Strong convergence for the dynamic mode decomposition based on the total least squares to noisy datasets

Kensuke Aishima

1Faculty of Computer and Information Sciences, Hosei University, 3-7-2 Kajino-cho, Koganei, Tokyo 184-8584, Japan
E-mail aishima@hosei.ac.jp

Received December 26, 2019, Accepted May 01, 2020

Abstract
Dynamic mode decomposition (DMD) is a popular technique for extracting important information of nonlinear dynamical systems. In this paper, we focus on the DMD based on the total least squares (TLS), which is experimentally efficient for noisy datasets for a dynamical system, while the asymptotic analysis is not given. We propose a statistical model of random noise, adapting to the Koopman operator associated with the DMD. Moreover, under reasonable assumptions, we prove strong convergence of random variables, corresponding to the eigenpairs computed by the DMD based on the TLS.

Keywords dynamic mode decomposition, eigenvalue problems, singular value decomposition, total least squares, strong convergence of random variables

Research Activity Group Algorithms for Matrix / Eigenvalue Problems and their Applications

1. Introduction
Mathematical modeling for nonlinear dynamics using data is greatly important in many scientific and engineering research fields such as fluid analysis, data science, image processing, and so forth. The dynamic mode decomposition (DMD) [1] has recently been developed as an efficient technique for the above purpose.

The DMD is a dimensionality reduction algorithm using matrix datasets $X, Y \in \mathbb{R}^{m \times n}$ regarded as the snapshots for the dynamical system. The aim is to find $r(\leq \min(m, n))$ eigenvalues and the corresponding eigenvectors of unknown matrix $A$ with rank$(A) \approx r$ such that $AX \approx Y$. Usually, matrix datasets $X, Y \in \mathbb{R}^{m \times n}$ are contaminated with noise. Under a statistical model of the noise, the total least squares (TLS) method is successfully applied to the DMD, which is experimentally verified in [2, 3], while the asymptotic analysis is not given in any literature as far as the author knows.

With such a background, we present a nice statistical model of the noise, where the strong convergence for the DMD based on the TLS can be straightforwardly proved. Moreover, under reasonable assumptions, we prove the strong convergence of the eigenpairs obtained by the DMD based on the TLS, in the asymptotic regime.

This paper is organized as follows. Section 2 is devoted to a description of the Koopman spectral analysis, as in [4], relevant to the DMD to clarify the issues of the noise model. In Section 3, the DMD based on the TLS [2, 3] and the corresponding convergence theorem [5] are explained. Section 4 is a mathematical discussion of the rank-deficient case [6], excluded in the original papers [2, 3] for the DMD based on the TLS. In Section 5, we propose a new statistical model for the noise and present a key lemma to prove convergence of the DMD based on the TLS. In Section 6, we prove the strong convergence of the eigenpairs computed by the DMD based on the TLS. Finally, concluding remarks are given in Section 7.

2. Koopman spectral analysis and DMD
As a background, we consider a nonlinear function $f$ on a manifold $M$, giving the following dynamical system:

$$z_k \in M, \quad z_{k+1} = f(z_k), \quad k = 0, 1, \ldots$$

Koopman spectral analysis in [4] is to introduce a linear operator $K$ for functions on $M$ to $\mathbb{R}$ as

$$g: M \to \mathbb{R}, \quad (Kg)(x) = g \circ f$$

where $g$ is any function, and $K$ gives the pullback. Let $\psi_i (i = 1, 2, \ldots, m)$ denote the eigenfunctions of $K$ such that $K(\psi_i) = \lambda_i \psi_i (i = 1, 2, \ldots)$. Since our purpose is to perform a statistical analysis in a simple situation we assume $g \in \text{span}(\psi_1, \ldots, \psi_r)$, reflecting typical examples in [4, Sections 2.1 and 2.2]. In physics, $g$ is an observable. In the following, we consider multiple observables $g_i: M \to \mathbb{R}$ ($1 \leq i \leq m$) and introduce a vector-valued function $g = [g_1, \ldots, g_m]^{\top}$. Analogously to the above case of the scalar-valued observable, suppose that $g_i \in \text{span}(\psi_1, \ldots, \psi_r)$ ($1 \leq i \leq m$) with $r \leq m$. Then $g(z_k) = \sum_{i=1}^{r} \psi_i(z_k) c_i$ with some $m$ dimensional vectors $c_i$ ($i = 1, \ldots, r$). Noting $g(z_{k+1}) = g \circ f(z_k)$, we have

$$g(z_{k+1}) = \sum_{i=1}^{r} \psi_i \circ f(z_k) c_i = \sum_{i=1}^{r} \lambda_i \psi_i(z_k) c_i.$$

By induction, we obtain

$$g(z_k) = \sum_{i=1}^{r} \lambda_i^k \psi_i(z_0) c_i \in \mathbb{R}^m.$$
Algorithm 1 Total least squares DMD [3]

Require: Matrices $X, Y \in \mathbb{R}^{m \times n}$, rank $r(\leq \min(m, n))$
1: $Z = [X^\top, Y^\top]^\top \in \mathbb{R}^{2m \times n}$
2: Compute the SVD of $Z$, and let $Q$ be an $n \times r$ matrix comprising the first $r$ right singular vectors
3: $\hat{X} = XQ, \hat{Y} = YQ$
4: Compute the SVD: $\hat{X} = U\Sigma V^\top$, where $U \in \mathbb{R}^{m \times r}$, $\Sigma \in \mathbb{R}^{r \times r}$, $V \in \mathbb{R}^{n \times r}$
5: $\hat{A}_{\text{tls}} = U^\top \hat{Y} V \Sigma^{-1}$
6: Compute the eigenpairs: $\hat{A}_{\text{tls}}w_{\text{tls}} = \lambda_{\text{tls}}w_{\text{tls}}$, and let $w_{\text{tls}} = Uw_{\text{tls}}$

Thus, there exists an $m \times m$ matrix $A$ such that
$$\text{rank}(A) = r, \quad g(z_k) = Ag(z_k) \quad (k = 0, 1, \ldots).$$

For some $k$, we observe $g(z_k), g(z_{k+1})$, for $\ell = 1, 2, \ldots, n$, let
$$x_\ell = g(z_k), \quad y_\ell = g(z_{k+1}),$$
$$\bar{X} = [x_1, \ldots, x_n], \quad \bar{Y} = [y_1, \ldots, y_n].$$

In general, $A \in \mathbb{R}^{m \times m}$ satisfying $AX = \bar{Y}$ has eigenvalues $\lambda_i (i = 1, \ldots, r)$, showing important features of the dynamics $z_{k+1} = f(z_k)$ $(k = 0, 1, \ldots)$. Basically, the aim of the DMD is to compute the above eigenpairs without explicitly constructing the large scale matrix $A$.

Usually, the observed snapshot datasets are contaminated with noise. Thus, we consider a simple noise model as follows. For some $k$, we observe perturbed $g(z_k), g(z_{k+1})$ with random noise. For $\ell = 1, 2, \ldots, n$, let
$$\bar{x}_\ell = g(z_k) + \xi_{k,\ell}, \quad \bar{y}_\ell = g(z_{k+1}) + \xi_{k+1,\ell},$$
$$\bar{X} = [\bar{x}_1, \ldots, \bar{x}_n], \quad \bar{Y} = [\bar{y}_1, \ldots, \bar{y}_n],$$
where $\xi_{j,\ell}$ are random variables. Hence, we have $AX \approx \bar{Y}$ for the above exact $A$ associated with the matrix datasets without noise. For computing the eigenpairs from the above noisy datasets, the total least squares DMD is efficient as described in numerical results in [2,3].

3. Total least squares DMD

For given $X, Y \in \mathbb{R}^{m \times n}$, we consider the total least squares (TLS) problem, i.e.,
$$\min_{A_{\text{tls}}: \Delta X, \Delta Y} \|\Delta X\|^2_F + \|\Delta Y\|^2_F$$
subject to $Y + \Delta Y = A_{\text{tls}}(X + \Delta X)$.

For computing $A_{\text{tls}}$, there is a sophisticated method using the singular value decomposition (SVD). In [2,3], such a method is successfully applied to the DMD, where $m \leq n$ is assumed. The method in [3] is presented in Algorithm 1. In this algorithm, let
$$\hat{X} = V\Sigma^{-1}U^\top, \quad \hat{A}_{\text{tls}} = \hat{Y}\hat{X}^\top.$$  (3)

Note that, in the case of $r = m \leq n$, the matrix $\hat{A}_{\text{tls}}$ is the solution of the TLS problem as in [2,3]. In general, for any $r \leq \min(m, n)$, it is easy to see that $\hat{A}_{\text{tls}}$ in (3) satisfies $\hat{A}_{\text{tls}}\hat{X} = \hat{Y}$ for $m \times r$ matrices $\hat{X}, \hat{Y}$ in line 3 of Algorithm 1, where we assume $\text{rank}(\hat{X}) = r$. More importantly,
$$\hat{A}_{\text{tls}} = U^\top \hat{A}_{\text{tls}}U.$$  

This implies that Algorithm 1 computes eigenpairs of $\hat{A}_{\text{tls}}$ by the Rayleigh-Ritz procedure using the $r$ dimensional subspace. All of $(\lambda_{\text{tls}}, w_{\text{tls}})$ are the exact eigenpairs of $\hat{A}_{\text{tls}}$. If $r = m$, $\hat{A}_{\text{tls}}$ and $\hat{A}_{\text{tls}}$ are similar.

From a viewpoint of asymptotic analysis, Algorithm 1 is designed under the next statistical model of the random noise.

Problem 1 Assume that unknown matrices $\bar{X}, \bar{Y} \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{m \times n}$ satisfy $AX = \bar{Y}$. Observed matrices are $X, Y \in \mathbb{R}^{m \times n}$ such that $X = \bar{X} + E, Y = \bar{Y} + F$, where $E, F \in \mathbb{R}^{m \times n}$ are random matrices. Find $m$ eigenvalues $\lambda_i (i = 1, \ldots, m)$ and the corresponding eigenvectors $w_i (i = 1, \ldots, m)$ of $A \in \mathbb{R}^{m \times m}$

We consider the next assumptions in the following asymptotic analysis.

Assumption 2 The columns of $[E^\top, F^\top]^\top$ are i.i.d. with common mean vector $0$ and common covariance matrix $\sigma^2I$. Moreover, $\lim_{n \to \infty} n^{-1}X^\top X$ exists.

Assumption 3 Under Assumption 2, the matrix $\lim_{n \to \infty} n^{-1}X^\top X$ is positive definite.

Gleser [5] proves strong convergence of $\hat{A}_{\text{tls}}$ as stated in the next lemma.

Lemma 4 (Strong convergence [5, Lemma 3.3]) Suppose that Algorithm 1 is applied to Problem 1 with $r = m$ under Assumptions 2 and 3. Then
$$\lim_{n \to \infty} \hat{A}_{\text{tls}} = \lim_{n \to \infty} \hat{Y}\bar{V}\Sigma^{-1}\bar{U}^\top = A$$
with probability one.

Although this lemma states the convergence of $\hat{A}_{\text{tls}}$ with $r = m$, the convergence property for $r < m$ is not considered in [3] that presents Algorithm 1. In addition, it is important to prove the convergence of eigenpairs of $\hat{A}_{\text{tls}}$ with mathematical rigor, while it is not discussed. In the following, we solve the above two issues. Firstly, we consider the rank-deficient case in the next section.

4. Rank-deficient case [6]

In [6], for the rank-deficient case of the TLS problem, the next variant is considered.
$$\min_{A_{\text{tls}}: \Delta X, \Delta Y} \|\Delta X\|^2_F + \|\Delta Y\|^2_F$$
subject to $Y + \Delta Y = A_{\text{tls}}(X + \Delta X)$,
$$Z = [X^\top, Y^\top]^\top, \quad \Delta Z = [\Delta X^\top, \Delta Y^\top]^\top,$$
$$\text{rank}(Z + \Delta Z) = r(< m).$$

The solution of this problem for sufficiently large $n$ can be used for solving the following problem.

Problem 5 Assume that unknown matrices $\bar{X}, \bar{Y} \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{m \times n}$ satisfy
$$AX = \bar{Y}, \quad Z = [X^\top, Y^\top]^\top, \quad \text{rank}(Z) = r(< m).$$
Observed matrices are $X, Y \in \mathbb{R}^{m \times n}$ such that $X = X + E, Y = Y + F$, where $E, F \in \mathbb{R}^{m \times n}$ are random matrices. Find $r$ eigenvalues $\lambda_i (i = 1, \ldots, r)$ and the corresponding eigenvectors $u_i (i = 1, \ldots, r)$ of $A \in \mathbb{R}^{m \times m}$.

Since $\tilde{A}_{1h}$ in (3) is the minimum norm solution for $A_{1h}$ with respect to the Frobenius norm, the strong convergence holds as in the next lemma.

Lemma 6 (Strong convergence [6, Lemma 4.6]) Let $\tilde{A}$ denote the minimum norm solution for $A$ of Problem 5 with respect to the Frobenius norm. Suppose that Algorithm 1 is applied to Problem 5 with $r(< m)$ under Assumption 2, where $r$ is the rank of $\lim_{n \to \infty} n^{-1} X X^T$. Then
\[
\lim_{n \to \infty} \tilde{A}_{1h} = \lim_{n \to \infty} \tilde{Y} V \Sigma^{-1} U^T = \tilde{A}
\]
with probability one.

From this lemma, the eigenpairs obtained by Algorithm 1 are expected to converge to the eigenpairs of the minimum norm solution $\tilde{A}$. However, we do not see whether the obtained eigenpairs are the required eigenpairs for the DMD associated with the Koopman spectral analysis. We solve this issue in the next section.

5. Proposed model and theory

In this section, we propose a noise model adapting to the Koopman spectral analysis in Section 2. Let $\mathcal{R}$ denote the range of the argument matrix. In addition, noting the features of $g$ and the definitions of $\tilde{X}, \tilde{Y}$, we have $\mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y})$ because, in general, they are the $r$ dimensional subspace spanned by the vectors $c_1, \ldots, c_r$. Therefore, we propose the next noise model.

Problem 7 Assume that unknown matrices $\tilde{X}, \tilde{Y} \in \mathbb{R}^{m \times n}$ and $A \in \mathbb{R}^{m \times m}$ satisfy
\[
A \tilde{X} = \tilde{Y}, \quad Z = [X^T, Y^T]^T, \quad \text{rank}(\tilde{Z}) = r(< m), \quad \mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y}).
\]
Observed matrices are $X, Y \in \mathbb{R}^{m \times n}$ such that $X = X + E, Y = Y + F$, where $E, F \in \mathbb{R}^{m \times n}$ are random matrices. Find $r$ eigenvalues $\lambda_i (i = 1, \ldots, r)$ and the corresponding eigenvectors $u_i (i = 1, \ldots, r)$ of $A \in \mathbb{R}^{m \times m}$.

From Lemma 6, we have the next corollary.

Corollary 8 Let $\tilde{A}$ denote the minimum norm solution $A$ of Problem 7 with respect to the Frobenius norm. Suppose that Algorithm 1 is applied to Problem 7 with $r(< m)$ under Assumption 2, where $r$ is the rank of $\lim_{n \to \infty} n^{-1} X X^T$. Then
\[
\lim_{n \to \infty} \tilde{A}_{1h} = \lim_{n \to \infty} \tilde{Y} V \Sigma^{-1} U^T = \tilde{A}
\]
with probability one.

Thus, the convergence is theoretically guaranteed. The question is whether the minimum norm solution $\tilde{A}$ is appropriate for the DMD. The key to answer the question is the next lemma.

Lemma 9 Any matrix $A$ of Problem 7 has $r$ eigenvectors (or generalized eigenvectors) in the $r$ dimensional subspace $\mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y})$. The $r$ eigenvectors (or generalized eigenvectors) are uniquely determined for any $A$. Similarly, the $r$ eigenvalues of $A$ associated with the above $r$ eigenvectors (or generalized eigenvectors) are uniquely determined.

Proof Noting
\[
r = \text{rank}(\tilde{Z}) = \text{rank}([X^T, Y^T]^T) = \text{rank}(\tilde{X})
\]
from $A \tilde{X} = \tilde{Y}$, we have $\text{rank}(\tilde{Y}) = r$ in view of $\mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y})$. Let $U \in \mathbb{R}^{m \times n}$ denote a matrix comprising basis vectors of the subspace $\mathcal{R}(\tilde{X})$. From $\mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y})$ and $AX = \tilde{Y}$, we see
\[
AU_X = U_X B,
\]
where $B \in \mathbb{R}^{r \times r}$ is nonsingular due to $\mathcal{R}(\tilde{Y}) = \mathcal{R}(AU_X)$. Note that $B$ is unique for the above fixed $U_X$, even though $A$ is not unique, proved as follows. For given $X, \tilde{Y}$, since an $n \times r$ matrix $V_X$ with $\text{rank}(V_X) = r$ satisfying $\tilde{X} = U_X V_X^T$ uniquely exists, we have
\[
\tilde{Y} = U_X B V_X^T,
\]
implying that $B$ is unique due to $\text{rank}(U_X) = \text{rank}(V_X) = r$ for given $X, \tilde{Y}$.

In the above discussion, although $U_X \in \mathbb{R}^{m \times r}$ is not unique, any $n \times r$ matrix associated with $\mathcal{R}(\tilde{X})$ is given by $U_X C$ with a fixed $U_X$ and a nonsingular $C \in \mathbb{R}^{r \times r}$. Therefore, the above equation (4) is established for any $n \times r$ matrix associated with $\mathcal{R}(\tilde{X})$.

Hence, we consider the Jordan decomposition
\[
B = U_B J^{-1} U_B^T,
\]
where $J$ is a Jordan matrix of $B$. Thus, we see
\[
AU_X U_B = U_X U_B J,
\]
showing that the columns of $U_X U_B$ are eigenvectors or generalized eigenvectors of $A$ because $J$ is a block diagonal matrix. In addition, every diagonal block of $J$ is a Jordan block of $A$ or the leading principal submatrix of its Jordan block. Thus, the diagonal elements of $J$ are the eigenvalues of $A$ associated with the above $r$ eigenvectors (or generalized eigenvectors).

(QED)

In the above proof, one might think that $J$ comprises the Jordan block of $A$. However, there is a counter example as follows. Suppose that
\[
A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \mathcal{R}(\tilde{X}) = \mathcal{R}(\tilde{Y}) = \text{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}.
\]
Then $J = 1$ in the proof. Thus, $J$ is not the Jordan block of $A$ but the submatrix of the Jordan block.

Note that, if $\mathcal{R}(\tilde{X}) \neq \mathcal{R}(\tilde{Y})$, the uniqueness in Lemma 9 does not hold. For example, suppose that
\[
\mathcal{R}(\tilde{X}) = \text{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}, \quad \mathcal{R}(\tilde{Y}) = \text{span}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}.
\]
Then we see
\[
A = \begin{pmatrix} 0 & \alpha \\ \beta & \gamma \end{pmatrix} \quad (\beta \neq 0).
\]
Thus, the eigenpairs of $A$ are not unique.

We know that Algorithm 1 computes the minimum norm solution $\tilde{A}$ as stated in Corollary 8. Thus, from
Lemma 9, the eigenpairs computed by Algorithm 1 are expected to converge to the target eigenpairs. Our goal is to prove this convergence.

6. Convergence analysis of eigenpairs

In this section, we prove the strong convergence of the eigenpairs computed by Algorithm 1. Since we consider the Jordan block due to multiple eigenvalues, let \( \tilde{w}_{\text{dmd}} \) in Algorithm 1 include the generalized eigenvector of \( \bar{A}_{\text{dmd}} \).

First of all, note that, in general, the strong convergence of random variables is defined as follows:

**Definition 10** (Strong convergence) The sequence of random variables \( X^{(n)}(\omega) \) converges to \( X \) with probability one if and only if

\[
P \left( \left\{ \omega \in \Omega \mid \lim_{n \to \infty} X^{(n)}(\omega) = X(\omega) \right\} \right) = 1,
\]

where \( P \) is a probability measure on a sample space \( \Omega \).

Thus, the strong convergence of eigenpairs can be proved by the perturbation theory as shown in [7, Section 7.2]. We explain the details below. It is easy to see that, since \( \bar{A}_{\text{dmd}} \) is convergent with probability one, the eigenvalues are also convergent with probability one due to the discontinuities. In addition, for any simple eigenvalue, the accumulation points of the corresponding eigenvector are the eigenvectors of the convergent matrix, with probability one. For multiple eigenvalues of \( A \), the eigenvectors (or generalized eigenvectors) of a perturbed matrix go toward the corresponding invariant subspace, result in the next theorem.

**Theorem 11** Suppose that Algorithm 1 is applied to Problem 1 with \( r = m \) under Assumptions 2 and 3, where \( \lambda_{\text{dmd}}^{(n)} \), \( w_{\text{dmd}}^{(n)} \) denote \( \lambda_{\text{dmd}}, w_{\text{dmd}} \) in Algorithm 1 for \( n \). Then each eigenvalue \( \lambda_{\text{dmd}}^{(n)} \) is convergent to an eigenvalue of \( A \) with probability one, and the accumulation points of \( w_{\text{dmd}}^{(n)} \) are in the invariant subspace corresponding to \( \lambda_{\text{dmd}}^{(n)} \) with probability one.

Although the above theorem is easily proved from the perturbation theory, it is important to note the following fact from the mathematical viewpoints.

The subspace spanned by all the \( m \) vectors \( w_{\text{dmd}}^{(n)} \) does not necessarily converge to the \( m \) dimensional space. Clearly, if \( A \) has the \( m \) eigenvectors, then all the \( m \) vectors \( w_{\text{dmd}}^{(n)} \) are individually convergent to the eigenvectors of \( A \), result in the \( m \) dimensional space. However, the Jordan block that is not diagonalizable requires some care. For example, let

\[
\epsilon \neq 0, \quad A(\epsilon) = \begin{pmatrix} 0 & 1 \\ 1 & 1 + \epsilon \end{pmatrix}.
\]

We see that the eigenvalues are \( 1, 1 + \epsilon \) and the corresponding eigenvectors are \( (1, 0)^T, (1, \epsilon)^T \), respectively. Thus, the space spanned by the above two vectors is reduced to one dimension as \( \epsilon \to 0 \), corresponding to the Jordan matrix limit as \( \epsilon \to 0 \), \( A(\epsilon) \) that is not diagonalizable.

Similarly to the above theorem, we obtain the next theorem corresponding to Problem 5 and Lemma 6.

**Theorem 12** Let \( A \) denote the minimum norm solution \( A \) of Problem 5 with respect to the Frobenius norm. Suppose that Algorithm 1 is applied to Problem 5 with \( r < m \) under Assumption 2, where \( r \) is the rank of \( \lim_{n \to \infty} n^{-1}X^\top X \) and \( \lambda_{\text{dmd}}^{(n)}, w_{\text{dmd}}^{(n)} \) denote \( \lambda_{\text{dmd}}, w_{\text{dmd}} \) in Algorithm 1 for \( n \). Then each eigenvalue \( \lambda_{\text{dmd}}^{(n)} \) is convergent to an eigenvalue of \( A \) with probability one, and the accumulation points of \( w_{\text{dmd}}^{(n)} \) are in the invariant subspace corresponding to \( \lambda_{\text{dmd}}^{(n)} \) with probability one.

Obviously, the same convergence is proved to Problem 7 due to Corollary 8. It is worth noting that, from Lemma 9, we obtain the next theorem that states the convergence of the eigenpairs of any \( A \) in Problem 7.

**Theorem 13** Suppose that Algorithm 1 is applied to Problem 7 with \( r < m \) under Assumption 2, where \( r \) is the rank of \( \lim_{n \to \infty} n^{-1}X^\top X \) and \( \lambda_{\text{dmd}}^{(n)}, w_{\text{dmd}}^{(n)} \) denote \( \lambda_{\text{dmd}}, w_{\text{dmd}} \) in Algorithm 1 for \( n \). Then each eigenvalue \( \lambda_{\text{dmd}}^{(n)} \) is convergent to an eigenvalue of \( A \) with probability one, and the accumulation points of \( w_{\text{dmd}}^{(n)} \) are in the invariant subspace corresponding to \( \lambda_{\text{dmd}}^{(n)} \) with probability one.

Note that the subspace spanned by all the \( r \) vectors \( w_{\text{dmd}}^{(n)} \) does not necessarily converge to the \( r \) dimensional space in the same manner as the other theorems.

7. Concluding remarks

In this paper, we solved two issues of the total least squares DMD from the theoretical viewpoints. Firstly, we proposed a nice statistical model of noise to straightforwardly prove strong convergence for the total least squares DMD. Secondly, we proved the strong convergence of the eigenpairs for the total least squares DMD. Our theory covers multiple eigenvalues. In particular, we proved that the accumulation points of the corresponding eigenvectors are in the invariant subspace of the multiple eigenvalues, with probability one.

Acknowledgments

The author thanks the reviewer and Dr. Takeru Matsuda for their valuable comments. This study is supported by KAKENHI Nos. 17K14143 and 17H02826.

References

[1] P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech., 656 (2010), 5–28.
[2] 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, Exp. Fluids, 57 (2016), 1–19.
[3] 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, Theor. Comput. Fluid Dyn., 31 (2017), 349–368.
[4] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter and D. S. Henningson, Spectral analysis of nonlinear flows, J. Fluid Mech., 641 (2009), 115–127.
[5] L. J. Gleser, Estimation in a multivariate “errors in variables” regression model: large sample results, Ann. Statist., 9 (1981), 24–44.
[6] S. Park and D. P. O’Leary, Implicitly-weighted total least squares, Linear Algebra Appl., 435 (2011), 560–577.
[7] G. H. Golub and C. F. Van Loan, Matrix Computations, 4th ed., Johns Hopkins University Press, Baltimore, MD, 2013.