Approximation of risk-averse optimal feedback control

Philipp A. Guth[†] and Karl Kunisch[†,‡]

August 21, 2025

Abstract

The challenge of constructing feedback control laws for risk-averse optimal control of partial differential equations (PDEs) with random coefficients is addressed. The control objective composes a tracking-type cost with the nonlinear entropic risk measure. A sequential quadratic programming scheme is derived that iteratively solves linear quadratic subproblems obtained through second-order Taylor expansions of the objective functional, with each subproblem re-centered at the previous iterate. It is shown that this method converges locally quadratically to the unique risk-averse optimal control. This work provides the first rigorous feedback synthesis for risk-averse objectives subject to PDEs with random coefficients.

Keywords   risk-averse optimal control, feedback control under uncertainty, partial differential equations with random coefficients

MSC 2020   49J20, 49N35, 93B52, 93C20

1   Introduction

Optimization and optimal control of systems with uncertain parameters is a challenging task since already small parameter changes may change the characteristics of the system, such as its stability properties. If neglected, the uncertainty may lead to failure of the control objective and hence a careful treatment of the uncertainty is indispensable.

In uncertainty quantification, partial differential equations (PDEs) with random coefficients have become a popular way of incorporating uncertainty into models for engineering or physical processes. One reason behind their success is the ability to integrate knowledge of governing physical equations while accommodating for randomness, which may reflect, for instance, missing data, measurement uncertainties, material imperfections, or external influences.

If PDEs with random coefficients appear as constraints of an optimal control problem, the objective function becomes a random variable. In order to be able to perform optimization, risk measures are used, which map the random variable objective function to the real line. A widely used risk measure is the expected value, [2, 7, 13, 14, 19, 28, 34, 36]. The expected value is linear, Fréchet differentiable and it belongs to the class of coherent risk measures ([1]). Despite its widespread adoption, decisions based solely on the expected value (called risk-neutral decisions) are inadequate in many practical scenarios, as they fail to account for the inherent variability in uncertain outcomes. In many situations it is important to minimize the likelihood of extremely large cost rather than merely minimizing the expected cost. In fact, in real-world decision-making under uncertainty individuals typically prefer more predictable or less uncertain

[†]Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenbergerstraße 69, AT-4040 Linz, Austria.

[‡]Institute of Mathematics and Scientific Computing, Karl-Franzens University Graz, Heinrichstrasse 36, AT-8010 Graz, Austria.

EMAILS: `philipp.guth@ricam.oeaw.ac.at`, `karl.kunisch@uni-graz.at`

outcomes over those with greater volatility, even if both have the same expected return. This risk averse preferences can be reflected by risk measures which disproportionately penalize outcomes that involve high cost or variability.

Risk measures typically involve integrals over the possibly infinite-dimensional space of uncertain parameters. The significant computational expense associated with these problems has spurred research into the development of efficient numerical algorithms, such as those based on sparse grids [28, 29] or quasi-Monte Carlo (QMC) methods [19, 20, 21, 33]. The entropic risk measure is such a risk averse measure, which can be effectively computed using QMC methods, see [20, 32].

In this paper the application of sequential second-order Taylor approximations for the efficient approximation of a risk-averse feedback control law based on the entropic risk measure is investigated. To do so, given a possibly countably infinite sequence of uncertain parameters $\boldsymbol{\sigma} = (\sigma)_{j \geq 1} \in \mathfrak{S} \subseteq \mathbb{R}^{\mathbb{N}}$, let us consider the parameter-dependent optimal control problem

$$\min_{u,y} \mathcal{J}(u, y), \quad \mathcal{J}(u, y) := \frac{1}{2} \left(\int_0^T \left(\mathcal{R} \left(\|C(y(\cdot; t) - g(t))\|_H^2 \right) + \|u(t)\|_U^2 \right) \mathrm{d}t \right. \tag{1.1}$$
$$\left. + \mathcal{R} \left(\|P(y(\cdot; T) - g_T)\|_H^2 \right) \right),$$

subject to

$$\dot{y}(\boldsymbol{\sigma}; t) = A(\boldsymbol{\sigma}) y(\boldsymbol{\sigma}; t) + Bu(t) + f(t) \quad y(\boldsymbol{\sigma}; 0) = y_0, \tag{1.2}$$

for all $\boldsymbol{\sigma} \in \mathfrak{S}$. Given $T > 0$, the goal is to find a control input $u \in L^2(0, T; U)$ which steers the parameter-dependent state $y(\boldsymbol{\sigma})$ as close as possible to the targets $g \in \mathcal{C}([0, T]; H)$ and $g_T \in H$. The initial condition of the state $y(\boldsymbol{\sigma}; 0) = y_0 \in H$ is known. The time evolution of the system is described by the parameter-dependent operators $A(\boldsymbol{\sigma}) \in \mathcal{L}(V, V')$ acting on the current state, $B \in \mathcal{L}(U, H)$ acting on the input control, and an external forcing $f \in L^2(0, T; V')$. The operator $C \in \mathcal{L}(H)$ is an observation operator and $P \in \mathcal{L}(H)$. In principle, $\mathcal{R}$ can denote any risk measure. However, in this work, we focus on the entropic risk measure since it is smooth and it can be computed effectively for high-dimensional uncertain parameters.

Uncertain parameters Throughout this manuscript we assume that the domain of the uncertain variables is given as

$$\mathfrak{S} := [-1, 1]^{\mathbb{N}}.$$

Furthermore, the components σ_j of the parameter sequence $\boldsymbol{\sigma} = (\sigma_j)_{j \in \mathbb{N}}$ are independently and identically distributed random variables, each being uniformly distributed over $[-1, 1]$. That is, the sequence $\boldsymbol{\sigma}$ is distributed according to the countable product probability measure

$$\mathrm{d}\boldsymbol{\sigma} := \bigotimes_{j \in \mathbb{N}} \frac{\mathrm{d}\sigma_j}{2}.$$

We shall be interested in continuous functions $\boldsymbol{\sigma} \mapsto X(\boldsymbol{\sigma})$ taking values in some Banach space. In this case X is Bochner-integrable and its expected value can be written as infinite-dimensional integral

$$\mathbb{E}[X] = \int_{\mathfrak{S}} X(\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma}. \tag{1.3}$$

Entropic risk measure We shall focus on a particular choice of the risk measure $\mathcal{R}$: the entropic risk measure with risk-aversion parameter $\theta > 0$. For a random variable $X \in L^\infty(\mathfrak{S}; \mathbb{R})$, it is defined as

$$\mathcal{R}_\theta(X) := \frac{1}{\theta} \log\left(\mathbb{E}\left[e^{\theta X}\right]\right).$$

As a function of θ, the entropic risk measure is increasing and strictly increasing if X is not constant (a.s.). Moreover, it can be shown (see, e.g., [26, Thm. 1.3.2]) that

$$\lim_{\theta \to 0} \mathcal{R}_\theta(X) = \mathbb{E}[X] \qquad \text{and} \qquad \lim_{\theta \to \infty} \mathcal{R}_\theta(X) = \operatorname{ess\,sup}(X).$$

Due to its exponential form, the entropic risk measure in (1.1) assigns greater weight to scenarios that involve large tracking cost, thereby encouraging control actions that avoid these risks, even if the control cost might be higher. In large deviation theory the entropic risk measure is also called logarithmic moment generating function [11].

1.1 Notation

Given real numbers $r < s$ and separable Banach spaces $\mathcal{X}$ and $\mathcal{Y}$, the space of continuous functions from $[r, s]$ into $\mathcal{X}$ is denoted by $\mathcal{C}([r, s]; \mathcal{X})$ and the Bochner space of strongly measurable square integrable functions from the interval (r, s) into $\mathcal{X}$ is denoted by $L^2(r, s; \mathcal{X})$ and we also denote the subspace $W(r, s; \mathcal{X}, \mathcal{Y}) := \{v \in L^2(r, s; \mathcal{X}) \mid \dot{v} \in L^2(r, s; \mathcal{Y})\}$. Since the time horizon $T > 0$ will be fixed throughout this manuscript, to shorten the exposition, sometimes we shall denote

$$\mathcal{X}_T := L^2(0, T; \mathcal{X}) \quad \text{and} \quad W_T(\mathcal{X}, \mathcal{Y}) := W(0, T; \mathcal{X}, \mathcal{Y}).$$

By $\mathcal{L}(\mathcal{X}, \mathcal{Y})$ we denote the space of linear continuous mappings from $\mathcal{X}$ into $\mathcal{Y}$, and in case $\mathcal{X} = \mathcal{Y}$ we use the shorter $\mathcal{L}(\mathcal{X}) := \mathcal{L}(\mathcal{X}, \mathcal{X})$.

Throughout this manuscript, boldfaced symbols are used to denote multi-indices while the subscript notation m_j is used to refer to the jth component of a multi-index $\boldsymbol{m}$. Further, let

$$\mathscr{F} := \{\boldsymbol{m} \in \mathbb{N}_0^{\mathbb{N}} \mid |\boldsymbol{m}| < \infty\}$$

denote the set of finitely supported multi-indices, where the order of a multi-index $\boldsymbol{m}$ is defined as $|\boldsymbol{m}| := \sum_{j \geq 1} m_j$. Further, we write $\delta_{\boldsymbol{m}, \boldsymbol{0}} = 1$ if $m_j = 0 \ \forall j \geq 1$ and $\delta_{\boldsymbol{m}, \boldsymbol{0}} = 0$ otherwise.

2 Problem formulation

In order to show existence and uniqueness of a solution to (1.1) subject to (1.2), we make the following assumptions: the parameter-dependent operator $A(\boldsymbol{\sigma})$ in (1.2) can be associated with a continuous and V-H-coercive parameter-dependent bilinear form $a(\boldsymbol{\sigma}; \cdot, \cdot)$, that is

$$\langle A(\boldsymbol{\sigma})v, w\rangle_{V', V} := -a(\boldsymbol{\sigma}; v, w), \qquad \forall v, w \in V, \quad \forall \boldsymbol{\sigma} \in \mathfrak{S}, \tag{2.1}$$

where $\exists (\rho, \theta) \in \mathbb{R} \times (0, \infty)$ such that

$$a(\boldsymbol{\sigma}; v, v) + \rho\|v\|_H^2 \geq \theta\|v\|_V^2 \quad \forall v \in V \quad \text{and} \quad \forall \boldsymbol{\sigma} \in \mathfrak{S}. \tag{2.2}$$

In addition we assume that the operators have a uniform upper bound, that is

$$\|A(\boldsymbol{\sigma})\|_{\mathcal{L}(V, V')} \leq C_D \quad \forall \boldsymbol{\sigma} \in \mathfrak{S}. \tag{2.3}$$

For every $\boldsymbol{\sigma} \in \mathfrak{S}$, let us define the parameterized parabolic evolution operator $D(\boldsymbol{\sigma})$ as

$$\langle D(\boldsymbol{\sigma})w, (v_1, v_2)\rangle_{V_T' \times H, V_T \times H} := \langle \dot{w}, v_1\rangle_{V_T', V_T} - \langle A(\boldsymbol{\sigma})w, v_1\rangle_{V_T', V_T} + \langle w(0), v_2\rangle_H \tag{2.4}$$

for all $w \in W_T(V, V')$ and all $v = (v_1, v_2) \in V_T \times H$, where we recall that $V_T = L^2(0, T; V)$. The parameterized family of operators $\{D(\boldsymbol{\sigma}) \in \mathcal{L}(W_T(V, V'), V_T' \times H) : \boldsymbol{\sigma} \in \mathfrak{S}\}$ has uniformly bounded inverses, see, e.g., [21, 30].

Throughout the manuscript it will be assumed that

$$A : \mathfrak{S} \to \mathcal{L}(V, V'), \quad \boldsymbol{\sigma} \mapsto A(\boldsymbol{\sigma}) \quad \text{is continuous,}$$

which leads to the following regularity result for solutions of (1.2).

Lemma 2.1. *Given $u \in U_T$, $y_0 \in H$, and $f \in V_T'$, there is a unique solution $y \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$ of* (1.2).

Proof. Under the assumption that $\boldsymbol{\sigma} \mapsto A(\boldsymbol{\sigma}) \in \mathcal{L}(V, V')$ is continuous, the parametric mapping to the evolution operator $\boldsymbol{\sigma} \mapsto D(\boldsymbol{\sigma}) \in \mathcal{L}(W_T(V, V'), L_T^2(V') \times H)$, as defined in (2.4), is continuous. Then, with (2.2) and (2.3) it follows that $D(\boldsymbol{\sigma})$ is boundedly invertible for all $\boldsymbol{\sigma} \in \mathfrak{S}$, and hence from [15, Thm. 1.1.1] we conclude that $\boldsymbol{\sigma} \mapsto y \in W_T(V, V')$ is continuous. From the continuous embedding of $W_T(V, V') \hookrightarrow \mathcal{C}([0, T]; H)$ we have $y \in \mathcal{C}(\mathfrak{S}; \mathcal{C}([0, T]; H))$, and in particular $y \in C(\mathfrak{S} \times [0, T]; H)$, and consequently $y(\cdot; t, \cdot) \in \mathcal{C}(\mathfrak{S}; H)$ for every $t \in [0, T]$. $\square$

By substituting $y(\boldsymbol{\sigma}; t) = D(\boldsymbol{\sigma})^{-1}(Bu(t) + f(t), y_0)$ in (1.1) one obtains the reduced formulation of the optimal control problem $\min_u J(u)$, which only depends on the control input u. In the remainder of this section we will prove the existence and uniqueness of a minimizer of the reduced (and equivalently the original) problem and characterize it by a necessary and sufficient optimality condition.

It will sometimes be convenient to write (1.2) in the equivalent form

$$D(\boldsymbol{\sigma})y(\boldsymbol{\sigma}) = (D_1(\boldsymbol{\sigma})y(\boldsymbol{\sigma}), D_2(\boldsymbol{\sigma})y(\boldsymbol{\sigma})) = (Bu + f, y_0) \quad \text{in } V_T' \times H,$$

where we define

$$D_1(\boldsymbol{\sigma}) := \Lambda_1 D(\boldsymbol{\sigma}), \quad \Lambda_1 : V_T' \times H \to V_T', \quad \Lambda_1(v_1, v_2) := v_1, \quad \text{for all } v = (v_1, v_2) \in V_T' \times H,$$
$$D_2(\boldsymbol{\sigma}) := \Lambda_2 D(\boldsymbol{\sigma}), \quad \Lambda_2 : V_T' \times H \to H, \quad \Lambda_2(v_1, v_2) := v_2, \quad \text{for all } v = (v_1, v_2) \in V_T' \times H.$$

Meaningful (right-)inverse operators of $D_1(\boldsymbol{\sigma}) : W_T(V, V') \to V_T'$ and $D_2(\boldsymbol{\sigma}) : W_T(V, V') \to H$ are given by

$$D_1^\dagger(\boldsymbol{\sigma}) := D(\boldsymbol{\sigma})^{-1}\Xi_1, \quad \Xi_1 : V_T' \to V_T' \times H, \quad \Xi_1 v_1 := (v_1, 0), \quad \text{for all } v_1 \in V_T',$$
$$D_2^\dagger(\boldsymbol{\sigma}) := D(\boldsymbol{\sigma})^{-1}\Xi_2, \quad \Xi_2 : H \to V_T' \times H, \quad \Xi_2 v_2 := (0, v_2), \quad \text{for all } v_2 \in H.$$

It follows that, for all $\boldsymbol{\sigma} \in \mathfrak{S}$, the unique solution of (1.2) can be written as

$$y(\boldsymbol{\sigma}) = D(\boldsymbol{\sigma})^{-1}(Bu + f, y_0) = D_1^\dagger(\boldsymbol{\sigma})(Bu + f) + D_2^\dagger(\boldsymbol{\sigma})y_0 \quad \text{in } W_T(V, V').$$

By substituting this into the objective function (1.1), we arrive at the reduced problem

$$\min_u J(u), \quad J(u) := \mathcal{J}(u, D(\boldsymbol{\sigma})^{-1}(Bu + f, y_0)).$$

We show next that this reduced problem (and hence the original one) has a unique solution.

Existence of solutions We have that $0 \leq J(u)$, thus there is a sequence $(u_k)_{k \in \mathbb{N}}$ such that $\lim_{k \to \infty} J(u_k) = \inf_u J(u)$. Since $J(u)$ is coercive ($J(u) \geq \frac{1}{2} \int_0^T \|u\|_U^2 \, dt$), the infimizing sequence is bounded. The sequence takes values in $L^2(0, T; U)$ and thus it has a weakly convergent subsequence $u_\ell \rightharpoonup \bar{u}$. Since J is convex (see also next paragraph) and continuous, it is weakly lower semicontinuous and we have

$$J(\bar{u}) \leq \liminf_{\ell \to \infty} J(u_\ell) = \inf_u J(u).$$

Thus, $J(\bar{u}) = \inf_u J(u)$ and $\bar{u}$ is the sought minimizer.

Uniqueness of the solution First, observe that the two terms in the cost functional that involve the entropic risk measure, $\mathcal{R}_\theta(\|C(y(\cdot;t)-g(t))\|_H^2$ and $\mathcal{R}_\theta(\|P(y(\cdot;T)-g_T)\|_H^2$, are well-defined for solutions $y \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$ of (1.2). Given $\theta \geq 0$, for every fixed $t \in [0,T]$, the functional

$$u \mapsto \mathcal{R}_\theta(\|C(y(\cdot;t)-g(t))\|_H^2)$$

is convex. This follows from the fact, that $E_t : W_T(V,V') \to H$, which evaluates functions pointwise, is a bounded linear operator, and from the linear dependence of $y(\boldsymbol{\sigma})$ on u, the convexity of $z \mapsto \|z\|_H^2$, as well as the convexity and the monotonicity of $\mathcal{R}_\theta$. By the same reasoning the convexity of

$$u \mapsto \mathcal{R}_\theta(\|P(y(\cdot;T)-g_T)\|_H^2)$$

follows. Furthermore, we have that

$$u \mapsto \int_0^T \mathcal{R}_\theta(\|C(y(\cdot;t)-g(t))\|_H^2)\,\mathrm{d}t$$

is convex. The control cost $u \mapsto \frac{1}{2}\int_0^T \|u(t)\|_U^2\,\mathrm{d}t$ is strongly convex. Thus, $J(u)$ is strongly convex since it is the positive linear combination of two convex functions and a strongly convex function. Hence, the solution is unique.

Optimality condition Since the problem is strongly convex, first-order optimality conditions are necessary and sufficient. Thus, we shall be interested in the gradient of the reduced objective function

$$J(u) = \frac{1}{2}\left(\int_0^T \left(\frac{1}{\theta}\log\left(\mathbb{E}\left[\mathrm{e}^{\theta\|C(D(\cdot)^{-1}((Bu(t)+f(t)),y_0)-g(t))\|_H^2}\right]\right) + \|u(t)\|_U^2\right)\mathrm{d}t\right.$$
$$\left.+ \frac{1}{\theta}\log\left(\mathbb{E}\left[\mathrm{e}^{\theta\|P(E_T D(\cdot)^{-1}((Bu(t)+f(t)),y_0)-g_T)\|_H^2}\right]\right)\right).$$

Given $f \in V_T'$ and $y_0 \in H$, let $y_u(\boldsymbol{\sigma})$ denote the solution to (1.2) for a fixed $\boldsymbol{\sigma} \in \mathfrak{S}$, and a given input control u. Then, we have that $y_{u+\delta}(\boldsymbol{\sigma}) - y_u(\boldsymbol{\sigma}) = D(\boldsymbol{\sigma})^{-1}(B(u+\delta)+f,y_0) - D(\boldsymbol{\sigma})^{-1}(Bu+f,y_0) = D(\boldsymbol{\sigma})^{-1}(B\delta,0) = D_1^\dagger(\boldsymbol{\sigma})B\delta$. Using this result, the Fréchet differentiability of J at $u \in U_T = L^2(0,T;U)$ follows from the differentiability of both the entropic risk measure [24] and the squared norm, in combination with the chain rule and the Lebesgue dominated convergence theorem.

The gradient $\nabla J(u) \in U_T$ is identified with the Riesz representer of the Fréchet derivative at $u \in U_T$, that is $\frac{\partial}{\partial u}J(u)\delta = \langle \nabla J(u),\delta\rangle_{U_T}$ for all directions $\delta \in U_T$, and given by

$$
\nabla J(u) = u + \frac{\mathbb{E}\left[\mathrm{e}^{\theta\|C(y(\cdot)-g\|_H^2}B^*(D_1^\dagger(\cdot))^*\,(C^*C(y(\cdot)-g)\right]}{\mathbb{E}\left[\mathrm{e}^{\theta\|C(y(\cdot)-g\|_H^2}\right]}
$$
$$
+ \frac{\mathbb{E}\left[\mathrm{e}^{\theta\|P(E_T y(\cdot)-g_T)\|_H^2}B^*(D_1^\dagger(\cdot))^*\,(E_T^* P^* P(E_T y(\cdot)-g_T))\right]}{\mathbb{E}\left[\mathrm{e}^{\theta\|P(E_T y(\cdot)-g_T)\|_H^2}\right]},
$$

(2.5)

where E_T^* denotes the adjoint of the operator E_T.

3   Quadratic approximation of the entropic risk measure

Next we explore a quadratic approximation of the entropic risk measure leading to a linear quadratic optimal control problem. This approximation forms the basis of a sequential quadratic programming (SQP) method, which is employed to iteratively approximate the solution of the original optimal control problem involving the nonlinear entropic risk measure, that is (1.1) subject to (1.2).

To rigorously establish the Fréchet derivatives required for the quadratic approximation, we first revisit the regularity properties of the state y. Under the conditions of Lemma 2.1 we have $y \in L^2(\mathfrak{S}; V_T)$ and $\dot{y} \in L^2(\mathfrak{S}; V_T')$. Moreover, V_T as well as $L^2(\mathfrak{S})$ are separable Hilbert spaces, and hence it holds that $L^2(\mathfrak{S}; V_T) = L^2(\mathfrak{S}) \otimes V_T$, where $\otimes$ denotes the Hilbert tensor product, where equality (or canonical identification) is up to a unique isometric isomorphism, see [35, Thm. II.10]. Similarly, we have $V_T = L^2(0, T) \otimes V$ since both $L^2(0, T)$ and V are separable. Thus, we have

$$
\begin{aligned}
L^2(\mathfrak{S}; V_T) &= L^2(\mathfrak{S}) \otimes (L^2(0, T) \otimes V) = (L^2(\mathfrak{S}) \otimes L^2(0, T)) \otimes V \\
&= L^2(\mathfrak{S} \times [0, T]) \otimes V = (L^2(0, T) \otimes L^2(\mathfrak{S})) \otimes V \\
&= L^2(0, T) \otimes (L^2(\mathfrak{S}) \otimes V) = L^2(0, T; L^2(\mathfrak{S}; V)),
\end{aligned}
\tag{3.1}
$$

where we used the associativity (up to unique isometric ismorphism) of the tensor product, see [27, Prop. 2.6.5.]. Analogously, we have $\dot{y} \in L^2(0, T; L^2(\mathfrak{S}; V'))$, and consequently $y \in W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))$.

To compute Fréchet derivatives, we define

$$
Z := \{(u, y) \in U_T \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V')) \mid \dot{y} = \mathcal{A}y + Bu + f, \, y(0) = y_0\},
$$

where $\mathcal{A} \in \mathcal{L}(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))$ is the extension of the pointwise operator family $\{A(\boldsymbol{\sigma}) \in \mathcal{L}(V, V') \mid \boldsymbol{\sigma} \in \mathfrak{S}\}$ to the Lebesgue–Bochner space as $\mathcal{A}v := [\boldsymbol{\sigma} \mapsto A(\boldsymbol{\sigma})v(\boldsymbol{\sigma})]$. The dynamical system in the definition of Z is posed in $L^2(0, T; L^2(\mathfrak{S}; V'))$, which is identified with $L^2(\mathfrak{S}; V_T')$ by (3.1), with V_T replaced by V_T'. Thus, the dynamics in Z can be recast as equation in V_T' as follows:

$$
\dot{y}(\boldsymbol{\sigma}) = A(\boldsymbol{\sigma})y(\boldsymbol{\sigma}) + Bu + f, \quad y(\boldsymbol{\sigma}; 0, \cdot) = y_0
\tag{3.2}
$$

for almost every $\boldsymbol{\sigma} \in \mathfrak{S}$. On the other hand, given f, u, and y_0, the system (3.2) has a unique solution $\tilde{y}(\boldsymbol{\sigma}) \in W_T(V, V')$ for every $\boldsymbol{\sigma} \in \mathfrak{S}$, which by Lemma 2.1 admits the additional regularity $\tilde{y} \in \mathcal{C}(\mathfrak{S}; \mathcal{C}([0, T]; H))$. For the solution y to (3.2) we find that $y = \tilde{y}$ up to $\mathrm{d}\boldsymbol{\sigma}$-almost everywhere equivalence. We have thus shown the following result:

Lemma 3.1. *Given $f \in V_T'$ and $y_0 \in H$, the state component of tuples $(u, y) \in Z$ admits a representative $y \in \mathcal{C}(\mathfrak{S}; \mathcal{C}([0, T]; H))$.*

Hence, in the following, we shall make use of $y \in \mathcal{C}(\mathfrak{S}; \mathcal{C}([0, T]; H))$ for elements $y \in Z$ wherever continuity or pointwise evaluation is required.

Let $\bar{y} \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$, for example let $\bar{y}$ be the solution to (1.2) corresponding to some

control $\bar{u} \in U_T$. A quadratic approximation about $\bar{y}$ of the cost functional (1.1), is of the form

$$
\begin{aligned}
&\mathcal{J}_{\text{quad}}(u,y) \\
&= \frac{1}{2}\bigg(\int_0^T \Big(\|u(t)\|_U^2 + \mathcal{R}_\theta(\|C(\bar{y}(\cdot;t) - g(t))\|_H^2) \\
&\quad + \Big\langle \frac{\partial}{\partial y}\mathcal{R}_\theta(\|C(y(\cdot;t)) - g(t))\|_H^2)\big|_{y=\bar{y}(\cdot;t)}, y(\cdot;t) - \bar{y}(\cdot;t) \Big\rangle_{L^2(\mathfrak{S};H)} \\
&\quad + \frac{1}{2}\Big[\frac{\partial^2}{\partial y^2}\mathcal{R}_\theta(\|C(y(\cdot;t) - g(t))\|_H^2)\big|_{y=\bar{y}(\cdot;t)} \Big] (y(\cdot;t) - \bar{y}(\cdot;t), y(\cdot;t) - \bar{y}(\cdot;t)) \Big)\mathrm{d}t \qquad (3.3) \\
&\quad + \mathcal{R}_\theta(\|P(E_T\bar{y}(\cdot;t) - g_T)\|_H^2) \\
&\quad + \int_0^T \Big(\Big\langle \frac{\partial}{\partial y}\mathcal{R}_\theta(\|P(E_T y(\cdot;t) - g_T)\|_H^2)\big|_{y=\bar{y}(\cdot;t)}, y(\cdot;t) - \bar{y}(\cdot;t) \Big\rangle_{L^2(\mathfrak{S};H)} \\
&\quad + \frac{1}{2}\Big[\frac{\partial^2}{\partial y^2}\mathcal{R}_\theta(\|P(E_T y(\cdot;t) - g_T)\|_H^2)\big|_{y=\bar{y}(\cdot;t)} \Big] (y(\cdot;t) - \bar{y}(\cdot;t), y(\cdot;t) - \bar{y}(\cdot;t)) \Big)\mathrm{d}t \bigg),
\end{aligned}
$$

for $(u,y) \in U_T \times \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$.

For the remainder of the manuscript it will be useful to define

$$
\mathscr{C}(\boldsymbol{\sigma};t) := C^*C(y(\boldsymbol{\sigma};t) - g(t)) \quad \text{and} \quad \mathscr{P}(\boldsymbol{\sigma};T) := P^*P(y(\boldsymbol{\sigma};T) - g_T),
$$

as well as the weight functions

$$
\omega_{\theta,C}(y(\boldsymbol{\sigma};t)) := \frac{\mathrm{e}^{\theta\|C(y(\boldsymbol{\sigma};t) - g(t))\|_H^2}}{\mathbb{E}\big[\mathrm{e}^{\theta\|C(y(\cdot;t) - g(t))\|_H^2}\big]}, \qquad \text{and} \qquad \omega_{\theta,P}(y(\boldsymbol{\sigma};T)) = \frac{\mathrm{e}^{\theta\|P(y(\boldsymbol{\sigma};T) - g_T)\|_H^2}}{\mathbb{E}\big[\mathrm{e}^{\theta\|P(y(\cdot;T) - g_T)\|_H^2}\big]},
$$

which are well-defined for $y \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$. Furthermore, for an integrable element f taking values in a separable Banach space, we define the weighted expectations

$$
\mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))}[f] := \mathbb{E}[f(\cdot)\,\omega_{\theta,C}(y(\cdot;t))] \quad \text{and} \quad \mathbb{E}_{\omega_{\theta,P}(y(\boldsymbol{\sigma};T))}[f] := \mathbb{E}[f(\cdot)\,\omega_{\theta,P}(y(\cdot;T))].
$$

Using this notation, the first-order Fréchet derivatives at $y \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$ in direction $\delta_1 \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$ can be expressed as

$$
\begin{aligned}
&\int_0^T \Big\langle \frac{\partial}{\partial y}\mathcal{R}_\theta(\|C(y(\cdot,t) - g(t))\|_H^2), \delta_1(\cdot;t) \Big\rangle_{L^2(\mathfrak{S};H)} \mathrm{d}t \\
&\qquad = 2\int_0^T \frac{\mathbb{E}[\mathrm{e}^{\theta\|C(y(\cdot;t) - g(t))\|_H^2}\langle C^*C(y(\cdot;t) - g(t)), \delta_1(\cdot;t)\rangle_H]}{\mathbb{E}[\mathrm{e}^{\theta\|C(y(\cdot;t) - g(t))\|_H^2}]}\,\mathrm{d}t \\
&\qquad = 2\int_0^T \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))}[\langle\mathscr{C}(\cdot;t), \delta_1(\cdot;t)\rangle_H]\,\mathrm{d}t \\
&\int_0^T \Big\langle \frac{\partial}{\partial y}\mathcal{R}_\theta(\|P(E_T y(\cdot;t) - g_T)\|_H^2), \delta_1(\cdot;t) \Big\rangle_{L^2(\mathfrak{S};H)} \mathrm{d}t \\
&\qquad = 2\int_0^T \frac{\mathbb{E}[\mathrm{e}^{\theta\|P(E_T y(\cdot;t) - g_T)\|_H^2}\langle E_T^* P^* P(E_T y(\cdot;t) - g_T), \delta_1(\cdot;t)\rangle_H]}{\mathbb{E}[\mathrm{e}^{\theta\|P(E_T y(\cdot;t) - g_T)\|_H^2}]}\,\mathrm{d}t \\
&\qquad = 2\,\mathbb{E}_{\omega_{\theta,P}(y(\boldsymbol{\sigma};T))}[\langle\mathscr{P}(\cdot;T), \delta_1(\cdot;T)\rangle_H].
\end{aligned}
$$

This follows again from the chain rule and Lebesgue's dominated convergence theorem. We point out that both derivatives can be extended to bounded linear functionals on $L^2(0,T;L^2(\mathfrak{S};H))$. By Lemma 2.1 these derivatives are well-defined on Z, which will be used in Section 5 to prove that the solution of the quadratic approximation converges to the solution of the original

problem. Similarly, evaluated on Z, the second-order Fréchet derivatives in (3.3) are identified as the bounded bilinear forms

$$\int_0^T \left[\frac{\partial^2}{\partial y^2} \mathcal{R}_\theta(\|C(y(\cdot;t) - g(t))\|_H^2) \right] (\delta_2(\cdot;t), \delta_1(\cdot;t)) \, \mathrm{d}t$$

$$= \int_0^T 2 \, \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \left[\langle C^*C\delta_2(\cdot;t), \delta_1(\cdot;t) \rangle_H \right]$$

$$+ 4\theta \, \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \left[\langle \mathscr{C}(\cdot;t), \delta_2(\cdot;t) \rangle_H \langle \mathscr{C}(\cdot;t), \delta_1(\cdot;t) \rangle_H \right]$$

$$- 4\theta \, \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \left[\langle \mathscr{C}(\cdot;t), \delta_1(\cdot;t) \rangle_H \right] \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \left[\langle \mathscr{C}(\cdot;t), \delta_2(\cdot;t) \rangle_H \right] \mathrm{d}t$$

$$= \int_0^T 2 \, \mathbb{E}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \left[\langle C^*C\delta_2(\cdot;t), \delta_1(\cdot;t) \rangle_H \right]$$

$$+ 4\theta \, \mathrm{Cov}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \Big(\langle \mathscr{C}(\cdot;t), \delta_2(\cdot;t) \rangle_H, \langle \mathscr{C}(\cdot;t), \delta_1(\cdot;t) \rangle_H \Big) \mathrm{d}t,$$

for $\delta_1 \in L^2(0,T;L^2(\mathfrak{S};H))$ and $\delta_2 \in L^2(0,T;L^2(\mathfrak{S};H))$, and analogously

$$\int_0^T \left[\frac{\partial^2}{\partial y^2} \mathcal{R}_\theta(\|P(E_T y(\cdot;t) - g_T)\|_H^2) \right] (\delta_2(\cdot;t), \delta_1(\cdot;t)) \, \mathrm{d}t$$

$$= 2 \, \mathbb{E}_{\omega_{\theta,P}(y(\boldsymbol{\sigma};T))} \left[\langle P^*P\delta_2(\cdot;T), \delta_1(\cdot;T) \rangle_H \right]$$

$$+ 4\theta \, \mathrm{Cov}_{\omega_{\theta,P}(y(\boldsymbol{\sigma};T))} \Big(\langle \mathscr{P}(\cdot;T), \delta_2(\cdot;T) \rangle_H, \langle \mathscr{P}(\cdot;T), \delta_1(\cdot;T) \rangle_H \Big),$$

As a consequence, if $\delta_1 = \delta_2$, the second derivative is nonnegative and strictly positive if additionally $\mathscr{C}$ or $\mathscr{P}$ is not a constant (almost surely). Moreover, the second Fréchet derivatives are locally Lipschitz continuous with respect to $y \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$, which will be used in Section 5.

Remark 3.2. Several remarks are in order:

- We point out that the second-order Fréchet derivative differs from the risk neutral case not only by the additional covariance terms, but also the expected values are weighted by the normalized exponential functions $\omega_{\theta,C}(y(\boldsymbol{\sigma};t))$ and $\omega_{\theta,P}(y(\boldsymbol{\sigma};t))$. These weights reflect the risk averse control decision of the entropic risk measure by assigning exponentially more importance to realizations of the parameter sequence $\boldsymbol{\sigma} \in \mathfrak{S}$ which lead to large tracking cost. The exponential reweighting of a probability measure to make rare events more likely is known in large deviation theory ([5, Ch. 3.1]) and rare event simulation [37] as exponential tilting or exponential twisting.

- The weighted expected values are related to expected values with respect to the posterior measure in Bayesian inverse problems [38]. Specifically, in Bayesian inverse problems with forward model $\mathcal{G}$ and centered additive Gaussian noise on the data $Y_{\mathrm{data}} = G(\xi) + \eta$, $\eta \sim \mathcal{N}(0,\Gamma)$, the prior belief on the unknown parameter ξ is updated by a likelihood of the form $\exp(-\|G(\xi) - Y_{\mathrm{data}}\|_\Gamma^2) / \int \exp(-\|G(\xi) - Y_{\mathrm{data}}\|_\Gamma^2) \, \mathrm{d}\mathcal{N}(0,\Gamma)(\xi)$ resulting in a posterior probability distribution of the parameter given the data. This structural similarity between Bayesian inverse problems and optimal control problems with entropic risk measure has been used, e.g., in [32].

3.1 Quadratic approximation leads to risk-adjusted linear quadratic problem

The second-order Fréchet derivatives at $\bar{y} \in \mathcal{C}([0,T];\mathcal{C}(\mathfrak{S};H))$ are associated with the operators $\mathcal{Q}_{C,\omega}(\bar{y};t) \in \mathcal{L}(L^2(\mathfrak{S};H))$ and $\mathcal{Q}_{P,\omega}(\bar{y};T) \in \mathcal{L}(L^2(\mathfrak{S};H))$ as

$$\langle \mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t)\delta_1, \delta_2 \rangle_{L^2(\mathfrak{S};H)} = \mathbb{E}_{\omega_{\theta,C}(\bar{y}(\boldsymbol{\sigma};t))} \left[\langle C^*C\delta_2(\cdot;t), \delta_1(\cdot;t) \rangle_H \right]$$

$$+ 2\theta \, \mathrm{Cov}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))} \Big(\langle \overline{\mathscr{C}}(\cdot;t), \delta_2(\cdot;t) \rangle_H, \langle \overline{\mathscr{C}}(\cdot;t), \delta_1(\cdot;t) \rangle_H \Big)$$

8

for almost every $t \in [0, T]$ and

$$
\begin{aligned}
\langle \mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T) \delta_1, \delta_2 \rangle_{L^2(\mathfrak{S}; H)} = {} & \mathbb{E}_{\omega_{\theta, P}(\bar{y}(\boldsymbol{\sigma}; T))} \left[\langle P^* P \delta_2(\cdot; T), \delta_1(\cdot; T) \rangle_H \right] \\
& + 2\theta \operatorname{Cov}_{\omega_{\theta, P}(y(\boldsymbol{\sigma}; T))} \Big(\langle \overline{\mathscr{P}}(\cdot; T), \delta_2(\cdot; T) \rangle_H, \langle \overline{\mathscr{P}}(\cdot; T), \delta_1(\cdot; T) \rangle_H \Big),
\end{aligned}
$$

where we use $\overline{\mathscr{C}}(\boldsymbol{\sigma}; t) := C^* C(\bar{y}(\boldsymbol{\sigma}; t) - g(t))$ and $\overline{\mathscr{P}}(\boldsymbol{\sigma}; T) := P^* P(\bar{y}(\boldsymbol{\sigma}; T) - g_T)$. Furthermore, we include the factor $\frac{1}{2}$ for the second-order derivatives in the Taylor expansion (3.3). These two operators sum up to the second-order Fréchet derivative of the (pointwise in time) tracking error composed with the entropic risk measure, that is

$$
\langle L(\bar{y}(\cdot; t); t, T) \delta_1, \delta_2 \rangle_{L^2(\mathfrak{S}; H)} := \langle (\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t) + \mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T)) \delta_1, \delta_2 \rangle_{L^2(\mathfrak{S}; H)} \tag{3.4}
$$

for almost every $t \in [0, T]$ as well as for all directions $\delta_1 \in L^2(0, T; L^2(\mathfrak{S}; H))$ and $\delta_2 \in L^2(0, T; L^2(\mathfrak{S}; H))$. Given $\bar{y} \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$, for almost every $t \in [0, T]$, the operator $L(\bar{y}(\cdot; t); t, T) \in \mathcal{L}(L^2(\mathfrak{S}; H))$ is nonnegative and self-adjoint.

Let us assume that P and C are not both zero (otherwise (1.1)–(1.2) admits only the trivial solution). Moreover, we assume that there exist positive constants c_C and c_P such that $\|Cx\|_H \geq c_C \|x\|_H$ if $C \neq 0$ and $\|Px\|_H \geq c_P \|x\|_H$ for all $x \in H$ if $P \neq 0$. Then both operators $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)$ for almost every $t \in [0, T]$ and $\mathcal{Q}_{T,\omega}(\bar{y}(\cdot; T); T)$ are coercive (or zero), and thus also $L(\bar{y}(\boldsymbol{\sigma}; t); t, T)$ is coercive. Hence, for almost every $t \in [0, T]$ and any $\bar{y} \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$, there exist unique positive self-adjoint operators $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)^{1/2}$ and $\mathcal{Q}_{T,\omega}(\bar{y}(\cdot; T); T)^{1/2}$.

Next, for $\bar{y} \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$, we set

$$
\mathscr{R}_C(\bar{y}(\boldsymbol{\sigma}; t); t) := \overline{\mathscr{C}}(\cdot; t) \omega_{\theta, C}(\bar{y}(\boldsymbol{\sigma}; t)), \quad \text{and} \quad \mathscr{R}_P(\bar{y}(\boldsymbol{\sigma}; T); T) := \overline{\mathscr{P}}(\cdot; T) \omega_{\theta, P}(\bar{y}(\boldsymbol{\sigma}; T)),
$$

so that the quadratic approximation (3.3) of the cost functional (1.1) can be written as

$$
\begin{aligned}
\mathcal{J}_{\text{quad}}(u, y) = \frac{1}{2} \Bigg(& \mathcal{R}_\theta(\|P(\bar{y}(\cdot; T) - g_T)\|_H^2) + 2 \mathbb{E} \left[\langle \mathscr{R}_P(\bar{y}(\boldsymbol{\sigma}; T); T), y(\cdot; T) - \bar{y}(\cdot; T) \rangle_H \right] \\
& + \langle \mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T)(y(\cdot; T) - \bar{y}(\cdot; T)), (y(\cdot; T) - \bar{y}(\cdot; T)) \rangle_{L^2(\mathfrak{S}; H)} \\
& + \int_0^T \Big(\mathcal{R}_\theta(\|C(\bar{y}(\cdot; t) - g(t))\|_H^2) + 2 \mathbb{E} \left[\langle \mathscr{R}_C(\bar{y}(\boldsymbol{\sigma}; t); t), y(\cdot; t) - \bar{y}(\cdot; t) \rangle_H \right] \\
& + \langle \mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)(y(\cdot; t) - \bar{y}(\cdot; t)), (y(\cdot; t) - \bar{y}(\cdot; t)) \rangle_{L^2(\mathfrak{S}; H)} + \|u(t)\|_U^2 \Big) \mathrm{d}t \Bigg).
\end{aligned}
$$

To reformulate the objective function as a sum of quadratic terms only, we substitute $z_C(\boldsymbol{\sigma}; t) = \mathcal{Q}_{C,\omega}(\bar{y}(\boldsymbol{\sigma}; t); t)^{1/2}(y(\boldsymbol{\sigma}; t) - \bar{y}(\boldsymbol{\sigma}; t))$ and $z_P(\boldsymbol{\sigma}; T) = \mathcal{Q}_{P,\omega}(\bar{y}(\boldsymbol{\sigma}; T); T)^{1/2}(y(\boldsymbol{\sigma}; T) - \bar{y}(\boldsymbol{\sigma}; T))$, lead-

ing to

$$\mathcal{J}_{\text{quad}}(u, z)$$

$$= \frac{1}{2}\bigg(\mathcal{R}_\theta(\|P(\bar{y}(\cdot; T) - g_T)\|_H^2) + 2\,\mathbb{E}\left[\langle(\mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T)^{-1/2})^*\mathscr{R}_P(\bar{y}(\cdot; T); T), z_P(\cdot; T)\rangle_H\right]$$

$$\quad + \mathbb{E}\left[\|z_P(\cdot; T)\|_H^2\right] + \int_0^T \bigg(\|u(t)\|_U^2 + \mathcal{R}_\theta(\|C(\bar{y}(\cdot; t) - g(t))\|_H^2)$$

$$\quad + 2\,\mathbb{E}\left[\langle(\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)^{-1/2})^*\mathscr{R}_C(\bar{y}(\cdot; t); t), z_C(\cdot; t)\rangle_H\right] + \mathbb{E}\left[\|z_C(\cdot; t)\|_H^2\right]\bigg)\mathrm{d}t\bigg)$$

$$= \frac{1}{2}\bigg(\mathcal{R}_\theta(\|P(\bar{y}(\cdot; T) - g_T)\|_H^2) + \mathbb{E}\left[\|z_P(\cdot; T) + (\mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T)^{-1/2})^*\mathscr{R}_P(\bar{y}(\cdot; T); T)\|_H^2\right]$$

$$\quad - \mathbb{E}\left[\|(\mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T); T)^{-1/2})^*\mathscr{R}_P(\bar{y}(\cdot; T); T)\|_H^2\right] + \int_0^T \bigg(\|u(t)\|_U^2 + \mathcal{R}_\theta(\|C(\bar{y}(\cdot; t) - g(t))\|_H^2)$$

$$\quad + \mathbb{E}\left[\|z_C(\cdot; t) + (\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)^{-1/2})^*\mathscr{R}_C(\bar{y}(\cdot; t); t)\|_H^2\right]$$

$$\quad - \mathbb{E}\left[\|(\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)^{-1/2})^*\mathscr{R}_C(\bar{y}(\cdot; t); t)\|_H^2\right]\bigg)\mathrm{d}t\bigg).$$

In the following we will leave out the terms which do not depend on u or y since they have no effect on the minimizer of $\mathcal{J}_{\text{quad}}$. This leads to the linear quadratic optimal control problem $\min_{u,y} J_{\text{quad}}$, subject to (1.2), where

$$J_{\text{quad}}(u, y) = \frac{1}{2}\bigg(\int_0^T \bigg(\mathbb{E}\left[\|\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t))^{1/2}(y(\cdot; t) - \widetilde{g}(\cdot; t))\|_H^2\right] + \|u(t)\|_U^2\bigg)\mathrm{d}t \tag{3.5}$$

$$+ \mathbb{E}\left[\|\mathcal{Q}_{P,\omega}(\bar{y}(\cdot; T))^{1/2}(y(\cdot; T) - \widetilde{g}_T(\cdot))\|_H^2\right]\bigg)$$

with target $\widetilde{g}(\boldsymbol{\sigma}; t) = \bar{y}(\boldsymbol{\sigma}; t) - \mathcal{Q}_{C,\omega}(\bar{y}(\boldsymbol{\sigma}; t))^{-1}\mathscr{R}_C(\bar{y}(\boldsymbol{\sigma}; t); t)$ and terminal target $\widetilde{g}_T(\boldsymbol{\sigma}) = \bar{y}(\boldsymbol{\sigma}; T) - \mathcal{Q}_{P,\omega}(\bar{y}(\boldsymbol{\sigma}; T); T)^{-1}\mathscr{R}_P(\bar{y}(\boldsymbol{\sigma}; T); T)$. For $\theta \to 0$, the risk-adjusted quadratic objective functional (3.5) coincides with the risk-neutral case, that is with (1.1) with $\mathcal{R} = \mathbb{E}$.

4 Optimal feedback law of the quadratic approximation

We shall be interested in an optimal feedback law $\mathscr{K}(\cdot; t) : L^2(\mathfrak{S}; H) \to U$ for the linear quadratic probem $\min_{u,y} J_{\text{quad}}(u, y)$ subject to (1.2). It will turn out that this feedback law is affine, that is

$$\mathscr{K}(\cdot; t) = -B^*(\Pi(T - t)(\cdot) + h(t)),$$

where Π solves a Riccati equation. For our main result, we will adapt [4, Part IV, Ch. 2, Thm. 7.3] to the setting presented in this work. To this end, we recall the extension of the pointwise operator $A(\boldsymbol{\sigma}) \in \mathcal{L}(V, V')$ for $\boldsymbol{\sigma} \in \mathfrak{S}$ to Lebegue–Bochner spaces given by

$$\mathcal{A} \in \mathcal{L}(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V')), \quad \mathcal{A}v := [\boldsymbol{\sigma} \mapsto A(\boldsymbol{\sigma})v(\boldsymbol{\sigma})].$$

Secondly, we introduce the cone $\Omega(L^2(\mathfrak{S}; H))$ of bounded, linear, self-adjoint, and nonnegative operators in $L^2(\mathfrak{S}; H)$ endowed with the norm of $\mathcal{L}(L^2(\mathfrak{S}; H))$. The Riccati operators will be sought as strongly continuous operator-valued functions in the set $\mathcal{S} := \mathcal{C}_s([0, T]; \Omega(L^2(\mathfrak{S}; H)))$, which is endowed with the topology of strong convergence, i.e., $F_n \to F$ holds if and only if for all $x \in L^2(\mathfrak{S}; H)$ it holds that $F_n(\cdot)x \to F(\cdot)x$ in $\mathcal{C}([0, T]; L^2(\mathfrak{S}; H))$, see e.g., [4, Part IV, Ch. 2.1].

Thirdly, we verify that $t \mapsto \mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t)x \in \mathcal{C}([0,T];L^2(\mathfrak{S};H))$ for any $x \in L^2(\mathfrak{S};H)$. Indeed, for $x \in L^2(\mathfrak{S};H)$, we have

$$\mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t)x = \frac{\mathrm{e}^{\theta\|C(\bar{y}(\cdot;t)-g(t))\|_H^2}}{\mathbb{E}\left[\mathrm{e}^{\theta\|C(\bar{y}(\cdot;t)-g(t))\|_H^2}\right]}\left(2\,C^*Cx + 4\theta\,\overline{\mathscr{C}}(\cdot;t)\langle\overline{\mathscr{C}}(\cdot;t),x\rangle_H\right.$$
$$\left. - 4\theta\,\overline{\mathscr{C}}(\cdot;t)\mathbb{E}_{\omega_{\theta,C}(\bar{y}(\boldsymbol{\sigma};t))}\left[\langle\overline{\mathscr{C}}(\cdot;t),x\rangle_H\right]\right) \in \mathcal{C}([0,T];L^2(\mathfrak{S};H)).$$

Finally, the above mentioned result [4, Part IV, Ch. 2, Thm. 7.3] is stated for the case $\widetilde{g} = \widetilde{g}_T = 0$. A generalization to tracking problems with nontrivial targets can be found, e.g., in [31, Ch. III, Eqn. (4.88)]. The latter result can readily be generalized to nontrivial terminal penalizations (as in [4, Part IV, Ch. 2, Thm. 7.3]). Thus, we arrive at the following result.

Theorem 4.1. *The problem $\min_{u,y} J_{\mathrm{quad}}(u,y)$ subject to (1.2) has a unique optimal pair (u,y), and the optimal control $u \in \mathcal{C}([0,T];U)$ is related to the optimal state by the feedback formula*

$$u(t) = -B^*(\Pi(T-t)y(\cdot;t) + h(t)),$$

for $t \in [0,T]$, where $\Pi \in \mathcal{S}$ and $h \in W_T(L^2(\mathfrak{S};V);L^2(\mathfrak{S};V'))$ solve

$$-\dot{\Pi}(T-t) = \Pi(T-t)\mathcal{A} + \mathcal{A}^*\Pi(T-t) - \Pi(T-t)BB^*\Pi(T-t) + \mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t),$$
$$-\dot{h}(t) = (\mathcal{A}^* - \Pi(T-t)BB^*)h(t) + \Pi(T-t)f(t) - \mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t);t)\widetilde{g}(\cdot;t;\bar{y}(\cdot;t)),$$

with $\Pi(0) = \mathcal{Q}_{P,\omega}(\bar{y}(\cdot;T);T)$ and $h(T) = -\mathcal{Q}_{P,\omega}(\bar{y}(\cdot;T);T)\widetilde{g}_T(\cdot,\bar{y}(\cdot;T))$.

5 Convergence of the quadratic approximation

Recall that the unique optimal pair of $\min_{u,y} J_{\mathrm{quad}}$ subject to (1.2) coincides with the unique optimal pair of $\min_{u,y} \mathcal{J}_{\mathrm{quad}}$ subject to (1.2) since the two cost functionals differ only by a term which is independent of both the control u and state y.

We shall show next, that the sequence of minimizers $\{(u_\star^{(k)}, y_\star^{(k)})\}_{k\geq 0}$ generated by repeatedly solving the linear quadratic problem with updated expansion points $\bar{y}^{(k)} = y_\star^{(k-1)}$ converges with second order to the unique minimizer of the original problem (1.1)–(1.2) provided that the initial guess is sufficiently close to the minimizer, as will be made precise below.

To this end, we show that the optimality system of the quadratic approximation $\min_{u,y} \mathcal{J}_{\mathrm{quad}}$ subject to (1.2) is a Newton step for the original problem (1.1)–(1.2). Thus, repeatedly solving the quadratic approximation with updated expansion points results in a so-called sequential quadratic programming (SQP) method for solving the original problem. The SQP method is based on the Lagrangian

$$\mathscr{L}(u,y,q) = \mathcal{J}(u,y) + \langle q, e(u,y)\rangle_{L^2(0,T;L^2(\mathfrak{S};V))\times L^2(\mathfrak{S};H),L^2(0,T;L^2(\mathfrak{S};V'))\times L^2(\mathfrak{S};H)}.$$

where we abbreviated $e(u,y) = D(\cdot)y(\cdot) - (Bu + f, y_0)$. Let $q_\star = q(y_\star)$ denote the Lagrange multiplier at the solution $y_\star$ of (1.1)–(1.2), then a first-order necessary optimality condition is

$$\mathscr{L}'(u_\star, y_\star, q_\star) = 0, \quad e(u,y) = 0, \tag{5.1}$$

where $\mathscr{L}'$ denotes the derivative of $\mathscr{L}$ with respect to (u,y), that is

$$\mathscr{L}'(u,y,q) = \begin{bmatrix} e_u(u,y)^*q + \mathcal{J}_u(u,y) \\ e_y(u,y)^*q + \mathcal{J}_y(u,y) \end{bmatrix}, \quad \text{with} \quad e_u = \frac{\partial}{\partial u}e(u,y), \quad e_y = \frac{\partial}{\partial y}e(u,y).$$

11

The SQP method consists in a Newton method applied to the necessary optimality condition (5.1) in order to iteratively solve for the solution $(u_\star, y_\star, q_\star)$ of the original problem (1.1)–(1.2). Each Newton step results in the following linear system for the updates:

$$\begin{bmatrix} \mathscr{L}''(u^{(k)}, y^{(k)}, q^{(k)}) & e'(u^{(k)}, y^{(k)})^* \\ e'(u^{(k)}, y^{(k)}) & 0 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} u^{(k+1)} - u^{(k)} \\ y^{(k+1)} - y^{(k)} \end{bmatrix} \\ q^{(k+1)} - q^{(k)} \end{bmatrix} = - \begin{bmatrix} e_u(u^{(k)}, y^{(k)})^* q^{(k)} + \mathcal{J}_u(u^{(k)}, y^{(k)}) \\ e_y(u^{(k)}, y^{(k)})^* q^{(k)} + \mathcal{J}_y(u^{(k)}, y^{(k)}) \\ e(u^{(k)}, y^{(k)}) \end{bmatrix}.$$

A sufficient second-order condition for optimality of $(u_\star, y_\star, q_\star)$ is

$$\mathscr{L}''(u_\star, y_\star, q_\star)((v, x), (v, x)) \geq \kappa \|(v, x)\|^2_{U_T \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))}, \tag{5.2}$$

for $\kappa > 0$, and for all (v, x) in the kernel of $e'(u_\star, y_\star)$, which is given by $\{(v, x) \in U_T \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))) \mid \dot{x} = \mathcal{A}x + Bv, \ x(0) = 0\}$. By the linearity of the constraint $e(u, y)$ we have $e_{uu} = e_{uy} = e_{yu} = e_{yy} = 0$, and thus $\mathscr{L}''(u, y, q) = \mathcal{J}''(u, y)$. Further, from Section (3.1) we know that $\mathcal{J}''(u, y)((v, x), (v, x)) \geq \|v\|^2_{U_T}$. With the bound $\|x\|_{W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))} \leq C\|u\|_{U_T}$ for all (v, x) in the kernel of $e'(u_\star, y_\star)$, we further estimate $\mathcal{J}''(u, y)((v, x), (v, x)) \geq \frac{1}{2}\|v\|^2_{U_T} + \frac{1}{2C^2}\|x\|^2_{W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))}$. Thus, the sufficient condition (5.2) is satisfied for $\kappa = \min\{\frac{1}{2}, \frac{1}{2C^2}\}$. Moreover, the Newton step simplifies to

$$\begin{bmatrix} \mathbf{1}_U & 0 & e_u(u^{(k)}, y^{(k)})^* \\ 0 & \mathcal{J}_{yy}(u^{(k)}, y^{(k)}) & e_y(u^{(k)}, y^{(k)})^* \\ e_u(u^{(k)}, y^{(k)}) & e_y(u^{(k)}, y^{(k)}) & 0 \end{bmatrix} \begin{bmatrix} \begin{bmatrix} u^{(k+1)} - u^{(k)} \\ y^{(k+1)} - y^{(k)} \end{bmatrix} \\ q^{(k+1)} \end{bmatrix} = - \begin{bmatrix} \mathcal{J}_u(u^{(k)}, y^{(k)}) \\ \mathcal{J}_y(u^{(k)}, y^{(k)}) \\ e(u^{(k)}, y^{(k)}) \end{bmatrix},$$

which is the optimality system of (3.3) subject to (1.2). Consequently, in each Newton step one can equivalently solve the quadratic subproblem (3.3) (or (3.5)) subject to (1.2) with expansion trajectory $\bar{y} = y^{(k)}$.

In order to ensure that the Fréchet derivatives are well-defined at all iterates, we recall from Section 3 that this is ensured for pairs $(u, y) \in Z$, i.e., pairs with additional state regularity $y \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$. Thus, we choose an initial trajectory $\bar{y}^{(-1)} \in \mathcal{C}([0, T]; \mathcal{C}(\mathfrak{S}; H))$. From the Newton step above it is evident that $e(u_{k+1}, y_{k+1}) = 0$, and thus $(u_{k+1}, y_{k+1}) \in Z$, i.e., the iterates remain in Z. We next show that Z is a closed (affine) subspace of $X := L^2(0, T; U) \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))$: clearly $Z \subset X$. To verify that Z is an affine subspace, let

$$Z_0 := \{(u, y) \in L^2(0, T; U) \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V')) \mid \dot{y} - \mathcal{A}y - Bu = 0, \ y(0) = 0\}.$$

It is readily verified that Z_0 is a linear subspace of X. Furthermore, for arbitrary $f \in L^2(0, T; V')$ and $y_0 \in H$, there exists a pair (v_0, x_0) such that $\dot{x}_0 - \mathcal{A}x_0 - Bv_0 - f = 0$ with $x_0(0) = y_0$. Furthermore, any pair $(v, x) \in Z$ can be expressed as a translation $(v_0, x_0) + (\tilde{v}_0, \tilde{x}_0)$, where $(\tilde{v}_0, \tilde{x}_0) \in Z_0$, since $\dot{x} - \dot{x}_0 - \mathcal{A}(x - x_0) - B(v - v_0) = 0$, and $x(0) - x_0(0) = 0$. Thus, $Z = (v_0, x_0) + Z_0$ is an affine subspace of X. We will next show that Z is closed in X. For $n \in \mathbb{N}$ let $(u_n, y_n) \in Z$ be a sequence such that $(u_n, y_n) \to (u, y) \in U_T \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))$. Then, it holds that $Bu_n \to Bu$ since B is a bounded linear operator, $\dot{y}_n \to \dot{y}$, and $\mathcal{A}y_n \to \mathcal{A}y$ by continuity of $\mathcal{A}$. Thus, taking the limit of $\dot{y}_n(\boldsymbol{\sigma}) = \mathcal{A}y_n(\boldsymbol{\sigma}) + Bu_n + f$, $y_n(\boldsymbol{\sigma}; 0, \cdot) = y_0$, we arrive at

$$\dot{\bar{y}}(t) = \mathcal{A}\bar{y}(t) + B\bar{u}(t) + f(t), \quad \text{for a.e. } t \in [0, T], \quad \text{and} \quad \bar{y}(\cdot; 0, \cdot) = y_0,$$

that is, Z is closed. Further, $\mathcal{J}$ and e are smooth with locally Lipschitz continuous second derivatives, and $e'(u_\star, y_\star)$ is surjective. Under these conditions it is known (see, [25, Ch. 5.3]) that the SQP method is locally quadratic convergent to the unique global minimizer of (1.1)–(1.2), that is

$$\|(u^{(k+1)}, y^{(k+1)}, q^{(k+1)}) - (u_\star, y_\star, q_\star)\|_W \leq K\|(u^{(k)}, y^{(k)}, q^{(k)}) - (u_\star, y_\star, q_\star)\|^2_W,$$

for a constant K independent of k, provided that $\|(u^{(0)}, y^{(0)}, q^{(0)}) - (u_\star, y_\star, q_\star)\|_W$ with $W = U_T \times W_T(L^2(\mathfrak{S}; V), L^2(\mathfrak{S}; V'))) \times L^2(0, T; L^2(\mathfrak{S}; V)) \times L^2(\mathfrak{S}; H)$ is sufficiently small.

Algorithm 1. SQP Algorithm

Require: Tolerance $\varepsilon > 0$. Initial expansion trajectory $\bar{y}^{(-1)} \in \mathcal{C}([0,T]; \mathcal{C}(\mathfrak{S}; H))$.

Ensure: Approximate minimizer $(u_\star, y_\star)$ of (1.1)–(1.2) with control in feedback form.

1: Set $k \leftarrow 0$ and $y_\star^{(k-1)} \leftarrow \bar{y}^{(-1)}$.
2: **repeat**
3:      Given the expansion point $\bar{y} = y_\star^{(k-1)}$, solve the equations in Theorem 4.1 to find the optimal Riccati-based feedback control $u_\star^{(k)}$ with corresponding closed-loop state $y_\star^{(k)}$ of (3.5) s.t. (1.2).
4:      $k \leftarrow k + 1$.
5: **until** $\|\nabla J(u_\star^{(k)})\|_{U_T} < \varepsilon$.
6: **return** $(u_\star^{(k)}, y_\star^{(k)}) \approx (u_\star, y_\star)$, where the control is given in feedback form.

The optimal control that is computed using Algorithm 1 is in feedback form for the linear quadratic approximation of the original problem (1.1) subject to (1.2) evaluated at the optimal trajectory.

6   Discretization and numerical implementation

This section is devoted to the numerical implementation of the dynamical systems appearing in Theorem 4.1.

6.1   Numerical integration

The expected values involved in the presented problem are given as integrals over an infinite-dimensional parameter space $\mathfrak{S}$, see (1.3). For the numerical implementation we first truncate the infinite-dimensional parameter sequence $\boldsymbol{\sigma}$ to an s-dimensional vector $\boldsymbol{\sigma}_s$ by setting all components with index larger than s equal to zero. In the follwoing we will thus consider $\boldsymbol{\sigma} \approx (\boldsymbol{\sigma}_s, \mathbf{0}) = (\sigma_1, \sigma_2, \ldots, \sigma_s, 0, 0, \ldots)$. This allows to replace the infinite-dimensional integrals by integrals over the s-dimensional parameter space $\mathfrak{S} := [-1,1]^s$.

The numerical approximation of these integrals using full tensorgrids of univariate quadrature rules suffers from the curse of dimensionality, i.e., the number of nodes required to guarantee a prescribed error tolerance increases exponentially in the dimension s. For sufficiently smooth problems this issue can be alleviated using sparse grids [6], or quasi-Monte Carlo methods [12]. In this work we shall focus on sample average approximations of the respective integrals. This includes Monte Carlo as well as quasi-Monte Carlo methods and the presented results can straightforwardly be modified to other quadrature methods.

6.2   Polynomial chaos expansion

Let $L_n(\xi)$ denote the Legendre polynomial of degree $n \geq 0$ in $[-1,1]$, which is normalized such that

$$\int_{-1}^{1} |L_n(\xi)|^2 \frac{\mathrm{d}\xi}{2} = 1$$

Then we have $L_0 = 1$ and $\{L_n\}_{n \geq 0}$ is an orthonormal basis of $L^2(-1, 1)$. For $\boldsymbol{\nu} \in \mathscr{F}$ we introduce the tensorized Legendre polynomials

$$L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) = \prod_{j \geq 1} L_{\nu_j}(\sigma_j).$$

The tensorized Legendre polynomials $\{L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) \mid \boldsymbol{\nu} \in \mathscr{F}\}$ form a Riesz basis, i.e. a dense, orthonormal family in $L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma})$; in particular, each $\delta \in L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma}; H)$ admits an orthogonal

expansion

$$\delta(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\nu} \in \mathscr{F}} \delta_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), \quad \text{where} \quad \delta_{\boldsymbol{\nu}} = \int_{\mathfrak{S}} \delta(\boldsymbol{\sigma}) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma} \in H$$

In this manuscript we shall be interested in functions in $\delta \in L^2(0, T; L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma}; H))$, thus for almost every $t \in [0, T]$ the corresponding coefficients $\{\delta_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathscr{F}}$ are elements in H.

6.2.1 Stochastic Galerkin approximation of a parametric dynamical system

We wish to approximate the solution $y(\boldsymbol{\sigma}; t)$ of (1.2) using a (truncated) generalized polynomial chaos expansion

$$y(\boldsymbol{\sigma}; t) \approx \sum_{\boldsymbol{\nu} \in \Lambda} y_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), \tag{6.1}$$

where $\Lambda := \{\boldsymbol{\nu} \in \mathscr{F} : |\boldsymbol{\nu}| \leq p\}$ is a finite index set, and $L_{\boldsymbol{\nu}}(\boldsymbol{\sigma})$ are the tensorized orthonormal Legendre polynomials. The bound p on the cardinality of the multi-indices in Λ equals the highest degree of the polynomials, and hence the total number of coefficients in a truncated expansion of the form (6.1) is $K + 1 = \frac{(s+p)!}{s!p!}$. Thus, we can number the indices as $\Lambda = \{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \ldots, \boldsymbol{\nu}_{K+1}\}$. Furthermore, for almost every $t \in [0, T]$ the coefficients $\{y_{\boldsymbol{\nu}}(t)\}_{\boldsymbol{\nu} \in \mathscr{F}}$ are elements in H.

General Case Substituting the polynomial expansion into the equation (1.2) yields

$$\sum_{\boldsymbol{\nu} \in \Lambda} \dot{y}_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) = A(\boldsymbol{\sigma}) \sum_{\boldsymbol{\nu} \in \Lambda} y_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) + Bu(t) + f(t).$$

Next we project both sides onto the basis functions $\{L_{\boldsymbol{m}}\}_{\boldsymbol{m} \in \Lambda}$ by taking the inner product in $L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma})$

$$\left\langle \sum_{\boldsymbol{\nu} \in \Lambda} \dot{y}_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \right\rangle = \left\langle A(\boldsymbol{\sigma}) \sum_{\boldsymbol{\nu} \in \Lambda} y_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \right\rangle + \left\langle Bu(t) + f(t), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \right\rangle,$$

for $\boldsymbol{m} \in \Lambda$. Using the orthonormality of the basis, this simplifies to

$$\dot{y}_{\boldsymbol{m}}(t) = \sum_{\boldsymbol{\nu} \in \Lambda} A^{(\boldsymbol{\nu}, \boldsymbol{m})} y_{\boldsymbol{\nu}}(t) + \delta_{\boldsymbol{m}, \mathbf{0}}(Bu(t) + f(t)), \quad \boldsymbol{m} \in \Lambda,$$

where the operator-valued coefficients are defined as

$$A^{(\boldsymbol{\nu}, \boldsymbol{m})} := \langle A(\boldsymbol{\sigma}) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \rangle_{L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma})} \in \mathcal{L}(V, V'),$$

and

$$\delta_{\boldsymbol{m}, \mathbf{0}}(Bu(t) + f(t)) = \langle Bu(t) + f(t), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \rangle_{L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma})} = \begin{cases} Bu(t) + f(t), & \text{if } \boldsymbol{m} = \mathbf{0}, \\ 0, & \text{otherwise.} \end{cases},$$

since B, u, and f are deterministic. Denoting $A^{\mathrm{GPC}} := \left(A^{(\boldsymbol{\nu}, \boldsymbol{m})}\right)_{\boldsymbol{m}, \boldsymbol{\nu} \in \Lambda}$, the Galerkin system becomes the fully coupled system

$$\dot{\mathbf{y}}(t) = A^{\mathrm{GPC}} \mathbf{y}(t) + (Bu(t) + f(t)) \otimes e_0,$$

where $\mathbf{y}(t) = [y_{\boldsymbol{\nu}_1}(t), \ldots, y_{\boldsymbol{\nu}_{K+1}}(t)]^\top \in H^{K+1} = \times_{j=0}^{K} H$ is the stacked coefficient vector, and $(Bu(t) + f(t)) \otimes e_0 = [Bu(t) + f(t), 0, \ldots, 0]^\top$ is the projection onto the first polynomial.

Affine parameter dependence In many situations, such as in diffusion or reaction problems in which the random input field is parameterized in terms of a Karhunen–Loève expansion, the operator $A(\boldsymbol{\sigma})$ depends affinely on the parameters, that is

$$A(\boldsymbol{\sigma}) = A_0 + \sum_{j=1}^{s} \sigma_j A_j,$$

where $A_j \in \mathcal{L}(V, V')$ for $j = 0, \ldots, s$, for some $s \in \mathbb{N}$. Substituting this into the Galerkin projection and using orthonormality, we get

$$\dot{y}_{\boldsymbol{m}}(t) = A_0 y_{\boldsymbol{\nu}}(t) + \sum_{j=1}^{s} \sum_{\boldsymbol{\nu} \in \Lambda} A_j y_{\boldsymbol{\nu}}(t) \langle \sigma_j L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \rangle + \delta_{\boldsymbol{m},\mathbf{0}}(Bu(t) + f(t)), \quad \boldsymbol{m} \in \Lambda.$$

Since the basis functions are tensorized polynomials, and we use a product measure, we can factor the (s-dimensional) integral over the parameter domain and use the orthonormality

$$[M_j]_{\boldsymbol{\nu},\boldsymbol{\mu}} := \langle \sigma_j L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \rangle_{L^2(\mathfrak{S};\mathrm{d}\boldsymbol{\sigma})} = \prod_{i=1, i \neq j}^{s} \delta_{\nu_i, m_i} \cdot \int_{-1}^{1} \sigma_j L_{\nu_j}(\sigma_j) L_{m_j}(\sigma_j) \frac{\mathrm{d}\sigma_j}{2},$$

and thus only univariate integrals need to be computed. Then, the full Galerkin system becomes

$$\dot{\mathbf{y}}(t) = \left(A_0 \otimes I + \sum_{j=1}^{s} A_j \otimes M_j \right) \mathbf{y}(t) + (Bu(t) + f(t)) \otimes e_0,$$

where $\otimes$ denotes the tensor product.

6.3 Choice of the control operator B

The control input $u \in U_T$ is modeled as a deterministic function, that is, it is constant as a function of $\boldsymbol{\sigma} \in \mathfrak{S}$. The polynomial chaos expansion of the control takes the form

$$u(t) = u(\boldsymbol{\sigma}; t) = \sum_{\boldsymbol{\nu} \in \mathscr{F}} u_{\boldsymbol{\nu}}(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), \quad \text{where} \quad \begin{cases} u_{\boldsymbol{\nu}}(t) = u(t) \in U_T & \text{if } \boldsymbol{\nu} = \mathbf{0}, \\ u_{\boldsymbol{\nu}}(t) = 0 & \text{otherwise.} \end{cases}$$

Thus, in the (truncated) stochastic Galerkin basis the control operator $B : U \to H^{K+1}$ can be chosen to act only on the first polynomial, that is, the zeroth-order coefficients

$$\left[Bu_{\boldsymbol{\nu}_1}(t), Bu_{\boldsymbol{\nu}_2}(t), \ldots, Bu_{\boldsymbol{\nu}_{K+1}}(t) \right] = \left[Bu_{\boldsymbol{\nu}_1}(t), 0, \ldots, 0 \right] = \left[Bu(t), 0, \ldots, 0 \right],$$

where $\boldsymbol{\nu}_1 = \mathbf{0}$. Below, by slight abuse of notation we use the same symbol for the control operator $B : U \to H$, its extension to the Bochner space $B : U \to L^2(\mathfrak{S}; H)$, and the Galerkin discretization of the latter $B : U \to H^{K+1}$.

We will now show, that the adjoint operator of B naturally involves the expected value. Indeed, for $v \in L^2(\mathfrak{S}; H)$, we have

$$\langle u, B^* v \rangle_U = \langle Bu, v \rangle_{L^2(\mathfrak{S};H)} = \int_{\mathfrak{S}} \langle Bu, v(\boldsymbol{\sigma}) \rangle_H \, \mathrm{d}\boldsymbol{\sigma} = \int_{\mathfrak{S}} \langle u, B^* v(\boldsymbol{\sigma}) \rangle_H \, \mathrm{d}\boldsymbol{\sigma}$$

$$= \langle u, \int_{\mathfrak{S}} B^* v(\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma} \rangle_U = \langle u, B^* \int_{\mathfrak{S}} v(\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma} \rangle_U.$$

With this choice of the control operator, the optimal control of a linear quadratic problem in the polynomial chaos basis is then implemented as

$$u(t) = -B^*\Pi(T - t)y(\boldsymbol{\sigma}; t) \approx -\sum_{\boldsymbol{\nu} \in \Lambda} B^*\Pi_{\boldsymbol{\nu}}(T - t)y_{\boldsymbol{\nu}}(t)L_{\boldsymbol{\nu}}(\boldsymbol{\sigma})$$

$$= -B^*\left(\sum_{j=1}^{K+1} (\Pi_{\boldsymbol{\nu}_1, \boldsymbol{\nu}_j}(T - t)y_{\boldsymbol{\nu}_j}(t)\right) L_{\boldsymbol{\nu}_1}(\boldsymbol{\sigma}), \tag{6.2}$$

where for $\boldsymbol{\nu} \in \Lambda$ and all $t \in [0, T]$, we use the notation

$$\Pi_{\boldsymbol{\nu}}(T - t) = \begin{bmatrix} \Pi_{\boldsymbol{\nu}_1, \boldsymbol{\nu}_1} & \Pi_{\boldsymbol{\nu}_1, \boldsymbol{\nu}_2} & \cdots & \Pi_{\boldsymbol{\nu}_1, \boldsymbol{\nu}_{K+1}} \\ \Pi_{\boldsymbol{\nu}_2, \boldsymbol{\nu}_1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \Pi_{\boldsymbol{\nu}_{K+1}, \boldsymbol{\nu}_1} & \Pi_{\boldsymbol{\nu}_{K+1}, \boldsymbol{\nu}_2} & \cdots & \Pi_{\boldsymbol{\nu}_{K+1}, \boldsymbol{\nu}_{K+1}} \end{bmatrix} (T - t) \in \mathcal{L}(H^{K+1}, H^{K+1}).$$

The polynomial expansion of the feedback control in (6.2) contains only zeroth-order polynomials, if $\boldsymbol{\nu}_1 = \mathbf{0}$. Thus, the optimal control is a deterministic function, that is, it is constant as a function of the random parameters $\boldsymbol{\sigma} \in \mathfrak{S}$.

Related work Suppose there is a true but unknown parameter $\boldsymbol{\sigma} = \bar{\boldsymbol{\sigma}} \in \mathfrak{S}$. In this situation one can apply the following version of the above feedback: using the same B and $\{\Pi_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \Lambda}$, we project the state trajectory onto the first index $\boldsymbol{\nu}_1$ corresponding to the zeroth-order polynomials. For $t \in [0, T]$ and a given finite set of multi-indices $\Lambda \subset \mathscr{F}$, the resulting feedback takes the form $\{-B^*\Pi_{\boldsymbol{\nu}}(T - t)\mathcal{E}\}_{\boldsymbol{\nu} \in \Lambda} = -B^*\Pi_{\boldsymbol{\nu}_1, \boldsymbol{\nu}_1}(T - t)$, where $\mathcal{E} : H \to H^{K+1}$ maps $y \mapsto \begin{bmatrix} y^\top & 0 & \cdots & 0 \end{bmatrix}^\top$. This feedback law has been constructed and investigated in [23, 22] based on snapshots/samples, whereas in this manuscript we use a stochastic Galerkin ansatz.

6.4 Choice of the expansion point $\bar{y}$ and approximation of the weights

The expansion point $\bar{y}$, around which the entropic risk objective functional is approximated, is updated according to Algorithm 1. The initialization of Algorithm 1 can be constructed as follows. In practical applications, when observational data are available, $\bar{y}^{(-1)}$ may be constructed from filtered or regularized measurements, yielding a data-informed reference that incorporates empirical knowledge of the state. In the absence of data one might start, for instance, with the uncontrolled trajectories.

In every iteration of Algorithm 1, the weight functions are updated by evaluation at the current expansion point $\bar{y}$, which is given as a polynomial chaos surrogate (cp. (6.1)), that is

$$\bar{y}(\boldsymbol{\sigma}; t) = \sum_{\boldsymbol{\nu} \in \mathscr{F}} \bar{y}_{\boldsymbol{\nu}}(t)L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) \approx \sum_{\boldsymbol{\nu} \in \Lambda} \bar{y}_{\boldsymbol{\nu}}(t)L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}),$$

for a finite set of multi-indices $\Lambda \subset \mathscr{F}$.

We select a set of quadrature nodes $\{\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(N)}\}$, leading, for $t \in [0, T]$, to an ensemble of expansion trajectories $\{\bar{y}(\boldsymbol{\sigma}^{(1)}; t), \dots, \bar{y}(\boldsymbol{\sigma}^{(N)}; t)\}$, and to a collection of weights stored in a vector $\boldsymbol{\omega}$ of length N with components

$$\omega_i(t) := \omega_{\theta, C}(\bar{y}(\boldsymbol{\sigma}^{(i)}; t)) = \frac{e^{\theta\|C(\bar{y}(\boldsymbol{\sigma}^{(i)}; t) - g(t))\|_H^2}}{\frac{1}{\tilde{N}} \sum_{k=1}^{\tilde{N}} \left[e^{\theta\|C(\bar{y}(\boldsymbol{\sigma}^{(k)}; t) - g(t))\|_H^2} \right]}, \tag{6.3}$$

with a possibly different set of quadrature nodes $\{\boldsymbol{\sigma}^{(1)}, \dots, \boldsymbol{\sigma}^{(\tilde{N})}\}$, $\tilde{N} \in \mathbb{N}$. We point out that each node is an s-dimensional vector, that is $\boldsymbol{\sigma}^{(i)} \in [-1, 1]^s$ for all $1 \leq i \leq N$. The weights $\omega_{\theta, P}(\bar{y}(\boldsymbol{\sigma}^{(i)}; T))$ are computed analogously.

6.5 Approximation of the weighted covariance

Given $t \in [0,T]$ and a polynomial chaos expansion of the test functions $\delta^X(t) \in L^2(\mathfrak{S}; H)$ and $\delta^Y(t) \in L^2(\mathfrak{S}; H)$, the weighted sample average approximation of the covariance term in the bilinear form associated to the operator $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t); t)$ is of the form

$$\mathrm{Cov}_{\omega_{\theta,C}(y(\boldsymbol{\sigma};t))}\Big(\langle \overline{\mathscr{C}}(\cdot;t), \delta^X(\cdot;t)\rangle_H, \langle \overline{\mathscr{C}}(\cdot;t), \delta^Y(\cdot;t)\rangle_H\Big)$$

$$\approx \frac{1}{N-1} \sum_{i=1}^{N} \Bigg(\omega_i(t)\bigg(X_i(t) - \frac{1}{N}\sum_{j=1}^{N}\omega_j(t)X_j(t)\bigg)\bigg(Y_i(t) - \frac{1}{N}\sum_{j=1}^{N}\omega_j(t)Y_j(t)\bigg)\Bigg),$$

where, for $t \in [0,T]$, $X(t)$ and $Y(t)$ are two vectors of length N, which contain

$$X_i(t) := \sum_{\boldsymbol{m},\boldsymbol{\nu}\in\mathscr{F}} \langle (C^*C\bar{y}_{\boldsymbol{\nu}}(t) - \delta_{\boldsymbol{\nu},\mathbf{0}}g(t))L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)}), \delta_{\boldsymbol{m}}^X(t)L_{\boldsymbol{m}}(\boldsymbol{\sigma}^{(i)})\rangle_H$$

$$Y_i(t) := \sum_{\boldsymbol{m},\boldsymbol{\nu}\in\mathscr{F}} \langle (C^*C\bar{y}_{\boldsymbol{\nu}}(t) - \delta_{\boldsymbol{\nu},\mathbf{0}}g(t))L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)}), \delta_{\boldsymbol{m}}^Y(t)L_{\boldsymbol{m}}(\boldsymbol{\sigma}^{(i)})\rangle_H$$

for $1 \le i \le N$. In the remainder of this section, we shall omit the dependence on t for better readability whenever it is clear from the context. The (unbiased) weighted empirical covariance for fixed $t \in [0,T]$ admits a matrix representation, that is

$$\langle \mathfrak{Cov}(\omega)X, Y\rangle_{\mathbb{R}^N} = \frac{1}{N-1}\sum_{i=1}^{N}\Bigg(\omega_i\bigg(X_i - \frac{1}{N}\sum_{j=1}^{N}\omega_j X_j\bigg)\bigg(Y_i - \frac{1}{N}\sum_{j=1}^{N}\omega_j Y_j\bigg)\Bigg).$$

It is readily verified that this matrix representation is given by

$$\mathfrak{Cov}(\omega) := \frac{1}{N(N-1)}\big[c_{ij}\big] \quad \text{with} \quad c_{ij} := \begin{cases} N\omega_i - \omega_i^2 & \text{if } i = j, \\ -\omega_i\omega_j & \text{otherwise.} \end{cases}$$

Using this matrix representation, we will next derive a matrix representation of the operator associated with the covariance term in $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t); t)$.

Approximation of the operator $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot;t); t)$ For the numerical implementation, a spatial discretization, for instance by a finite element method (FEM), is required. For functions $v \in H$ we suppose a representation of the form $v = \sum_{i=1}^{d_{\mathrm{FEM}}} v_i\phi_i$ such that $\langle v, \tilde{v}\rangle_H = \boldsymbol{v}^\top M\tilde{\boldsymbol{v}}$, for $v \in H$ and $\tilde{v} \in H$ with $M_{ij} = \langle \phi_i, \phi_j\rangle_H$, i.e., $M \in \mathbb{R}^{d_{\mathrm{FEM}}\times d_{\mathrm{FEM}}}$.

Without loss of generality we assume that $\boldsymbol{\nu}$ and $\boldsymbol{m}$ both belong to Λ, and that $b_{\boldsymbol{m}}^X$ and $b_{\boldsymbol{m}}^Y$ depend on the same multi-index $\boldsymbol{m}$. Then, defining the FEM coefficient vectors as $\boldsymbol{a}_{\boldsymbol{\nu}_k} := \big[C^*C\bar{y}_{\boldsymbol{\nu}_k}(t) - \delta_{\boldsymbol{\nu}_k,\mathbf{0}}g(t)\big] \in \mathbb{R}^{d_{\mathrm{FEM}}}$ and $b_{\boldsymbol{m}_k}^X := \big[\delta_{\boldsymbol{m}_k}^X(t)\big] \in \mathbb{R}^{d_{\mathrm{FEM}}}$ for $1 \le k \le K+1$ allows to rewrite

$$X_i = \big[a_{\boldsymbol{\nu}_1}^\top \quad \cdots \quad a_{\boldsymbol{\nu}_{K+1}}^\top\big]\, \boldsymbol{L}_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)})\boldsymbol{M}\boldsymbol{L}_{\boldsymbol{m}}(\boldsymbol{\sigma}^{(i)})\begin{bmatrix} b_{\boldsymbol{m}_1}^X \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^X \end{bmatrix},$$

where $\boldsymbol{M} = \mathrm{blockdiag}(M) \in \mathbb{R}^{d_{\mathrm{FEM}}\cdot(K+1)\times d_{\mathrm{FEM}}\cdot(K+1)}$ and

$$\boldsymbol{L}_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)}) = \begin{bmatrix} [L_{\boldsymbol{\nu}_1}(\boldsymbol{\sigma}^{(i)})] & & \\ & \ddots & \\ & & [L_{\boldsymbol{\nu}_{K+1}}(\boldsymbol{\sigma}^{(i)})] \end{bmatrix} \in \mathbb{R}^{d_{\mathrm{FEM}}\cdot(K+1)\times d_{\mathrm{FEM}}\cdot(K+1)},$$

with

$$[L_{\boldsymbol{\nu}_k}(\boldsymbol{\sigma}^{(i)})] \in \mathbb{R}^{d_{\mathrm{FEM}} \times d_{\mathrm{FEM}}} \quad \text{with entries} \quad [L_{\boldsymbol{\nu}_k}(\boldsymbol{\sigma}^{(i)})]_{j\ell} = \begin{cases} L_{\boldsymbol{\nu}_k}(\boldsymbol{\sigma}^{(i)}) & \text{if } j = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

Analogously, with $b_{\boldsymbol{m}_k}^Y := [\delta_{\boldsymbol{m}_k}^Y(t)] \in \mathbb{R}^{d_{\mathrm{FEM}}}$ for $1 \le k \le K+1$ we get

$$Y_i = \begin{bmatrix} a_{\boldsymbol{\nu}_1}^\top & \cdots & a_{\boldsymbol{\nu}_{K+1}}^\top \end{bmatrix} \boldsymbol{L_\nu}(\boldsymbol{\sigma}^{(i)}) \boldsymbol{M} \boldsymbol{L_m}(\boldsymbol{\sigma}^{(i)}) \begin{bmatrix} b_{\boldsymbol{m}_1}^Y \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^Y \end{bmatrix}.$$

For the set of quadrature points $\{\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(N)}\}$ we define

$$\mathfrak{M} = \begin{bmatrix} \begin{bmatrix} a_{\boldsymbol{\nu}_1}^\top & \cdots & a_{\boldsymbol{\nu}_{K+1}}^\top \end{bmatrix} \boldsymbol{L_\nu}(\boldsymbol{\sigma}^{(1)}) \boldsymbol{M} \boldsymbol{L_\nu}(\boldsymbol{\sigma}^{(1)}) \\ \vdots \\ \begin{bmatrix} a_{\boldsymbol{\nu}_1}^\top & \cdots & a_{\boldsymbol{\nu}_{K+1}}^\top \end{bmatrix} \boldsymbol{L_m}(\boldsymbol{\sigma}^{(N)}) \boldsymbol{M} \boldsymbol{L_m}(\boldsymbol{\sigma}^{(N)}) \end{bmatrix} \in \mathbb{R}^{N \times d_{\mathrm{FEM}} \cdot (K+1)}.$$

such that we can write

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} = \mathfrak{M} \begin{bmatrix} b_{\boldsymbol{m}_1}^X \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^X \end{bmatrix} \in \mathbb{R}^N \quad \text{and} \quad Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \mathfrak{M} \begin{bmatrix} b_{\boldsymbol{m}_1}^Y \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^Y \end{bmatrix} \in \mathbb{R}^N.$$

This allows the following representation of the discretized covariance operator

$$\langle \mathfrak{Cov}(\omega) X, Y \rangle_{\mathbb{R}^N} = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix}^\top \mathfrak{Cov}(\omega) \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} b_{\boldsymbol{m}_1}^X \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^X \end{bmatrix}^\top \mathfrak{M}^\top \mathfrak{Cov}(\omega) \mathfrak{M} \begin{bmatrix} b_{\boldsymbol{m}_1}^X \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^X \end{bmatrix}.$$

where $\mathfrak{M}^\top \mathfrak{Cov}(\omega) \mathfrak{M}$ is of dimension $d_{\mathrm{FEM}} \cdot (K+1) \times d_{\mathrm{FEM}} \cdot (K+1)$.

Furthermore, for the constant term in $\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t)$ we have

$$\langle C^* C \delta^X, \delta^Y \rangle_{L^2(\mathfrak{S}; H)} = \sum_{\boldsymbol{m}, \boldsymbol{\nu} \in \Lambda} \langle C^* C \delta_{\boldsymbol{\nu}}^X, \delta_{\boldsymbol{m}}^Y \rangle_H \langle \omega_{\theta, C}(\bar{y}(\boldsymbol{\sigma}; t)) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), L_{\boldsymbol{m}}(\boldsymbol{\sigma}) \rangle_{L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma})}$$

$$= \begin{bmatrix} b_{\boldsymbol{\nu}_1}^X \\ \vdots \\ b_{\boldsymbol{\nu}_{K+1}}^X \end{bmatrix}^\top \left[\tfrac{1}{N} \sum_{i=1}^N \omega_i L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)}) \boldsymbol{C^* M C} L_{\boldsymbol{m}}(\boldsymbol{\sigma}^{(i)}) \right] \begin{bmatrix} b_{\boldsymbol{m}_1}^Y \\ \vdots \\ b_{\boldsymbol{m}_{K+1}}^Y \end{bmatrix},$$

where $\boldsymbol{C^* M C} = \mathrm{blockdiag}(C^* M C) \in \mathbb{R}^{d_{\mathrm{FEM}} \cdot (K+1) \times d_{\mathrm{FEM}} \cdot (K+1)}$.

Overall, for $\boldsymbol{\nu}, \boldsymbol{m} \in \Lambda$, $t \in [0, T]$ and $\{\boldsymbol{\sigma}^{(i)}\}_{i=1}^N$ we have

$$\mathcal{Q}_{C,\omega}(\bar{y}(\cdot; t); t) \approx \frac{1}{N} \sum_{i=1}^N \left(\omega_i(t) L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}^{(i)}) \boldsymbol{C^* M C} L_{\boldsymbol{m}}(\boldsymbol{\sigma}^{(i)}) \right) + 2\theta \, \mathfrak{M}^\top(t) \mathfrak{Cov}(\omega; t) \mathfrak{M}(t).$$

The operator $\mathcal{Q}_{P,\omega}$ can be computed analogously.

6.6 Remarks and outlook on the error analysis

In addition to the spatial and temporal discretization of a deterministic optimal control problem subject to PDEs, the involved uncertainty requires the following approximations:

1. Dimension truncation of the parameter space, leading to

$$\int_{\mathfrak{S}} X(\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma} \approx \int_{\mathfrak{S}_s} X(\boldsymbol{\sigma}_s) \, \mathrm{d}\boldsymbol{\sigma}_s, \qquad (6.4)$$

for some integrable quantity of interest X.

2. Numerical integration of the s-dimensional integrals

$$\int_{\mathfrak{S}_s} X(\boldsymbol{\sigma}_s) \, \mathrm{d}\boldsymbol{\sigma}_s \approx \sum_{i=1}^{N} w_i X(\boldsymbol{\sigma}^{(i)}), \qquad (6.5)$$

for some integrable quantity of interest X.

3. Truncation of the polynomial chaos expansion

$$\sum_{\boldsymbol{\nu} \in \mathscr{F}} \delta_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}) \approx \sum_{\boldsymbol{\nu} \in \Lambda} \delta_{\boldsymbol{\nu}} L_{\boldsymbol{\nu}}(\boldsymbol{\sigma}), \qquad (6.6)$$

for a finite set $\Lambda \subset \mathscr{F}$ and some $\delta \in L^2(\mathfrak{S}; \mathrm{d}\boldsymbol{\sigma}; H)$

For sufficiently smooth problems these errors can be controlled. More precisely, assuming that $\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} A(\boldsymbol{\sigma})\|_{\mathcal{L}(V,V')} \leq \boldsymbol{b}^{\boldsymbol{\nu}}$ for a monotonically decreasing sequence $\boldsymbol{b} = (b_j)_{j \in \mathbb{N}}$, which is p-summable for some $p \in (0,1)$, it can be shown that the solution of (1.2) depends analytically on the parameter sequence $\boldsymbol{\sigma} \in \mathfrak{S}$ with $\|\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} y(\boldsymbol{\sigma})\|_V \leq C|\boldsymbol{\nu}|! \boldsymbol{b}^{\boldsymbol{\nu}}$. Here we use the following notation: for a sequence $\boldsymbol{\sigma} := (\sigma_j)_{j \in \mathbb{N}}$ of real numbers and $\boldsymbol{\nu} \in \mathscr{F}$, we define

$$\partial_{\boldsymbol{\sigma}}^{\boldsymbol{\nu}} := \frac{\partial^{\nu_1}}{\partial \sigma_1} \frac{\partial^{\nu_2}}{\partial \sigma_2} \cdots, \qquad \text{and} \qquad \boldsymbol{\sigma}^{\boldsymbol{\nu}} := \prod_{j=1}^{\infty} \sigma_j^{\nu_j},$$

where we follow the convention $0^0 := 1$.

In [30] it is shown that the optimal control and optimal state-adjoint-state pair of a linear quadratic optimal control problem admit analytic regularity with respect to the uncertain parameters, provided that the system operator $A(\boldsymbol{\sigma})$ admits the regularity as described above. Based on this result, in [21] it is shown that a Riccati based feedback law for an autonomous system $(A(\cdot), B, C, P)$ admits a similar regularity bound. Such bounds are frequently obtained in the context of Karhunen–Loève expansions of random fields ([3, 10, 8, 16]), and can be used to derive convergence rates for the approximations (6.4), (6.5), and (6.6): precisely the integrated dimension truncation error (6.4) decays as $\mathcal{O}(s^{-2/p+1})$ ([18]), while the L^2-dimension truncation error decays as $\mathcal{O}(s^{-1/p+1/2})$ ([17]). Randomly shifted rank-1 lattice rules achieve a root mean square error of $\mathcal{O}(N^{-1+\varepsilon})$ ([20, Thm. 6.6]), while interlaced polynomial lattice rules achieve an integration error convergence $\mathcal{O}(N^{-\alpha})$, $\alpha \geq 1$ ([21, Sect. 5]). Moreover, there exists a sequence of index sets $(\Lambda_n)_{n \in \mathbb{N}}$ whose cardinality does not exceed n such that the best n-term approximation rate (in L^2) of the Legendre polynomials is $\mathcal{O}(n^{1/p+1/2})$, see [9], or [30, Thm. 7].

The extension of these results to the presented setting with a time-dependent term appearing in the Riccati equation as well as a parameter-dependent target of the quadratic subproblems, which are used to approximate the risk-averse optimal control problem, exceeds the scope of this paper and is postponed to future works.

7 Numerical experiments

Let us consider the parameterized diffusion-reaction equation under Neumann boundary conditions as follows

$$\dot{y}(\boldsymbol{\sigma};t,x) - 0.5\Delta y(\boldsymbol{\sigma};t,x) + c(\boldsymbol{\sigma};x)y(\boldsymbol{\sigma};t,x) = \sqrt{10}\sum_{i=1}^{N_a} u_i(t,x)\mathbf{1}_{O_i}(x) \quad (t,x) \in (0,T] \times D,$$

$$\frac{\partial y(\boldsymbol{\sigma};t,x)}{\partial \mathrm{n}} = 0 \qquad\qquad (t,x) \in [0,T] \times \partial D,$$

$$y(\boldsymbol{\sigma};t,x) = y_\circ \qquad\qquad (t,x) \in \{t=0\} \times D,$$

where $T = 0.5$, $D = (0,1)$ with boundary $\partial D = \{0,1\}$, and the functions $\mathbf{1}_{O_i}$ represent the support of the actuators, which are modelled as the characteristic functions associated to open sets $O_i \subset D$ for $1 \le i \le N_a$. In the following numerical experiments $N_a = 3$ actuators are used as $O_1 = [0.1,0.3]$, $O_2 = [0.4,0.6]$, and $O_3 = [0.7,0.9]$. While the diffusion is constant, the reaction term is modeled as

$$c(\boldsymbol{\sigma};x) = \bar{c}(x) + \sum_{j=1}^{s} \sigma_j \psi_j(x),$$

where the mean field is set to $\bar{c}(x) = 0.2$ and σ_j are independent and identically distributed (i.i.d.) uniformly in $[-1,1]$ for all $j = 1,\ldots,s = 2$. Furthermore, the parametric basis functions are chosen to be $\psi_{2j}(x) = (2j)^{-\vartheta}\sin(j\pi x)$ and $\psi_{2j-1}(x) = (2j-1)^{-\vartheta}\cos(j\pi x)$ with decay $\vartheta = 2$. We use tensorized Legendre polynomials up to degree $p = 2$ for the stochastic Galerkin method, $(\Delta t)^{-1} = 200$ time steps for the temporal discretization and piecewise linear finite elements with meshwidth $h = 2^{-5}$ for the spatial discretization. The weights (6.3) are approximated using 100 i.i.d. samples of the trajectories in a Monte Carlo approximation for the expected values.

We set $P = 0$ and $Q = \mathbf{1}_H$, where $\mathbf{1}_H$ denotes the identity operator in H, as well as the initial condition $y_\circ(x) = 4 - \cos(2\pi x)$. The initial expansion point $\bar{y}^{(-1)}$ for Algorithm 1 solves the uncontrolled equation with shifted initial condition $\bar{y}_\circ(t=0,x) = 1 - \cos(2\pi x)$. The target g solves the state equation with $c(\boldsymbol{\sigma};x) = 0$ for all $\boldsymbol{\sigma} \in \mathfrak{S}$ and all $x \in D$, and with initial data $g(t=0,x) = 1.25 - \cos(2\pi x)$.

Comparison to the risk-neutral case   In Figure 1, we present the empirical distributions of the optimal tracking errors at six selected time points, based on 10^4 random draws of the parameter $\boldsymbol{\sigma} \in \mathfrak{S}$. We compare the risk-averse ($\theta = 10$) and risk-neutral ($\theta = 0$) cases. As expected, the distribution of the optimal risk-averse tracking errors are shifted toward smaller errors compared to the optimal risk-neutral case. This behavior is observed uniformly across all selected time points.

Comparison to reference trajectories   As before we compute 10^4 realizations of the tracking errors of the uncontrolled state. At $t = 0$ the tracking errors coincide for all realizations since the initial condition is deterministic. Furthermore, we observe that the tracking errors increase over time for all realizations, see Figure 2a. Figure 2b shows 10^4 realizations of the tracking error corresponding to the optimal state, obtained with the proposed SQP method (Algorithm 1) alongside 10^4 realizations of the tracking error corresponding to an open-loop reference solution computed by a gradient descent algorithm. The displayed results corresponds to the final iterate after 20 SQP iterations and 200 gradient descent iterations, respectively. The SQP tracking errors are indistinguishable from the open-loop reference, indicating that the
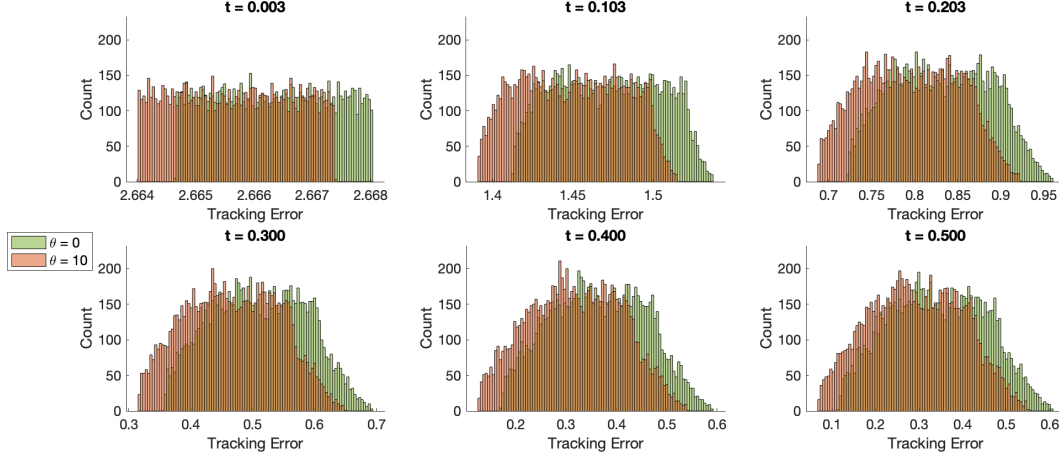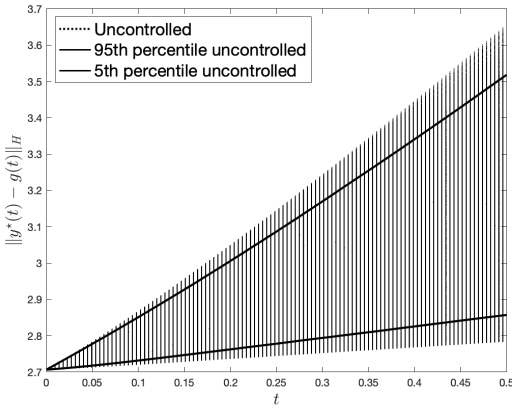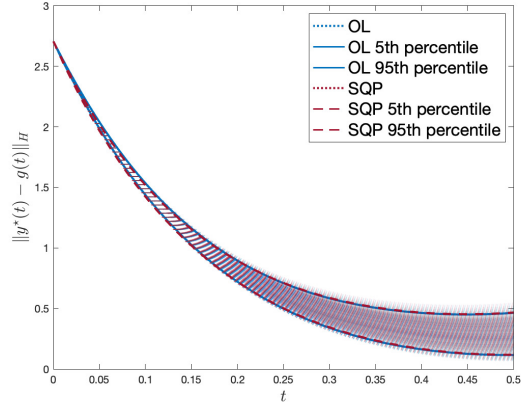
Figure 1: Histogram of the tracking errors at six selected time points based on 10^4 realizations of the optimal trajectory. Risk-averse case (green) and risk-neutral-case (orange).

(a) 10^4 realizations of the tracking error as a function of time for the uncontrolled state trajectories alongside the 95th and 5th percentiles.

(b) 10^4 realizations of the tracking error as a function of time for the optimal trajectories alongside the 95th and 5th percentiles (SQP and OL, respectively).

two methods converge approximately to the same solution, thus providing a validation for the numerical realizations.

Robustness against perturbed initial conditions Closed-loop solutions are expected to behave in a more robust manner against noise than open-loop controls. In this example we illustrate that this is also the case in the present setting. To demonstrate this behavior, the performance of the approximate feedback control obtained by the proposed SQP method is compared to an open-loop solution for varying initial conditions. To this end, we parameterize the initial condition by a noise level ℓ as

$$y_\ell^+ := y_0 + \ell + 0.01\ell\xi, \quad 0 \le \ell \le 2,$$
$$y_\ell^- := y_0 - \ell - 0.01\ell\xi, \quad 0 \le \ell \le 2.$$

where ξ is a vector of the same length as y_0 containing i.i.d. standard normally distributed random variables. The approximate feedback control as well as the open-loop solution are computed without noise, i.e., for $\ell = 0$ such that $y_0 = y_\ell^+ = y_\ell^-$, and then tested in the dynamics with perturbed initial conditions. The results for y_ℓ^+ and y_ℓ^- are displayed in Figure 3. It is observed that the approximate closed-loop control (CL) steers the state to approximately the same neighborhood of the target, i.e., the tracking errors have a similar magnitude across

21

different noise levels. The open-loop reference solution on the other hand clearly leads to larger tracking errors for increasing noise levels. Furthermore, with noise level $\ell = 0$ the CL and OL solutions are indistinguishable as expected.
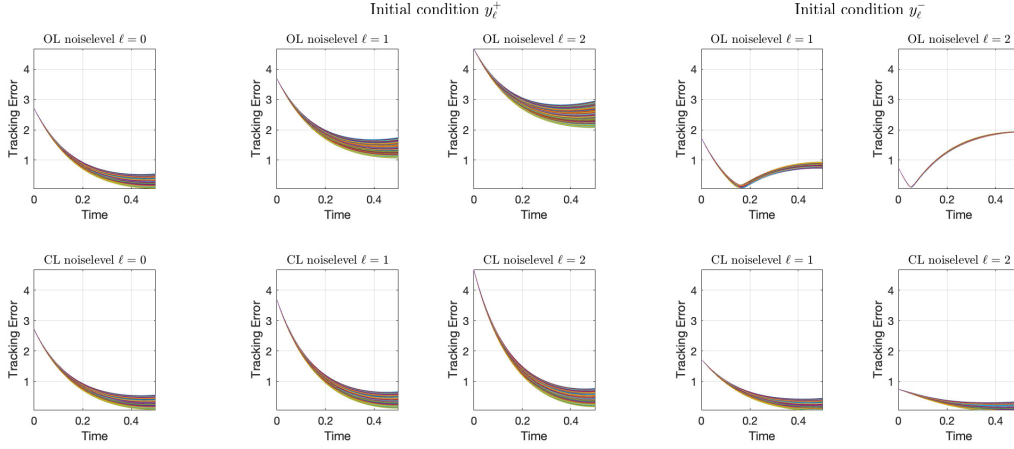

Figure 3: 10^4 realizations of the tracking error for the open-loop solution (top line) and the approximate closed-loop solution (bottom line) for different initial conditions. The plots in the left column correspond to the unperturbed initial condition y_0, i.e., to noise level $\ell = 0$. The plots in the center correspond to the perturbed intial condition y_ℓ^+, $0 \le \ell \le 2$, and the plots in the right-hand side block correspond to the perturbed intial condition y_ℓ^-, $0 \le \ell \le 2$.

8   Conclusions and outlook

We investigate a class of risk-averse optimal control problems for linear dynamical systems with uncertain evolution operators and develop an SQP algorithm to approximate the optimal feedback control. Such controls in feedback form can be constructed independently of the initial condition of the dynamics, which makes them particularly well-suited for time-critical applications involving uncertainty. Our approximate feedback is computed *a priori* and is therefore independent of any specific realization of the uncertain dynamics. We have shown that our approximation coincides with the unique optimal control along the optimal trajectory. Moreover, the numerical experiments suggest that our method yields a control that is significantly more robust to variations in initial conditions than the optimal open-loop control.

Future research directions include the investigation of computational strategies for large-scale problems with high-dimensional uncertainty, the extension to optimal control problems over an infinite time horizon, and a sequential update rule of the feedback driven by observational data.

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Math. Finance*, 9(3):203–228, 1999.

[2] B. Azmi, L. Herrmann, and K. Kunisch. Analysis of RHC for stabilization of nonautonomous parabolic equations under uncertainty. *SIAM J. Control Optim.*, 62(1):220–242, 2024.

[3] J. Beck, R. Tempone, F. Nobile, and L. Tamellini. On the optimal polynomial approximation of stochastic PDEs by Galerkin and collocation methods. *Math. Models Methods Appl. Sci.*, 22(09):1250023, 2012.

[4] A. Bensoussan, G. Da Prato, M. C. Delfour, and S. K. Mitter. *Representation and control of infinite dimensional systems*, volume 2. Springer, 2007.

[5] J. Bucklew. *Introduction to Rare Event Simulation*. Springer, 2014.

[6] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numer.*, 13:147–269, 2004.

[7] P. Chen and O. Ghattas. Taylor approximation for chance constrained optimization problems governed by partial differential equations with high-dimensional random parameters. *SIAM/ASA J. Uncertain. Quantif.*, 9(4):1381–1410, 2021.

[8] P. Chen, A. Quarteroni, and G. Rozza. Stochastic optimal Robin boundary control problems of advection-dominated elliptic equations. *SIAM J. Numer. Anal.*, 51(5):2700–2722, 2013.

[9] A. Cohen and R. DeVore. Approximation of high-dimensional parametric pdes. *Acta Numer.*, 24:1–159, 2015.

[10] A. Cohen, R. DeVore, and Ch. Schwab. Convergence rates of best n-term Galerkin approximations for a class of elliptic sPDEs. *Found. Comput. Math.*, 10(6):615–646, 2010.

[11] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer, Berlin, Heidelberg, 2 edition, 2010.

[12] J. Dick, F. Y. Kuo, and I. H. Sloan. High-dimensional integration: The quasi-monte carlo way. *Acta Numer.*, 22:133–288, 2013.

[13] D. B. Gahururu, M. Hintermüller, and T. M. Surowiec. Risk-neutral PDE-constrained generalized Nash equilibrium problems. *Math. Program.*, 198(2):1287–1337, 2023.

[14] C. Geiersbach and G. Ch. Pflug. Projected stochastic gradients for convex constrained problems in hilbert spaces. *SIAM J. Optim.*, 29(3):2079–2099, 2019.

[15] C. J. Gittelson. *Adaptive Galerkin methods for parametric and stochastic operator equations*. Dissertation, ETH Zurich, 2011.

[16] I. G. Graham, F. Y. Kuo, J. A. Nichols, R. Scheichl, Ch. Schwab, and I. H. Sloan. Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients. *Numer. Math.*, 131(2):329–368, 2015.

[17] P. A. Guth and V. Kaarnioja. Application of dimension truncation error analysis to high-dimensional function approximation in uncertainty quantification. In A. Hinrichs, P. Kritzer, and F. Pillichshammer, editors, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 297–312, Cham, 2024. Springer International Publishing.

[18] P. A. Guth and V. Kaarnioja. Generalized dimension truncation error analysis for high-dimensional numerical integration: Lognormal setting and beyond. *SIAM J. Numer. Anal.*, 62(2):872–892, 2024.

[19] P. A. Guth, V. Kaarnioja, F. Y. Kuo, C. Schillings, and I. H. Sloan. A quasi-Monte Carlo method for optimal control under uncertainty. *SIAM/ASA J. Uncertain. Quantif.*, 9(2):354–383, 2021.

[20] P. A. Guth, V. Kaarnioja, F. Y. Kuo, C. Schillings, and I. H. Sloan. Parabolic PDE-constrained optimal control under uncertainty with entropic risk measure using quasi-Monte Carlo integration. *Numer. Math.*, 156(2):565–608, 2024.

[21] P. A. Guth, P. Kritzer, and K. Kunisch. Quasi-Monte Carlo integration for feedback control under uncertainty. *SIAM/ASA J. Uncertain. Quantif.*, 13(3):1228–1264, 2025.

[22] P. A. Guth, K Kunisch, and S. S. Rodrigues. Tracking optimal feedback control under uncertain parameters. *Phys. D: Nonlinear Phenom.*, 467:134245, 2024.

[23] P. A. Guth, K. Kunisch, and S. S. Rodrigues. Ensemble feedback stabilization of linear systems. *Appl. Math. Optim.*, 92(1):21, 2025.

[24] J. Isohätälä and W. B. Haskell. Risk aware minimum principle for optimal control of stochastic differential equations. *IEEE Trans. Autom. Control*, 67(10):5102–5117, 2022.

[25] K. Ito and K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Society for Industrial and Applied Mathematics, USA, 2008.

[26] R. Kaas, M. Goovaerts, J. Dhaene, and M. Denuit. *Modern Actuarial Risk Theory Using R.* Springer, 1992.

[27] R. V. Kadison and J. R. Ringrose. *Fundamentals of the Theory of Operator Algebras*. Academic Press Inc., 1983.

[28] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. *SIAM J. Sci. Comput.*, 35(4):A1847–A1879, 2013.

[29] D. P. Kouri and T. M. Surowiec. Risk-averse PDE-constrained optimization using the conditional Value-At-Risk. *SIAM J. Optim.*, 26(1):365–396, 2016.

[30] A. Kunoth and Ch. Schwab. Analytic regularity and GPC approximation for control problems constrained by linear parametric elliptic and parabolic PDEs. *SIAM J. Control Optim.*, 51(3):2442–2471, 2013.

[31] J.-L. Lions. *Optimal control of systems governed by partial differential equations*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen; Bd. 170. Springer-Verlag, Berlin, 1971.

[32] M. Longo, Ch. Schwab, and A. Stein. A-posteriori QMC-FEM error estimation for Bayesian inversion and optimal control with entropic risk measure. In A. Hinrichs, P. Kritzer, and F. Pillichshammer, editors, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 431–450, Cham, 2024. Springer International Publishing.

[33] O. Melnikov and J. Milz. Randomized quasi-Monte Carlo methods for risk-averse stochastic optimization. *J. Optim. Theory Appl.*, 206(1):14, 2025.

[34] J. Milz. Consistency of Monte Carlo estimators for risk-neutral PDE-constrained optimization. *Appl. Math. Optim.*, 87(3):57, 2023.

[35] M. Reed and B. Simon. *Methods of Modern Mathematical Physics*. Elsevier Inc., 1972.

[36] W. Römisch and T. M. Surowiec. Asymptotic properties of Monte Carlo methods in elliptic PDE-constrained optimization under uncertainty. *Numer. Math.*, 156(5):1887–1914, 2024.

[37] G. Rubino and B. Tuffin, editors. *Rare Event Simulation Using Monte Carlo Methods*. John Wiley & Sons, Chichester, UK, 2009.

[38] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.