

Kernel-Based Optimal Control: An Infinitesimal Generator Approach

Petar Bevanda¹

PETAR.BEVANDA@TUM.DE

Nicolas Hoischen¹

NICOLAS.HOISCHEN@TUM.DE

Tobias Wittmann¹

T.WITTMANN@TUM.DE

Jan Brüdigam¹

JAN.BRUEDIGAM@TUM.DE

Sandra Hirche¹

HIRCHE@TUM.DE

Boris Houska²

BORISH@SHANGHAITECH.EDU.CN

¹*Chair of Information-oriented Control, Technical University of Munich, Germany*

²*School of Information Science and Technology, ShanghaiTech University*

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

This paper presents a novel operator-theoretic approach for optimal control of nonlinear stochastic systems within reproducing kernel Hilbert spaces. Our learning framework leverages data samples of system dynamics and stage cost functions, with only control penalties and constraints provided. The proposed method directly learns the infinitesimal generator of a controlled stochastic diffusion in an infinite-dimensional hypothesis space. We demonstrate that our approach seamlessly integrates with modern convex operator-theoretic Hamilton-Jacobi-Bellman recursions, enabling a data-driven solution to the optimal control problems. Furthermore, our learning framework includes nonparametric estimators for uncontrolled infinitesimal generators as a special case. Numerical experiments, ranging from synthetic differential equations to simulated robotic systems, showcase the advantages of our approach compared to both modern data-driven and classical nonlinear programming methods for optimal control.

Keywords: Controlled Diffusion Processes, Operator Learning, Optimal Control, RKHS, HJB

1. Introduction

Traditional optimal control theory and algorithms rely on first principle models to represent system dynamics, with parameters learned from data through system identification techniques (Ljung, 1998), either online or offline (Rawlings and Mayne, 2009). These models enable the formulation of nonlinear optimal control problems, which can be solved to global optimality using existing methods, such as dynamic programming (Beard et al., 1997), based on Hamilton-Jacobi-Bellman (HJB) theory (Fleming and Vermes, 1989; Crandall et al., 1992), or other types of global optimal control methods (Houska and Chachuat, 2014). However, due to their high complexity, these global optimal control methods are rarely used in practice (Houska and Chachuat, 2019); at least not for nonlinear systems with many states and controls. Instead, nonlinear programming approaches, such as interior point and sequential quadratic programming methods, are commonly used to solve nonlinear optimal control problems locally (Howell et al., 2019; Biegler, 2007; Bock and Plitt, 1984). Although these local methods are sensitive to suitable initial guesses, they have reached a high level of maturity (Diehl et al., 2002; Houska et al., 2011; Zavala and Biegler, 2009).

Recently, however, there has been a shift towards learning models directly from data. This shift leads to completely new types of system models as modern learning and system identification methods depart from relying on first principle models. While neural networks models are popular

and often effective in terms of capturing the nonlinear behavior of control systems (Beintema, 2024), such representations are highly nonlinear and difficult to exploit for optimal control (Lutter et al., 2020; Yang et al., 2024; Meng et al., 2024) due to modeling in a physical state-space.

In contrast to data-driven models in the physical state-space, modern operator-theoretic representations leverage linear operators, such as the Koopman operator (Mezić and Banaszuk, 2004; Mezić, 2005), to represent the dynamics of observables (functions of the state) (Bevanda et al., 2024a; Klus et al., 2020; Kostić et al., 2022; Vaidya and Tellez-Castro, 2024; Bruder et al., 2020; Bevanda et al., 2023), see Bevanda et al. (2021); Brunton et al. (2022) for recent reviews. These operator-theoretic models can be used as a basis to represent and learn system behavior from data. In the context of optimal control, the key advantages of operator-theoretic models is their compatibility with traditional stochastic control theory for diffusion processes (Fleming and Vermes, 1989; Crandall et al., 1992). This combination allows for the convexification of nonlinear optimal control problems using infinite-dimensional representations (Houska, 2025; Vinter, 1993). Furthermore, advances in synthesizing fully data-driven approaches for optimal control using kernel methods have shown promise in solving various types of optimal control problems (Bevanda et al., 2024b; Gopalakrishnan et al., 2022; Thorpe et al., 2022).

Contribution. This article contributes novel theory and an innovative algorithm for data-driven nonlinear system identification and globally optimal stochastic control with key contributions in:

- **System Identification.** We introduce a novel method for deriving non-parametric estimators of infinitesimal generators of controlled diffusions. Our approach leverages the properties of reproducing kernel Hilbert spaces (Lemma 1) to learn the adjoint of the infinitesimal generator of strongly parabolic Fokker-Planck-Kolmogorov (FPK) equation.
- **Stochastic Optimal Control.** With the data-driven estimates of the infinitesimal operator dynamics we formulate a tractable continuous-time Kernel Hamilton-Jacobi-Bellman (KHJB) (Proposition 2) that enables the computation of approximations of globally optimal solutions to stochastic optimal control problems via Algorithm 1. To demonstrate its performance, we apply this algorithm to both synthetic control systems and robotics benchmarks, showcasing its advantages over modern data-driven and classical nonlinear programming methods for optimal control.

Note that our estimators encompass nonparametric estimators for controlled infinitesimal generators of stochastic processes, which are notably absent from existing literature (Klus et al., 2016, 2020; Hou et al., 2023; Kostić et al., 2024). This distinction sets our work apart from finite-dimensional deterministic settings (Buzhardt and Tallapragada, 2023; Nüske et al., 2023) and unforced transfer operators (Kostić et al., 2022, 2024). Moreover, the continuous-time nature of our approach reduces the dependence on a specific time-lag of the data. This allows for the derivation of explicit controllers and value functions, and provides a counterpart to the discrete-time learning-based optimal control method presented in (Bevanda et al., 2024b).

Structure¹. After the problem statement in Section 2, we present an equivalent convex optimal con-

1. **Notation.** Let $\mu \in M_+(\mathbb{X})$ be a measure with full support, $\text{supp}(\mu) = \mathbb{X}$ and $M(\mathbb{X})$ the corresponding set of bounded signed Borel measures. The symbols $C^k(\mathbb{X})$, $L_\mu^k(\mathbb{X})$, $H_\mu^k(\mathbb{X})$ represent the set of k -times continuously differentiable, L_μ^k -integrable, and k -times weakly differentiable functions with L_μ^2 -integrable derivatives. For non-negative integers n and m , $[m, n] = \{m, m+1, \dots, n\}$ with $n \geq m$ gives an interval set $[n] := [1, n]$. Given a separable Hilbert space \mathcal{H} we let $\text{HS}(\mathcal{H})$ be a Hilbert space of Hilbert-Schmidt (HS) operators from \mathcal{H} to itself with norm $\|A\|_{\text{HS}}^2 \equiv \sum_{i \in \mathbb{N}} \|Ae_i\|_{\mathcal{G}}^2$ where $\{e_i\}_{i \in \mathbb{N}}$ is an orthonormal basis of \mathcal{H} .

trol formulation, using controlled Fokker-Plank-Kolmogorov (FPK) equations. Section 4 derives a novel operator regression in RKHS to approximate the infinitesimal generator for controlled diffusion, yielding feedback control policies through a simple dynamic programming recursion. Finally, Section 5 demonstrates our approach on a robotic swing-up task on the inverted pendulum and cart-pole and is validated by benchmark examples for synthetic ODEs and aforementioned simulated nonlinear systems.

2. Problem Statement

This work aims to learn an optimal feedback policy π^* for a control-affine nonlinear system, such that given the current state \mathbf{x} , the input $\mathbf{u}^* = \pi^*(\mathbf{x})$ solves the infinite horizon optimal control problem

$$\underset{\mathbf{x}, \mathbf{u}}{\text{minimize}} \int_0^\infty (q(\mathbf{x}) + r(\mathbf{u})) dt \quad \text{s.t.} \quad \dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u} =: \mathbf{f}_{\mathbf{u}}(\mathbf{x}), \quad \mathbf{u} \in \mathbb{U}, \quad (1)$$

with $\mathbf{f} \in C^1(\mathbb{X})^{n_x}$, $\mathbf{G} \in C^1(\mathbb{X})^{n_x \times n_u}$ and stage cost $q \in C^1(\mathbb{X})$. Here, we define $\mathbb{X} := \mathbb{R}^{n_x}$. Moreover, $\mathbb{U} := \{\mathbf{u} \in \mathbb{R}^{n_u} \mid \mathbf{u}_- \leq \mathbf{u} \leq \mathbf{u}_+\}$ denotes control bounds and $r \in C^1(\mathbb{U})$ is a strongly convex control penalty. Real systems, commonly modeled by ODEs, are often subject to process noise, which can be taken into account by replacing the deterministic models by a stochastic differential equation (SDE) (Øksendal, 2013). To that end, we build on the approach described in (Houska, 2025; Bevanda et al., 2024b) by considering the extension of (1) to dynamics subject to a small white noise disturbance, leading to the closed-loop process

$$d\mathbf{X}_t = (\mathbf{f}(\mathbf{X}_t) + \mathbf{G}(\mathbf{X}_t)\pi(\mathbf{X}_t)) dt + \sqrt{2\epsilon} d\mathbf{W}_t, \quad (\text{CSDE})$$

where \mathbf{W}_t is a \mathbb{R}^{n_x} Wiener process and $\epsilon > 0$ a diffusion parameter, modeling the amplitude of the process noise. Our goal is to obtain a (Lebesgue measurable) feedback $\pi : \mathbb{X} \rightarrow \mathbb{U}$ that minimizes the average *ergodic* cost,

$$\lim_{T \rightarrow \infty} \min_{\pi: \mathbb{X} \rightarrow \mathbb{U}} \mathbb{E} \left[\frac{1}{T} \int_0^T (q(\mathbf{X}_t) + r(\pi(\mathbf{X}_t))) dt \right] \quad \text{s.t.} \quad (\text{CSDE}). \quad (2)$$

The above stochastic optimal control problem formulation can either be viewed both as a *viscosity solution to (1)* for $\epsilon \rightarrow 0^+$, whenever this limit exists, or as an *effort to identify more robust control policies for noisy or uncertain dynamics* for $\epsilon > 0$. To ensure the above infinite-horizon problem is well-defined, we require the following assumption from (Bevanda et al., 2024b).

Assumption 1 *There exists a $\pi : \mathbb{X} \rightarrow \text{int}(\mathbb{U})$ with $\pi \in L^\infty(\mathbb{X})$ and a strongly convex $\mathcal{V} \in C^2(\mathbb{X})$ with bounded Hessian, and constants $0 < c_1, c_2 < \infty$ such that $(\mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\pi(\mathbf{x}))^* \nabla \mathcal{V}(\mathbf{x}) \leq c_1 - c_2(q(\mathbf{x}) + r(\pi(\mathbf{x})))$, for all $\mathbf{x} \in \mathbb{X}$. The set $\mathbb{U} \subseteq \mathbb{R}^{n_u}$ is closed, convex, and $\text{int}(\mathbb{U}) \neq \emptyset$. The control penalty $r \in C^1(\mathbb{U})$ is strongly convex.*

The above condition is often met in practical scenarios; see Houska (2025); Bevanda et al. (2024b) for an in-depth discussion. We require the following dataset to learn (CSDE).

Assumption 2 *There are state observations $\mathbf{X} := \{\mathbf{x}^{(i)}\}_{i \in [N]}$ of the nominal system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}$ under no excitation $\mathbf{u}_0 := \mathbf{0}$ and under “one-hot” (standard basis) input vectors $\{\mathbf{u}_j := \mathbf{e}_j\}_{j \in [n_u]}$, forming a dataset*

$$\mathbb{D}^N = \{\mathbb{D}_j^N\}_{j=0}^{n_u} \quad \text{where} \quad \mathbb{D}_j^N := \left\{ \mathbf{x}^{(i)}, \dot{\mathbf{x}}_{\mathbf{u}_j}^{(i)} := \mathbf{f}_{\mathbf{u}_j}(\mathbf{x}^{(i)}) \right\}_{i=1}^N. \quad (3)$$

While the above assumption may be restrictive for real-world data collection, it is still readily fulfilled in many settings, for example, when using existing physical or data-driven models (Umlauf et al., 2017) as well as data from high-fidelity simulators (Howell et al., 2022). Depending on whether the stage cost q is known or not, the vector $\mathbf{q}_X = [q(\mathbf{x}^{(i)})]_{i \in [N]}$ can subsequently be measured or computed on X . After a reformulation of a convex optimal control problem equivalent to (2), we propose learning infinitesimal generators of the associated controlled diffusion process, leading to a simple kernel-based dynamic programming recursion using the dataset (3).

3. Operator-Theoretic Dynamic Programming and HJB Recursions

As long as Assumption 1 holds, a martingale solution \mathbf{X}_t to (CSDE) exists for at least one feasible feedback π ; see Houska (2025) and (Bevanda et al., 2024b, Section II). Moreover, under the additional assumption that the stage cost of (2) has a bounded variance for the optimal ergodic limit distribution, existence of an optimal ergodic solution can be guaranteed for any $\epsilon > 0$; see (Houska, 2025, Thm. 1). The solution \mathbf{X}_t constitutes a time-homogeneous Markov diffusion process whose transition operator, $\Gamma_\pi(t)$, maps the probability density function $\rho_0 \in D_+(\mathbb{X})$ of \mathbf{X}_0 to the probability density function $\rho_t \in H^1(\mathbb{X})$ of \mathbf{X}_t so that $\rho_t = \Gamma_\pi(t)\rho_0$, where $D_+(\mathbb{X})$ is the set of non-negative bounded distributions on \mathbb{X} . It is well-known (Hinze et al., 2009; Øksendal, 2000) that—under mild regularity assumptions— $\Gamma_\pi(t)$ is for any given $\pi \in L^\infty((0, T) \times \mathbb{X})^{n_u}$ a bounded linear operator on $H^1(\mathbb{X})$.

It is worth recalling that the infinitesimal generator $\mathcal{L}_\pi : H^1(\mathbb{X}) \rightarrow H^1(\mathbb{X})$ associated to the Perron-Frobenius operator, is defined for every $\rho \in H^1(\mathbb{X})$ such that the limit $\mathcal{L}_\pi \rho := \lim_{t \rightarrow 0^+} \frac{\Gamma_\pi(t)\rho - \rho}{t}$ exists under mild regularity assumptions (Engel, 2000). The map \mathcal{L}_π is called *Fokker-Planck-Kolmogorov (FPK) operator* (Bogachev et al., 2015) or Kolmogorov’s infinitesimal generator (Froyland and Koltai, 2023). It describes the evolution of the probability density of the state of a diffusion process \mathbf{X}_t under arbitrary feedback laws $\pi \in L^\infty(\mathbb{X})$. Notice that the backward FPK operator \mathcal{L}_π^* , relates to the infinitesimal generator of the stochastic Koopman operator for autonomous stochastic differential equations (SDEs) (Kostić et al., 2024) that describes the evolution of observables $h \in H^1(\mathbb{X})$ under \mathbf{X}_t . Observing that \mathcal{L}_π is affine in π , will allow us to reformulate (2) as an equivalent convex optimization problem (Houska, 2025).

3.1. Convex Reformulation

If $\epsilon > 0$, we have $\rho_t = \Gamma_\pi(t)\rho_0 > 0$ for $t > 0$, since ρ_t is the state of a uniformly parabolic diffusion process (Bogachev et al., 2015). Exploiting that \mathcal{L}_π is affine in π allows to write the infinitesimal generator in the form

$$\dot{\rho} := \mathcal{L}_\pi \rho = \mathcal{A}\rho + \mathcal{B}(\pi\rho) \quad (4)$$

with linear-system shorthands $\mathcal{A}\rho = -\nabla \cdot (\mathbf{f}\rho) + \epsilon \nabla^2 \rho$ and $\mathcal{B}(\pi\rho) = -\nabla \cdot (\mathbf{G}\pi\rho)$, with the autonomous $\mathcal{A} : H^1(\mathbb{X}) \rightarrow H^1(\mathbb{X})$, and the control $\mathcal{B} : [H^1(\mathbb{X})]^{n_u} \rightarrow H^1(\mathbb{X})$ FPK operator. Then, the ergodic optimal control problem (2) can equivalently be cast as a convex PDE-constrained optimization problem through a change of variables, namely, $\nu = \pi\rho$ (Houska, 2025). In particular, for a finite time horizon $T < \infty$, and a given initial probability distribution of the state, $\rho(0) = \rho_0$ on \mathbb{X} , the finite-horizon stochastic optimal control problem is equivalent to solving the PDE-

constrained convex optimization problem

$$\mathcal{J}(T, \rho_0) := \min_{\rho, \nu} \int_0^T \int_{\mathbb{X}} \left(q + r \left(\frac{\nu}{\rho} \right) \right) \rho \, d\mathbf{x} \, dt \quad \text{s.t.} \quad \begin{cases} \dot{\rho} = \mathcal{A}\rho + \mathcal{B}\nu & \text{on } (0, T) \times \mathbb{X} \\ \rho(0) = \rho_0 & \text{on } \mathbb{X} \\ \nu \in \rho\mathbb{U} \end{cases} \quad (\text{PDE-OC})$$

Under the mentioned regularity assumptions, the ergodic limit in (2) is given by $\lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{J}(T, \rho)$ and invariant for $\epsilon > 0$ in the sense that it does not depend on the initial distribution ρ_0 , see (Houska, 2025, Thm. 1).

Hamilton-Jacobi-Bellman equations To exploit the convex duality of HJBs and FPKs, we first introduce the Fenchel conjugate $r^*(\lambda) := \sup_{\mathbf{u} \in \mathbb{U}} \{\langle \lambda, \mathbf{u} \rangle - r(\mathbf{u})\}$ of the control penalty r and define the following optimal policy form $\mathbf{u}^*(\cdot)$ and dual function $\mathcal{D}_r(\cdot) \equiv -r^*(-(\cdot))$, respectively

$$\mathbf{u}^*(\lambda) := \arg \min_{\mathbf{u} \in \mathbb{U}} \{r(\mathbf{u}) + \langle \lambda, \mathbf{u} \rangle\} \quad \text{and} \quad \mathcal{D}_r(\lambda) := \min_{\mathbf{u} \in \mathbb{U}} \{r(\mathbf{u}) + \langle \lambda, \mathbf{u} \rangle\}. \quad (5)$$

so $\mathbf{u}^*(\lambda) = -\nabla r^*(\lambda)$ and $\mathcal{D}_r(\lambda) = r(\mathbf{u}^*(\lambda)) + \langle \lambda, \mathbf{u}^*(\lambda) \rangle$. We construct an associated dual problem with the strongly measurable functional $V: [0, T] \rightarrow H_{\mu}^1(\mathbb{X})$ as a co-state, such that the infinite-dimensional optimization problem (PDE-OC) can be solved using the stochastic HJB equation

$$-\dot{V} = \mathcal{A}^*V + q + \mathcal{D}_r(\mathcal{B}^*V) \quad \text{s.t.} \quad V(T) = 0 \quad \text{on } \mathbb{X}_T \quad (\text{HJB-FVP})$$

which can be interpreted as a *final value problem* (FVP) in H_{μ}^1 .

The associated initial value $V(0)$ turns out to be a Riesz representation of the cost functional of (PDE-OC); that is, we have $\mathcal{J}(T, \rho_0) = \langle \rho_0, V \rangle$ for all initial probability distributions $\rho_0 \in H_{\mu}^1(\mathbb{X})$. In this context, one needs to introduce a suitable ergodic probability measure $\mu \in H^1(\mathbb{X})$ in order to define the weighted Sobolev space H_{μ}^1 ; see (Houska, 2025) for details. Conditions for which this argument holds (i.e., under strong duality) are nonrestrictive and can be found in (Bevanda et al., 2024b, Theorem 2). As the above stochastic HJB provides a Hilbert space FVP, a space discretization is usually required for practical computation, e.g., one coming from Galerkin methods as proposed in (Houska, 2025). Such discretizations are, however, computationally demanding in moderate to high dimensional state spaces. In the following section, we devise a flexible (nonparametric) framework and provide a data-driven approximation of (HJB-FVP) based on (3).

4. Generator Regression for Control-Affine Diffusions and HJB Approximation

Reproducing Kernel Hilbert Spaces We require the RKHS \mathcal{H} to be a dense subspace of the classical Sobolev space $H^1(\mathbb{X})$. Additionally, we recall that an RKHS is associated with a kernel function $k: \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$ that is symmetric positive definite. Let $\phi(\mathbf{x}) = k(\cdot, \mathbf{x}): \mathbb{X} \rightarrow \mathcal{H}$ be the canonical feature maps, denoted by $\phi(\mathbf{x})$, which can be assumed to satisfy $\phi(\mathbf{x}) \in H^1(\mathbb{X})$ for all $\mathbf{x} \in \mathbb{X}$. Moreover, $\forall \mathbf{x}, \mathbf{x}' \in \mathbb{X}$, we have that $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}} = \langle k(\cdot, \mathbf{x}), k(\cdot, \mathbf{x}') \rangle_{\mathcal{H}}$ and the reproducing property $h(\mathbf{x}) = \langle h, k(\cdot, \mathbf{x}) \rangle_{\mathcal{H}}$ holds for all $\mathbf{x} \in \mathbb{X}$ and all observables $h \in \mathcal{H}$. We introduce the (canonical) *embedding operator*, given for any $f \in L^2$ as $\mathcal{E}: f \mapsto \mathbb{E}_{\mathbf{X} \sim p}[f(\mathbf{X})\phi(\mathbf{X})]$, where $p \in \mathcal{M}_+(\mathbb{X})$ a probability measure. Its adjoint, the inclusion operator $\mathcal{E}^*: \mathcal{H} \rightarrow L^2$, is given by $(\mathcal{E}^*f)(\mathbf{x}) = f(\mathbf{x})$, for all $\mathbf{x} \in \mathbb{X}$ which we consider to be a Hilbert-Schmidt operator, which holds under very mild technical conditions (Steinwart and Christmann, 2008). As we are dealing with differential operators, we require the following assumption for well-posed regression.

Assumption 3 The RKHS \mathcal{H} is norm-equivalent to $H^s(\mathbb{X})$ with $s > \frac{n_x}{2} + 1$

This condition guarantees, via *Maurin's Theorem* (Adams and Fournier, 2003), that the canonical inclusion $\mathcal{H} \hookrightarrow H^1(\mathbb{X})$ is a Hilbert-Schmidt (HS) operator, allowing us to formulate *Hilbert-Schmidt regression* problems (Mollenhauer et al., 2022) to learn differential operators.²

Hilbert-Schmidt regression in infinite dimensions To approximate the operators $\mathcal{A}^*, \mathcal{B}^*$ in a nonparametric manner, we look for RKHS approximations $\{L_\pi^* : \mathcal{H} \rightarrow \mathcal{H}\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}$ based on its RKHS restrictions $\{\mathcal{L}_\pi^*|_{\mathcal{H}} : \mathcal{H} \rightarrow L^2\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}$. This is a valid strategy, as the control affinity of (CSDE) is inherited by its infinitesimal generator so that

$$\mathcal{A}^* \equiv \mathcal{L}_0^* \quad \text{and} \quad \mathcal{B}^* \equiv [(\mathcal{L}_{e_1}^* - \mathcal{L}_0^*) \cdots (\mathcal{L}_{e_{n_u}}^* - \mathcal{L}_0^*)], \quad (6)$$

we define the following set of risk functionals

$$\{\mathcal{R}(L_\pi^*) = \|\mathcal{L}_\pi^*|_{\mathcal{H}} - \mathcal{E}^* L_\pi^*\|_{\text{HS}(\mathcal{H}, L^2)}^2\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}, \quad (7)$$

to measure the mean square error for regression given the labels $\{\mathcal{L}_\pi^*|_{\mathcal{H}}\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}$. The latter is classical in the context of infinite-dimensional regression (Kostić et al., 2022, 2023; Li et al., 2022; Mollenhauer et al., 2022; Mollenhauer and Koltai, 2020). Based on mild regularity conditions (Houska, 2025) the FPK operator \mathcal{L}_π admits a bounded adjoint \mathcal{L}_π^* on $H^1 \subseteq L^2$. Then, by selecting a suitably regular RKHS to satisfy Assumption 3, and ensuring that the restriction $\mathcal{L}_\pi^*|_{\mathcal{H}} \in \text{HS}(\mathcal{H}, L^2)$, the optimization objective in (7) becomes well-defined.

Still, in practice, minimizing (7) may require solving a badly conditioned equation system. Thus, we formulate a Tikhonov-regularized problem

$$\left\{ L_\pi^* := \arg \min_{L_\pi^* \in \text{HS}(\mathcal{H})} \mathcal{R}(L_\pi^*) + \gamma \|L_\pi^*\|_{\text{HS}}^2 = C_\gamma^{-1} T_\pi \right\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}, \quad \gamma > 0 \quad (\text{KRR-FPK})$$

which corresponds to the *Kernel Ridge Regression* (KRR) approximation of \mathcal{L}_π over \mathcal{H} where operators T_π, C are defined as

$$C := \mathcal{E}\mathcal{E}^* = \mathbb{E}[\phi(\mathbf{X}) \otimes \phi(\mathbf{X})] \quad \{T_\pi := \mathcal{E}\mathcal{E}_\pi^* = \mathbb{E}[\phi(\mathbf{X}) \otimes \phi_{\mathcal{L}_\pi^*}(\mathbf{X})]\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}}, \quad (8)$$

where $\mathcal{E}_\pi^* = \mathcal{L}_\pi^*|_{\mathcal{H}}$. with the *regularized covariance* $C_\gamma = C + \gamma I_{\mathcal{H}}$. In practice, the risk (7) can only be evaluated on data (3) leading to the *empirical risk minimizations*

$$\left\{ \widehat{L}_\pi^* := \arg \min_{L_\pi^* \in \text{HS}(\mathcal{H})} \widehat{\mathcal{R}}(L_\pi^*) + \gamma \|L_\pi^*\|_{\text{HS}}^2 = \widehat{C}_\gamma^{-1} \widehat{T}_\pi = \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} \widehat{\mathcal{E}}_\pi^* \right\}_{\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}} \quad (\widehat{\text{KRR-FPK}})$$

to obtain a finite rank operators $\widehat{L}_0^* \cup \{\widehat{L}_{e_j}^*\}_{j \in [n_u]}$. This is due to the Gram matrix $\mathbf{K} := \widehat{\mathcal{E}}^* \widehat{\mathcal{E}} = [k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})]_{i,j \in [N]}$ in $\mathbf{K}_\gamma^{-1} := (\mathbf{K} + N\gamma I_N)^{-1}$ uncovered by the *Sherman-Morrison-Woodbury* formula and the reproducing property. The *data-based injection* operators are defined $\widehat{\mathcal{E}}^*, \widehat{\mathcal{E}}_\pi^* \in \text{HS}(\mathcal{H}, \mathbb{R}^N)$ as

$$\widehat{\mathcal{E}}^* : \mathcal{H} \rightarrow \mathbb{R}^N \text{ s.t. } \phi \mapsto [\phi^*(\mathbf{x}^{(i)})]_{i \in [N]} \quad \widehat{\mathcal{E}}_\pi^* : \mathcal{H} \rightarrow \mathbb{R}^N \text{ s.t. } \phi \mapsto [\phi_{\mathcal{L}_\pi^*}^*(\mathbf{x}^{(i)})]_{i \in [N]}. \quad (9)$$

2. The requirement stems from the condition $s > n_x/2 + 1$ for an embedding $H^s \hookrightarrow H^1$ to be Hilbert-Schmidt. The above assumption is fulfilled by popular kernels with sufficiently regular RKHS, e.g., Matern and Gaussian kernels.

and their adjoints, *data-based embeddings*,

$$\widehat{\mathcal{E}} : \mathbb{R}^N \rightarrow \mathcal{H} \text{ s.t. } \mathbf{w} \mapsto \sum_{i \in [N]} w_i \phi(\mathbf{x}_i) \quad \widehat{\mathcal{E}}_\pi : \mathbb{R}^N \rightarrow \mathcal{H} \text{ s.t. } \mathbf{w} \mapsto \sum_{i \in [N]} w_i \phi_{\mathcal{L}_\pi^*}(\mathbf{x}_i), \quad (10)$$

which can informally be considered as kernel-induced feature matrices. While ($\widehat{\text{KRR-FPK}}$) is defined over the RKHS \mathcal{H} , in practice, our computations will only require the Gram matrix \mathbf{K} and *target kernel matrices* $\mathbf{K}_0 := \widehat{\mathcal{E}}_0^* \widehat{\mathcal{E}}$, $\mathbf{K}_{e_j} := \widehat{\mathcal{E}}_{e_j}^* \widehat{\mathcal{E}} \in \mathbb{R}^{N \times N}$ for any $j \in [n_u]$ by plugging in the data (3) from Assumption 2, derived in the following lemma.

Lemma 1 *Let k be a Mercer kernel such that $k \in C^4(\mathbb{X} \times \mathbb{X})$ with corresponding RKHS \mathcal{H} and the system dynamics be described by (CSDE) under inputs $\pi \in \mathbf{0} \cup \{e_j\}_{j \in [n_u]}$. Then, the entries of the target kernel matrices are computed via*

$$(\mathbf{K}_\pi)_{ij} = \langle (\mathbf{f}(\mathbf{x}^{(i)}) + \mathbf{G}(\mathbf{x}^{(i)})\pi(\mathbf{x}^{(i)})), \nabla_{\mathbf{x}^{(i)}} k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \rangle + \epsilon \text{Tr}(\nabla_{\mathbf{x}^{(i)}}^2 k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})). \quad (11)$$

Proof First we use Itô formula (Arnold, 1974; Kostić et al., 2024) associated to (CSDE) for $\mathbf{u} \equiv \pi(\mathbf{x})$ to compute $\widehat{\mathcal{E}}_\pi$. After using (9)-(10) we have $(\mathbf{K}_\pi)_{ij} := (\widehat{\mathcal{E}}_\pi^* \widehat{\mathcal{E}})_{ij} = \langle \phi_{\mathcal{L}_\pi^*}(\mathbf{x}^{(j)}), \phi(\mathbf{x}^{(i)}) \rangle_{\mathcal{H}}$ so (11) is obtained after applying the *derivative reproducing property* (Zhou, 2008, Theorem 1). ■

With the control-affinity of the dynamics inherited by (11) for \mathbf{K}_π , we obtain matrices

$$\mathbf{A} := \mathbf{K}_\gamma^{-1} \mathbf{K}_0 \in \mathbb{R}^{N \times N} \quad \text{and} \quad \mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_{n_u}], \mathbf{B}_i := \mathbf{K}_\gamma^{-1} (\mathbf{K}_{e_i} - \mathbf{K}_0) \in \mathbb{R}^{N \times N}, \quad (12)$$

that are fully described using data from Assumption 2 after setting π to $\{\mathbf{u}_j\}_{j=0}^{n_u}$ in (11). With the help of the above lemma and control system matrices (12), we can state the following result.

Proposition 2 *Let the estimates for (HJB-FVP) be $\hat{A} := \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} \widehat{\mathcal{E}}_0^*$, $\{\hat{B}_i := \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} (\widehat{\mathcal{E}}_{e_m}^* - \widehat{\mathcal{E}}_0^*)\}_{m=1}^{n_u}$, $\hat{V} := \widehat{\mathcal{E}} \hat{v}$, $\hat{V} := \widehat{\mathcal{E}} \mathbf{v}$, $\hat{q} := \widehat{\mathcal{E}} \mathbf{q}$, $\hat{D}_r(\hat{B} \hat{V}) = \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} \widehat{\mathcal{E}}^* \mathcal{D}_r(\hat{B} \hat{V}) \in \mathcal{H}$, where \mathcal{H} fulfills the conditions of Lemma 1. Then, solving the infinite-dimensional*

$$\langle -\dot{\hat{V}} = \hat{A} \hat{V} + \hat{q} + \hat{D}_r(\hat{B} \hat{V}), k_{\mathbf{x}} \rangle \quad \text{s.t.} \quad \langle \hat{V}_T = 0, k_{\mathbf{x}} \rangle \quad \text{on } (0, T) \times \mathbb{X}.$$

amounts to N -dimensional final-value problem (FVP)

$$\langle -\dot{\mathbf{v}} = \mathbf{A} \mathbf{v} + \mathbf{q} + \mathbf{D}_r(\mathbf{B} \mathbf{v}), \mathbf{k}(\mathbf{x}) \rangle \quad \text{s.t.} \quad \langle \mathbf{v}_T = \mathbf{0}, \mathbf{k}(\mathbf{x}) \rangle \quad \text{on } (0, T) \times \mathbb{X}. \quad (\widehat{\text{HJB-FVP}})$$

where $\mathbf{k}(\mathbf{x}) = [k(\mathbf{x}^{(i)}, \mathbf{x})]_{i \in [N]} \in \mathbb{R}^N$ and $\mathbf{q} = \mathbf{K}_\gamma^{-1} \mathbf{q}_\mathbf{x}$, $\mathbf{D}_r(\mathbf{B} \mathbf{v}) = \mathbf{K}_\gamma^{-1} [\mathcal{D}_r(\langle \mathbf{B} \mathbf{v}, \mathbf{k}(\mathbf{x}^{(i)}) \rangle)]_{i \in [N]}$.

Proof The estimated (HJB-FVP) takes the following form in \mathcal{H}

$$\langle -\dot{\hat{V}} = \hat{A} \hat{V} + \hat{q} + \hat{D}_r(\hat{B} \hat{V}), k_{\mathbf{x}} \rangle \quad \text{s.t.} \quad \langle \hat{V}_T = 0, k_{\mathbf{x}} \rangle \quad \text{on } (0, T) \times \mathbb{X}.$$

After plugging the defined estimates, we obtain

$$\langle -\widehat{\mathcal{E}} \dot{\mathbf{v}} = \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} \widehat{\mathcal{E}}_0^* \widehat{\mathcal{E}} \mathbf{v} + \widehat{\mathcal{E}} \mathbf{q} + \widehat{\mathcal{E}} \mathbf{K}_\gamma^{-1} \widehat{\mathcal{E}}^* \mathcal{D}_r(\hat{B} \hat{V}), k_{\mathbf{x}} \rangle \quad \text{s.t.} \quad \langle \mathbf{v}_T = \mathbf{0}, \mathbf{k}(\mathbf{x}) \rangle \quad \text{on } (0, T) \times \mathbb{X}, \quad (13a)$$

$$\langle -\dot{\mathbf{v}} = \mathbf{A} \mathbf{v} + \mathbf{q} + \mathbf{D}_r(\mathbf{B} \mathbf{v}), \mathbf{k}(\mathbf{x}) \rangle \quad \text{s.t.} \quad \langle \mathbf{v}_T = \mathbf{0}, \mathbf{k}(\mathbf{x}) \rangle \quad \text{on } (0, T) \times \mathbb{X}, \quad (13b)$$

where (13b) is obtained by using the definitions (9)-(10), the derivative reproducing property of Lemma 1 and the reproducing property in (13a) for $\widehat{\mathcal{E}}^* k_{\mathbf{x}} = \mathbf{k}(\mathbf{x})$ and $[\hat{B} \hat{V}](\mathbf{x}) = \langle \mathbf{B} \mathbf{v}, \mathbf{k}(\mathbf{x}) \rangle$. ■

Algorithm 1 Infinitesimal Generator Kernel HJB Equation (IG-KHJB)

Require: Data \mathbb{D}^N (3) & samples for the state cost $q_X = [q(x^{(i)})]_{i=1}^N$, control penalty dual, diffusion $\epsilon > 0$, kernel k , regularizer $\gamma > 0$.
 Compute $K := [k(x^{(i)}, x^{(j)})]_{i,j \in [N]}$, $K_\gamma = (K + N\gamma I_N)$ and define $(\mathcal{D}_r(\hat{B}\hat{V}))_X := [\mathcal{D}_r(\langle Bv, k(x^{(i)}) \rangle)]_{i=1}^N$
 Compute A and B using (12) and $q = K_\gamma^{-1} q_X$, $\mathcal{D}_r(Bv) := K_\gamma^{-1} (\mathcal{D}_r(\hat{B}\hat{V}))_X$
function HJB-FVP($T, q, A, B, \mathcal{D}_r(\cdot)$)
 Initialize $v_T = 0$
 Integrate $-\dot{v} = Av + q + \mathcal{D}_r(Bv)$ from T to 0 ▷ e.g., using implicit Euler
 return v_0
end function
 $v_0 = \text{HJB-FVP}(T, q, A, B, \mathcal{D}_r(\cdot))$
 Compute $\hat{V}_0^*(x) = \langle v_0, k(x) \rangle$ and $\hat{\pi}^*(x) := u^*(\langle Bv_0, k(x) \rangle)$

System	Data grid	\mathbb{X}_S	$\max u $	σ	ϵ	γ	Δt	H
Unstable Oscillator	$\sqrt{N} \times \sqrt{N}$	$[\pm 3, \pm 3]$	—	$[5, 300]$	0.01	10^{-8}	0.01s	1000
Inverted Pendulum	50×50	$[\pm 0.99\pi, \pm 10]$	1.5 Nm	25	0.02	10^{-12}	0.02s	500
Cartpole	$9 \times 7 \times 23 \times 23$	$[\pm 2.5, \pm 3, \pm 0.99\pi, \pm 8]$	7 N	15	0.01	10^{-12}	0.01s	3000

Table 1: Parameter values used in Algorithm 1 for the experiments in Sections 5.2 and 5.3. We use an implicit integration scheme to solve IG-KHJB over a time-horizon $T = H\Delta t$.

5. Numerical Experiments

In this section, we present numerical examples to evaluate the performance of our IG-KHJB approach using Algorithm 1. We compare our approach to Bevanda et al. (2024b) for optimal control of an unstable oscillator and to state-of-the-art NMPC for a swing-up and stabilization task on the inverted pendulum and cartpole systems. The latter employ Dojo (Howell et al., 2022) for dynamics simulation and Altro (Howell et al., 2019) in a receding-horizon fashion (NMPC)³.

Implementation details In the unstable oscillator example (5.1), we use the ODE of the system directly as well as analytical formulations for the kernel partial derivatives. In the examples using Dojo, the infinitesimal measurements for (3) are not directly accessible and is therefore approximated via finite differences (FD). Additionally, to speed up computations, partial derivatives are also approximated by FD. We use an Euler-Implicit scheme to integrate (HJB-FVP) with a time step size Δt , with parameter values for each experiment described in Table 1. The controllers for inverted pendulum and cartpole were smoothed using $\frac{2 \max|u|}{\pi} \arctan(\hat{\pi}^*(x))$ according to the input constraint from Table 1.

5.1. Unstable Oscillator

We compare our IG-KHJB to Bevanda et al. (2024b) on the 2D Van der Pol Oscillator using identical dynamics and cost functions. The optimal infinite-horizon policy for $\epsilon \rightarrow 0^+$ can be analytically computed as $\pi_\infty^*(x) = -x_1 x_2$ (Villanueva et al., 2021). Both methods are compared using the root mean square error (RMSE) against this optimal policy, evaluated on 1000 uniformly sampled test points. Both methods employ an RBF kernel⁴ $e^{-\|x-x'\|^2/\sigma^2}$. In Figure 1 (left), we evaluate the

3. ALTRO’s parameters including initial guesses and OCP discretization were hand-tuned to best performance.

4. The discrete-time kHJB requires a diffused RBF kernel; see Bevanda et al. (2024b) for details.

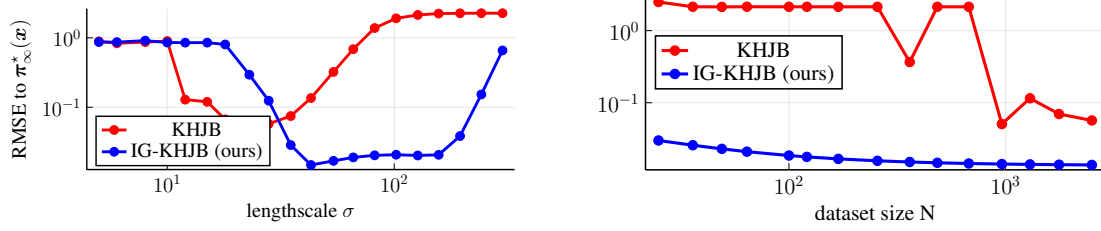


Figure 1: Comparison of RMSE to the known optimal policy $\pi_{\infty}^*(x)$ between KHJB (Bevanda et al. (2024b)) and our IG-KHJB approach for the Van der Pol Oscillator.

performance of KHJB (Bevanda et al., 2024b) against our proposed IG-KHJB for a varying lengthscale (Table 1) using $N = 2500$ datapoints. In Figure 1 (right), the optimal lengthscales are fixed to $\sigma_{\text{KHJB}} = 23$, $\sigma_{\text{IG-KHJB}} = 43$ and the RMSE is evaluated for both methods across $N \in \{5^2, \dots, 50^2\}$ data. The results show that our method is more robust to the kernel lengthscale choice and achieves up to three times lower RMSE with fewer data compared to KHJB. Notably, IG-KHJB achieves a lower RMSE than KHJB with only $N = 25$ datapoints.

5.2. Torque Limited Inverted Pendulum

We evaluate our method on a swing-up and stabilization task at the upright equilibrium $\theta = 0$ for the inverted pendulum. The value function and policy shown in Fig. 2 are learned using Algorithm 1 with data (3) generated from Dojo (Howell et al., 2022). We use a Laplace kernel $e^{-\|x-x'\|/\sigma}$, that is additionally smoothed to allow for stable derivative computations⁵. The employed parameters can be found in Table 1. To deal with wrap-around of θ we define the state $x := [c_\theta, s_\theta, \dot{\theta}]^\top$ using shorthands $s_\theta := \sin \theta$ and $c_\theta := \cos \theta$. The state cost is given by $q(x) = q_1 s_\theta^2 + q_2 (c_\theta - 1)^2 + q_v \dot{\theta}^2$ with $q_1, q_2 = 30, q_v = 1$ and control penalization $r(u) = \|u\|_{1/2}^2$, clipped at the control limits. We compare our policy with regard to the accumulated trajectory costs to state-of-the-art NMPC using Altro (Howell et al., 2019) for trajectory optimization. We use the same pendulum dynamics for Altro as in Dojo⁶. Both our policy, $\hat{\pi}^*(x)$, and Altro-NMPC are deployed with a control frequency of 50Hz. Figure 4(a) shows that our method achieves lower mean accumulated costs than Altro-NMPC under identical stage costs. Each lasting 5 seconds, the trajectories were simulated with initial states sam-

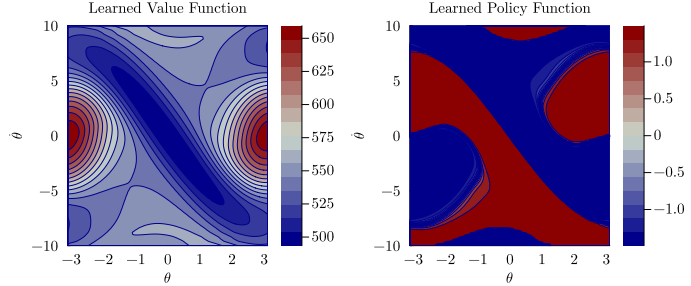
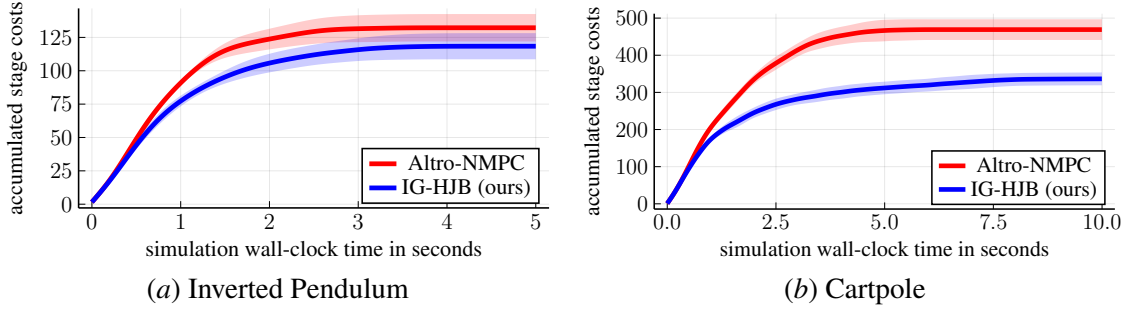


Figure 2: Contour plots of the value and controller functions learned from $2.5 \cdot 10^3$ samples using our IG-KHJB approach (upright at $(\dot{\theta}, \theta) = (0, 0)$).

5. We provide the derivative at 0 using those of an RBF kernel with $\sigma/100$.

6. With $m = 1.0\text{kg}$, $g = 9.81\text{m/s}^2$, $l = 1.0\text{m}$, $b = 0.1\text{kg.m}^2\text{s}^{-1}$, $I = 0.0842\text{kg.m}^2$ with m, g, l, b being respectively the mass, the gravitational constant and length of the pendulum, viscous damping and inertia around its mass center.

Figure 3: Accumulated stage costs using our learned policy $\hat{\pi}^*(x)$ and `Altro`-NMPC.

pled uniformly from $[\pm\pi, \pm 8]$. The mean accumulated costs are averaged over 10 runs, with 50 different initial positions each.

5.3. Cartpole: Inverted Pendulum on a Cart

We now consider the swing-up and stabilization task for a Cartpole system and compare it to `Altro`-NMPC. We again use the smoothed Laplace kernel, with the parameters in Table 1 and data from `Dojo`. Here, we also augment the state space representation to $x = [x, \dot{x}, c_\theta, s_\theta, \dot{\theta}]^\top$. To describe the task via stage costs, we penalize the Euclidean distance between the pendulum's endpoint (x_p, y_p) and the goal $(0, l)$ (McAllister and Rasmussen, 2017), defined as $d(x, c_\theta, s_\theta) := q_h x_p^2 + q_v (y_p - l)^2 = q_h (x - l s_\theta)^2 + q_v (c_\theta - 1)^2$, with a velocity penalty, this amounts to the stage cost $q(x) = d(x, c_\theta, s_\theta) + q_{vel} \dot{x}^2 + q_\omega \dot{\theta}^2$ and control penalty $r(u) = \|u\|_{1/5}^2$, where $q_h = 10$, $q_v = 100l^2$, $q_{vel}, q_\omega = 1$. For the cost comparison, trajectories with a duration of 10s are simulated with initial states sampled random uniformly on $[0, \pm 2, \pm\pi, \pm 6]$. Here, our controller's accumulated costs (see Figure 4(b)) are lower than those of `Altro`-NMPC⁷ by a large margin, computed over 10 runs with 50 different initial states per run and deployed at 200Hz. To eliminate confounding effects from model-mismatch, we ran `Altro` on an analytic dynamics model that we built to replicate the assumptions and parameterization used by `Dojo` in our experiments.

6. Conclusion

This article has introduced novel methods for data-driven nonlinear system identification and globally optimal stochastic control. In detail, we have derived non-parametric estimators of infinitesimal generators of optimally controlled diffusions, as summarized in Lemma 1. Moreover, in Proposition 2 we have formulated a continuous-time Kernel Hamilton-Jacobi-Bellman (KHJB) equation, which enables the computation of data-driven approximations of globally optimal solutions to stochastic optimal control problems. Our method, outlined in Algorithm 1, has been demonstrated to outperform modern data-driven and classical nonlinear programming methods for optimal control in both synthetic and robotics benchmarks.

7. The same dynamics as in `Dojo` are used, i.e. $M = 0.5\text{kg}$, $m = 0.5\text{kg}$, $l = 1.0\text{m}$, $b = 0.05\text{kg.m}^2\text{s}^{-1}$, $k = 0.05\text{kg.s}^{-1}$, $I = 0.0513\text{kg.m}^2$, being respectively the cart mass, pendulum mass, pendulum length, viscous rotational damping, viscous damping of the cart, and the pendulum inertia around the center of mass.

Acknowledgments

We thank Robert Lefringhausen and Max Beier for their valuable feedback while preparing this manuscript. This work was supported by the European Union’s Horizon Europe innovation action programme under grant agreement No. 101093822, “SeaClear2.0”, the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, sponsored by the Federal Ministry of Education and Research.

References

- Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140. Elsevier, 2003.
- Ludwig Arnold. *Stochastic differential equations: theory and applications*, volume 2. John Wiley & Sons, 1974.
- Randal W. Beard, George N Saridis, and John T Wen. Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation. *Automatica*, 33(12):2159–2177, 1997.
- Gerben Izaak Beintema. *Data-driven Learning of Nonlinear Dynamic Systems: A Deep Neural State-Space Approach*. Phd thesis, Eindhoven University of Technology, 2024.
- Petar Bevanda, Stefan Sosnowski, and Sandra Hirche. Koopman operator dynamical models: Learning, analysis and control. *Annu. Rev. Control*, 52:197–212, 2021.
- Petar Bevanda, Max Beier, Armin Lederer, Stefan Sosnowski, Eyke Hüllermeier, and Sandra Hirche. Koopman Kernel Regression. In *Neural Information Processing Systems*, volume 37, 2023.
- Petar Bevanda, Bas Driessen, Lucian Cristian Iacob, Roland Toth, Stefan Sosnowski, and Sandra Hirche. Nonparametric Control Koopman Operators. *arXiv preprint arXiv:2405.07312*, 2024a.
- Petar Bevanda, N. Hoischen, Stefan Sosnowski, Sandra Hirche, and Boris Houska. Data-driven optimal feedback laws via kernel mean embeddings. *arXiv preprint arXiv:2407.16407*, 2024b.
- Lorenz T. Biegler. An overview of simultaneous strategies for dynamic optimization. *Chem. Eng. Process. Process Intensif.*, 46:1043–1053, 2007.
- Hans Georg Bock and K.J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. *Proceedings 9th IFAC World Congress Budapest*, pages 243–247, 1984.
- Vladimir I. Bogachev, Nikolai V. Krylov, Michael Röckner, and Stanislav V. Shaposhnikov. *Fokker-Planck-Kolmogorov equations*. AMS, 2015.
- Daniel Bruder, Xun Fu, R Brent Gillespie, C David Remy, and Ram Vasudevan. Data-driven control of soft robots using koopman operator theory. *IEEE Transactions on Robotics*, 37(3):948–961, 2020.
- Steven L. Brunton, Marko Budišić, Eurika Kaiser, and J. Nathan Kutz. Modern koopman theory for dynamical systems. *SIAM Review*, 64(2):229–340, 2022.

- Jake Buzhardt and Phanindra Tallapragada. Controlled density transport using Perron Frobenius generators. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, pages 651–656. IEEE, 2023.
- Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bull. Am. Math. Soc.*, 27(1):1–67, 1992.
- Moritz Diehl, H.Georg Bock, Johannes P. Schlöder, Rolf Findeisen, Zoltan Nagy, and Frank Allgöwer. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *Journal of Process Control*, 12(4):577–585, 2002. ISSN 0959-1524.
- Klaus-Jochen Engel. *One-Parameter Semigroups for linear evolution equations*. Springer-Verlag, 2000.
- Wendell H. Fleming and Domokos Vermes. Convex duality approach to the control of diffusions. *SIAM Journal on Control and Optimization*, 27(5):1136–1155, 1989.
- Gary Froyland and Péter Koltai. Detecting the birth and death of finite-time coherent sets. *Communications on Pure and Applied Mathematics*, 76(12):3642–3684, 2023.
- Bharath Gopalakrishnan, Arun Kumar Singh, Madhava Krishna, and Dinesh Manocha. Solving Chance-Constrained Optimization Under Nonparametric Uncertainty Through Hilbert Space Embedding. *IEEE Transactions on Control Systems Technology*, 30(3):901–916, 2022.
- Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE Constraints*. Springer, 2009.
- Boya Hou, Sina Sanjari, Nathan Dahlin, Subhonmesh Bose, and Umesh Vaidya. Sparse learning of dynamical systems in RKHS: An operator-theoretic approach. In *International Conference on Machine Learning*, pages 13325–13352. PMLR, 2023.
- Boris Houska. Convex operator-theoretic methods in stochastic control. *Automatica*, 177:112274, 2025. ISSN 0005-1098. doi: <https://doi.org/10.1016/j.automatica.2025.112274>.
- Boris Houska and Benoit Chachuat. Branch-and-lift algorithm for deterministic global optimization in nonlinear optimal control. *Journal of Optimization Theory and Applications*, 162:208–248, 2014.
- Boris Houska and Benoit Chachuat. Global optimization in Hilbert space. *Mathematical Programming*, 173:221–249, 2019.
- Boris Houska, Hans Joachim Ferreau, and Moritz Diehl. An auto-generated real-time iteration algorithm for nonlinear mpc in the microsecond range. *Automatica*, 47(10):2279–2285, 2011. ISSN 0005-1098. doi: <https://doi.org/10.1016/j.automatica.2011.08.020>.
- Taylor A Howell, Brian E Jackson, and Zachary Manchester. Altro: A fast solver for constrained trajectory optimization. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7674–7679. IEEE, 2019.

- Igor Mezić and Andrzej Banaszuk. Comparison of systems with complex behavior. *Physica D: Nonlinear Phenomena*, 197(1-2):101–133, 2004.
- Mattes Mollenhauer and Péter Koltai. Nonparametric approximation of conditional expectation operators. *arXiv preprint arXiv:2012.12917*, 2020.
- Mattes Mollenhauer, Nicole Mücke, and T. J. Sullivan. Learning linear operators: Infinite-dimensional regression as a well-behaved non-compact inverse problem. *arXiv preprint arXiv:2211.08875*, 2022.
- Feliks Nüske, Sebastian Peitz, Friedrich Philipp, Manuel Schaller, and Karl Worthmann. Finite-data error bounds for koopman-based prediction and control. *Journal of Nonlinear Science*, 33(1):14, 2023.
- Bernt Øksendal. *Stochastic Differential Equations*. Springer, 2000.
- Bernt Øksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.
- James B. Rawlings and David Q. Mayne. *Model Predictive Control: Theory and Design*. Madison, WI: Nob Hill Publishing, 2009.
- Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Information Science and Statistics. Springer, New York, NY, first edition, 2008.
- Adam J. Thorpe, Thomas Lew, Meeko Mitsuko Karen Oishi, and Marco Pavone. Data-driven chance constrained control using kernel distribution embeddings. In *PMLR*, volume 168, pages 790–802, 2022.
- Jonas Umlauft, Armin Lederer, and Sandra Hirche. Learning stable gaussian process state space models. In *2017 American Control Conference (ACC)*, pages 1499–1504. IEEE, 2017.
- Umesh Vaidya and Duvan Tellez-Castro. Data-driven stochastic optimal control with safety constraints using linear transfer operators. *IEEE Trans. Autom. Control*, 69(4):2100–2115, 2024. doi: 10.1109/TAC.2023.3288623.
- Mario Eduardo Villanueva, Colin N. Jones, and Boris Houska. Towards global optimal control via Koopman lifts. *Automatica*, 132(109610), 2021.
- Richard Vinter. Convex duality and nonlinear optimal control. *SIAM Journal on Control and Optimization*, 31(2):518–538, 1993.
- Lujie Yang, Hongkai Dai, Zhouxing Shi, Cho-Jui Hsieh, Russ Tedrake, and Huan Zhang. Lyapunov-stable neural control for state and output feedback: A novel formulation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *PMLR*, pages 56033–56046. PMLR, 21–27 Jul 2024.
- Victor M. Zavala and Lorenz T. Biegler. The advanced-step nmmpc controller: Optimality, stability and robustness. *Automatica*, 45(1):86–93, 2009. ISSN 0005-1098. doi: <https://doi.org/10.1016/j.automatica.2008.06.011>.

Ding-Xuan Zhou. Derivative reproducing properties for kernel methods in learning theory. *Journal of Computational and Applied Mathematics*, 220(1):456–463, 2008. ISSN 0377-0427. doi: <https://doi.org/10.1016/j.cam.2007.08.023>.