

Global Convergence of Oja's Component Flow for General Square Matrices and Its Applications

Daiki Tsuzuki and Kentaro Ohki, *Member, IEEE*

Abstract—In this study, the global convergence properties of the Oja flow, a continuous-time algorithm for principal component extraction, was established for general square matrices. The Oja flow is a matrix differential equation on the Stiefel manifold designed to extract a dominant subspace. Although its analysis has traditionally been restricted to symmetric positive-definite matrices, where it acts as a gradient flow, recent applications have extended its use to general matrices. In this non-symmetric case, the flow extracts the invariant subspace corresponding to the eigenvalues with the largest real parts. However, prior convergence results have been purely local, leaving the global behavior as an open problem. The findings of this study fill this gap by providing a comprehensive global convergence analysis, establishing that the flow converges exponentially for almost all initial conditions. We also propose a modification to the algorithm that enhances its numerical stability. As an application of this theory, we developed novel methods for model reduction of linear dynamical systems and the synthesis of low-rank stabilizing controllers. The study advances the theoretical understanding of the Oja flow and demonstrates its potential as a reliable and versatile tool for analyzing and controlling complex linear systems.

Index Terms—Principal component analysis, subspace tracking, model reduction

I. INTRODUCTION

PRINCIPAL component analysis (PCA) is a foundational technique for dimensionality reduction and feature extraction, with broad applications across technology and science [1]–[3]. For large-scale and streaming datasets, online algorithms for PCA are essential. A prominent class of such algorithms is principal and minor component flows, which are recursive methods that have gained significant attention in statistics, machine learning [4]–[7], signal processing [8], [9], and control theory [10]–[18]. Although these algorithms are valued for their computational efficiency, the theoretical conditions guaranteeing their convergence, particularly for general non-symmetric matrices, remain incompletely understood.

This work was supported by JSPS KAKENHI Grant Numbers 23H01432 and 21K12097.

Daiki Tsuzuki is with the Graduate School of Informatics, Kyoto University, Kyoto, Japan (e-mail: tsuzuki.daiki.57a@bode.amp.i.kyoto-u.ac.jp)

Kentaro Ohki is with the Graduate School of Informatics, Kyoto University, Kyoto, Japan (e-mail: ohki@i.kyoto-u.ac.jp).

A. Background of the Oja Flow

A canonical example of a principal component flow is the Oja flow [19]–[21]:

$$\varepsilon \frac{d}{dt} U(t) = (I_n - U(t)U(t)^\top)AU(t), \quad U(0) \in \text{St}(r, n), \quad (1)$$

where $r \leq n$, $A \in \mathbb{R}^{n \times n}$, $\varepsilon \in (0, 1]$ is a rate-controlling parameter, I_n is the $n \times n$ identity matrix, and $\text{St}(r, n) := \{X \in \mathbb{R}^{n \times r} \mid X^\top X = I_r\}$ is the Stiefel manifold. The corresponding minor component flow, which seeks the least dominant subspace, is obtained by replacing A with $-A$.

Although the Oja flow is widely used in machine learning to extract the dominant singular subspaces of symmetric positive-definite matrices, a case supported by strong theoretical guarantees [11], [22]–[24], its analysis and application for general square matrices have been limited [25].

This paper provides a comprehensive convergence analysis of the Oja flow for general matrices and explores its applications in control theory. Our previous work [17] showed that the Oja flow can extract the invariant subspace corresponding to the r eigenvalues with the largest real parts. However, that analysis was restricted to local convergence, and the estimate of the domain of attraction was conservative. This work extends these results to establish global convergence properties.

B. Related Work

Prior studies on the Oja flow have primarily focused on symmetric positive-definite matrices, yielding several key results:

- 1) The existence and uniqueness of the solution to the (1) were established in [26].
- 2) For any full-rank initial matrix $U(0) \in \mathbb{R}^{n \times r}$, the solution $U(t)$ converges to the Stiefel manifold $\text{St}(r, n)$ [26].
- 3) For almost all initial conditions, the solution converges to the subspace spanned by the eigenvectors corresponding to the r largest eigenvalues, provided the eigenvalues are distinct [22].

These results extend to any symmetric matrix by a spectral shift, $A \rightarrow A + aI_n$, $a > 0$, which renders the matrix positive-definite without altering the eigenvectors. Extensions have broadly followed two directions: modifying the Oja flow itself or analyzing the original flow for a more general class of matrices. This paper pursues the latter direction, following preliminary work in [10], [14].

Concurrently with our work, Kuo et al. [25] have also established the existence and uniqueness of the solution for general square matrices and analyzed its exponential convergence on the Stiefel manifold under a certain spectral gap condition. Although this significant work addresses convergence on the Stiefel manifold, a rigorous analysis of the global domain of attraction and the solution behavior for initial conditions outside the manifold remains an open area. The aim of our study also included addressing these specific gaps.

C. Contributions

The main contributions of this paper are as follows:

- 1) A comprehensive convergence analysis of the Oja flow for general real square matrices (Theorems 9 and 12). We prove exponential convergence to the dominant invariant subspace under a mild eigenvalue separation condition and propose a modification that ensures numerical stability by guaranteeing convergence to the Stiefel manifold.
- 2) Estimation of the domain of attraction (Theorem 17). We show that the domain of attraction encompasses almost the entire manifold, indicating that the flow converges for almost all initial values.
- 3) Applications to control theory. We developed a novel framework for model reduction and low-rank controller synthesis. We show that the proposed reduction method preserves key system properties (observability and controllability) and can be used to design stabilizing controllers for large-scale systems with low-dimensional unstable manifolds.

Although we focus on real matrices, our results extend directly to complex matrices.

D. Organization and Notation

This paper is organized as follows. Section II reviews existing results. Section III presents our main theoretical contributions on the convergence of the Oja flow, including its numerical stability and domain of attraction. Section IV demonstrates applications to model reduction and controller synthesis. Key theoretical results are followed by numerical examples for validation.

Notation: The sets of real and complex numbers are $\mathbb{R}$ and $\mathbb{C}$. The set of $n \times m$ real matrices is $\mathbb{R}^{n \times m}$. I_n is the $n \times n$ identity matrix, and $0_{n,m}$ is the $n \times m$ zero matrix. For simplicity, we denote $0_n = 0_{n,n}$; if the dimension is trivial, 0 is also used. $A^\top$ and $A^\dagger$ denote the transpose and Hermitian conjugate of a matrix A . $\|x\|$ is the Euclidean norm for a vector x and $\|A\|_{\text{ind}}$ is its induced norm for a matrix A . For a symmetric matrix A , $A > 0$ ($A \geq 0$) indicates it is positive-definite (semidefinite). $A^{1/2}$ denotes the unique positive-semidefinite square root. The eigenvalues of a square matrix $A \in \mathbb{C}^{n \times n}$ are ordered such that $\text{Re}(\lambda_1(A)) \geq \dots \geq \text{Re}(\lambda_n(A))$. The corresponding (generalized) eigenvector is $\psi_i(A)$, and the matrix of eigenvectors is $\Psi(A) := [\psi_1(A), \dots, \psi_n(A)]$. The Stiefel manifold is $\text{St}(r, n) := \{X \in \mathbb{R}^{n \times r} \mid X^\top X = I_r\}$.

II. SUMMARY OF EXISTING RESULTS

Existing theoretical work on the key properties of the Oja flow (1) can be broadly divided according to whether the initial value $U(0)$ lies on the Stiefel manifold $\text{St}(r, n)$.

A. Convergence Analysis in Euclidean Space

The first category of results concerns the solution behavior of the Oja flow (1) in Euclidean space. It is well established that if the initial condition $U(0)$ is on the Stiefel manifold, i.e., $U(0) \in \text{St}(r, n)$, then the solution $U(t)$ remains on $\text{St}(r, n)$ for all $t \geq 0$ and any matrix $A \in \mathbb{R}^{n \times n}$ [17, §II.B]. In practice, however, numerical errors can cause $U(t)$ to deviate from the manifold. If $\text{St}(r, n)$ were a stable invariant set, a simple numerical integrator such as the forward Euler scheme would keep the solution in a neighborhood of the manifold. However, the stability conditions for general matrices have not been fully investigated. For the specific case of symmetric positive-semidefinite matrices, the following result was shown in [26]:

Proposition 1 ([26, Prop. 3.1]): Let $U(0) \in \mathbb{R}^{n \times r}$ be of full rank. Then, for any symmetric positive-definite matrix $A \in \mathbb{R}^{n \times n}$, the solution $U(t)$ of (1) converges exponentially to $\text{St}(r, n)$ as $t \rightarrow \infty$.

This result implies that for positive-definite matrices A , $\text{St}(r, n)$ is a stable invariant set for the Oja flow. In contrast to Proposition 1, if A is not positive-definite, the Oja flow may fail to converge to $\text{St}(r, n)$ for an initial condition $U(0) \notin \text{St}(r, n)$. The following example illustrates this idea.

Example 2: Consider a system with

$$A = \text{diag}(1, -1), \quad U(0) = \begin{bmatrix} 0 & a \end{bmatrix}^\top, \quad a > 1.$$

The initial derivative for the Oja flow (1) is

$$\left. \frac{d}{dt} U(t) \right|_{t=0} = \frac{1}{\varepsilon} \begin{bmatrix} 1 & 0 \\ 0 & a^2 - 1 \end{bmatrix} U(0) = \frac{1}{\varepsilon} \begin{bmatrix} 0 \\ a(a^2 - 1) \end{bmatrix},$$

which indicates that the second component of the solution $U(t)$ monotonically increases from its initial value. Hence, $U(t)$ does not converge to $\text{St}(r, n)$.

Chen et al. [22] extended Proposition 1 to positive-semidefinite matrices and rank-deficient initial conditions, showing that the Oja flow converges to a point in $\mathbb{R}^{n \times r}$ that depends on $U(0)$. Hasan [10, Variation 4] claimed that convergence to $\text{St}(r, n)$ could be ensured for any square matrix A and any non-zero initial value $U(0)$ by replacing A with $A + aI_n$ for a sufficiently large constant $a > 0$. This modification does not alter the vector field of the Oja flow (1) when restricted to $U(t) \in \text{St}(r, n)$. However, a non-zero initial value is not a sufficient condition for convergence. For instance, if the columns of $U(0) \in \mathbb{R}^{n \times 2}$ are identical, i.e., $U(0) = x_0 \begin{bmatrix} 1 & 1 \end{bmatrix}$, then the solution retains this structure, $U(t) = x(t) \begin{bmatrix} 1 & 1 \end{bmatrix}$, where $x(t)$ evolves according to

$$\frac{d}{dt} x(t) = (I_n - 2x(t)x(t)^\top)Ax(t), \quad x(0) = x_0.$$

Therefore, $U(t)$ preserves its initial rank and cannot converge to $\text{St}(r, n)$. The strategy of shifting the matrix spectrum is compelling, but further analysis is needed because $\text{St}(r, n)$

is generally not a stable invariant set for non-positive-definite matrices. As demonstrated in Example 2 and discussed in [23], if A is negative-definite, solutions initiating outside the Stiefel manifold may diverge. The Stiefel manifold is a generalization of a sphere, and this potential instability implies that numerical implementations of the Oja flow require periodic normalization to mitigate the accumulation of numerical errors. To address this limitation, Chen et al. [23], [27] proposed an alternative algorithm for principal and minor component extraction that does not require normalization. In this study, however, we focused on the Oja flow (1) because of its lower computational complexity compared to the algorithm in [27].

The stability properties of (1) in $\mathbb{R}^{n \times r}$, particularly for general non-symmetric matrices, remain incompletely understood. The aim of this study was to fill this gap by providing a rigorous analysis of the invariance and stability of $\text{St}(r, n)$ for any $A \in \mathbb{R}^{n \times n}$, as described in Section III-A.

B. Convergence Analysis on the Stiefel Manifold

As a principal component extraction algorithm, the Oja flow is expected to converge to the subspace spanned by the eigenvectors corresponding to the dominant eigenvalues. This property is known to hold for positive-definite matrices.

Proposition 3 ([26]):

- 1) ([26, Theorem 5.1]) If A is symmetric positive-definite with $\lambda_r > \lambda_{r+1}$, then for any initial condition $U(0)$ satisfying $\det(U(0)^\top \Psi_r) \neq 0$, the solution of the Oja flow converges to the set

$$\mathcal{U}'_r := \left\{ \Psi(A) \begin{bmatrix} K_r \\ 0_{n-r, r} \end{bmatrix} \in \text{St}(r, n) \mid K_r \in \mathbb{R}^{r \times r} \right\}, \quad (2)$$

where $\Psi(A)$ is the matrix of eigenvectors of A , and $\Psi_r := [\psi_1(A), \dots, \psi_r(A)]$.

- 2) ([26, Corollary 5.1]) The initial condition $\det(U(0)^\top \Psi_r) \neq 0$ holds for almost all $U(0) \in \text{St}(r, n)$.

Notably, these results also apply for symmetric matrices through the use of a spectral shift. Recently, these results were extended to general matrices $A \in \mathbb{R}^{n \times n}$.

Proposition 4 ([17], [18], [25]): Assume that the eigenvalues of A are ordered such that $\text{Re}(\lambda_r) > \text{Re}(\lambda_{r+1})$. Then, the following properties hold:

- 1) ([17, Prop. 2]) The equilibrium sets of (1) are given by

$$\mathcal{U}_{r,P} := \left\{ \Psi(A)P \begin{bmatrix} K_r \\ 0_{n-r, r} \end{bmatrix} \in \text{St}(r, n) \mid K_r \in \mathbb{C}^{r \times r} \right\}, \quad (3)$$

where $P \in \mathbb{R}^{n \times n}$ is a permutation matrix such that $P^\top \Lambda P$ is a block-diagonal with blocks of size $r \times r$ and $(n-r) \times (n-r)$. For simplicity, we denote $\mathcal{U}_r = \mathcal{U}_{r, I_n}$. Note that for any $\bar{U} \in \mathcal{U}_{r,P}$, the matrix $\bar{U}W$ is also in $\mathcal{U}_{r,P}$ for any orthogonal matrix $W \in \mathbb{R}^{r \times r}$.

- 2) ([17, Prop. 3]) For any $\bar{U} \in \mathcal{U}_{r,P}$, the eigenvalues of $\bar{U}^\top A \bar{U}$ are a permutation of the first r eigenvalues of A , i.e., $\{\lambda_i(\bar{U}^\top A \bar{U})\}_{i=1}^r = \{\lambda_{\mathcal{I}_P(i)}(A)\}_{i=1}^r$, where $\mathcal{I}_P$

is the permutation associated with P . In particular, for $\bar{U} \in \mathcal{U}_r$, we have $\lambda_i(\bar{U}^\top A \bar{U}) = \lambda_i(A)$ for $i = 1, \dots, r$.

- 3) ([18, Lemma 1]) For any $\bar{U} \in \mathcal{U}_{r,P}$ and any integer $p \geq 0$, the subspace spanned by the columns of $\bar{U}$ is an invariant subspace of A , satisfying $A^p \bar{U} = \bar{U}(\bar{U}^\top A \bar{U})^p$. This result implies $e^{A\bar{U}} = \bar{U}e^{\bar{U}^\top A \bar{U}}$.
- 4) ([17, Thm. 1]) The set $\mathcal{U}_r$ is the unique asymptotically stable equilibrium set, which means any trajectory starting sufficiently close to $\mathcal{U}_r$ remains in its neighborhood and converges to $\mathcal{U}_r$.
- 5) ([25]) For any initial condition U_0 in the set

$$\mathcal{V}_r := \left\{ \Psi(A) \begin{bmatrix} K_r \\ K_\perp \end{bmatrix} \in \text{St}(r, n) \mid K_r \in \mathbb{C}^{r \times r}, \right. \\ \left. K_\perp \in \mathbb{C}^{(n-r) \times r}, \text{rank}(K_r) = r \right\}, \quad (4)$$

the solution $U(t)$ of (1) converges exponentially to $\mathcal{U}_r$.

However, these properties are insufficient for a complete convergence analysis for two main reasons:

- 1) In many practical applications, the strict spectral gap condition at the desired subspace dimension, $\text{Re}(\lambda_r) > \text{Re}(\lambda_{r+1})$, cannot be guaranteed or is difficult to verify. A convergence analysis under a milder assumption is therefore required.
- 2) The existence of other invariant sets, such as limit cycles, has not been ruled out [17].

The milder assumption was addressed in this study by investigation of the case where a spectral gap exists at an index $m \geq r$, such that $\text{Re}(\lambda_m) > \text{Re}(\lambda_{m+1})$. Furthermore, the analysis of the domain of attraction shows that no other meaningful invariant set exists.

Example 5 (Visualization of the set $\mathcal{V}_r$): Consider the Oja flow with the matrix:

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

The eigenvectors of A are

$$\psi_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \psi_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \psi_3 = \frac{1}{3} \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}.$$

These vectors are not mutually orthogonal. For $(n, r) = (3, 1)$, the stable equilibrium set is $\mathcal{U}_1 = \{\pm\psi_1\}$, whereas the unstable equilibria are $\{\pm\psi_2\}$ and $\{\pm\psi_3\}$. Fig. 1 depicts the stable equilibria $\mathcal{U}_1$ (red markers) and the unit circle formed by the intersection of $\text{St}(1, 3)$ and the plane $\text{span}\{\psi_2, \psi_3\}$ (blue line). The initial set (4) in Prop. 4 is then the unit sphere without the blue line, which means that almost all initial values converge to $\mathcal{U}_1$. This geometric observation suggests that the domain of attraction has full measure on the manifold, a claim we formalized as discussed in Section III-C.

C. Other Related Work

Several variants of the Oja flow have been proposed [5], [28]–[30]. A key property of the standard Oja flow (1) is

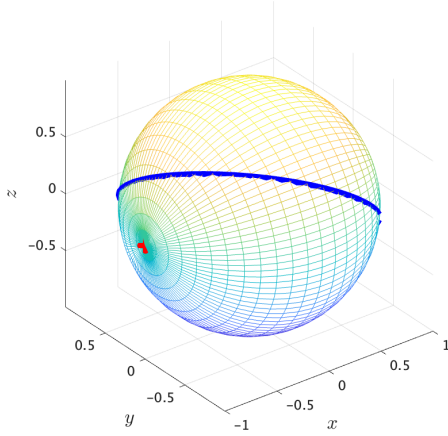


Fig. 1. Sphere representing $\text{St}(1, 3)$. The red markers represent $\mathcal{U}_1 = \{\pm\psi_1\}$ and the blue line represents the unit circle $\text{St}(1, 3) \cap \text{span}\{\psi_2, \psi_3\}$ in Example 5.

its invariance under right-multiplication by an orthogonal matrix. Specifically, if $U(t)$ is a solution, then for any differentiable orthogonal matrix $W(t) \in \mathbb{R}^{r \times r}$, the trajectory $U'(t) := U(t)W(t)$ spans the same subspace as $U(t)$ at every instant, because $U'(t)U'(t)^\top = U(t)W(t)W(t)^\top U(t)^\top = U(t)U(t)^\top$. Consequently, $U'(t)$ converges to the same invariant subspace as $U(t)$. The dynamics of $U'(t)$ can be expressed by augmenting the Oja flow with a term related to the derivative of $W(t)$. Because $W(t)$ is orthogonal, its derivative satisfies $\varepsilon \frac{d}{dt} W(t) = W(t)S(t)$ for some skew-symmetric matrix $S(t) = -S(t)^\top \in \mathbb{R}^{r \times r}$. The resulting dynamics for $U'(t)$ are given by

$$\varepsilon \frac{d}{dt} U'(t) = (I_n - U'(t)U'(t)^\top)AU'(t) + U'(t)S(t).$$

This formulation enables modifications that impose additional structure on the solution without altering the fundamental subspace dynamics.

An important application of this principle is the continuous-time reduced QR algorithm [31], [32]. Setting $\varepsilon = 1$ for simplicity, this algorithm is described by the following coupled differential equations:

$$\frac{d}{dt} U_{\text{qr}}(t) = (I_n - U_{\text{qr}}(t)U_{\text{qr}}(t)^\top)AU_{\text{qr}}(t) + U_{\text{qr}}(t)S(t), \quad (5)$$

$$\frac{d}{dt} R(t) = B(t)R(t), \quad (6)$$

with initial conditions $U_{\text{qr}}(0) \in \text{St}(r, n)$ and $R(0) \in \mathbb{R}^{r \times r}$, where $R(0)$ is an upper triangular matrix. The skew-symmetric matrix $S(t)$ is chosen at each instant such that the matrix $B(t) := U_{\text{qr}}(t)^\top AU_{\text{qr}}(t) - S(t)$ is forced to be upper triangular. If A has real eigenvalues, the diagonal elements of the solution $R(t)$ converge to the r eigenvalues of A with the largest real parts.

This property makes the reduced QR algorithm a widely used tool for estimating Lyapunov exponents and dynamical

system spectra, in a manner analogous to the full continuous-time QR algorithm (i.e., the case where $r = n$) [33]–[35]. The algorithm defined by (5) and (6) has found applications in Kalman–Bucy filtering for linear time-varying systems and in nonlinear observers [36]. Although this algorithm is designed for general square matrices and extensive numerical evidence suggests that it effectively extracts the dominant subspace, a rigorous theoretical analysis of its convergence properties is still lacking, even for the time-invariant case.

III. CONVERGENCE ANALYSIS FOR OJA FLOW

This section provides the theoretical results of the Oja flow (1). First, we demonstrate how to stabilize the Stiefel manifold $\text{St}(r, n)$ and establish the convergence rate in Euclidean space $\mathbb{R}^{n \times r}$. Next, we clarify the convergence on $\text{St}(r, n)$, and subsequently establish the domain of attraction. Finally, we provide guidance on how to efficiently increase or decrease r .

A. Convergence to the Stiefel Manifold

In this section, we extend the convergence results for the Oja flow from symmetric positive-definite matrices, as presented in Proposition 1, to the general case of arbitrary square matrices. First, we present a lemma that generalizes Lemma 2.2 of [26]. Although Hasan previously established the asymptotic convergence of (1) for general square matrices [10, Variation 4], our analysis results demonstrated exponential convergence. The following result from the literature is instrumental to our proof.

Lemma 6 ([37]): Consider the Riccati differential equation:

$$\frac{d}{dt} P(t) = AP(t) + P(t)A^\top - P(t)QP(t) + R,$$

where $A, Q = Q^\top, R = R^\top \in \mathbb{R}^{n \times n}$, with $Q \geq 0, R \geq 0$, and $P(0) = P_0 \geq 0$. Let $\bar{P} \in \mathbb{R}^{n \times n}$ be an equilibrium solution. Then, the solution is given by $P(t) = Y(t)X(t)^{-1}$ for all $t \in [0, t_{\max})$, where

$$\begin{aligned} X(t) &= e^{-\tilde{A}^\top t} \left(I_n + \int_0^t e^{\tilde{A}^\top s} R e^{\tilde{A}s} ds (P_0 - \bar{P}) \right), \\ Y(t) &= \bar{P} e^{-\tilde{A}^\top t} \\ &\quad + \left(\bar{P} \int_0^t e^{-\tilde{A}^\top (t-s)} R e^{\tilde{A}s} ds + e^{\tilde{A}t} \right) (P_0 - \bar{P}), \end{aligned}$$

with $\tilde{A} := A - \bar{P}R$, and $t_{\max} := \inf\{t \geq 0 \mid \det(X(t)) = 0\}$.

Using Lemma 6, we first establish that the rank of the solution matrix is preserved.

Lemma 7: Assume that the symmetric part of $A \in \mathbb{R}^{n \times n}$ is positive-definite, i.e., $A_{\text{sym}} > 0$. Let $U(0) \in \mathbb{R}^{n \times r}$ be of full rank. Then, the solution $U(t)$ of (1) remains full rank for all $t \geq 0$.

Proof: Let $P(t) := U(t)U(t)^\top$. Without loss of generality, we set $\varepsilon = 1$. The evolution of $P(t)$ is governed by the Riccati differential equation:

$$\frac{d}{dt} P(t) = AP(t) + P(t)A^\top - 2P(t)A_{\text{sym}}P(t),$$

with the initial condition $P(0) = P_0 := U(0)U(0)^\top$. Notably, $\bar{P} = 0_{n,n}$ is an equilibrium solution. Applying Lemma 6 with $\bar{P} = 0_n$, $Q = 2A_{\text{sym}}$, and $R = 0_n$ yields the solution components

$$X(t) = e^{-A^\top t} (I_n + G(t)P_0), \quad Y(t) = e^{At}P_0,$$

where $G(t) := 2 \int_0^t e^{A^\top s} A_{\text{sym}} e^{As} ds$. Because $A_{\text{sym}} > 0$, the integrand is positive-definite, and thus $G(t) > 0$ for any $t > 0$. The eigenvalues of $G(t)P_0$ are the same as those of the symmetric matrix $G(t)^{1/2}P_0G(t)^{1/2}$, which are non-negative. This result implies that all eigenvalues of $I_n + G(t)P_0$ are greater than or equal to 1 for all $t \geq 0$. Therefore, $X(t)$ is invertible for all $t \geq 0$, and the solution is

$$P(t) = Y(t)X(t)^{-1} = e^{At}P_0(I_n + G(t)P_0)^{-1}e^{A^\top t}. \quad (7)$$

Because e^{At} is always invertible, the rank of $P(t)$ is equal to the rank of P_0 . This result implies that the rank of $U(t)$ is preserved for all $t \geq 0$. ■

The following lemma, a generalization of Theorem 2.2 in [26], characterizes the evolution of the singular values of $U(t)$.

Lemma 8: Assume that $A_{\text{sym}} > 0$. Let $U(0) \in \mathbb{R}^{n \times r}$ have full rank, and let $\sigma_1(t) \geq \dots \geq \sigma_r(t)$ denote the singular values of $U(t)$. Then, the following hold:

- 1) $\lim_{t \rightarrow \infty} \sigma_i(t) = 1$ for all $i = 1, \dots, r$.
- 2) If $\sigma_i(0) \geq 1$ for all i , then each $\sigma_i(t)$ is a non-increasing function of t .
- 3) If $\sigma_i(0) \leq 1$ for all i , then each $\sigma_i(t)$ is a non-decreasing function of t .
- 4) $\sigma_r(t) \geq \alpha$ for all $t \geq 0$, where $\alpha := \min\{1, \sigma_r(0)\}$.

Proof: Let $P(t) = U(t)U(t)^\top$, with its solution given by (7). Without loss of generality, set $\varepsilon = 1$. The matrix $G(t)$ in (7) can be rewritten as $G(t) = \int_0^t \frac{d}{ds}(e^{A^\top s}e^{As})ds = H(t) - I_n$, where $H(t) := e^{A^\top t}e^{At}$. Take an orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ such that

$$QU(0)U(0)^\top Q^\top = \text{blk-diag}(L, 0_{n-r}), \\ L := \text{diag}(\sigma_1(0)^2, \dots, \sigma_r(0)^2).$$

Then, $P(t)$ takes the form

$$P(t) = e^{At}Q^\top \text{blk-diag}(L, 0_{n-r}) \\ \times Q\{I_n + (H(t) - I_n)P_0\}^{-1}Q^\top Qe^{A^\top t} \\ = e^{At}Q^\top \text{blk-diag}(L, 0_{n-r}) \\ \times \{I_n + (\Sigma(t) - I_n)\text{blk-diag}(L, 0_{n-r})\}^{-1}Qe^{A^\top t} \\ = e^{At}Q^\top \begin{bmatrix} \{L^{-1} - I_r + \Sigma_{11}(t)\}^{-1} & 0 \\ 0 & 0 \end{bmatrix} Qe^{A^\top t},$$

where $\Sigma(t) := QH(t)Q^\top$ and $\Sigma_{11}(t) \in \mathbb{R}^{r \times r}$ is the top-left block of $\Sigma(t)$. Because $P(t)$ is symmetric positive semidefinite, its nonzero singular values are the square roots of its nonzero eigenvalues. These eigenvalues are the same as those of the following similarity transformed matrix:

$$Q^\top e^{A^\top t}e^{At}Q \begin{bmatrix} \{L^{-1} - I_r + \Sigma_{11}(t)\}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \\ = \begin{bmatrix} \Sigma_{11}(t)\{L^{-1} - I_r + \Sigma_{11}(t)\}^{-1} & 0 \\ \Sigma_{21}(t)\{L^{-1} - I_r + \Sigma_{11}(t)\}^{-1} & 0 \end{bmatrix},$$

where $\Sigma_{21}(t) \in \mathbb{R}^{(n-r) \times r}$ is the bottom-left block of $\Sigma(t)$. Because $\Sigma(t) > 0$, $\Sigma_{11}(t) > 0$. Thus, the nonzero eigenvalues of $P(t)$ are the eigenvalues of

$$\Sigma_{11}(t)\{L^{-1} - I_r + \Sigma_{11}(t)\}^{-1} \\ = \{(L^{-1} - I_r)\Sigma_{11}(t)^{-1} + I_r\}^{-1}.$$

Recall that $\text{Re}(\lambda_i(A)) \geq \lambda_n(A_{\text{sym}}) > 0$ for all i from the assumption $A_{\text{sym}} > 0$. Therefore, $H(t)$ is monotonically increasing in the Loewner order and will diverge as $t \rightarrow \infty$. This result implies that $\Sigma_{11}(t)$ also diverges, making $\Sigma_{11}(t)^{-1}$ approach the zero matrix as $t \rightarrow \infty$. This result immediately shows that all the nonzero singular values of $U(t)$ approach 1, and hence, statement 1 holds true.

The monotonicity of $\Sigma_{11}(t)$ is then used to establish statements 2, 3, and 4 by following the same arguments as in the proof of Theorem 2.2 of [26]. ■

The condition $A_{\text{sym}} > 0$ may seem restrictive, but as shown in the next theorem, it can always be satisfied by a simple modification of the Oja flow (1) that does not alter its dynamics on the Stiefel manifold.

Theorem 9: For a given $A \in \mathbb{R}^{n \times n}$, choose a scalar $a \geq 0$ such that $A_{\text{sym}} + aI_n$ is positive-definite. Let the initial condition $U(0) \in \mathbb{R}^{n \times r}$ be a full-rank matrix. Then, the solution of the modified Oja flow

$$\varepsilon \frac{d}{dt}U(t) = (I_n - U(t)U(t)^\top)(A + aI_n)U(t)$$

converges exponentially to $\text{St}(r, n)$.

Proof: The proof follows that of Proposition 3.1 in [26]. Let $B := A + aI_n$ and consider the function $z(t) := \|I_r - U(t)^\top U(t)\|_F^2$, where $\|\bullet\|_F$ is the Frobenius norm. Its time derivative satisfies

$$\frac{d}{dt}z(t) = -2\text{Tr} \left[(I_r - U(t)^\top U(t)) \frac{d}{dt}(U(t)^\top U(t)) \right] \\ = -\frac{2}{\varepsilon} \text{Tr} \left[U(t)(I_r - U(t)^\top U(t))^2 U(t)^\top (B^\top + B) \right] \\ \leq -\frac{4}{\varepsilon} \text{Tr} \left[U(t)^\top U(t)(I_r - U(t)^\top U(t))^2 \right] \lambda_n(B_{\text{sym}}) \\ \leq -\frac{4}{\varepsilon} \lambda_r(U(t)^\top U(t)) \lambda_n(B_{\text{sym}}) z(t).$$

By construction, $\lambda_n(B_{\text{sym}}) > 0$. From Lemma 8, we know that the smallest singular value of $U(t)$ is bounded below by $\alpha > 0$, which implies $\lambda_r(U(t)^\top U(t)) \geq \alpha^2 > 0$. Thus, $z(t)$ satisfies

$$\frac{d}{dt}z(t) \leq -\frac{4}{\varepsilon} \alpha^2 \lambda_n(B_{\text{sym}}) z(t)$$

which proves exponential convergence of $z(t)$ to 0. This result is equivalent to the exponential convergence of $U(t)$ to $\text{St}(r, n)$. ■

The convergence rate is tunable by the parameters a and ε . A simple choice to ensure $A_{\text{sym}} + aI_n > 0$ is to select $a > \|A\|_F = \sqrt{\text{Tr}[A^\top A]}$. However, an exceedingly large value of a makes the system dynamics stiff, requiring a smaller time step for stable numerical integration.

To demonstrate the stabilizing effect of the spectral shift aI_n proposed in Theorem 9, we revisit the setting of Example 5. The matrix A in this example is not positive-definite.

Consequently, simulating (1) with a standard forward Euler integrator (time step $h = 0.1$) can cause the solution $U(t)$ to drift off the Stiefel manifold because of numerical errors, as illustrated by the red dashed line in Fig. 2.

In contrast, applying the same numerical scheme to the modified flow with $A + aI_n$ for $a = 2$ and $a = 4$ yields stable trajectories. Because $\lambda_3(A_{\text{sym}}) \approx -1.47$, both values of a satisfy the theorem's condition. As shown by the blue and green lines in Fig. 2, the solution may briefly leave the manifold because of initial numerical errors but is actively driven back. A larger value of a results in faster convergence to the manifold; however, an excessively large a can make the system stiff, requiring a prohibitively small time step h for numerical stability. For comparison, the black solid line shows that periodic re-normalization also keeps the solution on the manifold. However, this approach incurs additional computational cost at each step; for instance, a Householder QR decomposition requires $O(nr^2)$ flops [38, Algorithm 5.2.1], which can be demanding if r is not small.

Fig. 3 illustrates the case where the initial condition $U(0)$ starts outside the Stiefel manifold, plotted on a semi-logarithmic scale. The blue and green lines again demonstrate exponential convergence to the manifold. The trajectory with the larger value ($a = 4$, green line) exhibits faster initial convergence. After $t = 1$, the convergence rates appear similar, an artifact of the relatively large time step h . Using a smaller step size would show the superior convergence rate of the larger a over a longer time interval.

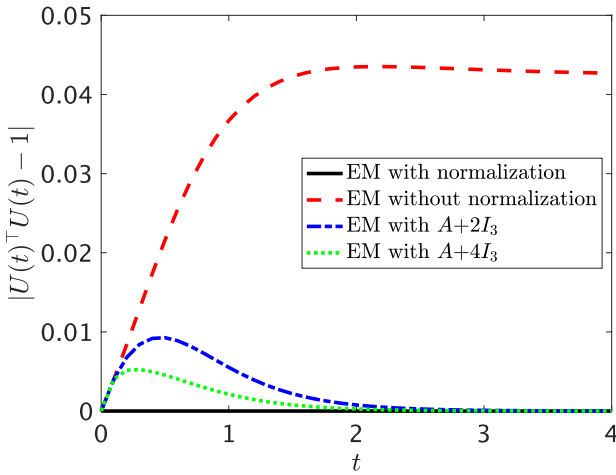


Fig. 2. Plot of $|U(t)^T U(t) - 1|$ for each time t by the forward Euler method with normalization (black solid line), without normalization (red dashed line), with $A + 2I_3$ (blue chain line), and with $A + 4I_3$ (green dotted line) under the conditions in Example 5 with $U(0) = (\psi_2 + \psi_3)/\|\psi_2 + \psi_3\|$.

Example 10 (Singular Subspace Extraction Algorithm):

Theorem 9 can be used to improve the numerical stability of online singular subspace extraction algorithms. The Weingessel–Hornik (WH) algorithm [39] for extracting the singular subspaces of a matrix $A \in \mathbb{R}^{n \times m}$ is given by a set of coupled ordinary differential equations. This algorithm can be shown to be equivalent to the Oja flow for an augmented

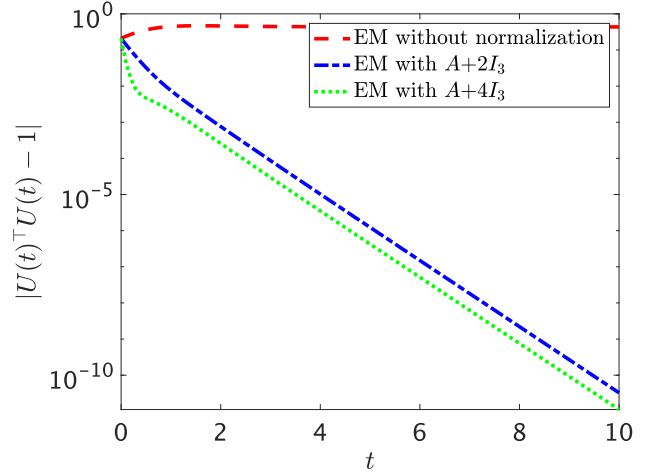


Fig. 3. Plot of $|U(t)^T U(t) - 1|$ for each time t by the forward Euler method without normalization (red dashed line), with $A + 2I_3$ (blue chain line), and with $A + 4I_3$ (green dotted line) under the conditions in Example 5 with $U(0) = \frac{1}{10}(\psi_2 + \psi_3)/\|\psi_2 + \psi_3\|$.

system [40]:

$$\varepsilon \frac{d}{dt} X(t) = (I_{n+m} - X(t)X(t)^T)AX(t), \quad (8)$$

where $X := [U^T, V^T]^T / \sqrt{2}$ and $A := \begin{bmatrix} 0_m & A^T \\ A & 0_n \end{bmatrix}$. Because A is symmetric but generally not positive-definite, a naive numerical implementation may be unstable if $X(t)$ deviates from the Stiefel manifold. To ensure numerical stability, we can apply Theorem 9 by using the shifted matrix $B := A + aI_{n+m}$ for a sufficiently large $a > 0$. This approach yields a modified WH algorithm with an additional term in each equation, which actively drives the solution back to the Stiefel manifold, eliminating the need for periodic re-normalization steps that can be computationally expensive. A detailed comparison with other truncated singular value decomposition algorithms is left for future work.

B. Convergence of Oja Flow on the Stiefel Manifold

We then analyzed the convergence of the Oja flow on the Stiefel manifold. The convergence problem can be reformulated by examining the dynamics of the projection matrix $P(t) := U(t)U(t)^T$. If $U(t)$ solves the Oja flow (1), then $P(t)$ evolves according to the matrix Riccati equation:

$$\varepsilon \frac{d}{dt} P(t) = AP(t) + P(t)A^T - P(t)(A + A^T)P(t). \quad (9)$$

To analyze this equation, we introduce an auxiliary linear system whose solution trajectory spans the same subspace:

$$\varepsilon \frac{d}{dt} Z(t) = AZ(t), \quad Z(0) \in \mathbb{R}^{n \times r}, \quad (10)$$

where $Z(0)$ is assumed to be of full rank. The solution is $Z(t) = e^{At/\varepsilon} Z(0)$, which preserves the initial rank. The projection matrix onto the subspace spanned by $Z(t)$ is $P'(t) := Z(t)(Z(t)^T Z(t))^{-1} Z(t)^T$. A direct calculation shows that $P'(t)$ also satisfies the Riccati equation (9). By setting $Z(0) = U(0)$, we ensure $P'(t) = P(t)$ for all t .

This equivalence is powerful because $P'(t)$ has an explicit analytical form:

$$P'(t) = e^{At/\varepsilon} Z(0) \left(Z(0)^\top e^{A^\top t/\varepsilon} e^{At/\varepsilon} Z(0) \right)^{-1} \times Z(0)^\top e^{A^\top t/\varepsilon}.$$

This expression enables us to analyze the Oja flow's asymptotic behavior by studying the underlying linear system.

Remark 11: The trajectory $U(t)$ of the Oja flow is not, in general, a simple normalization of $Z(t)$. An additional time-varying rotation is involved, such that $U(t) = Z(t)(Z(t)^\top Z(t))^{-1/2} R(t)$ for some orthogonal matrix $R(t) \in \mathbb{R}^{r \times r}$. Characterizing this rotation is non-trivial, which motivates our analysis of the projection matrix $P(t)$ instead of $U(t)$ directly.

Theorem 12: Assume the eigenvalues of $A \in \mathbb{R}^{n \times n}$ satisfy the gap condition $\text{Re}(\lambda_m(A)) > \text{Re}(\lambda_{m+1}(A))$ for some integer m with $r \leq m < n$. Then, for any initial condition U_0 in the set

$$\mathcal{V}_{m,r} := \left\{ \Psi(A) \begin{bmatrix} K_{m,r} \\ K_{m,\perp} \end{bmatrix} \in \text{St}(r, n) \mid K_{m,r} \in \mathbb{C}^{m \times r}, K_{m,\perp} \in \mathbb{C}^{(n-m) \times r}, \text{rank}(K_{m,r}) = r \right\},$$

the solution $U(t)$ of (1) converges exponentially to the invariant set

$$\mathcal{U}_{m,r} := \left\{ \Psi(A) \begin{bmatrix} K_{m,r} \\ 0_{n-m,r} \end{bmatrix} \in \text{St}(r, n) \mid K_{m,r} \in \mathbb{C}^{m \times r} \right\}.$$

The convergence rate is governed by the exponent $-(\text{Re}(\lambda_m) - \text{Re}(\lambda_{m+1}) - \delta)/\varepsilon$, where $\delta > 0$ is an arbitrarily small constant. In the particular case where $r = m$, any solution starting in $\mathcal{V}_r := \mathcal{V}_{r,r}$ converges exponentially to an element of $\mathcal{U}_r := \mathcal{U}_{r,r}$.

Proof: Let $\Lambda = \Psi^{-1}A\Psi$ be the Jordan normal form of A , where $\Psi = \Psi(A)$. We analyze the convergence of the projection matrix $P(t) = P'(t) = Z(t)(Z(t)^\top Z(t))^{-1} Z(t)^\top$, using its explicit form derived from the auxiliary system (10) with $Z(0) = U_0 \in \mathcal{V}_{m,r}$.

The key idea is to analyze the system in a scaled coordinate frame that factors out the dominant exponential growth. Let $Z'(t) = e^{-\text{Re}(\lambda_m)t/\varepsilon} Z(t)$. The projection matrix can be written as $P(t) = Z'(t)(Z'(t)^\top Z'(t))^{-1} Z'(t)^\top$. Then, $Z'(t)$ becomes

$$Z'(t) = \Psi \begin{bmatrix} e^{(\Lambda_m - \text{Re}(\lambda_m)I_m)t/\varepsilon} K_{m,r} \\ e^{(\Lambda_\perp - \text{Re}(\lambda_m)I_{n-m})t/\varepsilon} K_{m,\perp} \end{bmatrix}$$

and $\Lambda_m \in \mathbb{C}^{m \times m}$ and $\Lambda_\perp \in \mathbb{C}^{(n-m) \times (n-m)}$ are Jordan blocks forming $\Lambda = \text{blk-diag}(\Lambda_m, \Lambda_\perp)$, corresponding to the dominant and non-dominant eigenvalues. Because of the eigenvalue gap condition, the components of the solution associated with $\Lambda_\perp$ decay exponentially relative to the components associated with Λ_m . Specifically, one can show that

$$\|e^{\Lambda_\perp t/\varepsilon} K_{m,\perp}\|_{\text{ind}} e^{-\text{Re}(\lambda_m)t/\varepsilon} \leq c e^{(\text{Re}(\lambda_{m+1}) - \text{Re}(\lambda_m) + \delta)t/\varepsilon} \rightarrow 0,$$

for an arbitrarily small $\delta > 0$ and a constant $c > 0$. Thus, $P(t) = P'(t)$ converges to the subset $\{\bar{U}\bar{U}^\top \mid \bar{U} \in \mathcal{U}_{m,r}\}$. This result proves that $U(t)$ converges to $\mathcal{U}_{m,r}$.

For the special case $r = m$, the analysis can be refined to show exponential convergence to a specific projector, which implies convergence of $U(t)$ to the set $\mathcal{U}_r$. Because $K_r := K_{r,r}$ is square and full-rank, we can consider $P'(t) = \hat{Z}(t)(\hat{Z}(t)^\dagger \hat{Z}(t))^{-1} \hat{Z}(t)^\dagger$, where

$$\hat{Z}(t) = Z(t)(e^{\Lambda_r t/\varepsilon} K_r)^{-1} = \Psi \begin{bmatrix} I_r \\ e^{\Lambda_\perp t/\varepsilon} K_{r,\perp} K_r^{-1} e^{-\Lambda_r t/\varepsilon} \end{bmatrix}.$$

Using norm inequality,

$$\|e^{\Lambda_\perp t/\varepsilon} K_{r,\perp} K_r^{-1} e^{-\Lambda_r t/\varepsilon}\|_{\text{ind}} \leq c' e^{(\text{Re}(\lambda_{m+1}) - \text{Re}(\lambda_m) + \delta)t/\varepsilon}$$

holds, where $\delta > 0$ is an arbitrarily small constant and $c' > 0$. Thus, $\lim_{t \rightarrow \infty} \hat{Z}(t) = \Psi[I_r, 0_{r, n-r}]^\top$, proving that $U(t)$ converges to $\mathcal{U}_r$. Because any element of $\mathcal{U}_r$ is an equilibrium solution of (1), $U(t)$ converges to an element of $\mathcal{U}_r$. ■

Theorem 12 establishes convergence for the general case where $r \leq m$, thereby extending the recent result of Kuo et al. [25], which addresses the specific case where $r = m$. A consequence of the proof is that the projection matrix $\bar{U}\bar{U}^\top$ is unique for any $\bar{U} \in \mathcal{U}_r$. The initial condition requires that the subspace spanned by U_0 have a non-trivial projection onto the dominant invariant subspace of A . As discussed in Section III-C, this condition is not overly restrictive.

Fig. 4 shows the convergence of the Oja flow (1) to the dominant mode $\mathcal{U}_1$. The matrix A is defined in Example 5, and the forward Euler scheme is used with time step $h = 0.1$ and $\varepsilon = 1$. The black line shows the upper bound of the convergence rate $-(\text{Re}(\lambda_1(A)) - \text{Re}(\lambda_2(A))) = -1$. The figure shows that the convergence rate of $U(t)$ is slightly faster than the upper bound. To emphasize that the actual convergence rate is faster than the upper bound, we use $0.7e^{-t}$ rather than using the theoretical upper bound e^{-t} .

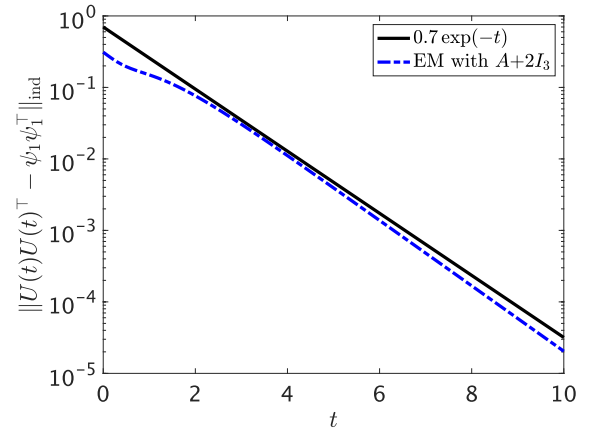


Fig. 4. Plot of $\|U(t)U(t)^\top - \psi_1 \psi_1^\top\|_{\text{ind}}$ for each time t using the forward Euler method with $A + 2I_3$ (blue chain line) under the conditions in Example 5 with $U(0) = (\psi_1 + \psi_2 + \psi_3)/\|\psi_1 + \psi_2 + \psi_3\|$. The black line shows the upper bound of the convergence rate.

Remark 13 (On the case $r < m$): When the desired subspace dimension r is less than the number of dominant modes

m , the eigenvalues of the projected matrix $\bar{U}^\top A \bar{U}$ for an equilibrium point $\bar{U} \in \mathcal{U}_{m,r}$ are not necessarily a subset of the eigenvalues of A . Notably, $\bar{U}$ is written as $\bar{U} = \Psi[K_{m,r}^\top, 0_{r,n-m}]^\top$ and the QR decomposition yields $K_{m,r} = Q_{m,r}K_r$, where $Q_{m,r} \in \mathbb{C}^{m \times r}$ has orthonormal columns and $K_r \in \mathbb{C}^{r \times r}$ is invertible. Let $Q := [Q_{m,r}^\top, 0_{r,(n-m)}]^\top \in \mathbb{C}^{n \times r}$ and $Q_\perp \in \mathbb{C}^{n \times (n-r)}$ so that $[Q, Q_\perp]$ is a unitary matrix. Following similar arguments to the proof of [17, Prop. 3],

$$\begin{aligned} & \bar{U}^\top A \bar{U} \\ &= K_r^\dagger Q^\dagger \Psi^\dagger \Psi \\ & \times [QK_r, Q_\perp] \begin{bmatrix} K_r^{-1} Q^\dagger \\ Q_\perp^\dagger \end{bmatrix} \begin{bmatrix} \Lambda_m Q_{m,r} K_r \\ 0_{(n-r),r} \end{bmatrix} \\ &= [I_r, K_r^\dagger Q^\dagger \Psi^\dagger \Psi Q_\perp] \begin{bmatrix} K_r^{-1} Q^\dagger \Lambda_m Q_{m,r} K_r \\ Q_\perp^\dagger [(\Lambda_m Q_{m,r} K_r)^\top, 0_{(n-r),r}^\top]^\top \end{bmatrix} \\ &= K_r^{-1} Q_{m,r}^\dagger \Lambda_m Q_{m,r} K_r \\ & + K_r^\dagger Q^\dagger \Psi^\dagger \Psi Q_\perp Q_\perp^\dagger [(\Lambda_m Q_{m,r} K_r)^\top, 0_{(n-r),r}^\top]^\top \end{aligned}$$

holds. Consequently, the eigenvalues of $\bar{U}^\top A \bar{U}$ will generally not coincide with any of the eigenvalues of A for $r < m$.

Conversely, for the case where $r > m$, if the solution is initialized within a larger invariant subspace, it will converge while preserving its dominant components. This idea is formalized in the following discussion.

Proposition 14: Assume there exist integers m and m' such that $1 \leq m < r \leq m' < n$ and that the eigenvalue gap conditions $\text{Re}(\lambda_m) > \text{Re}(\lambda_{m+1})$ and $\text{Re}(\lambda_{m'}) > \text{Re}(\lambda_{m'+1})$ both hold. If the initial matrix $U(0)$ belongs to the set

$$\left\{ \Psi(A) \begin{bmatrix} K_{m,r} \\ K_{m,\perp} \end{bmatrix} \in \mathcal{V}_{m',r} \mid K_{m,r} \in \mathbb{C}^{m \times r}, \right. \\ \left. \text{rank}(K_{m,r}) = m \right\},$$

then the solution $U(t)$ converges to an element of the set

$$\left\{ \Psi(A) \begin{bmatrix} K'_{m,r} \\ K'_{m'-m,\perp} \\ 0_{n-m',r} \end{bmatrix} \in \mathcal{U}_{m',r} \mid K'_{m,r} \in \mathbb{C}^{m \times r}, \right. \\ \left. \text{rank}(K'_{m,r}) = m, K'_{m'-m,\perp} \in \mathbb{C}^{(m'-m) \times r} \right\}.$$

In essence, this proposition states that if the initial condition is contained within the m' -dominant invariant subspace and already spans the m -dominant invariant subspace, then the flow will converge to the m' -dominant subspace while preserving the span of the m -dominant component.

Proof: From the proof of Theorem 12, the projection matrix $P(t) = U(t)U(t)^\top$ converges to the set of projectors onto the subspace $\mathcal{U}_{m',r}$. The asymptotic behavior of $P(t)$ is determined by the projection matrix $S(t) := V(t)(V(t)^\top V(t))^{-1}V(t)^\top$, where

$$V(t) = e^{At/\varepsilon} \Psi \begin{bmatrix} K_{m,r} \\ K'_{m'-m,\perp} \\ 0_{n-m',r} \end{bmatrix} = \Psi \begin{bmatrix} e^{\Lambda_m t/\varepsilon} K_{m,r} \\ e^{\Lambda_{m',m'/t/\varepsilon} K'_{m'-m,\perp} \\ 0_{n-m',r} \end{bmatrix}.$$

Here, $\Lambda_m \in \mathbb{C}^{m \times m}$ and $\Lambda_{m',m'} \in \mathbb{C}^{(m'-m) \times (m'-m)}$ are Jordan blocks containing the eigenvalues $\{\lambda_i\}_{i=1}^m$ and $\{\lambda_i\}_{i=m+1}^{m'}$, respectively. The initial condition ensures that the matrix $[K_{m,r}^\top, K_{m'-m,\perp}^\top]^\top$ has full rank r , which guarantees that $V(t)^\top V(t)$ is invertible for all t . Furthermore, because $K_{m,r}$ has full rank m and $e^{\Lambda_m t/\varepsilon}$ is non-singular, the block $e^{\Lambda_m t/\varepsilon} K_{m,r}$ retains rank m for all $t \geq 0$. This result confirms that the component of the solution spanning the m -dominant subspace does not lose rank, which proves the proposition. ■

C. Evaluation of the Domain of Attraction

Next, we investigated the size of the domain of attraction, $\mathcal{V}_{m,r}$, within the Stiefel manifold. We consider only $m < n$; otherwise, $\mathcal{V}_{n,r} = \text{St}(r, n)$. Intuitively, an element $U \in \text{St}(r, n)$ is determined by its coordinate representation $K = \Psi^{-1}U$. The condition for convergence, as stated in Theorem 12, is that a specific sub-block of K , denoted $K_{m,r}$, must have full rank. For a randomly chosen U , it is highly improbable that this sub-block would be rank-deficient. This result suggests that the set of initial conditions leading to convergence, $\mathcal{V}_{m,r}$, should encompass almost the entire manifold. We formalize this claim in the following propositions.

Proposition 15: The set $\mathcal{V}_{m,r}$ is dense in $\text{St}(r, n)$. That is, its closure is the entire manifold: $\overline{\mathcal{V}_{m,r}} = \text{St}(r, n)$.

Proof: To prove that $\mathcal{V}_{m,r}$ is dense in $\text{St}(r, n)$, we must show that for any $U \in \text{St}(r, n)$ and any open neighborhood $\mathcal{N}_\delta(U)$ of U , the intersection $\mathcal{N}_\delta(U) \cap \mathcal{V}_{m,r}$ is non-empty [41, Thm 17.5]. We define a neighborhood as

$$\mathcal{N}_\delta(U) := \{V \in \text{St}(r, n) \mid \|U - V\|_{\text{ind}} < \delta\}, \quad \delta > 0. \quad (11)$$

If $U \in \mathcal{V}_{m,r}$, the condition is trivially satisfied. Therefore, we consider the case where $U \in \text{St}(r, n) \setminus \mathcal{V}_{m,r}$, which implies that its coordinate representation $K = \Psi^{-1}U = [K_{m,r}^\top, K_{m,\perp}^\top]^\top$ has a rank-deficient sub-block, i.e., $\text{rank}(K_{m,r}) < r$.

Notably, for any orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ and any $U \in \text{St}(r, n)$, $QU \in \text{St}(r, n)$ holds. Consider a skew-symmetric matrix $W = -W^\top \in \mathbb{R}^{n \times n}$ with $\|W\|_{\text{ind}} = 1$ and $Q = \exp(\bar{\delta}W)$ for a small $\bar{\delta} > 0$. We show that we can choose $\bar{\delta} > 0$ such that $QU \in \mathcal{N}_\delta(U)$ for a given $\delta > 0$. Using the operator norm inequality yields

$$\begin{aligned} \|U - QU\|_{\text{ind}} &\leq \|I_n - \exp(\bar{\delta}W)\|_{\text{ind}} \|U\|_{\text{ind}} \\ &= \|I_n - \exp(\bar{\delta}W)\|_{\text{ind}} \end{aligned}$$

and because $\|W\|_{\text{ind}} = 1$,

$$\|I_n - \exp(\bar{\delta}W)\|_{\text{ind}} \leq \sum_{k=1}^{\infty} \frac{\bar{\delta}^k}{k!} = e^{\bar{\delta}} - 1$$

holds. Taking $\bar{\delta} > 0$ such that $\bar{\delta} < \ln(\delta + 1)$ ensures $QU \in \mathcal{N}_\delta(U)$.

Our strategy is to construct a matrix V that is arbitrarily close to U but lies in $\mathcal{V}_{m,r}$. We construct V by applying a small orthogonal rotation to U , i.e., $V = QU$, where $Q = \exp(\bar{\delta}W)$ for a small scalar $\bar{\delta} > 0$ and a skew-symmetric matrix W with $\|W\|_{\text{ind}} = 1$.

The crucial step is to choose W such that the resulting matrix V is in $\mathcal{V}_{m,r}$. Let $K'_{m,r} \in \mathbb{C}^{m \times r}$ be a matrix of full rank

r such that $\Psi_m K'_{m,r} \in \text{St}(r, n)$, where $\Psi_m = [\psi_1, \dots, \psi_m] \in \mathbb{C}^{n \times m}$. We construct the rotation generator as $W = c(XY^\dagger - YX^\dagger)$, where $X = \Psi_m K'_{m,r} \in \text{St}(r, n)$, $Y = U = \Psi K$, and c is a normalization constant. Notably, $\text{rank}(K'_{m,r}) = r$. This choice of W perturbs U in a direction that mixes its components with those of a full-rank element.

The coordinate representation of the perturbed matrix is $\Psi^{-1}V = \Psi^{-1}QU$. A first-order Taylor expansion yields

$$\begin{aligned} \Psi^{-1}V &= \Psi^{-1} \exp(\bar{\delta}W) \Psi K = K + \bar{\delta} \Psi^{-1}W \Psi K + O(\bar{\delta}^2) \\ &= K + c\bar{\delta} \left(\begin{bmatrix} K'_{m,r} \\ 0_{n-m,r} \end{bmatrix} - K(\Psi_m K'_{m,r})^\dagger U \right) + O(\bar{\delta}^2). \end{aligned}$$

The top $m \times r$ block of the perturbed coordinates, which we denote as $K'_{m,r}(\bar{\delta})$, is

$$K'_{m,r}(\bar{\delta}) = K_{m,r}(I_r - c\bar{\delta}(X^\dagger U)) + c\bar{\delta}K'_{m,r} + O(\bar{\delta}^2).$$

Because $K_{m,r}$ is rank-deficient and $K'_{m,r}$ is full-rank, the first-order term $c\bar{\delta}K'_{m,r}$ perturbs $K_{m,r}$ in a direction that restores its rank. For a sufficiently small $\bar{\delta} > 0$, the higher-order terms do not alter this outcome, so $\text{rank}(K'_{m,r}(\bar{\delta})) = r$. This result implies that $V = QU \in \mathcal{V}_{m,r}$. Because we can find such a V in any arbitrarily small neighborhood of any U , we conclude that $\mathcal{V}_{m,r}$ is dense in $\text{St}(r, n)$. ■

Proposition 15 establishes that $\mathcal{V}_{m,r}$ is dense, but this result is not sufficient to make claims about its volume. For example, the set of rational numbers is dense in the real line but has zero Lebesgue measure. On the Stiefel manifold, there exists a uniform invariant measure [42, §1.4.3]. To show that $\mathcal{V}_{m,r}$ has the same volume as $\text{St}(r, n)$, the following proposition fills the gap.

Proposition 16: For any $U \in \mathcal{V}_{m,r}$,

- 1) $\mathcal{N}_\delta(U)$ is a path-connected set for $\delta \in (0, 2)$.
- 2) There exists $\delta > 0$ such that $\mathcal{N}_\delta(U) \subset \mathcal{V}_{m,r}$, i.e., $\mathcal{V}_{m,r}$ is an open subset of $\text{St}(r, n)$.

Proof: 1) Path-connectedness of neighborhoods: For any $U \in \mathcal{V}_{m,r}$, the singular value decomposition is $U = Q[I_r, 0_{r,n-r}]^\top$, where $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. Hence, for any $U_1, U_2 \in \text{St}(r, n)$, there exist orthogonal matrices $Q \in \mathbb{R}^{n \times n}$ such that $U_2 = QU_1$. Hence, $V \in \mathcal{N}_\delta(U)$ is represented as $V = QU$, and from the definition of $\mathcal{N}_\delta(U)$, we can take Q that satisfies

$$\|U - QU\|_{\text{ind}} \leq \|I_n - Q\|_{\text{ind}} = \max_{i=1, \dots, n} |1 - \lambda_i(Q)| < \delta.$$

Because $Q \in \mathbb{R}^{n \times n}$ is orthogonal, all the eigenvalues are on the unit circle within the δ -ball from 1, and therefore, Q does not contain its eigenvalue on -1 if we take $\delta < 2$. This result means that $\det(Q) = 1$, i.e., Q is in a special orthogonal group $SO(n)$. Because $SO(n)$ is a simple Lie group, $\mathcal{N}_\delta(U)$ is path-connected for small $\delta > 0$.

2) Openness of $\mathcal{V}_{m,r}$: We must show that for any $U \in \mathcal{V}_{m,r}$, there exists a neighborhood $\mathcal{N}_\delta(U)$ that is entirely contained within $\mathcal{V}_{m,r}$. If $U \in \mathcal{V}_{m,r}$, its coordinate block $K_{m,r}$ has full rank. The rank of a full-rank matrix does not decrease its rank under small perturbations. Therefore, for any sufficiently small perturbation of U to a nearby point $V \in \mathcal{N}_\delta(U)$, the corresponding coordinate block $K'_{m,r}$ will also have full rank.

This result implies there exists a $\delta > 0$ such that $\mathcal{N}_\delta(U) \subset \mathcal{V}_{m,r}$, proving that $\mathcal{V}_{m,r}$ is an open set. ■

The Stiefel manifold $\text{St}(r, n)$ is a compact manifold equipped with a uniform invariant measure, giving it a finite, well-defined volume [42, §1.4.4]. Because $\mathcal{V}_{m,r}$ is an open and dense subset of $\text{St}(r, n)$, its complement must be a closed set with an empty interior, which has measure zero. This leads to our main result on the domain of attraction.

Theorem 17: The volume of the set $\mathcal{V}_{m,r}$ is equal to the volume of the Stiefel manifold $\text{St}(r, n)$.

A consequence of Theorems 12 and 17 implies that there is no other stable invariant set for the Oja flow (1). Theorem 17 implies that if an initial condition $U(0)$ is chosen uniformly at random from $\text{St}(r, n)$, it will belong to the domain of attraction $\mathcal{V}_{m,r}$ with probability one. Combining this result with those described in Section III-A, one can confidently initialize the Oja flow from an arbitrary full-rank matrix in $\mathbb{R}^{n \times r}$. By using a suitable shift $a > 0$ to ensure $A_{\text{sym}} + aI_n > 0$, the trajectory will first converge to the Stiefel manifold and then, with probability one, proceed to converge to the dominant invariant subspace $\mathcal{U}_{m,r}$.

D. Change in the Number of Dominant Component Subspaces

In this analysis, we assume that a solution $\bar{U}_r \in \mathcal{U}_r$ spanning an r -dimensional dominant invariant subspace has been found. We now explore efficient methods for obtaining related subspaces of either higher or lower dimension.

1) *Expanding the Subspace:* The following proposition provides a computationally efficient method for finding the principal components corresponding to a larger, $(r + \ell)$ -dimensional subspace, given the solution for the r -dimensional one.

Proposition 18 (Subspace Expansion): Assume the eigenvalues of $A \in \mathbb{R}^{n \times n}$ satisfy the gap conditions $\text{Re}(\lambda_r(A)) > \text{Re}(\lambda_{r+1}(A))$ and $\text{Re}(\lambda_{r+\ell}(A)) > \text{Re}(\lambda_{r+\ell+1}(A))$ for $r \geq 1$ and $1 \leq \ell < n - r$. Given a solution $\bar{U}_r \in \mathcal{U}_r$, let $\bar{U}_\perp \in \text{St}(n - r, n)$ be an orthonormal basis for the orthogonal complement of the subspace spanned by $\bar{U}_r$, such that $\bar{U}_\perp \bar{U}_r^\top = I_n - \bar{U}_r \bar{U}_r^\top$. Consider the reduced-order Oja flow:

$$\varepsilon \frac{d}{dt} u(t) = (I_{n-r} - u(t)u(t)^\top) A_{\bar{U}_\perp} u(t), \quad (12)$$

where $u(t) \in \mathbb{R}^{(n-r) \times \ell}$ and $A_{\bar{U}_\perp} := \bar{U}_\perp^\top A \bar{U}_\perp$. Then, for an initial condition $u(0)$ such that $[\bar{U}_r, \bar{U}_\perp u(0)] \in \mathcal{V}_{r+\ell}$, the combined solution $U_{r+\ell}(t) := [\bar{U}_r, \bar{U}_\perp u(t)]$ converges to an element of $\mathcal{U}_{r+\ell}$.

Proof: Because $\bar{U}_r$ spans an invariant subspace of A , the matrix A is block upper-triangular in the basis $[\bar{U}_r, \bar{U}_\perp]$, meaning $\bar{U}_\perp^\top A \bar{U}_r = 0$. Consequently, the eigenvalues of the projected matrix $A_{\bar{U}_\perp}$ are precisely the remaining $n - r$ eigenvalues of A , i.e., $\{\lambda_{r+1}(A), \dots, \lambda_n(A)\}$. From Theorem 12, the solution $u(t)$ of the reduced-order Oja flow (12) converges to the ℓ -dimensional dominant invariant subspace of $A_{\bar{U}_\perp}$.

Here, we verify that the dynamics of the composite matrix $U_{r+\ell}(t) := [\bar{U}_r, \bar{U}_\perp u(t)]$ are equivalent to the full $(r + \ell)$ -dimensional Oja flow. The time derivative is $\frac{d}{dt} U_{r+\ell}(t) =$

$[0_{n,r}, \bar{U}_\perp \frac{d}{dt} u(t)]$. Substituting the dynamics from (12) yields

$$\frac{d}{dt} U_{r+\ell}(t) = \frac{1}{\varepsilon} [0_{n,r}, \bar{U}_\perp (I_{n-r} - u(t)u(t)^\top) A \bar{U}_\perp u(t)].$$

Conversely, the full Oja flow for $U_{r+\ell}(t)$ is $(I_n - U_{r+\ell}(t)U_{r+\ell}(t)^\top)AU_{r+\ell}(t)$. Using the facts that $\bar{U}_\perp^\top A \bar{U}_r = 0$ and $I_n - U_{r+\ell}(t)U_{r+\ell}(t)^\top = \bar{U}_\perp (I_{n-r} - u(t)u(t)^\top) \bar{U}_\perp^\top$, one can show that this expression simplifies to the same result. Therefore, the trajectory of $U_{r+\ell}(t)$ is identical to that of an $(r + \ell)$ -dimensional Oja flow, and its convergence to $\mathcal{U}_{r+\ell}$ is guaranteed by Theorem 12. ■

Because (12) is a lower-dimensional Oja flow, the stabilization technique from Theorem 9 can be applied to ensure robust numerical computation. Although constructing the basis $\bar{U}_\perp$ for the orthogonal complement may be non-trivial, it can often be performed as an offline computation.

2) Reducing the Subspace: Next, we consider methods for extracting a lower-dimensional, $\tilde{r}$ -dominant subspace (with $\tilde{r} < r$) from a given solution $\bar{U}_r \in \mathcal{U}_r$.

a) Method 1: Eigendecomposition: As shown in [17, Prop. 3], the projected matrix $\bar{U}_r^\top A \bar{U}_r$ is related to the dominant Jordan block of A by a similarity transform: $\bar{U}_r^\top A \bar{U}_r = K_r^{-1} \Lambda_r K_r$. The matrix of eigenvectors of this small $r \times r$ problem gives the coordinate transformation K_r . The first r eigenvectors of the original matrix A can then be recovered as $\Psi_r = \bar{U}_r K_r$. To obtain the $\tilde{r}$ -dominant subspace, one can simply select the first $\tilde{r}$ columns of Ψ_r to form $\Psi_{\tilde{r}}$, and then re-orthonormalize to obtain $\bar{U}_{\tilde{r}} \in \mathcal{U}_{\tilde{r}}$. This approach is highly efficient for small r , but the cost of the $r \times r$ eigendecomposition can become significant for larger r .

b) Method 2: Recursive Oja Flow: An alternative iterative approach is to apply the Oja flow recursively. Because the small matrix $A_{\bar{U}_r} := \bar{U}_r^\top A \bar{U}_r \in \mathbb{R}^{r \times r}$ is readily available, we can solve the following reduced-order Oja flow on its domain:

$$\varepsilon \frac{d}{dt} \tilde{U}(t) = (I_r - \tilde{U}(t)\tilde{U}(t)^\top) A_{\bar{U}_r} \tilde{U}(t),$$

where $\tilde{U}(0) \in \text{St}(\tilde{r}, r)$ is chosen from the domain of attraction for this smaller system. The solution $\tilde{U}(t)$ converges to an element $\tilde{U}_\infty \in \mathcal{U}_{\tilde{r}}^r$, where $\mathcal{U}_{\tilde{r}}^r$ is the set of $\tilde{r}$ -dominant invariant subspaces of $A_{\bar{U}_r}$.

The key insight is that the solution of this small-scale problem directly provides the projection needed to extract the desired subspace from the original solution $\bar{U}_r$. An element of $\mathcal{U}_{\tilde{r}}^r$ has the form $K_r^{-1}[\tilde{K}_{\tilde{r}}^\top, 0]^\top$. Therefore, the desired $\tilde{r}$ -dimensional subspace of the original system is obtained by the simple matrix product:

$$\bar{U}_{\tilde{r}} := \bar{U}_r \tilde{U}_\infty \in \mathcal{U}_{\tilde{r}}.$$

This method avoids a full eigendecomposition and can be advantageous when r is moderately large.

IV. APPLICATIONS TO CONTROL PROBLEMS

The extraction of dominant eigenvalues and eigenvectors is essential for analyzing large-scale dynamical systems and recursive algorithms. We demonstrate how to utilize the Oja flow (1) for model reduction, low-rank controller synthesis, and singularly perturbed systems. We also provide some theoretical guarantees for these applications.

A. Model Reduction for Linear Dynamical Systems

In practical control problems, model reduction is crucial for designing controllers for large and complex systems [13], [43]–[45]. Many existing methods, such as balanced truncation, primarily focus on stable, linear, time-invariant systems [46]–[48]. Balanced truncation relies on computing controllability and observability Gramians, a task that can be computationally prohibitive for large-scale systems. One of its key advantages is that the modeling error can be bounded in the H^∞ or H^2 norm. Although extensions such as time-limited and linear-quadratic-Gaussian (LQG) balanced truncation exist for unstable systems [49]–[52], they still require solving large-scale algebraic Riccati equations to obtain the necessary Gramians. Other prominent techniques, such as Hankel norm approximation and Krylov subspace methods, are also often computationally demanding or are restricted to stable systems [46], [53]. In this section, we propose model reduction methods and a related controller synthesis framework based on the Oja flow, which avoids the need to solve for Gramians directly.

For convenience, given a matrix $Y \in \mathbb{R}^{n \times q}$, we adopt the standard notation for projected systems: $A_Y := Y^\top A Y$, $B_Y := Y^\top B$, and $C_Y := C Y$.

1) Properties of Oja Flow Associated with A and $A^\top$: Assume that for a given matrix $A \in \mathbb{R}^{n \times n}$, the eigenvalue gap condition $\text{Re}(\lambda_r(A)) > \text{Re}(\lambda_{r+1}(A))$ holds. The Oja flow (1) applied to A and $A^\top$ can then extract the dominant right and left invariant subspaces, respectively. We denote the resulting steady-state solutions as $\bar{U} \in \mathcal{U}_r(A)$ and $\bar{V} \in \mathcal{U}_r(A^\top)$, where $\mathcal{U}_r(X)$ is the stable equilibrium set for the Oja flow associated with matrix X . Let $\bar{U}_\perp \in \text{St}(n-r, n)$ and $\bar{V}_\perp \in \text{St}(n-r, n)$ be orthonormal bases for the orthogonal complements of the subspaces spanned by $\bar{U}$ and $\bar{V}$, respectively. Using the orthogonal matrices $Q_{\bar{U}} := [\bar{U}, \bar{U}_\perp]$ and $Q_{\bar{V}} := [\bar{V}, \bar{V}_\perp]$, we can subject A to similarity transformations that reveal a block-triangular structure:

$$Q_{\bar{U}}^\top A Q_{\bar{U}} = \begin{bmatrix} A_{\bar{U}} & \bar{U}^\top A \bar{U}_\perp \\ 0_{n-r,r} & A_{\bar{U}_\perp} \end{bmatrix}, \quad (13)$$

$$Q_{\bar{V}}^\top A Q_{\bar{V}} = \begin{bmatrix} A_{\bar{V}} & 0_{r,n-r} \\ \bar{V}_\perp^\top A \bar{V} & A_{\bar{V}_\perp} \end{bmatrix}. \quad (14)$$

The $(2, 1)$ block of (13) is zero because $\bar{U}$ is an equilibrium point of its Oja flow, which implies $(I_n - \bar{U}\bar{U}^\top)A\bar{U} = 0_{n,r}$, and therefore $\bar{U}_\perp^\top A \bar{U} = 0_{n-r,r}$. Similarly, the $(1, 2)$ block of (14) is zero, which means $\bar{V}^\top A \bar{V}_\perp = 0_{r,n-r}$.

The quadruple $(\bar{U}, \bar{U}_\perp, \bar{V}, \bar{V}_\perp)$ exhibits the following important properties.

Lemma 19: The matrix $\bar{U}_\perp$ spans the minor left invariant subspace of A ; specifically, $\bar{U}_\perp \in \mathcal{U}_{n-r}(-A^\top)$. Similarly, $\bar{V}_\perp \in \mathcal{U}_{n-r}(-A)$. Furthermore, the matrix $\bar{V}^\top \bar{U}$ is non-singular.

Proof: Any $\bar{U} \in \mathcal{U}_r(A)$ has the representation

$$\bar{U} = \Psi(A) \begin{bmatrix} K_r \\ 0_{n-r,r} \end{bmatrix}, \quad K_r \in \mathbb{C}^{r \times r}, \quad \det(K_r) \neq 0.$$

From the definition, $\bar{U}_\perp^\top \bar{U} = 0_{n-r,r}$. The columns of $\bar{U}_\perp$ span the complementary subspace of the dominant r -dimensional

eigensubspace. Thus, $\bar{U}_\perp$ is associated with the remaining $n-r$ eigenvalues, and its form is

$$\bar{U}_\perp = \Psi(A)^{-\dagger} \begin{bmatrix} 0_{r,n-r} \\ \tilde{K}_\perp \end{bmatrix}, \quad \tilde{K}_\perp \in \mathbb{C}^{(n-r) \times (n-r)},$$

$$\det(\tilde{K}_\perp) \neq 0.$$

Because $A = \Psi(A)\Lambda\Psi(A)^{-1}$, we have

$$A^\top = \Psi(A)^{-\dagger} \Lambda^\dagger \Psi(A)^\dagger = \Psi(A^\top) \Lambda \Psi(A^\top)^{-1},$$

where $\Psi(A^\top) = \Psi(A)^{-\dagger} P$ and $P \in \mathbb{R}^{n \times n}$ is a permutation matrix. From the eigenvalue gap condition, $P = \text{blk-diag}(P_r, P_\perp)$, where $P_r \in \mathbb{R}^{r \times r}$ and $P_\perp \in \mathbb{R}^{(n-r) \times (n-r)}$ are permutation matrices corresponding to the dominant eigenvalues; $\bar{V} \in \mathcal{U}_r(A^\top)$ is given by

$$\bar{V} = \Psi(A^\top) \begin{bmatrix} L'_r \\ 0_{n-r,r} \end{bmatrix} = \Psi(A)^{-\dagger} \begin{bmatrix} L_r \\ 0_{n-r,r} \end{bmatrix},$$

where $L'_r \in \mathbb{C}^{r \times r}$ is non-singular and $L_r = P_r L'_r$. From the definition of $\mathcal{U}_r(\bullet)$, $\bar{U}_\perp \in \mathcal{U}_{n-r}(-A^\top)$ and $\bar{V}_\perp \in \mathcal{U}_{n-r}(-A)$.

Using these representations, the product $\bar{V}^\top \bar{U}$ is

$$\begin{aligned} \bar{V}^\top \bar{U} &= \left(\Psi(A)^{-\dagger} \begin{bmatrix} L_r \\ 0_{n-r,r} \end{bmatrix} \right)^\top \left(\Psi(A) \begin{bmatrix} K_r \\ 0_{n-r,r} \end{bmatrix} \right) \\ &= [L_r^\dagger \quad 0] \Psi(A)^{-1} \Psi(A) \begin{bmatrix} K_r \\ 0 \end{bmatrix} = L_r^\dagger K_r. \end{aligned}$$

Because $\det(L_r) \neq 0$ and $\det(K_r) \neq 0$, the matrix $\bar{V}^\top \bar{U}$ is non-singular. ■

According to Lemma 19, the Oja flow can be expressed as a set of coupled equations for a basis and its orthogonal complement:

$$\begin{aligned} \varepsilon \frac{d}{dt} U(t) &= U_\perp(t) U_\perp(t)^\top A U(t), \\ \varepsilon \frac{d}{dt} U_\perp(t) &= -U(t) U(t)^\top A^\top U_\perp(t). \end{aligned}$$

We now relate the two reduced-order matrices, $A_{\bar{U}}$ and $A_{\bar{V}}$.

Lemma 20: The matrices $A_{\bar{U}}$ and $A_{\bar{V}}$ are related by a similarity transformation:

$$A_{\bar{U}} = (\bar{V}^\top \bar{U})^{-1} A_{\bar{V}} (\bar{V}^\top \bar{U}).$$

Consequently, they share the same eigenvalues, which are the r dominant eigenvalues of A .

Proof: From their coordinate representations, we know that $A_{\bar{U}} = K_r^{-1} \Lambda_r K_r$ and $A_{\bar{V}} = L_r^\dagger \Lambda_r L_r^{-\dagger}$. From the proof of Lemma 19, we have $\bar{V}^\top \bar{U} = L_r^\dagger K_r$. A direct substitution shows that the similarity transformation holds. ■

2) Model Reduction and its Properties: The Oja flow extracts the invariant subspace corresponding to the r eigenvalues with the largest real parts. This property renders it a natural tool for identifying and isolating the dominant dynamics of a linear time-invariant (LTI) system, which is the cornerstone of model reduction. Consider the LTI system:

$$\frac{d}{dt} x(t) = Ax(t) + Bu(t), \quad y(t) = Cx(t), \quad (15)$$

where $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $y \in \mathbb{R}^p$ are the state, input, and output, respectively. Throughout the analysis discussed in this

section, we assume the eigenvalue gap condition $\text{Re}(\lambda_r(A)) > \text{Re}(\lambda_{r+1}(A))$ holds for a fixed dimension $r < n$.

Projecting the system (15) onto the subspace spanned by $\bar{U} \in \mathcal{U}_r(A)$ yields a reduced-order model $(A_{\bar{U}}, B_{\bar{U}}, C_{\bar{U}})$. Although the basis $\bar{U}$ is unique only up to an orthogonal transformation, the resulting input-output behavior of the reduced model is unique. This result can be seen by examining the reduced model's impulse response, $C_{\bar{U}} e^{A_{\bar{U}} t} B_{\bar{U}}$. Using the property that for an invariant subspace, $e^{At} \bar{U} = \bar{U} e^{A_{\bar{U}} t}$, this result becomes

$$C_{\bar{U}} e^{A_{\bar{U}} t} B_{\bar{U}} = C \bar{U} e^{A_{\bar{U}} t} \bar{U}^\top B = C e^{At} \bar{U} \bar{U}^\top B,$$

which implies the systems $(A_{\bar{U}}, B_{\bar{U}}, C_{\bar{U}})$ and $(A, \bar{U} \bar{U}^\top B, C)$ are identical. Because the projector $\bar{U} \bar{U}^\top$ is unique for the subspace $\mathcal{U}_r(A)$, the impulse response is independent of the specific choice of basis $\bar{U}$. Because the eigenvalues of $A_{\bar{U}}$ are the r dominant eigenvalues of A , this reduction method preserves stability. Furthermore, it preserves observability.

Proposition 21 ([17, Proposition 5]): If the pair (A, C) is observable, then the reduced pair $(A_{\bar{U}}, C_{\bar{U}})$ is also observable.

By duality, if the pair (A, B) is controllable, then projecting onto the dominant left invariant subspace preserves this property. That is, for $\bar{V} \in \mathcal{U}_r(A^\top)$, the reduced pair $(A_{\bar{V}}, B_{\bar{V}})$ is controllable. The inheritance of these properties thus depends on whether the projection is onto the right or left dominant invariant subspace.

This outcome can be verified using Gramians. Projecting the observability Gramian integral using $\bar{U}$ and the controllability Gramian using $\bar{V}$ for $T > 0$ yields the Gramians of the respective reduced-order systems:

$$\begin{aligned} G_o(\bar{U}, T) &:= \bar{U}^\top \left(\int_0^T e^{A_{\bar{U}}^\top t} C_{\bar{U}}^\top C_{\bar{U}} e^{A_{\bar{U}} t} dt \right) \bar{U} \\ &= \int_0^T e^{A_{\bar{U}}^\top t} C_{\bar{U}}^\top C_{\bar{U}} e^{A_{\bar{U}} t} dt, \end{aligned} \quad (16)$$

$$\begin{aligned} G_c(\bar{V}, T) &:= \bar{V}^\top \left(\int_0^T e^{A_{\bar{V}} t} B_{\bar{V}} B_{\bar{V}}^\top e^{A_{\bar{V}}^\top t} dt \right) \bar{V} \\ &= \int_0^T e^{A_{\bar{V}} t} B_{\bar{V}} B_{\bar{V}}^\top e^{A_{\bar{V}}^\top t} dt. \end{aligned} \quad (17)$$

Positive definiteness of the original Gramians implies positive definiteness of the projected ones, confirming the preservation of observability and controllability. However, these two reduced models, based on $\bar{U}$ and $\bar{V}$, exist in different state-space coordinate systems. To create a single reduced model that preserves both properties, we must relate them. Using the similarity transformation from Lemma 20,

$$\begin{aligned} &(\bar{V}^\top \bar{U})^{-1} G_c(\bar{V}, T) (\bar{V}^\top \bar{U})^{-\top} \\ &= \int_0^T e^{A_{\bar{U}} t} ((\bar{V}^\top \bar{U})^{-1} B_{\bar{V}}) ((\bar{V}^\top \bar{U})^{-1} B_{\bar{V}})^\top e^{A_{\bar{U}}^\top t} dt. \end{aligned}$$

Because $G_c(\bar{V}, T) > 0$ implies $(\bar{V}^\top \bar{U})^{-1} G_c(\bar{V}, T) (\bar{V}^\top \bar{U})^{-\top} > 0$, the pair $(A_{\bar{U}}, B_{\bar{V}(\bar{V}^\top \bar{U})^{-1}})$ is controllable.

Similarly, similarity transformation of $G_o(\bar{U}, T)$ by $(\bar{V}^\top \bar{U})^{-\top}$ shows that if $G_o(\bar{U}, T) > 0$, the pair

$(A_{\bar{V}}, C_{\bar{U}(\bar{V}^\top \bar{U})^{-1}})$ is observable. This argument is concluded in the following result:

Proposition 22: If the original system (A, B, C) is a minimal realization (i.e., controllable and observable), then the following two reduced-order models are also minimal:

- 1) $(A_{\bar{U}}, (\bar{V}^\top \bar{U})^{-1} B_{\bar{V}}, C_{\bar{U}})$
- 2) $(A_{\bar{V}}, B_{\bar{V}}, C_{\bar{U}}(\bar{V}^\top \bar{U})^{-1})$

The two minimal models from Proposition 22 are simply different state-space representations of the same input-output system.

Proposition 23: The two reduced-order models from Proposition 22 have the same transfer function.

Proof: Let $R := \bar{V}^\top \bar{U}$. The transfer function for model 1 is $C_{\bar{U}}(sI_r - A_{\bar{U}})^{-1} R^{-1} B_{\bar{V}}$. Using the identity $A_{\bar{V}} R = R A_{\bar{U}}$ from Lemma 20, we have $(sI_r - A_{\bar{U}})^{-1} R^{-1} = R^{-1}(sI_r - A_{\bar{V}})^{-1}$. Substituting this yields $C_{\bar{U}} R^{-1}(sI_r - A_{\bar{V}})^{-1} B_{\bar{V}}$, which is the transfer function for model 2. ■

Let $P_{\bar{U}}(s)$, $P_{\bar{V}}(s)$, and $P_{rd}(s)$ be the transfer functions of the observability-preserving, controllability-preserving, and minimal reduced models, respectively. They are related by

$$\begin{aligned} P_{rd}(s) &= P_{\bar{U}}(s) + C_{\bar{U}}(sI_r - A_{\bar{U}})^{-1} (\bar{V}^\top \bar{U})^{-1} \bar{V}^\top \bar{U}_\perp B_{\bar{U}_\perp} \\ &= P_{\bar{V}}(s) + C_{\bar{V}_\perp} \bar{V}_\perp^\top \bar{U} (\bar{V}^\top \bar{U})^{-1} (sI_r - A_{\bar{V}})^{-1} B_{\bar{V}}. \end{aligned}$$

This result shows that the minimal model $P_{rd}(s)$ augments the observability-preserving model $P_{\bar{U}}(s)$ with a term that incorporates controllability information from the truncated subspace. A similar expression relates $P_{rd}(s)$ to $P_{\bar{V}}(s)$. These three reduced models are generally distinct, as illustrated in the following example.

Example 24: Recall $A \in \mathbb{R}^{3 \times 3}$ in Example 5 with the following matrices.

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}.$$

The transfer function $P(s)$ of the original system with (A, B, C) and the reduced transfer functions of the model $(A_{\bar{U}}, B_{\bar{U}}, C_{\bar{U}})$ and $(A_{\bar{V}}, B_{\bar{V}}, C_{\bar{V}})$ for $\bar{U} \in \mathcal{U}_2(A)$ and $\bar{V} \in \mathcal{U}_2(A^\top)$ are

$$P(s) = \frac{2s+1}{s^3-s}, \quad P_{\bar{U}}(s) = 0, \quad P_{\bar{V}}(s) = \frac{1}{9} \frac{2s+1}{s(s-1)}.$$

Note that

$$\psi_1(A^\top) = \frac{1}{\sqrt{17}} \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}, \quad \psi_2(A^\top) = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \psi_3(A^\top) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Because the eigenvectors $\psi_1(A)$ and $\psi_2(A)$ denoted in Example 5 are orthogonal to B , $P_{\bar{U}}(s) = 0$ and $P_{\bar{V}}(s) \neq 0$.

Conversely, the reduced transfer function of the model $(A_{\bar{U}}, B_{\bar{V}(\bar{V}^\top \bar{U})^{-1}}, C_{\bar{U}})$ is

$$P_{rd}(s) = \frac{1}{2} \frac{2s+1}{s(s-1)}.$$

In this example, each reduced transfer function is different from the others. Fig. 5 shows the Bode diagrams of $P(s)$, $P_{\bar{V}}(s)$, and $P_{rd}(s)$. $P_{\bar{V}}(s)$ and $P_{rd}(s)$ well approximate $P(s)$

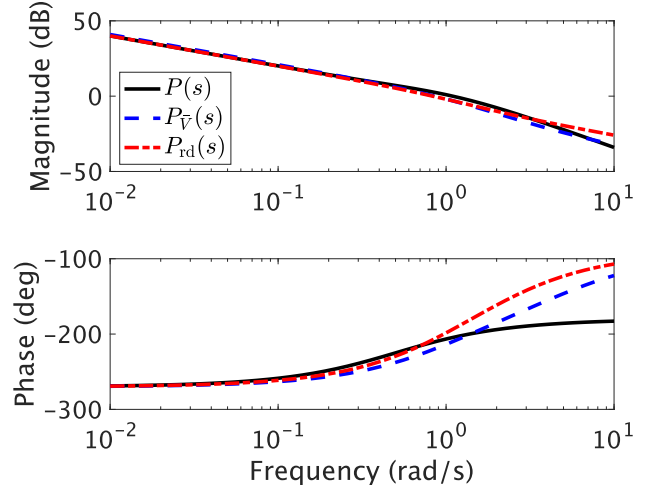


Fig. 5. Bode diagrams of $P(s)$, $P_{\bar{V}}(s)$, and $P_{rd}(s)$.

in the low-frequency range, and the phase differs in the high-frequency range. A detailed analysis of the approximation error will be studied in the future.

Although we have a controllability- and observability-preserving reduced model $(A_{\bar{U}}, B_{\bar{V}(\bar{V}^\top \bar{U})^{-1}}, C_{\bar{U}})$, as mentioned in Example 24, it is unclear whether the reduction is more effective than the other ones. For example, as demonstrated here, the others have intuitive approximation errors:

Proposition 25: Consider a system with (A, B, C) . Then, for $\bar{U} \in \mathcal{U}_r(A)$ and $\bar{V} \in \mathcal{U}_r(A^\top)$, the transfer function is

$$\begin{aligned} P(s) &= P_{\bar{U}}(s) + C_{\bar{U}_\perp}(sI_{n-r} - A_{\bar{U}_\perp})^{-1} B_{\bar{U}_\perp} \\ &\quad + C_{\bar{U}}(sI_r - A_{\bar{U}})^{-1} \bar{U}^\top A \bar{U}_\perp (sI_{n-r} - A_{\bar{U}_\perp})^{-1} B_{\bar{U}_\perp} \\ &= P_{\bar{V}}(s) + C_{\bar{V}_\perp}(sI_{n-r} - A_{\bar{V}_\perp})^{-1} B_{\bar{V}_\perp} \\ &\quad + C_{\bar{V}_\perp}(sI_{n-r} - A_{\bar{V}_\perp})^{-1} \bar{V}_\perp^\top A \bar{V} (sI_r - A_{\bar{V}})^{-1} B_{\bar{V}}, \end{aligned}$$

where $\bar{U}_\perp \in \mathcal{U}_{n-r}(-A^\top)$ and $\bar{V}_\perp \in \mathcal{U}_{n-r}(-A)$.

Recall that $\bar{U}_\perp \bar{U}_\perp^\top = I_n - \bar{U} \bar{U}^\top$ and $\bar{V}_\perp \bar{V}_\perp^\top = I_n - \bar{V} \bar{V}^\top$.

Proof: Using (13),

$$\begin{aligned} P(s) &= C Q_{\bar{U}}(sI_n - Q_{\bar{U}}^\top A Q_{\bar{U}})^{-1} Q_{\bar{U}}^\top B \\ &= \begin{bmatrix} C_{\bar{U}}^\top \\ C_{\bar{U}_\perp}^\top \end{bmatrix}^\top \begin{bmatrix} (sI_r - A_{\bar{U}})^{-1} & P_{ur}(s) \\ 0_{n-r,r} & (sI_{n-r} - A_{\bar{U}_\perp})^{-1} \end{bmatrix} \begin{bmatrix} B_{\bar{U}} \\ B_{\bar{U}_\perp} \end{bmatrix}, \end{aligned}$$

where the upper right block is $P_{ur}(s) := (sI_r - A_{\bar{U}})^{-1} \bar{U}^\top A \bar{U}_\perp (sI_{n-r} - A_{\bar{U}_\perp})^{-1}$. Hence,

$$\begin{aligned} P(s) &= C_{\bar{U}}(sI_r - A_{\bar{U}})^{-1} B_{\bar{U}} + C_{\bar{U}_\perp}(sI_{n-r} - A_{\bar{U}_\perp})^{-1} B_{\bar{U}_\perp} \\ &\quad + C_{\bar{U}}(sI_r - A_{\bar{U}})^{-1} \bar{U}^\top A \bar{U}_\perp (sI_{n-r} - A_{\bar{U}_\perp})^{-1} B_{\bar{U}_\perp}. \end{aligned}$$

Similarly,

$$\begin{aligned} P(s) &= C_{\bar{V}}(sI_r - A_{\bar{V}})^{-1} B_{\bar{V}} + C_{\bar{V}_\perp}(sI_{n-r} - A_{\bar{V}_\perp})^{-1} B_{\bar{V}_\perp} \\ &\quad + C_{\bar{V}_\perp}(sI_{n-r} - A_{\bar{V}_\perp})^{-1} \bar{V}_\perp^\top A \bar{V} (sI_r - A_{\bar{V}})^{-1} B_{\bar{V}} \end{aligned}$$

holds. ■

Fig. 6 shows the block diagram of $P(s)$ and $P_{\bar{U}}(s)$ in Proposition 25. If one considers the approximation error $P(s) - P_{rd}(s)$, it is also calculated from the relationship

C. Singularly Perturbed Linear Systems

Consider a stable LTI system whose dynamics exhibit a two-time-scale property. This system is characterized by a Hurwitz matrix A whose eigenvalues are separated into a group of r slow and $n - r$ fast eigenvalues, such that there is a significant gap between them: $|\operatorname{Re}(\lambda_r(A))| \ll |\operatorname{Re}(\lambda_{r+1}(A))|$. The Oja flow provides a natural method for separating these modes. The subspace spanned by $\bar{U} \in \mathcal{U}_r(A)$ corresponds to the slow dynamics, whereas the orthogonal complement, spanned by $\bar{U}_\perp$, corresponds to the fast dynamics.

This separation is the foundation of the singular perturbation method, which decouples a system to derive a simplified model of the slow dynamics [54]. Applying this idea to system (15), we define the slow states as $x_s(t) := \bar{U}^\top x(t)$ and the fast states as $x_f(t) := \bar{U}_\perp^\top x(t)$. With the block-triangular decomposition from (13), the state equations become

$$\begin{aligned} \frac{d}{dt} x_s(t) &= A_{\bar{U}} x_s(t) + (\bar{U}^\top A \bar{U}_\perp) x_f(t) + B_{\bar{U}} u(t), \\ \frac{d}{dt} x_f(t) &= A_{\bar{U}_\perp} x_f(t) + B_{\bar{U}_\perp} u(t). \end{aligned}$$

The core assumption of singular perturbation is that the fast dynamics reach a quasi-steady state much more rapidly than the slow dynamics evolve, provided the input $u(t)$ also varies slowly. By setting $\frac{d}{dt} x_f(t) \approx 0$, we can solve for the quasi-steady-state value of the fast mode: $x_f(t) \approx -A_{\bar{U}_\perp}^{-1} B_{\bar{U}_\perp} u(t)$. Substituting this back into the equation for the slow dynamics yields the reduced-order model:

$$\frac{d}{dt} x_s(t) = A_{\bar{U}} x_s(t) + \left(B_{\bar{U}} - (\bar{U}^\top A \bar{U}_\perp) A_{\bar{U}_\perp}^{-1} B_{\bar{U}_\perp} \right) u(t).$$

Although the singular perturbation method is powerful, it can be challenging to identify the transformation that separates the slow and fast modes for large and complex systems. The Oja flow offers a significant advantage by providing a systematic and computationally tractable way to find this decomposition automatically.

V. CONCLUSION

This paper presents a global convergence analysis of Oja's component flow for extracting the r -dominant invariant subspace of a general square matrix A , which corresponds to the r eigenvalues with the largest real parts. The stable invariant set, $\mathcal{U}_r$, and its domain of attraction were characterized. Crucially, the algorithm was shown to ensure exponential convergence for almost all initial conditions, providing a strong theoretical guarantee for its reliability. We also investigated the tracking performance for time-varying matrices and provided numerical examples validating the theoretical results.

The primary implication of these findings is that the Oja flow is a theoretically sound and practically viable tool for analyzing dominant dynamics in a much broader class of linear systems than previously considered, extending beyond symmetric positive-definite matrices. The applications presented demonstrate this utility. We propose methods for model reduction and low-rank controller synthesis that leverage the Oja flow's ability to isolate key subspaces. The properties of the resulting reduced-order models, including preservation

of observability, controllability, and stabilization using output feedback are discussed. Furthermore, we demonstrated the flow's applicability to mode identification in singularly perturbed systems by exploiting its capacity to separate slow and fast dynamics. These concepts offer promising extensions to time-varying and discrete-time systems.

Future research directions include: 1) accelerating the convergence rate for enhanced practical performance; 2) establishing rigorous analytical guarantees for time-varying matrices, essential for model reduction of more general systems; and 3) further characterizing the properties of the derived reduced-order models. Regarding the time-varying case, although applications to state estimation exist [36], performance is sensitive to the parameter ε , often requiring small integration steps and increasing computational cost. A more detailed analysis of tracking error is therefore required, alongside broader exploration of estimation and control applications. Finally, refining the understanding of the reduced-model properties remains a challenging yet important problem because of its direct impact on practical controller design and performance guarantees.

REFERENCES

- [1] Bruce C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control*, 26(1):17–32, 1981.
- [2] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer New York, 2006.
- [3] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.
- [4] Elad Plaut. From Principal Subspaces to Principal Components with Linear Autoencoders. *Preprint arXiv:1804.10253v3*, April 2018.
- [5] Edmond Cunningham, Adam D. Cobb, and Susmit Jha. Principal component flows. In *International Conference on Machine Learning*, pages 4492–4519. PMLR, 2022.
- [6] John Wright and Yi Ma. *High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications*. Cambridge University Press, 2022.
- [7] Martin J. Wainwright. *High-Dimensional Statistics: A non-Asymptotic Viewpoint*, volume 48. Cambridge University Press, 2019.
- [8] Toshihisa Tanaka. Generalized weighted rules for principal components tracking. *IEEE Transactions on Signal Processing*, 53(4):1243–1253, 2005.
- [9] Dezhong Peng and Zhang Yi. Convergence analysis of a deterministic discrete time system of Feng's MCA learning algorithm. *IEEE Transactions on Signal Processing*, 54(9):3626–3632, 2006.
- [10] Mohammed A. Hasan. Stability analysis of dynamical systems for minor and principal component analysis. In *2006 American Control Conference*, pages 6–pp. IEEE, 2006.
- [11] Jonathan H. Manton, Uwe Helmke, and Iven M. Y. Mareels. A dual purpose principal and minor component flow. *Systems & Control Letters*, 54(8):759–769, 2005.
- [12] V. Borkar and Sean P. Meyn. Oja's algorithm for graph clustering, Markov spectral decomposition, and risk sensitive control. *Automatica*, 48(10):2512–2519, 2012.
- [13] Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Review*, 57(4):483–531, 2015.
- [14] Sivere Bonnabel and Rodolphe Sepulchre. The geometry of low-rank Kalman filters. In *Matrix Information Geometry*, pages 53–68. Springer, 2012.
- [15] Shuto Yamada and Kentaro Ohki. On a New Low-Rank Kalman-Bucy Filter and Its Convergence Property. In *The 52nd ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS '20)*, volume 2021, pages 16–20, Osaka, 2020.

- [16] Shuto Yamada and Kentaro Ohki. Comparison of Estimation Error between Two Different Low-Rank Kalman-Bucy Filters. In *2021 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, September 2021.
- [17] Daiki Tsuzuki and Kentaro Ohki. Low-rank approximated Kalman-Bucy filters using Oja's principal component flow for linear time-invariant systems. *IEEE Control Systems Letters*, 8:1583–1588, 2024.
- [18] Daiki Tsuzuki and Kentaro Ohki. Low-rank approximated Kalman filter using Oja's principal component flow for discrete-time linear systems. In *2024 SICE Festival with Annual Conference (SICE FES)*, pages 242–247, July 2024.
- [19] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.
- [20] Erkki Oja and Juha Karhunen. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. *Journal of Mathematical Analysis and Applications*, 106(1):69–84, 1985.
- [21] Erkki Oja. Neural networks, principal components, and subspaces. *International Journal of Neural Systems*, 1(01):61–68, 1989.
- [22] Tianping Chen, Yingbo Hua, and Wei-Yong Yan. Global convergence of Oja's subspace algorithm for principal component extraction. *IEEE Transactions on Neural Networks*, 9(1):58–67, 1998.
- [23] Tianping Chen and Shun-ichi Amari. Unified stabilization approach to principal and minor components extraction algorithms. *Neural Networks*, 14(10):1377–1387, 2001.
- [24] Shintaro Yoshizawa, Uwe Helmke, and Konstantin Starkov. Convergence analysis for principal component flows. *International Journal of Applied Mathematics and Computer Science*, 11(1):223–236, 2001.
- [25] Yueh-Cheng Kuo, Huey-Er Lin, and Shih-Feng Shieh. Asymptotic analysis of generalized orthogonal flows. *Linear Algebra and its Applications*, 717(15):1–25, 2025.
- [26] Wei-Yong Yan, Uwe Helmke, and John B. Moore. Global analysis of Oja's flow for neural networks. *IEEE Transactions on Neural Networks*, 5(5):674–683, 1994.
- [27] Tianping Chen, Shun Ichi Amari, and Qin Lin. A unified algorithm for principal and minor components extraction. *Neural Networks*, 11(3):385–390, 1998.
- [28] Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *Advances in Neural Information Processing Systems*, 33:2503–2515, 2020.
- [29] Silvere Bonnabel, Marc Lambert, and Francis Bach. Low-rank plus diagonal approximations for Riccati-like matrix differential equations. *SIAM Journal on Matrix Analysis and Applications*, 45(3):1669–1688, 2024.
- [30] Yueh-Cheng Kuo, Huey-Er Lin, and Shih-Feng Shieh. The orthogonal flows for orthogonal iteration. *Linear Algebra and its Applications*, 679(15):67–85, 2023.
- [31] Jason Frank and Sergiy Zhuk. A detectability criterion and data assimilation for nonlinear differential equations. *Nonlinearity*, 31(11):5235, 2018.
- [32] Markus Tranninger, Richard Seeber, Sergiy Zhuk, Martin Steinberger, and Martin Horn. Detectability analysis and observer design for linear time varying systems. *IEEE Control Systems Letters*, 4(2):331–336, 2019.
- [33] Luca Dieci, Robert D. Russell, and Erik S. Van Vleck. On the computation of Lyapunov exponents for continuous dynamical systems. *SIAM Journal on Numerical Analysis*, 34(1):402–423, 1997.
- [34] Luca Dieci and Erik S. Van Vleck. Lyapunov and Sacker–Sell spectral intervals. *Journal of Dynamics and Differential Equations*, 19:265–293, 2007.
- [35] Thomas J. Bridges and Sebastian Reich. Computing Lyapunov exponents on a Stiefel manifold. *Physica D: Nonlinear Phenomena*, 156(3–4):219–238, 2001.
- [36] Markus Tranninger, Richard Seeber, Martin Steinberger, Martin Horn, and Christian Pötzsche. Detectability Conditions and State Estimation for Linear Time-Varying and Nonlinear Systems. *SIAM Journal on Control and Optimization*, 60(4):2514–2537, 2022.
- [37] Tetsushi Sasagawa. On the finite escape phenomena for matrix Riccati equations. *IEEE Transactions on Automatic Control*, 27(4):977–979, 1982.
- [38] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, 4 edition, 2012.
- [39] Andreas Weingessel and Kurt Hornik. SVD algorithms: APEX-like versus subspace methods. *Neural Processing Letters*, 5:177–184, 1997.
- [40] Simone Fiori. Singular value decomposition learning on double stiefel manifold. *International Journal of Neural Systems*, 13(03):155–170, 2003.
- [41] James R. Munkres. *Topology*. Peason Education, 2 edition, 2018.
- [42] Yasuko Chikuse. *Statistics on Special Manifolds*, volume 174 of *Lecture Notes in Statistics*. Springer New York, 2003.
- [43] Goro Obinata and Brian D. O. Anderson. *Model Reduction for Control System Design*. Communication and Control Engineering. Springer, 2000.
- [44] A. C. Antoulas. An overview of approximation methods for large-scale dynamical systems. *Annual Reviews in Control*, 29(2):181–190, 2005.
- [45] Thomas J. Snowden, Piet H. van der Graaf, and Marcus J. Tindall. Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends. *Bulletin of Mathematical Biology*, 79(7):1449–1486, June 2017.
- [46] A. C. Antoulas. *Approximation of Large-Scale Dynamical Systems*. Society for Industrial Mathematics, 2005.
- [47] Serkan Gugercin, Athanasios C. Antoulas, and Christopher Beattie. $\mathcal{H}_2$ model reduction for large-scale linear dynamical systems. *SIAM Journal on Matrix Analysis and Applications*, 30(2):609–638, 2008.
- [48] Kazuhiro Sato and Hiroyuki Sato. Structure-Preserving H^2 Optimal Model Reduction Based on the Riemannian Trust-Region Method. *IEEE Transactions on Automatic Control*, 63(2):505–512, 2018.
- [49] Kemin Zhou, Gregory Salomon, and Eva Wu. Balanced realization and model reduction for unstable systems. *International Journal of Robust and Nonlinear Control*, 9(3):183–198, 1999.
- [50] P. H. Opdenacker and Edmond A. Jonckheere. LQG balancing and reduced LQG compensation of symmetric passive systems. *International Journal of Control*, 41(1):73–109, 1985.
- [51] Pawan Goyal and Martin Redmann. Time-limited $\mathcal{H}^2$ -optimal model order reduction. *Applied Mathematics and Computation*, 355:184–197, 2019.
- [52] Kasturi Das, Srinivasan Krishnaswamy, and Somanath Majhi. H_2 Optimal Model Order Reduction Over a Finite Time Interval. *IEEE Control Systems Letters*, 6:2467–2472, 2022.
- [53] S. Kung and D. Lin. Optimal Hankel-norm model reductions: Multivariable systems. *IEEE Transactions on Automatic Control*, 26(4):832–852, 1981.
- [54] Petar V. Kokotovic and P. Sannuti. Singular Perturbation Method for Reducing the Model Order in Optimal Control Design. *IEEE Transactions on Automatic Control*, 13(4):377–384, 1968.