Journal of
Nonlinear
Science

Finite-Data Error Bounds for Koopman-Based Prediction and Control

Feliks Nüske[1,2] (iD) · **Sebastian Peitz[1]** (iD) · **Friedrich Philipp[3]** (iD) ·
Manuel Schaller[3] (iD) · **Karl Worthmann[3]** (iD)

Abstract

The Koopman operator has become an essential tool for data-driven approximation of dynamical (control) systems, e.g., via extended dynamic mode decomposition. Despite its popularity, convergence results and, in particular, error bounds are still scarce. In this paper, we derive probabilistic bounds for the approximation error and the prediction error depending on the number of training data points, for both ordinary and stochastic differential equations while using either ergodic trajectories or i.i.d. samples. We illustrate these bounds by means of an example with the Ornstein–Uhlenbeck process. Moreover, we extend our analysis to (stochastic) nonlinear control-affine systems. We prove error estimates for a previously proposed approach that exploits the linearity of the Koopman generator to obtain a bilinear surrogate control system

✉ Sebastian Peitz
    sebastian.peitz@upb.de

    Feliks Nüske
    nueske@mpi-magdeburg.mpg.de

    Friedrich Philipp
    friedrich.philipp@tu-ilmenau.de

    Manuel Schaller
    manuel.schaller@tu-ilmenau.de

    Karl Worthmann
    karl.worthmann@tu-ilmenau.de

[1] Paderborn University, Paderborn, Germany

[2] Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

[3] Technische Universität Ilmemau, Ilmenau, Germany

🖄 Springer

and, thus, circumvents the curse of dimensionality since the system is not autonomized by augmenting the state by the control inputs. To the best of our knowledge, this is the first finite-data error analysis in the stochastic and/or control setting. Finally, we demonstrate the effectiveness of the bilinear approach by comparing it with state-of-the-art techniques showing its superiority whenever state and control are coupled.

Keywords Koopman operator · Nonlinear systems · Error bounds · Data-driven control

Mathematics Subject Classification 37A50 · 37C30 · 37M10 · 47D03 · 47D06 · 93B28 · 93B30

1 Introduction

The Koopman framework (Koopman 1931) is the operator-theoretic basis for a wide range of data-driven methodologies to predict the evolution of nonlinear dynamical systems using linear techniques, see, e.g., Mezić (2005), Rowley et al. (2009) or the recent survey (Brunton et al. 2022) and the references therein. The underlying concept is that observables, which may also be understood as outputs from the systems-and-control perspective, can be propagated forward in time using the linear yet infinite-dimensional Koopman operator or its generator, instead of simulating the nonlinear system and evaluating the observable functions. Its recent success is closely linked to numerically tractable approximation techniques like extended Dynamic Mode Decomposition (eDMD), see, e.g., Williams et al. (2015), Klus et al. (2016), Korda and Mezić (2018b), Klus et al. (2020) for numerical techniques and convergence results.

While the Koopman framework is well established, approximation results are typically only established in the infinite-data limit, i.e., if *sufficient* data are available. Recently, Lu and Tartakovsky (2020) discussed error bounds w.r.t. DMD invoking the seminal work (Korda and Mezić 2018b) by Korda and Mezić. While the authors numerically demonstrate the effectiveness of their approach even for nonlinear parabolic Partial Differential Equations (PDEs), see also their extension (Lu and Tartakovsky 2021), there remains a significant gap from a more theoretical point of view since the approximation error is assumed to be zero for *finite data*, see Lu and Tartakovsky (2020, Remark 3.1). Mamakoukas et al. (2021) mimic a Taylor-series expansion based on a particular set of observables to approximate the system dynamics of an Ordinary Differential Equation (ODE). This work may be understood as a promising approach to incorporate (local) knowledge on the system dynamics in the Koopman framework. However, a bound on the prediction error in terms of data is not deduced. Error bounds for Koopman eigenvalues in terms of the finite-data estimation error were derived in Webber et al. (2021), but the estimation error itself was not quantified. In Mollenhauer et al. (2020), concentration inequalities were applied to bound the estimation error for the co-variance and cross-covariance operators involved in Koopman estimation. In the exhaustive preprint (Kurdila and Bobade 2018), the authors treat the projection error for different approximation spaces such as, e.g., reproducing kernel Hilbert spaces

and wavelets. The estimation error is also discussed briefly in Sect. 8.5. In Zhang and Zuazua (2021), besides providing a finite-data error bound on the approximation of the Koopman operator in the context of ODEs, the authors estimate the projection error by means of finite-element analysis. In conclusion, to the best of our knowledge,[1] Zhang and Zuazua (2021), Kurdila and Bobade (2018) are the only works providing rigorous error bounds for Koopman-based approximations of a dynamical system governed by a nonlinear ODE.

In this paper, we rigorously derive probabilistic bounds on the approximation error (or finite-data estimation error) and the (multi-step) prediction error for nonlinear Stochastic Differential Equations (SDEs). This, of course, also includes nonlinear ODEs. The deduced bounds on the approximation error and prediction accuracy explicitly depend on the number of data points used in eDMD. To this end, besides using mass concentration inequalities and a numerical error analysis to deal with the error propagation in time, we employ substantially different techniques in comparison to Kurdila and Bobade (2018), Zhang and Zuazua (2021) to provide an additional alternative assumption based on ergodic sampling tailored to stationary SDEs. Our results in this setting focus on the concept of asymptotic variance, see Lelièvre and Stoltz (2016) and the references therein. In contrast to most concentration inequalities, the asymptotic variance is a genuinely dynamical quantity. Even though it cannot be directly accessed for most complex systems, it provides a solid basis for further theoretical study of the estimation error. For instance, it was shown in Duncan et al. (2016) that a spectral analysis of the generator can be used to speed up convergence to equilibrium and, by extension, the convergence of empirical estimators. In this study, we use a simple Ornstein–Uhlenbeck process to illustrate our error bounds in practice and show that they are surprisingly sharp. This serves as additional motivation to continue the study of the sampling error for ergodic sampling by means of asymptotic variances. Let us stress that we do not fully analyze the projection error; in other words, how much the Koopman generator fails to be invariant on the approximation subspace, referring to the existing literature Kurdila and Bobade (2018) and Zhang and Zuazua (2021) in the autonomous case and our follow-up work (Schaller et al. 2022) in the control setting.

W.r.t. the application of Koopman theory in control, a lot of research has been invested over the past years, beginning with the popular *DMD with control* (Proctor et al. 2016), which was later used in Model Predictive Control (MPC) (Korda and Mezić 2018a). Another popular method is to use a coordinate transformation into Koopman eigenfunctions (Kaiser et al. 2021) or the already mentioned component-wise Taylor series expansion (Mamakoukas et al. 2021). In Lu et al. (2020), the prediction error of the method proposed in Proctor et al. (2016) was estimated using the convergence result of Korda and Mezić (2018b). However, the result is of purely asymptotic nature, i.e., it does not state a convergence rate in terms of data points. All approaches mentioned until now yield linear surrogate models of the form $Ax + Bu$, i.e., the control enters linearly. For general control-affine systems, numerical simulation studies indicate that bilinear surrogate models are better suited, see Goswami and Paley (2017), Peitz et al. (2020), Bruder et al. (2021), Peitz and Bieker (2021). The technique proposed in

---

[1] We are already referring to two authoritative references on preprint servers supporting our claim that finite-data error bounds are still missing; thanks to one of the unknown referees for drawing our attention to the still unpublished work (Kurdila and Bobade 2018).

Peitz and Klus (2019); Peitz et al. (2020) constructs its surrogate model from $n_c + 1$ autonomous Koopman operators, where n_c is the control dimension. The key feature is that the state-space dimension is not augmented by the number of control inputs, which counteracts the curse of dimensionality in comparison with the more widespread approach introduced in Korda and Mezić (2018a). Compared to Peitz et al. (2020), we present a detailed analysis of the accuracy regarding both the dictionary size and the amount of training data. Even though the bound is rather coarse on the operator level, we demonstrate that it correctly captures the qualitative behavior. In this context, we provide a probabilistic bound on the approximation error of the projected Koopman generator, the projected Koopman semigroup and the respective trajectories. To this end, we extend our results toward nonlinear control systems. Besides a rigorous bound on the approximation error, we present estimates on the (auto-regressive) prediction accuracy, i.e., in an open-loop prediction (without feedback). For control systems, we also refer to the follow-up work (Schaller et al. 2022), where we obtained the following two extensions: On the one hand, we deduced quantitative estimates of the projection error depending on the (finite) dictionary size. Combining this with the error bounds depending on the (finite) number of data points proposed in this work yields a complete analysis of the approximation error. On the other hand, we further elaborated the estimates such that the error bounds uniformly hold for a set of admissible control functions rendering the approach applicable for optimal and predictive control.

The paper is structured as follows. Firstly, in Sect. 2, we deduce a rigorous bound on the approximation error for nonlinear SDEs. Then, we extend our analysis to nonlinear control-affine systems in Sect. 3. In Sect. 4, two numerical simulation studies for the Ornstein–Uhlenbeck system (SDE) and the controlled Duffing equation (nonlinear control-affine system) are presented before conclusions are drawn in Sect. 5.

2 Finite-Data Bounds on the Approximation Error: Nonlinear SDEs

In this section, we analyze the approximation quality of extended Dynamic Mode Decomposition (eDMD) with finitely many data points for the finite-dimensional stochastic differential equation

$$\mathrm{d}X_t = F(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{SDE}$$

where $X_t \in \mathbb{X} \subset \mathbb{R}^d$ is the state, $F : \mathbb{X} \to \mathbb{R}^d$ is the drift vector field, $\sigma : \mathbb{X} \to \mathbb{R}^{d \times d}$ is the diffusion matrix field, and W_t is a d-dimensional Brownian motion. We assume that F, σ satisfy standard Lipschitz properties to ensure global existence of solutions to (SDE), see the textbook (Oksendal 2013) for an introduction to this class of systems. We stress that the deterministic case is included by simply setting $\sigma \equiv 0$, leading to the ordinary differential equation

$$\tfrac{\mathrm{d}}{\mathrm{d}t} x(t) = F(x(t)).$$

The state space is assumed to be a measure space $(\mathbb{X}, \Sigma_{\mathbb{X}}, \mu)$ with Borel σ-algebra $\Sigma_{\mathbb{X}}$ and probability measure μ. In case of an ODE, the set $\mathbb{X}$ is often assumed to be

compact and forward-invariant and the probability measure is the standard Lebesgue measure, cf. Zhang and Zuazua (2021).

Definition 1 (Koopman operator) Let X_t satisfy (SDE) for $t \geq 0$. The Koopman operator semigroup associated with (SDE) is defined by

$$\mathcal{K}^t f(x_0) = \mathbb{E}^{x_0}[f(X_t)] = E[f(X_t)|X_0 = x_0]$$

for all bounded measurable functions f.

In case of ergodic sampling, that is, obtaining data points from a single long trajectory, we will assume invariance of the measure μ w.r.t. the stochastic process X_t.

Definition 2 (Invariant measure with positive density) A probability measure μ is called invariant if it satisfies

$$\int_{\mathbb{X}} \mathcal{K}^t f \, \mathrm{d}\mu = \int_{\mathbb{X}} f \, \mathrm{d}\mu$$

for all bounded measurable functions f and all $t \geq 0$. Further, μ has an everywhere positive density $\rho : \mathbb{X} \to \mathbb{R}$ if $\mu(A) = \int_A \rho(x) \, \mathrm{d}x$ holds for all $A \in \Sigma_{\mathbb{X}}$.

We can now formulate our assumption on the underlying dynamics.

Assumption 3 Let either of the following hold:

(a) The set $\mathbb{X}$ is compact and forward invariant ($\forall x^0 \in \mathbb{X} : \mathbb{P}^{x_0}(X_t \in \mathbb{X}) = 1$ for all $t \geq 0$) and μ is the normalized Lebesgue measure. Moreover, the Koopman operator can be extended to a strongly continuous semigroup on the Hilbert space $L^2_\mu(\mathbb{X})$.
(b) The probability measure is an invariant measure in the sense of Definition 2.

We briefly comment on this assumption and first note that forward invariance of $\mathbb{X}$ can be weakened, if one is only interested in estimates for states contained in $\mathbb{X}$, see also Zhang and Zuazua (2021, Section 3.2). Moreover, if the dynamics obey an ODE, it was shown that the Koopman operator can indeed be extended to a strongly continuous semigroup on $L^2_\mu(\mathbb{X})$, see also Zhang and Zuazua (2021). Second, the assumption of invariance of the underlying probability measure is satisfied for a broad class of SDEs, see, e.g., Risken (1996). It can be checked that μ is then invariant for X_t, that is, $\mathbb{P}(X_t \in A) = \mu(A)$ holds for all $A \in \Sigma_{\mathbb{X}}$ and $t \geq 0$, provided X_0 is distributed according to μ. Under Assumption 3(b), Definition 1 can be extended to the Lebesgue spaces $L^p_\mu(\mathbb{X})$, $1 \leq p < \infty$, i.e., the Banach spaces of all (equivalence classes of) measurable functions $f : \mathbb{X} \to \mathbb{R}$ with $\int_{\mathbb{X}} |f|^p \, \mathrm{d}\mu < \infty$. Then, the Koopman operators $\mathcal{K}^t$ form a strongly continuous semigroup of contractions on all spaces $L^p_\mu(\mathbb{X})$, see Bakry et al. (2013). The functions in any of these spaces are often referred to as *observables*.

Next, we recall the definition of the generator associated with the semigroup $\mathcal{K}_t$:

Definition 4 (Koopman generator) The infinitesimal generator $\mathcal{L}$ is defined via

$$\mathcal{L}f := \lim_{t \to 0} \frac{(\mathcal{K}^t - \mathrm{Id})f}{t} \tag{1}$$

for all $f \in D(\mathcal{L})$, where $D(\mathcal{L})$ is the set of functions for which limit (1) exists in the appropriate topology.

For sufficiently smooth functions f, Ito's lemma (Oksendal 2013) shows that the generator acts as a second-order differential operator, defined in terms of the coefficients of (SDE), i.e.,

$$\mathcal{L} = F \cdot \nabla + \tfrac{1}{2}\sigma\sigma^\top : \nabla^2 \tag{2}$$

with $A : B := \sum_{i,j} a_{i,j} b_{i,j}$ being the standard Frobenius inner product for matrices. In what follows, we will focus exclusively on the Koopman semigroup on the Hilbert space $L^2_\mu(\mathbb{X})$ with inner product $\langle f, g \rangle_\mu = \int_{\mathbb{X}} fg \, d\mu$. As the semigroup is strongly continuous on $L^2_\mu(\mathbb{X})$ by our assumptions, by standard semigroup theory, the domain $D(\mathcal{L})$ together with the graph norm forms a dense Banach space in $L^2_\mu(\mathbb{X})$.

2.1 Extended Dynamic Mode Decomposition

In this part we introduce the data-driven finite-dimensional approximation by eDMD of the Koopman generator defined in (1), see, e.g., Williams et al. (2015), Klus et al. (2016), Klus et al. (2018). To this end, for a fixed set of linearly independent observables $\psi_1, \ldots, \psi_N \in D(\mathcal{L})$, we consider the finite-dimensional subspace

$$\mathbb{V} := \mathrm{span}\{\{\psi_j\}_{j=1}^N\} \subset D(\mathcal{L}).$$

Let $P_{\mathbb{V}}$ denote the orthogonal projection onto $\mathbb{V}$. We define the Galerkin projection of the Koopman generator by $\mathcal{L}_{\mathbb{V}} := P_{\mathbb{V}}\mathcal{L}|_{\mathbb{V}}$. Note that this is not the restriction of $\mathcal{L}$ onto $\mathbb{V}$, as the image is also projected back onto $\mathbb{V}$. If $\mathbb{V}$ is an invariant set under the action of the generator, then $\mathcal{L}_{\mathbb{V}} = \mathcal{L}|_{\mathbb{V}}$ holds. As $\dim \mathbb{V} = N$, the linear operator $\mathcal{L}_{\mathbb{V}} : \mathbb{V} \to \mathbb{V}$ may be represented by a matrix. In what follows, we denote the matrix representation of $\mathcal{L}_{\mathbb{V}}$ in terms of the basis functions $\psi_1, \ldots, \psi_N$ by the same symbol $\mathcal{L}_{\mathbb{V}}$ as the operator itself in a slight abuse of notation. Thus, using Klus et al. (2020), we get

$$\mathcal{L}_{\mathbb{V}} = C^{-1}A$$

with $C, A \in \mathbb{R}^{N \times N}$ defined by $C_{i,j} = \langle \psi_i, \psi_j \rangle_{L^2_\mu(\mathbb{X})}$ and $A_{i,j} = \langle \psi_i, \mathcal{L}\psi_j \rangle_{L^2_\mu(\mathbb{X})}$. The norm of the isomorphism from $\mathbb{V}$ to $\mathbb{R}^N$ depends on the smallest resp. largest eigenvalues of C, cf. Proposition 21 in "Appendix A.1."

Consider data points $x_0, \ldots, x_{m-1} \in \mathbb{X}$. In the following, these data are either drawn from a trajectory of an ergodic system or sampled independent and identically

distributed (i.i.d.). We state this as the following assumption, using the notation:

$$L^2_{\mu,0}(\mathbb{X}) := \{f \in L^2_\mu(\mathbb{X}) : \langle f, 1 \rangle_\mu = 0\}.$$

Assumption 5 Let Assumption 3 hold and assume either of the following.

(iid) The data are drawn i.i.d. from the measure specified via Assumption 3.
(erg) Assumption 3.(b) holds and the data are obtained as snapshots from a single ergodic trajectory, that is, from a single long trajectory of dynamics (SDE) with x_0 drawn from the unique invariant measure μ. Further assume the Koopman semigroup is exponentially stable on $L^2_{\mu,0}(\mathbb{X})$, i.e., $\|\mathcal{K}^t\|_{L^2_{\mu,0}(\mathbb{X})} \leq Me^{-\omega t}$ for some $M \geq 1$, $\omega > 0$.

The second assumption (erg) is satisfied for a broad class of ergodic SDEs that are considered widely in, for example, statistical physics and molecular simulation. However, it should also be noted that it is not at all universal. For instance, for ergodic ODEs, the Koopman operator is unitary and hence cannot be exponentially stable. In this case one can still resort to i.i.d. sampling.
Let us form the transformed data matrices

$$\Psi(X) := \left(\left. \begin{pmatrix} \psi_1(x_0) \\ \vdots \\ \psi_N(x_0) \end{pmatrix} \right| \cdots \left| \begin{pmatrix} \psi_1(x_{m-1}) \\ \vdots \\ \psi_N(x_{m-1}) \end{pmatrix} \right. \right)$$

$$\mathcal{L}\Psi(X) := \left(\left. \begin{pmatrix} (\mathcal{L}\psi_1)(x_0) \\ \vdots \\ (\mathcal{L}\psi_N)(x_0) \end{pmatrix} \right| \cdots \left| \begin{pmatrix} (\mathcal{L}\psi_1)(x_{m-1}) \\ \vdots \\ (\mathcal{L}\psi_N)(x_{m-1}) \end{pmatrix} \right. \right).$$

The evaluation of $\mathcal{L}$ can be realized via representation (2). If the coefficients of dynamics (SDE) are not available in explicit form, they need to be approximated, for example, using Kramers–Moyal expansions. The analysis of this source of error is beyond the scope of this study. Furthermore, there might occur challenges when evaluating the time derivatives of the observables to compute $\mathcal{L}\Psi(X)$ if the system is not explicitly given. This is a well-known problem and may be addressed by various numerical differentiation techniques (Brunton et al. 2016; van Breugel et al. 2020). Alternatively, one can resort to the finite-time Koopman operator as performed in Sect. 2.3, which has been observed to provide robust results in various applications, cf., e.g., Peitz et al. (2020); Klus et al. (2022). The empirical estimator for the Galerkin projection $\mathcal{L}_\mathbb{V}$ is then given by

$$\tilde{\mathcal{L}}_m = \tilde{C}_m^{-1} \tilde{A}_m$$

with $\tilde{C}_m = \frac{1}{m}\Psi(X)\Psi(X)^\top$, $\tilde{A}_m = \frac{1}{m}\Psi(X)\mathcal{L}\Psi(X)^\top \in \mathbb{R}^{N \times N}$. In all scenarios of Assumption 5, we have with probability one that

(1) $\tilde{\mathcal{L}}_m$ is well-defined for large enough m, that is, $\tilde{C}_m$ is invertible, and
(2) $\tilde{\mathcal{L}}_m$ converges to $\mathcal{L}_\mathbb{V}$ for $m \to \infty$, see, e.g., Klus et al. (2018, 2020).

For the case of a long trajectory, this result follows from ergodic theory, which is concerned with the convergence of time averages to spatial averages as the data size grows to infinity (Beck and Schwartz 1957). Ergodic theory particularly applies to systems with a unique invariant measure.

2.2 Error Bounds on Approximations of Projected Koopman Generator And Operator

Next, we quantify the approximation quality of the data-driven finite-dimensional approximation of the Koopman generator, i.e., for a given linear space $\mathbb{V}$ of observables and data $x_0, \ldots, x_{m-1} \in \mathbb{X}$, we aim to estimate

$$\|\mathcal{L}_{\mathbb{V}} - \tilde{\mathcal{L}}_m\|_F = \|C^{-1}A - \tilde{C}_m^{-1}\tilde{A}_m\|_F.$$

2.2.1 Concentration Bounds for Random Matrices

We start by deriving entry-wise error bounds for the data-driven mass and stiffness matrix, respectively. Since most of the arguments are significantly simpler for i.i.d. sampling, cf. Remark 11 at the end of this subsection, we first consider the more involved situation, i.e., ergodic sampling. This is of particular interest as simulation data of dynamics (SDE) can, then, be directly used.

For $x \in \mathbb{X}$, consider a centered scalar random variable

$$\phi : \mathbb{X} \mapsto \mathbb{R}, \quad \int_{\mathbb{X}} \phi(x)\,\mathrm{d}\mu(x) = 0.$$

We denote its variance with respect to the invariant measure by

$$\sigma_\phi^2 = \mathbb{E}^\mu[\phi^2] = \|\phi\|_{L_\mu^2}^2.$$

Moreover, we set $\phi_k = \phi(x_k)$ for given data points x_k, $k \in \{0, 1, \ldots, m-1\}$ and define the averaged random variable

$$\bar{\phi}_m := \frac{1}{m} \sum_{k=0}^{m-1} \phi_k.$$

In Lemma 6 below, we quantify the variance of the averaged random variable $\bar{\phi}_m$. The key point is the decomposition of the variance into an asymptotic contribution, independent of m, and a second contribution, which decays with an explicitly given (polynomial) dependence on the amount of data m.

Lemma 6 *Let Assumption 5.(erg) hold. Then we have*

$$\sigma_{\bar{\phi}_m}^2 = \frac{1}{m}\left[\sigma_{\phi,\infty}^2 - R_\phi^m\right]. \tag{3}$$

The asymptotic variance $\sigma_{\phi,\infty}^2$ and the remainder term R_ϕ^m are given by

$$\sigma_{\phi,\infty}^2 = \sigma_\phi^2 + 2\sum_{l=1}^{\infty}\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu, \quad R_\phi^m = 2\sum_{l=m}^{\infty}\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu + \frac{2}{m}\sum_{l=1}^{m-1}l\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu.$$

Proof We repeat the proof given in Lelièvre and Stoltz (2016, Section 3.1.2) for the sake of illustration:

$$
\begin{aligned}
\sigma_{\bar{\phi}_m}^2 &= \frac{1}{m^2}\sum_{k,l=0}^{m-1}\mathbb{E}^\mu\left[\phi_k\,\phi_l\right] = \frac{1}{m}\sigma_\phi^2 + \frac{2}{m^2}\sum_{k=0}^{m-1}\sum_{l=k+1}^{m-1}\mathbb{E}^\mu\left[\phi_k\,\phi_l\right] \\
&= \frac{1}{m}\left[\sigma_\phi^2 + \frac{2}{m}\sum_{k=0}^{m-1}\sum_{l=k+1}^{m-1}\mathbb{E}^\mu\left[\phi_0\,\phi_{l-k}\right]\right] = \frac{1}{m}\left[\sigma_\phi^2 + \frac{2}{m}\sum_{k=0}^{m-1}\sum_{l=1}^{m-k-1}\mathbb{E}^\mu\left[\phi_0\,\phi_l\right]\right] \\
&= \frac{1}{m}\left[\sigma_\phi^2 + \frac{2}{m}\sum_{l=1}^{m-1}(m-l)\mathbb{E}^\mu\left[\phi_0\,\phi_l\right]\right] = \frac{1}{m}\left[\sigma_\phi^2 + 2\sum_{l=1}^{m-1}(1-\tfrac{l}{m})\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu\right].
\end{aligned}
$$

The result follows by adding and subtracting the term $2\sum_{l=m}^{\infty}\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu$. □

Remark 7 The assumption of exponential stability is satisfied, for example, if the generator $\mathcal{L}$ is self-adjoint (also known as detailed balance or reversibility) and additionally satisfies a Poincaré or spectral gap inequality (Lelièvre and Stoltz 2016). The requirement $\langle f, 1\rangle_\mu = 0$ is necessary, as the constant function is invariant for $\mathcal{K}^t$.

Remark 8 The proof of Lemma 6 shows that $\sigma_{\phi,\infty}^2 = \lim_{m\to\infty}\sigma_{\bar{\phi}_m}^2 \geq 0$; hence, it can indeed be interpreted as a variance.

For reversible systems, we have $\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu \geq 0$ by symmetry of the Koopman operator. Therefore, $\sigma_{\phi,\infty}^2 \geq \sigma_\phi^2 > 0$ is guaranteed in this case, and the variance $\sigma_{\bar{\phi}_m}^2$ approaches $\frac{1}{m}\sigma_{\phi,\infty}^2$ from below.

Next, we derive an estimate for the remainder term in terms of the number m of data points.

Lemma 9 *Let Assumption 5.(erg) hold, and set $q = e^{-\omega\Delta_t} < 1$. Then*

$$|R_\phi^m| \leq \frac{2\sigma_\phi^2}{m}\frac{q}{(1-q)^2}.$$

Proof We first observe that by the Cauchy Schwarz inequality

$$|\langle\phi,\ \mathcal{K}^{l\Delta_t}\phi\rangle_\mu| \leq \|\mathcal{K}^{l\Delta_t}\|_{L(L_{\mu,0}^2(\mathbb{X}),L_\mu^2(\mathbb{X}))}\|\phi\|_{L_\mu^2(\mathbb{X})}^2 \leq e^{-\omega\Delta_t l}\sigma_\phi^2,$$

and therefore:

$$|R_\phi^m| \le 2\sigma_\phi^2 \left[\sum_{l=m}^\infty e^{-\omega\Delta_t l} + \frac{1}{m}\sum_{l=1}^{m-1} l e^{-\omega\Delta_t l} \right]$$

$$= 2\sigma_\phi^2 \left[\frac{q^m}{1-q} + \frac{1}{m}\frac{(m-1)q^{m+1} - mq^m + q}{(1-q)^2} \right]$$

$$= \frac{2\sigma_\phi^2}{m}\frac{q(1-q^m)}{(1-q)^2} \le \frac{2\sigma_\phi^2}{m}\frac{q}{(1-q)^2}.$$

In the second line, we have used the geometric series for the first term and a similar identity for the sum $\sum_{l=1}^\infty l q^l$, $q < 1$. The third line is obtained by direct simplification. $\qquad\square$

We can now combine the results of Lemmas 6 and 9 in order to obtain a concentration bound for a centered, matrix-valued random variable. To this end, we consider an $N \times N$ random matrix Φ with all entries $\phi_{ij} \in L_{\mu,0}^2$ centered. We define Φ_k and $\overline{\Phi}_m$ as for the scalar case, i.e., $\Phi_k = \Phi(x_k)$ and $\overline{\Phi}_m = \frac{1}{m}\sum_{k=0}^{m-1} \Phi_k$.

Proposition 10 *Let Assumption 5.(erg) hold,, set $q = e^{-\omega\Delta_t}$, and assume $\sigma_{\phi_{ij},\infty}^2 > 0$ for all (i, j). Let $\Phi \in \mathbb{R}^{N \times N}$ be a centered, matrix-valued random variable in L_μ^2. Denote the matrices of all entry-wise variances and asymptotic variances by*

$$\Sigma_\Phi = \left(\sigma_{\phi_{ij}} \right)_{i,j=1}^N, \qquad\qquad \Sigma_{\Phi,\infty} = \left(\sigma_{\phi_{ij},\infty} \right)_{i,j=1}^N$$

Then, for any given $\delta > 0$, and $m \in \mathbb{N}$, we have with probability at least $1 - \delta$ that

$$\|\overline{\Phi}_m\|_F \le \frac{N}{\sqrt{m\delta}} \left[\|\Sigma_{\Phi,\infty}\|_F^2 + \frac{2q}{m(1-q)^2}\|\Sigma_\Phi\|_F^2 \right]^{1/2}. \tag{4}$$

For reversible systems, we obtain the simplified bound

$$\|\overline{\Phi}_m\|_F \le \frac{N}{\sqrt{m\delta}}\|\Sigma_{\Phi,\infty}\|_F. \tag{5}$$

Proof Noting that $[\overline{\Phi}_m]_{ij} = [\overline{\phi}_{ij}]_m$, the scalar Chebyshev inequality and the result of Lemma 6, yield for all (i, j) :

$$\mathbb{P}\left([\overline{\Phi}_m]_{ij}^2 \le \varepsilon^2 \right) \ge 1 - \frac{\sigma_{[\overline{\phi}_{ij}]_m}^2}{\varepsilon^2} = 1 - \frac{\frac{1}{m}[\sigma_{\phi_{ij},\infty}^2 - R_{\phi_{ij}}^m]}{\varepsilon^2}$$

$$\ge 1 - \frac{1}{m\varepsilon^2}\left[\sigma_{\phi_{ij},\infty}^2 + \frac{2\sigma_{\phi_{ij}}^2 q}{m(1-q)^2} \right].$$

The second term on the right-hand side does not exceed $\frac{\delta}{N^2}$ if

$$\varepsilon^2 \geq \frac{N^2}{m\delta} \left[\sigma_{\phi_{ij},\infty}^2 + \frac{2\sigma_{\phi_{ij}}^2 q}{m(1-q)^2} \right],$$

in other words, there is a set of trajectories of probability at least $1 - \frac{\delta}{N^2}$ such that

$$[\overline{\Phi}_m]_{ij}^2 \leq \frac{N^2}{m\delta} \left[\sigma_{\phi_{ij},\infty}^2 + \frac{2\sigma_{\phi_{ij}}^2 q}{m(1-q)^2} \right].$$

On the intersection of these sets, we have that

$$\|\overline{\Phi}_m\|_F \leq \frac{N}{\sqrt{m\delta}} \left[\|\Sigma_{\Phi,\infty}\|_F^2 + \frac{2q}{m(1-q)^2} \|\Sigma_\Phi\|_F^2 \right]^{1/2},$$

and the probability of the intersection is at least $1 - \delta$ by Lemma 22. In the reversible case, we know that $R_{\phi_{ij}}^m \geq 0$ for all (i, j), and therefore

$$\mathbb{P}\left([\overline{\Phi}_m]_{ij}^2 \leq \varepsilon^2 \right) \geq 1 - \frac{1}{m\varepsilon^2} \sigma_{\phi_{ij},\infty}^2. \tag{6}$$

Simplified bound (5) follows by repeating the above argument starting from this inequality. □

Remark 11 (I.i.d. sampling) If the data are sampled i.i.d., that is, Assumption 5.(iid) holds instead of Assumption 5.(erg), then by standard results, one has $\sigma_{\phi_m}^2 = \frac{1}{m}\sigma_\phi^2$. The bounds from Proposition 10 simplify significantly in this case. By the Chebyshev inequality:

$$\mathbb{P}\left([\overline{\Phi}_m]_{ij}^2 \leq \varepsilon^2 \right) \geq 1 - \frac{\frac{1}{m}\sigma_{\phi_{ij}}^2}{\varepsilon^2},$$

which leads to the following error estimate for fixed $m \in \mathbb{N}$ and $\delta > 0$:

$$\|\overline{\Phi}_m\|_F \leq \frac{N}{\sqrt{m\delta}} \|\Sigma_\Phi\|_F. \tag{7}$$

The setting of sampling via the Lebesgue measure on a compact set $\mathbb{X}$ was thoroughly considered in Zhang and Zuazua (2021).

2.2.2 Error Bound for the Projected Generator

Next, we deduce our first main result by applying the probabilistic bounds obtained in Proposition 10 to estimate the error for the data-driven Galerkin projection $\tilde{\mathcal{L}}_m$.

Theorem 12 (Approximation error: probabilistic bound) *Let Assumption 5 hold. Then, for any error bound $\tilde{\varepsilon} > 0$ and probabilistic tolerance $\tilde{\delta} \in (0, 1)$, we have*

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbb{V}} - \tilde{\mathcal{L}}_m\|_F \leq \tilde{\varepsilon} \right) \geq 1 - \tilde{\delta} \tag{8}$$

for any amount $m \in \mathbb{N}$ of data points such that the following hold with

$$\varepsilon = \min\left\{ 1, \frac{1}{\|A\|\|C^{-1}\|} \right\} \cdot \frac{\|A\|\tilde{\varepsilon}}{2\|A\|\|C^{-1}\| + \tilde{\varepsilon}} \quad and \quad \delta = \frac{\tilde{\delta}}{3}.$$

- *In case of ergodic sampling, i.e., Assumption 5.(erg),*

$$m \geq \frac{N^2}{\delta \varepsilon^2} \left[\|\Sigma_{\Phi,\infty}\|_F^2 + \frac{2q}{m(1-q)^2} \|\Sigma_\Phi\|_F^2 \right]$$

- *In case of ergodic sampling, i.e., Assumption 5.(erg), of a reversible system*

$$m \geq \frac{N^2}{\delta \varepsilon^2} \|\Sigma_{\Phi,\infty}\|_F^2.$$

- *In case of i.i.d. sampling, i.e., Assumption 5.(iid),*

$$m \geq \frac{N^2}{\delta \varepsilon^2} \|\Sigma_\Phi\|_F^2.$$

Proof In this proof, we will omit the subscript for the norm and set $\|\cdot\| = \|\cdot\|_F$. Let us introduce the centered matrix-valued random variables

$$\Phi_C(x) := \Psi(x)\Psi(x)^\top - C \quad \text{and} \quad \Phi_A(x) := \Psi(x)\mathcal{L}\Psi(x)^\top - A,$$

where $\Psi = [\psi_1, \ldots, \psi_N]^\top$. Then $\tilde{C}_m - C = [\overline{\Phi_C}]_m$ and $\tilde{A}_m - A = [\overline{\Phi_A}]_m$. Hence, we may apply Proposition 10 to these matrix-valued random variables. First, by the choice of m above we have

$$\mathbb{P}\left(\|C - \tilde{C}_m\| \leq \frac{R}{\|A\|\|C^{-1}\|} \right) \geq 1 - \frac{\tilde{\delta}}{3} \quad \text{and} \quad \mathbb{P}\left(\|\tilde{A}_m - A\| \leq R \right) \geq 1 - \frac{\tilde{\delta}}{3},$$

where

$$R := \frac{\|A\|\tilde{\varepsilon}}{2\|A\|\|C^{-1}\| + \tilde{\varepsilon}} = \frac{\tilde{\varepsilon}}{2\left(\|C^{-1}\| + \frac{\tilde{\varepsilon}}{2\|A\|} \right)}.$$

Moreover, we compute

$$\|\tilde{C}_m^{-1} - C^{-1}\| = \|\tilde{C}_m^{-1}(C - \tilde{C}_m)C^{-1}\| \leq \|C^{-1}\|\|C - \tilde{C}_m\| \left(\|\tilde{C}_m^{-1} - C^{-1}\| + \|C^{-1}\| \right)$$

which implies

$$\|\tilde{C}_m^{-1} - C^{-1}\| \leq \frac{\|C^{-1}\|^2 \|C - \tilde{C}_m\|}{1 - \|C^{-1}\| \|C - \tilde{C}_m\|}.$$

Hence, by straightforward computations we obtain

$$\mathbb{P}\left(\|\tilde{C}_m^{-1} - C^{-1}\| \leq \tfrac{\tilde{\varepsilon}}{2\|A\|}\right) \geq \mathbb{P}\left(\frac{\|C^{-1}\|^2 \|C - \tilde{C}_m\|}{1 - \|C^{-1}\| \|C - \tilde{C}_m\|} \leq \tfrac{\tilde{\varepsilon}}{2\|A\|}\right)$$

$$= \mathbb{P}\left(\|C - \tilde{C}_m\| \leq \frac{R}{\|A\| \|C^{-1}\|}\right) \geq 1 - \tfrac{\tilde{\delta}}{3}.$$

and

$$\mathbb{P}\left(\|\tilde{A}_m - A\| \leq \tfrac{\tilde{\varepsilon}}{2\|\tilde{C}_m^{-1}\|}\right) \geq \mathbb{P}\left(\|\tilde{A}_m - A\| \leq \frac{\tilde{\varepsilon}}{2\left(\|C^{-1}\| + \|\tilde{C}_m^{-1} - C^{-1}\|\right)}\right)$$

$$\geq \mathbb{P}\left(\|\tilde{A}_m - A\| \leq \frac{\tilde{\varepsilon}}{2\left(\|C^{-1}\| + \frac{\tilde{\varepsilon}}{2\|A\|}\right)} \ \wedge \ \|\tilde{C}_m^{-1} - C^{-1}\| \leq \tfrac{\tilde{\varepsilon}}{2\|A\|}\right)$$

$$\geq \left(1 - \tfrac{\tilde{\delta}}{3}\right) + \left(1 - \tfrac{\tilde{\delta}}{3}\right) - 1 = 1 - \tfrac{2\tilde{\delta}}{3}.$$

Thus, we conclude

$$\mathbb{P}(\|C^{-1}A - \tilde{C}_m^{-1}\tilde{A}_m\| \leq \tilde{\varepsilon}) = \mathbb{P}\left(\|\tilde{C}_m^{-1}(A - \tilde{A}_m) + \left(C^{-1} - \tilde{C}_m^{-1}\right) A\| \leq \tilde{\varepsilon}\right)$$

$$\geq \mathbb{P}\left(\|\tilde{C}_m^{-1}\| \|A - \tilde{A}_m\| + \|C^{-1} - \tilde{C}_m^{-1}\| \|A\| \leq \tilde{\varepsilon}\right)$$

$$\geq \mathbb{P}\left(\|A - \tilde{A}_m\| \leq \tfrac{\tilde{\varepsilon}}{2\|\tilde{C}_m^{-1}\|} \ \wedge \ \|C^{-1} - \tilde{C}_m^{-1}\| \leq \tfrac{\tilde{\varepsilon}}{2\|A\|}\right)$$

$$\geq (1 - \tfrac{2\tilde{\delta}}{3}) + (1 - \tfrac{\tilde{\delta}}{3}) - 1 = 1 - \tilde{\delta},$$

which is (8). □

A similar result as Theorem 12 was obtained for ODE systems in Zhang and Zuazua (2021) under the assumption that the data are drawn i.i.d.

An immediate consequence of the estimate on the generator approximation error is a bound on the error of the trajectories. To this end, consider the systems

$$\dot{z} = \mathcal{L}_{\mathbb{V}}z \qquad\qquad z(0) = z_0 \qquad\qquad (9)$$

$$\dot{\tilde{z}} = \tilde{\mathcal{L}}_m\tilde{z} \qquad\qquad \tilde{z}(0) = z_0. \qquad\qquad (10)$$

where $z_0 \in \mathbb{R}^n$, which represents an ODE in terms of the coefficients in the basis representation of elements of $\mathbb{V}$. We will leverage the error bound obtained in Theorem 12 to derive an estimate on the resulting prediction error in the observables, i.e.,

$\|z(t) - \tilde{z}(t)\|_2$. Note that in view of the isomorphism $\mathbb{V} \simeq \mathbb{R}^N$ this also directly translates to an error estimate for trajectories in $\mathbb{V}$.

Corollary 13 *Let Assumption 5 hold. Then for any $T > 0$ and $\delta, \varepsilon > 0$ there is $m_0 \in \mathbb{N}$ such that for $m \geq m_0$ data points we have*

$$\min_{t \in [0,T]} \mathbb{P}\big(\|z(t) - \tilde{z}(t)\|_2 \leq \varepsilon\big) \geq 1 - \delta.$$

Proof See "Appendix A.3."   □

A sufficient amount of data m_0 can be easily specified by combining the calculations displayed in the proof of Corollary 13, i.e., Gronwall's inequality and Condition (4). Under additional assumptions on the Koopman semigroup generated by $\mathcal{L}_{\mathbb{V}}$, e.g., stability, one can refine this estimate or render it uniform in T, cf. Corollary 24 in "Appendix A.3."

2.3 Error Bound for the Projected Koopman Operator

Similar to the derivation of the probabilistic bound on the projected generator, a bound on the Koopman operator is possible. We briefly sketch the main steps of the argumentation. Let $t = l\Delta_t$ for some $l \in \mathbb{N}$ and again choose a subspace $\mathbb{V} = \operatorname{span}\{\{\psi_j\}_{j=1}^N\} \subset L_\mu^2(\mathbb{X})$ (which, in contrast to the generator-based setting, is not required to be contained in the domain). The restricted Koopman operator on this subspace is defined via

$$\mathcal{K}_{\mathbb{V}}^t := P_{\mathbb{V}} \mathcal{K}^t|_{\mathbb{V}} = C^{-1} A,$$

where

$$C_{i,j} = \langle \psi_i, \psi_j \rangle_{L_\mu^2(\mathbb{X})} \quad \text{and} \quad A_{i,j} = \langle \psi_i, \mathcal{K}^t \psi_j \rangle_{L_\mu^2(\mathbb{X})}.$$

Define the data matrices

$$\Psi(X) := \left(\begin{pmatrix} \psi_1(x_0) \\ \vdots \\ \psi_N(x_0) \end{pmatrix} \middle| \cdots \middle| \begin{pmatrix} \psi_1(x_{m-l-1}) \\ \vdots \\ \psi_N(x_{m-l-1}) \end{pmatrix} \right)$$

$$\Psi(Y) := \left(\begin{pmatrix} \psi_1(x_l) \\ \vdots \\ \psi_N(x_l) \end{pmatrix} \middle| \cdots \middle| \begin{pmatrix} \psi_1(x_{m-1}) \\ \vdots \\ \psi_N(x_{m-1}) \end{pmatrix} \right).$$

The empirical estimator is then defined similarly to the generator setting via

$$\tilde{\mathcal{K}}_m^t = \tilde{C}_m^{-1} \tilde{A}_m$$

with

$$\tilde{C}_m = \frac{1}{m} \Psi(X) \Psi(X)^\top \quad \text{and} \quad \tilde{A}_m = \frac{1}{m} \Psi(X) \Psi(Y)^\top.$$

We now present the analogue to Theorem 12 for the Koopman operator which follows by straightforward adaptations of the results of Sect. 2.2.

Theorem 14 *Let Assumption 5 hold. Then, for $t \geq 0$, any error bound $\varepsilon > 0$ and probabilistic tolerance $\delta \in (0, 1)$ there is $m_0 \in \mathbb{N}$ such that for any $m \geq m_0$,*

$$\mathbb{P}\left(\|\mathcal{K}_{\mathbb{V}}^t - \tilde{\mathcal{K}}_m^t\|_F \leq \varepsilon \right) \geq 1 - \delta.$$

A sufficient amount of data m_0 can be specified analogously to Theorem 12.

3 Extension to Control Systems

In this section, we derive probabilistic bounds on the approximation error of nonlinear control-affine SDE systems of the form

$$\mathrm{d}X_t = \left(F(X_t) + \sum_{i=1}^{n_c} G_i(X_t)u_i \right) \mathrm{d}t + \sigma(X_t)\, \mathrm{d}W_t, \tag{11}$$

with input $u \in \mathbb{R}^{n_c}$ and state $X_t \in \mathbb{X}$, where $F : \mathbb{X} \to \mathbb{R}^n$ and $G_i : \mathbb{X} \to \mathbb{R}^n$, $i = 1, \ldots, n_c$, are locally Lipschitz-continuous vector fields. In the deterministic case $\sigma \equiv 0$ the controlled SDE reduces to the control-affine ODE system

$$\dot{x} = F(x) + \sum_{i=1}^{n_c} G_i(x)u_i. \tag{12}$$

We will describe how one can apply the bounds on the generators of autonomous (SDE) systems obtained in Sect. 2 in order to obtain bounds for prediction of control systems, either for i.i.d. or ergodic sampling. Again, we will analyze the error on the finite dictionary resulting from finitely many data points, depending on the chosen control variable. In the i.i.d. setting, we analyzed the projection error using a finite element dictionary in the follow-up work (Schaller et al. 2022). Further, also in Schaller et al. (2022), we derived uniform bounds for data requirements and dictionary size w.r.t. the control variable, assuming that the control is constrained to a compact subset.

Central in this part is the fact that the Koopman generators for control-affine systems are control-affine. More precisely, if $\mathcal{L}^{\bar{u}}$ denotes the Koopman generator for a control-affine system with constant control $\bar{u} \in \mathbb{R}^{n_c}$ and $\bar{u} = \sum_{i=1}^{r} \alpha_i \bar{u}_i$ is a linear combination of constant controls $\bar{u}_i \in \mathbb{R}^{n_c}$, we have

$$\mathcal{L}^{\bar{u}} = \mathcal{L}^0 + \sum_{i=1}^{n_c} \alpha_i \left(\mathcal{L}^{\bar{u}_i} - \mathcal{L}^0 \right). \tag{13}$$

This easily follows from representation (2) of the Koopman generator, see also Peitz et al. (2020, Theorem 3.2) for the special (deterministic) case $\sigma \equiv 0$.

We will utilize this property to invoke our results from Sect. 2 to approximate the Koopman generator corresponding to basis elements of the control space, that is, $\mathcal{L}^{e_i}$, $i = 1, \ldots, n_c$, and $\mathcal{L}^0$ corresponding to the drift term to form a bilinear control system in the observables.

Analogously to Assumption 5 we have the following two cases for the collected data and the underlying measure.

Assumption 15 Let either of the following hold:

(iid) The data for each autonomous system with control $u = e_i$, $i = 0, \ldots, n_c$, are sampled i.i.d. from either the normalized Lebesgue measure and contained in a compact set $\mathbb{X}$ or from an invariant measure μ_i in the sense of Definition 2.

(erg) The data for each autonomous system with control $u = e_i$, $i = 0, \ldots, n_c$, satisfy Assumption 5.(erg), i.e., is drawn from a single ergodic trajectory, the probability measure μ_i of the resulting autonomous SDE is invariant in the sense of Definition 2 and the Koopman semigroup is exponentially stable on $L^2_{\mu_i,0}(\mathbb{X})$.

It is important to note that in the first case of (iid), we did not make any assumption of invariance of the set $\mathbb{X}$ for all autonomous systems corresponding to the constant controls e_i, $i = 0, \ldots, n_c$, as this would be very restrictive. Hence, we have to ensure that the state trajectories remain (with probability one in stochastic setting (11)) in the set $\mathbb{X}$. Sufficient conditions are, e.g., controlled forward invariance of the set $\mathbb{X}$ or knowing that the initial condition is contained in a suitable sub-level set of the optimal value function of a respective optimal control problem, see, e.g., Boccia et al. (2014) or Esterhuizen et al. (2020) for an illustrative application of such a technique in showing recursive stability of Model Predictive Control (MPC) without stabilizing terminal constraints for discrete- and continuous-time systems, respectively.

In the following, we set $\mathcal{O}_i = L^2_{\mu_i}(\mathbb{X})$, $i = 1, \ldots, n_c$, and consider the generators $\mathcal{L}^{e_i}$ in these spaces, respectively. Further, let $\psi_1, \ldots, \psi_N : \mathbb{X} \to \mathbb{R}$ be N linearly independent observables whose span $\mathbb{V} = \text{span}\{\psi_1, \ldots, \psi_N\}$ satisfies

$$\mathbb{V} \subset D(\mathcal{L}^{e_0}) \cap D(\mathcal{L}^{e_1}) \cap \ldots \cap D(\mathcal{L}^{e_{n_c}}), \tag{14}$$

where e_i, $i = 1, \ldots, n_c$, denote the standard basis vectors of $\mathbb{R}^{n_c}$ and $e_0 := 0$. We now discuss two cases of sampling, one corresponding to the approach of Sect. 2 and one to the standard case of i.i.d. sampling as in Zhang and Zuazua (2021).

As the original system and the Koopman generator are control affine, the remainder of this section is split up into two parts. First, we derive error estimates corresponding to autonomous systems driven by $n_c + 1$ constant controls. Second, we use these estimates and control affinity to deduce a result for general controls.

In accordance with the notation in Sect. 2 we define $\mathcal{L}^{e_i}_{\mathbb{V}} := P_{\mathbb{V}} \mathcal{L}^{e_i}|_{\mathbb{V}}$ and also use this symbol to denote the matrix representation of this linear operator w.r.t. to the basis $\{\psi_1, \ldots, \psi_N\}$ of $\mathbb{V}$. Its approximation based on the data $x_0, \ldots, x_{m-1} \in \mathbb{X}$ will be denoted by $\tilde{\mathcal{L}}^{e_i}_m$.

Proposition 16 *Let $i \in \{0, \ldots, n_c\}$ be given and Assumption 15 hold. Then, for any pair consisting of a desired error bound $\varepsilon > 0$ and a probabilistic tolerance $\delta \in (0, 1)$,*

there are a number of data points m_i such that for any $m \geq m_i$, we have the estimate

$$\mathbb{P}\big(\|\mathcal{L}_{\mathbb{V}}^{e_i} - \tilde{\mathcal{L}}_m^{e_i}\|_F \leq \varepsilon\big) \geq 1 - \delta.$$

The minimal amount of data m_i is given by the formulas of Theorem 12.

Proof The claim follows immediately from applying Theorem 12. □

Having obtained an estimate for the autonomous systems corresponding to the constant controls e_i, $i = 0, \ldots n_c$, we can leverage the control affinity of the system to formulate the corresponding results for arbitrary controls. To this end, for any control $u(t) = \sum_{i=1}^{n_c} \alpha_i(t) e_i \in L^\infty(0, T; \mathbb{R}^{n_c})$, we define the projected Koopman generator and its approximation corresponding to the nonautonomous system with control u by

$$\mathcal{L}_{\mathbb{V}}^u(t) := \mathcal{L}_{\mathbb{V}}^0 + \sum_{i=1}^{n_c} \alpha_i(t)\big(\mathcal{L}_{\mathbb{V}}^{e_i} - \mathcal{L}_{\mathbb{V}}^0\big),$$

$$\tilde{\mathcal{L}}_m^u(t) := \tilde{\mathcal{L}}_m^0 + \sum_{i=1}^{n_c} \alpha_i(t)\big(\tilde{\mathcal{L}}_m^{e_i} - \tilde{\mathcal{L}}_m^0\big).$$

Theorem 17 *Let Assumption 15 hold. Then, for any pair consisting of a desired error bound $\tilde{\varepsilon} > 0$ and probabilistic tolerance $\tilde{\delta} \in (0, 1)$, prediction horizon $T > 0$, and control function $u \in L^\infty(0, T; \mathbb{R}^{n_c})$ we have*

$$\operatorname{ess\,inf}_{t \in [0,T]} \mathbb{P}\big(\|\mathcal{L}_{\mathbb{V}}^u(t) - \tilde{\mathcal{L}}_m^u(t)\|_F \leq \tilde{\varepsilon}\big) \geq 1 - \tilde{\delta},$$

provided that the number m of data points exceeds $\max_{i=0,\ldots,n_c} m_i$ with m_i defined as in Proposition 16 with

$$\varepsilon = \frac{\tilde{\varepsilon}}{(n_c+1)\big(1+\sum_{i=1}^{n_c} \|\alpha_i\|_{L^\infty(0,T)}\big)} \quad and \quad \delta = 1 - \frac{\tilde{\delta}}{n_c+1}.$$

Proof Again, we omit the subscript of the norm and set $\|\cdot\| = \|\cdot\|_F$. Using the result of Proposition 16 and our choice of m_0, we have

$$\mathbb{P}\left(\|\tilde{\mathcal{L}}_m^0 - \mathcal{L}_{\mathbb{V}}^0\| \leq \frac{\tilde{\varepsilon}}{(n_c+1)\big(1+\sum_{i=1}^{n_c} \|\alpha_i\|_{L^\infty(0,T)}\big)}\right) \geq 1 - \frac{\tilde{\delta}}{n_c+1},$$

and for all $i \in 1, \ldots n_c$

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbb{V}}^{e_i} - \tilde{\mathcal{L}}_m^{e_i}\| \leq \frac{\tilde{\varepsilon}}{(n_c+1)\|\alpha_i\|_{L^\infty(0,T)}}\right) \geq 1 - \frac{\tilde{\delta}}{n_c+1}.$$

Then we compute for *a.e.* $t \in [0, T]$

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbb{V}}^u(t) - \tilde{\mathcal{L}}_m^u(t)\| \leq \tilde{\varepsilon}\right)$$

$$\geq \mathbb{P}\left(\left\|\left(1-\sum_{i=1}^{n_c}\alpha_i(t)\right)\left(\mathcal{L}_{\mathbb{V}}^0-\tilde{\mathcal{L}}_m^0\right)\right\|+\sum_{i=1}^{n_c}\left\|\alpha_i(t)\left(\tilde{\mathcal{L}}_m^{e_i}-\mathcal{L}_{\mathbb{V}}^{e_i}\right)\right\|\leq\tilde{\varepsilon}\right)$$

$$\geq \mathbb{P}\left(\left\|\left(1-\sum_{i=1}^{n_c}\alpha_i(t)\right)\left(\mathcal{L}_{\mathbb{V}}^0-\tilde{\mathcal{L}}_m^0\right)\right\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\wedge\bigwedge_{i=1}^{n_c}:\left\|\alpha_i(t)\left(\tilde{\mathcal{L}}_m^{e_i}-\mathcal{L}_{\mathbb{V}}^{e_i}\right)\right\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\right).$$

Next, we use Lemma 22 from "Appendix A.2" with $d = n_c + 1$,

$$A_0 = \left\{\left\|\left(1-\sum_{i=1}^{n_c}\alpha_i(t)\right)\left(\mathcal{L}_{\mathbb{V}}^0-\tilde{\mathcal{L}}_m^0\right)\right\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\right\}\quad\text{and}\quad A_i=\left\{\left\|\alpha_i(t)\left(\tilde{\mathcal{L}}_m^{e_i}-\mathcal{L}_{\mathbb{V}}^{e_i}\right)\right\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\right\}$$

for $i = 1, \ldots, n_c$. This yields

$$\mathbb{P}\left(\|\mathcal{L}_{\mathbb{V}}^u(t)-\tilde{\mathcal{L}}_m^u(t)\|\leq\tilde{\varepsilon}\right)$$

$$\geq \mathbb{P}\left(\left\|\left(1-\sum_{i=1}^{n_c}\alpha_i(t)\right)\left(\mathcal{L}_{\mathbb{V}}^0-\tilde{\mathcal{L}}_m^0\right)\right\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\right)$$

$$+\sum_{i=1}^{n_c}\mathbb{P}\left(\|\alpha_i(t)(\tilde{\mathcal{L}}_m^{e_i}-\mathcal{L}_{\mathbb{V}}^{e_i})\|\leq\frac{\tilde{\varepsilon}}{n_c+1}\right)-n_c$$

$$\geq \mathbb{P}\left(\|\tilde{\mathcal{L}}_m^0-\mathcal{L}_{\mathbb{V}}^0\|\leq\frac{\tilde{\varepsilon}}{(1+\sum_{i=1}^{n_c}\|\alpha_i\|_{L^\infty(0,T)})(n_c+1)}\right)$$

$$+\sum_{i=1}^{n_c}\mathbb{P}\left(\|\tilde{\mathcal{L}}_m^{e_i}-\mathcal{L}_{\mathbb{V}}^{e_i}\|\leq\frac{\tilde{\varepsilon}}{(n_c+1)\|\alpha_i\|_{L^\infty(0,T)}}\right)-n_c$$

$$\geq 1-\frac{\tilde{\delta}}{n_c+1}+\sum_{i=1}^{n_c}\left(1-\frac{\tilde{\delta}}{n_c+1}\right)-n_c=1-\tilde{\delta}.$$

Taking the essential infimum yields the result.   □

In the previous result of Theorem 17, the data requirements depend on the chosen control. If the values of the control function are constrained to a compact subset, one can derive uniform data requirements w.r.t. the control, cf. our follow-up work (Schaller et al. 2022). Finally, similar as in the previous section, we obtain a bound on trajectories via Gronwall, if the state response is contained in $\mathbb{X}$.

Corollary 18 *Let Assumption 15 hold. Let $T, \varepsilon > 0$ and $\delta \in (0,1)$, $z_0 \in \mathbb{R}^N$ and $u \in L^\infty(0, T; \mathbb{R}^{n_c})$ such that the solution of (SDE) is contained in $\mathbb{X}$ with probability one. Then there is $m_0 \in \mathbb{N}$ such that for $m \geq m_0$ the solutions $z, \tilde{z}$ of*

$$\dot{z}(t) = \mathcal{L}_{\mathbb{V}}^u(t)z \qquad\qquad z(0) = z_0$$
$$\dot{\tilde{z}}(t) = \tilde{\mathcal{L}}_m^u(t)\tilde{z} \qquad\qquad \tilde{z}(0) = z_0$$

satisfy

$$\min_{t \in [0,T]} \mathbb{P}\big(\|z(t) - \tilde{z}(t)\|_2 \leq \varepsilon\big) \geq 1 - \delta.$$

Proof See "Appendix A.3." □

As in Corollary 13, m_0 can explicitly be computed by combining Theorem 17 with the constants in Gronwall's inequality.

We conclude this section with a final corollary regarding the optimality of the solution obtained using an error-certified Koopman model. To this end, we consider the optimal control problem with $x_0 \in \mathbb{X}$ and a stage cost $\ell : \mathbb{R}^n \times \mathbb{R}^{n_c} \to \mathbb{R}$:

$$\min_{u \in L^\infty(0,T;\mathbb{R}^{n_c})} \int_0^T \ell(x(t), u(t))\, \mathrm{d}t$$
$$\text{s.t.} \quad \dot{x} = F(x) + \sum_{i=1}^{n_c} G_i(x)u_i, \qquad x(0) = x_0. \tag{15}$$

In what follows, we compare the optimal value of the Koopman representation of (15) projected onto the subspace of observables $\mathbb{V}$ with initial datum $z_0 = \Psi(x_0)$

$$\min_{\alpha \in L^\infty(0,T;\mathbb{R}^{n_c})} \int_0^T \ell(P(z(t)), \alpha(t))\, \mathrm{d}t$$
$$\text{s.t.} \quad \dot{z}(t) = \left[\mathcal{L}_{\mathbb{V}}^0 + \sum_{i=1}^{n_c} \alpha_i(t) \left(\mathcal{L}_{\mathbb{V}}^{e_i} - \mathcal{L}_{\mathbb{V}}^0 \right) \right] z(t), \qquad z(0) = z_0, \tag{16}$$

to the optimal value of the surrogate-based control problem:

$$\min_{\tilde{\alpha} \in L^\infty(0,T;\mathbb{R}^{n_c})} \int_0^T \ell(P(\tilde{z}(t)), \tilde{\alpha}(t))\, \mathrm{d}t$$
$$\text{s.t.} \quad \dot{\tilde{z}}(t) = \left[\tilde{\mathcal{L}}_m^0 + \sum_{i=1}^{n_c} \tilde{\alpha}_i(t) \left(\tilde{\mathcal{L}}_m^{e_i} - \tilde{\mathcal{L}}_m^0 \right) \right] \tilde{z}(t), \qquad \tilde{z}(0) = z_0, \tag{17}$$

where P maps a trajectory of observables to a trajectory in the state space, which in practice is frequently realized by including the coordinates of the identity function in the dictionary Ψ of observables.

Corollary 19 *Let $T, \varepsilon > 0$, $\delta \in (0, 1)$, $z_0 \in \mathbb{R}^N$, let J be locally Lipschitz continuous and let Assumption 15 hold. Furthermore, let (z^*, α^*) be an optimal solution of problem (16) such that the state response of (15) emanating from the control α^* is contained in $\mathbb{X}$. Then there is $m_0 \in \mathbb{N}$ such that for $m \geq m_0$ data points contained in $\mathbb{X}$, there exists a tuple $(\tilde{z}, \tilde{\alpha})$ which is feasible for (17) such that for the cost, we have the estimate*

$$\mathbb{P}\left(\left| \int_0^T \ell(P(\tilde{z}(t)), \tilde{\alpha}(t)) - \ell(P(z^*(t)), \alpha^*(t))\, \mathrm{d}t \right| \leq \varepsilon \right) \geq 1 - \delta.$$

4 Numerical Examples

In this section, we first present numerical experiments on the derived error bound for the Koopman generator and then discuss the implications for optimal control. In particular, we emphasize that the bilinear Koopman model from Sect. 3 appears to be the best approach for a straightforward transfer of predictive error bounds to the control setting.

4.1 Generator Error Bounds: Ornstein–Uhlenbeck Process

We begin by investigating the validity and accuracy of the error bounds for the Galerkin matrices of a single SDE system, as derived in Proposition 10. To this end, we consider the one-dimensional reversible Ornstein–Uhlenbeck (OU) process

$$\mathrm{d}X_t = -X_t \mathrm{d}t + \mathrm{d}W_t. \tag{18}$$

As the spectrum of the generator $\mathcal{L}$ of the OU process, as well as its invariant density, is known in analytical form, we can exactly calculate the Galerkin matrices C, A, all variances $\sigma^2_{\Phi_{ij}}$, and asymptotic variances $\sigma^2_{\Phi_{ij},\infty}$, if we consider a basis set comprised of monomials, see "Appendix A.4."

We consider monomials of maximal degree four (i.e., $N = 4$) and set the discrete integration time step to $\Delta_t = 10^{-3}$. For a range of different data sizes m and confidence levels δ, we estimate the minimal error ε that can be achieved with probability $1 - \delta$ for a variety of quantities of interest. We calculate ε for all individual entries C_{ij} and A_{ij} using inequality (6). Moreover, we also calculate ε for the Frobenius norm errors in C and A by means of (5).

In order to compare our bound to the real error, we conduct 500 identical experiments. For each experiment, we generate an independent simulation of OU process (18), with initial condition drawn from the invariant distribution. For each trajectory and each of the data sizes m considered, we estimate the matrices $\tilde{C}_m$, $\tilde{A}_m$. We then calculate the absolute entry-wise errors to C and A, as well as the Frobenius norm errors $\|\tilde{C}_m - C\|_F$ and $\|\tilde{A}_m - A\|_F$. Finally, we numerically compute the $1 - \delta$-percentile of each of these errors for all confidence levels δ considered above (i.e., the error ε below which 450 of the 500 repeated experiments lie). These can be directly compared to the probabilistic bounds ε obtained from our theoretical estimates.

The results are shown in Fig. 1. We can see in panels B and C that our estimates for individual entries of the Galerkin matrices C and A are quite accurate, as the data-based error is over-estimated by only a factor of two to three. Our estimates for Frobenius norm errors are less accurate, with approximately one order of magnitude difference between theoretical and data-based errors. It can be concluded that the factor N in (5) is too coarse in this example, as the actual Frobenius norm error only marginally exceeds the maximal entry-wise error. Nevertheless, the qualitative behavior of all theoretical error bounds is confirmed by the data.
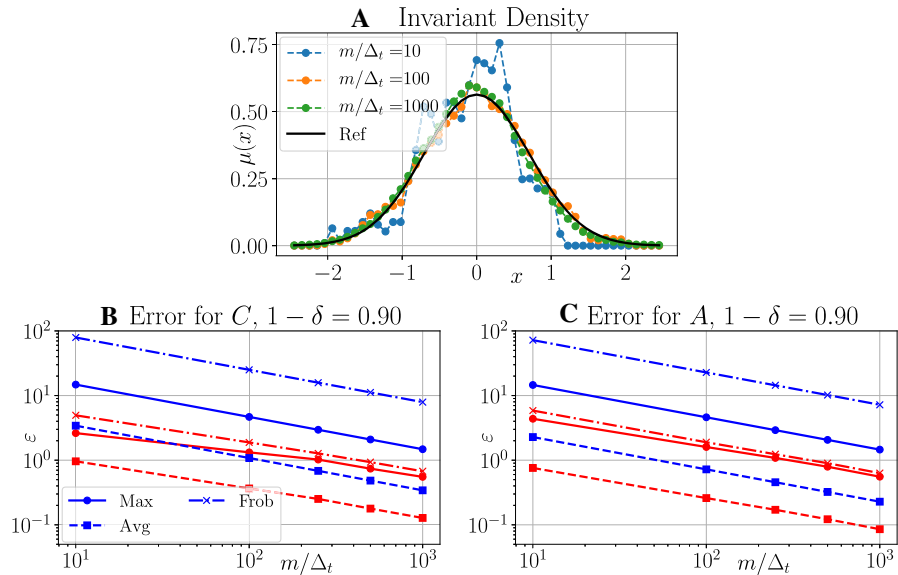
Fig. 1 Numerical Results for one-dimensional OU Process (18). A: Exact invariant density μ in black, compared to histograms of the first m points of an exemplary trajectory, for various data sizes m. B: Error bounds for C corresponding to confidence level $1 - \delta = 0.9$. We show both the theoretical estimates obtained in Proposition 10 (blue), as well as the data-based estimates obtained as described in the text (red). We show the maximal error over all entries C_{ij} (dots), the average error over all matrix entries (squares), and the Frobenius norm errors $\|\tilde{C}_m - C\|_F$. C: The same as B for the matrix A

4.2 Extension to Control Systems

In this section, we illustrate our findings for deterministic as well as stochastic systems regarding prediction and control. We compare the solution of the exact model to the bilinear system

$$
\begin{aligned}
\dot{z}(t) &= \left[\tilde{\mathcal{L}}_m^0 + \sum_{i=1}^{n_c} u_i(t) \left(\tilde{\mathcal{L}}_m^{e_i} - \tilde{\mathcal{L}}_m^0 \right) \right] z(t) \\
z(t_0) &= \psi(x(t_0)),
\end{aligned}
\tag{19}
$$

where n_c is the dimension of the control input u. Besides bilinear model (19), we also compare the true solution to the linear model obtained via eDMD with control, see Proctor et al. (2016); Korda and Mezić (2018a) for details. Optimality of the computed trajectories from a theoretical standpoint will not be addressed here, as the error bounds for $\tilde{\mathcal{L}}_m$ are still too large. However, the principled approach is to choose an m such that Corollary 19 holds.

For the numerical discretization, we use eDMD with a finite lag time to obtain a discrete-time version of (19) in case of the Duffing system, which corresponds to an explicit Euler discretization (Peitz et al. 2020). For the Ornstein–Uhlenbeck example, we calculate the generator using gEDMD (Klus et al. 2020) and then obtain

the resulting discrete-time version by taking the matrix exponential. In the case of eDMD with control, we use the standard algorithm from Korda and Mezić (2018a), which also results in a forward Euler version of the linear system $\dot{z} = \hat{A}z + \hat{B}u$, i.e.,

$$
\begin{aligned}
z_{i+1} &= Az_i + Bu_i, \\
z_0 &= \psi(x(t_0)).
\end{aligned}
\tag{20}
$$

Remark 20 Note that one can drastically improve the predictive accuracy—in particular for longer time horizons—by introducing an intermediate *project-and-lift* step, which only makes a difference if the space $\mathbb{V}$ spanned by the $\{\{\psi_k\}_{k=1}^N\}$ is not a *Koopman-invariant subspace* (Proctor et al. 2018). Moreover, it becomes less and less important the more the dynamics of the $\tilde{\mathcal{L}}_m$ are truly restricted to $\mathbb{V}$, or—alternatively— if we are not interested in long-term predictions, for instance in the MPC setting.

Considering this intermediate step, the bilinear discrete-time systems become

$$
\begin{aligned}
\widehat{z}_i &= \psi(P(z_i)) \\
z_{i+1} &= \left(K_0 + \sum_{j=1}^{n_c} (K_j - K_0) u_{j,i} \right) \widehat{z}_i \\
z_0 &= \psi(x(t_0)),
\end{aligned}
\tag{21}
$$

where P is the projection of the lifted state z onto the full state $x \in \mathbb{X}$. In the same manner, the DMDc model reads

$$
\begin{aligned}
\widehat{z}_{i+1} &= \psi(P(z_{i+1})) \\
z_{i+1} &= A\widehat{z}_i + Bu_i, \\
z_0 &= \psi(x(t_0)).
\end{aligned}
\tag{22}
$$

However, this comes at the cost of losing the bilinear or linear structure of the DMD-based models, respectively.

4.2.1 Duffing Equation (ODE)

The first system we study is the Duffing oscillator:

$$
\frac{dx}{dt} = \begin{pmatrix} x_2 \\ -\delta x_2 - \alpha x_1 u - 2\beta x_1^3 \end{pmatrix}, \quad x(t_0) = x_0.
\tag{23}
$$

with $\alpha = -1$, $\beta = 1$ and $\delta = 0$. Note that the control does not enter linearly, which is a well-known challenge for DMDc (Peitz et al. 2020).

As the dictionary ψ, we choose monomials with varying maximal degrees, and we also include square and cubic roots for comparison. For the data collection process, we simulate the system with constant control inputs $u = 0$ and $u = 1$ using the standard Runge–Kutta scheme of fourth order with time step $h = 0.005$. As the final time, we choose $T = n_{lag}h$ seconds, where n_{lag} is the integer number of time steps we step
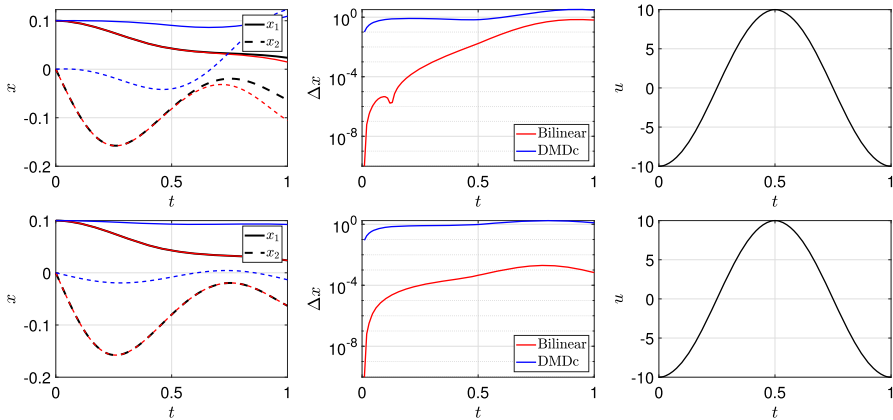
Fig. 2 Comparison of ODE solution, the bilinear surrogate model and the linear model obtained via eDMDc for system (23). Top row shows systems (19) and (20); bottom row uses project-and-lift versions (21) and (22)

forward by the discrete-time Koopman operator model. We perform experiments for both $n_{lag} = 1$ and $n_{lag} = 10$. Each trajectory yields one tuple $(x, y) = (x(0), x(T))$, and we sample various numbers m of data points with uniformly distributed random initial conditions over the rectangle $[-1.5, 1.5]^2$.

Figure 2 shows the prediction accuracy for $m = 10000$ and $n_{lag} = 1$, where excellent agreement is observed for the bilinear surrogate model. In particular the relative error

$$\Delta x(t) = \frac{\|x(t) - \tilde{x}(t)\|_2}{\|x(t)\|_2},$$

where $\tilde{x}(t) = P(z(t))$ is the solution obtained via the surrogate model, is below 1 percent for the first second (i.e., 200 steps), whereas the eDMDc approach has a large error from the start.

To study the influence of the size of the training data set, Fig. 3 shows boxplots of the one-step prediction accuracy for various m. Each boxplot was obtained by performing 20 trainings of a bilinear system according to the procedure described above. After each training, a single time step was made for 1000 initial conditions $x_0 \in [-1.5, 1, 5]^2$ and control inputs $u \in [0, 1]$, both drawn uniformly. Consequently, each boxplot consists of $2 \cdot 10^4$ data points. We see that, as expected, the training error decreases for larger m. However, what is really surprising is that a saturation can be observed already at $m = 30$ for an ODE system. Beyond that, no further improvement can be seen, which demonstrates the advantage of (i) the linearity of the Koopman approach and (ii) the usage of autonomous systems for the model reduction process.

Interestingly, the lag time between two consecutive data points has a critical impact on the maximal accuracy in the control case. This is due to the fact that the bilinear surrogate model is only exact for the Koopman generator (Peitz et al. 2020). For a finite lag time, the bilinear model is a first-order approximation such that smaller lag
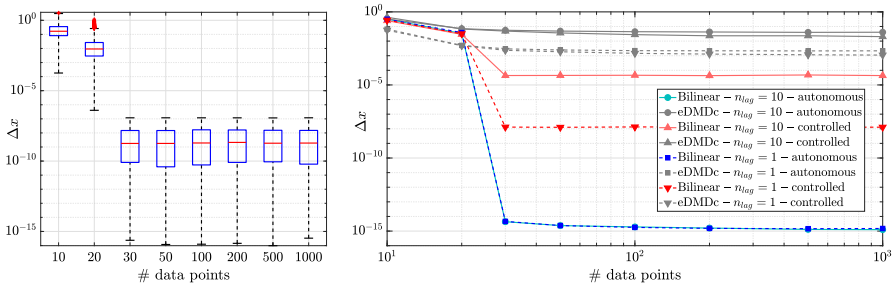
Fig. 3 Left: Boxplot of the relative one-step prediction error over 20 training runs and 1000 different samples (x_0, u) in each run for a dictionary of monomials up to degree at most five and $n_{lag} = 1$. Right: The influence of the lag time as well as the control input on the mean accuracy (the dashed line with triangle symbols corresponds to the left plot). We see that the lag time plays an important role in the control setting
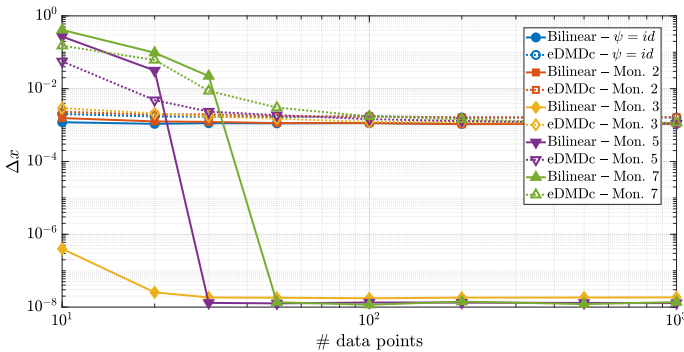
Fig. 4 Mean relative one-step prediction errors for various dictionaries and data set sizes m

times are advantageous. Nevertheless, the accuracy still significantly supersedes the eDMDc approach.

Another interesting observation can be made with respect to the choice of the dictionary ψ. Figure 4 shows a comparison of the mean errors (analogous to the red bars in Fig. 3 for various dictionaries. We observe excellent performance for monomials with degree three or larger. The addition of roots of x is not beneficial at all, and in particular, smaller dictionaries are favorable in terms of the data requirements, which is in agreement with our error analysis and which was also reported in Peitz and Klus (2020).

Next, we study the stabilization of system (23) for $h = 0.01$ and the final time $T = 1.5$. Using the time discretization as above and a straight-forward single-shooting method, this yields a 150-dimensional optimization problem similar to Problem (17) from Corollary 19:

$$\min_u \int_0^5 \| P(z(t)) - x^{\text{ref}}(t) \|^2 \tag{24}$$

$$\text{s.t.} \quad (19)$$
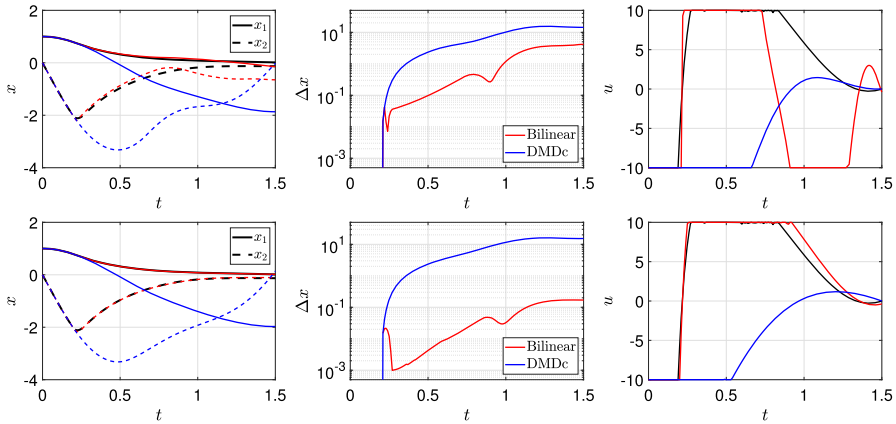
Fig. 5 Open-loop control performance (stabilization of the origin) using the true ODE model as well as the bilinear and DMDc surrogate models. Top row shows systems (19) and (20); bottom row uses project-and-lift versions (21) and (22)

where x^{ref} is the reference trajectory to be tracked. Figure 5 demonstrates the performance for $x^{\text{ref}} = 0$ with models using $M = 5$ and only $m = 200$ training samples – 100 for each model in the bilinear setting and 200 for eDMDc. We see very good performance for the bilinear system even without the intermediate projection step. In contrast, the eDMDc approximation fails for System (23), even when initializing with the optimal solution from the full system.
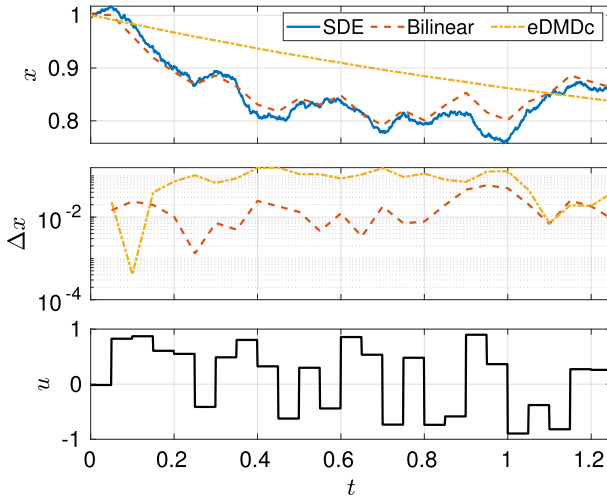
4.2.2 Ornstein–Uhlenbeck Process (SDE)

For the stochastic setting, we consider an Ornstein–Uhlenbeck process with a control input:

$$\mathrm{d}X_t = -\alpha(uX_t)\mathrm{d}t + \sqrt{2\beta^{-1}}\mathrm{d}W_t. \tag{25}$$

with $\alpha = 1$, $\beta = 2$ and $u(t) \in [0, 1]$. The system is simulated numerically using an Euler-Maruyama integration scheme with a time step of 10^{-3} as in Sect. 4.1. For both systems, we calculate the Koopman operator corresponding to $u = 0$ and $u = 1$, respectively, using the gEDMD procedure presented in Klus et al. (2020) with monomials up to degree five. We then calculate the corresponding Koopman operators for the time step $h = 0.05$ using the matrix exponential.

To study the prediction performance (cf. Fig. 6), we proceed in the same way as for the Duffing system, except that we now compare the expected values, approximated by averaging over 100 SDE simulations. The results are very similar to the deterministic case, where the performance of both surrogate modeling techniques is comparable when the control enters linearly, and very poor for eDMDc otherwise. Even though the Ornstein–Uhlenbeck process is stochastic, the linearity is highly favorable for the data requirements. We do not observe any considerable deterioration even in the very low data limit.

Fig. 6 Prediction accuracy for the expected value of the Ornstein–Uhlenbeck process (approximated by averaging over 100 simulations) of the bilinear system and eDMDc, respectively
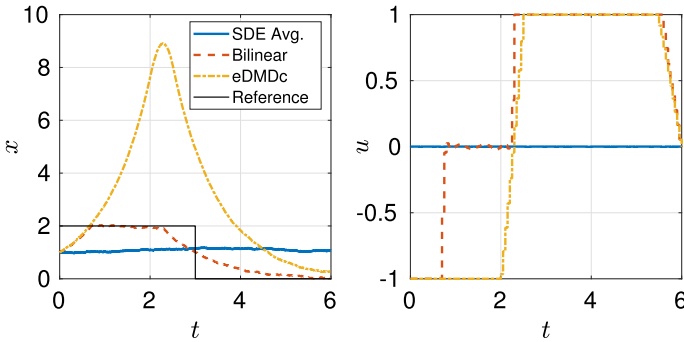
Fig. 7 Control of the expected value of the Ornstein–Uhlenbeck process (approximated by averaging over 100 simulations using the optimal control input shown in the bottom plots). In the SDE-based control, we have used 20 simulations in each objective function evaluation

Finally, in the control setting, we aim at tracking the expected value $\mathbb{E}[X_t]$, which is precisely the quantity that is predicted by the Koopman operator. Thus, Problem (24) can directly be applied to SDEs as well. In order to compare the results to the full system, we average over 20 simulations in the evaluation of the objective function value when using the SDE. However, this appears to be insufficient, as the performance is inadequate, cf. Fig. 7. The bilinear surrogate model, on the other hand, shows very good performance with a small amount of $m = 100$ training data points.

5 Conclusions

We presented the first rigorously derived probabilistic bounds on the finite-data approximation error for the Koopman generator of SDEs and nonlinear control systems. Furthermore, by using slightly more advanced techniques from probability theory, we also relaxed the assumption of i.i.d. data invoked in Zhang and Zuazua (2021) in the ODE setting. Moreover, we also provided an analysis for the error propagation to estimate the prediction accuracy in terms of the data size. A novelty for SDEs and in the control setting is that our bounds explicitly depend on the number of data points (and not only in the infinite-data limit). Further, the proposed techniques provide the theoretical foundation for the Koopman-based approach (Peitz et al. 2020) to control-affine systems, which seems to be superior for control and particularly well-suited for MPC, since it avoids the curse of dimensionality w.r.t. the control dimension. In future work, we will focus on the application of the derived bounds in a optimal and predictive control, in particular in combination with the recently obtained (control-uniform) projection error bounds of our follow-up work (Schaller et al. 2022).

Appendix

A.1 Norm of the Isomorphism $\mathbb{V} \simeq \mathbb{R}^n$

Proposition 21 *Let* $\mathbb{V} = \mathrm{span}\{\{\psi_j\}_{j=1}^N\} \subset L_\mu^2(\mathbb{X})$, $\mathcal{B} \in L(\mathbb{V}, \mathbb{V})$ *and* $B \in \mathbb{R}^{n \times n}$ *be its corresponding matrix representation. Then*

$$\sqrt{\frac{\lambda^{\min}(C)}{\lambda^{\max}(C)}}\|B\|_2 \leq \|\mathcal{B}\|_{L(\mathbb{V},\mathbb{V})} \leq \sqrt{\frac{\lambda^{\max}(C)}{\lambda^{\min}(C)}}\|B\|_2$$

where $C_{i,j} = \langle \psi_i, \psi_j \rangle_{L_\mu^2(\mathbb{X})}$.

Proof This follows from the identity

$$\left\|\sum_{i=1}^n \alpha_i \psi_i\right\|_{L_\mu^2(\mathbb{X})}^2 = \sum_{i,j=1}^N \alpha_i \alpha_j \langle \psi_i, \psi_j \rangle_{L_\mu^2(\mathbb{X})} = \alpha^\top C \alpha,$$

which shows the equivalence of the vector norms. This induces the desired equivalence of the operator norms.   □

A.2 A Technical Lemma

Lemma 22 *Let A_i, $i = 1, \ldots, d$, be measurable sets. Then*

$$\mathbb{P}\left(\bigcap_{i=1}^{d} A_i\right) = \sum_{i=1}^{d} \mathbb{P}(A_i) - \sum_{i=1}^{d-1} \mathbb{P}\left(A_i \cup \bigcap_{j=i+1}^{d} A_j\right).$$

Moreover, if $\mathbb{P}(A_i) \geq 1 - \delta$ for all $i = 1, \ldots, d$, then

$$\mathbb{P}\left(\bigcap_{i=1}^{d} A_i\right) \geq 1 - d\delta.$$

Proof Inductively applying the classical formula

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cup A_2)$$

yields

$$\mathbb{P}\left(\bigcap_{i=1}^{d} A_i\right) = \mathbb{P}\left(A_1 \cap \bigcap_{i=2}^{d} A_i\right) = \mathbb{P}(A_1) + \mathbb{P}\left(\bigcap_{i=2}^{d} A_i\right) - \mathbb{P}\left(A_1 \cup \bigcap_{i=2}^{d} A_i\right)$$

$$= \sum_{i=1}^{d} \mathbb{P}(A_i) - \sum_{i=2}^{d-1} \mathbb{P}\left(A_i \cup \bigcap_{j=i+1}^{d} A_j\right) - \mathbb{P}\left(A_1 \cup \bigcap_{i=2}^{d} A_i\right)$$

$$= \sum_{i=1}^{d} \mathbb{P}(A_i) - \sum_{i=1}^{d-1} \mathbb{P}\left(A_i \cup \bigcap_{j=i+1}^{d} A_j\right),$$

which proves the first claim. The second claim follows by estimating the first sum by $d(1 - \delta)$ from below and the second sum by $-(d - 1)$ from below. □

A.3 Proof of the Error Bound on the Trajectories

Lemma 23 *Let z and $\tilde{z}$ solve (9) and (10), respectively. Then for all $t \geq 0$*

$$\|z(t) - \tilde{z}(t)\|_2 \leq \|\tilde{\mathcal{L}}_m - \mathcal{L}_\mathbb{V}\|_2 \|\tilde{z}\|_{L^1(0,t;\mathbb{R}^N)} e^{t\|\mathcal{L}_\mathbb{V}\|_2}$$

Proof Denoting $e = z - \tilde{z}$, subtracting (10) from (9) and integrating over a time interval $[0, t]$ with $t \geq 0$ we obtain that

$$
\begin{aligned}
e(t) &= \int_0^t \mathcal{L}_{\mathbb{V}} z(s) - \tilde{\mathcal{L}}_m \tilde{z}(s) \, \mathrm{d}s \\
&= \int_0^t \mathcal{L}_{\mathbb{V}} e(s) - \left(\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}} \right) \tilde{z}(s) \, \mathrm{d}s
\end{aligned}
$$

This implies using Gronwall's inequality, cf. Chicone (2006, Theorem 2.1), that

$$
\begin{aligned}
\|e(t)\|_2 &\leq \int_0^t \|\mathcal{L}_{\mathbb{V}}\|_2 \|e(s)\|_2 + \|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}(s)\|_2 \, \mathrm{d}s \\
&\leq e^{t\|\mathcal{L}_{\mathbb{V}}\|_2} \int_0^t \|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}(s)\|_2 \, ds \\
&= e^{t\|\mathcal{L}_{\mathbb{V}}\|_2} \|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}\|_{L^1(0,t;\mathbb{R}^N)}.
\end{aligned}
$$

$\square$

Proof (Corollary 13) Using the bound of Lemma 23 we obtain

$$
\begin{aligned}
\|z(t) - \tilde{z}(t)\|_2 &\leq \|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 t e^{t\|\tilde{\mathcal{L}}_m\|_2} e^{t\|\mathcal{L}_{\mathbb{V}}\|_2} \\
&= t\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 e^{t\left(\|\mathcal{L}_{\mathbb{V}}\|_2 + \|\tilde{\mathcal{L}}_m\|_2\right)}.
\end{aligned}
$$

We compute

$$
\begin{aligned}
\mathbb{P} &\left(\|z(t) - \tilde{z}(t)\|_2 \leq \varepsilon \right) \\
&\geq \mathbb{P}\left(t\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 e^{t\|\mathcal{L}_{\mathbb{V}}\|_2} e^{t\|\tilde{\mathcal{L}}_m\|_2} \|z_0\| \leq \varepsilon \right) \\
&\geq \mathbb{P}\left(t\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 e^{2t\|\mathcal{L}_{\mathbb{V}}\|_2} e^{t\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2} \|z_0\| \leq \varepsilon \right) \\
&\geq \mathbb{P}\left(T\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 e^{2T\|\mathcal{L}_{\mathbb{V}}\|_2} e^{T\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2} \|z_0\| \leq \varepsilon \right)
\end{aligned}
$$

By Theorem 12 and $\| \cdot \|_2 \leq \| \cdot \|_F$, for any $\tilde{\varepsilon}$ we can choose m_0 such that $\mathbb{P}\left(\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \leq \tilde{\varepsilon} \right) \geq 1 - \delta$. Hence, there is m_0 only depending on T, z_0, $\mathcal{L}_{\mathbb{V}}$ and ε such that for any $t \geq 0$

$$
\mathbb{P}\left(\|z(t) - \tilde{z}(t)\|_2 \leq \varepsilon \right) \geq 1 - \delta.
$$

Taking the minimum over all $t \in [0, T]$ proves the claim.

Proof (Corollary 18) This proof follows with obvious modifications in the proof of Corollary 23 using the bound on then error of the time-dependent generators of Theorem 17.

Corollary 24 *If the Koopman semigroup generated by $\mathcal{L}_{\mathbb{V}}$ is bounded by M, then*

$$\|\tilde{z}(t) - z(t)\|_2 \leq M\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}\|_{L^1(0,t;\mathbb{R}^N)}.$$

If it is exponentially stable then

$$\|\tilde{z}(t) - z(t)\|_2 \leq Mc\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}\|_{L^p(0,t;\mathbb{R}^N)}$$

for any $1 \leq p \leq \infty$ with $M \geq 1$ and $c = c(p) \geq 0$ independent of t. If additionally the semigroup generated by $\tilde{\mathcal{L}}_m$ is exponentially stable, $\|\tilde{z}(t) - z(t)\|_2$ can be bounded uniformly in $t \geq 0$.

Proof Subtracting (10) from (9) and denoting $e(t) = \tilde{z}(t) - z(t)$ yields the system

$$\dot{e}(t) = \mathcal{L}_{\mathbb{V}}e(t) + (\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}})\tilde{z}(t).$$

Denoting by $\mathcal{K}_{\mathbb{V}}^t$ the Koopman semigroup generated by $\mathcal{L}_{\mathbb{V}}$ yields, using the variation of constants formula

$$e(t) = \int_0^t \mathcal{K}_{\mathbb{V}}^{t-s}\left(\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\right)\tilde{z}(s)\,\mathrm{d}s$$

and hence

$$\|e(t)\|_2 \leq \int_0^t \|\mathcal{K}_{\mathbb{V}}^{t-s}\|_2 \|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}(s)\|_2\,\mathrm{d}s.$$

If $\mathcal{K}_{\mathbb{V}}^t$ is bounded by M, i.e., $\|\mathcal{K}_{\mathbb{V}}^t\| \leq M$, we have

$$\|e(t)\|_2 \leq M\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}\|_{L^1(0,t;\mathbb{R}^N)}.$$

If $\mathcal{K}_{\mathbb{V}}^t$ is exponentially stable, i.e., $\|\mathcal{K}_{\mathbb{V}}^t\|_2 \leq Me^{-\omega t}$, we obtain

$$\|e(t)\|_2 \leq Mc\|\tilde{\mathcal{L}}_m - \mathcal{L}_{\mathbb{V}}\|_2 \|\tilde{z}\|_{L^p(0,t;\mathbb{R}^N)}$$

for any $1 \leq p \leq \infty$ with $c = c(p, \omega)$. If additionally, the semigroup generated by $\tilde{\mathcal{L}}_m$ is exponentially stable implying that $\|\tilde{z}(t)\|_2 \leq \tilde{M}e^{-\tilde{\omega}t}\|z_0\|_2$, this upper bound can be bounded uniformly in t. $\square$

A.4 Analytical Expressions for the OU Process

For one-dimensional SDE (18), the Koopman generator is given by:

$$\mathcal{L}\phi = -x\phi'(x) + \frac{1}{2}\phi''(x).$$

The eigenvalues of the generator are given by negative integers $\kappa_l = -l$, eigenvalues of the Koopman operator are their exponentials, as usual, $\lambda_l(t) = e^{-lt}$. The corresponding eigenfunctions are given by scaled physicist's Hermite polynomials. They are orthonormal with respect to the inner product with weight function μ, which is the density of a normal distribution with variance one half, yielding the relations:

$$\mu(x) = \frac{1}{\sqrt{\pi}} \exp(-x^2), \quad \psi_l = \frac{1}{\sqrt{2^l(l-1)!}} H_l(x), \quad \langle H_l, H_m \rangle_\mu = 2^l l! \delta_{lm}. \quad (26)$$

The monomial basis can be recovered from eigenfunction basis ψ_i by the representation formula:

$$x^n = \frac{n!}{2^n} \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} \frac{1}{k!(n-2k)!} H_{n-2k}(x). \quad (27)$$

For a basis set comprised of monomials up to maximal degree N, the Galerkin matrices C and A can be obtained as the moments of the normal distribution with variance 0.5:

$$C_{ij} = \begin{cases} \frac{(i+j)!}{2^{i+j}((i+j)/2)!} & (i+j) \text{ even,} \\ 0 & (i+j) \text{ odd,} \end{cases} \qquad A_{ij} = \begin{cases} -\frac{ij}{2} \frac{(i+j-2)!}{2^{i+j-2}((i+j-2)/2)!} & (i+j) \text{ even,} \\ 0 & (i+j) \text{ odd.} \end{cases}$$

For their numerical estimation, we consider centered random variables:

$$\phi_{ij}(x) = x^i x^j - C_{ij} \quad \text{for } C, \qquad \phi_{ij}(x) = -\frac{ij}{2} x^{i-1} x^{j-1} - A_{ij} \quad \text{for } A.$$

We calculate the asymptotic variance of the scalar random variable ϕ_{ij} if it is defined by either of the two expressions above. We also introduce the quantity $n := i + j$ for C or $n := i + j - 2$ for A. The analytical expressions for C_{ij}, A_{ij} above exactly equal the terms corresponding to H_0 in the general expansion for the monomial x^n in (27). As the random variables ϕ_{ij} are centered, no contribution from H_0 is left. Thereby, we obtain the decomposition for ϕ_{ij} (up to the factor $-\frac{ij}{2}$ for estimation of A):

$$\phi_{ij} = x^n - \mathbb{E}^\mu[x^n] = \frac{n!}{2^n} \sum_{k=0}^{\lceil \frac{n}{2}-1 \rceil} \frac{1}{k!(n-2k)!} H_{n-2k}(x). \quad (28)$$

Next, we calculate matrix elements with the Koopman operator at lag time $l\Delta_t$ by combining (28) with orthogonality relation (26):

$$
\begin{aligned}
\langle \phi_{ij}, \mathcal{K}^{l\Delta_t}\phi_{ij}\rangle_\mu &= \frac{(n!)^2}{2^{2n}}\sum_{k=0}^{\lceil \frac{n}{2}-1\rceil} \frac{1}{(k!(n-2k)!)^2} e^{-(n-2k)l\Delta_t} 2^{n-2k}(n-2k)! \\
&= \frac{(n!)^2}{2^{2n}}\sum_{k=0}^{\lceil \frac{n}{2}-1\rceil} \frac{2^{n-2k}}{(k!)^2(n-2k)!} e^{-(n-2k)l\Delta_t}.
\end{aligned}
$$

Finally, by setting $q_k = e^{-(n-2k)\Delta_t}$, we calculate the asymptotic variance according to the result in Lemma 6 (note that the contribution for $l=0$ appears only once, and that the result needs to be multiplied by $\frac{1}{4}ij$ for the estimation of A):

$$
\begin{aligned}
\sigma^2_{\phi_{ij},\infty} &= \langle \phi_{ij}, \phi_{ij}\rangle_\mu + 2\sum_{l=1}^\infty \langle \phi_{ij}, \mathcal{K}^{l\Delta_t}\phi_{ij}\rangle_\mu \\
&= \frac{(n!)^2}{2^{2n}}\sum_{k=0}^{\lceil \frac{n}{2}-1\rceil} \frac{2^{n-2k}}{(k!)^2(n-2k)!}\left[\sum_{l=0}^\infty q_k^l + \sum_{l=1}^\infty q_k^l\right] \\
&= \frac{(n!)^2}{2^{2n}}\sum_{k=0}^{\lceil \frac{n}{2}-1\rceil} \frac{2^{n-2k}}{(k!)^2(n-2k)!}\left[\frac{1}{1-q_k} + \frac{q_k}{1-q_k}\right] \\
&= \frac{(n!)^2}{2^{2n}}\sum_{k=0}^{\lceil \frac{n}{2}-1\rceil} \frac{2^{n-2k}}{(k!)^2(n-2k)!}\frac{1+q_k}{1-q_k}.
\end{aligned}
$$

References

Bakry, D., Gentil, I., Ledoux, M.: Analysis and geometry of Markov diffusion operators, vol. 348. Springer Science & Business Media, Berlin (2013)

Beck, A., Schwartz, J.T.: A vector-valued random ergodic theorem. Proc. Am. Math. Soc. **8**(6), 1049–1059 (1957)

Boccia, A., Grüne, L., Worthmann, K.: Stability and feasibility of state constrained MPC without stabilizing terminal constraints. Syst. Control Lett. **72**(8), 14–21 (2014)

Bruder, D., Fu, X., Vasudevan, R.: Advantages of bilinear Koopman realizations for the modeling and control of systems with unknown dynamics. IEEE Robot. Autom. Lett. **6**(3), 4369–4376 (2021)

Brunton, S.L., Budišić, M., Kaiser, E., Kutz, J.N.: Modern Koopman theory for dynamical systems. SIAM Rev. **64**(2), 229–340 (2022)

Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. **113**, 3932–3937 (2016)

Chicone, C.: Ordinary Differ. Equ. Appl., vol. 34. Springer Science & Business Media, New York (2006)

Duncan, A.B., Lelievre, T., Pavliotis, G.A.: Variance reduction using nonreversible Langevin samplers. J. Stat. Phys. **163**(3), 457–491 (2016)

Esterhuizen, W., Worthmann, K., Streif, S.: Recursive feasibility of continuous-time model predictive control without stabilising constraints. IEEE Control Syst. Lett. **5**(1), 265–270 (2020)

Goswami, D., Paley, D.A.: Global bilinearization and controllability of control-affine nonlinear systems: a Koopman spectral approach. In 2017 IEEE 56th Annual Conference on Decision and Control (CDC), pp. 6107–6112. IEEE (2017)

Kaiser, E., Kutz, J.N., Brunton, S.L.: Data-driven discovery of Koopman eigenfunctions for control. Mach. Learn. Sci. Technol. **2**, 035023 (2021)

Klus, S., Koltai, P., Schütte, C.: On the numerical approximation of the Perron–Frobenius and Koopman operator. J. Comput. Dyn. **3**(1), 51–79 (2016)

Klus, S., Nüske, F., Peitz, S.: Koopman analysis of quantum systems. J. Phys. A. **31**, 314002 (2022)

Klus, S., Nüske, F., Hamzi, B.: Kernel-based approximation of the Koopman Generator and Schrödinger Operator. Entropy **22**(7), 722 (2020)

Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F.: Data-driven model reduction and transfer operator approximation. J. Nonlinear Sci. **28**(3), 985–1010 (2018)

Klus, S., Nüske, F., Peitz, S., Niemann, J.-H., Clementi, C., Schütte, C.: Data-driven approximation of the Koopman generator: model reduction, system identification, and control. Phys. D **406**, 132416 (2020)

Koopman, B.O.: Hamiltonian systems and transformations in Hilbert space. Proc. Natl. Acad. Sci. **17**(5), 315–318 (1931)

Korda, M., Mezić, I.: Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control. Automatica **93**, 149–160 (2018)

Korda, M., Mezić, I.: On convergence of extended dynamic mode decomposition to the Koopman operator. J. Nonlinear Sci. **28**(2), 687–710 (2018)

Kurdila, A.J., Bobade, P.: Koopman theory and linear approximation spaces. arXiv:1811.10809 (2018)

Lelièvre, T., Stoltz, G.: Partial differential equations and stochastic methods in molecular dynamics. Acta Numer. **25**, 681–880 (2016)

Lu, H., Tartakovsky, D.M.: Predictive accuracy of dynamic mode decomposition. SIAM J. Sci. Comput. **42**(3), 1639–1662 (2020)

Lu, H., Tartakovsky, D.M.: Extended dynamic mode decomposition for inhomogeneous problems. J. Comput. Phys. **444**, 110550 (2021)

Lu, Q., Shin, S., Zavala, V.M.: Characterizing the predictive accuracy of dynamic mode decomposition for data-driven control. IFAC-PapersOnLine **53**(2), 11289–11294 (2020). 21th IFAC World Congress

Mamakoukas, G., Castano, M.L., Tan, X., Murphey, T.D.: Derivative-based koopman operators for real-time control of robotic systems. IEEE Trans. Robot. **37**(6), 2173–2192 (2021)

Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. Nonlinear Dyn. **41**, 309–325 (2005)

Mollenhauer, M., Klus, S., Schütte, C., Koltai, P.: Kernel autocovariance operators of stationary processes: estimation and convergence. arXiv:2004.00891, (2020)

Oksendal, B.: Stochastic Differential Equations: An Introduction with Applications. Springer Science & Business Media, New York (2013)

Peitz, S., Bieker, K.: On the universal transformation of data-driven models to control systems. Automatica arXiv:2021.04722, (2021)

Peitz, S., Klus, S.: Koopman operator-based model reduction for switched-system control of PDEs. Automatica **106**, 184–191 (2019)

Peitz, S., Klus, S.: Feedback control of nonlinear PDEs using data-efficient reduced order models based on the Koopman operator. In: Mauroy, A., Mezić, I., Suzuki, Y. (eds.) The Koopman Operator in Systems and Control: Concepts, Methodologies and Applications, pp. 257–282. Springer, Cham (2020)

Peitz, S., Otto, S.E., Rowley, C.W.: Data-driven model predictive control using interpolated Koopman generators. SIAM J. Appl. Dyn. Syst. **19**(3), 2162–2193 (2020)

Proctor, J.L., Brunton, S.L., Kutz, J.N.: Dynamic mode decomposition with control. SIAM J. Appl. Dyn. Syst. **15**(1), 142–161 (2016)

Proctor, J.L., Brunton, S.L., Kutz, J.N.: Generalizing Koopman Theory to allow for inputs and control. SIAM J. Appl. Dyn. Syst. **17**(1), 909–930 (2018)

Risken, H.: The Fokker–Planck Equation. Springer, Berlin (1996)

Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.S.: Spectral analysis of nonlinear flows. J. Fluid Mech. **641**, 115–127 (2009)

Schaller, M., Worthmann, K., Nüske, F., Peitz, S., Philipp, F.: Towards reliable data-based optimal and predictive control using extended DMD. IFAC-PapersOnLine arXiv:2202.09084, (2022)

van Breugel, F., Kutz, J.N., Brunton, B.W.: Numerical differentiation of noisy data: a unifying multi-objective optimization framework. IEEE Access **8**, 196865–196877 (2020)

Webber, R.J., Thiede, E.H., Dow, D., Dinner, A.R., Weare, J.: Error bounds for dynamical spectral estimation. SIAM J. Math. Data Sci. **3**(1), 225–252 (2021)

Williams, M.O., Kevrekidis, I.G., Rowley, C.W.: A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. J. Nonlinear Sci. **25**(6), 1307–1346 (2015)

Zhang, C., Zuazua, E.: A quantitative analysis of Koopman operator methods for system identification and predictions. (2021) arXiv:hal-0327.8445