University of Reading

Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution

Article

Accepted Version

Gurumoorthy, K. S., Grudzien, C., Apte, A., Carrassi, A. and Jones, C. K. R. T. (2017) Rank deficiency of Kalman error covariance matrices in linear time-varying system with deterministic evolution. SIAM Journal on Control and Optimization, 55 (2). pp. 741-759. ISSN 0363-0129 doi: https://doi.org/10.1137/15M1025839 Available at http://centaur.reading.ac.uk/90354/

Published version at: http://dx.doi.org/10.1137/15M1025839

To link to this article DOI: http://dx.doi.org/10.1137/15M1025839

Publisher: Society for Industrial and Applied Mathematics

SIAM J. CONTROL OPTIM.
Vol. 0, No. 0, pp. 000–000

RANK DEFICIENCY OF KALMAN ERROR COVARIANCE MATRICES IN LINEAR TIME-VARYING SYSTEM WITH DETERMINISTIC EVOLUTION*

KARTHIK S. GURUMOORTHY[†], COLIN GRUDZIEN[‡], AMIT APTE[†], ALBERTO CARRASSI[§], AND CHRISTOPHER K. R. T. JONES[¶]

Abstract. We prove that for-linear, discrete, time-varying, deterministic system (perfect-model) with noisy outputs, the Riccati transformation in the Kalman filter asymptotically bounds the rank of the forecast and the analysis error covariance matrices to be less than or equal to the number of nonnegative Lyapunov exponents of the system. Further, the support of these error covariance matrices is shown to be confined to the space spanned by the unstable-neutral backward Lyapunov vectors, providing the theoretical justification for the methodology of the algorithms that perform assimilation only in the unstable-neutral subspace. The equivalent property of the autonomous system is investigated as a special case.

Key words. Kalman filter, data assimilation, linear dynamics, control theory, covariance matrix, rank

AMS subject classifications. 93E11, 93C05, 93B05, 60G35, 15A03

DOI. 10.1137/15M1025839

1. Introduction.

The problem of estimating the state of an evolving system from an incomplete set of noisy observations is the central theme of the state estimation and optimal control theory [7], also referred to as data assimilation (DA) in geosciences [6, 20]. In the filtering procedure, based on the concept of recursive processing, measurements are utilized sequentially, as they become available [7]. For linear dynamics, and when a linear relation exists between measurements and the state variables, and when the errors associated to all sources of information are Gaussian, the solution can be expressed via the Kalman filter (KF) equations [8]. The KF provides a closed set of equations for the first two moments of the posterior probability density function of the system state, conditioned on the observations. In the case of nonlinear dynamics, the first order extension of the KF is known as the extended Kalman filter (EKF) [7], whereas a Monte Carlo approximation is the basis of a set of methods known as the ensemble Kalman filter, both of which have been studied extensively in geophysical contexts [13, 5].

Atmosphere and ocean are example of dissipative chaotic systems. This implies sensitivity to the initial condition [11] and the fact that the estimation error strongly

*Received by the editors June 15, 2015; accepted for publication (in revised form) November 23, 2016; published electronically DATE.
http://www.siam.org/journals/sicon/x-x/M102583.html
Funding: This work received support from the AIRBUS Group Corporate Foundation Chair in Mathematics of Complex Systems established in ICTS-TIFR. This work was partially funded by the project REDDA of the Norwegian Research Council under contract 250711. This work was supported by EU-FP7 project SANGOMA under grant contract 283580 and by funding from the Centre of Excellence EmblA of the Nordic Countries Research Council, NordForsk.
[†]International Center for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore, Karnataka, India (karthik.gurumoorthy@icts.res.in, apte@icts.res.in).
[‡]Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599, and Nansen Environmental and Remote Sensing Center, Bergen, Norway (cgrudz@email.unc.edu).
[§]Nansen Environmental and Remote Sensing Center, Bergen, Norway (alberto.carrassi@nersc.no).
[¶]Department of Mathematics, University of North Carolina, Chapel Hill, NC 27599 (ckrtj@email.unc.edu).

projects on the unstable manifold of the dynamics [18], which has inspired the development of a class of algorithms known as assimilation in the unstable subspace (AUS) [23]. In AUS, the span of the leading Lyapunov vectors (to be defined precisely in later sections), or a suitable approximation of this span, is used explicitly in the analysis step: the analysis update is confined to the unstable subspace [16]. AUS has been formalized in the framework of the EKF, the EKF-AUS [22], and the variational (smoothing) procedure, 4DVar-AUS [21]. Applications with atmospheric, oceanic, and traffic models [24, 3, 17] showed that even in high dimensional systems, an efficient error control is achieved by monitoring only a limited number of unstable directions, making AUS a computationally efficient alternative to standard procedures. The AUS methodology is based on and at the same time hints at a fundamental observation: the span of the estimation error covariance matrices asymptotically (in time) tends to the subspace spanned by the unstable-neutral Lyapunov vectors.

The search for a formal proof of this aforesaid property is the basic motivation of the present work, which is focused on linear nonautonomous and linear autonomous perfect-model dynamical systems. The main results of the paper are as follows. In Theorem 3.5 we show that the error covariance matrices, independent of the initial condition, asymptotically become rank deficient in time, and then in Theorem 3.7 we characterize their null spaces by proving that the restriction of the these matrices onto the stable backward Lyapunov vectors converges to zero in time. When restricted to the linear, autonomous system with the time invariant propagator A, we establish that the stable space of the time independent backward Lyapunov vectors equals the stable space of A^T—span of generalized eigen-vectors of A^T corresponding to eigen-values less than one in absolute magnitude—in Theorem B.3. Consequently, in Corollary 4.2 we show that the null space of the error covariance matrices contain the stable space of A^T asymptotically.

The paper is organized as follows. After describing the general notation in section 2, the nonautonomous case is considered in section 3. The assumptions used in proving our main result, other useful results such as the Oseledets theorem, and the concepts of observability and controllability for noiseless systems are described in sections 3.1, 3.2, and 3.3. Theorem 3.5 discussing the rank deficiency of error covariance matrices is presented in section 3.4 and the proof of Theorem 3.7 using the geometric viewpoint of Kalman filtering [2, 25, 1] is detailed in section 3.5. Section 3.6 presents some numerical results buttressing the theorem. Section 4 includes the proof of Corollary 4.2 along with a numerical illustration supporting the analytical findings for autonomous systems. We conclude in section 5.

Although the extension of these results to the general nonlinear case is the object of active research [19], the current findings already provide a formal justification to the AUS foundation and further motivate its use as a DA strategy in nonlinear chaotic dynamics.

2. Notation. The dimension of the state space is represented by d. For any square matrix $Z \in \mathbb{C}^{d\times d}$ let the set $\{\lambda_1(Z), \dots, \lambda_d(Z)\}$ represent the eigen-values of Z, where $|\lambda_1(Z)| \geq \cdots \geq |\lambda_d(Z)|$. Similarly, let the set $\{\sigma_1(Z), \dots, \sigma_d(Z)\}$ stand for the singular values of Z with $\sigma_1(Z) \geq \cdots \geq \sigma_d(Z)$. We define the column vectors of the matrix $V_Z = [\mathbf{v}_1(Z), \dots, \mathbf{v}_d(Z)]$ to be the generalized eigen-vectors of Z of satisfying the relation $ZV_Z = V_Z J(Z)$, where $J(Z)$ is the Jordan canonical form of Z. In the event that Z is diagonalizable ($J(Z)$ is diagonal), let the entries of the diagonal matrix $\Lambda_Z = J(Z)$ symbolize the eigen-values of Z and the columns of V_Z—the eigen-vectors—be of unit magnitude. Z^* denotes the adjoint of Z for the scalar product under consideration in $\mathbb{C}^d$ and $Z^\dagger$ represents the conjugate transpose of Z. For the canonical scalar product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^\dagger \mathbf{v}$ in $\mathbb{C}^d$, $Z^* = Z^\dagger$, and when confined to the real space $\mathbb{R}^d$ where $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \mathbf{v}$, $Z^* = Z^T$. Unless explicitly stated we assume a real vector space endowed with a canonical scalar product. The matrix norm $\|Z\|$ we consider is the largest singular value $\sigma_1(Z)$ of Z. The notation $Z > 0$ ($Z \geq 0$) is used when Z is symmetric positive-definite (positive-semidefinite). For any two symmetric matrices Z_1, Z_2, the notation $Z_1 \geq Z_2$ means $Z_1 - Z_2 \geq 0$. The following definitions are useful.

DEFINITION 2.1 (real span). *The real span of a complex vector* $\mathbf{w} = \mathbf{u} + i\mathbf{v}$ *where* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ *is the vector space* $\mathcal{T}_\mathbf{w} \subset \mathbb{R}^d$ *defined as*

$$\mathcal{T}_\mathbf{w} \equiv \{\alpha \mathbf{u} + \beta \mathbf{v} : \alpha, \beta \in \mathbb{R}\}.$$

DEFINITION 2.2 (α-eigenspace). *Given* $\alpha > 0$*, the* α*-eigenspace of a square matrix* Z *denoted by* $\mathcal{E}^\alpha(Z)$ *is the real span of the generalized eigen-vectors of* Z *corresponding to eigen-values* λ *with* $|\lambda| < \alpha$.

3. Nonautonomous systems.

3.1. Setup and assumptions.

We define the general linear nonautonomous dynamical system at time $n \geq 0$ by

$$\begin{aligned} \mathbf{x}_{n+1} &= A_{n+1}\mathbf{x}_n + F_{n+1}\mathbf{p}_{n+1}, \\ \mathbf{y}_{n+1} &= H_{n+1}\mathbf{x}_{n+1} + \mathbf{q}_{n+1}, \end{aligned} \tag{3.1}$$

where $\mathbf{x}_n \in \mathbb{R}^d$, $\mathbf{q}_n \in \mathbb{R}^q$, $\mathbf{p}_n \in \mathbb{R}^p$. The $\mathbf{x}_n$ are the state variables, $\mathbf{p}_n$ represents model noise, $\mathbf{y}_n$ represents observational variables, and $\mathbf{q}_n$ is the observational noise term. The basic random variables $\{\mathbf{x}_0, \mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{p}_1, \mathbf{p}_2, \ldots\}$ are all assumed to be independent and Gaussian with

$$\mathbf{x}_0 \sim \mathcal{N}\left(\mathbf{x}_{0|0}, \Delta_0\right), \quad \mathbf{q}_n \sim \mathcal{N}(0, Q_n), \quad \mathbf{p}_n \sim \mathcal{N}(0, I)$$

such that $\Delta_0 \in \mathbb{R}^{d \times d}$ is the initial error covariance matrix of the state variable $\mathbf{x}_0$, $Q_n \in \mathbb{R}^{q \times q}$ is the observation error covariance matrix at time n, and $F_n \in \mathbb{R}^{d \times p}$. The matrices $\Delta_0, Q_n, F_n, A_n, H_n$ are known for all time n. Further, A_n and Q_n are considered to be nonsingular, $\|A_n\| \leq c_A$, $\|Q_n\| \leq c_Q$, and $\|H_n\| \leq c_H \; \forall n \geq 1$, where c_A, c_Q, and c_H are positive constants. The model noise error covariance is given by $P_n \equiv F_n F_n^T$. Unless explicitly stated $\Delta_0 > 0$, i.e., its eigen-values are strictly positive.

Filtering theory deals with the properties of the conditional distribution, called the *analysis* in the context of DA, of the state $\mathbf{x}_n$ at time n conditioned on observations $Y_{0:n} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n]$ up to time n where the first observation $\mathbf{y}_1$ is assumed to occur at time $n = 1$. This conditional distribution provides an optimal state estimate in the least squares sense [7]. Under the assumptions of linearity and Gaussianity stated above, this conditional distribution is Gaussian, with mean and covariance denoted by $\mathbf{x}_{n|n}$, and Δ_n respectively:

$$\mathbf{x}_{n|n} = \mathbb{E}[\mathbf{x}_n \mid Y_{0:n}] \quad \text{and} \quad \Delta_n = \mathbb{E}[(\mathbf{x}_n - \mathbf{x}_{n|n})(\mathbf{x}_n - \mathbf{x}_{n|n})^T \mid Y_{0:n}]\,.$$

We also note that the conditional distribution, called the *forecast* in DA literature, of the state $\mathbf{x}_{n+1}$ conditioned on observations $Y_{0:n}$ up to time n is Gaussian with its mean and covariance denoted by $\mathbf{x}_{n+1|n}$ and Σ_{n+1}, respectively:

$$\mathbf{x}_{n+1|n} = \mathbb{E}[\mathbf{x}_{n+1} \mid Y_{0:n}] \quad \text{and} \quad \Sigma_{n+1} = \mathbb{E}[(\mathbf{x}_{n+1} - \mathbf{x}_{n+1|n})(\mathbf{x}_{n+1} - \mathbf{x}_{n+1|n})^T \mid Y_{0:n}]\,.$$

In this work we concern ourselves with systems that have no model error, i.e., $F_n \equiv 0 \ \forall n \geq 1$, and investigate the dynamics

$$\mathbf{x}_{n+1} = A_{n+1}\mathbf{x}_n \quad \text{and} \quad \mathbf{y}_{n+1} = H_{n+1}\mathbf{x}_{n+1} + \mathbf{q}_{n+1}\,. \tag{3.2}$$

We will be interested in asymptotic properties of the conditional error covariances Σ_n and Δ_n. The KF provides a closed form, iterative formula for obtaining these quantities [7]. Under the assumption of no model noise, the update equation for the forecast error covariance is

$$\Sigma_n = A_n \Delta_{n-1} A_n^T\,. \tag{3.3}$$

By defining the Kalman gain matrix K_n as

$$K_n \equiv \Sigma_n H_n^T \left[H_n \Sigma_n H_n^T + Q_n\right]^{-1}\,, \tag{3.4}$$

the analysis error covariance equals

$$\Delta_n = (I - K_n H_n)\Sigma_n\,. \tag{3.5}$$

The update equations for the means are given by

$$\mathbf{x}_{n+1|n} = A_{n+1}\mathbf{x}_{n|n}, \tag{3.6}$$

$$\mathbf{x}_{n+1|n+1} = \mathbf{x}_{n+1|n} + K_{n+1}\left(y_{n+1} - H_{n+1}\mathbf{x}_{n+1|n}\right). \tag{3.7}$$

Defining the sequence of matrices M_n as

$$M_1 \equiv (I - K_1 H_1)A_1, \quad M_n \equiv (I - K_n H_n) A_n M_{n-1} \tag{3.8}$$

and writing the propagator $B_{m:m+n}$ from time m to time $m+n$ by

$$B_{m:m+n} \equiv A_{m+n}A_{m+n-1}\cdots A_{m+1}, \tag{3.9}$$

the analysis covariance at time n can be expressed as

$$\Delta_n = (I - K_n H_n)A_n \cdots (I - K_1 H_1)A_1 \Delta_0 A_1^T \cdots A_n^T = M_n \Delta_0 B_{0:n}^T. \tag{3.10}$$

This equation clearly shows that the asymptotic properties of Δ_n are closely related to those of $B_{0:n}$ and M_n. The notation in (3.10) is suggestive of the line of argument we will take in the following sections. To outline, we may consider the singular-value decomposition of the propagator $B_{0:n}^T = V_n S_n U_n^T$ and decompose the error covariances into a basis of the left singular vectors. In particular, we know that this decomposition may be written as a function of the singular values, provided we have an appropriate bound on M_n in (3.10). Moreover, the left singular vectors of the propagator $B_{0:n}$ will become arbitrarily close to the backward Lyapunov vectors of the system.

The properties of $B_{0:n}$ are basically determined by the dynamical system and are discussed in the next section, while those of M_n are commonly discussed in the context of control theory and are discussed in section 3.3, where we prove a useful bound on its matrix norm in Lemma 3.3.

3.2. Oseledets theorem. Note that the boundedness condition on A_n implies the bound $\|B_{0:n}\| \le (c_A)^n \,\forall n$. Then the Oseledets multiplicative ergodic theorem in [15] states that for each nonzero vector $\mathbf{u} \in \mathbb{R}^d$ the limit

$$\mu = \lim_{n\to\infty} \frac{1}{n} \log \frac{\|B_{0:n}\mathbf{u}\|}{\|\mathbf{u}\|}$$

exists and assumes up to d distinct values $\mu_1 \ge \cdots \ge \mu_d$ which are called the Lyapunov exponents. We will assume

$$0 > \mu_{d_0+1} \tag{3.11}$$

so that exactly $d_0 < d$ of the Lyapunov exponents are nonnegative. Further, defining the matrices

$$E_n^b(m) \equiv [B_{m-n:m}(B_{m-n:m})^*]^{\frac{1}{2n}}\,, \quad E_n^f(m) \equiv [(B_{m:m+n})^* B_{m:m+n}]^{\frac{1}{2n}}\,, \tag{3.12}$$

the Oseledets theorem guarantees that the following limits exist, namely,

$$E^b(m) \equiv \lim_{n\to\infty} E_n^b(m), \tag{3.13}$$

$$E^f(m) \equiv \lim_{n\to\infty} E_n^f(m). \tag{3.14}$$

The eigen-vectors of $E^b(m)$ and $E^f(m)$ represented as the column vectors of $L^b(m) = [\mathbf{l}_1^b(m), \ldots, \mathbf{l}_d^b(m)]$ and $L^f(m) = [\mathbf{l}_1^f(m), \ldots, \mathbf{l}_d^f(m)]$, respectively, are defined as the *backward* and the *forward* Lyapunov vectors at time m [10]. We note that the asymptotic results in later sections will essentially use the backward Lyapunov vectors $L^b(m)$.

The convergence of the individual matrix entries in (3.13) and (3.14) guarantee the convergence of their characteristic polynomials—whose coefficients are well-defined functions of the matrix entries—the roots of which are the eigen-values. Therefore,

$$\lim_{n\to\infty} \Lambda_{E_n^b(m)} = \Lambda_{E^b(m)}, \quad \lim_{n\to\infty} \Lambda_{E_n^f(m)} = \Lambda_{E^f(m)},$$

where we recall that Λ_Z is a diagonal matrix comprising eigen-values of Z. Using the notation from section 2 we additionally find

$$\begin{aligned} \|\lambda_j\left(E^b(m)\right)\mathbf{v}_j\left(E_n^b(m)\right) - E^b(m)\mathbf{v}_j\left(E_n^b(m)\right)\| \le & \left|\lambda_j\left(E^b(m)\right) - \lambda_j\left(E_n^b(m)\right)\right| \\ & + \|E_n^b(m) - E^b(m)\| \end{aligned}$$

from which we can infer that

$$\lim_{n\to\infty} \|\lambda_j\left(E^b(m)\right)\mathbf{v}_j\left(E_n^b(m)\right) - E^b(m)\mathbf{v}_j\left(E_n^b(m)\right)\| = 0$$

leading to $\lim_{n\to\infty} V_{E_n^b(m)} = V_{E^b(m)} = L^b(m)$. Similarly, $\lim_{n\to\infty} V_{E_n^f(m)} = V_{E^f(m)} = L^f(m)$.

The Oseledets theorem also asserts the eigen-values of $E^b(m)$ or $E^f(m)$ do not depend on the initial time m, are the same for the forward and backward matrices, and relate to the Lyapunov exponents as

$$\mu_j = \log(\lambda_j(E)), \quad j \in \{1, \ldots, d\}, \tag{3.15}$$

where we deliberately drop the index m and the superscript b or f on E. However, the forward and backward Lyapunov vectors are different from each other and they also depend on the time m, i.e., $L^b(k) \neq L^b(m) \neq L^f(m) \neq L^f(k)$ for $k \neq m$.

Consider the singular-value decomposition $B_{0:n} \equiv U_n S_n (V_n)^T$ so that under the canonical inner product

$$E_n^f(0) = \left[(B_{0:n})^T B_{0:n}\right]^{\frac{1}{2n}} = \left[V_n (S_n)^2 (V_n)^T\right]^{\frac{1}{2n}} = V_n (S_n)^{\frac{1}{n}} (V_n)^T \,,$$

implying $V_{E_n^f(0)} = V_n$ and

$$\lim_{n\to\infty} \|\mathbf{v}_{j,n} - \mathbf{l}_j^f(0)\| = 0, \tag{3.16}$$

where $\mathbf{v}_{j,n}$ (and similarly $\mathbf{u}_{j,n}$ below) is the jth column vector of V_n (respectively, U_n). Likewise, we obtain

$$E_n^b(n) = \left[B_{0:n}(B_{0:n})^T\right]^{\frac{1}{2n}} = \left[U_n (S_n)^2 (U_n)^T\right]^{\frac{1}{2n}} = U_n (S_n)^{\frac{1}{n}} (U_n)^T,$$

from which we can deduce that $V_{E_n^b(n)} = U_n$ and

$$\lim_{n\to\infty} \|\mathbf{u}_{j,n} - \mathbf{l}_j^b(n)\| = 0. \tag{3.17}$$

We also infer that

$$(\sigma_j(B_{0:n}))^{\frac{1}{n}} = \lambda_j(E_n^b(n)) = \lambda_j(E_n^f(0)). \tag{3.18}$$

3.3. Controllability and observability for linear dynamics. The notions of observability and controllability are dual notions within filtering problems. Roughly, observability is the condition that given sufficiently many observations, the initial state of the system can be reconstructed by using a finite number of observations. Similarly, controllability can be described as the ability to move the system from any initial state to a desired state over a finite time interval. This is formally stated as follows.

DEFINITION 3.1. *The system* (3.1) *is defined to be* completely observable *if* $\forall n \geq 1$,

$$\det\left(\sum_{m=0}^{d-1} (B_{n:n+m})^T H_{n+m}^T Q_{n+m}^{-1} H_{n+m} B_{n:n+m}\right) \neq 0, \tag{3.19}$$

and it is defined to be completely controllable *if* $\forall n \geq 0$,

$$\det\left(\sum_{m=1}^{d} B_{n+m:n+d} F_{n+m} F_{n+m}^T \left(B_{n+m:n+d}\right)^T\right) \neq 0. \tag{3.20}$$

In addition we describe the system as uniformly completely observable *(respectively,* uniformly completely controllable*) if* (3.19) *(respectively,* (3.20)*) is bounded from zero uniformly in n.*

We will assume that the system in (3.2) is uniformly completely observable, i.e., the inequality (3.19) is uniformly bounded away from zero. Note, however, that this system *cannot* be controllable since the determinant in (3.20) is identically zero for a deterministic, perfect-model system as $F_n = 0\,\forall n$. The hypothesis of uniform complete observability ensures that the error covariance matrices remain bounded over time, as seen below.

LEMMA 3.2. *Suppose that the linear, nonautonomous system* (3.2) *where the initial state* $\mathbf{x}_0$ *has a Gaussian law with mean* $\mathbf{x}_{0|0}$ *and covariance* Δ_0 *is uniformly completely observable (Definition* 3.1*). Then the error covariance matrices remain bounded for all time, i.e., there exist constants* c_Σ *and* c_Δ *such that* $\forall n$, $\|\Delta_n\| \le c_\Delta$ *and* $\|\Sigma_n\| \le c_\Sigma$.

Proof. The result is proven for autonomous systems in Kumar and Varaiya [9, Chapter 7, equations (2.36) and (2.37)]. Extension to the nonautonomous case is straightforward by rehashing the steps and changing the constants of the autonomous system to their time-varying counterparts. ∎

One should note the recent work of Ni and Zhang [14] has demonstrated a stronger result: in continuous, perfect-model systems the assumption of uniform complete observability is sufficient to demonstrate the stability of the KF. In particular this shows that all solutions to the continuous Riccati equation for any choice of initial error covariance are bounded and converge to the same solution asymptotically. This strongly suggests the same can be shown for the discrete time system, and we will return to this point in our discussion of results in section 5.

Utilizing only the boundedness of the error covariance matrices, we demonstrate that the matrix M_n stays bounded in the following lemma.

LEMMA 3.3. *Consider the uniformly completely observable, perfect-model, linear, nonautonomous system* (3.2) *where the initial state* $\mathbf{x}_0$ *has a Gaussian law with covariance* $\Delta_0 > 0$. *Then the matrix* M_n *defined in* (3.8) *is uniformly bounded, i.e., there exists a constant* c_M *such that* $\|M_n\| \le c_M \, \forall n$.

Proof. We first show that the analysis error covariance matrix satisfies the recursive equation

$$\Delta_n = (I - K_n H_n) A_n \Delta_{n-1} A_n^T (I - K_n H_n)^T + K_n Q_n K_n^T. \tag{3.21}$$

Plugging in the Kalman update equations (3.3) and (3.3), the right-hand side of (3.21) equals $\Delta_n - (\Delta_n H_n^T - K_n Q_n) K_n^T$. Equation (4.29) in [4] establishes the equality $K_n = \Delta_n H_n^T Q_n^{-1}$ from which the recursion (3.21) follows, further implying that

$$\Delta_n \ge (I - K_n H_n) A_n \Delta_{n-1} A_n^T (I - K_n H_n)^T.$$

Recursively applying the above inequality gives $\Delta_n \ge M_n \Delta_0 M_n^T$. Decomposing $\Delta_0 = V_{\Delta_0} \Lambda_{\Delta_0} V_{\Delta_0}^T$ and employing Lemma 3.2 we find

$$\left\| M_n V_{\Delta_0} \Lambda_{\Delta_0}^{\frac{1}{2}} \right\|^2 \le \|\Delta_n\| \le c_\Delta.$$

As $\|M_n\| \le \|M_n V_{\Delta_0} \Lambda_{\Delta_0}^{\frac{1}{2}}\| \|\Lambda_{\Delta_0}^{-\frac{1}{2}} V_{\Delta_0}^T\|$ the result follows. Note that as $\Delta_0 > 0$ the matrix $\Lambda_{\Delta_0}^{-\frac{1}{2}}$ is well-defined. ∎

Bearing this bound in mind we shall proceed to discuss the asymptotic properties of the error covariance matrices.

3.4. The asymptotic rank deficiency of the error covariance. We begin by introducing a lemma which allows us to formally describe the collapse of the eigen-values of the error covariance matrix.

LEMMA 3.4. *For a given* $\epsilon > 0$, *let* $Z \in \mathbb{R}^{d \times d}$ *be a symmetric matrix such that there is a* $k \le d$ *dimensional subspace* $\mathcal{W} \subset \mathbb{R}^d$ *for which*

$$\sup\{\|Z\mathbf{u}\| : \|\mathbf{u}\| = 1, \mathbf{u} \in \mathcal{W}\} < \epsilon.$$

Then $\dim(\mathcal{E}^\epsilon(Z)) \geq k$*, where the subspace* $\mathcal{E}^\epsilon$ *is in accordance with Definition* 2.2.

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_d\}$ be an orthonormal eigen-vector basis for Z corresponding to $\mid \lambda_1(Z) \mid \geq \cdots \geq \mid \lambda_d(Z) \mid$, and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a basis for $\mathcal{W}$ of unit magnitude, such that we write

$$\mathbf{u}_l = \sum_{j=1}^{d} \beta_{l,j} \mathbf{v}_j; \quad l \in \{1, 2, \ldots, k\},$$

and the matrix of coefficients

$$\begin{pmatrix} \beta_{1,1} & \beta_{1,2} & \cdots & \beta_{1,d-k+1} & 0 & \cdots & 0 \\ \beta_{2,1} & \beta_{2,2} & \cdots & \beta_{2,d-k+1} & \beta_{2,d-k+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{k-1,1} & \beta_{k-1,2} & \cdots & \cdots & \cdots & \beta_{k-1,d-1} & 0 \\ \beta_{k,1} & \beta_{k,2} & \cdots & \cdots & \cdots & \beta_{k,d-1} & \beta_{k,d} \end{pmatrix}$$

is in column echelon form where for every column index $j > d - k + 1$, the entries

$$\beta_{1,j} = \cdots = \beta_{k+j-d-1,j} = 0$$

and for every row index $l \leq k$, $\sum_{j=1}^{d-k+l} \beta_{l,j}^2 = 1$ corresponding $\|u_l\| = 1$. Furthermore, as Z is symmetric its eigen-vectors form an orthonormal basis and hence $\|Z\mathbf{u}_l\|^2 = \sum_{j=1}^{d-k+l} \beta_{l,j}^2 \lambda_j^2(Z)$. For every $1 \leq l \leq k$, setting $s = k - l + 1$ we find

$$\epsilon^2 > \|Z\mathbf{u}_s\|^2 = \sum_{j=1}^{d-k+s} \beta_{s,j}^2 \lambda_j^2(Z) \geq \lambda_{d-k+s}^2(Z) = \lambda_{d-l+1}^2(Z).$$

Hence the k smallest eigen-values in absolute magnitude satisfy

$$\mid \lambda_d(Z) \mid \leq \cdots \leq \mid \lambda_{d-k+1}(Z) \mid < \epsilon$$

and the result follows. ∎

THEOREM 3.5. *Consider the uniformly completely observable, perfect-model, linear, nonautonomous system* (3.2) *where the initial state* $\mathbf{x}_0$ *has a Gaussian law with covariance* Δ_0*. Then* $\forall \epsilon > 0$*,* $\exists n_1 > 0$ *such that if* $n \geq n_1$*,* Σ_n *and* Δ_n *will each have at least* $d - d_0$ *eigen-values which are less than* ϵ *where* $d - d_0$ *is the number of* negative *Lyapunov exponents of the system* (3.2)*, i.e.,*

$$\dim(\mathcal{E}^\epsilon(\Sigma_n)) \geq d - d_0, \quad \textit{and} \quad \dim(\mathcal{E}^\epsilon(\Delta_n)) \geq d - d_0, \tag{3.22}$$

where the subspace $\mathcal{E}^\epsilon$ *is in accordance with Definition* 2.2.

Proof. As denoted earlier, let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_d$ be the Lyapunov exponents of the system (3.2) where $d_0 < d$ of them are nonnegative. The *forward stable* Lyapunov vectors based at time zero are the set $\{\mathbf{l}_j^f(0)\}_{j=d_0+1}^d$ which by definitions (3.13) and (3.15) satisfy

$$\lim_{n \to \infty} \frac{1}{n} \log\left(\left\| B_{0:n} \mathbf{l}_j^f(0) \right\| \right) = \mu_j. \tag{3.23}$$

Rewriting the analysis error covariance update equation in terms of the transpose

$$\Delta_n = M_n \Delta_0 B_{0:n}^T = B_{0:n} \Delta_0 M_n^T$$

we get $\Delta_n M_n^{-T} \Delta_0^{-1} = B_{0:n}$ and in particular

$$\Delta_n M_n^{-T} \Delta_0^{-1} \mathbf{l}_j^f(0) = B_{0:n} \mathbf{l}_j^f(0).$$

Let us therefore define the sequence of vectors

$$\mathbf{w}_{j,n} \equiv M_n^{-T} \Delta_0^{-1} \mathbf{l}_j^f(0). \tag{3.24}$$

By Lemma 3.3 we know that M_n is bounded above, so that the sequence of vectors $\mathbf{w}_{j,n} = M_n^{-T} \Delta_0^{-1} \mathbf{l}_j^f(0)$ must be bounded below. As such, there is a constant $c_{\mathbf{w}}$ such that $c_{\mathbf{w}} \leq \|\mathbf{w}_{j,n}\| \forall n$ and $j \in \{d_0 + 1, \ldots, d\}$. Choose a $\rho > 0$ such that for each $j \in \{d_0 + 1, \ldots, d\}$, $\rho + \mu_j < 0$. Define $\overline{\mathbf{w}}_{j,n} \equiv \frac{\mathbf{w}_{j,n}}{\|\mathbf{w}_{j,n}\|}$. Then for a given $\epsilon > 0$, $\exists n_1$ such that for $n \geq n_1$

$$\|\Delta_n \overline{\mathbf{w}}_{j,n}\| = \frac{1}{\|\mathbf{w}_{j,n}\|} \|B_{0:n} \mathbf{l}_j^f(0)\| \leq \frac{1}{c_{\mathbf{w}}} e^{(\mu_j + \rho)n} < \epsilon. \tag{3.25}$$

The theorem is therefore an immediate consequence of Lemma 3.4. The proof for Σ_n follows along similar lines. ∎

3.5. Null space characterization and assimilation in the unstable subspace. The sequence of subspaces defined by the span of $\{\mathbf{w}_{j,n}\}_{j=d_0+1}^d$ will be the object of study for the remainder of this section. In particular, we wish to establish the connection between this sequence of subspaces and AUS which utilizes the *backward Lyapunov vectors.*

DEFINITION 3.6. *Define* $\Lambda^s_{E_n^f(0)}$ *to be the* $d - d_0 \times d - d_0$ *diagonal matrix with diagonal entries given by* $\left\{\lambda_j\left(E_n^f(0)\right)\right\}_{j=d_0+1}^d$. *Also, let us define the following* $d \times d - d_0$ *operators:*

$$U_n^s = \left[\mathbf{u}_{d_0+1,n}, \ldots, \mathbf{u}_{d,n}\right], \tag{3.26}$$

$$V_n^s = \left[\mathbf{v}_{d_0+1,n}, \ldots, \mathbf{v}_{d,n}\right], \tag{3.27}$$

$$L_n^{bs} = \left[\mathbf{l}^b_{d_0+1}(n), \ldots, \mathbf{l}^b_d(n)\right]. \tag{3.28}$$

Note that (3.17) *implies that*

$$\lim_{n\to\infty} \|U_n^s - L_n^{bs}\| = 0. \tag{3.29}$$

Consider (3.10), namely, $\Delta_n = M_n \Delta_0 V_n S_n U_n^T$, for the analysis error covariance Δ_n at time n in terms of the matrix M_n and the singular-value decomposition of the propagator $B_{0:n}$. Noting that $B_{0:n}^T \mathbf{u}_{j,n} = \sigma_j(B_{0:n})\mathbf{v}_{j,n}$ and utilizing the relation (3.18) we get

$$\Delta_n U_n^s \left(U_n^s\right)^T = M_n \Delta_0 V_n^s \left(\Lambda^s_{E_n^f(0)}\right)^n \left(U_n^s\right)^T. \tag{3.30}$$

Likewise, recalling that $\Sigma_n = A_n \Delta_{n-1} A_n^T$, we can express the restriction of the forecast error covariances as

$$\Sigma_n U_n^s \left(U_n^s\right)^T = A_n M_{n-1} \Delta_0 V_n^s \left(\Lambda^s_{E_n^f(0)}\right)^n \left(U_n^s\right)^T. \tag{3.31}$$

Making use of the above relations we now prove one of our main results, which states that the norm of the restriction of the analysis and forecast error covariances onto the backward stable Lyapunov subspaces must tend to zero.

THEOREM 3.7. *Consider the uniformly completely observable, perfect-model, linear, nonautonomous system* (3.2) *where the initial state* $\mathbf{x}_0$ *has a Gaussian law with covariance* Δ_0*. The restriction of* Δ_n *and* Σ_n *into the span of the backward stable Lyapunov vectors,* $\{\mathbf{l}_j^b(n)\}_{j=d_0+1}^d$*, tends to zero as* $n \to \infty$*. That is,*

$$\lim_{n\to\infty} \|\Delta_n L_n^{bs} \left(L_n^{bs}\right)^T \| = 0, \tag{3.32}$$

$$\lim_{n\to\infty} \|\Sigma_n L_n^{bs} \left(L_n^{bs}\right)^T \| = 0. \tag{3.33}$$

Proof. By definition $\log(\lambda_j(E^f(0))) = \mu_j$, so that the eigen-values $\lambda_j(E^f(0)) < 1$ correspond to the stable Lyapunov exponents. Recalling that $\lambda_{d_0+1}(E_n^f(0)) \geq \cdots \geq \lambda_d(E_n^f(0))$ we find $\|\Lambda^s_{E_n^f(0)}\| = \lambda_{d_0+1}(E_n^f(0))$ and

$$\lim_{n\to\infty} \left\|\Lambda^s_{E_n^f(0)}\right\| = \lambda_{d_0+1}(E^f(0)) < 1. \tag{3.34}$$

Consequent to (3.34) we can choose a small $0 < \rho < 1$ and sufficiently large n_1 such that when $n \geq n_1$, $\|\Lambda^s_{E_n^f(0)}\| \leq 1 - \rho$.

The restriction of Δ_n into the span of the columns of U_n^s is given by (3.30). Note the column vectors of V_n^s and U_n^s are orthogonal and of unit norm, hence $\|V_n^s\| = \|U_n^s\| = 1$. We then find for $n \geq n_1$

$$\|\Delta_n U_n^s \left(U_n^s\right)^T \| \leq \left\|\Lambda^s_{E_n^f(0)}\right\|^n \|M_n\| \|\Delta_0\| \leq (1-\rho)^n c_M \|\Delta_0\|. \tag{3.35}$$

Consider

$$\|\Delta_n L_n^{bs} \left(L_n^{bs}\right)^T \| \leq \|\Delta_n\| \|L_n^{bs} \left(L_n^{bs}\right)^T - U_n^s \left(U_n^s\right)^T \| + \|\Delta_n U_n^s \left(U_n^s\right)^T \|, \tag{3.36}$$

and Lemma 3.2 states $\|\Delta_n\|$ is bounded. Therefore,

$$\lim_{n\to\infty} \|\Delta_n L_n^{bs} \left(L_n^{bs}\right)^T \| = 0 \tag{3.37}$$

by (3.17) and (3.35). This may be similarly stated for the forecast error covariance. ∎

The forecast and analysis error covariance matrices for a generic nonautonomous system in general do not converge, but the above results entail that asymptotically the only relevant directions for the error covariance matrices are the backward unstable-neutral Lyapunov directions validating the central hypothesis made by Trevisan and Palatella [22] in their proposed reduced rank Kalman filtering algorithms.

An intriguing consequence from (3.25) in Theorem 3.5 is the following corollary.

COROLLARY 3.8. *Suppose that for some* $\epsilon_0 > 0$*,* $N_0 > 0$*, and for every* $0 < \epsilon < \epsilon_0$*,* $n > N_0$*,*

$$\dim\left(\mathcal{E}^\epsilon(\Delta_n)\right) = d - d_0, \tag{3.38}$$

i.e., asymptotically the rank deficiency of the analysis error covariance Δ_n *is exactly of dimension* $d - d_0$*. Then the transformation* $M_n^{-T}\Delta_0^{-1}$ *asymptotically maps the forward stable vectors* $\{\mathbf{l}_j^f(0)\}_{j=d_0+1}^d$ *into the span of the backward stable vectors* $\{\mathbf{l}_j^b(n)\}_{j=d_0+1}^d$ *as* $n \to \infty$*.*

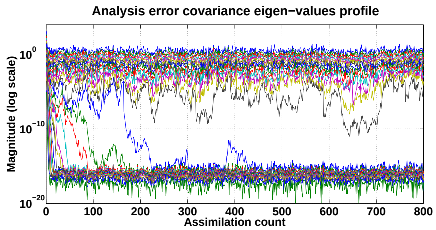

FIG. 1. *Profile of the eigen-values of* Δ_n*. Counting establishes that the bottom 16 eigen-values converge to zero.*

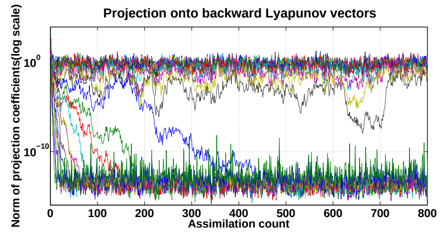

FIG. 2. *Norm of the projection coefficients* $\|\Delta_n \mathbf{u}_{j,n}\|$ *for varying observation time* n*.*

3.6. Numerical results for a 30-dimensional system. Below we provide an illustration for this *asymptotic rank deficiency* property of the error covariance matrices. The state space vector $\mathbf{x}_n$ and the observation vector $\mathbf{y}_n$ have dimension $d = 30$ and $q = 10$, respectively. This choice is arbitrary and our simulations with different d and q have shown qualitatively equivalent results.

The time-varying, invertible propagators $A_n \in \mathbb{R}^{30\times 30}$, the observation error covariance matrices $Q_n \in \mathbb{R}^{10\times 10}$, and the observation matrices $H_n \in \mathbb{R}^{10\times 30}$ were all randomly generated for sufficiently large n. We employed the QR method [10] to numerically compute the Lyapunov vectors and the Lyapunov exponents and it was found that the number of nonnegative Lyapunov exponents was $d_0 = 14$. Starting from a random positive-definite Δ_0, the sequence (Σ_n, Δ_n) was generated based on the Kalman update equations (3.3)–(3.5). For every n we computed the eigen-values of Δ_n sorted in descending order.

Figure 1 shows the eigen-values of Δ_n as a function of n. Barring the dominant 14 eigen-values, the rest converge to zero, serving as a visual testament to Theorem 3.5. Furthermore, we also calculated the norm $\|\Delta_n \mathbf{u}_{j,n}\|, j \in \{1, 2, \ldots, d\}, \forall n$ and plot them in Figure 2. These norm values are unsorted, meaning that the topmost line in Figure 2 represents the values $\|\Delta_n \mathbf{u}_{1,n}\|$ and the bottommost line denotes $\|\Delta_n \mathbf{u}_{d,n}\|$ for different values of n. For $j > d_0 = 14$, $\|\Delta_n \mathbf{u}_{j,n}\|$ approaches zero, suggesting that as $n \to \infty$, the row space of Δ_n (and also Σ_n) coincides the space spanned by the unstable-neutral, backward Lyapunov vectors, i.e., the bounds in inequalities (3.22) are saturated.

4. Autonomous linear dynamical systems.

4.1. Null space characterization for autonomous systems. The noiseless, linear autonomous system can be defined from (3.2), with the additional assumptions that $A_n \equiv A$, $H_n \equiv H$, $Q_n \equiv Q$ are fixed matrices $\forall n$—therefore the results about the asymptotic rank deficiency property of the error covariance matrices in section 3 also apply to autonomous systems. However, a stronger statement can be made for time invariant systems because the backward Lyapunov vectors will not vary in time. In fact, the result in this section is valid even for the case when only the dynamical system is autonomous ($A_n \equiv A$) but the observation process is time dependent (H_n and Q_n depend on n).

Akin to the nonautonomous case we define

$$E^b_n \equiv [A^n (A^n)^*]^{\frac{1}{2n}}, \quad E^f_n \equiv [(A^n)^* A^n]^{\frac{1}{2n}} \tag{4.1}$$

and the similarity with (3.12) can readily be seen by setting $B_{m:m+n} = A^n \, \forall m$ in (3.9) (hence the omission of the time index m). As before, the existence of the limits

$$E^b \equiv \lim_{n\to\infty} E_n^b\,, \quad E^f \equiv \lim_{n\to\infty} E_n^f \tag{4.2}$$

is guaranteed by the Oseledets theorem [10]. The eigen-vectors of E^b and E^f are called the backward and forward Lyapunov vectors, represented here as the column vectors of L^b and L^f ordered left to right from the most unstable direction—corresponding to the largest Lyapunov exponent—to the most stable direction—corresponding to the smallest Lyapunov exponent. Specifically, the Lyapunov vectors are defined globally and have *no* dependence on the time in the linear, autonomous case. Without the time dependence on the backward stable Lyapunov vectors, we obtain a stronger statement about the asymptotic null space of the covariance matrices.

Definition 4.1. *Let* $L^{bs} \equiv L_n^{bs} = \left[\mathbf{l}_{d_0+1}^b, \ldots, \mathbf{l}_d^b\right]$. *Note that Theorem* B.3 *proved in Appendix* B *states that the span of the columns of* L^{bs} *is equal to* $\mathcal{E}^1(A^T)$.

Corollary 4.2. *Consider the uniformly completely observable, perfect-model, linear, autonomous system defined from* (3.2) *where* $A_n \equiv A$, *but* H_n *and* Q_n *may depend on* n *and the initial state* $\mathbf{x}_0$ *has a Gaussian law with covariance* Δ_0. *Then the restriction of the analysis and forecast error covariances onto* $\mathcal{E}^1(A^T)$ *tend to zero as* $n \to \infty$. *That is,*

$$\lim_{n\to\infty} \|\Delta_n L^{bs} \left(L^{bs}\right)^T\| = 0, \tag{4.3}$$

$$\lim_{n\to\infty} \|\Sigma_n L^{bs} \left(L^{bs}\right)^T\| = 0. \tag{4.4}$$

Proof. Combining Theorem 3.7 with Theorem B.3 this is a straightforward consequence. ∎

In our numerical simulations with arbitrary (and completely observable) choices of A, H, and Q we have additionally observed *convergence* of Δ_n and Σ_n to a fixed Δ and Σ, respectively, and seen their null spaces contain $\mathcal{E}^1\left(A^T\right)$ as stated by Corollary 4.2 (refer to section 4.2). Considering the recent work of Ni and Zhang [14], this strongly suggests that the classical result of the stable Riccati equation for completely observable and controllable, discrete autonomous systems [9] has an analogue in the case of completely observable, perfect-model systems.

4.2. Numerical results for linear autonomous system. We choose a non-singular matrix $A \in \mathbb{R}^{30\times 30}$ $(d = 30)$ consisting of random entries and set $d_0 = 12$ of its eigen-values to be greater than or equal to one in absolute magnitude. We ran the Kalman filtering system long enough and observed that the analysis error covariances do converge to a fixed Δ and then projected Δ onto the generalized eigen-space of A^T. Figure 3 plots the absolute magnitude of eigen-values of A sorted in descending order $(\mid \lambda_1(A) \mid \geq \cdots \geq \mid \lambda_d(A) \mid)$ in blue and shows the Lyapunov exponents for this system in red, where we note that the number of nonnegative Lyapunov exponents is exactly 12 tantamount to the number of eigen-values of A greater than or equal to one in magnitude. Additionally, it can be verified that the Lyapunov exponents are just the logarithm (to the base e) of the absolute magnitude eigen-values of A. Recalling the definition of the Lyapunov exponents from (3.15), this equality also lends credence to our Theorem A.3. The plot in Figure 4 displays $\|\Delta\left(\mathbf{v}_j\left(A^T\right)\right)\|; j \in \{1, 2, \ldots, d\}$, where $\mathbf{v}_j\left(A^T\right)$ is the generalized eigen-vector of $\lambda_j(A)$. Observe that when $j > 12$, the norm of the projected coefficients is zero, rendering a visual confirmation to Corollary 4.2.

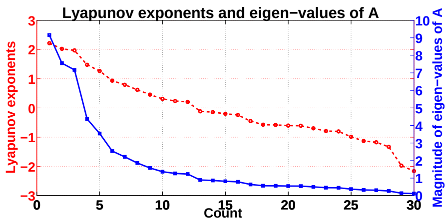

FIG. 3. *Lyapunov exponents in blue and the magnitude of the eigen-values of A in red.*

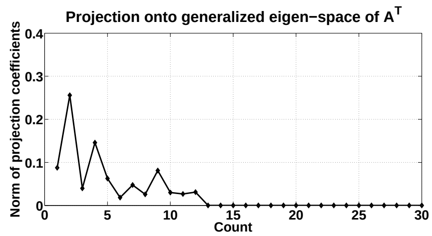

FIG. 4. *Norm of the projection coefficients onto the generalized eigen-space of A^T.*

5. Discussion. We have shown that under sequential Kalman filtering, the error covariance for a linear, perfect-model, conditionally Gaussian system asymptotically collapses to the subspaces spanned by the backward unstable Lyapunov vectors. This has been known to practitioners in the forecasting community [1] but had yet to be stated in precise mathematical terms. In particular, this foundational work validates the underlying assumptions and methodology of AUS.

At the same time, these results open many new questions for ongoing research related to AUS algorithms. For instance, the present results do not formally show the equivalence of a fully reduced-rank algorithm such as EKF-AUS applied in such a setting. The conditions that imply the convergence of the covariance matrices, given arbitrary low rank symmetric matrices chosen as initial conditions have yet to be established. Recent work strongly suggests that filter stability for discrete, perfect-model systems can be demonstrated under sufficient observability hypotheses alone [14]. Determining the necessary hypotheses for stability of the discrete with low rank initializations of the prior covariance matrix in perfect-model systems will be the subject of the sequel to our work.

Additionally there are conceptual issues to be resolved in bridging the results for linear systems to nonlinear settings, the former having the advantage of Lyapunov vectors being defined globally in space, whereas the formulation must change in a nonlinear setting, respecting the dependence on the underlying path. Both of these directions of inquiry open rich areas for mathematical research and future algorithm design.

While the ultimate goal of DA is a precise estimate of state for chaotic dynamics, it is critical to understand the uncertainty of the prediction. An exact calculation of the posterior distribution of states for a high dimensional, complex system is computationally intractable; as computational resources increase, so will model complexity and thus computational efficiency alone will not resolve this issue. This work provides an idealized but general framework for future investigations into low dimensional approximations for uncertainty calculation. We hope that a precise mathematical framework for understanding the nature of uncertainty for linear systems will lead to innovative research to surmount these challenges.

Appendix A. Eigen-values, singular values, and Lyapunov exponents of linear autonomous systems. The results established in this appendix and Appendix B should be treated as an independent body of work elucidating the relationship between various concepts in linear, autonomous systems and not restricted to the domain of DA and filtering theory. While these relationships are known and can be retrieved from multiple sources in the literature, we have explicitly proved them here for completeness. Readers familiar with these mathematical connections may choose to skip through these sections without any loss of continuity.

Based on the definition of the matrix E^f_n in (4.1) we find $\lambda_j(E^f_n) = [\sigma_j(A^n)]^{\frac{1}{n}}$. As $E^f_n \to E^f$ we also have

$$\lim_{n\to\infty} \lambda_j(E^f_n) = \lim_{n\to\infty} [\sigma_j(A^n)]^{\frac{1}{n}} = \lambda_j(E^f) \qquad j \in \{1,2,\dots,d\}, \tag{A.1}$$

where the eigen-values λ_j and singular values σ_j are ordered descending in norm. Dropping the label for brevity let $J = V_A^{-1} A V_A$ (instead of $J(A)$) be the Jordan canonical form of A. It is straightforward to see that $A^n = V_A J^n V_A^{-1}$ for any integer n. The following inequality stated in Theorem 9 of [12] is quite useful. For any two square matrices Z_1 and Z_2 we have

$$\sigma_j(Z_1)\sigma_d(Z_2) \quad\le\quad \sigma_j(Z_1 Z_2) \quad\le\quad \sigma_j(Z_1)\sigma_1(Z_2). \tag{A.2}$$

Since the singular values of both the matrix and its transpose are the same, it follows that

$$\sigma_d(Z_1)\sigma_j(Z_2) \quad\le\quad \sigma_j(Z_1 Z_2) \quad\le\quad \sigma_1(Z_1)\sigma_j(Z_2). \tag{A.3}$$

LEMMA A.1. *For any square matrix $Z = V_Z J(Z) V_Z^{-1}$*

$$\lim_{n\to\infty} [\sigma_j(Z^n)]^{\frac{1}{n}} = \lim_{n\to\infty} [\sigma_j(J(Z)^n)]^{\frac{1}{n}}.$$

Proof. Inequalities (A.2) and (A.3) lead to

$$\sigma_d(V_Z)\,\sigma_d\left(V_Z^{-1}\right)\sigma_j(J(Z)^n) \le \sigma_j(Z^n) \le \sigma_1(V_Z)\,\sigma_1\left(V_Z^{-1}\right)\sigma_j(J(Z)^n).$$

Raising each term to the power $1/n$ and letting $n \to \infty$ proves the result. ∎

COROLLARY A.2. *For any matrix A let E^f be defined as in* (4.2) *and J be the Jordan canonical form of A. Then $\lambda_j(E^f) = \lim_{n\to\infty} [\sigma_j(J^n)]^{\frac{1}{n}}, j \in \{1,2,\dots,d\}$.*

Proof. The results follow immediately when we employ Lemma A.1 setting $Z = A$ in conjunction with (A.1). ∎

The theorem below establishes the relation between the eigen-values of the time invariant propagator A and the limit matrix E^f.

THEOREM A.3. *For any matrix A let the matrix E^f be defined as in* (4.2)*. Then the eigen-values of E^f equal the absolute magnitude eigen-values of A, i.e., $\lambda_j(E^f) = |\lambda_j(A)|, j \in \{1,2,\dots,d\}$.*

Proof. We consider two different cases.

Case 1: *A is diagonalizable.* When J is diagonal then $\sigma_j(J) = |\lambda_j(J)| = |\lambda_j(A)|$. Recalling that $\lambda_j(J^n) = [\lambda_j(J)]^n\ \forall n$, we get $[\sigma_j(J^n)]^{\frac{1}{n}} = |\lambda_j(A)|$ and the result follows from Corollary A.2.

Case 2: *A is not diagonalizable.* Let $J_\lambda(A)$ denote the Jordan block of size $k \times k$ corresponding to an eigen-value λ of A of the form

$$J_\lambda(A) \equiv \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}. \tag{A.4}$$

The following lemma is useful in proving Theorem A.3.

LEMMA A.4. *For any matrix A let $J_\lambda(A)$ be a Jordan block corresponding to eigen-value λ of A as defined in* (A.4). *Then the singular values of $J_\lambda(A)$ respect the following equality, namely,*

$$\lim_{n\to\infty} \left[\sigma_j\left(J_\lambda^n\right)\right]^{\frac{1}{n}} = |\lambda|, \qquad j \in \{1, 2, \dots, k\}, \tag{A.5}$$

i.e., the limiting singular-values are the absolute magnitude of their respective eigen-values.

Proof. Following the standard proof technique for equality results we individually show that

$$\lim_{n\to\infty} \left[\sigma_j\left(J_\lambda^n\right)\right]^{\frac{1}{n}} \le |\lambda|, \qquad j \in \{1, 2, \dots, k\}, \tag{A.6}$$

and

$$\lim_{n\to\infty} \left[\sigma_j\left(J_\lambda^n\right)\right]^{\frac{1}{n}} \ge |\lambda|, \qquad j \in \{1, 2, \dots, k\}. \tag{A.7}$$

Let the Nilponent matrix $N \equiv J_\lambda - \lambda I$ with $N^k = \mathbf{0}$. When $n \ge k-1$ we get

$$J_\lambda^n \quad = \quad (\lambda I + N)^n \quad = \quad \sum_{r=0}^{k-1} \tbinom{n}{r} \lambda^{n-r} N^r.$$

Further, the highest singular-value $\sigma_1(N^r) = 1$ for $r \in \{0, 1, \dots, k-1\}$. If $\lambda = 0$, then $J_\lambda^n = \mathbf{0}$ when $n \ge k-1$ and the result is trivially true. Suppose $\lambda \ne 0$ define $\delta \equiv \frac{1}{\lambda}$. Using the identity that for any two matrices Z_1 and Z_2, $\sigma_1(Z_1+Z_2) \le \sigma_1(Z_1)+\sigma_1(Z_2)$ as stated in Theorem 6 of [12], we have

$$\sigma_1\left(J_\lambda^n\right) \le |\lambda|^n \left[\sum_{r=0}^{k-1} \binom{n}{r} |\delta|^r\right]. \tag{A.8}$$

Let $|\delta| = \epsilon\xi$ for any $0 < \epsilon \le |\delta|$. Then

$$\begin{aligned}
\sigma_1\left(J_\lambda^n\right) &\le |\lambda|^n \xi^k \left[\sum_{r=0}^{k-1} \binom{n}{r} \epsilon^r\right] \\
&\le |\lambda|^n \xi^k \left[\sum_{r=0}^{n} \binom{n}{r} \epsilon^r\right] = |\lambda|^n \xi^k (1+\epsilon)^n.
\end{aligned}$$

Raising to the power $1/n$ and taking the limit we get

$$\lim_{n\to\infty} \left[\sigma_1\left(J_\lambda^n\right)\right]^{\frac{1}{n}} \le |\lambda|(1+\epsilon).$$

The above inequality is also true for the rest of the singular values as $\sigma_1(.)$ is the largest. Since ϵ is *arbitrary* the first inequality (A.6) follows. If $\lambda = 0$ we get the desired, stronger equality result in (A.5) as the singular values by definition are non-negative. It suffices to focus on the case $\lambda \ne 0$, where J_λ is invertible.

To establish the reverse inequality (A.7), let T_λ be the Jordan canonical form of J_λ^{-1} given by

$$T_\lambda \equiv \begin{pmatrix} \frac{1}{\lambda} & 1 & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \frac{1}{\lambda} & 1 \\ 0 & 0 & 0 & 0 & \frac{1}{\lambda} \end{pmatrix}.$$

Lemma A.1 entails that

$$\lim_{n\to\infty}\left[\sigma_j\left(\left(J_\lambda^{-1}\right)^n\right)\right]^{\frac{1}{n}} = \lim_{n\to\infty}\left[\sigma_j(T_\lambda^n)\right]^{\frac{1}{n}}.$$

Applying the inequality (A.6) on T_λ gives us

$$\lim_{n\to\infty}\left[\sigma_j\left(T_\lambda^n\right)\right]^{\frac{1}{n}} \le \tfrac{1}{|\lambda|}, \qquad j\in\{1,2,\dots,k\}.$$

In particular,

$$\lim_{n\to\infty}\left[\sigma_1\left(\left(J_\lambda^{-1}\right)^n\right)\right]^{\frac{1}{n}} = \lim_{n\to\infty}\frac{1}{\left[\sigma_k\left(J_\lambda^n\right)\right]^{\frac{1}{n}}} \le \frac{1}{|\lambda|},$$

where the equality stems from the fact that for any invertible matrix Z of size $k\times k$

$$\sigma_j\left(Z^{-1}\right) = \frac{1}{\sigma_{k-j+1}\left(Z\right)}.$$

We then get

$$\lim_{n\to\infty}\left[\sigma_k(J_\lambda^n)\right]^{\frac{1}{n}} \ge |\lambda|. \tag{A.9}$$

Since $\sigma_k(.)$ is the smallest singular value the inequality (A.9) is also valid for the rest. ∎

Now to prove Theorem A.3 note that for any n

$$J^n = \begin{pmatrix} J_{\lambda_1}^n & 0 & \cdots & 0 \\ 0 & J_{\lambda_2}^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_{\lambda_l}^n \end{pmatrix}$$

is a block diagonal matrix and the eigen-(singular) values of J^n equal the *disjoint union* of eigen-(singular) values of individual Jordan blocks $J_{\lambda_1}^n,\dots,J_{\lambda_l}^n$. In accordance with Corollary A.2 and Lemma A.4 we find $\forall j\in\{1,2,\dots,d\}$,

$$\lambda_j(E^f) = \lim_{n\to\infty}\left[\sigma_j(J^n)\right]^{\frac{1}{n}} = |\lambda_j(J)| = |\lambda_j(A)|\,.$$ ∎

Appendix B. Eigen-spaces and Lyapunov vectors of linear autonomous systems. By a suitable coordinate transformation, namely, $\mathbf{z}_n = V_A^{-1}\mathbf{x}_n$, studying the dynamics $\mathbf{x}_{n+1} = A\mathbf{x}_n$ is tantamount to investigating $\mathbf{z}_{n+1} = J\mathbf{z}_n$, where $J = V_A^{-1}AV_A$ is the Jordan canonical form of A. Indeed,

$$\mathbf{z}_{n+1} = J\mathbf{z}_n = V_A^{-1}AV_AV_A^{-1}\mathbf{x}_n = V_A^{-1}\mathbf{x}_{n+1}.$$

Corresponding to the definitions of the matrices E_n^f and E^f in (4.1)–(4.2), let $G_n \equiv \left[(J^n)^*J^n\right]^{\frac{1}{2n}}$ and let $G \equiv \lim_{n\to\infty} G_n$.

We consider the two systems in the different d dimensional spaces $\mathbb{R}_A^d$ and $\mathbb{C}_J^d$, where the underlying propagators are A and J, respectively. Note that as the matrix V_A might be complex (though A is real) the dynamics for the propagator J is examined in a complex state space.

LEMMA B.1. *If the scalar product in* $\mathbb{C}_J^d$ *is the canonical one, namely,* $\langle \mathbf{u}, \mathbf{v} \rangle_J = \mathbf{u}^\dagger \mathbf{v}$*, then* $V_G = I_d$*, where* I_d *is the* $d \times d$ *identity matrix.*

Proof. We find it convenient to handle the following scenarios separately.

Case 1: *A is diagonalizable.* J is diagonal and so is J^n. In the canonical inner product setting the entries of the diagonal G_n are the absolute magnitude entries of J. It follows that G is diagonal and $V_G = V_J = I_d$.

Case 2: *A is not diagonalizable.* As before, consider the Jordan block J_λ given in (A.4) of size $k \times k$ corresponding to the eigen-value λ. Define $G_\lambda \equiv \lim_{n\to\infty} [(J_\lambda^n)^* J_\lambda^n]^{\frac{1}{2n}}$. Since G_λ is symmetric it is diagonalizable and by Theorem A.3 we have $\lambda_j(G_\lambda) = |\lambda|\, \forall j \in \{1, 2, \ldots, k\}$. As all the eigen-values of G_λ are equal, it is a scalar matrix and therefore we can choose $V_{G_\lambda} = I_k$. Since

$$G = \begin{pmatrix} G_{\lambda_1} & 0 & \cdots & 0 \\ 0 & G_{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_{\lambda_l} \end{pmatrix}$$

the result follows. ∎

LEMMA B.2. *Under the definition of the scalar products* $\langle \mathbf{u}, \mathbf{v} \rangle_J = \mathbf{u}^\dagger V_A^\dagger V_A \mathbf{v}$ *in* $\mathbb{C}_J^d$ *and* $\langle \mathbf{u}, \mathbf{v} \rangle_A = \mathbf{u}^T \mathbf{v}$ *in* $\mathbb{R}_A^d$*,* $V_G = V_A^{-1} V_{E^f}$*.*

Proof. For the aforesaid considerations of the scalar products in $\mathbb{C}_J^d$ and $\mathbb{R}_A^d$, $J^* = (V_A^\dagger V_A)^{-1} J^\dagger V_A^\dagger V_A$ and $A^* = A^T$, respectively. Recalling that $J = V_A^{-1} A V_A$ we have

$$(J^n)^* = \left(V_A^\dagger V_A\right)^{-1} V_A^\dagger \left(A^n\right)^T \left(V_A^{-1}\right)^\dagger V_A^\dagger V_A = V_A^{-1} \left(A^n\right)^T V_A$$

$$\Rightarrow G_n = \left[V_A^{-1} \left(A^n\right)^T V_A V_A^{-1} A^n V_A\right]^{\frac{1}{2n}} = \left[V_A^{-1} \left(A^n\right)^T A^n V_A\right]^{\frac{1}{2n}}.$$

As $\left(E_n^f\right)^{2n} = \left(A^n\right)^T A^n$ is symmetric, it is diagonalizable by an orthonormal matrix $V_{E_n^f}$ and carries a representation $\left(E_n^f\right)^{2n} = V_{E_n^f} (\Lambda_{E_n^f})^{2n} V_{E_n^f}^T$. We find $\Lambda_{G_n} = \Lambda_{E_n}$ and $V_{G_n} = V_A^{-1} V_{E_n}\, \forall n$ and the result follows by letting $n \to \infty$. ∎

Recall the real span $\mathcal{T}_\mathbf{w}$ from Definition 2.1 bearing in mind the complex generalized eigen-vectors of any matrix Z always occur in conjugate pairs $\{\mathbf{w}, \overline{\mathbf{w}}\}$ with $\mathcal{T}_\mathbf{w} = \mathcal{T}_{\overline{\mathbf{w}}}$. We have the following theorem.

THEOREM B.3 (eigenspace equality). *For any matrix* A *let the matrix* E^f *be defined as in* (4.2)*. Then for any* $\alpha \geq 0$ *the corresponding* α*-eigenspaces of* E^f *and* A *are the same, i.e.,* $\mathcal{E}^\alpha\left(E^f\right) = \mathcal{E}^\alpha(A)$*. Equivalently,* $\mathcal{E}^\alpha\left(E^b\right) = \mathcal{E}^\alpha\left(A^T\right)$*.*

Proof. By Theorem A.3 we have $\lambda_j(G) = |\lambda_j(J)| = |\lambda_j(A)| = \lambda_j(E^f)$. Recall that the eigen-values are ordered with $\lambda_1(G)$ and $\lambda_d(G)$ being the largest and the smallest, respectively. The Oseledets theorem states that there exists a sequence of embedded subspaces

$$0 \subset \mathcal{F}_d \subset \mathcal{F}_{d-1} \subset \cdots \subset \mathcal{F}_1 = \mathbb{C}_J^d$$

such that on the complement $\mathcal{F}_j \backslash \mathcal{F}_{j+1}$ of $\mathcal{F}_{j+1}$ in $\mathcal{F}_j$ the growth rate is at most $\lambda_j(G)$ [15]. The subspaces $\mathcal{F}_j$ can be obtained as the direct sum of the eigen-vectors $\mathbf{v}_j(G)$ as

$$\mathcal{F}_j = \mathbf{v}_d(G) \oplus \mathbf{v}_{d-1}(G) \oplus \cdots \oplus \mathbf{v}_j(G),$$

where $\mathbf{v}_j(G)$ is the eigen-vector of G corresponding to $\lambda_j(G)$. Further, though the eigen-vectors of G depend on the underlying scalar product in $\mathbb{C}^d_J$, the embedded subspaces $\mathcal{F}_j$ and the eigen-values $\lambda_j(G)$ are *independent* of it [10].

Corresponding to the two inner product definitions in $\mathbb{C}^d_J$, specifically $\langle \mathbf{u}, \mathbf{v}\rangle_J = \mathbf{u}^\dagger \mathbf{v}$ and $\langle \mathbf{u}, \mathbf{v}\rangle_J = \mathbf{u}^\dagger V_A^\dagger V_A \mathbf{v}$, we denote the respective eigen-vectors with the superscript symbols 1 and 2. By Lemma B.1 we have $V_G^1 = I_d = V_A^{-1} V_A$ and Lemma B.2 declares that $V_G^2 = V_A^{-1} V_{E^f}$, where V_{E^f} is computed using the canonical inner product in $\mathbb{R}^d_A$. For the given α let $q = \operatorname{argmin}_j \lambda_j(G) \le \alpha$. The invariance of the embedded subspace $\mathcal{F}_q$ to the underlying scalar product signifies that the real span of the vectors $\{V_A \mathbf{v}^1_d(G), \ldots, V_A \mathbf{v}^1_q(G)\}$ equals the real span of the vectors $\{V_A \mathbf{v}^2_d(G), \ldots, V_A \mathbf{v}^2_q(G)\}$. As $\forall j \in \{1, 2, \ldots, d\}, V_A \mathbf{v}^1_j(G) = \mathbf{v}_j(A)$ and $V_A \mathbf{v}^2_j(G) = \mathbf{v}_j(E^f)$, the result follows. □

REFERENCES

[1] S. Bonnabel and R. Sepulchre, *The geometry of low-rank Kalman filters*, in Matrix Information Geomentry, F. Nielsen and R. Bhatia, eds., Springer, Berlin, 2013, pp. 53–68.

[2] P. Bougerol, *Kalman filtering with random coefficients and contractions*, SIAM J. Control Optim., 31 (1993), pp. 942–959.

[3] A. Carrassi, A. Trevisan, L. Descamps, O. Talagrand, and F. Uboldi, *Controlling instabilities along a 3DVar analysis cycle by assimilating in the unstable subspace: A comparison with the EnKF*, Nonlinear Process. Geophys., 15 (2008), pp. 503–521.

[4] S. E. Cohn, *An introduction to estimation theory*, J. Meteor. Soc. Japan, 75 (1997), pp. 257–288.

[5] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer, New York, 2009.

[6] M. Ghil and P. Malanotte-Rizzoli, *Data assimilation in meteorology and oceanography*, Adv. Geophys., 33 (1991), pp. 141–266.

[7] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[8] R. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 35–45.

[9] P. R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, NJ, 1986.

[10] B. Legras and R. Vautard, *A guide to Lyapunov vectors*, in Proceedings of ECWF Seminar, Vol. 1, T. Palmer, ed., 1996, pp. 135–146.

[11] E. N. Lorenz, *Deterministic non-periodic flow*, J. Atmos. Sci., 20 (1963), pp. 130–141.

[12] J. K. Merikoshi and R. Kumar, *Inequalities for spreads of matrix sums and products*, Appl. Math. E-Notes, 4 (2004), pp. 150–159.

[13] R. N. Miller, M. Ghil, and F. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmos. Sci., 51 (1994), pp. 1037–1056.

[14] B. Ni and Q. Zhang, *Stability of the Kalman filter for continuous time output error systems*, Systems Control Lett., 94 (2016), pp. 172–180.

[15] V. I. Oseledets, *Multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems*, Trans. Moscow Math. Soc., 19 (1968), pp. 197–231.

[16] L. Palatella, A. Carrassi, and A. Trevisan, *Lyapunov vectors and assimilation in the unstable subspace: Theory and applications*, J. Phys. A, 46 (2013), 254020.

[17] L. Palatella, A. Trevisan, and S. Rambaldi, *Nonlinear stability of traffic models and the use of Lyapunov vectors for estimating the traffic state*, Phys. Rev. E, 88 (2013), 022901.

[18] C. Pires, R. Vautard, and O. Talagrand, *On extending the limits of variational assimilation in nonlinear chaotic systems*, Tellus A, 48 (1996), pp. 96–121.

[19] D. Sanz-Alonso and A. M. Stuart, *Long-Time Asymptotics of the Filtering Distribution for Partially Observed Chaotic Dynamical Systems*, arXiv:1411.6510, 2014.

[20] O. Talagrand, *Assimilation of observations, an introduction*, J. Meteor. Soc. Japan, 75 (1997), pp. 191–209.

[21] A. Trevisan, M. D'Isidoro, and O. Talagrand, *Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension*, Quart. J. Roy. Meteor. Soc., 2010 (2010), pp. 487–496.

[22] A. Trevisan and L. Palatella, *On the Kalman filter error covariance collapse into the unstable subspace*, Nonlinear Process. Geophys., 18 (2011), pp. 243–250.

[23] A. Trevisan and F. Uboldi, *Assimilation of standard and targeted observations within the unstable subspace of the observation-analysis-forecast cycle*, J. Atmos. Sci., 61 (2004), pp. 103–113.

[24] F. Uboldi and A. Trevisan, *Detecting unstable structures and controlling error growth by assimilation of standard and adaptive observations in a primitive equation ocean model*, Nonlinear Process. Geophys., 16 (2006), pp. 67–81.

[25] M. P. Wojtowski, *Geometry of Kalman filters*, J. Geom. Symmetry Phys., 9 (2007), pp. 83–95.