Approximating the Koopman Operator using Noisy Data: Noise-Resilient Extended Dynamic Mode Decomposition

Masih Haseli and Jorge Cortés

Abstract— **This paper presents a data-driven method to find a finite-dimensional approximation for the Koopman operator using noisy data. The proposed method is a modification of Extended Dynamic Mode Decomposition which finds an approximation for the projection of the Koopman operator on a subspace spanned by a predefined dictionary of functions. Unlike the Extended Dynamic Mode Decomposition which is based on least squares method, the presented method is based on element-wise weighted total least squares which enables one to find a consistent approximation when the data come from a static linear relationship and the noise at different times are not identically distributed. Even though the aforementioned method is consistent, it leads to a nonconvex optimization problem. To alleviate this problem, we show that under some conditions the nonconvex optimization problem has a common minimizer with a different method based on total least squares for which one can find the solution in closed form.**

I. INTRODUCTION

In recent years, operator theoretic point of view regarding dynamical systems has gained a widespread attention since it enables researches to analyze high-dimensional nonlinear systems with acceptable computational complexity. Koopman operator plays a crucial role in the aforementioned point of view. Koopman operator is a linear infinite-dimensional operator which represents a dynamical systems. Despite the linearity of the operator, it cannot be implemented or analyzed directly using digital computers due to its infinite-dimensional nature. In order to alleviate this problem, one can find a finite-dimensional approximation for the Koopman operator using data-driven methods such as dynamic mode decomposition (DMD) and extended dynamic mode decomposition (EDMD) and their variants.

Since DMD and EDMD are data-driven methods, one must pay attention to the measurement noise. The noise problem in DMD has been addressed using total least squares methods. The measurement noise problem is much more complicated in EDMD since the noisy data go through a dictionary of nonlinear functions which distorts the noise distribution. This paper addresses the noise problem in EDMD.

Literature Review. The Koopman operator was first introduced in 1931 [1], [2]. Being a linear operator, it can be used to analyze a nonlinear dynamical system via the spectral properties of the operator [3]–[6]. This leads to a diverse range of applications including system identification [7], model validation [8], and control [9]–[12]. Even though the Koopman operator was introduced in 1931, its use has only become common recently due to the lack of practical ways to find representations for it. Two main numerical methods to approximate finite-dimensional representations for the Koopman operator are DMD and EDMD.

DMD was introduced to find coherent structures of nonlinear fluid flows from time-series data [13] originally. Later, it was modified to work with snapshot data without any predefined order [14]. DMD finds a linear relationship between the data snapshots acquired from the system and it is in a close relationship with the Koopman operator on one side and with the least squares method on the other side. The noise problem for DMD has been addressed in the case that the noise has known mean and is additive to the data matrices which are used directly in the DMD algorithm [15], [16].

EDMD is an extension of DMD such that the states of the system go through a generally nonlinear dictionary of functions to create dictionary snapshots, then a linear relationship between the dictionary snapshots will be found using least squares method [17]. The convergence of EDMD to the Koopman operator has been investigated in [18]. It is worth mentioning that to the best of the authors' knowledge, the measurement noise problem in the EDMD has not been fully investigated in the literature. However, one can find a case regarding random dynamical systems in [19].

Statement of Contributions: We consider the noise problem in EDMD[1]. In the EDMD method, the noisy data first go through a dictionary of nonlinear functions. This leads to distortion of the noise, i.e., the perturbations in the rows of data matrices will not remain identically distributed. Consequently, we cannot use the standard total least squares (TLS) methods used in [15], [16] to calculate a consistent approximation. In order to find a consistent approximation from noisy data, we propose a method based on element-wise weighted total least squares (EWTLS) which can find

This work was partially supported by ONR Award N00014-18-1-2828.

Masih Haseli and Jorge Cortés are with Department of Mechanical and Aerospace Engineering, University of California, San Diego, CA 92093, USA, {mhaseli,cortes}@ucsd.edu

[1]Throughout the paper, we use the following notation: $\mathbb{R}$ and $\mathbb{N}$ denote real and natural numbers, resp. For a matrix $A \in \mathbb{R}^{m\times n}$, we denote the Frobenius norm, pseudo-inverse, and ith row of A by $\|A\|_F$, $A^\dagger$, and A_i. When $m = n$, we denote the trace, inverse, and eigenvalues of A by $\mathrm{Tr}(A)$, A^{-1}, and $\lambda(A)$, resp. Given matrices $A_1, \dots, A_n$, we use $\mathrm{blkdiag}(A_1, \dots, A_n)$ to denote the block-diagonal matrix with diagonal blocks $A_1, \dots, A_n$. Given $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times d}$, we denote by $[A, B] \in \mathbb{R}^{m\times(n+d)}$ the matrix created by concatenating A and B side by side. We denote the real and imaginary parts of a complex number a by $\mathrm{Re}(a)$ and $\mathrm{Im}(a)$, resp. For a random vector $x \in \mathbb{R}^n$, we denote by $\mathbb{E}[x]$ and $\mathrm{Cov}[x]$ its expected value and covariance matrix. We denote the Euclidean norm by $\|\cdot\|_2$. Given functions $f : \mathcal{M} \to \mathcal{K}$ and $g : \mathcal{K} \to \mathcal{F}$, we let $g \circ f : \mathcal{M} \to \mathcal{F}$ denote its composition. Given a positive measure μ on $\mathcal{M}$, we denote by $L_2(\mu)$ the set of measurable functions $f : \mathcal{M} \to \mathbb{R}$ satisfying $\int_{\mathcal{M}} |f(x)|^2 \,\mathrm{d}\mu(x) < \infty$.

consistent approximation in cases that the noise is not identically distributed. The proposed method leads to a nonconvex optimization problem for which finding the global minimizer is arduous if not impossible. To solve this problem, we present an optimization method based on standard TLS for which one can find the closed-form solution, and we show that the solution of the two presented methods are the same under some conditions. Due to lack of space, the proofs are omitted and will appear elsewhere.

II. Preliminaries

We present preliminaries on total least squares, the Koopman operator, and extended dynamic mode decomposition.

A. Total Least Squares Method

Given matrices $\tilde{A} = A + \Delta A \in \mathbb{R}^{m\times n}, \tilde{B} = B + \Delta B \in \mathbb{R}^{m\times d}$ are the data matrices where $AX^* = B$ and $m > n$, one can use total least squares (TLS) to find approximate solutions to the overdetermined system of linear equations $\tilde{A}X \approx \tilde{B}$. Using variables $\Delta_1 \in \mathbb{R}^{m\times n}$, $\Delta_2 \in \mathbb{R}^{m\times d}$, and $X \in \mathbb{R}^{n\times d}$, the TLS optimization problem is defined as follows, see e.g. [20],

$$\begin{aligned} \underset{X,\Delta_1,\Delta_2}{\text{minimize}} \quad & \|[\Delta_1, \Delta_2]\|_F^2 \\ \text{subject to} \quad & (\tilde{A} + \Delta_1)X = (\tilde{B} + \Delta_2). \end{aligned} \tag{1}$$

Let $U \operatorname{diag}(\sigma_1, \dots, \sigma_{n+d})V^T$ be the singular value decomposition of $[\tilde{A}, \tilde{B}]$, with $\sigma_1 \geq \cdots \geq \sigma_{n+d}$. For convenience, we block-partition V as

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where $V_{11} \in \mathbb{R}^{n\times n}$, $V_{12} \in \mathbb{R}^{n\times d}$, $V_{21} \in \mathbb{R}^{d\times n}$, and $V_{22} \in \mathbb{R}^{d\times d}$. The solution of (1) exists if and only if V_{22} is nonsingular. Additionally, the solution is unique if and only if $\sigma_n \neq \sigma_{n+1}$. The unique solution is given by

$$\begin{aligned} \bar{X} &= -V_{12}V_{22}^{-1}, \\ \left[\bar{\Delta}_1, \bar{\Delta}_2\right] &= -U \operatorname{diag}(0, \dots, 0, \sigma_{n+1}, \dots, \sigma_{n+d})V^T. \end{aligned}$$

An appealing feature [21] of the TLS method is that the matrix $\bar{X}$ approximates X^* consistently when the rows of $[\Delta A, \Delta B]$ are zero-mean, independent and identically distributed (i.i.d.) measurement noise.

In the case that the noise is zero-mean but not identically distributed, one can use instead the element-wise weighted total least squares (EWTLS) method to find a consistent approximation [22], [23]. For general true data matrices $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times d}$ with $AX^* = B$, and noisy data matrices $\tilde{A} = A + \Delta A$, $\tilde{B} = B + \Delta B$, the EWTLS problem is defined as follows

$$\begin{aligned} \underset{X,\Delta_1,\Delta_2}{\text{minimize}} \quad & \sum_{i=1}^{N} \left\| [\Delta_1, \Delta_2]_i \, C_i^{-1/2} \right\|_2^2 \\ \text{subject to} \quad & (\tilde{A} + \Delta_1)X = (\tilde{B} + \Delta_2), \end{aligned}$$

where $C_i = \operatorname{Cov}\left([\Delta A, \Delta B]_i\right)$. The EWTLS cost function, scales the rows of the variables in order to account for the difference in the covariance of the rows of $[\Delta A, \Delta B]$. The EWTLS method provides a weakly consistent approximation for X^* [22]. It is important to note that unlike TLS, the EWTLS problem does not have a closed-form solution.

B. The Koopman Operator

We start by introducing the Koopman operator [5]. Consider a discrete-time autonomous dynamical system over the state space $\mathcal{M} \subseteq \mathbb{R}^{n_s}$

$$x^+ = T(x), \tag{2}$$

where $T : \mathcal{M} \to \mathcal{M}$. Let $\mathcal{F}$ be a vector space of functions (also called observables) defined on $\mathcal{M}$ such that $f \circ T \in \mathcal{F}$ for every $f \in \mathcal{F}$. One standard choice of $\mathcal{F}$ is $\mathcal{F} = L_2(\mu)$, where μ is a given positive measure on $\mathcal{M}$. The Koopman operator $\mathcal{K} : \mathcal{F} \to \mathcal{F}$ associated with the dynamical system (2) is defined as $\mathcal{K}(f) = f \circ T$.

It is important to note that, unlike the dynamical description (2) which defines the evolution of states in $\mathcal{M}$, the Koopman operator acts on functions in $\mathcal{F}$. The Koopman operator is linear even if the underlying dynamics (2) is nonlinear. Moreover, it is infinite-dimensional. The Koopman operator approach allows us to analyze nonlinear dynamical systems using the spectral properties of the operator while preserving the global features of the system. The viewpoint, centered around observables, is inherently global, whereas that of the traditional dynamics (2) is local, which has important consequences for assimilation.

The fact that Koopman operator is generally infinite dimensional makes it difficult to apply conventional linear methods to compute or analyze it. One efficient way to overcome this problem is by finding a finite-dimensional approximation using data-driven methods such as Extended Dynamic Mode Decomposition (EDMD) [17]. In this method, a dictionary of functions, typically nonlinear, is used to lift the states to a higher-dimensional space and then a linear relationship among the values of the lifted states for consecutive time steps is found. Formally, let the matrices $X, Y \in \mathbb{R}^{N\times n_s}$ encode N snapshots of data obtained from the system (2), meaning that $y_i = T(x_i)$ for $i \in \{1, \dots, N\}$. Where x_i^T and y_i^T denote the ith rows of X and Y, resp. Let $D : \mathbb{R}^{n_s} \to \mathbb{R}^{1\times N_d}$ with $D(x) = [d_1(x), d_2(x), \dots, d_{N_d}(x)]$ be a dictionary of N_d functions taken from the space $\mathcal{F}$. For convenience, we define $D_N : \mathbb{R}^{N\times n_s} \to \mathbb{R}^{N\times N_d}$ by $D_N(X) = [D(x_1)^T, \dots, D(x_N)^T]^T$.

The EDMD algorithm computes an approximation to the Koopman operator by solving the least squares optimization problem $\min_K \|D_N(Y) - D_N(X)K\|_F^2$. The closed-form solution takes the form

$$K_{\text{Approx}} = D_N(X)^\dagger D_N(Y). \tag{3}$$

Given the presence of the pseudo-inverse of $D_N(X) \in \mathbb{R}^{N\times N_d}$ in (3), the computational cost of the EDMD algorithm grows rapidly as N_d increases. One can address this problem by using kernel methods [24]. Equation (3) also shows that K_{Approx} depends on the choice of dictionary, which raises the possibility of optimizing this choice. Generally, if the dictionary does not span a subspace invariant under the Koopman operator, we loose some information about it (and consequently the system) by finding a finite-dimensional approximation. However, choosing an appropriate dictionary still results in capturing the important characteristics of the operator.

III. Problem Statement

In this paper, we are inspired by the observations that noise is always present when collecting data and that the EDMD algorithm does not explicitly account for this fact. EDMD relies on least squares optimization which, in case of noisy measurements, leads to inconsistent approximations when all data matrices are perturbed, see e.g. [20], [21]. Another point to keep in mind is that attempts to find an appropriate finite-dimensional approximation of the Koopman operator may not be successful using arbitrary dictionaries. Our aim here is to develop a computational procedure to construct approximations of the Koopman operator that takes measurement noise into consideration. Formally, let the measurement noise be a random vector in $\mathbb{R}^{1\times n_s}$ with arbitrary but known distribution. Consequently, the measured data takes the form

$$\tilde{X} = X + \Delta X, \tag{4a}$$

$$\tilde{Y} = Y + \Delta Y, \tag{4b}$$

where the rows of $\tilde{X}, \tilde{Y} \in \mathbb{R}^{N\times n_s}$ are N measured data snapshots, $X, Y \in \mathbb{R}^{N\times n_s}$ represent the corresponding true states of the system, and $\Delta X, \Delta Y \in \mathbb{R}^{N\times n_s}$ are matrices comprised of measurement noise realizations.

The objective of this paper is design an algorithm to find an approximate static relationship between dictionary snapshots in the form $D_N(Y) \approx D_N(X)K_N$ using the noisy dictionary snapshots $D_N(\tilde{X}), D_N(\tilde{Y})$. Moreover, if the dictionary spans an invariant Koopman subspace, i.e., if $D_N(Y) = D_N(X)\bar{K}$ for some $\bar{K} \in \mathbb{R}^{N_d\times N_d}$ and all $N \in \mathbb{N}$, then the solution provided by the algorithm must satisfy $K_N \to \bar{K}$.

Our strategy to tackle this starts with a discussion in Section IV of the limitations of EDMD in the presence of noisy data, which leads us to propose a method based on element-wise weighted total least squares (EWTLS). To circumvent the fact that EWTLS does not have a closed-form solution, we propose an alternative method based on standard total least squares (TLS) in Section V and show that, under reasonable assumptions, both methods have a common solution.

IV. Noise-Resilient Extended Dynamic Mode Decomposition

In this section, we start by analyzing the behavior of EDMD when employing noisy data. We rely on this analysis to propose an alternative algorithm based on total least squares optimization to overcome the limitations of EDMD.

A. Limitations of Extended Dynamic Mode Decomposition

Suppose that we feed noisy data $\tilde{X}, \tilde{Y}$ directly to the EDMD algorithm in order to find matrix K_{noisy} as the minimizer of the following optimization problem

$$\underset{K}{\text{minimize}} \quad \left\| D_N(\tilde{Y}) - D_N(\tilde{X})K \right\|_F^2. \tag{5}$$

In order to determine if K_{noisy} is close to $K_{\text{Approx}} = \operatorname{argmin} \|D_N(Y) - D_N(X)K\|_F^2$, we examine the cost function in (5) acting on exact and noisy data.

Lemma 4.1: (Cost Functions of EDMD Acting on Exact and Noisy Data): Let $J(K; X, Y) = \|D_N(Y) - D_N(X)K\|_F^2$ be the cost function of EDMD, where $K \in \mathbb{R}^{N_d\times N_d}$ and $X, Y, \Delta X, \Delta Y \in \mathbb{R}^{N\times n_s}$. Then,

$$\begin{aligned} &J(K; X+\Delta X, Y+\Delta Y) \\ &= J(K; X, Y) + J(K; \Delta D_N(X, \Delta X), \Delta D_N(Y, \Delta Y)) \\ &\quad + 2\operatorname{Tr}\big[(\Delta D_N(Y,\Delta Y) - \Delta D_N(X,\Delta X)K)^T \\ &\quad (D_N(Y) - D_N(X)K)\big], \end{aligned} \tag{6}$$

where $\Delta D_N(X, \Delta X)$ and $\Delta D_N(Y, \Delta Y)$ are given by

$$\Delta D_N(X, \Delta X) = D_N(X + \Delta X) - D_N(X), \tag{7a}$$

$$\Delta D_N(Y, \Delta Y) = D_N(Y + \Delta Y) - D_N(Y). \tag{7b}$$

□

Equation (6) clearly shows that applying the EDMD on the data while ignoring the measurement noise, i.e., minimizing the left-hand side of (6), does not lead to the correct answer, which would correspond to minimizing the first term on the right-hand side of (6).

Note that the EDMD optimization problem can be written in the following form

$$\begin{aligned} &\underset{K,\Delta}{\text{minimize}} && \|\Delta\|_F^2 \\ &\text{subject to} && D_N(Y) + \Delta = D_N(X)K. \end{aligned} \tag{8}$$

The optimization (8) makes it clear that the least squares algorithm employed in EDMD finds the smallest correction for $D_N(Y)$ to satisfy the constraint, assuming there is no error in $D_N(X)$. The asymmetry created by correcting $D_N(Y)$ while keeping $D_N(X)$ intact is responsible for the inconsistent answers obtained by least squares when X is noisy, see e.g. [15], [16], [21].

In addition, since in EDMD data go through a non-linear dictionary, the error matrix is not generally zero mean and i.i.d. In order to resolve this issue, one needs to investigate the statistical properties of $\Delta D_N(X, \Delta X)$ and $\Delta D_N(Y, \Delta Y)$. Considering the rows of these matrices as functions of measurement noise, we define for convenience,

$$X_{\text{bias}} = \mathbb{E}[\Delta D_N(X, \Delta X)], \tag{9a}$$

$$Y_{\text{bias}} = \mathbb{E}[\Delta D_N(Y, \Delta Y)], \tag{9b}$$

$$\begin{aligned} C_i^X = \mathbb{E}\Big[\, &[\Delta D_N(X, \Delta X) - X_{\text{bias}}]_i^T \\ &[\Delta D_N(X, \Delta X) - X_{\text{bias}}]_i \Big], \end{aligned} \tag{9c}$$

$$C_i^Y = \mathbb{E}\Big[[\Delta D_N(Y,\Delta Y) - Y_{\text{bias}}]_i^T \\ [\Delta D_N(Y,\Delta Y) - Y_{\text{bias}}]_i \Big], \tag{9d}$$

for each $i \in \{1,\dots,N\}$.

B. A Method Inspired by Element-Wise Weighted Total Least Squares

We propose here a method to incorporate measurement noise based on EWTLS. Formally, the noise-resilient extended dynamic mode decomposition (NREDMD) consists of solving the following optimization problem

$$\min_{K,\Delta_1,\Delta_2} \sum_{i=1}^{N} \big\| [\Delta_1,\Delta_2]_i \ \text{blkdiag}(C_i^X, C_i^Y)^{-1/2} \big\|_2^2 \tag{10a}$$

$$\text{subject to } D_N(\tilde{Y}) - Y_{\text{bias}} + \Delta_2 \\ = (D_N(\tilde{X}) - X_{\text{bias}} + \Delta_1)K. \tag{10b}$$

We investigate the consistency of NREDMD as $N \to \infty$ under the following assumptions.

Assumption 4.2: The rows of measurement noise matrix $[\Delta X, \Delta Y]$ are independent.

Assumption 4.3: There exist $\kappa,\mu,\eta > 0$ such that $\kappa \le \lambda(\text{blkdiag}(C_i^X, C_i^Y)) \le \mu$ and $\|[X_{\text{bias}}, Y_{\text{bias}}]_i\|_2^2 \le \eta$.

Assumption 4.4: There exists $r > N_d^2/2$ such that $\sup_{i\in\mathbb{N}, j\in\{1,2,\dots,2N_d\}} \mathbb{E}\left[[\Delta D_N(X,\Delta X), \Delta D_N(Y,\Delta Y)]_{ij}^{2r} \right] < \infty$.

Assumption 4.5: $\lim_{N\to\infty} \frac{1}{N} D_N(X)^T D_N(X)$ exists and is positive definite.

Assumptions 4.2-4.4 hold for typical choices of noise distribution and dictionary such as Gaussian noise with positive definite covariance matrix and polynomial or sinusoidal dictionaries. Assumption 4.5 is only about the dictionary and the initial conditions used to excite the system. This assumption requires both functions in the dictionary and the initial conditions used for acquiring the data to be diverse enough to capture the characteristics of the system over the domain of interest. A critical implication of this fact is that the functions in the dictionary must be linearly independent. Note that the rows of $[\Delta D_N(X,\Delta X), \Delta D_N(Y,\Delta Y)]$ are functions of the rows of $[\Delta X, \Delta Y]$. Consequently, using Assumption 4.2, one can conclude that the rows of $[\Delta D_N(X,\Delta X), \Delta D_N(Y,\Delta Y)]$ are independent.

Next, we state the consistency of NREDMD.

Theorem 4.6: (Consistency of NREDMD): Suppose there exists $K_{\text{Exact}} \in \mathbb{R}^{N_d\times N_d}$ such that $D_N(Y) = D_N(X)K_{\text{Exact}}$ for all $N \in \mathbb{N}$. Then, under Assumptions 4.2-4.5, problem (10) has a solution K_{NREDMD} with probability tending to 1 as $N \to \infty$. Moreover, this solution converges in probability to K_{Exact} as $N \to \infty$. □

Similar to the EDMD, the NREDMD method relies on the dictionary. Theorem 4.6 guarantees approximation in probability when the dictionary spans a Koopman invariant subspace, i.e., if the EDMD method finds the right answer using exact data, the NREDMD method approximates the true matrix consistently.

Remark 4.7: (Approximating (9)*):* Since we do not have access to the exact values X and Y due to noise, the quantities in (9) are not directly computable. Assuming the signal to noise ratio is high, one can use $\tilde{X}, \tilde{Y}$ instead of X, Y to approximate them. Our simulations show that the proposed algorithms are robust to the error in this approximation. □

We define $C_i = \text{blkdiag}(C_i^X, C_i^Y)$, where $i \in \{1,\dots,N\}$. As we have discussed above, problem (10) does not have a general closed-form solution. Even though there exist algorithms that can solve (10) numerically, see, e.g. [22], [23], it would be more desirable to have exact closed-form solutions. This is the task we tackle next.

V. Numerically Efficient NREDMD

Here, we propose an alternative algorithm based on TLS for which one can find the closed-form solution efficiently and establish the key fact that, under appropriate conditions, the proposed method shares a common minimizer with the NREDMD method proposed above.

For convenience, we use the shorthand notation $Z = [\Delta_1, \Delta_2]$. Consider the objective function

$$J_1^N(Z) = \frac{1}{N}\sum_{i=1}^{N} Z_i C_i^{-1} Z_i^T, \tag{11a}$$

Note that J_1^N is the normalized version of (10)'s cost function and since $N > 0$, this normalization does not change the minimizer. In the NREDMD method, each row of Z is normalized individually, which hinders calculating a closed-form solution for the optimization problem (10). Instead, here we propose to normalize the rows of Z using the sum of all covariance matrices. Therefore, let $C_\Sigma^X = \sum_{i=1}^N C_i^X$, $C_\Sigma^Y = \sum_{i=1}^N C_i^Y$, $C_\Sigma = \text{blkdiag}(C_\Sigma^X, C_\Sigma^Y)$ and consider the objective function

$$J_2^N(Z) = \|Z C_\Sigma^{-1/2}\|_F^2 = \sum_{i=1}^{N} Z_i C_\Sigma^{-1} Z_i^T. \tag{11b}$$

The method we propose to implement is then

$$\begin{aligned} &\underset{Z,K}{\text{minimize}}\ J_2^N(Z) \\ &\text{subject to (10b).} \end{aligned} \tag{12}$$

Remark 5.1: (Turning (12) *into Standard TLS):* Using the change of variables $Z' = Z C_\Sigma^{-1/2}$, $\Delta_1' = \Delta_1 (C_\Sigma^X)^{-1/2}$, $\Delta_2' = \Delta_2 (C_\Sigma^Y)^{-1/2}$, and $K' = (C_\Sigma^X)^{1/2} K (C_\Sigma^Y)^{-1/2}$, one can write (12) as

$$\begin{aligned} &\underset{Z',K'}{\text{minimize}} && \|Z'\|_F^2 \\ &\text{subject to} && \begin{bmatrix} A + Z' \end{bmatrix} \begin{bmatrix} K' \\ -I \end{bmatrix} = 0, \end{aligned} \tag{13}$$

where $A = [D_N(\tilde{X}) - X_{\text{bias}}, D_N(\tilde{Y}) - Y_{\text{bias}}] C_\Sigma^{-1/2}$. Problem (13) is a TLS and has a closed-form solution. □

Remark 5.1 justifies our focus in the rest of the section devoted to characterize the relationship between the cost functions of (10) and (12) as $N \to \infty$, and guaranteeing the existence of the limits. In order to tackle this, we constrain the variable Z which describes the perturbations in dictionary snapshot matrices, as follows

$$\begin{aligned} &\underset{Z,K}{\text{minimize}}\ J_1^N(Z) \\ &\text{subject to (10b)} \\ &\qquad \gamma \leq \|Z_i\|_2^2 \leq \Gamma, i \in \{1,\dots,N\}, \qquad (14) \\ &\underset{Z,K}{\text{minimize}}\ J_2^N(Z) \\ &\text{subject to (10b)} \\ &\qquad \gamma \leq \|Z_i\|_2^2 \leq \Gamma, i \in \{1,\dots,N\}. \qquad (15) \end{aligned}$$

Here the constants $0 < \gamma < \Gamma$ are design choices. By choosing γ sufficiently small and Γ sufficiently large, one can make the problems (14) and (15) have close solutions to the problems (10) and (12), resp.

The following result characterizes the behavior of J_1^N at the solutions of (14) as N goes to infinity.

Lemma 5.2: (Convergence of J_1^N to a Positive Limit at the Solutions of (14)*):* Suppose that Assumptions 4.2-4.5 hold. Let $\{Z^N\}$ and $\{K^N\}$ be sequences comprised of solutions of (14) using the first N data snapshots. If $\gamma \leq \|[D_N(\tilde{X})-X_{\text{bias}}, D_N(\tilde{Y})-Y_{\text{bias}}]_N\|_2^2 \leq \Gamma$ for every $N \in \mathbb{N}$ then $\lim_{N\to\infty} J_1^N(Z^N)$ exists and is positive. □

The following result characterizes the behavior of J_2^N at the solutions of (15) as N goes to infinity.

Lemma 5.3: (Convergence of J_2^N to a Positive Limit at the Solutions of (15)*):* Suppose that Assumptions 4.2-4.5 hold. Let $\{Z^N\}$ and $\{K^N\}$ be sequences comprised of solutions of (15) using the first N data snapshots. If $\gamma \leq \|[D_n(\tilde{X})-X_{\text{bias}}, D_n(\tilde{Y})-Y_{\text{bias}}]_N\|_2^2 \leq \Gamma$ for every $N \in \mathbb{N}$ then $\lim_{N\to\infty} J_2^N(Z^N)$ exists and is positive. □

Lemmas 5.2 and 5.3 facilitate connecting the cost functions of (14) and (15). Now, we are ready to establish the connection between the minimizers of (14) and (15) for a sufficiently large number of data snapshots.

Theorem 5.4: (Common Minimizer for (14) *and* (15)*):* Let Assumptions 4.2-4.5 hold and $\gamma \leq \|[D_N(\tilde{X}) - X_{\text{bias}}, D_N(\tilde{Y}) - Y_{\text{bias}}]_N\|_2^2 \leq \Gamma$ for every $N \in \mathbb{N}$. If the problems (14) and (15) have unique solutions for sufficiently large N, then for sufficiently large N those solutions are equal. □

Theorem 5.4 shows that the problems (14) and (15) are equivalent for sufficiently large number of data snapshots.

Remark 5.5: (Solving (15) *using a Standard TLS Problem):* Using the change of variables presented in Remark 5.1, problem (15) can be written as problem (13) with additional constraints $\gamma \leq \|[Z' C_\Sigma^{1/2}]_i\|_2^2 \leq \Gamma$, for $i \in \{1,\dots,N\}$. As a result, for most practical cases, one can make sure that this constraint is not active by choosing γ sufficiently small and Γ sufficiently large. Moreover, if the conditions in Theorem 5.4 are satisfied, the problems (14) and (15) will have a common minimizer. Consequently, the constraint $\gamma \leq \|[Z' C_\Sigma^{1/2}]_i\|_2^2 \leq \Gamma$, for $i \in \{1,\dots,N\}$ will be inactive at the solution of (14). Hence, the minimizer of (10) is the same as the minimizer of (13). □

Theorem 5.4 and Remark 5.5 are a bridge between TLS and EWTLS problems which is helpful since one can solve a standard TLS problem in closed-form instead of solving EWTLS problem using cumbersome numerical methods.

VI. Simulation Results

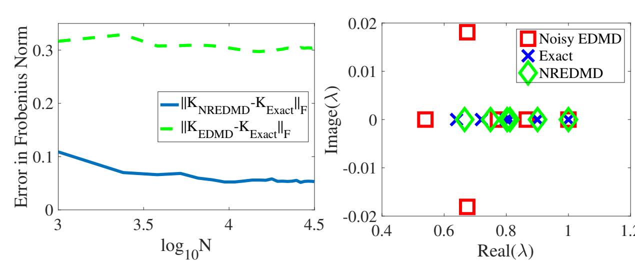
In this section, we illustrate the effectiveness of the NREDMD method using two examples from [17]. In order to focus on the advantages of NREDMD versus EDMD, we use examples for which the Koopman operator can be approximated via a reasonably simple dictionary. All the calculations regarding NREDMD are performed using the standard TLS presented in (13).

Example 6.1 (Linear System with Nonlinear Dictionary): Consider the following linear system

$$\begin{bmatrix} x_1(n+1) \\ x_2(n+1) \end{bmatrix} = \begin{bmatrix} 0.9 & -0.1 \\ 0 & 0.8 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix}. \qquad (16)$$

We use a dictionary comprised of distinct monomials of the form $\prod_{i=1}^{2} y_i$ where $y_i \in \{1, x_1, x_2\}$ for $i \in \{1,2,3\}$. Therefore, we have $n_s = 2$ and $N_d = 6$. Moreover, we confine the initial conditions to $[-2,2] \times [-2,2] \subset \mathbb{R}^2$. The measurement noise is zero-mean normally distributed with covariance matrix $W = 0.05\, I$.

Fig. 1a shows the difference between Koopman representation obtained from applying EDMD on true data and approximated representations obtained from applying EDMD and NREDMD on noisy data. It is important to note that the dictionary used in Example 6.1 spans an invariant Koopman subspace which leads to a linear relationship in the form of $D(Y) = D(X) K_{\text{Exact}}$. Fig. 1a indicates the effectiveness of NREDMD vs EDMD. Moreover, the eigenvalues of representations obtained from different methods using 4×10^4 data points are plotted in Fig. 1b which indicates reasonably accurate reconstruction of the eigenvalues using the NREDMD method. It is important to note that the only source of error that limits the accuracy of NREDMD method is the usage of $\tilde{X}, \tilde{Y}$ instead of X, Y in (9).



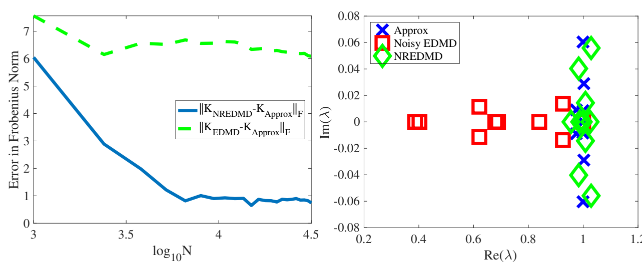
(a) (b)

Fig. 1: Comparison between the effectiveness of EDMD and NREDMD on noisy data acquired from the linear system (16): (a) The error in Frobenius norm between calculated matrices from noisy data and true Koopman representation. (b) Eigenvalues of corresponding matrix representations.

Example 6.2 (Duffing Equation with Nonlinear Dictionary): Consider the Duffing equation stated in (17)

$$\dot{x}_1 = x_2,$$
$$\dot{x}_2 = -0.5x_2 - x_1(x_1^2 - 1). \tag{17}$$

We use a dictionary comprised of distinct monomials of the form $\prod_{i=1}^{3} y_i$ where $y_i \in \{1, x_1, x_2\}$ for $i \in \{1, 2, 3\}$. Consequently, we have $n_s = 2$ and $N_d = 10$. Moreover, we confine the initial conditions to $[0, 2] \times [0, 2] \subset \mathbb{R}^2$. The measurement noise is zero-mean normally distributed with covariance matrix $W = 0.03\, I$. In this example, the subspace spanned by the dictionary is not an invariant Koopman subspace. Hence, one can only find an approximate representation for the Koopman operator in the form of $D(Y) \approx D(X)K_{\text{Approx}}$, i.e., the EDMD method minimizes the error $\|D(Y) - D(X)K_{\text{Approx}}\|_F$ which happens to be nonzero in this example. Consequently, we do not expect to accurately converge to K_{Approx} using NREDMD applied on noisy data. Fig. 2a shows the difference between Koopman



(a) (b)

Fig. 2: Comparison between the effectiveness of EDMD and NREDMD on noisy data acquired from Duffing equation (17): (a) The error in Frobenius norm between calculated matrices from noisy data and true Koopman representation. (b) Eigenvalues of corresponding matrix representations.

representation obtained from applying EDMD on true data and representations obtained from applying EDMD and NREDMD on noisy data. Fig. 2b illustrates the eigenvalues of approximate representation obtained by applying EDMD on true data, and representations produced by employing EDMD and NREDMD on noisy data. Fig. 2 shows nearly accurate performance of NREDMD method.

VII. CONCLUSIONS

We have presented a method based on EWTLS to find a finite-dimensional approximation for the Koopman operator using noisy data. Also, we have proposed an optimization problem based on TLS which is equivalent to the ETWLS based method under some conditions and leads to finding the closed-form solution for ETWLS based method. The effectiveness of the presented methods has been demonstrated using two simulation examples. Future work will characterize the minimal number of data snapshots needed for the two proposed methods to have a common minimizer, explore the extension of the results to handle streaming datasets, and design methods for the efficient selection of dictionaries.

REFERENCES

[1] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," *Proceedings of the National Academy of Sciences*, vol. 17, no. 5, pp. 315–318, 1931.

[2] B. O. Koopman and J. V. Neumann, "Dynamical systems of continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 18, no. 3, pp. 255–263, 1932.

[3] I. Mezić, "Spectral properties of dynamical systems, model reduction and decompositions," *Nonlinear Dynamics*, vol. 41, no. 1-3, pp. 309–325, 2005.

[4] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *Journal of Fluid Mechanics*, vol. 641, pp. 115–127, 2009.

[5] M. Budišić, R. Mohr, and I. Mezić, "Applied Koopmanism," *Chaos*, vol. 22, no. 4, p. 047510, 2012.

[6] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," *Annual Review of Fluid Mechanics*, vol. 45, pp. 357–378, 2013.

[7] A. Mauroy and J. Goncalves, "Linear identification of nonlinear systems: A lifting technique based on the Koopman operator," in *IEEE Conf. on Decision and Control*, Las Vegas, NV, Dec. 2016, pp. 6500–6505.

[8] B. Eisenhower, T. Maile, M. Fischer, and I. Mezić, "Decomposing building system data for model validation and analysis using the Koopman operator," in *Proceedings of SimBuild*, 2010, pp. 434–441.

[9] M. Korda and I. Mezić, "Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control," *Automatica*, vol. 93, pp. 149–160, 2018.

[10] H. Arbabi, M. Korda, and I. Mezic, "A data-driven Koopman model predictive control framework for nonlinear flows," *arXiv preprint arXiv:1804.05291*, 2018.

[11] I. Abraham, G. de la Torre, and T. Murphey, "Model-based control using Koopman operators," in *Proceedings of Robotics: Science and Systems*, Cambridge, Massachusetts, July 2017.

[12] S. Peitz and S. Klus, "Koopman operator-based model reduction for switched-system control of PDEs," *arXiv preprint arXiv:1710.06759*, 2017.

[13] P. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.

[14] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, "On dynamic mode decomposition: theory and applications," *arXiv preprint arXiv:1312.0041*, 2013.

[15] S. T. M. Dawson, M. S. Hemati, M. O. Williams, and C. W. Rowley, "Characterizing and correcting for the effect of sensor noise in the dynamic mode decomposition," *Experiments in Fluids*, vol. 57, no. 3, p. 42, 2016.

[16] M. S. Hemati, C. W. Rowley, E. A. Deem, and L. N. Cattafesta, "De-biasing the dynamic mode decomposition for applied Koopman spectral analysis of noisy datasets," *Theoretical and Computational Fluid Dynamics*, vol. 31, no. 4, pp. 349–368, 2017.

[17] 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, 2015.

[18] M. Korda and I. Mezić, "On convergence of extended dynamic mode decomposition to the Koopman operator," *Journal of Nonlinear Science*, vol. 28, no. 2, pp. 687–710, 2018.

[19] S. Sinha, B. Huang, and U. Vaidya, "Robust approximation of Koopman operator and prediction in random dynamical systems," in *American Control Conference*, Milwaukee, WI, 2018, pp. 5491–5496.

[20] I. Markovsky and S. V. Huffel, "Overview of total least-squares methods," *Signal processing*, vol. 87, no. 10, pp. 2283–2302, 2007.

[21] L. J. Gleser, "Estimation in a multivariate errors in variables regression model: large sample results," *The Annals of Statistics*, pp. 24–44, 1981.

[22] A. Kukush and S. V. Huffel, "Consistency of elementwise-weighted total least squares estimator in a multivariate errors-in-variables model AX=B," *Metrika*, vol. 59, no. 1, pp. 75–97, 2004.

[23] I. Markovsky, M. L. Rastello, A. Premoli, A. Kukush, and S. V. Huffel, "The element-wise weighted total least-squares problem," *Computational Statistics & Data Analysis*, vol. 50, no. 1, pp. 181–209, 2006.

[24] M. O. Williams, C. W. Rowley, and I. G. Kevrekidis, "A kernel-based approach to data-driven Koopman spectral analysis," *arXiv preprint arXiv:1411.2260*, 2014.