Diffusion Maps Kalman Filter

Tal Shnitzer, Student Member, IEEE, Ronen Talmon, Member, IEEE, Jean-Jacques Slotine

Abstract—In this paper, we propose a non-parametric method for state estimation of high-dimensional nonlinear stochastic dynamical systems. We combine diffusion maps, a manifold learning technique, with a linear Kalman filter and with concepts from Koopman operator theory. More concretely, using diffusion maps, we construct data-driven virtual state coordinates, which linearize the system model. Based on these coordinates, we devise a data-driven framework for state estimation using the Kalman filter. We demonstrate the strengths of our method with respect to both parametric and non-parametric algorithms in two object tracking problems. We show that the proposed method outperforms the competing non-parametric algorithms in the examined stochastic problem formulations. Additionally, we obtain results comparable to common parametric algorithms, which, in contrast to our method, are equipped with the hidden model knowledge.

Index Terms—Intrinsic modeling, manifold learning, Kalman filter, diffusion maps, non-parametric filtering.

I. INTRODUCTION

In many real applications, the system model is not accessible and some estimation of it is required. State estimation and characterization of stochastic, possibly nonlinear, dynamical systems are therefore widely studied problems. Commonly, such problems are addressed using classical algorithms, which rely on predefined parametric models. On the one hand, parametric models need to be sufficiently simple to allow accurate parameter estimation from measurements. On the other hand, too simple models often fail to accommodate the complexity of real systems. This facilitates the development of nonparametric methods [1]-[5]. Particularly in this paper, we take a non-parametric approach and propose a new method to derive the system model in a data-driven manner.

To demonstrate the primary idea, consider a classical nonlinear Simultaneous Localization and Mapping (SLAM) problem, where the goal is to track the 2D position $x = (x_1, x_2)^T$ of a moving object. Typically, in the bearing-only version of this problem, the accessible system measurements are given by the azimuth of an object, i.e. by $\phi = \text{arctan}(x_1/x_2)$. This nonlinear model, relating the state (position) of the system $x$ to the measurements $\phi$, poses a challenge for processing and analysis, since common linear methods cannot be applied. In [6], this nonlinearity is solved by constructing virtual measurements, $y$, and measurement mapping, $H$, based on the knowledge of the system properties and model, which allow for the formulation of a linear problem and the application of a linear time-varying Kalman filter. Briefly, since the system state is given by $x$, linearization is achieved by defining the following virtual measurements $y = Hx + v$, where $v$ is noise and:

$$H = \begin{bmatrix} \cos \phi & -\sin \phi \end{bmatrix}$$

such that $Hx = 0$. The resulting measurement equation is linear and can be constructed from the given measurements $\phi$. Analogously, in this work, we propose a computational method to construct the data-driven non-parametric counterparts of a virtual state, linearizing the system dynamics and measurement model. However, in contrast to [6], our construction of this virtual state is data-driven and does not require any explicit knowledge of the system properties or measurement modality, e.g. the knowledge that the measurements represent the azimuth in a 2D tracking problem.

One way to devise such a computational method, which has recently drawn significant research attention, is to address the problem of data-driven system analysis and state estimation from an operator-theoretic point of view. In this approach, the dynamical system is described by two dual operators, the Perron-Frobenius operator, which represents probability density evolution, and the Koopman Operator, which acts on some linear functional space of infinite dimension and describes the time-evolution of observables [7], [8].

In the context of empirical dynamical systems analysis, the main challenge is to approximate these operators from the system measurements. Several methods for estimating the Koopman Operator have been proposed in recent years [8]-[13]. For example, the Extended Dynamic Mode Decomposition (EDMD) [8], [9] approximates the Koopman eigenfunctions and modes based on two sets of points, related through system dynamics, and a set of dictionary elements. However, the optimal choice of the dictionary in EDMD depends on the data [9]. This framework for estimating the Koopman eigenfunctions and modes was later employed in [3] as part of a non-parametric Kalman filtering framework, where a linear Kalman filter is constructed based on the approximation of the Koopman Operator, along with its eigenvectors, eigenvalues and Koopman modes using EDMD. The Kalman filter propagates the system in the space spanned by the Koopman eigenvectors and then the resulting estimates are projected back to the state space. Other work on the analysis of nonlinear stochastic dynamical systems based on Koopman theory includes [10], [11], [14]. The work in [14] offers a formal definition and rigorous mathematical analysis for the generalization of the Koopman operator to stochastic dynamical systems. In addition, they propose a new framework for approximating the eigenfunctions and eigenvalues of this stochastic Koopman operator. A different approach is proposed in [10], [11], where the authors characterize the

Tal Shnitzer and Ronen Talmon are with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa, Israel 32000 (e-mail: shnitzer@campus.technion.ac.il; ronen@ee.technion.ac.il).

Jean-Jacques E. Slotine is with the Nonlinear Systems Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A. (e-mail: jjs@mit.edu).

This research was supported by DARPA under Contract HR0011-16-C-0116.
long term behavior of a system (asymptotic dynamics) based on time-averages of functions. In [10], invariant-measures are defined and it is shown that these measures coincide with the eigenfunctions of the Koopman Operator and can be simply calculated by the Fourier transform. In [11], this framework is extended to address dynamical systems which are not measure-preserving using Laplace averages. In both [10] and [11], several trajectories of the system are needed for the analysis.

Different related non-parametric frameworks for state estimation in stochastic dynamical systems based on geometry and manifold learning were presented in [5], [15]–[17]. In [5], [15] a probabilistic approach is proposed, in which the problem is projected onto coordinates constructed by diffusion maps [18], a manifold learning technique. In these diffusion maps coordinates, the probability density function of the system state can be propagated in time without prior knowledge of the system dynamics, yet, the underlying system state is assumed to be accessible. The work in [15] proposes to construct an ensemble Kalman filter based on delay embedding coordinates (Takens embedding [19]), used for dynamics estimation. The propagation in time is estimated based on nearest neighbors of the current time-lag frame. In [17], a framework for estimating the eigenfunctions of the Koopman Operator generator based on diffusion maps is presented. This work discusses the relationship between diffusion maps and the Koopman operator and proposes to estimate the Koopman generator eigenfunctions using the eigenfunctions of the Laplace-Beltrami operator approximated by diffusion maps.

The relationship between diffusion maps and the Koopman Operator is further discussed in [7]. There, it is shown that EDMD can also provide data-driven dimensionality reduction. Moreover, for systems in which the system state is described by a Markov process, the eigenfunctions of the backward Fokker-Planck operator can be approximated using EDMD, similarly to diffusion maps. The main benefit of such a manifold learning approach using EDMD is that both the dynamics and the geometry of the underlying state are taken into account.

In this paper, we present a framework for state estimation of stochastic dynamical systems, which reveals the system model with minimal prior assumptions, using diffusion maps [18] and the Kalman filter. We assume that we are given a set of noisy measurements from some unknown nonlinear function of a stochastic underlying state and show that a linear model describing the system can be revealed, even for highly non-linear systems. This is obtained in a completely data-driven manner, based on virtual state coordinates constructed by diffusion maps and their inherent dynamics [20]. A Kalman filter is then formulated based on the recovered system model and utilized for non-parametric state estimation. By constructing the Kalman filter based on the recovered model, we incorporate system dynamics into diffusion maps, combining geometry and dynamics, as in [7], from a new manifold learning standpoint. We further show that our method uncovers an operator describing the system dynamics, which is analogous to the Stochastic Koopman Operator [7], an extension of the Koopman Operator for stochastic systems. Moreover, our devised method is well suited for high-dimensional systems due to the nonlinear dimensionality reduction obtained by diffusion maps.

With respect to previous work, our method encompasses several key differences. First, it does not rely on accessibility to the state of the system (as assumed in [5], [15]). Second, it does not require predefined dictionary elements (as required in [9]). Third, our method addresses stochastic dynamical systems in contrast to the deterministic systems considered in [3], [16]. Finally, it does not require a training set of samples with known states, in contrast to common non-parametric filtering algorithms [1], [2].

The remainder of the paper is organized as follows. In Section II the general problem setting is presented. In Section III we first overview the key-points of our method and then describe the derivation of the Diffusion Maps Kalman filter framework in detail. In Section IV we demonstrate the strengths of our method on two object tracking problems. We compare our method with both parametric and non-parametric methods. Section V concludes the paper with a short summary.

II. PROBLEM FORMULATION

Consider an ergodic stochastic dynamical system with some non-linear generator \( T : \mathcal{M} \to \mathcal{M} \), where \( \mathcal{M} \) is a compact Riemannian manifold of dimension \( d \). The system is defined by:

\[
\dot{\theta}_t = T(\theta_t, \omega_t) \quad (2)
\]

\[
z_t = g(\theta_t) + v_t \quad (3)
\]

where \( \theta_t \in \mathcal{M} \) is the system state, \( \dot{\theta}_t \) is its time derivative, \( \omega_t \in \mathbb{R}^d \) is the process noise, \( z_t \in \mathbb{R}^m \) are the system measurements through some unknown non-linear function \( g \) and \( v_t \in \mathbb{R}^n \) is the measurement noise. The evolution in time of such a dynamical system can be described by the Stochastic Koopman Operator [7], which is defined by

\[
(U_{st}f)(\theta_t) = \mathbb{E}[f \circ T(\theta_t, \omega_t)] \quad (4)
\]

where \( \circ \) is the composition operator, \( f : \mathcal{M} \to \mathbb{R} \) are some observables from an infinite dimensional vector space, closed under composition with the generator \( T \), and \( \mathbb{E} \) denotes expectation.

One of the notable properties of the Koopman Operator, which has increased its usage in a line of recent work, is that it is linear in the space of observables, even for highly non-linear dynamical systems: \( U_{st}(\alpha_1f_1 + \alpha_2f_2) = \mathbb{E}[(\alpha_1f_1 + \alpha_2f_2) \circ T] = \alpha_1\mathbb{E}[f_1 \circ T] + \alpha_2\mathbb{E}[f_2 \circ T] = \alpha_1U_{st}f_1 + \alpha_2U_{st}f_2 \). However, the tradeoff is that even for finite dimensional dynamical systems, the Koopman Operator acts on an infinite dimensional space of observables.

In this work, we rely on the derivations in [21], which focus on the case in which the dynamics equation (2) takes the form of a Langevin equation:

\[
\dot{\theta}_t = -\nabla U(\theta_t) + \sqrt{\frac{2}{\beta}} \omega_t \quad (5)
\]

where \( U(\theta_t) \) is a smooth and bounded potential, \( \sqrt{\frac{2}{\beta}} \) is a constant diffusion coefficient, \( \omega_t \) is Brownian motion and \( \dot{\omega}_t \) is its time derivative. This model is mainly used for theoretical
III. DIFFUSION BASED KALMAN FILTER

In this section we lay the foundation for our suggested framework and present the theoretical results which allow for the construction of a data-driven linear Kalman filter even for highly non-linear dynamical systems, requiring only minimal assumptions on the dynamics of the system \( \theta \).

A. Overview

We will present a method based on diffusion maps that discovers a new coordinate system describing a model of the state of the system with linear drift in a completely data-driven manner. Based on this linear drift, our method constructs a linear operator, analogous to the Stochastic Koopman Operator, from measurements, without prior model knowledge. By exploiting the linearity of this operator, we will formulate a linear Kalman filter framework, which allows for estimation of trajectories of the underlying system state based on noisy non-linear measurements. Now we will briefly overview key points of our method, which is described in detail in the following subsections.

Given noisy measurements \( z_t \in \mathbb{R}^m \), we use diffusion maps to represent the system state \( \theta_t \), by a new set of \( k \) coordinates, denoted by \( \Psi_t \), commonly \( k < m \). We will show that in this new coordinate system, the system can be described by the following linear equations:

\[
\dot{\Psi}_t = F \Psi_t + Q_t \dot{\omega}_t \quad \text{(6)}
\]
\[
z_t = H \Psi_t + R_t \nu_t \quad \text{(7)}
\]

where \( F \) is a linear operator describing the linear drift of the dynamics of the new coordinates \( \Psi_t \), \( H \) is a linear lift operator from the new coordinates \( \Psi_t \) to the measurements \( z_t \), \( \dot{\omega}_t \) is a standard normally distributed noise process, \( \nu_t \) is measurement noise, \( Q_t \) and \( R_t \) are matrices determining the covariance of the driving and measurement noise processes, respectively.

We will further show that the linear operators \( F \) and \( H \), in (6) and (7) respectively, can be constructed using diffusion maps in a data-driven manner.

Diffusion maps provide a coarse approximation of the system, yet it ignores the inherent time-dependencies between consecutive samples. Therefore, we will formulate a linear Kalman filter using the constructed coordinates \( \Psi_t \) and the recovered system operators \( F \) and \( H \), to improve the state estimate by incorporating the dynamics into the diffusion maps coordinates. This leads to a data-driven linear filtering framework, which can be applied to non-linear systems with an unknown model, revealing a new representation of the system \( \Psi_t \), which is tightly related to the system state \( \theta_t \) as will be described in Subsection III-B. In addition, our framework allows for estimation of specific system trajectories based on measurements, in contrast to most existing work on the Stochastic Koopman Operator, which represent the average time-evolution.

The remainder of this section is described as follows. In Subsection III-B and Subsection III-C we reiterate the derivations presented in [21] for state and model recovery using diffusion maps [18]. In Subsection III-D we present our proposed Kalman filter framework and in Subsection III-E we elaborate on the relation between our framework and the Stochastic Koopman Operator.

B. Recovering the State

Our goal in this subsection is to find a data-driven representation for the unknown state \( \theta_t \), of a dynamical system, measured through an unknown non-linear measurement function \( g \) with additional noise as in (3). We address this problem using diffusion maps [18], a particular manifold learning method. Manifold learning methods are non-parametric, data-driven methods, which provide a new representation of the data based on their geometry, and therefore, are suitable for our purpose. The diffusion maps algorithm is described as follows.

Given noiseless system measurements, \( z_t = g(\theta_t) \), of the hidden system state, \( \theta_t \), the following kernel is defined

\[
k_{\text{eq}}(s,t) = \exp \left( \frac{-d^2(s, z_t)}{\epsilon^2} \right) \quad \text{(8)}
\]

where \( \epsilon \) is a kernel scale, commonly taken as the median of the distances between the measurements, and \( d^2(s, z_t) \) is a distance function between \( s \) and \( z_t \). In our case, we calculated this distance using a modified Mahalanobis distance first presented in [22]:

\[
d^2(s, z_t) = \frac{1}{2} (z_s - z_t) \left( C_s^{-1} + C_t^{-1} \right) (z_s - z_t)^T \quad \text{(9)}
\]

where \( C_s^{-1} \) and \( C_t^{-1} \) are the inverse covariance matrices of the measurements at times \( s \) and \( t \), respectively. In [22], it was shown that the modified Mahalanobis distance between the measurements provides an estimate of the Euclidean distance between the hidden states.

The constructed kernel is then normalized according to

\[
p(s,t) = \frac{k_{\text{eq}}(s,t)}{d_e(s)} \quad \text{(10)}
\]

where \( d_e(s) = \int k_{\text{eq}} (s,t) p_{eq}(\theta_t) \, dt \) and \( p_{eq}(\theta_t) = e^{-U(\theta_t)} \) is the equilibrium density of the hidden state \( \theta_t \).

We can then define the operator \( P_e \) on real functions of the hidden state \( \theta_t \) by

\[
(P_e f)(\theta_s) = \int p_e(s,t) f(\theta_t) p_{eq}(\theta_t) \, dt
\]

Based on \( P_e \) we define

\[
L_e = \frac{P_e - I}{\epsilon} \quad \text{(11)}
\]

where \( I \) is the identity operator. In [22] it was shown that given the dynamical system presented in (5), the operator \( L_e \)
converges to the backward Fokker-Planck operator \( \mathcal{L}^* \) defined on the manifold \( \mathcal{M} \), as \( \epsilon \to 0 \):
\[
\mathcal{L}^* f = \frac{1}{\beta} \Delta f - \nabla f \cdot \nabla U
\]
(12)

here \( f \) denotes functions in a subspace of observables defined on the system state, which describe averages of functions: \( f(\theta_t) = \mathbb{E} [h(\theta_t) | \theta_0 = a_0] \), where \( a_0 \) is some initial state, and \( h \) is smooth.

The operator \( L e \) has a discrete set of decreasing eigenvalues, \( \{ -\lambda^{(\ell)}(0) \} \in \mathbb{N}, 0 = \lambda^{(0)}(0) > \lambda^{(1)}(0) \geq \lambda^{(2)}(0) \geq ... \), and eigenfunctions, \( \{ \psi^{(\ell)}(0) \} \in \mathbb{N} \) [18]. In many stochastic systems, these eigenvalues exhibit a spectral gap, leaving only a few dominant eigenvalues (close to 0) [20]. In such systems, we can represent the hidden state using only the first \( k \) eigenfunctions, corresponding to the largest (non-trivial \( \ell \neq 0 \)) eigenvalues. The diffusion maps coordinates is then obtained by calculating the eigenvalue decomposition of \( L e \) and using the \( k \) eigenfunctions corresponding to the \( k \) largest eigenvalues:
\[
z_t \mapsto \Psi(\theta_t) = \left[ \psi^{(1)}(\theta_t), \psi^{(2)}(\theta_t), ..., \psi^{(k)}(\theta_t) \right]
\]
(13)

It was shown in [22] that when using a kernel based on the modified Mahalanobis distance [9], the eigenfunctions of \( L e \) converge to the eigenfunctions of the backward Fokker-Planck operator \( \mathcal{L}^* \), defined on the hidden state \( \theta_t \), as \( \epsilon \to 0 \).

The relation of the diffusion maps coordinates to the system state is based on the following derivation. Consider the adjoint of \( \mathcal{L}^* \), which is known as the forward Fokker-Planck operator. The forward Fokker-Planck operator exhibits two important properties. First, it describes the evolution of the transition probability density, \( p(\theta_t | \theta_0) \). Second, it is Hermitian, and therefore, its eigenfunctions \( \varphi^{(\ell)} \) form a basis for the space of real functions of the system state (with the equilibrium density as measure \( p_{eq}(\theta_t) = \exp^{-U(\theta_t)} \)). Combining these two properties enables us to represent the transition probability density of the underlying system state by [23]:
\[
p(\theta_t | \theta_0) = \sum_{\ell=0}^{\infty} c_{\ell} e^{-\lambda^{(\ell)}(0)} \varphi^{(\ell)}(\theta_t)
\]
(14)

where \( c_{\ell} \) are coefficients determined by the initial conditions. Since the backward and forward operators are adjoint, their eigenfunctions can be normalized to be bi-orthonormal, and [14] can be recast as
\[
p(\theta_t | \theta_0) = \sum_{\ell=0}^{\infty} b_{\ell} e^{-\lambda^{(\ell)}(0)} \psi^{(\ell)}(\theta_t)
\]
(15)

where \( b_{\ell} \) are the corresponding coefficients. Therefore, from this point on, we will address the eigenfunctions \( \psi^{(\ell)} \) as a new set of coordinates representing the hidden system state.

We note that the above results are obtained only when we have access to the clean measurements \( z_t = g(\theta_t) \). We address this issue in Subsection III-D.

### C. Recovering the Model

In the previous subsection we presented a data-driven derivation of the diffusion maps coordinates which are used as a new representation of the system state. This new representation provides access to the underlying system state due to the use of the modified Mahalanobis distance [22]. We will now show that the representation of the system state using the diffusion maps coordinates has additional important benefits, as it provides a data-driven representation for the system model as well.

The system model is composed of two main components, system dynamics, \( T \), and the measurement function, \( g \). In this subsection, we first start by showing that the diffusion maps coordinates evolve according to known dynamics, containing a linear drift component which is revealed by the diffusion maps algorithm. This allows for the construction of a linear model in the obtained new coordinates space. Second, we show that the lift function, from the diffusion maps coordinates to the measurements, can be represented by a linear mapping.

The Dynamics of the Diffusion Maps Coordinates: Consider the dynamical system in [5] measured through \( z_t = g(\theta_t) \). For such system equations, based on Itô’s Lemma, the eigenfunctions of the backward Fokker-Planck operator, \( \mathcal{L}^* \), obtained using the diffusion maps algorithm, evolve according to a stochastic differential equation (SDE) of known form [20]:
\[
\dot{\psi}^{(\ell)}(\theta_t) = -\lambda^{(\ell)}(\psi^{(\ell)}(\theta_t)) + \sqrt{2}||\nabla \psi^{(\ell)}(\theta_t)|| \omega^{(\ell)}_t
\]
(16)

where \( \lambda^{(\ell)}(\theta_t) \) is the \( \ell \)th eigenfunction of the backward Fokker-Planck operator, \( -\lambda^{(\ell)} \) is the corresponding eigenvalue and \( \omega^{(\ell)}_t \) is the \( \ell \)th coordinate of a multidimensional Brownian motion process.

This SDE depicts that the obtained diffusion maps coordinates evolve according to a linear drift, \( -\lambda^{(\ell)}(\psi^{(\ell)}(\theta_t)) \), and an additional diffusion component. We note that the linear drift component is fully known since we obtain both the eigenfunctions and the corresponding eigenvalues of the backward Fokker-Planck operator using diffusion maps.

Constructing the Lift Function: As stated in Subsection III-B the eigenfunctions of the backward Fokker-Planck operator, obtained by diffusion maps, form a basis for the space of real functions defined on the system state. Therefore, every real function of the system state can be written as an expansion in these eigenfunctions. Specifically, we can represent the measurement function in the following manner:
\[
z_t^{(i)} = g^{(i)}(\theta_t) = \sum_{\ell=0}^{\infty} \alpha_{i,\ell} \psi^{(\ell)}(\theta_t)
\]
(17)

where \( \alpha_{i,\ell} = \langle z_t^{(i)}, \psi^{(\ell)}(\theta_t) \rangle p_{eq}(\theta_t) = \int_{-\infty}^{\infty} z_t^{(i)} \psi^{(\ell)}(\theta_t) p_{eq}(\theta_t) \) \, dt.

When the eigenvalues of the backward Fokker-Planck operator exhibit a spectral gap, most of the energy is captured by the \( k \) eigenfunctions, corresponding to the \( k \) largest eigenvalues. In such cases we can approximate the mapping in (17) using only these \( k \) eigenfunctions:
\[
z_t^{(i)} \approx \sum_{\ell=0}^{k} \alpha_{i,\ell} \psi^{(\ell)}(\theta_t)
\]
(18)

We can now write expression (18) in matrix form:
\[
z_t \approx \alpha \Psi(\theta_t)
\]
(19)
where $\Psi(\theta_i)$ is defined in (13) and $\alpha$ is an $m \times k$ matrix in which $(\alpha)_{i,\ell} = \alpha_{i,\ell}$.

These results imply that through the representation of the measurement function using the diffusion maps eigenfunctions, we obtained a linear mapping between the eigenfunctions, $\Psi(\theta_i)$, and the system measurements, $z_i$. Thus, we set the linear lift function simply to be $\alpha$.

To conclude this subsection, we note that all of the theoretical derivations above are true for state equations of the form (5). However, it is not essential that specifically the state will exhibit such dynamics, but rather that the dynamics of some underlying system parameter are governed by the Langevin equation. In such systems, the diffusion maps coordinates, constructed using the modified Mahalanobis distance (9), capture significant properties and can be used as a foundation for the proposed time-series filtering framework, as described in the remainder of this paper. Since many natural phenomena are governed by dynamics that can be modeled using Langevin equation (5), our framework is applicable to a wide range of problems.

D. Diffusion Maps Kalman Filter

We now present our proposed method for time-series filtering by exploiting the properties of the diffusion maps coordinates described in Subsection III-B and Subsection III-C.

We begin by presenting the discrete counterpart of diffusion maps. Given discrete-time measurements $\{z_n\}_{n=1}^{N}$, the kernel matrix is calculated similarly to (8) by:

$$K(i, j) = \exp\left(-\frac{d^2(z_i, z_j)}{\epsilon^2}\right)$$

where $\epsilon$ is the kernel scale and $d^2(\cdot, \cdot)$ is the modified Mahalanobis distance (22).

The kernel matrix is then normalized to be row-stochastic

$$P(i, j) = \frac{K(i, j)}{D(i)}$$

where $D(i) = \sum_{j=1}^{N} K(i, j)$.

From the eigendecomposition of $P$ we obtain a set of eigenvectors and eigenvalues, $\{\psi^{(i)}\}_{i=0}^{k}$, $\{\mu^{(i)}\}_{i=0}^{k}$. It was shown in (18), that in this discrete setting, the diffusion maps eigenvectors approximate the continuous diffusion maps eigenfunctions discussed in the previous subsections. Moreover, it was shown in (22), that the eigenvalues of the discrete diffusion maps algorithm can be used to approximate the eigenvalues of the continuous operator according to $-\lambda^{(i)} = \frac{2}{\epsilon} \log \mu^{(i)}$.

Based on the Euler-Maruyama method, the dynamics of the diffusion maps coordinates can be discretized to

$$\psi^{(i)}(\theta_{n+1}) = \left(1 - \lambda^{(i)} \Delta t\right) \psi^{(i)}(\theta_n) + \sqrt{2} \|\nabla_{\theta_i}\psi^{(i)}(\theta_n)\| \Delta \omega^{(i)}_n$$

where $\Delta t$ is the time step, $\Delta \omega^{(i)}_n$ is a normally distributed noise process and $\nabla_{\theta_i}$ denotes the gradient according to the system state $\theta_n$. We can now write the state and measurement equations in the discrete setting using the diffusion maps coordinates to represent the state of the system:

$$\Psi_{n+1} = (I + \Lambda) \Psi_n + Q_n \Delta \omega_n$$
$$z_n = \alpha \Psi_n + R_n v_n$$

where $\Lambda$ is a diagonal matrix with $-\lambda^{(i)} \Delta t \frac{k}{\epsilon^2}$ as its diagonal elements, $I$ is the identity matrix, $Q_n$ is a matrix containing the coefficients of the second term in (22), $R_n$ is the covariance of the measurement noise and $\alpha$ is the lift function from the diffusion maps coordinates to the measurements. For simplicity of notation, we denote $\Psi_n = \Psi(\theta_n)$, omitting the dependency on $\theta_n$. Importantly, note that these system equations are approximately linear, even for highly non-linear systems. Therefore, using the diffusion maps coordinates, we obtain a virtual system state which linearizes the problem, in an entirely data-driven manner, given only the measurements.

In the discrete setting, the lift function is approximated by $\alpha_{i,\ell} = \sum_{n=1}^{N} z_n \psi^{(i)}(\theta_n)$.

Note that the noise in equation (24), $R_n v_n$, is different from the measurement noise in (3). The diffusion maps coordinates are constructed based on the noisy measurements $z_n$, and therefore, the noise in these coordinates is not additive. However, in many cases, the leading eigenvectors (corresponding to the largest eigenvalues) represent the state and are less affected by the noise. We approximate the remaining noise in $\Psi_n$ by the additive term in (24). In addition, this term can be used to represent deviations from the true model (24).

Due to the linearity of the derived system equations (23) and (24) (except for elements in $Q_n$ as discussed in a subsequent paragraph), we can construct a linear Kalman filter based on the diffusion maps coordinates. Using the Kalman filter framework we incorporate the theoretical linear dynamics (drift) of the diffusion maps eigenfunctions into their approximation by the eigenvectors of the discrete diffusion maps algorithm, thereby incorporating the inherent time-dependencies of the system samples into a manifold learning technique.

Our proposed framework improves the method in (21) in two main aspects. First, the constructed observer scheme is a deterministic framework which discards the stochastic term of the dynamics, whereas the Kalman filter takes it naturally into account. Second, the Kalman filter provides an adaptive optimal update of the fixed model parameters in (21). We show in Section IV that the adaptive parameter update as well as the stochastic framework significantly improves the state recovery and robustness to noise and outperforms competing methods.

We note that $Q_n$ is a non-linear function of the state and induces dependencies between different time-points of the system noise process, which is not suitable with the Kalman filter framework. However, in many applications, the leading diffusion maps eigenvectors, which are used as a low-dimensional representation of the system, are slowly varying functions of the system state. Therefore, the gradient of $\psi^{(i)}$ according to $\theta_n$ in (22) might be sufficiently low (or approximately constant) and allow the proper use of the Kalman filter. These properties are demonstrated in the simulations in Section IV.
An alternative approach to address the dependency of the driving noise on the state would be using a particle filter instead of the Kalman filter. Particle filters are designed to support a wider range of driving noise distributions. However, our empirical study showed that the Kalman filter significantly outperformed the particle filter, since the particle filter is sensitive to errors in the estimation of the gradient of $\psi^{(\ell)}$. Improving our methods by appropriately approximating the driving noise is a direction for future work.

The implemented Kalman filter framework is described as follows:

$$\begin{align*}
\hat{\Psi}_n &= F_n \tilde{\Psi}_{n-1} + \kappa_n \left( z_n - H_n F_n \tilde{\Psi}_{n-1} \right) \\
P_n &= (I - \kappa_n H_n) \left( F_n P_{n-1} F_n^T + Q_n \right) \\
\kappa_n &= \left( \left( F_n P_{n-1} F_n^T + Q_n \right) H_n^T \right)^{-1} (H_n F_n P_{n-1} F_n^T H_n^T + H_n Q_n H_n^T + R_n)
\end{align*}$$

where $\hat{\Psi}_n$ is the state estimate at time $n$, $z_n$ is the measurement at time $n$, $I$ is the identity matrix, $F_n = I + \Lambda$ represents the dynamics of the state, $H_n = \alpha$ is the lift function between the measurements and the state calculated in a data driven manner as described after equation (24), $Q_n$ is the covariance matrix of the state driving noise, and $R_n$ is the covariance matrix of the measurement noise.

**E. Koopman Operator**

Our presented framework is tightly related to the Koopman Operator. Specifically, we show the analogy of the revealed dynamics in the present work to the Stochastic Koopman Operator [7], $(U_{st} f)(\theta_n) = E[ f \circ T(\theta_n, \omega_n)]$. Given measurements from some stochastic non-linear system of the form (5), we project the problem onto the eigenvector space obtained by diffusion maps. By taking these eigenvectors as observables we obtain a space in which the evolution of the observables is represented by a known linear operator:

$$\begin{align*}
\tilde{\Psi}_n &= \mathbb{E} [\psi^{(\ell)}(\theta_n) | \Psi_n] \\
&= \mathbb{E} \left[ 1 - \lambda^{(\ell)} \Delta t \right] \psi^{(\ell)}(\theta_n) + \sqrt{2} \left\| \nabla_{\theta} \psi^{(\ell)}(\theta_n) \right\| \tilde{\omega}^{(\ell)}(\theta_n) | \Psi_n \\
&= \left( 1 - \lambda^{(\ell)} \Delta t \right) \psi^{(\ell)}(\theta_n)
\end{align*}$$

where $\tilde{U}_{st}$ is analogous to the Stochastic Koopman Operator. Note that in contrast to the standard Stochastic Koopman Operator, $\tilde{U}_{st}$ is conditioned on $\Psi_n$.

The use of diffusion maps for approximation of the Koopman Operator is thoroughly discussed in [17]. This work presents a Galerkin method for approximating the eigenfunctions of the Koopman generator using the diffusion maps eigenfunctions and eigenvalues. Furthermore, it was shown that for ergodic systems with pure-point spectra, the eigenfunctions of the Koopman generator can be robustly estimated from finite data using diffusion maps. Importantly, our method is completely different than the method presented in [17], since we approximate the Stochastic Koopman Operator, whereas in [17] the generator is approximated.

In our proposed framework, we combine the constructed linear operator and observables, obtained by diffusion maps, with a Kalman filter. This leads to two main benefits. First, instead of representing the average time evolution of the observables in stochastic systems, we obtain an estimation of specific trajectories based on the measurements. In addition, the Kalman filter compensates for the noise and deviations from the measurements. Second, due to the use of diffusion maps we obtain a data-driven dimensionality reduction and approximate the Koopman Operator based on a finite set of orthonormal functions, spanning the state space of the system [20].

A related work, combining the Koopman Operator and a Kalman filter is presented in [3]. The authors define the Koopman Observer Form (KOF) for noisless systems and the Koopman Kalman Filter (KKF) for systems with measurement noise. They construct a set of linear update equations based on the eigenvectors and modes of the Koopman Operator which provides a linear filtering framework for non-linear systems. In their work, they approximate the Koopman Operator of a given data-set using EDMD [9]. The EDMD algorithm requires a dictionary of basis functions which affects the resulting estimations [9]. Conversely, in the proposed work, we obtain the linear dynamics and observables based on the data, from the diffusion maps algorithm. Another difference is the problem setting, which now includes stochastic system dynamics rather than measurement noise only. We compare our suggested framework to the one presented in [3], in Subsection IV-A.

Our framework can also be partially related to the work presented in [4]. There, linear update equations are learned from the data for chaotic systems, using concepts from Koopman theory. In order to represent the chaotic dynamics using a finite approximation, a non-linear forcing term is added which represents the deviation from linearity. In our work, we rely on the dynamics of the diffusion maps eigenfunctions which can be divided into a linear drift component and a non-linear stochastic component. The stochastic component represents deviations from the simple linear dynamics and can be used for error analysis as well. In the proposed Kalman filter framework we consider the stochastic component as the system noise.

**IV. Experimental Results**

In this section we present two examples of object tracking based on different measurement modalities, both in non-linear settings with unknown system dynamics and measurement functions. We compare our proposed Diffusion Maps Kalman (DMK) to several competing algorithms, which are detailed in the following. We show that our DMK framework leads to improved state estimates compared with non-parametric algorithms and obtains results which are comparable with parametric methods which, in contrast to DMK, are provided with the system model.
A. Non-Linear Object Tracking

We first present a Gaussian setting, where the location of an object in a 2-dimensional space is measured through its radius and azimuth angle. The underlying process, describing the Cartesian position of the object at each time point is given by the following discrete time Langevin equations:

\[
\begin{align*}
\Delta \theta^{(1)}_{n+1} & = \left( c_1 - 0.1\theta^{(1)}_n \right) + \sqrt{2}\omega^{(1)}_n \\
\Delta \theta^{(2)}_{n+1} & = \left( c_2 - 0.1\theta^{(2)}_n \right) + \sqrt{2}\omega^{(2)}_n
\end{align*}
\]

where \(c_1\) and \(c_2\) are coefficients which determine the potential well location and therefore the steady state.

The object location is measured in polar coordinates, through the azimuth and radius, with additive Gaussian noise

\[
\begin{align*}
\phi_n & = \arctan\left( \frac{\theta_n^{(1)}}{\theta_n^{(2)}} \right) + \nu_n^{(\phi)} \\
r_n & = \sqrt{\left( \theta_n^{(1)} \right)^2 + \left( \theta_n^{(2)} \right)^2} + \nu_n^{(r)}
\end{align*}
\]

where \(\nu_n^{(\phi)}\) and \(\nu_n^{(r)}\) are Gaussian noise processes with variance \(\sigma_{\phi}^2\) and \(\sigma_r^2\), respectively. We denote the system measurements by \(z_n = [\phi_n, r_n]\). Note that when \(\theta_n^{(1)}/\theta_n^{(2)}\) is close to zero, \(\arctan\) is approximately linear. Moreover, when one of the underlying coordinates is close to zero, \(r_n\) is approximately linear as well. Based on these characteristics we examined two regimes by considering two different choices of \(c_1\) and \(c_2\): a non-linear regime with \(c_1 = 3, c_2 = 0.5\), and an approximately linear regime with \(c_1 = 0.001, c_2 = 3\).

For each regime, we created trajectories of 1000 samples with different Signal-to-Noise ratios. We used a time-step of \(\Delta t = 0.01\).

We applied the diffusion maps algorithm to the measurements, \(z_n, r_n\), described above, with \(\epsilon\) set as the median of the distances. We then constructed the Diffusion Maps Kalman (DMK) based on the first two largest eigenvalues and corresponding eigenvectors obtained by diffusion maps. The dimensionality of the diffusion maps coordinates was determined based on the existence of a significant spectral gap after the second coordinate.

Using this setting, we evaluated our DMK algorithm, which requires access only to the noisy measurements, \(z_n\), and compared it to several parametric algorithms: the Extended Kalman Filter (EKF), the Particle Filter (PF), the optimal Kalman Filter (Opt. KF) on a corresponding linear problem, and to a work by Tan et al. [6], which proposes to use virtual measurements in order to linearize the non-linear tracking problem and apply a linear time-varying Kalman filter. In [6], the measurement equations presented in (28) are linearized by defining the following virtual measurements and corresponding mapping:

\[
y_n = \begin{bmatrix} 0 \\ r_n \end{bmatrix}, \quad H_n = \begin{bmatrix} \cos \phi_n & -\sin \phi_n \\ \sin \phi_n & \cos \phi_n \end{bmatrix}
\]

such that \(y_n = H_n \theta_n + u_n\), where \(u_n\) is a suitable noise term described in [6]. We denoted the latter algorithm by Virtual Measurements Kalman (VMK). For evaluation purposes, all parametric algorithms were provided with the true system model, which is considered unknown in our setting.

In addition, we compared our results to two non-parametric algorithms: Gaussian Process filtering (GP-ADF) [1] and to the Kalman filter based algorithm described in [3] (KKF). Note that there is a fundamental difference between our method and the competing non-parametric methods. Both of the non-parametric algorithms require a subset of data pairs, \(\{\theta_n, z_n\}_{n=1}^N\), to construct the algorithm. We used a subset of \(N = 100\) samples covering the entire state space in our experiments. In contrast, our DMK framework can provide a new set of coordinates representing the system characteristics without any information on the underlying state. However, in order to obtain an estimate of a specific state representation, an alignment between the DMK coordinates and the underlying state is required and can also be obtained using a subset of pairs of the system state and corresponding measurement. We note that another option is to use the mapping defined in [19], \(\alpha\), and obtain an estimate of the clean system measurement, \(g(\theta_n)\). In the experiments presented in this section, we obtained an estimate for the clean measurements as described and then used the true coordinate transforms to calculate the error in the system state domain, only for the purpose of comparison between the algorithms.

Figure [1] presents plots of average root mean square error (RMSE) values in log scale, of the state estimates of \(\theta_n^{(1)}, \theta_n^{(2)}\) obtained by the compared algorithms and the measurement error (denoted by Meas.) as an upper bound. Each plot represents a different regime as described above, where plots (a) and (c) describe the non-linear regime, presenting the compared parametric algorithms, and plots (b) and (d) describe the approximately linear regime, presenting the compared non-parametric algorithms. These plots depict that, as expected, the optimal Kalman filter performs best and that all parametric algorithms, which have access to the true system equations, obtain results which are close to the optimal Kalman. Our DMK algorithm obtains results which are comparable to the parametric algorithms, even though it is based solely on the system measurements. Moreover, plots (a) and (b) imply that DMK is not affected by the measurement regime, in contrast to the EKF for example, whose performance degrades in the non-linear case (plot (a)). Compared with the non-parametric algorithms, in plots (c) and (d), our method significantly outperforms the Gaussian Process based algorithm (GP-ADF).

In addition, for most SNR values and both regimes, our method outperforms the KKF algorithm from [3], which is more suitable for a deterministic state equation.

We note that the examined system is especially suited to our framework, since its underlying state equation is a Langevin equation, specifically, an Ornstein-Uhlenbeck process, and the measurement noise is Gaussian. In addition, the eigenfunctions of the backward Fokker-Planck operator of such a process are the Hermite polynomials [25] and therefore, the norm element in (16) of the first two coordinates is constant or varies relatively slow. Yet, we believe that our method still reveals relevant properties in systems, which do not admit the full theoretical model. In the next subsection, we demonstrate a case where the measurement noise process is non-Gaussian.
Fig. 1: Average and standard deviation of RMSE of the system state estimates. The RMSE values were averaged based on 50 realizations of trajectories. Plots (a) and (c) depict a non-linear measurement regime for the parametric and non-parametric methods respectively, where $c_1 = 3, c_2 = 0.5$, and plots (b) and (d) depict an approximately linear measurement regime for the parametric and non-parametric methods respectively, where $c_1 = 0.001, c_2 = 3$.

In addition, we have examined different underlying state equations with quadratic potentials, for which the corresponding eigenfunctions are not the Hermite polynomials, and obtained similar results.

$$\Delta \theta_{n+1}^{(1)} = \left( \frac{\pi}{2} \cdot c - c \cdot \theta_n^{(1)} \right) + b \omega_n^{(1)}$$  \hspace{1cm} (30) \\
$$\Delta \theta_{n+1}^{(2)} = \left( \frac{\pi}{10} \cdot c - c \cdot \theta_n^{(2)} \right) + b \omega_n^{(2)}$$  \hspace{1cm} (31)

B. Non-Gaussian Non-Linear Object Tracking

We now present an example which is based on the setting given in [21], [26]. In this setting, the location of a radiating object moving on a 3D sphere is estimated based on measurements from 3 sensors, $s^{(1)}, s^{(2)}, s^{(3)}$, which are modeled as “Geiger Counters”. The movement of the object is defined by two underlying Langevin processes, describing the elevation and azimuth angles as follows

where $b$ is the diffusion coefficient and $c$ is the drift rate parameter. In our experiments, $b$ and $c$ were set to 0.05 and 0.05 respectively. We note that other values of $b$ and $c$ led to similar results.

The 3D location of the object at the $n$th time step is
computed by:
\[
x_n^{(1)} = \cos(\theta_n^{(2)}) \sin(\theta_n^{(1)}) \\
x_n^{(2)} = \sin(\theta_n^{(2)}) \sin(\theta_n^{(1)}) \\
x_n^{(3)} = \cos(\theta_n^{(1)})
\] (32)

We mark the 3D position of the object by \( x_n = [x_n^{(1)}, x_n^{(2)}, x_n^{(3)}] \).

The system measurements are given by 3 Poisson processes, with a rate parameter that is based on the 3D location of the object.
\[
y_n^{(j)} \sim \text{Pois} \left( r_n^{(j)} \right) \quad j = 1, 2, 3
\] (33)

where \( r_n^{(j)} = \exp \left( -\| s^{(j)} - x_n \| \right) \).

Finally, a Poisson noise process with a fixed rate parameter, \( v_n^{(j)} \sim \text{Pois} (\lambda_v) \), is added to each sensor
\[
z_n^{(j)} = y_n^{(j)} + v_n^{(j)}
\] (34)

where \( z_n^{(j)} \) are the accessible system measurements.

Note that the presented setting is non-Gaussian and therefore the Kalman filter assumptions are not held. However, we show that DMK still provides good results and improves the observer framework presented in [21].

In order to obtain an estimated representation of the system state (the azimuth and elevation angles) from the noisy measurements, \( \{ x_n \}_{n=1}^N \), we apply DMK.

We simulated 300, 000 time samples of the two underlying angles \( \theta_n^{(1)}, \theta_n^{(2)} \) with \( \Delta t = 1 \) and constructed the measurements according to (32) and (34). After obtaining the system measurements \( z_n \), we first perform a pre-processing stage, similarly to [21], [26]. This includes constructing histograms for overlapping frames of 60 time-samples of \( z_n \) and then calculating the modified Mahalanobis distance between pairs of histograms. At this point we are left with 300, 000/60 = 5000 system measurements.

The diffusion maps algorithm, described in Subsection III-B, is applied to the measurements using the calculated Mahalanobis, with an empirical choice of \( \epsilon \) as the median of the Euclidean distances. We obtain a set of eigenvectors and eigenvalues representing intrinsic properties of the system. However, these eigenvectors do not necessarily correspond to the true system state, \( \theta_n \), and can represent some linear combination of the state coordinates [27]. Therefore, for evaluation purposes, we perform a linear regression on 4 eigenvectors, corresponding to the largest eigenvalues, and the true system state, based on the first 1000 samples.

Based on the resulting eigenvectors and eigenvalues we construct the Kalman filter, described in Subsection III-D. The covariance matrices of the measurement noise, \( R \), and the state noise, \( Q \), were estimated from the data (according to the variance of the histogram measurements and the covariance of the obtained eigenvectors). We note that our estimates can be improved by using the adaptive estimation for \( Q \) and \( R \) described in [28], for unknown system dynamics and measurement function.

We compare our results to the observer framework, described in [21], with a choice of \( \gamma = 0.25 \) (which led to the best results in this case). Figure 2 presents an example of a normalized trajectory of the elevation angle \( \theta_n^{(1)} \) estimated by DMK (in dashed red) compared with the estimated trajectory obtained by the observer framework (in dotted green) and the true underlying elevation angle (in blue). We observe that the estimated DMK trajectory describes the true angle more accurately in this example. Two exemplary locations highlighting the benefit are marked with black arrows.

In Figure 3 a comparison between the DMK, the observer framework and the diffusion maps coordinates (without additional analysis) is presented. This figure contains 6 identical scatter plots, each presenting the true underlying angles, \( \theta_n^{(1)}, \theta_n^{(2)} \). Each plot is colored according to a different coordinate, plots (a) and (d) are colored according to the first and second estimated coordinates of the observer framework, plots (b) and (e) are colored according to the DMK estimation and plots (c) and (f) are colored according to the diffusion maps coordinates (DM). The color gradients in Figure 3 depict that the DMK significantly improves the estimation of the two underlying angles, compared with the diffusion maps coordinates and the observer framework. Moreover, the coordinates obtained by the observer framework suffer from inaccuracies at the boundaries of the data. This is visible for example, in plot (a), when \( \theta_n^{(1)} < 0.5 \) and is due to the inaccuracy of the linear lift function at the boundaries. These inaccuracies are not apparent in the DMK coordinates which recover the true underlying angles more accurately even at the boundaries of the data.

These results are summarised in Table I which presents the RMSE between the true underlying angles, \( \theta_n^{(1)}, \theta_n^{(2)} \), and their estimates, using the DMK framework, the observer framework and the diffusion maps coordinates. The RMSE values are averaged over 50 realizations of angle trajectories, described in [22] (simulated with different initial conditions).

V. CONCLUSION

In this work we presented a non-parametric filtering framework for high-dimensional nonlinear systems, in which a linear Kalman filter is constructed based on diffusion maps coordinates and their inherent dynamics. Our devised method is especially well-suited for stochastic systems with an unknown model. We showed that the constructed framework reveals a
new set of meaningful coordinates related to the underlying system state, and obtains an improved representation of dynamical systems compared with the diffusion maps coordinates. We further showed that our method outperforms other non-parametric methods in a class of stochastic dynamical systems settings.

In the future, we plan to extend our method and address multi-modal data-sets arising from stochastic dynamical systems. We will devise methods for revealing the underlying common dynamics based on measurements from different sensors.

### TABLE I: Average and standard deviation of RMSE values (over 50 realizations) calculated for the DMK coordinates, $\psi_{DMK}$, the coordinates of the observer framework, $\psi_{Observer}$, and the diffusion maps coordinates, $\psi_{DM}$.

|                | $\psi_{DMK}$ | $\psi_{Observer}$ | $\psi_{DM}$ |
|----------------|--------------|-------------------|------------|
| RMSE $\theta_{(1)}$ | 0.76 ± 0.088 | 0.92 ± 0.035 | 1.05 ± 0.025 |
| RMSE $\theta_{(2)}$ | 0.61 ± 0.02  | 0.64 ± 0.02  | 0.99 ± 0.02  |

Fig. 3: Scatter plots of the azimuth and elevation angles, colored by the coordinate estimates. The plots are colored according to the the first and second state estimates of the observer framework (plots (a) and (d)), the first and second state estimates of our suggested DMK filter (plots (b) and (e)) and the first and second coordinates obtained by diffusion maps (plots (c) and (f)).

**REFERENCES**

[1] M. P. Deisenroth, R. D. Turner, M. F. Huber, U. D. Hanebeck, and C. E. Rasmussen, “Robust filtering and smoothing with gaussian processes,” *IEEE Transactions on Automatic Control*, vol. 57, no. 7, pp. 1865–1871, 2012.

[2] Y. Wang, M. A. Brubaker, B. Chaib-draa, and R. Urtasun, “Bayesian filtering with online gaussian process latent variable models,” in *UAI*, 2014, pp. 849–857.

[3] A. Surana and A. Banaszuk, “Linear observer synthesis for nonlinear systems using koopman operator framework,” *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 716–723, 2016.

[4] S. L. Brunton, B. W. Brunton, J. L. Proctor, E. Kaiser, and J. N. Kutz, “Chaos as an intermittently forced linear system,” *Nature Communications*, vol. 8, 2017.

[5] T. Berry, D. Giannakis, and J. Harlim, “Nonparametric forecasting of low-dimensional dynamical systems,” *Physical Review E*, vol. 91, no. 3, pp. 032915, 2015.

[6] F. Tan, W. Lohmiller, and J. J. Slotine, “Analytical slam without linearization,” in *Robotics Research*, pp. 89–105. Springer, 2018.

[7] I. Mezić, “Spectral properties of dynamical systems, model reduction and decompositions,” *Nonlinear Dynamics*, vol. 41, no. 1, pp. 309–325, 2005.

[8] M. Budišić, R. Mohr, and I. Mezić, “Applied koopmanism a,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, no. 4, pp. 047510, 2012.

[9] 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*, vol. 25, no. 6, pp. 1307–1346, December 2015.

[10] I. Mezić and A. Banaszuk, “Comparison of systems with complex behavior,” *Physica D: Nonlinear Phenomena*, vol. 197, no. 1, pp. 101–113, 2004.

[11] R. Mohr and I. Mezić, “Construction of eigenfunctions for scalar-type operators via laplace averages with connections to the koopman operator,” arXiv preprint arXiv:1512.05997, 2015.

[12] S. M. Ulam, “Problems of modern mathematics,” *Science Editions, New York, 1964. Originally published as: A collection of mathematical problems*, 1960.

[13] S. Klus, P. Koltai, and C. Schütte, “On the numerical approximation of the perron-frobenius and koopman operator,” arXiv preprint arXiv:1512.05997, 2015.

[14] N. Črnjarić-Žic, S. Mačeskić, and I. Mezić, “Koopman operator spectrum for random dynamical systems,” arXiv preprint arXiv:1711.03146, 2017.

[15] T. Berry and J. Harlim, “Nonparametric uncertainty quantification for stochastic gradient flows,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 3, no. 1, pp. 484–508, 2015.
[16] F. Hamilton, T. Berry, and T. Sauer, “Ensemble kalman filtering without a model,” Physical Review X, vol. 6, no. 1, pp. 011021, 2016.
[17] D. Giannakis, “Data-driven spectral decomposition and forecasting of ergodic dynamical systems,” Applied and Computational Harmonic Analysis, in press, 2017.
[18] R. Coifman and S. Lafon, “Diffusion maps,” Appl. Comput. Harmon. Anal., vol. 21, pp. 5–30, Jul. 2006.
[19] F. Takens, “Dynamical systems and turbulence (detecting strange attractors in fluid turbulence),” 1981.
[20] R.R. Coifman, I.G. Kevrekidis, S. Lafon, M. Maggioni, and B. Nadler, “Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems,” Multiscale Modeling & Simulation, vol. 7, no. 2, pp. 842–864, 2008.
[21] T. Shnitzer, R. Talmon, and J. J. E. Slotine, “Manifold learning with contracting observers for data-driven time-series analysis,” IEEE Transactions on Signal Processing, vol. 65, no. 4, pp. 904–918, 2017.
[22] A. Singer and R. Coifman, “Non-linear independent component analysis with diffusion maps,” Appl. Comput. Harmon. Anal., vol. 25, pp. 226–239, 2008.
[23] B. Nadler, S. Lafon, R. Coifman, and I. G. Kevrekidis, “Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators,” Neural Information Process. Systems (NIPS), vol. 18, 2005.
[24] Y. Bulut and O. Bayat, “Kalman filtering with model uncertainties,” in Topics in Modal Analysis I, Volume 5, pp. 447–455. Springer, 2012.
[25] H. Risken, Fokker-planck equation, Springer, 1984.
[26] R. Talmon and R.R. Coifman, “Empirical intrinsic geometry for nonlinear modeling and time series filtering,” Proc. Nat. Acad. Sci., vol. 110, no. 31, pp. 12535–12540, 2013.
[27] C.J. Dsilva, R. Talmon, R.R. Coifman, and I.G. Kevrekidis, “Parsimonious representation of nonlinear dynamical systems through manifold learning: A chemotaxis case study,” Applied and Computational Harmonic Analysis, 2015, In Press.
[28] T. Berry and T. Sauer, “Adaptive ensemble kalman filtering of non-linear systems,” Tellus A: Dynamic Meteorology and Oceanography, vol. 65, no. 1, pp. 20331, 2013.