PAMM

WILEY-VCH

SPECIAL ISSUE 94th Annual Meeting of the International Association of Applied Mathematics and Mechanics (GAMM)

PAMM Proceedings in Applied Mathematics and Mechanics

RESEARCH ARTICLE OPEN ACCESS

Invariance of Gaussian RKHSs Under Koopman Operators of Stochastic Differential Equations With Constant Matrix Coefficients

Friedrich M. Philipp[1] | Manuel Schaller[1] | Karl Worthmann[1] | Sebastian Peitz[2] | Feliks Nüske[3,4]

[1]Optimization-Based Control Group, Institute of Mathematics, Technische Universität Ilmenau, Ilmenau, Germany | [2]Department of Computer Science, Paderborn University, Paderborn, Germany | [3]Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany | [4]Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany

Correspondence: Friedrich M. Philipp (friedrich.philipp@tu-ilmenau.de)

Received: 22 May 2024 | **Revised:** 18 November 2024 | **Accepted:** 19 November 2024

Funding: This research was supported by the Carl Zeiss Foundation within the project *DeepTurb—Deep Learning in and from Turbulence*, the free state of Thuringia and the German Federal Ministry of Education and Research (BMBF) within the project *THInKI—Thüringer Hochschulinitiative für KI im Studium*, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project ID: 507037103.

ABSTRACT

We consider the Koopman operator semigroup $(K^t)_{t\geq 0}$ associated with stochastic differential equations of the form $dX_t = AX_t\,dt + B\,dW_t$ with constant matrices A and B and Brownian motion W_t. We prove that the reproducing kernel Hilbert space $\mathbb{H}_C$ generated by a Gaussian kernel with a positive definite covariance matrix C is invariant under each Koopman operator K^t if the matrices A, B, and C satisfy the following Lyapunov-like matrix inequality: $AC^2 + C^2A^\top \leq 2BB^\top$. In this course, we prove a characterization concerning the inclusion $\mathbb{H}_{C_1} \subset \mathbb{H}_{C_2}$ of Gaussian RKHSs for two positive definite matrices C_1 and C_2. The question of whether the sufficient Lyapunov-condition is also necessary is left as an open problem.

1 | Introduction

The Koopman operator [1] of a (stochastic) dynamical system is a linear operator, which is defined on a linear function space of so-called *observables*. For deterministic systems, it is simply defined as a composition operator with the flow F of the system, that is, $(K^t f)(x) = f(F(x,t))$. For stochastic systems, $(K^t f)(x)$ is defined as the conditional expectation of $f(F(\cdot,t))$, given that the trajectory starts at x. Since the (generally unknown) Koopman operator is linear and provides complete information about the (expected) process behavior, it is very appealing to approximate this operator using sample data in the fashion of modern machine learning, see, for example, the monograph [2] and the references therein.

One of the most popular methods for Koopman operator learning is certainly Extended Dynamic Mode Decomposition (EDMD) [3], a data-driven approach propagating a finite number of predefined observable functions along the flow, which results in a well-interpretable surrogate model for analysis, prediction, and control. EDMD has been successfully applied in a number of highly relevant applications such as molecular dynamics [4], turbulent flows [5], neuroscience [6], or climate prediction [7], just to name a few. Rigorous error analyses for EDMD have been conducted in [8–11].

Since kernel methods play an important role in approximation theory and machine learning, it is not surprising that there exist kernel-based approaches for Koopman operator learning, one of them being a variant of EDMD, called kernel EDMD (kEDMD) [12, 13]. Kernel-based methods have been rigorously analyzed in, for example, [14–17] and in [18] for deterministic systems, where uniform error bounds have been provided. As it turns out in the analysis, for providing error bounds on Koopman approximations in the kernel-based setting, it is highly beneficial if the reproducing kernel Hilbert space (RKHS) generated by the kernel at hand is invariant under the Koopman operator. Therefore, an important task in kernel-based Koopman operator learning is to figure out which kernels and which systems are a good match in the sense that the generated RKHS is invariant under the Koopman operator of the system.

In this article, we consider the invariance of Gaussian RKHSs under the Koopman operator of dynamical systems driven by a stochastic differential equation (SDE) of the form $dX_t = AX_t\,dt + BdW_t$ with constant coefficient matrices A and B, where (A, B) is controllable and A is Hurwitz. To be more precise, we prove that the RKHS $\mathbb{H}_C$ generated by the Gaussian kernel $k(x, y) = \exp(-\|C^{-1}(x - y)\|^2)$ with a positive definite matrix C is Koopman-invariant if the matrices A, B, and C satisfy the Lyapunov-like inequality $AC^2 + C^2A^\top \le 2BB^\top$, see Theorem 2.1. It is left open whether this sufficient condition is also necessary. In the course of proving Theorem 2.1, we also prove that $\mathbb{H}_{C_1} \subset \mathbb{H}_{C_2}$ holds for two positive definite matrices C_1 and C_2 if and only if $C_1^2 \ge C_2^2$, see Proposition 3.3. This fact is well known [19] for the scalar case, where $C_1, C_2 \in (0, \infty)$, but—to the best of the authors' knowledge—seems to be unknown in the general matrix case.

The paper is arranged as follows. In the next section, we briefly introduce the reader to the Koopman operator of SDEs and present the main result, Theorem 2.1. In Section 3, we consider Gaussian kernels and their corresponding RKHSs and make use of the auxiliary results therein in Section 4, which is dedicated to the proof of Theorem 2.1.

2 | Setting and Main Result

Recall that a pair of matrices (A, B) with $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times m}$, $m \le d$, is called *controllable* if $(B, AB, \dots, A^{d-1}B)$ has full rank d. The notion is due to the fact that a controlled system $\dot{x}(t) = Ax(t) + Bu(t)$ can be steered from any state $x(0) = x_0$ in any time $t > 0$ to any state $z \in \mathbb{R}^d$ by a control function u if and only if (A, B) is controllable. Also recall that a square matrix A is called *Hurwitz* if the real parts of the eigenvalues of A are all negative. This implies that $\|e^{At}\| \le Me^{-\omega t}$ for all $t \ge 0$ with some constants $M, \omega > 0$.

Let us consider a d-dimensional SDE with constant matrix coefficients

$$dX_t = AX_t\,dt + B\,dW_t, \tag{1}$$

where W_t is m-dimensional Brownian motion, $m \le d$, and $A \in \mathbb{R}^{d\times d}$, $B \in \mathbb{R}^{d\times m}$. Here, we assume that (A, B) is controllable and that A is Hurwitz. The solution process X_t emerging from the SDE (1) is also called a d-dimensional *Ornstein–Uhlenbeck process*. It has the unique stationary distribution

$$d\mu(x) = (2\pi)^{-d/2}(\det\Sigma)^{-1/2}\exp(-\tfrac{1}{2}x^\top\Sigma^{-1}x)\,dx,$$

where $\Sigma \in \mathbb{R}^{d\times d}$ is the positive definite matrix

$$\Sigma = \int_0^\infty e^{As}BB^\top e^{A^\top s}\,ds.$$

Note that the limit Σ exists since A is Hurwitz and that Σ is positive definite thanks to controllability of (A, B). In linear control theory, the matrix Σ is called the *controllability Gramian* and is a solution of the matrix equation $AX + XA^\top = -BB^\top$.

The solution process X_t is a time-homogeneous Markov process. Recall that every such process has a so-called Markov transition kernel $\rho_t(x, F) = \mathbb{P}(X_t \in F \mid X_0 = x)$, where $x \in \mathbb{R}^d$ and F is a Borel set. Here, the transition kernel is absolutely continuous w.r.t. Lebesgue measure, and its density is given by the following (see [20]):

$$\rho_t(x, dy) = (2\pi)^{-d/2}(\det\Sigma(t))^{-1/2} \times \exp\left(-\tfrac{1}{2}(y - e^{At}x)^\top\Sigma(t)^{-1}(y - e^{At}x)\right)dy,$$

where

$$\Sigma(t) = \int_0^t e^{As}BB^\top e^{A^\top s}\,ds.$$

Again, $\Sigma(t)$ is positive definite as (A, B) is controllable.

The *Koopman operator* $K^t : L^2(\mu) \to L^2(\mu)$ at time $t \ge 0$ corresponding to the SDE (1) is defined by the following:

$$(K^t f)(x) := \mathbb{E}[f(X_t) \mid X_0 = x] = \int_{\mathbb{R}^d} f(y)\,\rho_t(x, dy), \quad f \in L^2(\mu). \tag{2}$$

It is well known (see, e.g., [14, Proposition 2.8]) that $(K^t)_{t\ge0}$ forms a strongly continuous semigroup of contractions on $L^2(\mu)$.

Here, we consider Gaussian kernels on $\mathbb{R}^d$ with a positive definite covariance matrix $C \in \mathbb{R}^{d\times d}$, that is,

$$k^C(x, y) = \exp\left[-(x - y)^\top C^{-2}(x - y)\right] = \exp\left[-\|C^{-1}(x - y)\|^2\right].$$

Note that for $C = \sigma \cdot I_d$, $\sigma > 0$, we obtain the usual RBF kernel $k^\sigma(x, y) = \exp(-\|x - y\|^2/\sigma^2)$. By $\mathbb{H}_C$, we denote the RKHS generated by the kernel k^C. The Hilbert space norm on $\mathbb{H}_C$ will be denoted by $\|\cdot\|_C$. For $\sigma > 0$, we simply write $\mathbb{H}_\sigma$ and $\|\cdot\|_\sigma$ instead of $\mathbb{H}_{\sigma I}$ and $\|\cdot\|_{\sigma I}$, respectively.

The main result of this note reads as follows.

Theorem 2.1. *Let $C \in \mathbb{R}^{d\times d}$ be a symmetric positive definite matrix such that*

$$\tfrac{1}{2}\left(AC^2 + C^2A^\top\right) \le BB^\top. \tag{3}$$

Then the RKHS $\mathbb{H}_C$ is invariant under each Koopman operator K^t, $t \ge 0$, associated with the SDE (1), and we have the following:

$$\|K^t f\|_C \le e^{t\cdot\mathrm{Tr}(-A)/2} \cdot \|f\|_C, \qquad f \in \mathbb{H}_C.$$

Let us briefly discuss the Lyapunov-like condition (3) on C.

Remark 2.2.

a. Since A is assumed to be Hurwitz, for every positive definite matrix $Q \in \mathbb{R}^{d\times d}$, there exists a positive definite matrix $P \in \mathbb{R}^{d\times d}$ such that $AP + PA^\top = -Q$. Hence, the set of positive definite matrices C satisfying the matrix inequality (3) is not empty. For example, $C = \alpha\Sigma^{1/2}$ is a solution for every $\alpha > 0$.

b. If $m = d$ and B is invertible, then for any given positive definite matrix C, there exists $\tau > 0$ such that τC satisfies Equation (3). Indeed, if this was not the case, there would exist a sequence $(x_n) \subset \mathbb{R}^d$ such that $\frac{1}{n}\langle (AC^2 + C^2A^\top)x_n, x_n\rangle > \|B^\top x_n\|^2$. It is no restriction to assume $\|x_n\| = 1$ and $x_n \to x$ as $n \to \infty$ for some $x \in \mathbb{R}^d$, $\|x\| = 1$. But then, the above implies $B^\top x = 0$ and hence $x = 0$, a contradiction.

c. It is not clear whether the condition (3) is also necessary for Koopman invariance of $\mathbb{H}_C$. We discuss this at the end of Section 4.

Corollary 2.3. *Let $\sigma > 0$ and assume that $\frac{1}{2}(A + A^\top) \le \frac{1}{\sigma^2}BB^\top$ (e.g., if A is dissipative[1]). Then the RKHS $\mathbb{H}_\sigma$ is invariant under each Koopman operator K^t, $t \ge 0$, associated with the SDE (1), and we have the following:*

$$\|K^t f\|_\sigma \le e^{t\cdot \mathrm{Tr}(-A)/2}\|f\|_\sigma, \qquad f \in \mathbb{H}_\sigma.$$

We conclude this section by tying up some notations. First of all, we agree on the convention that a positive definite kernel is always symmetric, that is, $k(x,y) = k(y,x)$, and real-valued. Also, the term "positive definite" only refers to real, symmetric matrices. Let X be a set. For $x \in X$ and a positive definite kernel k on X, set $k_x(y) := k(x,y)$, $y \in X$. That is, $x \mapsto k_x$ is the canonical feature map of the kernel k. For two positive definite kernels k_1 and k_2 on X, we write $k_1 \preceq k_2$ if

$$\sum_{i,j=1}^n \alpha_i\alpha_j k_1(x_i,x_j) \le \sum_{i,j=1}^n \alpha_i\alpha_j k_2(x_i,x_j)$$

for any choice of $n \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, and $x_j \in X$, $j = 1,\dots,n$. We also write $V \overset{c}{\hookrightarrow} W$ (continuous embedding) for two normed vector spaces V and W to indicate that $V \subset W$ and the identity map from V to W is continuous, that is, if there exists $K > 0$ such that $\|v\|_W \le K\|v\|_V$ for all $v \in V$.

3 | Some Properties of Gaussian Kernels

In this section, we collect the statements on Gaussian kernels that are utilized in the proof of Theorem 2.1. The next lemma shows that $\mathbb{H}_C$ can be regarded as a dense subspace of $L^2(\mu)$ and that the $\mathbb{H}_C$-norm is stronger than the $L^2(\mu)$-norm on $\mathbb{H}_C$.

Lemma 3.1. *Let $C \in \mathbb{R}^{d\times d}$ be positive definite. Then the RKHS $\mathbb{H}_C$ is densely and continuously embedded in $L^2(\mu)$.*

Proof. The fact that $\mathbb{H}_C \overset{c}{\hookrightarrow} L^2(\mu)$ is the first part of [21, Theorem 4.26]. Moreover, if $f \in \mathbb{H}_C$ and $f = 0$ μ-a.e., then $f(x) = 0$ for all $x \in \mathbb{R}^d$ as f is continuous and the density $d\mu/dx$ is positive. Therefore, we may regard $\mathbb{H}_C$ as a subspace of $L^2(\mu)$. Concerning the density of $\mathbb{H}_C$ in $L^2(\mu)$, we have to show that the integral operator $K : L^2(\mu) \to \mathbb{H}_C$, defined by the following:

$$Kf(x) = \int_{\mathbb{R}^d} k^C(x,y)f(y)\,d\mu(y), \qquad f \in L^2(\mu),$$

has trivial kernel in $L^2(\mu)$, see [21, Theorem 4.26]. So, let $f \in L^2(\mu)$ such that $Kf = 0$, that is,

$$(\phi * \psi)(x) = \int_{\mathbb{R}^d} \underbrace{\exp(-\|C^{-1}(x-y)\|^2)}_{=\phi(x-y)} \cdot \underbrace{f(y)\exp(-\|Py\|^2)}_{=\psi(y)}\,dy = 0, \qquad x \in \mathbb{R}^d,$$

where $P = (2\Sigma)^{-1/2}$. Applying the Fourier transform to $\phi * \psi$, we obtain $\widehat{\phi}(\omega)\cdot\widehat{\psi}(\omega) = 0$ for a.e. $\omega \in \mathbb{R}^d$. But $\widehat{\phi}(\omega) > 0$ for all $\omega \in \mathbb{R}^d$, hence $\widehat{\psi} = 0$ and thus $\psi = 0$. □

The following lemma is well-known. We will apply it in the proof of Proposition 3.3 below, which can be seen as a generalization of Lemma 3.2 to the multidimensional case.

Lemma 3.2. *For $\sigma > 0$, consider the Gaussian kernel k^σ on $\mathbb{R}$. If $\sigma_1 \ge \sigma_2 > 0$, then $\mathbb{H}_{\sigma_1} \overset{c}{\hookrightarrow} \mathbb{H}_{\sigma_2}$ with $\|f\|_{\sigma_2} \le (\frac{\sigma_1}{\sigma_2})^{1/2}\|f\|_{\sigma_1}$ for $f \in \mathbb{H}_{\sigma_1}$, and $k^{\sigma_1} \preceq \frac{\sigma_1}{\sigma_2}k^{\sigma_2}$.*

Proof. The first part is [19, Corollary 6], and the second follows from Aronszajn's inclusion theorem, Theorem A1. □

Proposition 3.3. *Let $C_1, C_2 \in \mathbb{R}^{d\times d}$ be positive definite. Then $\mathbb{H}_{C_1} \overset{c}{\hookrightarrow} \mathbb{H}_{C_2}$ if and only if $C_1^2 \ge C_2^2$. In this case, we have the following:*

$$\|f\|_{C_2} \le (\tfrac{\det C_1}{\det C_2})^{1/2}\|f\|_{C_1}, \qquad f \in \mathbb{H}_{C_1},$$

and $k^{C_1} \preceq \frac{\det C_1}{\det C_2}\cdot k^{C_2}$.

Proof. The proof of Proposition 3.3 is divided into three parts.

1. Denote by $I = I_d$ the $d\times d$-identity matrix. We first prove that for a positive diagonal matrix $D \in \mathbb{R}^{d\times d}$, we have $\mathbb{H}_D \overset{c}{\hookrightarrow} \mathbb{H}_I$ if and only if $D \ge I$. For this, let $D = \mathrm{diag}(\sigma_1,\dots,\sigma_d)$ with $\sigma_i > 0$ for all $i = 1,\dots,d$. Observe that for $x,y \in \mathbb{R}^d$, we have the following (for the tensor product of kernels see (A1)):

$$\begin{aligned} k^D(x,y) &= \exp\left(-(x-y)^\top D^{-2}(x-y)\right) \\ &= \exp\left(-\sum_{k=1}^d \sigma_k^{-2}(x_k-y_k)^2\right) \\ &= \prod_{k=1}^d \exp\left[-\frac{(x_k-y_k)^2}{\sigma_k^2}\right] \\ &= \prod_{k=1}^d k^{\sigma_k}(x_k,y_k) = (k^{\sigma_1}\otimes\cdots\otimes k^{\sigma_d})(x,y), \end{aligned}$$

[1] Recall that a matrix $A \in \mathbb{R}^{d\times d}$ is called *dissipative* if $A + A^\top \le 0$.

where the x_k and y_k are the Cartesian coordinates of x and y, respectively.
Hence, if $D \geq I$ (i.e., $\sigma_i \geq 1$ for all $i = 1, \dots, d$), Lemmas A2 and 3.2 imply

$$k^D = k^{\sigma_1} \otimes \cdots \otimes k^{\sigma_d} \preceq \sigma_1 k^1 \otimes \cdots \otimes \sigma_d k^1$$
$$= \sigma_1 \dots \sigma_d \cdot k^I = \det D \cdot k^I,$$

and Aronszajn's inclusion theorem, Theorem A1, implies $\mathbb{H}_D \overset{c}{\hookrightarrow} \mathbb{H}_I$.
Conversely, let $\mathbb{H}_D \overset{c}{\hookrightarrow} \mathbb{H}_I$ and suppose that $\sigma_1 < 1$. By Theorem A1, we have $k^D \preceq \alpha k^I$ with some $\alpha > 0$. By setting $x_k = y_k = 0$ for $k = 2, \dots, d$, we see that $k^D(x, y) = k^{\sigma_1}(x_1, y_1)$ and $k^I(x, y) = k^1(x_1, y_1)$. Hence, we have $k^{\sigma_1} \preceq \alpha k^1$. On the other hand, $1 > \sigma_1$ and Lemma 3.2 imply $k^1 \preceq \sigma_1^{-1} k^{\sigma_1}$. This implies $\mathbb{H}_{\sigma_1 I} = \mathbb{H}_I$, which contradicts [19, Corollary 7ii)].

2. Next, we prove that for a positive definite matrix $C \in \mathbb{R}^{d\times d}$, we have $\mathbb{H}_C \overset{c}{\hookrightarrow} \mathbb{H}_I$ if and only if $C \geq I$. For this, let $U \in \mathbb{R}^{d\times d}$ be an orthogonal matrix such that $C = U^\top D U$ with a diagonal matrix $D \in \mathbb{R}^{d\times d}$. Then $k^C = k^D \circ U$ and $k^I = k^I \circ U$. If $C \geq I$, then

$$k^C = k^D \circ U \preceq \det D \cdot k^I \circ U = \det C \cdot k^I.$$

Conversely, if $\mathbb{H}_C \overset{c}{\hookrightarrow} \mathbb{H}_I$, then $k^C \preceq \alpha k^I$ with some $\alpha > 0$, which implies $k^D \preceq \alpha k^I$ and hence $D \geq I$, which is equivalent to $C \geq I$.

3. Finally, let $C_1, C_2 \in \mathbb{R}^{d\times d}$ be arbitrary symmetric positive definite matrices. Note that $C_1^2 \geq C_2^2$ is equivalent to $C := (C_2^{-1} C_1^2 C_2^{-1})^{1/2} \geq I$. Furthermore, note that

$$k^C(x, y) = k^{C_1}(C_2 x, C_2 y) \quad \text{and}$$
$$k^I(x, y) = k^{C_2}(C_2 x, C_2 y), \quad x, y \in \mathbb{R}^d. \tag{4}$$

Therefore, if $C_1^2 \geq C_2^2$, by the second part of the proof,

$$k^{C_1} = k^C \circ C_2^{-1} \preceq \det C \cdot (k^1 \circ C_2^{-1}) = \tfrac{\det C_1}{\det C_2} \cdot k^{C_2},$$

which yields $\mathbb{H}_{C_1} \overset{c}{\hookrightarrow} \mathbb{H}_{C_2}$. Conversely, if $\mathbb{H}_{C_1} \overset{c}{\hookrightarrow} \mathbb{H}_{C_2}$, then there exists $\alpha > 0$ such that $k^{C_1} \preceq \alpha k^{C_2}$. In view of Equation (4), this gives $k^C \circ C_2^{-1} = k^{C_1} \preceq \alpha k^{C_2} = \alpha k^I \circ C_2^{-1}$ and hence $k^C \preceq \alpha k^I$. By Part 2, this implies $C \geq I$. □

Remark 3.4. Note that $C_1 \geq C_2$ does in general not imply $C_1^2 \geq C_2^2$ for symmetric positive definite matrices C_1 and C_2. On the other hand $C_1 \geq C_2$ implies $C_1^{1/2} \geq C_2^{1/2}$ and $C_1^{-1} \leq C_2^{-1}$.

We conclude this section with the following lemma.

Lemma 3.5. *Let $C_1, C_2 \in \mathbb{R}^{d\times d}$ be positive definite and $z, w \in \mathbb{R}^d$. Then,*

$$\int_{\mathbb{R}^d} k_z^{C_1}(y) \cdot k_w^{C_2}(y)\, dy = \frac{\pi^{d/2}}{\det(C_1^{-2} + C_2^{-2})^{1/2}} \cdot k^C(z, w),$$

where $C = (C_1^2 + C_2^2)^{1/2}$.

Proof. Let $M := (C_1^{-2} + C_2^{-2})^{1/2}$. We have $k_z^{C_1}(y) k_w^{C_2}(y) = e^{-L(y)}$, where

$$\begin{aligned}
L(y) &= (y - z)^\top C_1^{-2}(y - z) + (y - w)^\top C_2^{-2}(y - w) \\
&= y^\top (C_1^{-2} + C_2^{-2}) y - 2\big(C_1^{-2} z + C_2^{-2} w\big)^\top y + z^\top C_1^{-2} z + w^\top C_2^{-2} w \\
&= \|My\|^2 - 2\big\langle M^{-1}\big(C_1^{-2} z + C_2^{-2} w\big), My\big\rangle + \|C_1^{-1} z\|^2 + \|C_2^{-1} w\|^2 \\
&= \Big\| My - M^{-1}(C_1^{-2} z + C_2^{-2} w)\Big\|^2 - K(z, w),
\end{aligned}$$

where

$$\begin{aligned}
K(z, w) &= \Big\| M^{-1}(C_1^{-2} z + C_2^{-2} w)\Big\|^2 - \|C_1^{-1} z\|^2 - \|C_2^{-1} w\|^2 \\
&= \big\langle (C_1^{-2} + C_2^{-2})^{-1}(C_1^{-2} z + C_2^{-2} w), (C_1^{-2} z + C_2^{-2} w)\big\rangle \\
&\quad - \|C_1^{-1} z\|^2 - \|C_2^{-1} w\|^2 \\
&= \big\langle z + M^{-2}(C_2^{-2} w - C_2^{-2} z), C_1^{-2} z + C_2^{-2} w\big\rangle \\
&\quad - \|C_1^{-1} z\|^2 - \|C_2^{-1} w\|^2 \\
&= \langle z, C_2^{-2} w\rangle + \big\langle C_2^{-2}(w - z), (C_1^{-2} + C_2^{-2})^{-1}(C_1^{-2} z + C_2^{-2} w)\big\rangle \\
&\quad - \|C_2^{-1} w\|^2 \\
&= \langle z, C_2^{-2} w\rangle + \big\langle C_2^{-2}(w - z), w + M^{-2}(C_1^{-2} z - C_1^{-2} w)\big\rangle \\
&\quad - \|C_2^{-1} w\|^2 \\
&= \big\langle C_2^{-2}(w - z), M^{-2} C_1^{-2}(z - w)\big\rangle \\
&= \big\langle (C_2^2 M^2 C_1^2)^{-1}(w - z), (z - w)\big\rangle \\
&= -\langle (C_1^2 + C_2^2)^{-1}(z - w), (z - w)\big\rangle \\
&= -(z - w)^\top C^{-2}(z - w).
\end{aligned}$$

Thus, we obtain the following:

$$\begin{aligned}
&\int_{\mathbb{R}^d} k_z^{C_1}(y) \cdot k_w^{C_2}(y)\, dy \\
&\quad = e^{K(z,w)} \int_{\mathbb{R}^d} \exp(-\|My - M^{-1}(C_1^{-2} z + C_2^{-2} w)\|^2)\, dy \\
&\quad = k^C(z, w) \cdot \frac{\pi^{d/2}}{\det M},
\end{aligned}$$

and the lemma is proved. □

4 | Proof of Theorem 1.1

We first prove the following proposition on the action of the Koopman operator of Equation (1) on $\mathbb{H}_C$.

Proposition 4.1. *Let $t \geq 0$, let $C \in \mathbb{R}^{d\times d}$ be symmetric positive definite, and define the following:*

$$C_t = \Big[e^{-At}(C^2 + 2\Sigma(t)) e^{-A^\top t}\Big]^{1/2} \quad \text{and} \quad \tau_t = \frac{\det C}{(\det(C^2 + 2\Sigma(t)))^{1/2}}. \tag{5}$$

Then the Koopman operator K^t, associated with the SDE (1) and defined in Equation (2), maps $\mathbb{H}_C$ boundedly into $\mathbb{H}_{C_t}$ with norm not exceeding $\tau_t^{1/2}$.

Proof. Let $z \in \mathbb{R}^d$. By Lemma 3.5, we have the following:

$$(K^t k_z^C)(x) = (2\pi)^{-d/2}(\det \Sigma(t))^{-1/2}$$
$$\times \int_{\mathbb{R}^d} k_z^C(y) \exp\left(-\tfrac{1}{2}(y - e^{At}x)\Sigma(t)^{-1}(y - e^{At}x)\right) dy$$
$$= (2\pi)^{-d/2}(\det \Sigma(t))^{-1/2} \int_{\mathbb{R}^d} k_z^C(y) k_{e^{At}x}^{(2\Sigma(t))^{1/2}}(y)\, dy$$
$$= (2\pi)^{-d/2}(\det \Sigma(t))^{-1/2}$$
$$\times \frac{\pi^{d/2}}{\det(C^{-2} + (2\Sigma(t))^{-1})^{1/2}} k^{(C^2+2\Sigma(t))^{1/2}}(z, e^{At}x)$$
$$= \frac{(\det \Sigma(t))^{-1/2}}{\det(2C^{-2} + \Sigma(t)^{-1})^{1/2}} \cdot k_{e^{-At}z}^{C_t}(x) = \tau_t \cdot k_{e^{-At}z}^{C_t}(x),$$

hence, $K^t k_z^C = \tau_t \cdot k_{e^{-At}z}^{C_t} \in \mathbb{H}_{C_t}$. Therefore, if $f = \sum_{j=1}^n \alpha_j k_{x_j}^C$ with $\alpha_j \in \mathbb{R}$ and $x_j \in \mathbb{R}^d$ $(j = 1, \dots, n)$, we obtain the following:

$$\|K^t f\|_{C_t}^2 = \left\| \sum_{j=1}^n \alpha_j K^t k_{x_j}^C \right\|_{C_t}^2$$
$$= \tau_t^2 \left\| \sum_{j=1}^n \alpha_j k_{e^{-At}x_j}^{C_t} \right\|_{C_t}^2 = \tau_t^2 \sum_{i,j=1}^n \alpha_i \alpha_j k^{C_t}(e^{-At}x_i, e^{-At}x_j)$$
$$= \tau_t^2 \sum_{i,j=1}^n \alpha_i \alpha_j k^{(C^2+2\Sigma(t))^{1/2}}(x_i, x_j) \le \tau_t^2$$
$$\cdot \frac{\det(C^2 + 2\Sigma(t))^{1/2}}{\det C} \sum_{i,j=1}^n \alpha_i \alpha_j k^C(x_i, x_j) = \tau_t \|f\|_C^2,$$

where the inequality is due to Proposition 3.3.

This shows that K^t maps $\mathbb{H}_{0,C} = \operatorname{span}\{k_x^C : x \in \mathbb{R}\} \subset \mathbb{H}_C$ boundedly into $\mathbb{H}_{C_t}$. Since $\mathbb{H}_{0,C}$ is dense in $\mathbb{H}_C$, it follows that $K^t|_{\mathbb{H}_{0,C}}$ extends to a bounded operator $T : \mathbb{H}_C \to \mathbb{H}_{C_t}$. In order to see that $Tf = K^t f$ for $f \in \mathbb{H}_C$, let $(f_n) \subset \mathbb{H}_{0,C}$ such that $f_n \to f$ in $\mathbb{H}_C$. Then $K^t f_n = Tf_n \to Tf$ in $\mathbb{H}_{C_t}$. Since $\mathbb{H}_C \overset{c}{\hookrightarrow} L^2(\mu)$, we have $f_n \to f$ in $L^2(\mu)$ and thus $K^t f_n \to K^t f$ in $L^2(\mu)$. Also, $K^t f_n \to Tf$ in $L^2(\mu)$. Hence, $K^t f = Tf$ μ-a.e. on $\mathbb{R}^d$. But as both $K^t f$ and Tf are continuous and μ is absolutely continuous w.r.t. Lebesgue measure with a positive density, we conclude that $K^t f = Tf \in \mathbb{H}_C$. □

We are now in the position to prove the main result of this note—that the RKHS $\mathbb{H}_C$ is invariant under the Koopman operator of the SDE (1) if the Lyapunov-like condition (3) on the interplay of the matrices A, B, and C is satisfied.

Proof of Theorem 2.1. Let C_t and τ_t be defined as in Proposition 4.1. We prove that $C_t^2 \ge C^2$ for all $t \ge 0$ if any only if Equation (3) is satisfied. For $x \in \mathbb{R}^d$ and $t \ge 0$, we have the following:

$$f_x(t) := \langle C_t^2 x, x\rangle = \langle e^{-At}(C^2 + 2\Sigma(t))e^{-A^\top t}x, x\rangle$$
$$= \|Ce^{-A^\top t}x\|^2 + 2\langle e^{-At}\Sigma(t)e^{-A^\top t}x, x\rangle$$
$$= \|Ce^{-A^\top t}x\|^2 + 2\left\langle \left(\int_0^t e^{-As}BB^\top e^{-A^\top s}\, ds\right)x, x\right\rangle$$
$$= \|Ce^{-A^\top t}x\|^2 + 2\int_0^t \left\|B^\top e^{-A^\top s}x\right\|^2 ds.$$

Moreover,

$$\dot{f}_x(t) = \left\langle Ce^{-A^\top t}x, \tfrac{d}{dt}Ce^{-A^\top t}x\right\rangle + \left\langle \tfrac{d}{dt}Ce^{-A^\top t}x, Ce^{-A^\top t}x\right\rangle$$
$$+ 2\left\|B^\top e^{-A^\top t}x\right\|^2$$
$$= -\left\langle AC^2 e^{-A^\top t}x, e^{-A^\top t}x\right\rangle - \left\langle C^2 A^\top e^{-A^\top t}x, e^{-A^\top t}x\right\rangle$$
$$+ 2\left\|B^\top e^{-A^\top t}x\right\|^2$$
$$= \left\langle \left[2BB^\top - (AC^2 + C^2A^\top)\right]e^{-A^\top t}x, e^{-A^\top t}x\right\rangle.$$

Note that $f_x(0) = \|Cx\|^2$. Hence, if Equation (3) holds, we have $\dot{f}_x(t) \ge 0$ for all $t \ge 0$ and all $x \in \mathbb{R}^d$, which yields $C_t^2 \ge C^2$ for all $t \ge 0$. Conversely, if $C_t^2 \ge C^2$ for all $t \ge 0$, that is, $f_x(t) \ge f_x(0)$ for all $x \in \mathbb{R}^d$ and all $t \ge 0$, then $\dot{f}_x(0) \ge 0$ for all $x \in \mathbb{R}^d$, which is Equation (3).

We may thus apply Proposition 3.3 to conclude that $\mathbb{H}_{C_t} \overset{c}{\hookrightarrow} \mathbb{H}_C$ with $\|f\|_C^2 \le (\frac{\det C_t}{\det C})\|f\|_{C_t}^2$ for $f \in \mathbb{H}_{C_t}$ and all $t \ge 0$. Thus,

$$\|K^t f\|_C^2 \le \frac{\det C_t}{\det C}\|K^t f\|_{C_t}^2 \le \tau_t \frac{\det C_t}{\det C}\|f\|_C^2$$
$$= \det e^{-At}\|f\|_C^2 = e^{t\cdot \operatorname{Tr}(-A)}\|f\|_C^2,$$

which concludes the proof of the theorem. □

Remark 4.2. It is not clear whether the statement of Theorem 2.1 is actually an equivalence. For the other direction, one would have to show that $K^t\mathbb{H}_C \subset \mathbb{H}_C$ implies $\mathbb{H}_{C_t} \overset{c}{\hookrightarrow} \mathbb{H}_C$ as the latter was shown to be equivalent to Equation (3) in the proof of Theorem 2.1. Note that $K^t\mathbb{H}_C \subset \mathbb{H}_C$ at least implies $\mathbb{H}_{0,C_t} = \operatorname{span}\{k_x^{C_t} : x \in \mathbb{R}^d\} \subset \mathbb{H}_C$, see Proposition 4.1.

Acknowledgments

F. M. Philipp was funded by the Carl Zeiss Foundation within the project *DeepTurb—Deep Learning in and from Turbulence*. He was further supported by the free state of Thuringia and the German Federal Ministry of Education and Research (BMBF) within the project *THInKI—Thüringer Hochschulinitiative für KI im Studium*. K. Worthmann gratefully acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project ID: 507037103.

Open access funding enabled and organized by Projekt DEAL.

References

1. B. O. Koopman, "Hamiltonian Systems and Transformation in Hilbert Space, " *Proceedings of the National Academy of Sciences of the United States of America* 17, no. 5, (1931): 315–318.

2. A. Mauroy, I. Mezić, and Y. Susuki, *The Koopman Operator in Systems and Control*, Lecture Notes in Control and Information Sciences (Cham, Switzerland: Springer, 2020).

3. M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A Data-Driven Approximation of the Koopman Operator: Extending Dynamic Mode Decomposition, " *Journal of Nonlinear Science* 25 (2015): 1307–1346.

4. C. Schütte, P. Koltai, and S. Klus, "On the Numerical Approximation of the Perron-Frobenius and Koopman Operator, " *Journal of Computational Dynamics* 3, no. 9 (2016): 1–12.

5. D. Giannakis, A. Kolchinskaya, D. Krasnov, and J. Schumacher, "Koopman Analysis of the Long-Term Evolution in a Turbulent Convection Cell, " *Journal of Fluid Mechanics* 847 (2018): 735–767.

6. B. W. Brunton, L. A. Johnson, J. G. Ojemann, and J. N. Kutz, "Extracting Spatial–Temporal Coherent Patterns in Large-Scale Neural Recordings Using Dynamic Mode Decomposition, " *Journal of Neuroscience Methods* 258 (2016): 1–15.

7. O. Azencot, N. B. Erichson, V. Lin, and M. W. Mahoney, "Forecasting Sequential Data Using Consistent Koopman Autoencoders," in *Proceedings of the 37th International Conference on Machine Learning (ICML)* (2020), 475–485.

8. I. Mezić, "On Numerical Approximations of the Koopman Operator, " *Mathematics* 10 (2022): 1180.

9. C. Zhang and E. Zuazua, "A Quantitative Analysis of Koopman Operator Methods for System Identification and Predictions, " *Comptes Rendus Mécanique* 351, no. S1 (2023): 1–31.

10. F. Nüske, S. Peitz, F. Philipp, M. Schaller, and K. Worthmann, "Finite-Data Error Bounds for Koopman-Based Prediction and Control, " *Journal of Nonlinear Science* 33, no. 1, (2023): 14.

11. F. Philipp, M. Schaller, S. Boshoff, S. Peitz, F. Nüske, and K. Worthmann, "Variance Representations and Convergence Rates for Data-driven Approximations of Koopman Operators," *Arxiv preprint, arxiv: 2402.02494* (2024).

12. M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, "A Kernel-Based Method for Data-Driven Koopman Spectral Analysis, " *Journal of Computational Dynamics* 2, no. 2 (2015): 247–265.

13. S. Klus, I. Schuster, and K. Muandet, "Eigendecompositions of Transfer Operators in Reproducing Kernel Hilbert Spaces, " *Journal of Nonlinear Science* 30 (2020): 283–315.

14. F. M. Philipp, M. Schaller, K. Worthmann, S. Peitz, and F. Nüske, "Error Bounds for Kernel-Based Approximations of the Koopman Operator, " *Applied and Computational Harmonic Analysis* 71 (2024): 101657.

15. F. Philipp, M. Schaller, K. Worthmann, S. Peitz, and F. Nüske, "Error Analysis of Kernel EDMD for Prediction and Control in the Koopman Framework," *Arxiv preprint, arxiv: 2312.10460* (2023).

16. V. R. Kostic, K. Lounici, P. Novelli, and M. Pontil, "Sharp Spectral Rates for Koopman Operator Learning," in *Advances in Neural Information Processing Systems*, vol. 36 (2023).

17. V. Kostic, P. Novelli, A. Maurer, C. Ciliberto, L. Rosasco, and M. Pontil, "Learning Dynamical Systems via Koopman Operator Regression in Reproducing Kernel Hilbert Spaces, " *Advances in Neural Information Processing Systems* 35 (2022): 4017–4031.

18. F. Köhne, F. Philipp, M. Schaller, A. Schiela, and K. Worthmann, "l^∞-Error Bounds for Approximations of the Koopman Operator by Kernel Extended Dynamic Mode Decomposition," to appear in *SIAM Journal on Applied Dynamical Systems*, ArXiv preprint, arxiv: 2403.18809 (2024).

19. I. Steinwart, D. Hush, and C. Scovel, "An Explicit Description of the Reproducing Kernel Hilbert Spaces of Gaussian RBF Kernels, " *IEEE Transactions on Information Theory* 52 (2006): 4635–4643.

20. P. Vatiwutipong and N. Phewchean, "Alternative Way to Derive the Distribution of the Multivariate Ornstein-Uhlenbeck Process, " *Advances in Difference Equations* 2019 (2019): 276.

21. I. Steinwart and A. Christmann, *Support Vector Machines* (New York: Springer Science + Business Media, LLC, 2008).

22. V. I. Paulsen and M. Raghupathi, *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces*, Cambridge Studies in Advanced Mathematics, vol. 152 (Cambridge, UK: Cambridge University Press, 2016).

Appendix A: Some Background on RKHS Theory

Theorem A1 (Aronszajn's inclusion theorem, [22, Theorem 5.1]). *Let k_1 and k_2 be two symmetric positive definite kernels on a set X. Denote their corresponding RKHS's by $\mathbb{H}_1$ and $\mathbb{H}_2$, respectively. Then $\mathbb{H}_1 \subset \mathbb{H}_2$ if and only if $\mathbb{H}_1 \overset{c}{\hookrightarrow} \mathbb{H}_2$. Moreover, for $c > 0$ we have $\mathbb{H}_1 \overset{c}{\hookrightarrow} \mathbb{H}_2$ with $\|f\|_2 \le c\|f\|_1$ for $f \in \mathbb{H}_1$ if and only if $k_1 \le c^2 k_2$.*

Given two symmetric positive definite kernels k_X on X and k_Y on Y, we define the *tensor product* (cf. [22, Definition 5.12]) $k_X \otimes k_Y : (X \times Y)^2 \to \mathbb{R}$ of k_X and k_Y by the following:

$$(k_X \otimes k_Y)\big((x,y),(x',y')\big) = k_X(x,x') \cdot k_Y(y,y'). \tag{A1}$$

It follows from the positive semi-definiteness of the Schur product of two positive semi-definite matrices (see, e.g., [22, Theorem 4.8]) that $k_X \otimes k_Y$ is a symmetric positive definite kernel on $X \times Y$.

Lemma A2. *Let k_X, k_X' be kernels on X and k_Y, k_Y' kernels on Y and assume that $k_X \le k_X'$ and $k_Y \le k_Y'$. Then also $k_X \otimes k_Y \le k_X' \otimes k_Y'$.*

Proof. Denote the RKHS corresponding to k_X, k_X', k_Y, and k_Y' by $\mathbb{H}_X$, $\mathbb{H}_X'$, $\mathbb{H}_Y$, and $\mathbb{H}_Y'$, respectively. Let $(e_s)_{s\in S}$ and $(f_t)_{t\in T}$ be orthonormal bases of $\mathbb{H}_Y$ and $\mathbb{H}_X'$, respectively. Then we have $(x_1, x_2 \in X, y_1, y_2 \in Y)$

$$k_Y(y_1,y_2) = \sum_{s\in S} e_s(y_1)e_s(y_2) \qquad \text{and} \qquad k_X'(x_1,x_2) = \sum_{t\in T} f_t(x_1)f_t(x_2)$$

with pointwise convergence, see [22, Theorem 2.4]. Let $n \in \mathbb{N}$ and $\alpha_j \in \mathbb{R}$, $(x_j, y_j) \in X \times Y$, $j = 1, \dots, n$. Then,

$$\begin{aligned}
&\sum_{i,j=1}^n \alpha_i\alpha_j (k_X \otimes k_Y)\big((x_i,y_i),(x_j,y_j)\big)\\
&\quad= \sum_{i,j=1}^n \alpha_i\alpha_j k_X(x_i,x_j)k_Y(y_i,y_j) = \sum_{s\in S}\sum_{i,j=1}^n [\alpha_i e_s(y_i)][\alpha_j e_s(y_j)]k_X(x_i,x_j)\\
&\quad\le \sum_{s\in S}\sum_{i,j=1}^n [\alpha_i e_s(y_i)][\alpha_j e_s(y_j)]k_X'(x_i,x_j) = \sum_{i,j=1}^n \alpha_i\alpha_j k_X'(x_i,x_j)k_Y(y_i,y_j)\\
&\quad= \sum_{t\in T}\sum_{i,j=1}^n [\alpha_i f_t(x_i)][\alpha_j f_t(x_j)]k_Y(y_i,y_j) \le \sum_{i,j=1}^n \alpha_i\alpha_j k_X'(x_i,x_j)k_Y'(x_i,y_j)\\
&\quad= \sum_{i,j=1}^n \alpha_i\alpha_j (k_X' \otimes k_Y')\big((x_i,y_i),(x_j,y_j)\big),
\end{aligned}$$

which was to be proven. □