in both the control operator B and the source term E. While a broader scope, including uncertainty in the operator A and nonlinear dependencies on control and/or state in the source term, would span a wider range of applications, the sensitivity analysis for optimization problems with probabilistic functions relies heavily on the nature of dependency with respect to uncertainty, even in finite-dimensional optimization scenarios. In this study, we adopt a Hilbertian framework capable of accommodating both finite and infinite-dimensional settings. We focus on a linear case already encompassing a diverse set of compelling problems [18]. We will use a *convexity* structure pertaining to uncertainty to derive explicit optimality conditions of the control problem. Further investigation of more general cases is slated for future research works.

The problems of designing a single ω-independent control strategy u for controlling an ensemble of nonlinear systems arise in many real applications. Questions of controllability (i.e., steering the ensemble systems from an initial configuration to a prescribed final state) have been addressed in [9], where a criterion for controllability has been derived for a large class of control-linear systems in finite-dimensional space $\mathbb{H} = \mathbb{R}^N$ (Ensemble version of Rashevsky-Chow theorem).

The ensemble controllability has been also considered in the context of quantum systems (see for instance [11, 13, 42]).

The aim of this work is to obtain strong optimality conditions for the control problem with a cost functional (and with a final constraint) in probabilistic form. In the deterministic setting, optimality conditions are usually derived in the form of a so-called Pontryagin's principle, [16, 46]. This principle states that any optimal control and its corresponding state, satisfies a Hamiltonian system, which is a two-point boundary value problem, plus a maximum condition of the control Hamiltonian. These necessary conditions become sufficient under certain convexity conditions on the objective and constraint functions. In presence of uncertainties, KKT-type optimality conditions have been studied in [21, 23] for control problems with uncertainties and almost-sure constraints. Here, we consider a different setting where the risk is defined in terms of probability of success. The proof of Pontryagin principle relies on differentiability properties of the probability functions. As mentioned earlier, differentiability of probability functions has received quite some attention. Results relevant for the structures appearing in this work are [2, 3, 7, 19, 28]. However, in this work we are dealing with an infinite dimensional "decision vector space", unlike most of the previous references where the decision vector w.r.t. which the derivative was to be computed was simply $\Re^n$. The work [28] does present "first order" results in infinite dimensions, but of a more abstract kind and not immediately applicable to our setting. For this reason, in section 3 we have undertaken the task of laying down the various pieces in a consistently and clearly presented framework. This effectively extends the previous investigations to the infinite dimensional setting, all while providing easy to verify conditions ensuring the applicability of the results.

The paper is organized as follows. The control problem is described in Sect. 2 where some concrete examples are presented. Section 3 is devoted to the analysis of differential calculus of probability functions. In Sect. 4, we study optimality conditions for some control problems governed by ordinary differential equations or by Partial differential equations. Finally, we discuss a simple numerical example.

2 Formulation of the Problem

Let us start by establishing some notations that will be used in this paper. For any measure space $(S, \Sigma; \mu)$, a Banach space $(X, \|\cdot\|_X)$, and $r \in [1, +\infty]$, the Bochner space $L^r(S; X)$ consists of all measurable functions $y : \mathcal{S} \longrightarrow X$ whose norm

$$\|y\|_{L^r(\mathcal{S};X)} := \begin{cases} \left(\int_{\mathcal{O}} \|y(s)\|_X^r \, d\mu(s) \right)^{\frac{1}{r}} & \text{if } r < \infty, \\ \text{ess sup}_{s \in \mathcal{S}} \|y(s)\|_X & \text{if } r = \infty \end{cases}$$

is finite, and where functions which agree μ-almost everywhere are identified.

Throughout the paper, we assume that we are given a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For any X-valued random variable $y : \Omega \longrightarrow X$ that is Bochner integrable, the expectation of y, denoted by $\mathbb{E}[y]$, is defined by

$$\mathbb{E}[y] = \int_\Omega y(\omega) d\mathbb{P}(\omega).$$

It is worth noting that $\mathbb{E}[y]$ defines an element of X, whenever $y : \Omega \longrightarrow X$ is Bochner integrable.

A real function $L : X \times \Omega \longrightarrow \mathbb{R}$ is said to be Carathéodory if $L(x, \cdot)$ is measurable with respect to (w.r.t.) the variable ω for every $x \in X$ and if $L(\cdot, \omega)$ is continuous w.r.t. x for every $\omega \in \Omega$. For any measurable function, $y : \Omega \longrightarrow X$, and any Carathéodory function $L : X \times \Omega \longrightarrow \mathbb{R}$, the composition $\omega \longmapsto L(y(\omega), \omega)$ is also a measurable function.

Let $T > 0$ be a fixed final time, we shall denote by $C([0, T]; X)$ the Banach space that consists of all continuous functions $y : [0, T] \longrightarrow X$. This space is endowed with the usual norm

$$\|y\|_{C([0,T];X)} := \max_{t\in[0,T]} \|y(t)\|_X.$$

Finally, for any Banach spaces X and Y, we shall denote $L(X, Y)$ the space of linear continuous operators from X into Y. This space will be simply denoted by $L(X)$ when $Y = X$. The dual of X will be denoted X'.

Let $(\mathbb{H}, \langle\cdot, \cdot\rangle_{\mathbb{H}})$ and $(\mathbb{U}, \langle\cdot, \cdot\rangle_{\mathbb{U}})$ be two Hilbert spaces, which are accordingly identified with their duals. Let U be a closed convex subset of $\mathbb{U}$ and consider the set of square integrable functions $\boldsymbol{u} : [0, T] \longrightarrow \mathbb{U}$ satisfying $\boldsymbol{u}(s) \in U$ for almost every (a.e.) $s \in [0, T]$

$$\mathcal{U} := \{\boldsymbol{u} \in L^2([0, T]; \mathbb{U}) \text{ and } \boldsymbol{u}(s) \in U \text{ a.e. on } [0, T]\}.$$

In the sequel $\mathcal{U}$ will be referred to as the set of admissible control functions.

For every parameter $z \in \mathbb{R}^n$, we consider the differential equation

$$\dot{\mathbf{X}}_z(t) = A\mathbf{X}_z(t) + B(z)\boldsymbol{u}(t) + E(t, z), \quad \mathbf{X}_z(0) = x_0, \tag{3}$$

where the control input u belongs to $\mathcal{U}$ and $x_0 \in \mathbb{H}$. Here, $A : \mathcal{D}(A) \subset \mathbb{H} \longrightarrow \mathbb{H}$ is a linear (unbounded) operator generating a strongly continuous analytic semi-group, denoted e^{At}, on the Hilbert space $\mathbb{H}$. The mapping $B(z) : \mathbb{U} \longrightarrow [\mathcal{D}(A)]'$ is a linear operator, and $E : [0, T] \times \mathbb{R}^n \longmapsto \mathbb{H}$ is a given source term.

In the sequel, we will assume that $\mathbb{H}$ and $\mathbb{U}$ are either of finite or infinite dimension. In both cases, the operator A and B and the source term E are supposed to satisfy some standing assumptions (that will be made precise later) in order to guarantee, for every $\boldsymbol{u} \in L^2([0, T]; \mathbb{U})$ and for every $z \in \mathbb{R}^n$, the existence and well-posedness of a solution $\mathbf{X}_z^{\boldsymbol{u}} \in C([0, T]; \mathbb{H})$ to the state equation (3). This solution is considered in the *mild sense,* meaning that $\mathbf{X}_z^{\boldsymbol{u}}(0) = x_0$ and, for $t \in [0, T]$,

$$\mathbf{X}_z^{\boldsymbol{u}}(t) = e^{At}x_0 + \int_0^t e^{A(t-s)}B(z)\boldsymbol{u}(s)\,ds + \int_0^t e^{A(t-s)}E(s,z)\,ds. \tag{4}$$

We introduce the mapping $\mathcal{G} : L^2([0,T];\mathbb{U}) \times \mathbb{R}^n \longrightarrow C([0,T];\mathbb{H})$ defined, for $\boldsymbol{u} \in z \in \mathbb{R}^n$, by

$$\mathcal{G}(\boldsymbol{u}, z) := e^{A\cdot}x_0 + \int_0^{\cdot} e^{A(\cdot-s)}B(z)\boldsymbol{u}(s)\,ds + \int_0^{\cdot} e^{A(\cdot-s)}E(s,z)\,ds.$$

With this notation, the solution $\mathbf{X}_z^{\boldsymbol{u}}$ is given by $\mathbf{X}_z^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u}, z)$.

Now, Let ξ be a given n-dimensional random variable in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and consider the controlled system subject to uncertainties:

$$\dot{\mathbf{x}}_\omega(t) = A\mathbf{x}_\omega(t) + B(\xi(\omega))\boldsymbol{u}(t) + E(t, \xi(\omega)), \quad \mathbf{x}_\omega(0) = x_0. \tag{6}$$

The state equation associated with a control $\boldsymbol{u} \in L^2([0,T];\mathbb{H})$ is then given by $\mathbf{x}_\omega^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u}, \xi(\omega))$, for every $\omega \in \Omega$. We will also use the notation $\mathbf{x}^{\boldsymbol{u}}$ for the bundle of trajectories $\{\mathbf{x}_\omega^{\boldsymbol{u}} \mid \omega \in \Omega\}$.

In the sequel, we will assume that the uncertainty enters in the operator B and the source term E in a *structured* manner. In particular, we assume that B and E are affine with respect to the variable z

$$B(z) = B_0 + B_1(z) \qquad \text{and } E(t,z) = E_0(t) + E_1(t,z), \tag{7}$$

where $E_1(t,\cdot)$ and $B_1(\cdot)$ are linear maps. Hereafter, we outline the assumptions that will be considered throughout the paper concerning the mappings B and E as well as the operator A.

$(\mathbf{H}_S)$

(a) If $\mathbb{H} = \mathbb{R}^d$ and $\mathbb{U} = \mathbb{R}^r$ are finite dimensional spaces : A is a $d \times d$-matrix. The mapping B_0 is a real $d \times r$ matrix. Moreover, the application $z \longmapsto B_1(z)$ is linear from $\mathbb{R}^n$ into $\mathbb{R}^{d\times r}$.

(b) If $\mathbb{H}$ and $\mathbb{U}$ are infinite dimensional spaces. We assume that 0 belongs to the resolvent of A. Then the fractional powers $(-A)^\gamma$, $0 < \gamma < 1$, are well defined, and we have $\|(-A)^\gamma e^{At}\|_{\mathcal{L}(\mathbb{H})} \le C_\gamma t^{-\gamma}$ for $t > 0$ (see [36, p. 74]), where $C_\gamma > 0$ denotes a positive constant. For every $z \in \mathbb{R}^n$, the mapping $B(z) : \mathbb{U} \to [\mathcal{D}(A)]'$ is a linear continuous operator,[1] such that $A^{-\bar{\alpha}}B(z) \in \mathcal{L}(\mathbb{U},\mathbb{H})$ for some $\bar{\alpha} \in [0, \frac{1}{2}[$:

$$\|A^{-\bar{\alpha}}B(z)\|_{\mathcal{L}(\mathbb{U},\mathbb{H})} = \|B^*(z)(A^*)^{-\bar{\alpha}}\|_{\mathcal{L}(\mathbb{H},\mathbb{U})} \le c_{\bar{\alpha}}(1 + \|z\|), \tag{8}$$

for some $c_{\bar{\alpha}} > 0$, and where $\langle B(z)u, y\rangle_{\mathbb{H}} = \langle u, B^*(z)y\rangle_{\mathbb{U}}$ (B^* being the $\mathbb{H}$-adjoint operator). Moreover, the mapping $z \longmapsto B_1(z)$ is linear and continuous on $\mathbb{R}^n$.

[1] $B(z)$ may be unbounded as an operator from $\mathbb{U}$ to $\mathbb{H}$

We assume also that both components B_0 and $B_1(z)$ also independently verify the same estimate (8).

(c) The function $E : [0, T] \times \mathbb{R}^n \to \mathbb{H}$ is continuous. For every $t \in [0, T]$, the mapping $E_1(t, \cdot)$ is a linear continuous operator from $\mathbb{R}^n$ into $\mathbb{H}$.

Let us emphasize that by requiring, in $(\mathbf{H}_S)$, that $\bar{\alpha}$ belongs to $[0, \frac{1}{2}[$, we ensure that for every $z \in \mathbb{R}^n$ and for every control input $\boldsymbol{u} \in \mathcal{U}$, equation (3) admits a unique mild solution $\mathcal{G}(\boldsymbol{u}, z) \in L^2([0, T]; \mathbb{H}) \cap C([0, T]; H)$ (see [33, Chapter 3]).

Now, consider a terminal cost function $\Psi : \mathbb{H} \to \mathbb{R}$ that satisfies the following assumption.

$(\mathbf{H}_\Psi)$ Ψ is convex and continuously Fréchet differentiable on $\mathbb{H}$ into $\mathbb{R}$

There exists $C_\Psi > 0$ such that

$$|\Psi(x)| \le C_\Psi(1 + \|x\|_{\mathbb{H}}^m) \quad \text{and} \quad \|\nabla\Psi(x)\| \le C_\Psi(1 + \|x\|_{\mathbb{H}}^{m-1}) \quad \forall x \in \mathbb{H},$$

for some $m \ge 1$.

Consider also a distributed cost function $\ell : [0, T] \times \mathbb{H} \times \mathbb{U} \times \Omega \to \mathbb{R}$ that is a Carathéodory function satisfying the following requirement

$(\mathbf{H}_\ell)$ For every $(t, x, u) \in [0, T] \times \mathbb{H} \times \mathbb{U}$, the function $\ell(t, x, u, \cdot)$ is measurable on Ω.

For every $\omega \in \Omega$, $\ell(\cdot, \cdot, \cdot, \omega)$ is continuous on $[0, T] \times \mathbb{H} \times \mathbb{U}$.

For every $(t, \omega) \in [0, T] \times \Omega$, $\ell(t, \cdot, \cdot, \omega)$ is continuously Fréchet differentiable on $\mathbb{H} \times \mathbb{U}$.

Moreover, there exists $C_\ell : \Omega \to \mathbb{R}_+$ such that

$$|\ell(t, x, u, \omega)| \le C_\ell(\omega)(1 + \|x\|_{\mathbb{H}}^m)(1 + \|u\|_{\mathbb{U}}^2),$$
$$\|\ell'_x(t, x, u, z)\|_{\mathbb{H}} + \|\ell'_u(t, x, u, \omega)\|_{\mathbb{U}}$$
$$\le C_\ell(\omega)(1 + \|x\|_{\mathbb{H}}^{m-1})(1 + \|u\|_{\mathbb{U}}),$$

where the real function $C_\ell : \Omega \to \mathbb{R}_+$ is Bochner integrable (i.e., $\mathbb{E}[C_\ell] < \infty$).

The control problem aims at determining a control law $\boldsymbol{u}(\cdot) \in \mathcal{U}$ that optimizes a certain cost function. This cost function encompasses a distributed cost over the time interval $[0, T]$ as well as the probability of a specific event occurring at the final time T and is defined defined over a set of trajectories parametrized by the elementary events $\omega \in \Omega$.

$$\text{Maximize } \left\{ \mathbb{E}\Big[\int_0^T \ell(t, \mathbf{x}^{\boldsymbol{u}}(t), \boldsymbol{u}(t), \cdot)\, dt \Big] + \mathbb{P}\big[\Psi(\mathbf{x}^{\boldsymbol{u}}(T)) \le 0\big],\ \boldsymbol{u} \in \mathcal{U} \right\}. \tag{9}$$

In the sequel problem (9) will be referred to as $(\mathcal{P}_0)$. By the control re-parametrization, the control problem can be also formulated as

$$\text{Maximize } \left\{ \mathbb{E}\Big[\int_0^T \ell(t, \mathcal{G}(\boldsymbol{u}, \xi)(t), \boldsymbol{u}(t), \cdot)\, dt \Big] + \mathbb{P}\big[\Psi(\mathcal{G}(\boldsymbol{u}, \xi)(T)) \le 0\big],\ \boldsymbol{u} \in \mathcal{U} \right\}.$$

The form of the cost function provides a trade-off between an average cost and a probability cost. A simple example for the function ℓ corresponds to the case where $\ell = 0$. In this case, the cost consists solely of the probability term. A second simple example corresponds to $\ell(t, x, u, \omega) = \frac{\delta}{2}\|u\|^2$. In this case, the control problem is simply

$$\text{Maximize } \left\{ \frac{\delta}{2} \int_0^T \|\boldsymbol{u}(t)\|^2 \, dt + \mathbb{P}\big[\Psi(\mathcal{G}(\boldsymbol{u}, \xi)(T)) \leq 0\big], \ \boldsymbol{u} \in \mathcal{U} \right\},$$

where $\delta > 0$ is a scaling parameter between the probability of the final event and the cost associated with the control law $\boldsymbol{u}$.

The main focus in this paper is to derive the optimality conditions of (9) in the form of Pontryagin's principle. As we will see the main difficulty in this problem comes from the probabilistic term, and more precisely from the differentiability of this term with respect to the state and control variables. The differentiability tools that will be developed in this paper also allow us to consider problems with a probability constraint

$$\text{Maximize } \left\{ \mathbb{E}\Big[\int_0^T \ell(t, \mathcal{G}(\boldsymbol{u}, \xi)(t), \boldsymbol{u}(t), \cdot) \, dt \Big], \ \boldsymbol{u} \in \mathcal{U} \text{ and} \right.$$
$$\left. \mathbb{P}\big[\Psi(\mathcal{G}(\boldsymbol{u}, \xi(T)) \leq 0\big] \geq p_0 \right\}, \tag{10}$$

where the level of success p_0 is a given number in [0, 1]. This chance-constrained problem will be referred to as $(\mathcal{P}_1)$. A discussion on the optimality conditions for this problem will also be given in this paper.

Before, we start the discussion about the optimality conditions, let us mention that the setting described in this section includes finite-dimensional control systems governed by linear equations (in this case, $\mathbb{H} = \mathbb{R}^d$ and $\mathbb{U} = \mathbb{R}^r$ for some $d, r \geq 1$, A is a $d \times d$ matrix and B is a $d \times r$ matrix). Additionally, our framework accommodates some systems governed by partial differential equations (PDEs). Here are some examples of PDEs where assumption $(\mathbf{H}_S)$ holds.

Example 2.1 (*Heat equation with Neumann Boundary control*) Let $\mathcal{O}$ be an open bounded subset of $\mathbb{R}^N$ (for $N \geq 1$) of class $C^{2,\beta}$, for some $\beta > 0$ (that is, the boundary $\partial\mathcal{O}$ of $\mathcal{O}$ is an $(N-1)$-dimensional manifold of class $C^{2,\beta}$ such that $\mathcal{O}$ lies on one side of $\partial\mathcal{O}$). Consider the parabolic system where the control input acts in the Neumann boundary condition:

$$\begin{cases} \partial_t \mathbf{x}(t) = \Delta \mathbf{x}(t) + a\mathbf{x}(t) + E(t, z) & \text{in } (0, T) \times \mathcal{O}, \\ \frac{\partial \mathbf{x}(t)}{\partial \nu}\Big|_{\partial\mathcal{O}} = b(z)\boldsymbol{u}(t) & \text{in } (0, T) \times \partial\mathcal{O}, \\ \mathbf{x}(0) = x_0 & \text{in } \mathcal{O}, \end{cases} \tag{11}$$

where $a > 0$ is a constant, b is an affine function defined on $\mathbb{R}^n$ by $b(z) = b_0 + \sum_{i=1}^{n} b_i z_i$, with $b_0, \cdots b_n$ are given constants. The function $E : [0, T] \times \mathbb{R} \to L^2(\mathcal{O})$ is defined as $E(t, z) = E_0(t) + \sum_i E_i(t) z_i$ with $E_0, \cdots, E_n$ are given functions in $L^2(\mathcal{O})$. The initial condition is supposed to be a given function $x_0 \in L^2(\mathcal{O})$. We assume that a is not an eigenvalue of the Laplacian operator.

To put this example in the abstract setting, we introduce

$$Ay = \Delta y + ay, \quad \mathcal{D}(A) = \left\{ y \in H^2(\mathcal{O}),\ \frac{\partial y}{\partial \nu}\Big|_{\partial \mathcal{O}} = 0 \right\}.$$

We select the spaces and operator B as follows: $\mathbb{H} = L^2(\mathcal{O})$, $\mathbb{U} = L^2(\partial\mathcal{O})$ and B is defined by

$$B(z) : \mathbb{U} \to [\mathcal{D}(A)]', \quad B(z)\boldsymbol{u} = -b(z)AN\boldsymbol{u}, \tag{12}$$

where N is the Neumann operator defined as

$$Nf = g \longrightarrow \Delta g + ag = 0 \text{ in } \mathcal{O},\ \frac{\partial g}{\partial \nu}\Big|_{\partial \mathcal{O}} = f.$$

In the light of elliptic equations theory (see [34, p. 187]), the linear operator N is well defined and is continuous as

$$N : \mathbb{U} \to H^{\frac{3}{2}}(\mathcal{O}) \quad \text{and more generally } N : H^s(\partial\mathcal{O}) \to H^{s+\frac{3}{2}}(\mathcal{O}),\ s \in \mathbb{R}.$$

From [32, p. 196] and [33, p. 364], B satisfies assumption $(\mathbf{H}_S)$(a) with $\alpha = \frac{1}{4} + \varepsilon$ for $\varepsilon > 0$.

Example 2.2 (*A heat equation with pointwise control - dimension 1*) Consider now the case when the control input is concentrated at fixed space points $s_1, \cdots, s_n \in \mathcal{O}$, where $\mathcal{O}$ is an open interval of $\mathbb{R}$. Let Here $\delta(\cdot - s_i)$ be the Dirac δ-function concentrated at s_i. The controlled system is

$$\begin{cases} \partial_t \mathbf{x}(t) = \Delta \mathbf{x}(t) + a\mathbf{x}(t) + \sum_{i=1}^n z_i u_i(t)\delta(\cdot - s_i) & \text{in } (0, T) \times \mathcal{O}, \\ \mathbf{x}(t)|_{\partial\mathcal{O}} = 0 & \text{in } (0, T) \times \partial\mathcal{O}, \\ \mathbf{x}(0) = x_0 & \text{in } \mathcal{O}. \end{cases} \tag{13}$$

In this new setting, we select the spaces $\mathbb{H} = L^2(\mathcal{O})$ and $\mathbb{U} = \mathbb{R}^n$. The operator $A : \mathcal{D}(A) \to \mathbb{H}$ is defined as

$$Ay = \Delta y + ay, \quad \mathcal{D}(A) = H^2(\mathcal{O}) \cap H_0^1(\mathcal{O}).$$

The constant $a > 0$ is assumed not to be an eigenvalue of the Laplacian. We define the operator $B(z) = \sum_i^n z_i \delta(\cdot - s_i)$. From [33, p. 365], the operator B satisfies assumption $(\mathbf{H}_S)$(a) with $\alpha = \frac{1}{4} + \varepsilon$ for $\varepsilon > 0$.

In this example and the previous one, the Laplacian operator can be replaced by a more general second-order uniformly elliptic operator with smooth variable coefficients as in [33, 40, 41].

3 Values and Derivatives of Probability Functions

In this section, we shall derive a representation as a spheric integral of probability functions and their derivative associated with a single random inequality. More precisely, we introduce the probability function

$$\varphi(x) := \mathbb{P}[g(x, \xi) \le 0], \tag{14}$$

where $g : X \times \mathbb{R}^n \to \mathbb{R}$, X is a Banach space and ξ is an n-dimensional random vector, i.e., $\xi : \Omega \to \Re^n$ a measurable function. In this work we will focus on the situation wherein ξ will be elliptically symmetric, which is a large class, also offering a concise presentation. Extensions to more general, almost arbitrary laws could likely be carried out along the lines of the investigation in [?].

3.1 Elliptical distributions

We recall the definition of an *elliptical distribution*:

Definition 3.1 An n-dimensional random distribution is called elliptical, if it admits a density of the form

$$f(z) = c \cdot \mathrm{k}\left((z-\mu)^T \Sigma^{-1} (z-\mu)\right) \quad \left(z \in \mathbb{R}^n\right),$$

where c is some normalizing constant, $\mu \in \mathbb{R}^n$, Σ is a $n \times n$ positive definite matrix and $\mathrm{k} : \mathbb{R}_+ \to \mathbb{R}_+$ is the generator, i.e., $\int_0^\infty r^{\frac{n}{2}} \mathrm{k}(r) dt < \infty$. We shall write $\mathcal{E}(\mu, \Sigma, \mathrm{k})$ for an elliptical distribution with parameters μ and Σ.

We note that μ is the expectation of the distribution (if it exists) and Σ is proportional to the covariance matrix of the distribution (if it exists). The class of elliptical distributions includes many prominent multivariate distributions such as Gaussian, Student or t-, symmetric Laplace or logistics distributions. Moreover, the following presentation of values and gradients for the probability function (14) can be extended to related non-elliptical distributions like log-normal, truncated Gaussian or Gaussian mixture upon performing a corresponding transformation of the inequality g. For example, a multivariate Gaussian distribution with expectation μ and covariance matrix Σ has an elliptical distribution $\mathcal{E}(\mu, \Sigma, \mathrm{k})$ with generator $\mathrm{k}(t) := e^{-t/2}$. We shall then use the common notation $\mathcal{N}(\mu, \Sigma)$ rather than $\mathcal{E}(\mu, \Sigma, \mathrm{k})$. It is well known (e.g., [4, eq. (11) and (13)]) that the probability of an elliptical random vector $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ to take values in a Lebesgue-measurable set M can be represented as the spherical integral

$$\mathbb{P}(\xi \in M) = \int_{w \in \mathbb{S}^{n-1}} \tilde{\nu}(\{r \ge 0 | \mu + rLw \in M\}) \, d\nu_{\mathrm{U}}(w),$$

where L is a Cholesky factor of Σ (i.e., $\Sigma = LT$), ν_{U} is the uniform distribution on the unit sphere $\mathbb{S}^{n-1}$ of $\mathbb{R}^n$ and $\tilde{\nu}$ is a one-dimensional probability distribution with density

$$\tilde{f}(t) := \tilde{c} \cdot t^{n-1} \mathrm{k}\left(t^2\right) \quad (t \geq 0), \tag{15}$$

with another normalizing constant $\tilde{c}$. For instance, if $\xi \sim \mathcal{N}(\mu, \Sigma)$, then

$$\tilde{f}(t) = \tilde{c} \cdot t^{n-1} e^{-t^2/2} \quad (t \geq 0),$$

which is the density of the Chi-distribution with n degrees of freedom. Applying the spherical representation above to (14), where g is supposed to be Lebesgue measurable in its second argument, we infer that

$$\varphi(x) = \int\limits_{w \in \mathbb{S}^{n-1}} \tilde{\nu}\left(\{r \geq 0 | g(x, \mu + rLw) \leq 0\}\right) d\nu_{\mathrm{U}}(w). \tag{16}$$

3.2 Continuity of the Probability Function and Representation as a Spherical Integral

As an application of (16), we obtain the following result on the probability function:

Proposition 3.1 *In (14), let g be a continuous function which in addition is convex in its second variable. Let ξ have an elliptic distribution according to $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$. Assume that $g(\bar{x}, \mu) < 0$ for some given $\bar{x} \in X$. Then, the probability function in (14) is continuous at $\bar{x}$ and has the representation*

$$\varphi(\bar{x}) = \int\limits_{w \in \mathbb{S}^{n-1}} e(\bar{x}, w)\, d\nu_{\mathrm{U}}(w), \tag{17}$$

where the radial probability function $e : X \times \mathbb{R}^n \to \mathbb{R}$ is defined as

$$e(x, w) := \begin{cases} 1 & if \quad g(x, \mu + rLw) < 0 \ \forall r \geq 0 \\ \tilde{F}(\rho(x, w)) & else \end{cases}$$

via the cumulative distribution

$$\tilde{F}(t) = \int_{-\infty}^{t} \tilde{f}(s)\, ds, \tag{18}$$

associated with the density $\tilde{f}$ from (15) and the radius function $\rho : X \times \mathbb{R}^n \to \bar{\mathbb{R}}$ defined by

$$\rho(x, w) := \sup_{r \geq 0} \{r \geq 0 | g(x, \mu + rLw) \leq 0\}.$$

Proof Whenever $(x, w) \in X \times \mathbb{R}^n$ are such that $g(x, \mu) < 0$ and $\rho(x, w) < \infty$, then, by convexity in the second variable of the continuous function g, $\rho(x, w)$ is the unique solution in $r \geq 0$ of the equation $g(x, \mu + rLw) = 0$ and it holds that

$$\{r \geq 0 | g(x, \mu + rLw) \leq 0\} = [0, \rho(x, w)].$$

On the other hand, if $g(x, \mu) < 0$ and $\rho(x, w) = \infty$, then

$$\{r \geq 0 | g(x, \mu + rLw) \leq 0\} = \mathbb{R}_+.$$

Since in (15), the density $\tilde{f}$ of the probability measure $\tilde{\nu}$ is nonzero only on $\mathbb{R}_+$, it follows that

$$\tilde{\nu}(\{r \geq 0 | g(x, \mu + rLw) \leq 0\}) = \begin{cases} \tilde{F}(\rho(x, w)) & if\, \rho(x, w) < \infty \\ 1 & \text{if } \rho(x, w) = \infty \end{cases}.$$

Combining this with (16), we arrive at the representation (17).

It has been shown in [2, Corollary 3.4] that the radial probability function e is continuous at all $(x, w) \in X \times \mathbb{S}^{n-1}$ with $g(x, \mu) < 0$. While this result and the auxiliary result it relies on ([2, Lemma 3.3]), were formulated in a setting where X is finite-dimensional and ξ has a Gaussian distribution, their proofs do nowhere exploit these properties and, hence, remain valid in our framework of X being a Banach space and ξ having a general elliptical distribution. Now, in order to verify the continuity of φ at $\bar{x}$, consider a sequence $x_k \to \bar{x}$. By continuity of g, we have that $g(x_k, \mu) < 0$ for k sufficiently large. Therefore, the assumption of this lemma is satisfied at x_k as well and, hence, the representation (17) holds true with $\bar{x}$ replaced by x_k as well. From the stated continuity of e, it follows that $e(x_k, w) \to e(\bar{x}, w)$ for all $w \in \mathbb{S}^{n-1}$. Moreover $e(x_k, w) \leq 1$ for all k and $w \in \mathbb{S}^{n-1}$. Since the dominating function 1 is integrable with respect to the uniform measure ν_{U} on the sphere, Lebesgue's dominated convergence theorem yields the asserted continuity of φ:

$$\begin{aligned}\lim_{x_k \to \bar{x}} \varphi(x_k) &= \lim_{x_k \to \bar{x}} \int\limits_{w \in \mathbb{S}^{n-1}} e(x_k, w)\, d\nu_{\mathrm{U}}(w) \\ &= \int\limits_{w \in \mathbb{S}^{n-1}} \left(\lim_{x_k \to \bar{x}} e(x_k, w) \right) d\nu_{\mathrm{U}}(w) = \varphi(\bar{x}).\end{aligned}$$

□

For a numerical approximation of $\varphi(\bar{x})$, one would replace the spherical integral in (17) by a finite sum

$$\varphi(\bar{x}) = K^{-1} \sum_{k=1}^{K} e\left(\bar{x}, w^{(k)}\right),$$

where $\left\{w^{(k)}\right\}_{k=1}^{K}$ is a sample of the uniform distribution on the sphere $\mathbb{S}^{n-1}$. We conclude this section with the remark that the additional condition $g(\bar{x}, \mu) < 0$ is not very restrictive. Indeed, as shown in [2, Prop. 3.11], it will always hold true if there exists at all some $z \in \mathbb{R}^n$ with $g(\bar{x}, z) < 0$ (Slater point) and if, in addition, the value $\varphi(\bar{x})$ of the probability function in $\bar{x}$ is not smaller than 0.5 (note that one is usually interested in probabilities close to one). In particular, according to Proposition 3.1, the condition $g(\bar{x}, \mu) < 0$ already implies that the probability function φ is (strongly) continuous. However, it does not have to be differentiable yet, despite the fact that the function g is supposed to be so [2, Prop. 2.2].

3.3 Differentiability of the Radial Probability Function

A crucial step in showing that φ is continuously differentiable consists in verifying the same property for the radial probability function e with respect to its first argument. We recall that the matrix L occuring in (16) is regular as a root of the positive definite matrix Σ.

Lemma 3.1 *In* (14)*, let* g *be a continuously differentiable function which in addition is convex in its second variable. Let* ξ *have an elliptic distribution according to* $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ *with continuous generator* k*. At some* $(\bar{x}, \bar{w}) \in X \times \mathbb{S}^{n-1}$*, assume that* $g(\bar{x}, \mu) < 0$ *and* $\rho(\bar{x}, \bar{w}) < \infty$*. Then, the radial probability function* e *is continuously differentiable with respect to its first argument in a neighbourhood of* $(\bar{x}, \bar{w})$ *and it holds that*

$$D_x e(x, w) \cdot h = -\frac{\tilde{f}(\rho(x, w))}{D_z g(x, \mu + \rho(x, w) Lw)(Lw)} D_x g(x, \mu + \rho(x, w) Lw) \tag{19}$$

for (x, w) *locally around* $(\bar{x}, \bar{w})$.

Proof According to [2, Lemma 3.1], the continuous differentiability of g and its convexity (in the second argument) imply the inequality

$$D_z g(\bar{x}, \mu + rL\bar{w})(L\bar{w}) \geq -\frac{g(\bar{x}, \mu)}{r} > 0 \tag{20}$$

for the unique $r > 0$ satisfying the equation $g(\bar{x}, \mu + rL\bar{w}) = 0$ (i.e., $r = \rho(\bar{x}, \bar{w})$). The inequality has been established in the reference for the centered case $\mu = 0$, but is evident for arbitrary μ as well. Now, (20) allows one to apply the implicit function theorem in order to derive that ρ is continuously differentiable in a neighbourhood of $(\bar{x}, \bar{w})$ with derivative

$$D_x \rho(x, w) = -\frac{1}{D_z g(x, \mu + \rho(x, w) Lw)(Lw)} D_x g(x, \mu + \rho(x, w) Lw) \tag{21}$$

for (x, w) locally around $(\bar{x}, \bar{w})$. By definition, e in Proposition 3.1 has the representation $e = \tilde{F} \circ \rho$ locally around $(\bar{x}, \bar{w})$. Since k is a continuous generator, the function $\tilde{f}$ in (15) is continuous too and, hence, $\tilde{F}$ is continuously differentiable with $\tilde{F}' = \tilde{f}$ by (18). It follows along with (21) that e is continuously differentiable with respect to its first argument in a neighbourhood of $(\bar{x}, \bar{w})$ with derivative (19). □

The analogous result of Lemma 3.1 in the alternative case of $\rho(\bar{x}, \bar{w}) = \infty$ is more delicate to handle. Here, one has to impose an additional growth condition.

Definition 3.2 In (14) let g be continuously differentiable and $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$. We say that the function g satisfies a distribution-adapted growth condition at $\bar{x} \in X$ if there exists a non-decreasing function $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ and a constant $c > 0$ such that the following hold true:

$$\lim\nolimits_{r\to\infty} r^n \mathrm{k}\left(r^2\right) \psi(\delta r) = 0 \quad \forall \delta \geq 0 \tag{22a}$$

$$\|D_x g(x, z)\|_{X'} \leq c\psi(\|z\|) \quad \forall x : \|x - \bar{x}\|_X \leq c^{-1} \quad \forall z : \|z\| \geq c. \tag{22b}$$

Observe that the growth condition (22b) requires the inequality to hold with possibly large modulus and for possibly large norm of z in a sufficiently small neighbourhood of $\bar{x}$.

Example 3.1 Assume that $\xi \sim \mathcal{N}(\mu, \Sigma)$. Then, $\mathrm{k}(r) = e^{-r/2}$ and condition (22b) reduces to

$$\lim_{r\to\infty} r^n e^{-r^2/2} \psi(\delta r) = 0 \quad \forall \delta \geq 0.$$

A possible candidate for ψ is then $\psi(r) := e^r$ for $r \geq 0$ which is non-decreasing and satisfies

$$\lim_{r\to\infty} r^n e^{-r^2/2} \psi(\delta r) = \lim_{r\to\infty} r^n e^{\delta r - r^2/2} = 0 \quad \forall \delta \geq 0.$$

Hence, in the Gaussian case, it is sufficient to verify the exponential growth condition

$$\|D_x g(x, z)\|_{X'} \leq c e^{\|z\|} \quad \forall x : \|x - \bar{x}\|_X \leq c^{-1} \quad \forall z : \|z\| \geq c.$$

This growth condition allows us to formulate the following technical result:

Lemma 3.2 *In* (14)*, let g be a continuously differentiable function which in addition is convex in its second variable. Let ξ have an elliptic distribution according to $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ with continuous generator k. At some $(\bar{x}, \bar{w}) \in X \times \mathbb{S}^{n-1}$, assume that $g(\bar{x}, \mu) < 0$ and $\rho(\bar{x}, \bar{w}) = \infty$. Suppose, moreover, that g satisfies a distribution-adapted growth condition at $\bar{x}$. Then, for every sequence $\{(x_k, w_k)\} \subseteq X \times \mathbb{R}^n$ with $(x_k, w_k) \to (\bar{x}, \bar{w})$ and $\rho(x_k, w_k) < \infty$ it holds that $D_x e(x_k, w_k) \to 0$.*

Proof Observe first that $g(x_k, \mu) < 0$ for k sufficiently large and $\rho(x_k, w_k) < \infty$ by assumption. Hence, we may apply Lemma 3.1 to the points (x_k, w_k) rather than $(\bar{x}, \bar{w})$ and verify that not only $D_x e(x_k, w_k)$ exists and calculates as

$$D_x e(x_k, w_k) = -\frac{\tilde{f}(\rho(x_k, w_k))}{D_z g(x_k, \mu + \rho(x_k, w_k) L w_k)(L w_k)} D_x g(x_k, \mu + \rho(x_k, w_k) L w_k), \tag{23}$$

but also the relation corresponding to (20) holds true:

$$D_z g(x_k, \mu + \rho(x_k, w_k) L w_k)(L w_k) \geq -\frac{g(x_k, \mu)}{\rho(x_k, w_k)} > 0. \tag{24}$$

By continuity, $g(x_k, \mu) < g(\bar{x}, \mu)/2 < 0$ for large enough k and it follows from (24) that

$$D_z g(x_k, \mu + \rho(x_k, w_k) L w_k)(L w_k) \geq \frac{|g(\bar{x}, \mu)|}{2\rho(x_k, w_k)}. \tag{25}$$

Exploiting the facts that $\rho(x_k, w_k) < \infty$, $x_k \to_k \bar{x}$ and $\rho(\bar{x}, \bar{w}) = \infty$, we may refer to [2, Lemma 3.3] in order to derive that $\rho(x_k, w_k) \to \infty$. Then, with L regular and $\|w_k\| \geq 1/2$ for k large enough, it follows that $\|\mu + \rho(x_k, w_k) L w_k\| \to \infty$. Consequently,

$$\|x_k - \bar{x}\|_X \leq c^{-1}, \quad \|\mu + \rho(x_k, w_k) L w_k\| \geq c$$

hold true for the constant c from (22b) and all k large enough. This allows us to combine (23), (25) and (22b), in order to verify that, for k large enough,

$$\|D_x e(x_k, w_k)\|_{X'} \leq 2c \frac{\tilde{f}(\rho(x_k, w_k))\, \rho(x_k, w_k)}{|g(\bar{x}, \mu)|} \psi(\|\mu + \rho(x_k, w_k) L w_k\|).$$

From $\|\rho(x_k, w_k) L w_k\| \to \infty$ and $\|w_k\| \leq 2$ for k large enough infer that

$$\|\mu + \rho(x_k, w_k) L w_k\| \leq 2\rho(x_k, w_k) \|L w_k\| \leq 4\rho(x_k, w_k) \|L\|$$

for k large enough. Therefore, we may exploit (15) along with the fact that ψ is required to be non-decreasing, in order to verify that, for k large enough,

$$\|D_x e(x_k, w_k)\|_{X'} \leq 2c\tilde{c} \frac{\rho^n(x_k, w_k)\, \mathrm{k}(\rho^2(x_k, w_k))}{|g(\bar{x}, \mu)|} \psi(4\|L\|\, \rho(x_k, w_k)).$$

Since $\rho(x_k, w_k) \to \infty$, the right-hand side of the inequality above tends to zero thanks to (22b) and the assertion of the lemma follows. □

We are now in a position to formulate the complementary result to Lemma 3.1:

Lemma 3.3 *Under the assumptions of Lemma 3.2, e is differentiable with respect to its first argument at $(\bar{x}, \bar{w})$ and it holds that $D_x e(\bar{x}, \bar{w}) = 0$.*

Proof Due to $\rho(\bar{x}, \bar{w}) = \infty$, the definition of e implies that $e(\bar{x}, \bar{w}) = 1$. We show first that e is differentiable with respect to its first argument at the point $(\bar{x}, \bar{w})$ itself with $D_x e(\bar{x}, \bar{w}) = 0$. Indeed, should this not be the case, then we find $\varepsilon > 0$ and a sequence $x_k \to \bar{x}$ such that

$$\frac{e(x_k, \bar{w}) - e(\bar{x}, \bar{w})}{\|x_k - \bar{x}\|_X} < -\varepsilon \quad \forall k, \tag{26}$$

since $e(\bar{x}, \bar{w}) = 1$ is a maximum for the probability function e. This entails $e(x_k, \bar{w}) < 1$ and, hence, $\rho(x_k, w) < \infty$ for all k. By continuity of g, our assumption $g(\bar{x}, \mu) < 0$ implies that $g(x, \mu) < 0$ for all x in some δ-ball $\mathbb{B}_\delta(\bar{x})$ around $\bar{x}$. For k large enough, the whole line segment $[x_k, \bar{x}]$ is contained in the ball $\mathbb{B}_\delta(\bar{x})$. Fix an arbitrary such k and define

$$\alpha := \inf \{\tau \geq 0 \mid e(x_k + \tau(\bar{x} - x_k), \bar{w}) = 1\}.$$

Clearly, $\alpha \in [0, 1]$. As stated in the proof of Proposition 3.1, the function e is continuous at all $(x, w) \in X \times \mathbb{S}^{n-1}$ with $g(x, \mu) < 0$. Hence, it is continuous at $(x, \bar{w})$ for all $x \in [x_k, \bar{x}]$. Along with $e(x_k, \bar{w}) < 1$, this entails that $\alpha > 0$ and

$$e(x_k + \alpha(\bar{x} - x_k), \bar{w}) = 1. \tag{27}$$

Therefore, the interval $(0, \alpha)$ is nonempty and

$$g(x_k + \tau(\bar{x} - x_k), \mu) < 0, \quad e(x_k + \tau(\bar{x} - x_k), \bar{w}) < 1 \quad \forall \tau \in (0, \alpha). \tag{28}$$

The second relation implies by definition of e that $\rho(x_k + \tau(\bar{x} - x_k), \bar{w}) < \infty$ for all $\tau \in (0, \alpha)$. Along with the first relation above, this entails that for all $\tau \in (0, \alpha)$ the point $x_k + \tau(\bar{x} - x_k)$ satisfies the same assumptions as did the point $\bar{x}$ in the proof of Lemma 3.1. We may thus conclude that the function $e(\cdot, w)$ is continuously differentiable at the points $x_k + \tau(\bar{x} - x_k)$ for all $\tau \in (0, \alpha)$. Since it is also continuous at these points even for the closed interval $\tau \in [0, \alpha]$ (by continuity of e at (x, w) for all $x \in [x_k, \bar{x}]$), we infer that the real function $\tau \mapsto e(x_k + \tau(\bar{x} - x_k), \bar{w})$ is continuous on $\tau \in [0, \alpha]$ and continuously differentiable on $(0, \alpha)$. Now, the mean value theorem allows us to identify some $\tau_k \in [0, \alpha]$ (in particular, $\tau_k \leq 1$) such that, along with (27),

$$\begin{aligned} D_x e(x_k + \tau_k(\bar{x} - x_k), \bar{w})(\bar{x} - x_k) &= \frac{e(x_k + \alpha(\bar{x} - x_k), \bar{w}) - e(x_k, \bar{w})}{\alpha} \\ &= \frac{1 - e(x_k, \bar{w})}{\alpha}. \end{aligned}$$

Recalling that $e(\bar{x}, \bar{w}) = 1$, we derive from (26) that

$$\|D_x e(x_k + \tau_k(\bar{x} - x_k), \bar{w})\|_{X'} \|\bar{x} - x_k\|_X \geq \frac{\varepsilon}{\alpha} \|\bar{x} - x_k\| \geq \varepsilon \|\bar{x} - x_k\| .$$

Since k had been fixed arbitrarily, we have thus shown that, for all k sufficiently large,

$$\|D_x e(y_k, \bar{w})\| \geq \varepsilon > 0 \quad (y_k := x_k + \tau_k(\bar{x} - x_k)). \tag{29}$$

From $\rho(y_k, \bar{w}) < \infty$ and $y_k \to_k \bar{x}$ due to $\tau_k \leq 1$, we infer that the sequence $(y_k, \bar{w})$ satisfies the assumptions of the sequence (x_k, w_k) in Lemma 3.2. Accordingly, we derive the contradiction $D_x e(y_k, \bar{w}) \to 0$ with (29). This shows our assertion $D_x e(\bar{x}, \bar{w}) = 0$. □

Corollary 3.2 *In* (14)*, let g be a continuously differentiable function which in addition is convex in its second variable. Let* ξ *have an elliptic distribution according to* $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ *with continuous generator* k*. At some* $\bar{x} \in X$*, assume that* $g(\bar{x}, \mu) < 0$ *and that g satisfies a distribution-adapted growth condition at* $\bar{x}$*. Then, there exists a neighbourhood* $\mathcal{V}$ *of* $\bar{x}$ *such that e is differentiable with respect to its first argument in* $\mathcal{V} \times \mathbb{S}^{n-1}$ *with the respective derivatives* $D_x e$ *indicated in Lemmas 3.1 and 3.3. Moreover,* $D_x e$ *is continuous on* $\mathcal{V} \times \mathbb{S}^{n-1}$.

Proof Observe first that the assumptions of Lemmas 3.1 and 3.3 are stable, i.e., the conditions $g(\bar{x}, \mu) < 0$ and the distribution-adapted growth condition at $\bar{x}$ pertain to hold in a neighbourhood $\mathcal{V}$ of $\bar{x}$. Hence, the differentiability of e with respect to its first argument in $\mathcal{V} \times \mathbb{S}^{n-1}$ follows from these two Lemmas along with the corresponding formulae for $D_x e$. We check the continuity of the derivative $D_x e$ at a point (x', w') in $\mathcal{V} \times \mathbb{S}^{n-1}$. This is clear from Lemma 3.1 in case of $\rho\left(x', w'\right) < \infty$ because then $D_x e$ is continuous locally around (x', w'). Otherwise, $\rho\left(x', w'\right) = \infty$. Lemma 3.3 entails that $D_x e(x', w') = 0$. If continuity of $D_x e$ at (x', w') failed, then there would exist some $\varepsilon > 0$ along with a sequence $(x_k, w_k) \to \left(x', w'\right)$ such that $\|D_x e(x_k, w_k)\| \geq \varepsilon$. Then, $x_k \in \mathcal{V}$ for k large enough and $\rho(x_k, w_k) < \infty$, because otherwise $D_x e(x_k, w_k) = 0$ by Lemma 3.3. Now, Lemma 3.2 yields the contradiction $D_x e(x_k, w_k) \to 0$. □

3.4 Differentiability of the Probability Function and Representation of the Derivative as a Spherical Integral

With the preliminary work laid down, we can now present our main result providing insights into the differentiability of the studied probability function. Several subsequent corollaries will allow us to explore various further practical settings.

Theorem 3.3 *In* (14)*, let g be a continuously differentiable function which in addition is convex in its second variable. Let* ξ *have an elliptic distribution according to* $\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ *with continuous generator* k*. Assume that* $g(\bar{x}, \mu) < 0$ *for some given* $\bar{x} \in X$ *and that g satisfies a distribution-adapted growth condition at* $\bar{x}$*. Then,* φ *is continuously differentiable in a neighbourhood of* $\bar{x}$ *and its derivative applied to an* *arbitrary $h \in X$ is given by*

$$D\varphi(\bar{x})(h) = -\int\limits_{\{w \in \mathbb{S}^{n-1} | \rho(\bar{x},w)<\infty\}} \frac{\tilde{f}(\rho(\bar{x},w))}{D_z g(\bar{x},\mu+\rho(\bar{x},w)Lw)(Lw)} D_x g(\bar{x},\mu+\rho(\bar{x},w)Lw)(h)\, d\nu_{\mathrm{U}}(w) \tag{30}$$

with $\tilde{f}$ from (15).

Proof Let $\mathcal{V}$ be some convex neighbourhood of $\bar{x}$ such that the (open) conditions of the Theorem persist to hold, i.e., $g(x,\mu) < 0$ and g satisfies a distribution-adapted growth condition at x for all $x \in \mathcal{V}$. By the continuity of $D_x e$ on $\mathcal{V} \times \mathbb{S}^{n-1}$ (see Corollary 3.2), the maximum $K := \max\limits_{w \in \mathbb{S}^{n-1}} \|D_x e(\bar{x},w)\|$ is attained and, moreover, after possibly shrinking $\mathcal{V}$,

$$\max_{w \in \mathbb{S}^{n-1}} \|D_x e(x,w)\| \le K+1 \quad \forall x \in \mathcal{V}.$$

We show first that φ is differentiable at $\bar{x}$. Observe that the operator P defined by

$$P(h) := \int\limits_{w \in \mathbb{S}^{n-1}} D_x e(\bar{x},w)(h)\, d\nu_{\mathrm{U}}(w) \quad (h \in X) \tag{31}$$

is evidently linear and also continuous due to $|P(h)| \le K\|h\|$ for all $h \in X$. We claim that $D\varphi(\bar{x}) = P$ which at the same time would establish the asserted derivative formula thanks to Lemmata 3.1 and 3.3. To proceed, let $x_k \to \bar{x}$ be an arbitrary sequence and define the sequence of functions $\alpha_k : \mathbb{S}^{n-1} \to \mathbb{R}$ as

$$\alpha_k(w) := \frac{e(x_k,w) - e(\bar{x},w) - D_x e(\bar{x},w)(x_k - \bar{x})}{\|x_k - \bar{x}\|} \quad (w \in \mathbb{S}^{n-1}).$$

Then, $\alpha_k(w) \to 0$ for all $w \in \mathbb{S}^{n-1}$ by Corollary 3.2. Clearly, for k sufficiently large,

$$\frac{|D_x e(\bar{x},w)(x_k - \bar{x})|}{\|x_k - \bar{x}\|} \le K \quad \forall w \in \mathbb{S}^{n-1}.$$

On the other hand, the Mean-Value-Theorem yields the existence of a sequence $\tilde{x}_k \in [x_k, \bar{x}] \subseteq \mathcal{V}$ (by convexity of $\mathcal{V}$) such that

$$\frac{|e(x_k,w) - e(\bar{x},w)|}{\|x_k - \bar{x}\|} = \frac{|D_x e(\tilde{x}_k)(x_k - \bar{x})|}{\|x_k - \bar{x}\|} \le K+1 \quad \forall w \in \mathbb{S}^{n-1}.$$

Altogether, this yields that $|\alpha_k| \le 2K+1$ for all $w \in \mathbb{S}^{n-1}$. Since constant functions are integrable with respect to the uniform distribution on the sphere, we may apply

Lebesgue's dominated convergence Theorem in order to derive along with Proposition 3.1 that

$$0 = \int_{w\in\mathbb{S}^{n-1}} \left(\lim_{k\to\infty} \alpha_k(w)\right) d\nu_{\mathrm{U}}(w) = \lim_{k\to\infty} \int_{w\in\mathbb{S}^{n-1}} \alpha_k(w)\, d\nu_{\mathrm{U}}(w)$$
$$= \lim_{k\to\infty} \frac{\varphi(x_k) - \varphi(\bar{x}) - P(x_k - \bar{x})}{\|x_k - \bar{x}\|}.$$

This shows that φ is differentiable at $\bar{x}$ with derivative $D\varphi(\bar{x}) = P$. From the representation of P the asserted derivative formula follows from Lemmas 3.1 and 3.3. Moreover, since the assumptions of the Theorem persist to hold for all $x \in \mathcal{V}$, we conclude from (31) that

$$D\varphi(x)(h) = \int_{w\in\mathbb{S}^{n-1}} D_x e(x, w)(h)\, d\nu_{\mathrm{U}}(w) \quad \forall x \in \mathcal{V}\ \forall h \in X. \tag{32}$$

It remains to verify the continuous differentiability of φ in a neighbourhood of $\bar{x}$. To this aim, consider an arbitrary sequence $x_k \to \bar{x}$ and define the function

$$\beta(x, w) := \|D_x e(x, w) - D_x e(\bar{x}, w)\| \quad (x \in \mathcal{V};\ \ w \in \mathbb{S}^{n-1}).$$

Then, β is continuous, thanks to Corollary 3.2 and it holds that $\beta(\bar{x}, w) = 0$ for all $w \in \mathbb{S}^{n-1}$. Now, defining the maximum function

$$\beta^{\max}(x) := \max_{w\in\mathbb{S}^{n-1}} \beta(x, w) \quad (x \in \mathcal{V}),$$

we observe that $\beta^{\max}$ is continuous and $\beta^{\max}(\bar{x}) = 0$. Consequently, for some arbitrarily given $\varepsilon > 0$ one has that $\beta^{\max}(x_k) \le \varepsilon$ for k large enough. It follows from (32) that

$$|D\varphi(x_k)(h) - D\varphi(\bar{x})(h)| \le \|h\| \int_{w\in\mathbb{S}^{n-1}} \beta(x_k, w)\, d\nu_{\mathrm{U}}(w)$$
$$\le \|h\|\, \beta^{\max}(x_k) \le \|h\|\, \varepsilon \quad \forall h \in X.$$

Hence, $\|D\varphi(x_k) - D\varphi(\bar{x})\| \le \varepsilon$ for k large enough and, because ε had been chosen arbitrarily, we arrive at $D\varphi(x_k) \to D\varphi(\bar{x})$. □

Corollary 3.4 *The statement of Theorem 3.3 remains true if the growth condition is replaced by the assumption that the set* $\{z \in \Re^n\ : g(\bar{x}, z) \le 0\}$ *is compact.*

Proof First we observe that the multifunction $M(x) := \{z \in \Re^n\ : g(x, z) \le 0\}$ is upper semicontinuous as a consequence of our assumptions [12, Theorem 3.2.1]. Then, with $M(\bar{x})$ required to be compact, there exist some $R \ge 0$ such that $M(x) \subseteq \mathbb{B}(0, R)$ for x near $\bar{x}$. As a result, $\rho(x, w) < \infty$ holds true for x sufficiently close to $\bar{x}$ and for all $w \in \mathbb{S}^{n-1}$. Hence, the proof of Corollary 3.2 (and thus of Theorem 3.3) does not require the growth condition (by appealing to Lemmas 3.2 and 3.3) but follows directly from Lemma 3.1. □

The previous observation allows us to adapt Theorem 3.3) to truncated elliptical distributions.

Definition 3.3 A n-dimensional random vector ξ is said to be truncated elliptical with parameters $(\mu, \Sigma, \mathrm{k}, C)$ if $C \subseteq \Re^n$ is a Lebesgue measurable set of positive measure and there exists $\eta \sim \mathcal{E}(\mu, \Sigma, \mathrm{k})$ such that the density of ξ relates with that of η via

$$f_\xi(z) = f_\eta(z) \cdot \mathbf{1}_C(z)/\mathbb{P}(\eta \in C),$$

where $\mathbf{1}_C$ denotes the characteristic function of C. Note that $\mathbb{P}(\eta \in C) > 0$. As a consequence,

$$\mathbb{P}(\xi \in M) = \frac{\mathbb{P}(\eta \in M \cap C)}{\mathbb{P}(\eta \in C)},$$

for all Lebesgue measurable sets $M \subseteq \Re^n$.

Corollary 3.5 *In (14), let g be a continuously differentiable function which in addition is convex in its second variable. Let ξ have a truncated elliptic distribution with parameters $(\mu, \Sigma, \mathrm{k}, C)$, with continuous generator k and compact set C having nonempty interior. Assume that, for some given $\bar{x} \in X$, it holds that $g(\bar{x}, \mu) < 0$ and*

$$\left\{z \in \Re^n \ : g(\bar{x}, z) \leq 0\right\} \subseteq \operatorname{int} C. \tag{33}$$

Then, φ is continuously differentiable in a neighbourhood of $\bar{x}$ and its derivative applied to an arbitrary $h \in X$ is given by

$$D\varphi(\bar{x})(h) = -\frac{1}{c}\int\limits_{\mathbb{S}^{n-1}} \frac{\tilde{f}(\rho(\bar{x}, w))}{D_z g(\bar{x}, \mu + \rho(\bar{x}, w) Lw)(Lw)} D_x g(\bar{x}, \mu + \rho(\bar{x}, w) Lw)(h)\, d\nu_{\mathrm{U}}(w) \tag{34}$$

with $\tilde{f}$ from (15) and $c = \mathbb{P}(\eta \in C)$ for $\eta \sim \mathcal{E}(\mu, \Sigma, k)$.

Proof Inclusion (33) enforces the compactness of the left-hand side. Then, with the upper semicontinuity argument already employed in the proof of Corollary 3.4, we infer the existence of some neighbourhood $\mathcal{V}$ of $\bar{x}$ such that

$$\left\{z \in \Re^n \ : g(x, z) \leq 0\right\} \subseteq \operatorname{int} C \quad \forall x \in \mathcal{V}.$$

This inclusion implies along with Definition 3.3 that

$$\phi(x) = \mathbb{P}(g(x, \xi) \leq 0) = \frac{1}{c}\mathbb{P}(\eta \in C,\ g(x, \eta) \leq 0) = \frac{1}{c}\mathbb{P}(g(x, \eta) \leq 0) \quad \forall x \in \mathcal{V}.$$

On the other hand, the probability function $\varphi^{(\eta)} := \mathbb{P}(g(x,\eta) \leq 0)$ satisfies the assumptions of Corollary 3.4 at $\bar{x}$, so that the assertion follows from Theorem 3.3 upon observing that $D\varphi(x) = c^{-1} D\varphi^{(\eta)}(x)$ for all $x \in \mathcal{V}$. □

In some models, the random variable takes values in a compact set. The following result indicates a sufficient condition under which the function φ is still differentiable.

Corollary 3.6 *In (14), let g be a continuously differentiable function which in addition is convex in its second variable. Let ξ have a truncated elliptic distribution with parameters $(\mu, \Sigma, \mathrm{k}, C)$, with continuous generator k and compact convex set C, itself representable as $C = \left\{z \in \Re^n \ : \hat{g}(z) \leq 0\right\}$, with $\hat{g} : \Re^n \to \Re$ convex and continuously differentiable. Assume that, for some given $\bar{x} \in X$, it holds that $g(\bar{x}, \mu) < 0$ and*

$$rank \left\{\nabla_z g(x,z), \nabla \hat{g}(z)\right\} = 2, \ \forall z \ s.t. \ \hat{g}(z) = 0 = g(\bar{x}, z). \tag{35}$$

for x in a neighbourhood of $\bar{x}$. Then, φ is continuously differentiable in a neighbourhood of $\bar{x}$ and its derivative applied to an arbitrary $h \in X$ is given by

$$D\varphi(\bar{x})(h) = -\frac{1}{c}\int_{I_\rho} \frac{\tilde{f}\left(\hat{\rho}(\bar{x},w)\right)}{D_z g\left(\bar{x}, \mu + \hat{\rho}(\bar{x},w)Lw\right)(Lw)} D_x g\left(\bar{x}, \mu + \hat{\rho}(\bar{x},w)Lw\right)(h)\, d\nu_{\mathrm{U}}(w) \tag{36}$$

with $\tilde{f}$ from (15), $c = \mathbb{P}(\eta \in C)$ for $\eta \sim \mathcal{E}(\mu, \Sigma, k)$, $I_\rho = \{w \in \mathbb{S}^{n-1} \ : \mu + \rho(x,w)Lw \in C\}$ and $\hat{\rho}(x,w) = \min\left\{\rho(x,w), \sup_{r\ :\mu + rLw \in C} r\right\}$.

Proof Condition (35) is called the rank-2 constraint qualification condition in [3], which as a result of Lemma 4.3 therein (relying on convexity and arguments in $\Re^n$) enables us to establish

$$\mu_\zeta(\left\{w \in \mathbb{S}^{n-1} \ : \hat{g}(\mu + \rho(x,w)Lw) = 0\right\}) = 0.$$

Since moreover C is bounded, $\mathcal{D}\mathrm{om}(\hat{\rho}(x,.)) = \mathbb{S}^{n-1}$ and in fact $\hat{\rho}(x,.)$ is bounded uniformly in x. Now, together with the just made observation we can employ Lemma 3.1 μ_ζ almost everywhere. We can now distinguish over I_ρ and $\left\{w \in \mathbb{S}^{n-1} \ : \mu + \rho(x,w)Lw \notin C\right\}$, on which $D_x e(x,w) = 0$. The arguments justifying the interchange of integration and differentation of Theorem 3.3 can be used again, but this time simplified since there can be no sequence on which $\hat{\rho}$ becomes arbitrarily large. This then allows us to arrive at the desired formula. □

Remark 3.1 In the just given result, condition (35) can be seen in the light of the well known LICQ condition. Should however the set C be defined by multiple inequalities, condition (35) would essentially remain unchanged - indeed it suffices to request linear independence of active gradients two by two. In that setting the given condition is weaker than LICQ. This condition is largely preferable over the abstract zero-measure requirement that it entails. It is not clear how one is to concretely verify the latter, whereas conditions, let alone constraint qualifications, on the nominal data are reasonable.

4 Analysis of the Control Problem

The general framework described in the previous section will be now used to analyze the control problems ($\mathcal{P}$) and ($\mathcal{P}_1$) defined in (9) and (10). Under assumption ($\mathbf{H}_S$), the mappings $\mathcal{G} : L^2([0,T];\mathbb{U}) \times \Re^n \to C([0,T];\mathbb{H})$ with

$$\mathcal{G}(\boldsymbol{u}, z) := e^{A\cdot}x_0 + \int_0^{\cdot} e^{A(\cdot - s)}E(s,z)\,ds + \int_0^{\cdot} e^{A(\cdot - s)}B(z)\boldsymbol{u}(s)\,ds \tag{37}$$

is well defined. Observe that for every $z \in \mathbb{R}^n$, the operator $\mathcal{G}(\cdot, z)$ is affine and continuous. Indeed, if we set $Y_z = \mathcal{G}(\boldsymbol{u}, z) - \mathcal{G}(\boldsymbol{v}, z)$, then it comes that Y_z satisfies the equation $\dot{Y}_z(t) = AY_z(t) + B(z)(\boldsymbol{u}(t) - \boldsymbol{v}(t))$ with $Y_z(0) = 0$. Then, assumption ($\mathbf{H}_S$) guarantees that there exists a constant $C_0 > 0$ such that, for every $z \in \mathbb{R}^n$, for every $\boldsymbol{u}, \boldsymbol{v} \in L^2([0,T];\mathbb{U})$ we have (see [33] for instance)

$$\|\mathcal{G}(\boldsymbol{u}, z) - \mathcal{G}(\boldsymbol{v}, z)\|_{C([0,T];\mathbb{H})} \le C_0(1 + \|z\|_{\mathbb{R}^n})\|\boldsymbol{u} - \boldsymbol{v}\|_{L^2([0,T];\mathbb{U})}, \tag{38a}$$

$$\|\mathcal{G}(\boldsymbol{u}, z)\|_{C([0,T];\mathbb{H})} \le C_0(1 + \|z\|_{\mathbb{R}^n})(1 + \|x_0\|_{\mathbb{H}} + \|\boldsymbol{u}\|_{L^2([0,T];\mathbb{U})}). \tag{38b}$$

Now, define $G : L^2([0,T];\mathbb{U}) \times \mathbb{R}^n \longrightarrow \mathbb{R}$ by

$$G(\boldsymbol{u}, z) := \Psi(\mathcal{G}(\boldsymbol{u}, z)(T)) \quad \forall z \in \mathbb{R}^n, \forall \boldsymbol{u} \in L^2([0,T];\mathbb{U}).$$

Lemma 4.1 *Assume that ($\mathbf{H}_S$) and ($\mathbf{H}_\Psi$) hold. For every $z \in \mathbb{R}^n$, the mapping $\boldsymbol{u} \longmapsto G(\boldsymbol{u}, z)$ is continuously differentiable and for every $\boldsymbol{u}, \boldsymbol{v} \in L^2([0,T];\mathbb{U})$, we have:*

$$D_u G(\boldsymbol{u}, z) \cdot \boldsymbol{v} = \int_0^T \langle B^*(z)\mathbf{y}_z^{\boldsymbol{u}}(s), \boldsymbol{v}(s)\rangle\,ds$$

where $\mathbf{y}_z^{\boldsymbol{u}} \in C([0,T;\mathbb{H})$ is the unique solution of the equation $-\dot{\mathbf{y}}_z^{\boldsymbol{u}}(s) = A^*\mathbf{y}_z^{\boldsymbol{u}}(s)$ and $\mathbf{y}_z^{\boldsymbol{u}}(T) = \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T))$.

Proof As mentioned earlier,for every $z \in \mathbb{R}^n$, the mapping $\boldsymbol{u} \longmapsto \mathcal{G}(\boldsymbol{u}, z)$ is affine and continuous. The derivative $D_u\mathcal{G}(\boldsymbol{u}, z) : L^2([0,T];\mathbb{U}) \to C([0,T];\mathbb{H})$ is well defined and is given by

$$D_u\mathcal{G}(\boldsymbol{u}, z) \cdot \boldsymbol{v} = \int_0^{\cdot} e^{A(\cdot - s)}B(z)\boldsymbol{v}(s)\,ds \quad \forall \boldsymbol{v} \in L^2([0,T];\mathbb{U}).$$

Since Ψ is differentiable, by the chain rule argument and a straightforward calculus, we get that

$$D_u G(\boldsymbol{u}, z) \cdot \boldsymbol{v} = \langle \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T)), \int_0^T e^{A(T-s)}B(z)\boldsymbol{v}(s)\,ds\rangle_{\mathbb{H}}.$$

This equality leads to

$$D_u G(\boldsymbol{u}, z) \cdot \boldsymbol{v} = \int_0^T \langle \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T)), e^{A(T-s)} B(s, z)\boldsymbol{v}(s)\rangle_{\mathbb{H}}\, ds$$
$$= \int_0^T \langle B^*(z) e^{A^*(T-s)} \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T)), \boldsymbol{v}(s)\rangle_{\mathbb{U}}\, ds. \quad (39)$$

Now, introduce the solution $\mathbf{y}_z^{\boldsymbol{u}} \in C([0, T; \mathbb{H})$ of the equation $-\dot{\mathbf{y}}_z^{\boldsymbol{u}}(s) = A^*\mathbf{y}_z^{\boldsymbol{u}}(s)$ with the final condition $\mathbf{y}_z^{\boldsymbol{u}}(T) = \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T))$. This solution is explicitly given by $\mathbf{y}_z^{\boldsymbol{u}}(s) = e^{A^*(T-s)}\nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T))$, which combined to (39) concludes the proof. □

Lemma 4.2 *Assume that* $(\mathbf{H}_S)$ *and* $(\mathbf{H}_\Psi)$ *hold. Then, the function* G *is convex w.r.t its second variable.*

Moreover, for every $\boldsymbol{u} \in L^2([0, T]; \mathbb{U})$*, the mapping* $G(\boldsymbol{u}, \cdot)$ *is continuously differentiable and for every* $z, h \in \mathbb{R}^n$*, we have*

$$D_z G(\boldsymbol{u}, z) \cdot h = \langle \nabla\Psi(\mathcal{G}(\boldsymbol{u}, z)(T)), \mathbf{y}_h(T)\rangle_{\mathbb{H}},$$

where $\mathbf{y}_h \in C([0, T]; \mathbb{H})$ *is the unique solution of the linear system:*

$$\dot{\mathbf{y}}_h(t) = A\mathbf{y}_h(t) + B_1(h)\boldsymbol{u}(s) + E_1(t, h) \quad \textit{for } t \in [0, T], \qquad \mathbf{y}_h(0) = 0 \quad (40)$$

(or equivalently,

$$\mathbf{y}_h(t) = \int_0^t e^{A(t-s)} \Big(B_1(h)\boldsymbol{u}(s) + E_1(s, h) \Big)\, ds).$$

Proof By its definition and by assumptions $(\mathbf{H}_S)$, the mapping $z \longmapsto \mathcal{G}(\boldsymbol{u}, z)(T)$ is affine. Hence, the function G is a composition of the convex function Ψ (by $(\mathbf{H}_\Psi)$) and the affine function $\mathcal{G}$, which implies that G is convex on its second variable (and even separately in its both variables). Moreover, using the linearity of the maps $B_1(\cdot)$ and $E_1(s, \cdot)$, for every $s \in [0, T]$, we get $\mathcal{G}(\boldsymbol{u}, z + h) - \mathcal{G}(\boldsymbol{u}, z) = \mathbf{y}_h$, where $\mathbf{y}_h$ is the solution of (40). Therefore, the mapping $z \longmapsto \mathcal{G}(\boldsymbol{u}, z)$ is continuously differentiable. Finally, by using the chain rule argument, we conclude the proof. □

Recall that the uncertainty enters in the controlled system in a structured manner (see (7)) through a n-dimensional random vector ξ that is distributed according to

$$\xi \sim \mathcal{E}(\mu, \Sigma, \mathrm{k}), \quad \mu \in \mathbb{R}^n, \ \Sigma \in \mathbb{R}^{n\times n},$$

with Σ positive definite. For every $\boldsymbol{u} \in L^2([0, T]; \mathbb{U})$ and every $\eta \in \mathbb{S}^{n-1}$, we define the possibly extended valued radius function $\rho(\boldsymbol{u}, \eta)$ by

$$\rho(\boldsymbol{u}, \eta) := \sup_{r\geq 0} \{r \geq 0 | \Psi(\mathcal{G}(\boldsymbol{u}, \mu + rL\eta)(T)) \leq 0\}.$$

Introduce the cost function $\mathcal{J}_1 : L^2([0,T];\mathbb{U}) \to [0,1]$ defined by

$$\mathcal{J}_1(\boldsymbol{u}) := \mathbb{P}\left[\Psi(\mathbf{x}^{\boldsymbol{u}}(T)) \leq 0\right] = \mathbb{P}\left[\Psi(\mathcal{G}(\boldsymbol{u},\xi)(T)) \leq 0\right] \quad \text{for every } \boldsymbol{u} \in L^2([0,T];\mathbb{U}).$$

In the sequel, we make the following hypothesis:

($\mathbf{H}_G$) The model uncertainty ξ has an elliptical distribution according to $\mathcal{E}(\mu, \Sigma, \mathrm{k})$ such that Gsatisfies
a distribution adapted growth (recall Definition
3.2) condition at all $u \in L^2([0,T];\mathbb{U})$.

Assumption ($\mathbf{H}_G$) is not a restrictive requirement. It indicates that there must be an interplay between the growth of the cost function $\nabla\Psi(x)$ and the decay of the kernel k. Under assumption ($\mathbf{H}_\Psi$), $\nabla\Psi(x)$ has a polynomial growth of degree $m-1$, then one can consider any choice of distribution whose kernel $\kappa(r)$ decays faster than the function r^{-q} with $2q > n+m$. Indeed, recall that from (39), we have

$$\nabla_u G(\boldsymbol{u},z)(s) = B^*(z)e^{A^*(T-s)}\nabla\Psi(\mathcal{G}(\boldsymbol{u},z)(T))$$
$$\forall \boldsymbol{u} \in L^2([0,T];\mathbb{R}^m)\ \forall z \in \mathbb{R}^n\ \forall s \in [0,T].$$

By hypothesis ($\mathbf{H}_S$), there exists some constant $C_1 >$ such that

$$\|\nabla_u G(\boldsymbol{u},z)\|_{L^2([0,T];\mathbb{U})} \leq C_1(1+\|z\|)\|\nabla\Psi(\mathcal{G}(\boldsymbol{u},z)(T))\|_{\mathbb{H}}.$$

Combining this with our assumed growth condition for Ψ, one obtains that

$$\|\nabla_u G(\boldsymbol{u},z)\|_{L^2([0,T];\mathbb{U})} \leq C_1 C_\Psi(1+\|z\|)(1+\|\mathcal{G}(\boldsymbol{u},z)(T)\|_{\mathbb{H}}^{m-1}).$$

Then, using the estimate (38), one obtains that

$$\|\nabla_u G(\boldsymbol{u},z)\|_{L^2([0,T];\mathbb{U})} \leq C(1+|z|)^m(1+\|\boldsymbol{u}\|_{L^2([0,T];\mathbb{U})})^{m-1}$$
$$\forall \mathbf{u} \in L^2([0,T];\mathbb{R}^m), \quad \forall z \in \mathbb{R}^n$$

for some constant $C > 0$. Assumption ($\mathbf{H}_G$) is then satisfied whenever the generator function k satisfies

$$\lim_{r\to\infty} r^n \mathrm{k}(r^2)(1+|r|)^m = 0. \tag{41}$$

Remark 4.1 The multivariate Gaussian distribution has generator $\mathrm{k}(t) = \exp(-t/2)/(2\pi)^{\frac{n}{2}}$ and evidently the growth condition (41) will then hold true. Likewise the logistic and "exponential power" families are readily seen to be compatible. One can easily verify that if ξ follows a multivariate Student distribution that the above growth condition will hold true whenever the degrees of freedom $\nu > m$. Indeed, the multivariate student random vector has as generator:

$$\mathrm{k}(t) = \frac{\Gamma(\frac{n+\nu}{2})}{\Gamma(\frac{\nu}{2})}(\pi\nu)^{-n/2}(1+\frac{t}{\nu})^{-\frac{n+\nu}{2}}.$$

Finally, let us mention that assumption ($\mathbf{H}_G$), along with the convexity property of G stated in Lemma 4.2, is needed to ensure that function $\mathcal{J}_1$ is differentiable. A precise statement is given in the next Theorem.

Theorem 4.1 *Assume ($\mathbf{H}_S$), ($\mathbf{H}_\Psi$) and ($\mathbf{H}_G$). Let $\boldsymbol{u} \in L^2([0,T];\mathbb{U})$ be such that*

$$G(\boldsymbol{u},\mu) = \Psi(\mathcal{G}(\boldsymbol{u},\mu)(T)) < 0.$$

Introduce the set $I_\rho := \{\eta \in \mathbb{S}^{n-1} \mid \rho(\boldsymbol{u},\eta) < \infty\}$ and the function $\alpha(\eta) := \mu + \rho(\boldsymbol{u},\eta)\, L\eta$. Then, $\mathcal{J}_1$ is continuously differentiable at $\boldsymbol{u}$ and its derivative is given by

$$D\mathcal{J}_1(\boldsymbol{u})\cdot\boldsymbol{v} = \int_0^T \int_{I_\rho} \langle B^*(\alpha(\eta))\boldsymbol{p}_\eta^{\boldsymbol{u}}(s), \boldsymbol{v}(s)\rangle_{\mathbb{U}}\, d\nu_{\mathrm{U}}(\eta) ds \qquad \forall \boldsymbol{v} \in L^2([0,T];\mathbb{U}),$$

where the adjoint state $\boldsymbol{p}_\eta^{\boldsymbol{u}} \in C([0,T];\mathbb{H})$ is solution of the equation:

$$-\dot{\boldsymbol{p}}(s) = A^*\boldsymbol{p}(s), \qquad \boldsymbol{p}(T) = -\frac{\widetilde{f}(\rho(\boldsymbol{u},\eta))}{\left\langle \nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)), \mathbf{y}_{L\eta}(T)\right\rangle_{\mathbb{H}}}\nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)),$$

with $\mathbf{y}_{L\eta}$ defined as in (40) (for $h = L\eta$), and the density $\widetilde{f}$ is given in (15).

Proof First, notice that by convexity of $G(\boldsymbol{u},\cdot)$ and by definition of $\rho(\boldsymbol{u},\eta)$, for every $\eta \in I_\rho$, it comes that

$$G(\boldsymbol{u},\mu) \geq G(\boldsymbol{u},\alpha(\eta)) + D_z G(\boldsymbol{u},\alpha(\eta))\cdot(\mu-\alpha(\eta)) = -\rho(\boldsymbol{u},\eta) D_z G(\boldsymbol{u},\alpha(\eta))\cdot L\eta.$$

Using the assumption that $G(\boldsymbol{u},\mu) < 0$ and Lemma 4.2, we get that

$$\langle \nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)), \mathbf{y}_{L\eta}(T)\rangle_{\mathbb{H}} > 0 \qquad \text{for every } \ \eta \in I_\rho.$$

Let $\boldsymbol{v} \in L^2([0,T];\mathbb{U})$ be given. A direct application of Theorem 3.3 leads to

$$D\mathcal{J}_1(\boldsymbol{u})\cdot\boldsymbol{v} = -\int_{I_\rho} \frac{\tilde{f}(\rho(\boldsymbol{u},\eta))}{D_z G(\boldsymbol{u},\alpha(\eta))\cdot L\eta}\left[D_u G(\boldsymbol{u},\alpha(\eta))\cdot\boldsymbol{v}\right] d\nu_{\mathrm{U}}(\eta).$$

By using Lemmas 4.1 & 4.2, we obtain

$$D\mathcal{J}_1(\boldsymbol{u})\cdot\boldsymbol{v} = -\int_{I_\rho} \frac{\tilde{f}(\rho(\boldsymbol{u},\eta))}{\left\langle \nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)), \mathbf{y}_{L\eta}(T)\right\rangle_{\mathbb{H}}}$$

$$\langle \nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)), [D_u\mathcal{G}(\boldsymbol{u},\alpha(\eta))\cdot \boldsymbol{v}](T)\rangle_{\mathbb{H}}\, d\nu_{\mathrm{U}}(\eta)$$
$$= \int\limits_{I_\rho} \left\langle \boldsymbol{p}_\eta^{\boldsymbol{u}}(T), [D_u\mathcal{G}(\boldsymbol{u},\alpha(\eta))\cdot \boldsymbol{v}](T)\right\rangle_{\mathbb{H}} d\nu_{\mathrm{U}}(\eta)$$

Notice that, by assumption $(\mathbf{H}_S)$, for every $\eta \in I_\rho$, the function $\boldsymbol{p}_\eta^{\boldsymbol{u}}$ belongs to $C([0,T],\mathbb{H})$ and $\boldsymbol{p}_\cdot^{\boldsymbol{u}} \in L^2(\mathbb{S}^n; C([0,T];\mathbb{H}))$. Besides, we have

$$D_u\mathcal{G}(\boldsymbol{u},\alpha(\eta))\cdot \boldsymbol{v} = \int_0^{\cdot} e^{A(\cdot-s)}B(\alpha(\eta))\boldsymbol{v}(s)\,ds.$$

Therefore,

$$D\mathcal{J}_1(\boldsymbol{u})\cdot \boldsymbol{v} = \int\limits_{I_\rho}\int_0^T \langle B^*(\alpha(\eta))e^{A^*(T-s)}\boldsymbol{p}_\eta^{\boldsymbol{u}}(T), \boldsymbol{v}(s)\rangle_{\mathbb{U}}\, ds d\nu_{\mathrm{U}}(\eta)$$
$$= \int\limits_{I_\rho}\int_0^T \langle B^*(\alpha(\eta))\boldsymbol{p}_\eta^{\boldsymbol{u}}(s), \boldsymbol{v}(s)\rangle_{\mathbb{U}}\, ds d\nu_{\mathrm{U}}(\eta)$$

for every $\boldsymbol{v} \in L^2([0,T];\mathbb{U})$. Since the measure $d\nu_{\mathrm{U}} \times ds$ is σ-finite on the compact set $\mathbb{S}^{n-1} \times [0,T]$, Tonelli-Fubini's theorem yields the asserted formula. □

Theorem 4.1 indicates that the gradient of $\mathcal{J}_1$, at $\boldsymbol{u}$, is given by the following expression

$$\nabla\mathcal{J}_1(\boldsymbol{u}) := \int_{I_\rho} B^*(\alpha(\eta))\boldsymbol{p}_\eta^{\boldsymbol{u}}(\cdot)\, d\nu_{\mathrm{U}}(\eta) \in L^2([0,T];\mathbb{U}).$$

In this expression, we used the radial decomposition of the random variable ξ. Convexity with respect to uncertainty is crucial here to determine the derivative expressed solely in terms of directions η belonging to the set I_ρ, that is, the directions η associated with a finite radius value $\rho(\boldsymbol{u},\eta)$. The same calculus can be done if instead of $(\mathbf{H}_G)$, we assume that the random variable ξ has a truncated elliptical distribution with parameters $(\mu, \Sigma, \mathrm{k}, \mathbb{B}(0, M))$ where $M > 0$. In this case, the differentiability can still be analysed by using Corollary 3.5 or 3.6.

In the control problem $(\mathcal{P})$ (given in (9)), the cost function is constituted by a sum of a probability cost $\mathcal{J}_1$ and the expectation of a running cost, denotes as $\mathcal{J}_2 : L^2([0,T];\mathbb{U}) \longrightarrow \mathbb{R}$ and defined by

$$\mathcal{J}_2(\boldsymbol{u}) := \mathbb{E}\Big[\int_0^T \ell(s, \mathbf{x}^{\boldsymbol{u}}(s), \boldsymbol{u}(s), \cdot)\, ds\Big] \quad \forall \boldsymbol{u} \in L^2([0,T];\mathbb{U}).$$

It is noteworthy that under assumption $(\mathbf{H}_S)$ that $\mathbf{x}_\omega^{\boldsymbol{u}} \in C([0,T],\mathbb{H})$ for every $\boldsymbol{u} \in L^2(0,T;\mathbb{U})$ and for every $\omega \in \Omega$. Additionally, due to assumption $(\mathbf{H}_\ell)$, the cost $\mathcal{J}_2(\boldsymbol{u})$ is finite.

Theorem 4.2 *Assume* ($\mathbf{H}_S$)*,* ($\mathbf{H}_\ell$)*. Let* $\boldsymbol{u} \in L^2([0,T];\mathbb{U})$ *and its associated state* $\mathbf{x}^{\boldsymbol{u}}$ *(i.e., solution of* (6)*). Then,* $\mathcal{J}_2$ *is continuously differentiable around* $\boldsymbol{u}$ *and its derivative is given by*

$$D\mathcal{J}_2(\boldsymbol{u}) \cdot \boldsymbol{v} = \mathbb{E}\Big[\int_0^T \langle \ell_u(t, \mathbf{x}^{\boldsymbol{u}}(t), \boldsymbol{u}(t), \cdot) + B^*(\xi)\boldsymbol{q}^{\boldsymbol{u}}(t), \boldsymbol{v}(t)\rangle_{\mathbb{U}}\, dt\Big] \quad \forall \boldsymbol{v} \in L^2([0,T];\mathbb{U}),$$

where the co-state $\boldsymbol{q}_\omega^{\boldsymbol{u}} \in C([0,T];\mathbb{H})$ *is solution of the adjoint equation:*

$$-\dot{\boldsymbol{q}}_\omega^{\boldsymbol{u}}(s) = A^*\boldsymbol{q}_\omega^{\boldsymbol{u}}(s) + \ell'_x(s, \mathbf{x}_\omega^{\boldsymbol{u}}(s), \boldsymbol{u}(s), \omega), \qquad \boldsymbol{q}(T) = 0.$$

Proof Notice that for $\boldsymbol{u} \in L^2([0,T];\mathbb{U})$

$$\mathcal{J}_2(\boldsymbol{u}) = \int_\Omega \int_0^T \ell(t, \mathcal{G}(\boldsymbol{u}, \xi(\omega))(t), \boldsymbol{u}(t), \omega) dt d\mathbb{P}(\omega).$$

Since ℓ is of class C^1 and satisfies ($\mathbf{H}_\ell$), by superposition principle, the function $\mathcal{J}_2$ is differentiable and, for every $\boldsymbol{v} \in L^2([0,T];\mathbb{U})$, we have

$$\begin{aligned} D\mathcal{J}_2(\boldsymbol{u}) \cdot \boldsymbol{v} = \int_\Omega \int_0^T & \Big[\langle \ell_x(t, \mathcal{G}(\boldsymbol{u}, \xi(\omega))(t), \boldsymbol{u}(t), \omega), \\ & \int_0^t e^{A(t-s)} B(\xi(\omega))\boldsymbol{v}(s)\, ds\rangle_{\mathbb{H}}\Big]\, dt d\mathbb{P}(\omega) \\ & + \int_\Omega \int_0^T \langle \ell_u(t, \mathcal{G}(\boldsymbol{u}, \xi(\omega))(t), \boldsymbol{u}(t), \omega), \boldsymbol{v}(t)\rangle_{\mathbb{U}}\, dt d\mathbb{P}(\omega). \end{aligned}$$

By introducing the adjoint state $\boldsymbol{q}_\omega^{\boldsymbol{u}}$ (for every $\omega \in \Omega$), we notice that:

$$\begin{aligned} &\int_0^T \langle \ell_x(t, \mathbf{x}_\omega^{\boldsymbol{u}}(t), \boldsymbol{u}(t), \omega), \int_0^t e^{A(t-s)} B(\xi(\omega))\boldsymbol{v}(s)\, ds\rangle_{\mathbb{H}}\, dt \\ &\quad = \int_0^T \langle B^*(\xi(\omega))\boldsymbol{q}_\omega^{\boldsymbol{u}}(t), \boldsymbol{v}(t)\rangle_{\mathbb{U}}\, dt, \end{aligned}$$

which concludes the proof. □

Unlike the cost function $\mathcal{J}_1$, the differentiability of the function $\mathcal{J}_2$ does not require any convexity property with respect to the uncertainty. For this function, the gradient at any $\boldsymbol{u} \in \mathcal{U}$, is identified to the function defined on $[0,T]$ by $\nabla J_2(\boldsymbol{u})(t) = \mathbb{E}\big[\ell_u(t, \mathcal{G}(\boldsymbol{u}, \xi)(t), \boldsymbol{u}(t), \cdot) + B^*(\xi)\boldsymbol{q}^{\boldsymbol{u}}(t)\big]$. This function involves an expectation over all elementary events in Ω. Now, we can state the optimality condition for the control problem with uncertainties.

Theorem 4.3 *Assume* ($\mathbf{H}_S$)*,* ($\mathbf{H}_\Psi$)*,* ($\mathbf{H}_\ell$) *and* ($\mathbf{H}_G$)*. Let* $\boldsymbol{u}$ *be an optimal control law and* $\mathbf{x}^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u}, \xi)$ *its associated optimal state. We assume that*

$$G(\boldsymbol{u}, \mu) = \Psi(\mathcal{G}(\boldsymbol{u}, \mu)(T)) < 0.$$

Introduce the set $I_\rho := \left\{\eta \in \mathbb{S}^{n-1} \mid \rho(\boldsymbol{u},\eta) < \infty\right\}$ *and the function* $\alpha(\eta) := \mu + \rho\,(\boldsymbol{u},\eta)\,L\eta$, *for* $\eta \in I_\rho$.

There exist perturbed adjoint states $\boldsymbol{p} = \{\boldsymbol{p}_\eta,\ \eta \in I_\rho\} \subset C([0,T];\mathbb{H})$ *and* $\boldsymbol{q} = \{\boldsymbol{q}_\omega,\ \omega \in \Omega\} \subset C([0,T];\mathbb{H})$ *satisfying*

$$-\dot{\boldsymbol{p}}_\eta(t) = A^*\boldsymbol{p}_\eta(t), \qquad \boldsymbol{p}_\eta(T) = \gamma(\eta)\nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)), \tag{42a}$$

$$-\dot{\boldsymbol{q}}_\omega(t) = A^*\boldsymbol{q}_\omega(t) + \ell_x(t,\mathcal{G}(\boldsymbol{u},\xi(\omega))(t),\boldsymbol{u}(t),\omega), \qquad \bar{\boldsymbol{q}}_\omega(T) = 0, \tag{42b}$$

where for every $\eta \in I_\rho$, $\gamma(\eta) = \dfrac{-\tilde{f}\,(\rho\,(\boldsymbol{u},\eta))}{\left\langle\nabla\Psi(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)),\mathbf{y}_{L\eta}(T)\right\rangle}$, *and*

$$\dot{\mathbf{y}}_{L\eta}(t) = A\mathbf{y}_{L\eta}(t) + B_1(L\eta)\boldsymbol{u}(t) + E_1(t,L\eta), \qquad \mathbf{y}_{L\eta}(0) = 0. \tag{42c}$$

Moreover, for a.e $t \in (0,T)$ *and for every* $v \in U$, *we have*

$$\Big\langle\mathbb{E}\big[\ell_u(t,\mathcal{G}(\boldsymbol{u},\xi)(t),\boldsymbol{u}(t),\cdot) + B^*(\xi)\boldsymbol{q}(t)\big] + \int_{I_\rho} B^*(\alpha(\eta))\boldsymbol{p}_\eta(t)\,d\nu_{\mathrm{U}}(\eta),\boldsymbol{u}(t)-v\Big\rangle_{\mathbb{U}} \geq 0. \tag{42d}$$

Proof By combining Theorems 4.1 and 4.2, and by convexity of U, we obtain for every $\boldsymbol{v} \in \mathcal{U}$

$$\begin{aligned}0 &\leq D(\mathcal{J}_1+\mathcal{J}_2)(\boldsymbol{u}).(\boldsymbol{u}-\boldsymbol{v})\\ &\leq \int_0^T \mathbb{E}\big[\left\langle\ell_u(t,\mathcal{G}(\boldsymbol{u},\xi)(t),\boldsymbol{u}(t),\cdot) + B^*(\xi)\boldsymbol{q}(t),\boldsymbol{u}(t)-\boldsymbol{v}(t)\right\rangle_{\mathbb{U}}\big]\,dt\\ &\quad + \int_0^T\int_{I_\rho}\left\langle B^*(\alpha(\eta))\boldsymbol{p}_\eta(t),\boldsymbol{u}(t)-\boldsymbol{v}(t)\right\rangle_{\mathbb{U}}\,d\nu_{\mathrm{U}}(\eta)dt.\end{aligned} \tag{43}$$

To conclude the proof, we use the *spike perturbation* techniques. Let $v \in U$ and denote by $\Gamma_0 \subset [0,T]$ the set of Lebesgue points of the function

$$t \longmapsto \Big\langle\mathbb{E}\big[\ell_u(t,\mathcal{G}(\boldsymbol{u},\xi)(t),\boldsymbol{u}(t),\cdot) + B^*(\xi)\boldsymbol{q}(t)\big] + \int_{I_\rho} B^*(\alpha(\eta))\boldsymbol{p}_\eta(t)\,d\nu_{\mathrm{U}}(\eta),\boldsymbol{u}(t)-v\Big\rangle_{\mathbb{U}}.$$

This application being continuous on $[0,T]$, the set Γ_0 is of full measure. Let $t_0 \in \Gamma_0$, for $\varepsilon > 0$, consider the perturbation

$$\boldsymbol{v}_\varepsilon(s) := \begin{cases} v & \text{if } s \in (t_0-\varepsilon,t_0+\varepsilon)\\ \boldsymbol{u}(s) & \text{otherwise.}\end{cases}$$

Using this perturbation in (43) and dividing by 2ε, we get

$$0 \leq \frac{1}{2\varepsilon}\int_{t_0-\varepsilon}^{t_0+\varepsilon} \Big\langle \mathbb{E}\big[\ell_u(t, \mathcal{G}(\boldsymbol{u}, \xi)(t), \boldsymbol{u}(t), \cdot) + B^*(\xi)\boldsymbol{q}(t)\big] + \int\limits_{I_\rho} B^*(\alpha(\eta))\boldsymbol{p}_\eta(t)\, d\nu_{\mathrm{U}}(\eta), \boldsymbol{u}(t) - v\Big\rangle\, dt.$$

By letting ε goes to 0, we obtain

$$0 \leq \Big\langle \mathbb{E}\big[\ell_u(t_0, \mathcal{G}(\boldsymbol{u}, \xi)(t_0), \boldsymbol{u}(t_0), \cdot) + B^*(\xi)\boldsymbol{p}(t_0)\big] + \int\limits_{I_\rho} B^*(t_0, \alpha(\eta))\boldsymbol{p}_\eta(t_0)\, d\nu_{\mathrm{U}}(\eta), \boldsymbol{u}(t_0) - v\Big\rangle,$$

for any $t_o \in \Gamma_0$, which concludes the proof. □

4.1 Chance-Constrained Control Problems

The key point in deriving the optimality conditions for problem (9) is the derivation formula for the probability functional. The same analysis can be extended to other control problems with chance constraints on the state at the final time. For instance, consider the following control problem:

$$(\mathcal{P}_1) \qquad \text{Maximize}\ \Big\{ \mathcal{J}_2(\boldsymbol{u}) \mid \boldsymbol{u} \in \mathcal{U} \ \ \text{and}\ \mathbb{P}\big[\Psi(\mathcal{G}(\boldsymbol{u}, \xi)(T)) \leq 0\big] \geq c \Big\}.$$

Theorem 4.4 *Consider the same setting as in Theorem 4.3. Let $\boldsymbol{u}$ be an optimal control law of $(\mathcal{P}_1)$ and $\mathbf{x}^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u}, \xi)$ its associated optimal state. We assume that*

$$G(\boldsymbol{u}, \mu) = \Psi(\mathcal{G}(\boldsymbol{u}, \mu)(T)) < 0.$$

Introduce the set $I_\rho := \big\{\eta \in \mathbb{S}^{n-1} \mid \rho(\boldsymbol{u}, \eta) < \infty\big\}$ and the function $\alpha(\eta) := \mu + \rho(\boldsymbol{u}, \eta)\, L\eta$, for $\eta \in I_\rho$. There exist $(\lambda_0, \lambda) \in \{0, 1\} \times \mathbb{R}^-$, and perturbed adjoint states $\boldsymbol{p} = \{\boldsymbol{p}_\eta,\ \eta \in I_\rho\} \subset C([0, T]; \mathbb{H})$ and $\boldsymbol{q} = \{\boldsymbol{q}_\omega,\ \omega \in \Omega\} \subset C([0, T]; \mathbb{H})$ satisfying

$$(\lambda_0, \lambda) \neq 0, \tag{44a}$$

$$\lambda\left(\mathbb{P}\big[\Psi(\mathbf{x}^{\boldsymbol{u}}_\omega(T)) \leq 0\big] - c\right) = 0, \quad \textit{and}\ \mathbb{P}\big[\Psi(\mathbf{x}^{\boldsymbol{u}}_\omega(T)) \leq 0\big] \geq c, \tag{44b}$$

$$-\dot{\boldsymbol{p}}_\eta(t) = A^*\boldsymbol{p}_\eta(t), \qquad \bar{\boldsymbol{p}}_\eta(T) = \gamma(\eta)\nabla\Psi(\mathcal{G}(\boldsymbol{u}, \alpha(\eta))(T)), \tag{44c}$$

$$-\dot{\boldsymbol{q}}_\omega(t) = A^*\boldsymbol{q}_\omega(t) + \lambda_0\ell_x(t, \mathbf{x}_\omega(t), \boldsymbol{u}(t), \omega), \qquad \bar{\boldsymbol{q}}_\omega(T) = 0, \tag{44d}$$

where for every $\eta \in I_\rho$, $\gamma(\eta) = \dfrac{-\tilde{f}\,(\rho\,(\boldsymbol{u}, \eta))}{\big\langle \nabla\Psi(\mathcal{G}(\boldsymbol{u}, \alpha(\eta))(T)), \mathbf{y}_{L\eta}(T)\big\rangle}$, with

$$\dot{\mathbf{y}}_{L\eta}(t) = A\mathbf{y}_{L\eta}(t) + B_1(L\eta)\boldsymbol{u}(t) + E_1(t, L\eta), \qquad \mathbf{y}_{L\eta}(0) = 0. \tag{44e}$$

Moreover, for a.e $t \in (0, T)$ and for every $v \in U$, we have

$$\Big\langle \lambda_0 \mathbb{E}\big[\ell_u(t, \mathbf{x}^{\boldsymbol{u}}(t), \boldsymbol{u}(t), \cdot) + B^*(\xi)\boldsymbol{q}(t)\big] + \lambda \int_{I_\rho} B^*(\xi)\boldsymbol{p}_\eta(t)\, d\nu_{\mathrm{U}}(\eta), \boldsymbol{u}(t) - v\Big\rangle_{\mathbb{U}} \geq 0. \tag{44f}$$

Proof By general results of optimization theory, the maximal solution $\boldsymbol{u}$ is associated with a non trivial pair of multipliers $(\lambda_0, \lambda) \in \{0, 1\} \times \mathbb{R}^+$ such that for all $v \in \mathcal{U}$

$$0 \leq \lambda_0 D\mathcal{J}_2(\boldsymbol{u}).(\boldsymbol{u} - \boldsymbol{v}) + \lambda D\mathcal{J}_1(\boldsymbol{u}).(\boldsymbol{u} - \boldsymbol{v}). \tag{45}$$

The rest of the proof is based on the differentiability formulas for $\mathcal{J}_1$ (in Theorem 4.1) and for $\mathcal{J}_2$ (in Proposition 4.2) and follows similar arguments as in the proof of Theorem 4.3. □

At this stage the question can be raised whether the necessary optimality conditions are sufficient or not. In particular, could the probabilistic function $\mathcal{J}_1$ be a concave functional? Although the controlled system is linear and the function Ψ is convex, the cost function $\mathcal{J}_1$ can not be expected to be convex in general. However, it could be the case that some upper-level sets of $\mathcal{J}_1$ are convex. The following result provides an insight into the matter.

Proposition 4.1 *Assume that the hypothesis* $(\mathbf{H}_S)$*,* $(\mathbf{H}_\Psi)$ *and* $(\mathbf{H}_G)$ *hold true. In addition assume that*

- *the mapping $\boldsymbol{u} \mapsto \rho(\boldsymbol{u}, \eta)$ is β-concave*[2] *for each η (see, e.g., Definition 3.2 in* [4]*).*
- *the random vector ξ is multivariate Gaussian.*

Then the upper level sets of $\mathcal{J}_1$ are convex for all levels beyond $c^* = \Phi(\sqrt{n - \beta})$. Moreover $\mathcal{J}_1$ is concave on any of the latter upper level sets. Here Φ stands for the 1 dimensional standard Gaussian distribution function.

Proof Notice that G is convex in its both arguments. Since the underlying space L^2 is reflexive, the result of the proposition is a direct application of [31, Theorem 11] along the lines of [4, Corollary 4.3]. The concavity of $\mathcal{J}_1$ results from [4, Theorem 3.1]. □

Under the assumptions of Proposition 4.1 it is always true that ρ is quasi-concave, i.e., $\beta = -\infty$. If G is jointly convex in both arguments, then $\beta = 1$, i.e., ρ is concave. In other situations the specific structure of G has to be explored to learn of an appropriate β value. However when G is jointly convex, one can leverage Prékopa's celebrated log-concavity results to ensure that all upper level sets of $\mathcal{J}_1$ are convex, not just beyond a given level - $\mathcal{J}_1$ being log-concave itself. Proposition 4.1 does ensure however the stronger concavity of $\mathcal{J}_1$ on the exhibited upper level sets.

[2] for $\beta < 0$, this means $\boldsymbol{u} \mapsto \rho^\beta(\boldsymbol{u}, \eta)$ is convex.

Corollary 4.5 *Assume the same setting as in Proposition 4.1. Assume that the distributed cost ℓ is concave on both state and control variables . Let $\boldsymbol{u} \in \mathcal{U}$ be an admissible control law and $\mathbf{x}^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u}, \xi)$ its associated optimal state. Assume that*

$$G(\boldsymbol{u}, \mu) = \Psi(\mathcal{G}(\boldsymbol{u}, \mu)(T)) < 0.$$

Assume also that the probability level $c > c^*$. Then, $\boldsymbol{u}$ is an optimal control of $(\mathcal{P}_1)$ if and only if there exists a nontrivial pair of multipliers $(\lambda_0, \lambda) \neq 0$ such the optimality system (44) is satisfied.

Remark 4.2 (*On Gaussian random vectors*) The restriction to Gaussian random vectors in Proposition 4.1 is only to make the statement less involved. A very similar result would indeed hold for essentially any Elliptically symmetrically distributed random vector ξ such that assumption $(\mathbf{H}_G)$ holds. We refer the reader to Table 1 [4]. To provide an example, if ξ would be multi-variate Student with ν degrees of freedom, then the resulting threshold would be:

$$c^* = (\frac{1}{2} - q)F_{n,\nu}\left(\frac{\nu(n-\alpha)}{\delta(q)^2(n\nu - n)}\right) + q + \frac{1}{2}$$

with $\delta(q)$ the unique solution (in δ) to the equation

$$\mathfrak{B}_i\left(\frac{n-1}{2}, \frac{1}{2}, 1-\delta^2\right) = (1-2q)\mathfrak{B}_c\left(\frac{n-1}{2}, \frac{1}{2}\right),$$

where $\mathfrak{B}_i$ ($\mathfrak{B}_c$) refers to the incomplete (resp. complete) Beta function, $F_{n,\nu}$ is the Fisher-Snedecor distribution with n and ν degrees of freedom and $q \in (0, \frac{1}{2})$ is a free parameter.

Remark 4.3 (*On the threshold*) The given threshold in the previous results is to be understood as a conservative estimate - and is by no means tight. As a result, in concrete applications it may well be that the upper level sets of $\mathcal{J}_1$ are convex beyond $c^* = \frac{1}{2}$ even though this can only be asserted theoretically for a much larger c^*.

4.2 Some Examples

The results from the previous section apply to a class of optimal control problems with a linear state equation. This limitation arises from the need for convexity to ensure the differentiability of the probability function $\mathcal{J}_1$. The linear control class is compelling and encompasses various applications in control theory. Here, we present two simple examples.

Example in finite dimensional space Consider first the simplest case in finite dimensional spaces $\mathbb{H} = \mathbb{R}^d$ and $\mathbb{U} = \mathbb{R}^m$, with a final cost defined on $\mathbb{R}^d$ by $\Psi(x) := \frac{1}{2}\|x - \bar{x}\|_2^2 - r_0^2$ for a fixed radius r_0. Let A be a $d \times d$ matrix and $z \longmapsto B(z)$ a linear application from $\mathbb{R}^n$ to $\mathbb{R}^{d\times m}$. Given a level of success $c \in [0, 1]$ and a fixed function

$\bar{u} \in L^2([0,T];\mathbb{R}^m)$, the control problem is as follows

$$
\begin{aligned}
&\text{Minimize } \frac{1}{2}\int_0^T \|u(t)-\bar{u}(t)\|^2_{\mathbb{R}^m}\,dt\\
&\qquad \dot{\mathbf{x}}^u_\omega(t) = A\mathbf{x}^u_\omega(t) + B(\xi(\omega))\boldsymbol{u}(t) \quad \text{for } t \in [0,T],\\
&\qquad \mathbf{x}^u_\omega(0) = x_0 \in \mathbb{R}^d,\\
&\qquad u(t) \in \mathbb{R}^m \quad \text{a.e. } t \in [0,T],\\
&\qquad \mathbb{P}\Big[\Psi(\mathbf{x}^{\boldsymbol{u}}(T)) \le 0\Big] \ge c.
\end{aligned}
$$

For simplicity, we also consider here the special case of a multivariate centred Gaussian distribution for the random vector ξ. Therefore, all the assumptions ($\mathbf{H}_S$), ($\mathbf{H}_\Psi$), ($\mathbf{H}_\ell$) and ($\mathbf{H}_G$) are fulfilled.

Let $\boldsymbol{u}$ be an optimal control law of the above problem and $\mathbf{x}^{\boldsymbol{u}} = \mathcal{G}(\boldsymbol{u},\xi)$ its associated optimal state. We assume that

$$
\frac{1}{2}\|\mathcal{G}(\boldsymbol{u},\mu)(T)) - \bar{x}\|^2_{\mathbb{R}^d} < r_0^2.
$$

Introduce the set $I_\rho := \big\{\eta \in \mathbb{S}^{n-1} \mid \rho(\boldsymbol{u},\eta) < \infty\big\}$ and the function $\alpha(\eta) := \rho\,(\boldsymbol{u},\eta)\,L\eta$, for $\eta \in I_\rho$. By Theorem 4.4, there exist $(\lambda_0,\lambda) \in \{0,1\}\times\mathbb{R}^-$, and perturbed adjoint states $\boldsymbol{p} = \{\boldsymbol{p}_\eta,\ \eta \in I_\rho\} \subset C([0,T];\mathbb{R}^d)$ satisfying

$$
(\lambda_0,\lambda) \neq 0, \tag{46a}
$$

$$
\begin{aligned}
&\lambda\left(\mathbb{P}\Big[\frac{1}{2}\|\mathcal{G}(\boldsymbol{u},\mu)(T)) - \bar{x}\|^2_{\mathbb{R}^d} \le r_0^2\Big] - c\right) = 0, \quad \text{and}\\
&\quad \mathbb{P}\Big[\frac{1}{2}\|\mathcal{G}(\boldsymbol{u},0)(T)) - \bar{x}\|^2_{\mathbb{R}^d} \le r_0^2\Big] \ge c,
\end{aligned} \tag{46b}
$$

$$
-\dot{\boldsymbol{p}}_\eta(t) = A^*\boldsymbol{p}_\eta(t), \qquad \bar{\boldsymbol{p}}_\eta(T) = \gamma(\eta)\Big[\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T) - \bar{x}\Big], \tag{46c}
$$

where for every $\eta \in I_\rho$, $\gamma(\eta) = \dfrac{-\tilde{f}\,(\rho\,(\boldsymbol{u},\eta))}{\langle \mathcal{G}(\boldsymbol{u},\alpha(\eta))(T) - \bar{x}, \mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)\rangle}$. Moreover, for a.e $t \in (0,T)$ and for every $v \in U$, we have

$$
-\lambda_0(\boldsymbol{u}(t) - \bar{u}(t)) + \lambda \int\limits_{I_\rho} B^*(\xi)\boldsymbol{p}_\eta(t)\,d\nu_{\mathrm{U}}(\eta) = 0. \tag{46d}
$$

Heat equation with Neumann Boundary control In the context of infinite dimensional control problem, consider the parabolic controlled system described in Example 2.1. For simplicity, we assume here that $b_0 = 0$, and $E_0 = 0$. Consider again the special case of a multivariate centred Gaussian distribution for the random vector ξ.

The control problem is the following

$$
\text{Maximize } \mathbb{E}\Big[\frac{-\lambda_Q}{2}\int_0^T\int_{\mathcal{O}} |\mathcal{G}(\boldsymbol{u},\xi)(t,y) - \mathbf{x}_Q(t,y)|^2\,dydt\Big]
$$

$$
\begin{aligned}
&+\frac{-\lambda_\Sigma}{2}\int_0^T\int_{\partial\mathcal{O}} |u(t,y)|^2\,dydt \\
&+\mathbb{P}\Big[\frac{1}{2}\|\mathcal{G}(\boldsymbol{u},\xi(\omega))(T)-\mathbf{x}_\mathcal{O}\|^2 - R^2 \le 0\Big], \\
&\text{such that } \boldsymbol{u}\in L^2([0,T];L^2(\partial\mathcal{O})),
\end{aligned}
$$

with $\lambda_Q > 0$, $\lambda_\Sigma > 0$, and the radius $R > 0$ are given constants. In this example, the hypothesis $(\mathbf{H}_S)$, $(\mathbf{H}_\Psi)$, $(\mathbf{H}_\ell)$ hold true. Assume the control problem has an optimal solution $\boldsymbol{u}\in\mathcal{U}$ such that

$$
\frac{1}{2}\|\mathcal{G}(\boldsymbol{u},0)(T)-\mathbf{x}_\mathcal{O}\|^2 < R^2.
$$

Then, Theorem 4.3 provides the first optimality optimality conditions . Introduce the set $I_\rho := \{\eta\in\mathbb{S}^{n-1} \mid \rho(\boldsymbol{u},\eta) < \infty\}$ and the function $\alpha(\eta) := \rho(\boldsymbol{u},\eta)\,L\eta$, for $\eta\in I_\rho$.

There exist perturbed adjoint states $\boldsymbol{p}\in L^2([0,T];\mathbb{H})$ and $\boldsymbol{q}\in L^2([0,T];\mathbb{H})$ satisfying

$$
\begin{cases}
-\partial_t \boldsymbol{p}_\eta(t,y) = \Delta\boldsymbol{p}_\eta(t,y) + a(y)\boldsymbol{p}_\eta(t,y) & \text{in } (0,T)\times\mathcal{O}, \\
\frac{\partial \boldsymbol{p}_\eta}{\partial\nu}\Big|_{\partial\mathcal{O}}(t,y) = 0 & \text{in } (0,T)\times\partial\mathcal{O}, \\
\boldsymbol{p}_\eta(T,y) = \gamma(\eta)(\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T,y) - x_\mathcal{O}(y)) & \text{in } \mathcal{O},
\end{cases} \tag{47a}
$$

$$
\begin{cases}
-\partial_t \boldsymbol{q}_\omega(t,y) = \Delta\boldsymbol{q}_\omega(t,y) + a(y)\boldsymbol{q}_\omega(t,y) & \\
\qquad\qquad +\lambda_Q(\mathcal{G}(\boldsymbol{u},\xi(\omega)(\eta))(t,y) - \mathbf{x}_Q(t,y)) & \text{in } (0,T)\times\mathcal{O}, \\
\frac{\partial \boldsymbol{q}_\omega}{\partial\nu}\Big|_{\partial\mathcal{O}}(t,y) = 0 & \text{in } (0,T)\times\partial\mathcal{O}, \\
\boldsymbol{q}_\omega(T,y) = 0 & \text{in } \mathcal{O}.
\end{cases} \tag{47b}
$$

where $\gamma(\eta) = \dfrac{-\tilde{f}\,(\rho\,(\boldsymbol{u},\eta))}{\langle\mathcal{G}(\boldsymbol{u},\alpha(\eta))(T)-\mathbf{x}_\mathcal{O},\mathcal{G}(\boldsymbol{u},L\eta)(T)\rangle}$, for every $\eta\in I_\rho$. Moreover, for a.e $t\in(0,T)$, we have

$$
\int_{\partial\mathcal{O}}\Big(-\lambda_\Sigma\boldsymbol{u}(t,y) + \mathbb{E}\big[\boldsymbol{q}(t,y)\big] + \int\limits_{I_\rho}\boldsymbol{p}_\eta(t,y)d\nu_{\mathrm{U}}(\eta)\Big]dy = 0.
$$

5 Conclusion

In conclusion, in this paper we have investigated the optimality conditions for a control problem subject to a probabilistic constraint or with a probabilistic cost. The key point is the differentiability of the probabilistic cost function. The differentiability result that is derived in the present work depends on a convexity structure with respect to the random variable. For a class of linear control problems, we show that the optimality conditions can be expressed in the form of a Pontryagin principle. To the best of our knowledge, the results obtained in this paper are new in the context of optimal control problems under uncertainties.

The obtained result constitute a starting point for future studies on optimality conditions and for developing efficient numerical methods for solving probabilistic constrained optimal control problems. More precisely, it should be emphasized that the requisite convexity structure is required only with respect to the uncertainty. In the framework considered in this paper, we have a joint convexity concerning both control and uncertainty. It is expected that our results can be extended to more general situations beyond the linear framework, provided that the convexity structure pretaining to the uncertainty is maintained. Furthermore, in cases where the problem lacks a convexity structure, it is still possible to prove some sub-differentiability of the probabilistic function. This sub-differentiability may be adequate for deriving a nonsmooth Pontryagin's principle for a broader class of control problems. These questions will be interesting to investigate in the future works.

Funding The authors acknowledge support by the FMJH Program Gaspard Monge in optimization and operations research including support to this program by EDF. The second author thanks the Deutsche Forschungsgemeinschaft for their support within project B04 in the "Sonderforschungsbereich/Transregio 154 Mathematical Modelling, Simulation and Optimization using the Example of Gas Networks". The third author acknowledges support by Normandy Region and the European Union through ERDF research and innovation program, under the grant for "chaire d'excellence COPTI".

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. van Ackooij, W.: A discussion of probability functions and constraints from a variational perspective. Set-Valued Var. Anal. **28**(4), 585–609 (2020)
2. van Ackooij, W., Henrion, R.: Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions. SIAM J. Opt. **24**(4), 1864–1889 (2014)
3. van Ackooij, W., Henrion, R.: (Sub-) Gradient formulae for probability functions of random inequality systems under Gaussian distribution. SIAM J. Uncertain. Quantif. **5**(1), 63–87 (2017)
4. van Ackooij, W., Malick, J.: Eventual convexity of probability constraints with elliptical distributions. Math. Program. **175**(1), 1–27 (2019)
5. van Ackooij, W., Pérez-Aros, P.: Generalized differentiation of probability functions: parameter dependent sets given by intersections of convex sets and complements of convex sets. Appl. Math. Opt. **85**(2), 1–39 (2022)
6. van Ackooij, W., Henrion, R., Möller, A., Zorgati, R.: Joint chance constrained programming for hydro reservoir management. Opt. Eng. **15**, 509–531 (2014)
7. van Ackooij, W., Aleksovska, I., Munoz Zuniga, M.: (sub-)differentiability of probability functions with elliptical distributions. Set Valued Var. Anal. **26**(4), 887–910 (2018)
8. van Ackooij, W., Henrion, R., Pérez-Aros, P.: Generalized gradients for probabilistic/robust (probust) constraints. Optimization **69**(7–8), 1451–1479 (2020)
9. Agrachev, A., Baryshnikov, Y., Sarychev, A.: Ensemble controllability by lie algebraic methods. ESAIM: Cont. Opt. Calc. Var. **22**(4), 921–938 (2016)
10. Assellaou, M., Bokanowski, O., Zidani, H.: Error estimates for second order hamilton-jacobi equations approximations of probabilistic reachable sets. Discrete Contin. Dyn. Syst. Ser. A **35**(9), 3933–3964 (2015)

11. Augier, N., Boscain, U., Sigalotti, M.: Adiabatic ensemble control of a continuum of quantum systems. SIAM J. Control Optim. **56**(6), 4045–4068 (2018)
12. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: *Non-Linear Parametric Optimization*. Birkhäuser Basel (1982)
13. Beauchard, K., Rouchon, J.M. Coron et P.: Controllability issues for continuous-spectrum systems and ensemble controllability of Bloch equations. Comm. Math. Phys. **296**(2), 525–557 (2010)
14. Bremer, I., Henrion, R., Möller, A.: Probabilistic constraints via SQP solver: Application to a renewable energy management problem. Comput. Manag. Sci. **12**, 435–459 (2015)
15. Caillau, J.B., Cerf, M., Sassi, A., Trélat, E., Zidani, H.: Solving chance constrained optimal control problems in aerospace via kernel density estimation. Optim. Control Appl. Methods **39**(5), 1818–1832 (2018)
16. Clarke, F.H.: Functional Analysis. Calculus of Variations and Optimal Control. Springer, Berlin (2013)
17. Curtis, F.E., Wächter, A., Zavala, V.: A sequential algorithm for solving nonlinear optimization problems with chance constraints. SIAM J. Opt. **28**, 930–958 (2018)
18. Dambrine, M., Harbrecht, H., Puig, B.: Incorporating knowledge on the measurement noise in electrical impedance tomography*. ESAIM: COCV **25**, 84 (2019)
19. Farshbaf-Shaker, M.H., Henrion, R., Hömberg, D.: Properties of chance constraints in infinite dimensions with an application to PDE constrained optimization. Set Valued Var. Anal. **26**(4), 821–841 (2018)
20. Farshbaf-Shaker, M.H., Gugat, M.H., Heitsch, H., Henrion, R.: Optimal Neumann boundary control of a vibrating string with uncertain initial data and probabilistic terminal constraints. SIAM J. Cont. Opt. **58**, 2288–2311 (2020)
21. Geiersbach, C., Hintermüller, M.: Optimality conditions and Moreau–Yosida regularization for almost sure state constraints (2021)
22. Geiersbach, C., Scarinci, T.: A stochastic gradient method for a class of nonlinear PDE-constrained optimal control problems under uncertainty (2021)
23. Geiersbach, C., Wollner, W.: Optimality conditions for convex stochastic optimization problems in banach spaces with almost sure state constraints. SIAM J. Opt. **31**(4), 2455–2480 (2021)
24. Geletu, A., Hoffmann, A., Schmidt, P., Li, P.: Chance constrained optimization of elliptic PDE systems with a smoothing convex approximation. *ESAIM: Cont. Opt. Calc. Var.***26**, (2020)
25. Geletu, A., Hoffmann, A., Klöppel, M., Li, P.: An inner-outer approximation approach to chance constrained optimization. SIAM J. Opt. **27**(3), 1834–1857 (2017)
26. Göttlich, S., Kolb, O., Lux, K.: Chance-constrained optimal inflow control in hyperbolic supply systems with uncertain demand. Optim. Control Appl. Methods **42**, 566–589 (2021)
27. Gotzes, C., Heitsch, H., Henrion, R., Schultz, R.: On the quantification of nomination feasibility in stationary gas networks with random load. Math. Methods Oper. Res. **84**, 427–457 (2016)
28. Hantoute, A., Henrion, R., Pérez-Aros, P.: Subdifferential characterization of continuous probability functions under Gaussian distribution. Math. Program. **174**(1–2), 167–194 (2019)
29. Hong, L.J., Yang, Y., Zhang, L.: Sequential convex approximations to joint chance constrained programed: A monte carlo approach. Oper. Res. **3**(59), 617–630 (2011)
30. Kouri, D. P., Staudigl, M., Surowiec, T. M.: A relaxation-based probabilistic approach for pde-constrained optimization under uncertainty with pointwise state constraints. *Comput. Opt. Appl.*, (2023)
31. Laguel, Y., van Ackooij, W., Malick, J., Matiussi Ramalho, G.: On the convexity of level-sets of probability functions. J. Convex Anal. **29**(2), 1–32 (2022)
32. Lasiecka, I., Triggiani, R.: Control Theory for Partial Differential Equations: Volume 1, Abstract Parabolic Systems: Continuous and Approximation Theories. Cambridge University Press, Cambridge (2000)
33. Li, X., Yong, J.: *Optimal Control Theory for Infinite Dimensional Systems*. Birkhäuser (1995)
34. Lions, J.-L., Magenes, E.: Non-homogeneous boundary value problems and applications, vol. 1. Springer-Verlag, New York- Heidelberg (1972)
35. Luedtke, J., Ahmed, S., Nemhauser, G.: An integer programming approach for linear programs with probabilistic constraints. Math. Program. **122**(2), 247–272 (2010)
36. Pazy, A.: Semigroups of linear operators and applications to partial differential equations, vol. 44. Springer-Verlag, New York (1983)
37. Peña-Ordieres, A., Luedtke, J.R., Wächter, A.: Solving chance-constrained problems via a smooth sample-based nonlinear approximation. SIAM J. Opt. **30**(3), 2221–2250 (2020)

38. Periago Esparza, F., Martinez-Frutos, J.: Optimal Control of PDEs under Uncertainty. Springer, Cham (2018)
39. Prékopa, A.: Stochastic programming. Kluwer Academic Publishers, Dordrecht (1995)
40. Raymond, J.P., Zidani, H.: Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. SIAM J. Cont. Opt. **36**(6), 1853–1879 (1998)
41. Raymond, J.P., Zidani, H.: Hamiltonian pontryagin's principles for control problems governed by semilinear parabolic equations. Appl. Math. Opt. **39**(2), 143–177 (1999)
42. Robin, R., Augier, N., Boscain, U., Sigalotti, M.: Ensemble Qubit controllability with a single control via adiabatic and rotating wave approximations. J. Differ. Equ. **318**(5), 414–442 (2022)
43. Schuster, M., Strauch, E., Gugat, M., Lang, J.: Probabilistic constrained optimization on flow networks. Opt. Eng. **23**, 1–50 (2022)
44. Shapiro, A., Dentcheva, D., A., Ruszczyński, A.: *Lectures on Stochastic Programming: Modeling and Theory*. MPS-SIAM series on optimization 9, (2009)
45. Uryas'ev, S.: Derivatives of probability and integral functions: General theory and examples. In: Floudas, C.A., Pardalos, P.M. (eds.) Encyclopedia of Optimization, 2nd edn., pp. 658–663. Springer - Verlag, Berlin (2009)
46. Vinter, R.B.: Optimal Control. Springer, Berlin (2000)