University of
Reading

Degenerate Kalman filter error covariances and their convergence onto the unstable subspace

Article

Accepted Version

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Degenerate Kalman filter error covariances and their convergence onto the unstable subspace

MARC BOCQUET*, KARTHIK S. GURUMOORTHY†, AMIT APTE†, ALBERTO CARRASSI‡, COLIN GRUDZIEN§, AND CHRISTOPHER K. R. T. JONES§

Abstract. The characteristics of the model dynamics are critical in the performance of (ensemble) Kalman filters. In particular, as emphasized in the seminal work of Anna Trevisan and co-authors, the error covariance matrix is asymptotically supported by the unstable-neutral subspace only, *i.e.* it is spanned by the backward Lyapunov vectors with non-negative exponents. This behavior is at the core of algorithms known as Assimilation in the Unstable Subspace, although its formal proof was still missing.

This paper provides the analytical proof of the convergence of the Kalman filter covariance matrix onto the unstable-neutral subspace when the dynamics and the observation operator are linear and when the dynamical model is error-free, for any, possibly rank-deficient, initial error covariance matrix. The rate of convergence is provided as well. The derivation is based on an expression that explicitly relates the error covariances at an arbitrary time to the initial ones. It is also shown that if the system is sufficiently observed and if the column space of the initial covariance matrix has a non-zero projection onto all of the forward Lyapunov vectors associated to the unstable and neutral directions of the dynamics, the covariance matrix of the Kalman filter collapses onto an asymptotic sequence which is independent of the initial covariances. Numerical results are also shown to illustrate and support the theoretical findings.

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

AMS subject classifications. 93E11, 93C05, 93B07, 60G35, 15A03

1. Introduction.

1.1. Context and objectives. Filtering methods are the techniques of estimation theory that process measurements sequentially as they become available. In a probabilistic Bayesian framework, the output of a filter is a probability density function (pdf), the conditional posterior pdf $p(\mathbf{x}|\mathbf{y})$, of the process $\mathbf{x}$, given the data $\mathbf{y}$ and a prior distribution $p(\mathbf{x})$. The posterior pdf fully characterizes the state's estimation and quantifies the uncertainty of the estimate. However, its exact calculation is extremely difficult in practice, and most often computationally intractable in high-dimensional, complex systems, such as numerical climate and weather models.

For linear dynamics, measurements with a linear dependence on the state variables, and Gaussian errors, the Kalman filter (KF) is the optimal filtering solution [13]. The Gaussian hypothesis implies an enormous simplification: the pdfs are all completely characterized by their first and second moments. In this case, the error covariance matrix quantifies the uncertainty of the state's estimate represented by the mean. The KF has been extremely successful for decades in numerous fields including navigation, economy, robotics, tracking objects, adaptive optics and many computer vision applications.

A Monte Carlo formulation of the KF leads to the introduction of a class of *Gaussian* algorithms referred to as ensemble Kalman filters (EnKF) [9]. They have been widely applied in atmospheric and oceanic contexts, where all methods designed

for filtering or smoothing are referred to as data assimilation (DA). In the EnKF the transition probability of the process, as well as all the error covariances entering the assimilation of observations, are approximated using an ensemble of realizations (members in the EnKF jargon) of the model dynamics. The EnKF and its variants are currently among the most popular approaches for DA in high-dimensional systems. Evidence has emerged that a small number of members, typically 100, is sufficient in many applications, especially when using localization techniques [21, and references therein], hence making the EnKF feasible in situations where the forward step of DA is computationally expensive. The choice of the ensemble members is critical and a key aspect in the EnKF setup. While a large ensemble is generally desirable to explain and represent the actual uncertainty in the most realistic manner, their number is limited by the computational resources at disposal. In the absence of localization, the EnKF error covariances are thus degenerate (or rank-deficient) by construction and it is then relevant to adequately choose these few (much fewer than the system's dimension) members so as to maximize the representation of the actual uncertainty.

For nonlinear chaotic dynamics, the assimilation in the unstable subspace (AUS), introduced by Anna Trevisan and collaborators [26, 6, 23, 24, 18], has shed light on an efficient way to operate the assimilation of observations by using the unstable subspace to describe the uncertainty in the estimate. AUS is based on two key properties of deterministic, typically dissipative, chaotic systems: (i) the perturbations tend to project on the unstable manifold of the dynamics, and (ii) the dimension of the unstable manifold is typically much smaller than the full phase-space dimension. Applications to atmospheric, oceanic, and traffic models [7, 27, 19] showed that even in high-dimensional systems, an efficient error control is achieved by monitoring only the unstable directions, and sometimes even a subset of them, making AUS a computationally efficient alternative to standard procedures.

The AUS approach has recently motivated a research effort toward a proper mathematical formulation and assessment of its driving idea, *i.e.* the span of the estimation error covariance matrices asymptotically (in time) tends to the subspace spanned by the unstable and neutral backward Lyapunov vectors. A proper statement of this latter property in precise mathematical terms is of vast importance for the design of efficient reduced-order uncertainty quantification and DA methods.

The first recent result along this line is given in [11]. It is proved that for linear, discrete, autonomous and non-autonomous, deterministic system (perfect model) with noisy observations, the covariance equations in the KF asymptotically bounds the rank of the forecast and the analysis error covariance matrices to be less than or equal to the number of non-negative 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 and neutral backward Lyapunov vectors. The results in [11] were obtained assuming a full rank covariance matrix at initial time. The conditions that imply the convergence, for possibly degenerate (low rank) initial matrices remained unresolved, yet they are fundamental to link these mathematical findings with concrete reduced-rank DA methods, particularly the EnKF.

This is the subject of the present work, which studies the convergence in the general setting of degenerate covariance matrices. The main result is the analytic proof of the KF covariance collapse, for any initial error covariance (of arbitrary rank), onto the unstable-neutral subspace. We also provide rigorous mathematical results for the rate of convergence on the stable subspace, for the asymptotic rank of the error covariance matrix and, under an observability condition, provide an upper bound for

the error covariance matrix as the minimum between the unconstrained propagated initial error covariance and the observational error. Finally, a more heuristic approach is used to derive an expression for the asymptotic error covariance matrix as a function of the initial one. This in turn allows us to prove, under certain observability conditions, the existence of an asymptotic sequence of error covariance matrix independent of the initial condition.

In the following, we set up the notations and discuss the organization of the paper.

1.2. Problem formulation. The purpose is the estimation of the unknown state of a system based on partial and noisy observations. The dynamical and observational models are both assumed to be linear, and expressible as

$$\mathbf{x}_k = \mathbf{M}_k \mathbf{x}_{k-1} + \mathbf{w}_k, \tag{1.1}$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{1.2}$$

with $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^d$ being the system's state and observation respectively, related via the linear observation operator $\mathbf{H}_k : \mathbb{R}^n \mapsto \mathbb{R}^d$. Throughout the entire text the conventional notation $k = 0, 1, 2, \ldots$ is adopted to indicate that the quantity is defined at time t_k. The matrix $\mathbf{M}_{k:l}$ is taken to represent the resolvent of the linear forward model from time t_l to time t_k, and is assumed to be non-singular throughout this paper. In particular $\mathbf{M}_{k:k} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix (of size $n \times n$ in this case). Also, we designate $\mathbf{M}_k$ as the 1-step matrix resolvent of the forward model from t_{k-1} to t_k: $\mathbf{M}_k \triangleq \mathbf{M}_{k:k-1}$ and, consequently $\mathbf{M}_{k:l} = \prod_{i=l+1}^{k} \mathbf{M}_i$, with the symbol $\triangleq$ used to signify that the expression is a definition. We will assume that the Lyapunov spectrum of the dynamics defined by $\mathbf{M}_{k:0}$ is non-degenerate, *i.e.* the Lyapunov exponents are all distinct. This assumption substantially simplifies the derivations that follow. Nonetheless, most of the results in this paper can be generalized to the degenerate case.

The model and observation noise, $\mathbf{w}_k$ and $\mathbf{v}_k$, are assumed mutually independent, unbiased Gaussian white sequences with statistics

$$\mathrm{E}[\mathbf{v}_k \mathbf{v}_l^{\mathrm{T}}] = \delta_{k,l} \mathbf{R}_k, \quad \mathrm{E}[\mathbf{w}_k \mathbf{w}_l^{\mathrm{T}}] = \delta_{k,l} \mathbf{Q}_k, \quad \mathrm{E}[\mathbf{v}_k \mathbf{w}_l^{\mathrm{T}}] = \mathbf{0}, \tag{1.3}$$

where $\mathbf{R}_k \in \mathbb{R}^{d \times d}$ is the observation error covariance matrix at time t_k, and $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$ stands for the model error covariance matrix. The error covariance matrix $\mathbf{R}_k$ can be assumed invertible without loosing generality.

The forecast error covariance matrix $\mathbf{P}_k$ of the Kalman filter satisfies the following recurrence equation [13, 11]

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1} \left(\mathbf{I} + \mathbf{P}_k \mathbf{\Omega}_k \right)^{-1} \mathbf{P}_k \mathbf{M}_{k+1}^{\mathrm{T}} + \mathbf{Q}_{k+1}, \tag{1.4}$$

where

$$\mathbf{\Omega}_k \triangleq \mathbf{H}_k^{\mathrm{T}} \mathbf{R}_k^{-1} \mathbf{H}_k \tag{1.5}$$

is the precision matrix of the observations transferred in state space. To avoid pathological behaviors, we will assume in this paper that the $\{\mathbf{\Omega}_k\}_{k=0,1,\ldots}$ are uniformly bounded from above, which is a very mild hypothesis.

Equation (1.4) highlights that the error covariance matrix, $\mathbf{P}_{k+1}$, is the result of a two-step process, consisting of the *update* or *analysis* step at time t_k leading to the analysis error covariance matrix $\mathbf{P}_k^{\mathrm{a}}$,

$$\mathbf{P}_k^{\mathrm{a}} = \left(\mathbf{I} + \mathbf{P}_k \mathbf{\Omega}_k \right)^{-1} \mathbf{P}_k, \tag{1.6}$$

and the *forecast* step which consists of the forward propagation of the analysis error covariance,

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1}\mathbf{P}_k^a\mathbf{M}_{k+1}^T + \mathbf{Q}_{k+1}. \tag{1.7}$$

It is worth mentioning that Eq. (1.4) still holds when $\mathbf{P}_k$ is degenerate, *i.e.* we have $\text{rank}(\mathbf{P}_k) < n$. This is the typical circumstance encountered in the ensemble Kalman filter [9, and references therein]. In this case, assuming that the model is perfect ($\mathbf{Q}_k = \mathbf{0}$) and under the same hypotheses of linear observation and evolution operators as well as of Gaussian statistics for the initial condition and observational errors, Eq. (1.4) will apply to the ensemble Kalman filter too.

1.3. Outline of the paper. In the rest of the paper, we will refer to Eq. (1.4) as the recurrence equation for $\mathbf{P}_k$, although we will mostly study the *perfect dynamical model case*, in which $\mathbf{Q}_k = \mathbf{0}$. In section 2 we demonstrate a relation between $\mathbf{P}_k$ at any arbitrary time, $t_k > t_0$, and the initial error covariance matrix $\mathbf{P}_0$, in the general case with $\mathbf{P}_0$ possibly being degenerate. An alternative proof based on the linear symplectic representation of the KF is proposed in appendix A. In the following section 3, we derive a useful bound that plays a central role in all results and derivations discussed in this study. Then in section 4 we study the asymptotic behavior of $\mathbf{P}_k$ (for $k \to \infty$) along with other relevant properties. Section 5 provides a heuristic proof, using a condition on the initial $\mathbf{P}_0$ and certain observability conditions, that the error covariances collapse onto an asymptotic sequence which is independent of the initial covariance matrix $\mathbf{P}_0$. Section 6 describes the numerical results corroborating and illustrating the theoretical findings while the conclusions are drawn in section 7.

2. Computation of the forecast error covariance matrix $\mathbf{P}_k$. In this section, we consider the perfect model case, *i.e.* $\mathbf{Q}_k = \mathbf{0}$ for all k. The stochastic model case, $\mathbf{Q}_k \neq \mathbf{0}$, is briefly considered in section 3.

The recurrence Eq. (1.4) is rational in $\mathbf{P}_k$. Furthermore if we assume that the $\mathbf{P}_k$ are invertible, we can take the inverse of both sides of the recurrence and obtain

$$\mathbf{P}_{k+1}^{-1} = \mathbf{M}_{k+1}^{-T}\left(\mathbf{P}_k^{-1} + \mathbf{\Omega}_k\right)\mathbf{M}_{k+1}^{-1}, \tag{2.1}$$

which shows that $\mathbf{P}_{k+1}^{-1}$ is an affine function of $\mathbf{P}_k^{-1}$. This relation is usually called the *information filter* [22, section 3.2].

However, a relevant situation in applications is when the $\mathbf{P}_k$ are degenerate. In this case the inverse of both sides of Eq. (2.1) are undefined, and a suitable generalization of Eq. (2.1) is required. To that end, we introduce an analytic continuation of Eq. (1.4). A regularized $\mathbf{P}_0$ is defined as

$$\mathbf{P}_0(\varepsilon) \triangleq \mathbf{P}_0 + \varepsilon\mathbf{I} \tag{2.2}$$

with $\varepsilon > 0$ and we define the subsequent $\mathbf{P}_k(\varepsilon)$ via the recurrence

$$\mathbf{P}_{k+1}(\varepsilon) \triangleq \mathbf{M}_{k+1}\left(\mathbf{I} + \mathbf{P}_k(\varepsilon)\mathbf{\Omega}_k\right)^{-1}\mathbf{P}_k(\varepsilon)\mathbf{M}_{k+1}^T. \tag{2.3}$$

From Eqs. (2.2,2.3), $\mathbf{P}_k(\varepsilon)$ is seen to be full-rank. Moreover, taking the limit $\varepsilon \to 0^+$, leads continuously back to $\mathbf{P}_0$ and Eq. (1.4), so that we have

$$\lim_{\varepsilon \to 0^+} \mathbf{P}_k(\varepsilon) = \mathbf{P}_k(0) = \mathbf{P}_k. \tag{2.4}$$

4

Then, we take the inverse of both sides of Eq. (2.3),

$$\mathbf{P}_{k+1}^{-1}(\varepsilon) = \mathbf{M}_{k+1}^{-\mathrm{T}}\mathbf{P}_k^{-1}(\varepsilon)\left(\mathbf{I} + \mathbf{P}_k(\varepsilon)\mathbf{\Omega}_k\right)\mathbf{M}_{k+1}^{-1}$$
$$= \mathbf{M}_{k+1}^{-\mathrm{T}}\mathbf{P}_k^{-1}(\varepsilon)\mathbf{M}_{k+1}^{-1} + \mathbf{M}_{k+1}^{-\mathrm{T}}\mathbf{\Omega}_k\mathbf{M}_{k+1}^{-1}. \tag{2.5}$$

This recurrence can be easily solved and it yields

$$\mathbf{P}_k^{-1}(\varepsilon) = \mathbf{M}_{k:0}^{-\mathrm{T}}\mathbf{P}_0^{-1}(\varepsilon)\mathbf{M}_{k:0}^{-1} + \mathbf{\Gamma}_k, \tag{2.6}$$

where

$$\mathbf{\Gamma}_k \triangleq \sum_{l=0}^{k-1}\mathbf{M}_{k:l}^{-\mathrm{T}}\mathbf{\Omega}_l\mathbf{M}_{k:l}^{-1}. \tag{2.7}$$

This matrix is a measure of the observability of the system since it propagates the precision matrices $\mathbf{\Omega}_l$ up to t_k, and Eq. (2.6) states that the precision in the state estimate is the sum of the forecast precision in the initial condition plus the precision of the observations transferred into the model space.

Let us now recall the partial order defined in the cone $\mathcal{C}^n$ of the symmetric positive semi-definite matrices in $\mathbb{R}^n$, of which we will do large use in this study. Similarly the partial order acts in the cone $\mathcal{C}_+^n$ of the symmetric positive definite matrices in $\mathbb{R}^n$. We will refer to this partial order using the standard comparison symbols. In appendix B, its definition is provided along with some additional properties that we rely on in this study. From Eq. (2.6) and using this partial order, we have

$$\mathbf{P}_k^{-1}(\varepsilon) \geq \mathbf{\Gamma}_k. \tag{2.8}$$

Let us assume that the system is observable, a condition defined here as $\det(\mathbf{\Gamma}_k) \neq 0$ according to [12] and references therein. This yields $\mathbf{P}_k(\varepsilon) \leq \mathbf{\Gamma}_k^{-1}$ (see appendix B, point 3), which implies

$$\mathbf{P}_k \leq \mathbf{\Gamma}_k^{-1}. \tag{2.9}$$

By taking the inverse of both sides of Eq. (2.6) we have

$$\mathbf{P}_k(\varepsilon) = \left(\mathbf{M}_{k:0}^{-\mathrm{T}}\mathbf{P}_0^{-1}(\varepsilon)\mathbf{M}_{k:0}^{-1} + \mathbf{\Gamma}_k\right)^{-1}$$
$$= \left[\left(\mathbf{I} + \mathbf{\Gamma}_k\mathbf{M}_{k:0}\mathbf{P}_0(\varepsilon)\mathbf{M}_{k:0}^{\mathrm{T}}\right)\mathbf{M}_{k:0}^{-\mathrm{T}}\mathbf{P}_0^{-1}(\varepsilon)\mathbf{M}_{k:0}^{-1}\right]^{-1}$$
$$= \mathbf{M}_{k:0}\mathbf{P}_0(\varepsilon)\mathbf{M}_{k:0}^{\mathrm{T}}\left(\mathbf{I} + \mathbf{\Gamma}_k\mathbf{M}_{k:0}\mathbf{P}_0(\varepsilon)\mathbf{M}_{k:0}^{\mathrm{T}}\right)^{-1}. \tag{2.10}$$

The limit $\varepsilon \to 0^+$ finally leads to

$$\mathbf{P}_k = \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}\left(\mathbf{I} + \mathbf{\Gamma}_k\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}\right)^{-1}. \tag{2.11}$$

Equation (2.11) is extremely important as it directly relates $\mathbf{P}_k$ to $\mathbf{P}_0$. In particular it shows that $\mathbf{P}_k$ depends on two concurring factors, the matrix $\mathbf{\Gamma}_k$ encoding all information about the observability of the system, and the matrix $\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}$ representing the free forecast of the initial covariances. The latter exemplifies the uncertainty propagation under the model dynamics, the former the ability of the observations to counteract the error growth.

We now use the matrix shift lemma that asserts that for any matrices $\mathbf{A} \in \mathbb{R}^{l\times m}$ and $\mathbf{B} \in \mathbb{R}^{m\times l}$, we have $\mathbf{A}f(\mathbf{BA}) = f(\mathbf{AB})\mathbf{A}$, with $x \mapsto f(x)$ being any function

that can be expressed as a formal power series. A derivation is recalled in appendix C. Here, we choose $f(x) = (1+x)^{-1}$, $\mathbf{A} = \mathbf{M}_{k:0}^{\mathrm{T}}$ and $\mathbf{B} = \boldsymbol{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0$, to obtain an alternative formulation of Eq. (2.11)

$$\mathbf{P}_k = \mathbf{M}_{k:0} \mathbf{P}_0 \left[\mathbf{I} + \mathbf{M}_{k:0}^{\mathrm{T}} \boldsymbol{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0 \right]^{-1} \mathbf{M}_{k:0}^{\mathrm{T}} \tag{2.12}$$

or, in a more condensed form,

$$\mathbf{P}_k = \mathbf{M}_{k:0} \mathbf{P}_0 \left[\mathbf{I} + \boldsymbol{\Theta}_k \mathbf{P}_0 \right]^{-1} \mathbf{M}_{k:0}^{\mathrm{T}} \tag{2.13}$$

where

$$\boldsymbol{\Theta}_k \triangleq \mathbf{M}_{k:0}^{\mathrm{T}} \boldsymbol{\Gamma}_k \mathbf{M}_{k:0} = \sum_{l=0}^{k-1} \mathbf{M}_{l:0}^{\mathrm{T}} \boldsymbol{\Omega}_l \mathbf{M}_{l:0}. \tag{2.14}$$

This matrix, known as the information matrix [12], is also related to the observability of the system but pulled back at the initial time t_0.

A more general, albeit less straightforward, proof of the expressions for $\mathbf{P}_k$ as a function of $\mathbf{P}_0$ can be obtained using the underlying symplectic structure of the Kalman filter and is described in appendix A.

3. Free forecast of $\mathbf{P}_0$ as an upper bound. In this section we demonstrate that an upper bound for $\mathbf{P}_k$ is given by the *free* forecast of $\mathbf{P}_0$; the term free is used in this study to mean without the observational forcing applied at analysis times. It is worth mentioning already that, although the existence of this bound is indeed very intuitive, its formal proof is provided here because it plays a pivotal role in all the convergence results that follow. The bound can be derived directly from the general expression for $\mathbf{P}_k$, Eq. (2.13), but we opted for showing a different approach, independent from Eq. (2.13), that better highlights the relevance of the bound for the results that follow.

The error covariance matrix $\mathbf{P}_k$ is symmetric and our purpose is to make the recurrence equation look patently symmetric as well so that we can derive inequalities using the partial ordering in $\mathcal{C}^n$. As a positive semi-definite matrix, $\mathbf{P}_k$ can be decomposed into $\mathbf{P}_k = \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}}$ using, for instance, a Choleski decomposition, with $\mathbf{X}_k \in \mathbb{R}^{n \times m}$ ($m \leq n$). Here, as opposed to the rest of the paper and for the sake of generality, we consider the presence of model noise given that it only represents a minor complication. The recurrence equation can be written as

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1} \left(\mathbf{I} + \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} \boldsymbol{\Omega}_k \right)^{-1} \mathbf{X}_k \mathbf{X}_k^{\mathrm{T}} \mathbf{M}_{k+1}^{\mathrm{T}} + \mathbf{Q}_{k+1}. \tag{3.1}$$

We use again the matrix shift lemma but this time with $f(x) = (1+x)^{-1}$, $\mathbf{A} = \mathbf{X}_k$ and $\mathbf{B} = \mathbf{X}_k^{\mathrm{T}} \boldsymbol{\Omega}_k$ so that Eq. (1.4) becomes

$$\mathbf{P}_{k+1} = \mathbf{M}_{k+1} \mathbf{X}_k \left(\mathbf{I} + \mathbf{X}_k^{\mathrm{T}} \boldsymbol{\Omega}_k \mathbf{X}_k \right)^{-1} \mathbf{X}_k^{\mathrm{T}} \mathbf{M}_{k+1}^{\mathrm{T}} + \mathbf{Q}_{k+1}. \tag{3.2}$$

Using the partial order in $\mathcal{C}^n$, we have from

$$\left(\mathbf{I} + \mathbf{X}_k^{\mathrm{T}} \boldsymbol{\Omega}_k \mathbf{X}_k \right)^{-1} \leq \mathbf{I} \tag{3.3}$$

and from Eq. (3.2) that

$$\mathbf{P}_{k+1} \leq \mathbf{M}_{k+1} \mathbf{P}_k \mathbf{M}_{k+1}^{\mathrm{T}} + \mathbf{Q}_{k+1}. \tag{3.4}$$

Hence $\mathbf{P}_k$ is bounded by the free forecast $\widetilde{\mathbf{P}}_k$ which satisfies the recurrence

$$\widetilde{\mathbf{P}}_{k+1} = \mathbf{M}_{k+1}\widetilde{\mathbf{P}}_k\mathbf{M}_{k+1}^{\mathrm{T}} + \mathbf{Q}_{k+1} \tag{3.5}$$

whose solution is for $k \geq 0$

$$\widetilde{\mathbf{P}}_k = \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}} + \boldsymbol{\Xi}_k \tag{3.6}$$

where

$$\boldsymbol{\Xi}_0 \triangleq \mathbf{0} \quad \text{and for } k \geq 1 \quad \boldsymbol{\Xi}_k \triangleq \sum_{l=1}^{k} \mathbf{M}_{k:l}\mathbf{Q}_l\mathbf{M}_{k:l}^{\mathrm{T}} \tag{3.7}$$

is known as the controllability matrix [12]. Therefore

$$\mathbf{P}_k \leq \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}} + \boldsymbol{\Xi}_k. \tag{3.8}$$

In particular, in the perfect model case, we obtain the pivotal inequality

$$\mathbf{P}_k \leq \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}. \tag{3.9}$$

Under the aforementioned assumptions on linear dynamical and observational models and Gaussian error statistics, the inequalities Eq. (3.8) and Eq. (3.9) state that data assimilation will always reduce, and in the worst case left unchanged, the state's estimate uncertainty with respect to the free run. This effect was already discussed in the context of nonlinear dynamics and in relation to the stability properties of data assimilation systems in [6], although an analytic proof in the nonlinear case is not provided either in that work or in the present one.

4. Convergence of the error covariance matrix: theoretical results. This section describes some of the implications of the recurrence equation and bounds described in the previous section that are relevant for the design of reduced-order formulations of the Kalman filter with unstable dynamics. We will assume here again to be in the perfect model scenario, $\mathbf{Q}_k = \mathbf{0}$.

4.1. Rank of $\mathbf{P}_k$. From the inequality Eq. (3.9), it is clear that the column space of $\mathbf{P}_k$, *i.e.* the subspace $\mathrm{Im}(\mathbf{P}_k) = \{\mathbf{P}_k\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\}$ satisfies

$$\mathrm{Im}(\mathbf{P}_k) \subseteq \mathbf{M}_{k:0}\left(\mathrm{Im}(\mathbf{P}_0)\right). \tag{4.1}$$

Moreover, since from Eq. (1.4) (with $\mathbf{Q}_k = 0$)

$$\mathrm{rank}(\mathbf{P}_{k+1}) = \mathrm{rank}(\mathbf{P}_k), \tag{4.2}$$

we infer that

$$\mathrm{Im}(\mathbf{P}_k) = \mathbf{M}_{k:0}\left(\mathrm{Im}(\mathbf{P}_0)\right). \tag{4.3}$$

The morale is that the Kalman filter merely operates within the subspaces of the sequence $\mathbf{M}_{k:0}\left(\mathrm{Im}(\mathbf{P}_0)\right)$, which does not depend on the observations. In the absence of model error, the rank of $\mathbf{P}_k$ cannot exceed that of $\mathbf{P}_0$ even if the dynamics is degenerate.

4.2. Collapse of the error covariance matrices onto the unstable-neutral subspace. The unstable-neutral subspace is defined as the subspace $\mathcal{U}_k$ spanned by the n_0 backward Lyapunov vectors at t_k whose exponents, λ_i with $i = 1, ..., n_0$, are non-negative. The stable subspace $\mathcal{S}_k$ is defined as the subspace spanned by the $n-n_0$ backward Lyapunov vectors at t_k associated with negative exponents. The inequality Eq. (3.9), $\mathbf{P}_k \leq \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}$, provides the convergence onto $\mathcal{U}_k$ in a sense that is made clear below. It also gives the rate of such convergence as shown in section 4.3.

Let us write the singular value decomposition (SVD) of $\mathbf{M}_{k:0} = \mathbf{U}_{k:0}\mathbf{\Sigma}_{k:0}\mathbf{V}_{k:0}^{\mathrm{T}}$, where $\mathbf{U}_{k:0}$ and $\mathbf{V}_{k:0}$ are both orthogonal matrices in $\mathbb{R}^{n \times n}$, and $\mathbf{\Sigma}_{k:0}$ in $\mathcal{C}_+^n$ is the diagonal matrix of the singular values. The left singular vectors are the columns of $\mathbf{U}_{k:0} = [\mathbf{u}_1^{k:0}, \dots, \mathbf{u}_n^{k:0}]$ and when $k \to \infty$, they converge to the backward Lyapunov vectors defined at t_k, denoted here as $\mathbf{u}_i^k$. The right singular vectors are the columns of $\mathbf{V}_{k:0} = [\mathbf{v}_1^{k:0}, \dots, \mathbf{v}_n^{k:0}]$ which converge to the forward Lyapunov vectors at time t_0 as $k \to \infty$ denoted here as $\mathbf{v}_i^0$ [14, 25]. Let us write $[\mathbf{\Sigma}_{k:0}]_{i,i} = \exp(\lambda_i^k k)$, with λ_i^k being real numbers and for large k ordered as $\lambda_1^k > \cdots > \lambda_{n_0}^k \geq 0 > \lambda_{n_0+1}^k > \cdots > \lambda_n^k$, which is justified by the non-degeneracy hypothesis on the Lyapunov spectrum. Using the SVD we have

$$\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}} = \mathbf{U}_{k:0}\mathbf{\Sigma}_{k:0}\mathbf{V}_{k:0}^{\mathrm{T}}\mathbf{P}_0\mathbf{V}_{k:0}\mathbf{\Sigma}_{k:0}\mathbf{U}_{k:0}^{\mathrm{T}}. \tag{4.4}$$

Define $\mathfrak{S}_k$ as the set of indices i for which $\lambda_i^k < 0$, and $\mathfrak{S}_k^s$ to be the set of indices corresponding to the s smallest singular values in $\mathbf{\Sigma}_{k:0}$. Note also that $\mathfrak{S}_k^{n-n_0} = \mathfrak{S}_k$ for large k. Let the subspace $\mathcal{S}_{k:0}^s$ be the span of the left singular vectors $\mathbf{u}_i^{k:0}$ where $i \in \mathfrak{S}_k^s$. Let $\Pi_{\mathcal{S}_{k:0}^s}$ be the orthogonal projector onto $\mathcal{S}_{k:0}^s$ which, owing to the orthonormality of the left singular vectors, reads

$$\Pi_{\mathcal{S}_{k:0}^s} = \sum_{i \in \mathfrak{S}_k^s} \mathbf{u}_i^{k:0} \left(\mathbf{u}_i^{k:0}\right)^{\mathrm{T}}. \tag{4.5}$$

For large enough k, $\mathfrak{S}_k$ gets progressively closer and eventually coincides with the set $\mathfrak{S}$ of indices i for which $\lambda_i < 0$ and each of its subset $\mathfrak{S}_k^s$ approaches its corresponding subset $\mathfrak{S}^s$ defined similarly. Note that $\mathfrak{S}^{n-n_0} = \mathfrak{S}$. Furthermore, the subspace $\mathcal{S}_{k:0}^s$ converges to $\mathcal{S}_k^s$ which is the span of the s most stable backward Lyapunov vectors.

We are now interested in an upper bound in $\mathcal{C}^n$ for $\left(\mathbf{V}_{k:0}^s\right)^{\mathrm{T}}\mathbf{P}_0\mathbf{V}_{k:0}^s$, with $\mathbf{V}_{k:0}^s = [\mathbf{v}_{n-s+1}^{k:0}, \dots, \mathbf{v}_n^{k:0}]$ to be jointly used with Eq. (4.4). For this purpose we define

$$\alpha_s^k = \max_{\mathbf{h} \in \mathrm{Im}\left(\mathbf{V}_{k:0}^s\right), \|\mathbf{h}\|=1} \mathbf{h}^{\mathrm{T}}\mathbf{P}_0\mathbf{h}, \tag{4.6}$$

where $\|.\|$ denotes the Euclidean norm. As a consequence, we have

$$\left(\mathbf{V}_{k:0}^s\right)^{\mathrm{T}}\mathbf{P}_0\mathbf{V}_{k:0}^s \leq \alpha_s^k\mathbf{I}. \tag{4.7}$$

From this inequality and from Eq. (4.4), we infer

$$\Pi_{\mathcal{S}_{k:0}^s}\mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}\Pi_{\mathcal{S}_{k:0}^s} \leq \alpha_s^k\Pi_{\mathcal{S}_{k:0}^s}\mathbf{U}_{k:0}\mathbf{\Sigma}_{k:0}^2\mathbf{U}_{k:0}^{\mathrm{T}}\Pi_{\mathcal{S}_{k:0}^s}. \tag{4.8}$$

Note that, if σ_1^0 is the largest eigenvalue of $\mathbf{P}_0$, we have the uniform bound $\alpha_s^k \leq \sigma_1^0$ for any k and s (see appendix B, point 5). Hence, we can define a finite bound

$$\alpha_s = \sup_{k \geq 0} \alpha_s^k \tag{4.9}$$

8

which satisfies for any k and s: $\alpha_s^k \leq \alpha_s \leq \sigma_1^0$. Using this uniform bound, in conjunction with Eq. (3.9) and Eq. (4.8), we obtain

$$\Pi_{\mathcal{S}_{k:0}^s} \mathbf{P}_k \Pi_{\mathcal{S}_{k:0}^s} \leq \alpha_s \sum_{i \in \mathfrak{S}_k^s} \exp\left(2\lambda_i^k k\right) \mathbf{u}_i^{k:0} \left(\mathbf{u}_i^{k:0}\right)^{\mathrm{T}}. \tag{4.10}$$

Hence, for every unit vector $\mathbf{h} \in \mathcal{S}_{k:0}^s$

$$\mathbf{h}^{\mathrm{T}} \mathbf{P}_k \mathbf{h} \leq \alpha_s \exp\left(2\lambda_{n-s+1}^k k\right). \tag{4.11}$$

In particular, if $i \in \mathfrak{S}$ then $\left(\mathbf{u}_i^{k:0}\right)^{\mathrm{T}} \mathbf{P}_k \mathbf{u}_i^{k:0} \to 0^+$ as $k \to \infty$. This defines a *weak* form of collapse of $\mathbf{P}_k$ onto the unstable-neutral subspace $\mathcal{U}_k$. A *strong* form of collapse is defined by the stable subspace $\mathcal{S}_k$ being in the null space of $\mathbf{P}_k$. This can be obtained under the hypothesis that $\mathbf{P}_k$ is uniformly bounded, which can in turn be satisfied if the system is sufficiently observed. Indeed, if $\mathbf{P}_k$ is uniformly bounded and owing to its positive semi-definiteness, it can be shown that $\left\|\mathbf{P}_k \mathbf{u}_i^{k:0}\right\| \to 0$ as $k \to \infty$ (see appendix B, point 5). Hence, asymptotically, the stable subspace $\mathcal{S}_k$ is in the null space of $\mathbf{P}_k$. As described in the introduction, this property is at the core of the class of data assimilation algorithms referred to as *assimilation in the unstable subspace* [17, and references therein].

4.3. Rate of convergence of the eigenvalues. In the case of weak – a fortiori strong – collapse, the rate of convergence of each of the eigenvalues of $\mathbf{P}_k$ can be determined from Eq. (4.11) as follows. Let σ_i^k, for $i = 1, \dots, n$ denote the eigenvalues of $\mathbf{P}_k$ ordered as $\sigma_1^k \geq \sigma_2^k \cdots \geq \sigma_n^k$. Equation (4.11) guarantees that (appendix B, point 6) $\mathbf{P}_k$ has at least s of its eigenvalues less than or equal to $\alpha_s \exp\left(2\lambda_{n-s+1}^k k\right)$. It follows that

$$\sigma_i^k \leq \alpha_i \exp\left(2\lambda_i^k k\right) \tag{4.12}$$

which gives us an upper bound for all eigenvalues of $\mathbf{P}_k$ and a rate of convergence for the $n - n_0$ smallest ones.

4.4. Asymptotic rank of the error covariance matrix. A consequence of Eq. (4.10) is the upper bound of the asymptotic rank of the error covariance matrix $\mathbf{P}_k$. In fact, the asymptotic rank of $\mathbf{P}_k$ is bounded by the minimum between the rank of $\mathbf{P}_0$ and n_0. This mathematically reads

$$\lim_{k \to \infty} \operatorname{rank}(\mathbf{P}_k) \leq \min\left\{\operatorname{rank}(\mathbf{P}_0), n_0\right\}. \tag{4.13}$$

4.5. Observability and boundedness of the error statistics. As mentioned in section 2, we define the system to be observable if $\det(\mathbf{\Gamma}_k) \neq 0$ or, equivalently, given that $\mathbf{M}_{k:0}$ is assumed to be non-singular, $\det(\mathbf{\Theta}_k) \neq 0$ [12]. If the system is observable, the inequalities Eq. (3.9) and Eq. (2.9) can be combined to obtain

$$\mathbf{P}_k \leq \min\left\{\mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^{\mathrm{T}}, \mathbf{\Gamma}_k^{-1}\right\}, \tag{4.14}$$

where the accurate definition of the minimum is given in appendix B (point 4). We note that if $\mathbf{\Gamma}_k$ is bounded by $\mathbf{L}$ in $\mathcal{C}_+^n$, $\mathbf{\Gamma}_k \geq \mathbf{L}$, we have $\mathbf{P}_k \leq \mathbf{\Gamma}_k^{-1} \leq \mathbf{L}^{-1}$ which bounds the error covariances. The existence of the bound $\mathbf{L}$ in $\mathcal{C}_+^n$ guarantees the observability of the system; it forces the precision of the observations to be spread in space and time. Interestingly, the inequality Eq. (4.14) reveals that the uncertainty in the state estimate cannot exceed that associated with the most precise ingredient of the assimilation, the forecast initial conditions or the observations. This is further explored in the following section.

5. Asymptotic behavior of $\mathbf{P}_k$ and its independence from $\mathbf{P}_0$. In this section, we study the asymptotic behavior of the forecast error covariance matrix $\mathbf{P}_k$ when $k \to \infty$. In particular, we are interested in the conditions for which the asymptotic sequence of $\mathbf{P}_k$ becomes independent of $\mathbf{P}_0$. In contrast to the rigor of the previous sections, a heuristic approach is followed here. We provide however a detailed description of the ideas and mathematical techniques of the derivation.

Recall that $\mathbf{P}_0$ is possibly degenerate of rank $r_0 \leq n$. As in section 3, $\mathbf{P}_0$ can be factorized into $\mathbf{P}_0 = \mathbf{X}_0 \mathbf{X}_0^{\mathrm{T}}$ where $\mathbf{X}_0$ is a matrix in $\mathbb{R}^{n \times r_0}$. The SVD of $\mathbf{M}_{k:0}$ has been defined in section 4.2. Using the partition

$$\mathbf{U}_{k:0} = [\mathbf{U}_{+,k:0}, \mathbf{U}_{-,k:0}], \quad \mathbf{V}_{k:0} = [\mathbf{V}_{+,k:0}, \mathbf{V}_{-,k:0}], \tag{5.1}$$

where $\mathbf{U}_{+,k:0}$ contains the left singular vectors that converge to the backward Lyapunov vectors of the unstable-neutral subspace $\mathcal{U}_k$, while $\mathbf{U}_{-,k:0}$ contains the left singular vectors that converge to the backward Lyapunov vectors of the stable subspace $\mathcal{S}_k$, the dynamics are conformably decomposed

$$\mathbf{M}_{k:0} = \mathbf{U}_{+,k:0} \boldsymbol{\Sigma}_{+,k:0} \mathbf{V}_{+,k:0}^{\mathrm{T}} + \mathbf{U}_{-,k:0} \boldsymbol{\Sigma}_{-,k:0} \mathbf{V}_{-,k:0}^{\mathrm{T}}. \tag{5.2}$$

Note that the columns of the singular and Lyapunov vector matrices are implicitly and consistently ordered, by decreasing value of the Lyapunov exponents.

Our goal is to heuristically prove that three conditions are sufficient for the existence of an asymptotic sequence of matrices, independent of $\mathbf{P}_0$, toward which $\mathbf{P}_k$ converges. These conditions follow.

Condition 1. Let $\mathbf{V}_{+,0} = \lim_{k \to \infty} \mathbf{V}_{+,k:0} \in \mathbb{R}^{n \times n_0}$ whose columns are the forward Lyapunov vectors at t_0 associated to the unstable and neutral directions. The condition reads

$$\mathrm{rank}\left(\mathbf{X}_0^{\mathrm{T}} \mathbf{V}_{+,0}\right) = n_0. \tag{5.3}$$

The idea is to make the column space of $\mathbf{P}_0$ large enough so that it has non-zero projections onto all of the unstable and neutral forward Lyapunov vectors at t_0. In that way, the column space of $\mathbf{P}_k$ will asymptotically contains the unstable-neutral subspace. Note that Eq. (5.3) implies $r_0 \geq n_0$.

Following the partition Eq. (5.1), we have

$$\mathbf{V}_{k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{k:0} = \begin{pmatrix} \mathbf{V}_{+,k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{+,k:0} & \mathbf{V}_{+,k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{-,k:0} \\ \mathbf{V}_{-,k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{+,k:0} & \mathbf{V}_{-,k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{-,k:0} \end{pmatrix} \tag{5.4}$$

which is positive semi-definite, and whose upper-left block $\mathbf{V}_{+,k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{+,k:0}$ is positive definite for k large enough by condition 1. It is shown in appendix B point 7, that, as a consequence, there exists $\sigma > 0$ such that $\mathbf{V}_{k:0}^{\mathrm{T}} \mathbf{P}_0 \mathbf{V}_{k:0} \geq \sigma \mathbf{V}_{+,k:0}^{\mathrm{T}} \mathbf{V}_{+,k:0}$ which yields from Eq. (4.4) and for k large enough

$$\mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^{\mathrm{T}} \geq \sigma \mathbf{U}_{+,k:0} \boldsymbol{\Sigma}_{+,k:0}^2 \mathbf{U}_{+,k:0}^{\mathrm{T}}. \tag{5.5}$$

In a complementary way, we know from section 4.2 that there exists $\alpha > 0$ such that

$$\alpha \mathbf{U}_{k:0} \boldsymbol{\Sigma}_{k:0}^2 \mathbf{U}_{k:0}^{\mathrm{T}} \geq \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^{\mathrm{T}}. \tag{5.6}$$

Condition 2. The model is sufficiently observed so that the unstable and neutral directions remain under control, that is

$$\mathbf{U}_{+,k}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{+,k} > \varepsilon\mathbf{I} \tag{5.7}$$

where $\mathbf{U}_{+,k} = \lim_{l\to-\infty}\mathbf{U}_{+,k:l}$ and $\varepsilon > 0$ is a positive number. Condition 2 amounts to a uniform observability condition on the unstable and neutral directions.

Condition 3. For any neutral backward Lyapunov vector $\mathbf{u}_k$, we have

$$\lim_{k\to\infty}\mathbf{u}_k^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{u}_k = \infty. \tag{5.8}$$

In practice, and looking at Eq. (2.7), a regularly frequent observation of the neutral directions should suffice to satisfy Eq. (5.8). Condition 3 complements condition 2.

Let us now infer bounds for $\mathbf{P}_k$ from the bounds of the free forecast of the initial error covariances, Eqs. (5.5,5.6). For $\mathbf{A},\mathbf{B}\in\mathcal{C}^n$, let us introduce the function $\mathbf{F}_k$ that maps $[0,1]$ in $\mathbb{R}^{n\times n}$ according to

$$\mathbf{F}_k(t) = (\mathbf{A} + t\mathbf{B})\left(\mathbf{I} + \boldsymbol{\Gamma}_k\left(\mathbf{A} + t\mathbf{B}\right)\right)^{-1}. \tag{5.9}$$

Denoting $\mathbf{F}_k'(t)$ the derivative of $\mathbf{F}_k$, we have

$$\mathbf{F}_k'(t) = \left(\mathbf{I} + (\mathbf{A} + t\mathbf{B})\,\boldsymbol{\Gamma}_k\right)^{-1}\mathbf{B}\left(\mathbf{I} + \boldsymbol{\Gamma}_k\left(\mathbf{A} + t\mathbf{B}\right)\right)^{-1}. \tag{5.10}$$

Since $\mathbf{F}_k'(t) \geq \mathbf{0}$, we have

$$\mathbf{F}_k(1) - \mathbf{F}_k(0) = \int_0^1\mathbf{F}_k'(t)\,\mathrm{d}t \geq \mathbf{0}. \tag{5.11}$$

Choosing $\mathbf{A} = \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}} - \sigma\mathbf{U}_{+,k:0}\boldsymbol{\Sigma}_{+,k:0}^2\mathbf{U}_{+,k:0}^{\mathrm{T}}$ and $\mathbf{B} = \sigma\mathbf{U}_{+,k:0}\boldsymbol{\Sigma}_{+,k:0}^2\mathbf{U}_{+,k:0}^{\mathrm{T}}$, we obtain from Eq. (2.11) and Eq. (5.5)

$$\mathbf{P}_k \geq \sigma\mathbf{U}_{+,k:0}\boldsymbol{\Sigma}_{+,k:0}^2\mathbf{U}_{+,k:0}^{\mathrm{T}}\left[\mathbf{I} + \sigma\boldsymbol{\Gamma}_k\mathbf{U}_{+,k:0}\boldsymbol{\Sigma}_{+,k:0}^2\mathbf{U}_{+,k:0}^{\mathrm{T}}\right]^{-1}$$
$$\geq \mathbf{U}_{+,k:0}\left[\boldsymbol{\Sigma}_{+,k:0}^{-2}/\sigma + \mathbf{U}_{+,k:0}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{+,k:0}\right]^{-1}\mathbf{U}_{+,k:0}^{\mathrm{T}}, \tag{5.12}$$

where we used appendix C in the second line. Given both observability conditions Eqs. (5.7,5.8), we have the asymptotic equivalence

$$\boldsymbol{\Sigma}_{+,k:0}^{-2}/\sigma + \mathbf{U}_{+,k:0}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{+,k:0} \underset{k\to\infty}{\sim} \mathbf{U}_{+,k}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{+,k} \tag{5.13}$$

whose right-hand-side matrix is invertible by condition 2, so that

$$\lim_{k\to\infty}\left\{\mathbf{P}_k - \boldsymbol{\Phi}_k\right\} \geq \mathbf{0}, \quad \text{with} \quad \boldsymbol{\Phi}_k \triangleq \mathbf{U}_{+,k}\left[\mathbf{U}_{+,k}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{+,k}\right]^{-1}\mathbf{U}_{+,k}^{\mathrm{T}}. \tag{5.14}$$

Conversely, we would like to show that $\lim_{k\to\infty}\left\{\mathbf{P}_k - \boldsymbol{\Phi}_k\right\} \leq \mathbf{0}$. Choosing $\mathbf{A} = \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}$ and $\mathbf{B} = \alpha\mathbf{U}_{k:0}\boldsymbol{\Sigma}_{k:0}^2\mathbf{U}_{k:0}^{\mathrm{T}} - \mathbf{M}_{k:0}\mathbf{P}_0\mathbf{M}_{k:0}^{\mathrm{T}}$, we obtain from Eq. (2.11), Eq. (5.6) and from appendix C

$$\mathbf{P}_k \leq \alpha\mathbf{U}_{k:0}\boldsymbol{\Sigma}_{k:0}^2\mathbf{U}_{k:0}^{\mathrm{T}}\left[\mathbf{I} + \alpha\boldsymbol{\Gamma}_k\mathbf{U}_{k:0}\boldsymbol{\Sigma}_{k:0}^2\mathbf{U}_{k:0}^{\mathrm{T}}\right]^{-1}$$
$$\leq \mathbf{U}_{k:0}\left(\boldsymbol{\Sigma}_{k:0}^{-2}/\alpha + \mathbf{U}_{k:0}^{\mathrm{T}}\boldsymbol{\Gamma}_k\mathbf{U}_{k:0}\right)^{-1}\mathbf{U}_{k:0}^{\mathrm{T}}. \tag{5.15}$$

Let us focus on the right-hand side of Eq. (5.15) and decompose it over the subspaces $\mathcal{U}_{k:0}$ and $\mathcal{S}_{k:0}$, using the notations $+$ for the entries in $\mathcal{U}_{k:0}$ and $-$ for the entries in $\mathcal{S}_{k:0}$. We denote $\widehat{\mathbf{\Gamma}}_k \triangleq \mathbf{U}_{k:0}^{\mathrm{T}} \mathbf{\Gamma}_k \mathbf{U}_{k:0}$ and we will often drop references to time t_k in the following, for the sake of clarity. The restriction of the right-hand side of Eq. (5.15) to $\mathcal{U}_{k:0}$ can be obtained using the Schur complement derived from the block decomposition, and from appendix B point 3

$$\left[\mathbf{U}_{+,k:0}^{\mathrm{T}} \mathbf{P}_k \mathbf{U}_{+,k:0}\right]^{-1} \geq \mathbf{\Sigma}_+^{-2}/\alpha + \widehat{\mathbf{\Gamma}}_{++} - \widehat{\mathbf{\Gamma}}_{+-}\left(\mathbf{\Sigma}_-^{-2}/\alpha + \widehat{\mathbf{\Gamma}}_{--}\right)^{-1}\widehat{\mathbf{\Gamma}}_{-+}, \qquad (5.16)$$

which implies

$$\left[\mathbf{U}_{+,k:0}^{\mathrm{T}} \mathbf{P}_k \mathbf{U}_{+,k:0}\right]^{-1} \geq \widehat{\mathbf{\Gamma}}_{++} - \alpha\widehat{\mathbf{\Gamma}}_{+-}\mathbf{\Sigma}_-^2\widehat{\mathbf{\Gamma}}_{-+} \qquad (5.17)$$

whose right-hand-side must be studied in the limit $k \to \infty$. Using the definition of $\mathbf{\Gamma}_k$ from Eq. (2.7) and the SVD decomposition of $\mathbf{M}_{k:l}$, we unfold

$$\widehat{\mathbf{\Gamma}}_k = \sum_{l=0}^{k-1} \mathbf{U}_{k:0}^{\mathrm{T}}\mathbf{U}_{k:l}\mathbf{\Sigma}_{k:l}^{-1}\mathbf{V}_{k:l}^{\mathrm{T}}\mathbf{\Omega}_l\mathbf{V}_{k:l}\mathbf{\Sigma}_{k:l}^{-1}\mathbf{U}_{k:l}^{\mathrm{T}}\mathbf{U}_{k:0} \qquad (5.18)$$

whose convergence can be heuristically studied. In the sum the convergence is controlled by the matrices of the singular values. At fixed l and for k large enough, $\lim_{k\to\infty} \{\mathbf{U}_{k:0} - \mathbf{U}_{k:l}\} = \mathbf{0}$ so that $\mathbf{U}_{k:0}^{\mathrm{T}}\mathbf{U}_{k:l}$ converges to the identity matrix. Hence, we have

$$\left[\widehat{\mathbf{\Gamma}}_k\right]_{ij} \sim \sum_{l=0}^{k-1} \gamma_l e^{-(\lambda_i+\lambda_j)(k-l)} \qquad (5.19)$$

where γ_l are non-negative uniformly bounded constants since the sequence $\mathbf{\Omega}_k$ was assumed uniformly bounded from above (section 1.2). The asymptotic trend of the terms $\widehat{\mathbf{\Gamma}}_{++}$ and $\widehat{\mathbf{\Gamma}}_{+-}\mathbf{\Sigma}_-$ can now be inferred. From Eq. (5.19), we have for i, j indexes related to unstable or neutral directions

$$\left[\widehat{\mathbf{\Gamma}}_{++}\right]_{ij} \sim \sum_{l=0}^{k-1} \gamma_l e^{-(\lambda_i+\lambda_j)(k-l)}, \qquad (5.20)$$

which converges to a finite value if i or j corresponds to an unstable direction. If i and j correspond to neutral directions, then by condition Eq. (5.8), the term goes to ∞. If i is an index related to an unstable or neutral direction and j an index related to a stable direction, we have

$$\left[\widehat{\mathbf{\Gamma}}_{+-}\mathbf{\Sigma}_-\right]_{ij} \sim \sum_{l=0}^{k-1} \gamma_l e^{-(\lambda_i+\lambda_j)(k-l)+\lambda_j k}. \qquad (5.21)$$

The exponent is bounded from above by $-\min\{\lambda_i, -\lambda_j\}(k-1)$ which implies that the entry converges to 0 unless i is related to a neutral direction, in which case it converges to a finite value. Consequently, $\left[\widehat{\mathbf{\Gamma}}_{+-}\mathbf{\Sigma}_-^2\widehat{\mathbf{\Gamma}}_{-+}\right]_{ij}$ asymptotically vanishes unless i and j are related to neutral directions, in which case it converges to a finite value. In both cases, because of the observability conditions Eqs. (5.3,5.7), the first term of the lower bound dominates the second term, which reads

$$\widehat{\mathbf{\Gamma}}_{++} - \alpha\widehat{\mathbf{\Gamma}}_{+-}\mathbf{\Sigma}_-^2\widehat{\mathbf{\Gamma}}_{-+} \underset{k\to\infty}{\sim} \widehat{\mathbf{\Gamma}}_{++} \qquad (5.22)$$

which implies from Eq. (5.17) and from appendix B, point 3,

$$\lim_{k\to\infty}\left\{\mathbf{U}_{+,k}^{\mathrm{T}}\mathbf{P}_k\mathbf{U}_{+,k}-\left[\mathbf{U}_{+,k}^{\mathrm{T}}\mathbf{\Gamma}_k\mathbf{U}_{+,k}\right]^{-1}\right\}\le\mathbf{0}. \tag{5.23}$$

Given that $\mathbf{P}_k$ is a positive semi-definite matrix and that $\mathbf{U}_{+,k}^{\mathrm{T}}\mathbf{P}_k\mathbf{U}_{+,k}$ converges to a finite matrix while $\mathbf{U}_{-,k}^{\mathrm{T}}\mathbf{P}_k\mathbf{U}_{-,k}$ converges to $\mathbf{0}$, we can furthermore conclude that the off-diagonal blocks of $\mathbf{P}_k$ converge to $\mathbf{0}$ as well. Hence, we have $\lim_{k\to\infty}\left\{\mathbf{P}_k-\mathbf{\Phi}_k\right\}\le\mathbf{0}$ which, with Eq. (5.14), finally yields

$$\lim_{k\to\infty}\left\{\mathbf{P}_k-\mathbf{\Phi}_k\right\}=\mathbf{0}, \tag{5.24}$$

where the asymptotic sequence $\mathbf{\Phi}_k$ defined in Eq. (5.14) is independent of $\mathbf{P}_0$.

If condition Eq. (5.7) is replaced with a full and uniform observability condition $\mathbf{\Gamma}_k\ge\varepsilon\mathbf{I}$ with $\varepsilon>0$, then the inverse of $\mathbf{\Gamma}_k$ exists and it could further be shown that

$$\lim_{k\to\infty}\left\{\mathbf{P}_k-\mathbf{\Gamma}_k^{-1}\right\}=\mathbf{0}. \tag{5.25}$$

Recall from its definition, Eq. (2.7), that $\mathbf{\Gamma}_k$ measures the precision of the observations transferred into the model space, thus the asymptotic limit, Eqs. (5.24,5.25), shows that the system's state error covariance matrix will converge to a matrix related to the observational error.

It is finally worth mentioning the particular role played by the neutral modes. In section 4, the exponential convergence to $\mathbf{0}$ of $\mathbf{P}_k$ for all stable directions was proven. Provided the three conditions Eqs. (5.3,5.7,5.8) are met, the above discussion points to an exponential convergence of $\mathbf{P}_k$ onto $\mathbf{\Phi}_k$ for all unstable directions. Nevertheless, the discussion also suggests a much slower convergence of $\mathbf{P}_k$ to $\mathbf{0}$ for all neutral directions, possibly in k^{-1}, since $\mathbf{\Gamma}_k$ is expected to grow linearly with k for uniformly bounded $\mathbf{\Omega}_k$. The critical importance of the neutral modes was originally observed by [23] in numerical experiments with assimilation in the unstable-neutral subspace in the context of variational data assimilation with nonlinear dynamics. They numerically showed that it was necessary to include the neutral direction within the subspace where the assimilation was performed in order to efficiently control the error growth. The analysis carried out in the present section further corroborates their findings. Moving away from the linear hypothesis towards nonlinear dynamics and in connection with this slow convergence of the neutral modes, it was recently argued [3] that the region of the Lyapunov spectrum around the neutral modes is critical in the convergence of the EnKF. The mis-estimation of the uncertainty in this region of the spectrum was shown to be the reason why the *ad-hoc* technique known as inflation meant to stabilize the filter is very often required.

Finally, let us mention that it is possible to extend the derivation of this section to the case $r_0<n_0$, by modifying the three conditions and using the projection on the r_0 most unstable backward Lyapunov vectors. The conclusion still stands but with the new projector. This situation is nonetheless physically and practically less interesting because several unstable directions are left uncontrolled; a situation that often leads to filter divergence.

6. Numerical results. We present here numerical results on the asymptotic properties of the analysis error covariance $\mathbf{P}_k^{\mathrm{a}}$ that corroborate and illustrate the theoretical findings. The convergence results obtained for $\mathbf{P}_k$ can easily be transferred to $\mathbf{P}_k^{\mathrm{a}}$ by Eq. (1.7), or by applying $\mathbf{P}_k^{\mathrm{a}}\le\mathbf{P}_k$ which is readily obtained using the matrix shift lemma to Eq. (1.6) as in Eq. (3.2).

Three different experimental setups are considered, with different choices of the dynamical and observational models in Eqs. (1.1,1.2). In all cases the perfect model hypothesis is employed, $\mathbf{Q}_k = \mathbf{0}$.

Exp1: Autonomous system. The state- and observation-space dimensions are $n = 30$ and $d = 10$ respectively. The time-invariant matrices $\mathbf{M}_k \triangleq \mathbf{M} \in \mathbb{R}^{n \times n}$, $\mathbf{H}_k \triangleq \mathbf{H} \in \mathbb{R}^{d \times n}$, and $\mathbf{R}_k \triangleq \mathbf{R} \in \mathbb{R}^{d \times d}$ are chosen randomly, *i.e.* with entries which are *iid* standard normal random variables.

Exp2: Random non-autonomous system. The state- and observation-space dimensions are $n = 30$ and $d = 10$ respectively. The time-varying, invertible, propagators $\mathbf{M}_k \in \mathbb{R}^{n \times n}$, the observation error covariance matrices $\mathbf{R}_k \in \mathbb{R}^{d \times d}$ and the observation matrices $\mathbf{H}_k \in \mathbb{R}^{d \times n}$ are all randomly generated, *i.e.* the entries of these matrices are *iid* standard normal random variables.

Exp3: Non-autonomous system obtained by linearization around a trajectory of the Lorenz-95 model. The entries of the observation error covariance, $\mathbf{R}_k$ and $\mathbf{H}_k$ are generated as in Exp2 but with the state- and observation-space dimensions being $n = 40$ and $d = 15$ respectively. The propagators $\mathbf{M}_k$ are taken to be the linearization around a trajectory on the attractor of the $n = 40$-dimensional Lorenz-95 model [15], which is very commonly used in DA literature, see *e.g.* [8, and references therein]. The equations read

$$\frac{\mathrm{d}x_j}{\mathrm{d}t} = x_{j-1}\left(x_{j+1} - x_{j-2}\right) - x_j + F \qquad j = 1, ..., n \qquad (6.1)$$

with periodic boundary conditions, $x_0 = x_n$, $x_{-1} = x_{n-1}$ and $x_{n+1} = x_1$. The standard value of the forcing, $F = 8$, is used in the following experiments. The observation interval is $\Delta t = 0.1$.

In another numerical experiment, we used a simpler observational network, by choosing $\mathbf{H}_k = [1, 0, \dots, 0]$, corresponding to observation of only the first component of the state vector. The numerical results for Exp3 (with randomly chosen elements of $\mathbf{H}_k$ of dimension 15×40) and for this much simpler observational network were qualitatively the same and thus the latter have not been presented here.

It must be emphasized that this case (Exp3) does not coincide with the nonlinear filtering problem of the Lorenz-95 model but it makes use of the linearization of the model to build up the propagator which is then used as a linear model in Eq. (1.1).

Note furthermore that, in an extended Kalman filter (EKF, [12]) applied to a nonlinear system such as the Lorenz-95, the only place where the state estimate enters the computation of the covariance matrices is in the linearization of the model dynamics in which one needs to estimate the Jacobian of the model dynamics evaluated on the system's state. Therefore, for the Lorenz-95 model, the analysis and forecast covariances of the EKF will show asymptotic behavior similar to what is presented below. While this behavior was already observed and exploited in a reduced-order formulation of the EKF based on the unstable subspace [24], it does not give much hints about the asymptotic behavior of a fully nonlinear filter.

Each of the three experimental setups is representative of a class of systems. Numerical results (not shown) for other choices of the system and observational dimension as well as for other realizations of the random matrices $\mathbf{M}_k, \mathbf{H}_k, \mathbf{R}_k$ were found to be qualitatively equivalent to the results reported below.

For the three numerical experiments described above, it is very difficult to check the observability condition $\det(\boldsymbol{\Gamma}_k) \neq 0$ because the matrices $\boldsymbol{\Gamma}_k$ soon become very

ill-conditioned. But we expect that the system would be observable with probability 1 for Exp1 and Exp2, while in Exp3 (the case of Lorenz-95 model), we expect the system to be observable even with a single variable being observed, since each variable is coupled to those around it. Note that we are unable to numerically verify the above statements.

The QR method [20, 14] is adopted to numerically compute the Lyapunov vectors and exponents. Starting from a random positive semi-definite $\mathbf{P}_0^{\mathrm{a}}$, the sequence $(\mathbf{P}_k, \mathbf{P}_k^{\mathrm{a}})$ for $k > 0$ of forecast and analysis error covariance matrices was generated based on the Kalman filter Eqs. (1.6,1.7).

Recall that n_0 stands for the number of non-negative Lyapunov exponents: in most cases, n_0 will correspond to the number of positive plus one zero exponent. Numerically, this zero Lyapunov exponent will not be exactly zero but it will fluctuate around it. Also recall that r_0 is the rank of the initial covariance matrices $\mathbf{P}_0$, or $\mathbf{P}_0^{\mathrm{a}}$.

6.1. Rate of convergence of the eigenvalues. The following numerical experiments show the relation between the rates of convergence of eigenvalues, σ_i^k, of the error covariance matrix, $\mathbf{P}_k$, and the Lyapunov exponents of the dynamical system of Eqs. (1.1,1.2). The eigenvalues are ordered so that $\sigma_1^k \geq \sigma_2^k \cdots \geq \sigma_n^k$.

When $r_0 < n_0$, the rank of $\mathbf{P}_k$ as $k \to \infty$ generically remains r_0 for almost all initial conditions with no eigenvalues decaying to zero. Thus we consider here the relevant situation: $r_0 \geq n_0$. From section 4.1, we know that in this case, $\sigma_{r_0+1}^k = \cdots = \sigma_n^k = 0$. Moreover, from section 4.4, Eq. (4.13), we know that $\sigma_1^k, \ldots, \sigma_{n_0}^k$ will remain non-zero even in the limit $k \to \infty$ – except maybe for the neutral directions as discussed in section 5 – whereas $\sigma_{n_0+1}^k, \ldots, \sigma_{r_0}^k$ will decay to zero. Recall from section 4.2 that $\exp(\lambda_i^k k)$ is a singular value of $\mathbf{M}_{k:0}$ so that λ_i^k approaches the Lyapunov exponent λ_i as $k \to \infty$; the Lyapunov exponents are ordered so that the first n_0 are non-negative, $\lambda_1 > \lambda_2 > \cdots > \lambda_{n_0} \geq 0$, whereas the rest are negative with decreasing value, $0 > \lambda_{n_0+1} > \lambda_{n_0+2} > \cdots > \lambda_n$.

The results of section 4.3 and the inequality (4.12) can be used to derive the rate of convergence of the smallest $r_0 - n_0$ eigenvalues σ_i^k with $i = n_0 + 1, \ldots, r_0$. Using the largest eigenvalue at the initial time, σ_1^0, as for α_i in Eq. (4.12) we have

$$\sigma_i^k \leq \sigma_1^0 \exp\left(2\lambda_i^k k\right) \tag{6.2}$$

which implies that, asymptotically,

$$\ln(\sigma_i^k) \leq \ln(\sigma_1^0) + 2\lambda_i^k k \underset{k \to \infty}{\sim} \ln(\sigma_1^0) + 2\lambda_i k. \tag{6.3}$$

The equivalence in Eq. (6.3) is valid in the limit $k \to \infty$ since $\lambda_i^k \to \lambda_i$ as $k \to \infty$. Therefore for $i = n_0 + 1, \ldots, r_0$, the eigenvalues σ_i^k of $\mathbf{P}_k$ decay to zero exponentially fast with the exponential decay rate asymptotically being at least twice the Lyapunov exponent λ_i. Note that, as mentioned above, $\mathbf{P}_k^{\mathrm{a}} \leq \mathbf{P}_k$, so the aforementioned decay rate is also valid for the eigenvalues of the analysis error covariance matrix, $\mathbf{P}_k^{\mathrm{a}}$.

Figure 6.1(a) illustrates the decay of some of the eigenvalues of the analysis error covariance matrix in Exp3. Similar graphs have been obtained for Exp1 and Exp2 (not shown). Figure 6.1(b) shows the slopes of the best fit lines for the semi-log plot of σ_i^k versus k for the full rank $\mathbf{P}_0$ (red dots), for all three experiments described above. The blue dots show the values of twice the absolute value of the corresponding negative Lyapunov exponents. We see that indeed the inequality (6.3) is saturated.
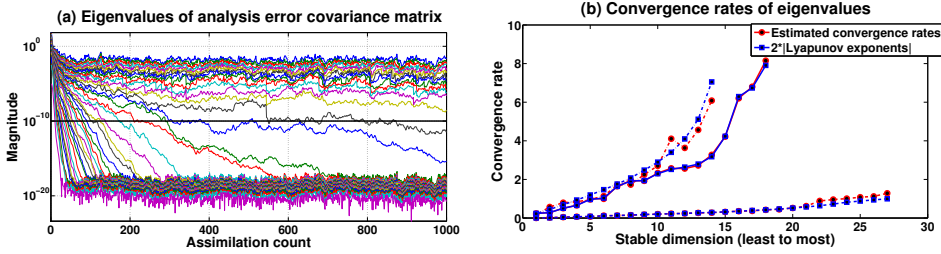
FIG. 6.1. *Panel (a) shows the eigenvalues of $\mathbf{P}_k^a$ in Exp3 and the decay of part of its spectrum. Panel (b) shows the comparison between the decay rate of the eigenvalues of the analysis covariance matrix (red lines) with twice the absolute value of the negative Lyapunov exponents (blue lines), for the autonomous system (solid line, $n = 30, n_0 = 16$), and for two examples of non-autonomous systems with random propagators (dash-dot line, $n = 30, n_0 = 16$) and with propagators derived from the Lorenz-95 system (dashed line, $n = 40, n_0 = 14$), for full rank $\mathbf{P}_0$.*

6.2. Existence of asymptotic sequences of low-rank covariance matrices. The next set of numerical results corroborate the results in section 4.2 about the projections of the covariance matrices onto the stable subspace vanishing and the results in section 5 about their asymptotic behavior.

Figure 6.2 plots the rank of $\mathbf{P}_k^a$ as a function of k, where various choices of the rank r_0 of $\mathbf{P}_0^a$ are shown by various colors in the figure. Note that we actually plot the number of eigenvalues greater than a threshold of 10^{-10}.
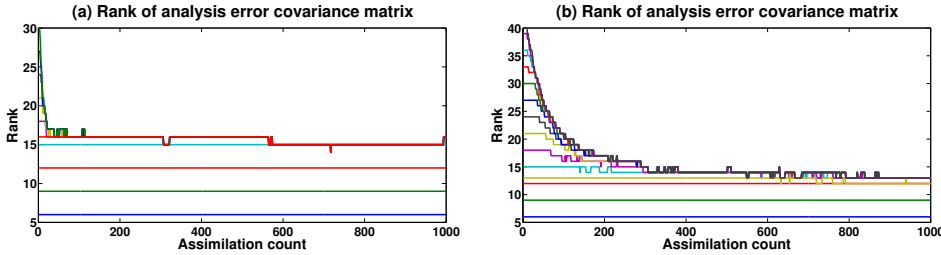

FIG. 6.2. *Rank of $\mathbf{P}_k^a$ as a function of k for several choices of the rank r_0 of $\mathbf{P}_0^a$ (various colors) for the two systems, one with random propagators $\mathbf{M}_k$ (a) with $n_0 = 16$ and another with propagators which are linearization of Lorenz-95 around a trajectory on the attractor (b) with $n_0 = 14$.*

Panel (a) of Fig. 6.2 shows the case of random propagators (Exp2) $\mathbf{M}_k$, which has $n = 30$ and $n_0 = 16$, *i.e.* the number of non-negative Lyapunov exponents is 16. Panel (b) refers to the case Exp3 of linearization of Lorenz-95 with $F = 8$ around a trajectory on its attractor with $n = 40$ and $n_0 = 14$. We see that if $r_0 < n_0$, then the rank of $\mathbf{P}_k^a$ is constant and equal to the initial rank r_0. On the other hand, if $r_0 \geq n_0$, then $r_0 - n_0$ eigenvalues values approach zero, $n_0 - 1$ eigenvalues remain non-zero while one eigenvalue fluctuates and it is unclear whether it will approach zero or indeed remain non-zero. It very likely corresponds to the neutral direction along which convergence of $\mathbf{P}_k$ is very slow even if well observed, as discussed in section 5.

In the next numerical experiment, we generate two sequences of analysis covariances $\mathbf{P}_k^a$ and $\mathbf{P}_k^{'a}$ starting from two different initial conditions $\mathbf{P}_0^a$ and $\mathbf{P}_0^{'a}$, respectively. Figure 6.3 shows the Frobenius norm of the difference between analysis covariances, *i.e.* $\|\mathbf{P}_k^a - \mathbf{P}_k^{'a}\|$, as a function of time k. Four cases are considered in Fig. 6.3:
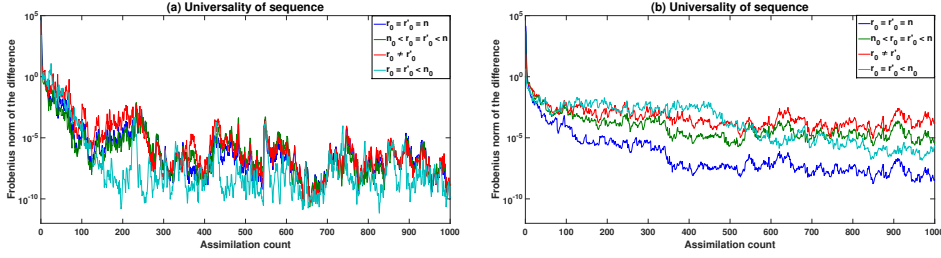
16

FIG. 6.3. *Frobenius norm of the difference, i.e.* $\|\mathbf{P}_k^{\mathrm{a}} - \mathbf{P}_k^{'\mathrm{a}}\|$ *for two sequences of analysis covariances matrices starting with different initial conditions* $\mathbf{P}_0^{\mathrm{a}}$ *and* $\mathbf{P}_0^{'\mathrm{a}}$ *for the case of random propagators (a) with* $n_0 = 16$ *and Lorenz-95 linearization (b) with* $n_0 = 14$.

1. $r_0 = r'_0 = n$ when the initial ranks are the same and equal to the state dimension (blue line),
2. $n_0 < r_0 = r'_0 < n$ when the initial covariance matrices are rank-deficient with same ranks greater than n_0 (green line),
3. $r_0 \neq r'_0$ and $n_0 < r_0, r'_0 < n$ when the initial ranks are unequal but both ranks are greater than n_0 (red line),
4. $r_0 = r'_0 < n_0$ when the initial ranks are the same and less than n_0 (teal line).

In all these cases, we see that the norm of the difference approaches zero within the numerical accuracy, fluctuating between 10^{-8} and 10^{-3}, *i.e.* for large k, $\mathbf{P}_k^{\mathrm{a}} \approx \mathbf{P}_k^{'\mathrm{a}}$. Thus the sequence $\mathbf{P}_k^{\mathrm{a}}$ is equivalent to a sequence of covariance matrices all of rank $s = \min\{r_0, n_0\}$, independent of the initial condition $\mathbf{P}_0^{\mathrm{a}}$, but of course dependent on the dynamics $\mathbf{M}_k$, the observations $\mathbf{H}_k$ and their error covariances $\mathbf{R}_k$.

The asymptotic covariance matrices are most easily represented in the basis of the backward Lyapunov vectors. As proven mathematically in section 4.2 in the case of strong collapse which occurs here because the systems are sufficiently observed, these covariance matrices have column spaces corresponding to the span of the most unstable backward Lyapunov vectors and their null space subsumes the span of the stable backward Lyapunov vectors. This can be seen by looking at the projection of these covariance matrices $\mathbf{P}_k^{\mathrm{a}}$ onto the backward Lyapunov vectors $\mathbf{u}_1^k, \cdots, \mathbf{u}_n^k$ at time t_k.

Figure 6.4 shows these projections for four different values of $k = 2500$, 3000, 3500, 4000 for the cases $r_0 \geq n_0$ (top row) and $r_0 < n_0$ (middle row). The Exp2 and Exp3 cases are displayed in the left and right column panels respectively. Note that the Lyapunov vectors are ordered from the largest to the smallest Lyapunov exponents. This is also clearly seen from the bottom row of the same Fig. 6.4 which shows these projections at a fixed time $k = 5000$ for various initial ranks r_0 which are equal or less than n_0.

6.3. Low-rank asymptotic covariance for autonomous systems. The last set of numerical results illustrates the asymptotic convergence of the analysis covariances for the case of autonomous systems. The results are very similar to those of the non-autonomous systems and a summary is presented in Fig. 6.5. The left panel shows the Frobenius norm of the difference $\mathbf{P}_{k+1}^{\mathrm{a}} - \mathbf{P}_k^{\mathrm{a}}$ of the analysis covariance matrices at consecutive time instances. The figure clearly shows this difference going to zero and thus by Cauchy's convergence criterion, the sequence of the analysis covariance matrices converges. Different lines are meant for cases of different initial ranks. The right panel shows the projections onto the backward Lyapunov vectors which also
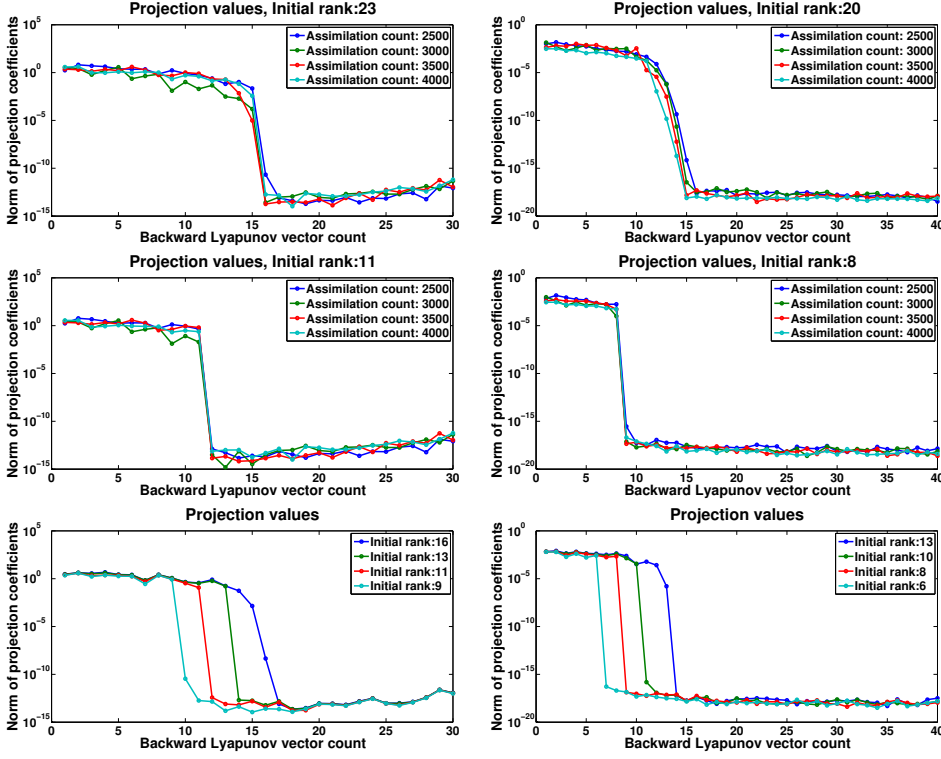
Fig. 6.4. *Projections of covariance matrices* $\mathbf{P}_k^{\mathrm{a}}$ *onto the backward Lyapunov vectors* $\mathbf{u}_1^k, \cdots, \mathbf{u}_n^k$ *for system with random propagators (left column,* $n = 30, n_0 = 16$*) and linearization of Lorenz-95 (right column,* $n = 40, n_0 = 14$*).*

span the generalized eigenspace of $\mathbf{M}^{\mathrm{T}}$ [11], for four cases with different initial rank r_0, and these results are very similar to those shown in the bottom row of Fig. 6.4.
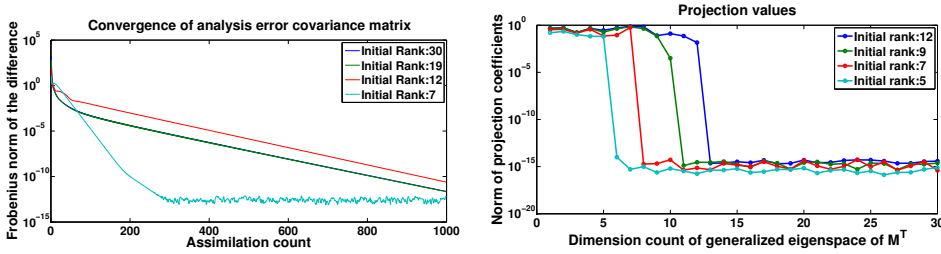

Fig. 6.5. *Frobenius norm of the consecutive difference, i.e.* $\|\mathbf{P}_k^{\mathrm{a}} - \mathbf{P}_{k-1}^{\mathrm{a}}\|$ *for several choices of rank* r_0 *of the initial condition* $\mathbf{P}_0^{\mathrm{a}}$ *(left panel) and projections onto the generalized eigenspace of* $\mathbf{M}^{\mathrm{T}}$ *(right panel) for the autonomous system with* $n = 30, n_0 = 13$*.*

7. Conclusion. We have shown that, for perfect linear dynamics and observation operator, and for any initial error covariance matrix, the solution of the Kalman filter covariance equation converges onto the unstable-neutral subspace of the dynamics. The rate of such convergence has also been provided. Moreover, we have proved the existence of a universal sequence, independent of the initial condition, to-

ward which the Kalman filter error covariance converges if the system is sufficiently observed and if the column space of the initial error covariance has a non-zero projection on all the unstable and neutral forward Lyapunov vectors. These results were obtained after proving an analytical expression of the covariances at any time in terms of the initial covariances. Numerical experiments were used to further corroborate and illustrate the mathematical statements. These results complete and generalize those in [11] and altogether lay the mathematical foundation of the methods that rely on the assimilation in the unstable subspace [17].

Yet, this work leaves unresolved some key issues that are worth investigating in the perspective of the design of reduced-order algorithms applicable to practical situations. Specific lines of development include the treatment of model error and the extension to nonlinear dynamics. This latter problem stimulates indeed an intriguing, albeit necessary, direction of study whose main difficulty stands on the fact that the unstable subspace is, in this nonlinear case, no longer globally defined but a function of the underlying trajectory. Both lines of research may lead to interesting methodological and mathematical developments and are central in data assimilation.

In our view the present results are also relevant to the field of ensemble-based data assimilation algorithms for the geosciences, or more generally to the uncertainty quantification and data assimilation methods in complex high-dimensional and big data problems, of which data assimilation for the geosciences is a prototypical example. We believe so for two distinctive reasons. First the present findings on the error covariance projection onto the unstable-neutral subspace provide a natural rationale to interpret a stream of numerical evidences that relates the minimum ensemble size to achieve a satisfactorily estimate of the system's state, with the number of unstable directions of the underlying dynamics [8, 16, 4]. Second, this study encourages a research effort toward EnKF formulations that incorporate the information on the unstable subspace explicitly in the design and choice of the ensemble, possibly in combination with localization techniques widely used to artificially increase the rank of the ensemble-based error covariance matrices.

It is finally worth mentioning another appealing research direction: the extension of the present framework to fully Bayesian data assimilation methods typically preferable in the presence of strong nonlinearities and/or non-Gaussian error [2]. Besides the aforesaid difficulty inherent to nonlinear dynamics, the additional problem here is on how to link the geometrical (in the phase-space) features of the unstable subspace to the conditioning of a probability density function, that is the generalization to the fully Bayesian framework, of projecting the error covariances onto the unstable subspace.

Appendix A. Deriving $\mathbf{P}_k$ using the symplectic symmetry.

This appendix gives an account of the linear representation of the recurrence Eq. (1.4), which had initially been developed as a way to solve the Riccati equation in the autonomous case [1, and references therein]. We use it to give an alternative derivation of Eq. (2.11) and to discuss in more details the analytic expression in the autonomous case. The underlying symplectic structure of the Kalman filter has been for instance explored in [5, 28].

A.1. General linear representation using symplectic matrices. Let us rewrite the recurrence Eq. (1.4)

$$
\begin{aligned}
\mathbf{P}_k &= \mathbf{M}_k\left(\mathbf{I} + \mathbf{P}_{k-1}\boldsymbol{\Omega}_{k-1}\right)^{-1}\mathbf{P}_{k-1}\mathbf{M}_k^{\mathrm{T}} + \mathbf{Q}_k \\
&= \mathbf{M}_k\mathbf{P}_{k-1}\left(\mathbf{I} + \boldsymbol{\Omega}_{k-1}\mathbf{P}_{k-1}\right)^{-1}\mathbf{M}_k^{\mathrm{T}} + \mathbf{Q}_k \\
&= \mathbf{M}_k\mathbf{P}_{k-1}\left(\mathbf{M}_k^{-\mathrm{T}} + \mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1}\mathbf{P}_{k-1}\right)^{-1} + \mathbf{Q}_k \\
&= \left(\mathbf{M}_k\mathbf{P}_{k-1} + \mathbf{Q}_k\left\{\mathbf{M}_k^{-\mathrm{T}} + \mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1}\mathbf{P}_{k-1}\right\}\right)\left(\mathbf{M}_k^{-\mathrm{T}} + \mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1}\mathbf{P}_{k-1}\right)^{-1} \\
&= \left(\left\{\mathbf{M}_k + \mathbf{Q}_k\mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1}\right\}\mathbf{P}_{k-1} + \mathbf{Q}_k\mathbf{M}_k^{-\mathrm{T}}\right)\left(\mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1}\mathbf{P}_{k-1} + \mathbf{M}_k^{-\mathrm{T}}\right)^{-1} \\
&\triangleq \left(\mathbf{A}_k\mathbf{P}_{k-1} + \mathbf{B}_k\right)\left(\mathbf{C}_k\mathbf{P}_{k-1} + \mathbf{D}_k\right)^{-1}, \tag{A.1}
\end{aligned}
$$

where we used the matrix shift lemma from the first to the second line, and we defined block matrices $\mathbf{A}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$, $\mathbf{D}_k$ in the fourth line. Let us define [1]

$$
\mathbf{Z}_k \triangleq \begin{pmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{C}_k & \mathbf{D}_k \end{pmatrix} = \begin{pmatrix} \mathbf{M}_k + \mathbf{Q}_k\mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1} & \mathbf{Q}_k\mathbf{M}_k^{-\mathrm{T}} \\ \mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1} & \mathbf{M}_k^{-\mathrm{T}} \end{pmatrix}. \tag{A.2}
$$

which is valid in the presence of model noise. This matrix belongs to the symplectic group $\mathrm{Sp}(2n, \mathbb{R})$ since $\mathbf{Z}_k^{-1} = -\mathbf{J}\mathbf{Z}_k^{\mathrm{T}}\mathbf{J}$, where $\mathbf{J} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix}$. It has a simple expression in the perfect model case

$$
\mathbf{Z}_k \triangleq \begin{pmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{C}_k & \mathbf{D}_k \end{pmatrix} = \begin{pmatrix} \mathbf{M}_k & \mathbf{0} \\ \mathbf{M}_k^{-\mathrm{T}}\boldsymbol{\Omega}_{k-1} & \mathbf{M}_k^{-\mathrm{T}} \end{pmatrix}. \tag{A.3}
$$

Furthermore, let us introduce the following matrix in $\mathbb{R}^{2n \times n}$

$$
\mathbf{W}_k = \begin{pmatrix} \mathbf{X}_k \\ \mathbf{Y}_k \end{pmatrix} \tag{A.4}
$$

where $\mathbf{Y}_k$ is assumed to be invertible, which can and will be checked a posterior, and we define the ratio $\boldsymbol{\omega}_k = \mathbf{X}_k\mathbf{Y}_k^{-1}$ in $\mathbb{R}^{n \times n}$. The $\mathbf{W}_k$ are related by the defining recurrence

$$
\mathbf{W}_{k+1} \triangleq \mathbf{Z}_k\mathbf{W}_k. \tag{A.5}
$$

We explicitly have

$$
\begin{pmatrix} \mathbf{X}_{k+1} \\ \mathbf{Y}_{k+1} \end{pmatrix} \triangleq \mathbf{Z}_k\mathbf{W}_k = \begin{pmatrix} \mathbf{A}_k & \mathbf{B}_k \\ \mathbf{C}_k & \mathbf{D}_k \end{pmatrix}\begin{pmatrix} \mathbf{X}_k \\ \mathbf{Y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{A}_k\mathbf{X}_k + \mathbf{B}_k\mathbf{Y}_k \\ \mathbf{C}_k\mathbf{X}_k + \mathbf{D}_k\mathbf{Y}_k \end{pmatrix}, \tag{A.6}
$$

from which it is possible to infer the following recurrence on $\boldsymbol{\omega}_k$

$$
\begin{aligned}
\boldsymbol{\omega}_{k+1} &= \mathbf{X}_{k+1}\mathbf{Y}_{k+1}^{-1} = (\mathbf{A}_k\mathbf{X}_k + \mathbf{B}_k\mathbf{Y}_k)(\mathbf{C}_k\mathbf{X}_k + \mathbf{D}_k\mathbf{Y}_k)^{-1} \\
&= (\mathbf{A}_k\mathbf{X}_k\mathbf{Y}_k^{-1} + \mathbf{B}_k)(\mathbf{C}_k\mathbf{X}_k\mathbf{Y}_k^{-1} + \mathbf{D}_k)^{-1} \\
&= (\mathbf{A}_k\boldsymbol{\omega}_k + \mathbf{B}_k)(\mathbf{C}_k\boldsymbol{\omega}_k + \mathbf{D}_k)^{-1}. \tag{A.7}
\end{aligned}
$$

Hence, we can represent the nonlinear update of $\boldsymbol{\omega}_k$ by the linear recurrence Eq. (A.5).

Now, we choose

$$
\mathbf{X}_0 = \mathbf{P}_0 \quad \text{and} \quad \mathbf{Y}_0 = \mathbf{I}_0, \tag{A.8}
$$

in order to have $\boldsymbol{\omega}_k = \mathbf{P}_k$ for all $k \geq 0$, which implies that the nonlinear recurrence on $\mathbf{P}_k$ can be represented by the linear recurrence Eq. (A.5).

In so far, no assumption on the rank of $\mathbf{P}_k$ was required and, even in presence of model noise, the linear representation implies that $\mathbf{P}_k$ has the following dependence on $\mathbf{P}_0$

$$\mathbf{P}_k = (\mathbf{A}^{(k)}\mathbf{P}_0 + \mathbf{B}^{(k)})(\mathbf{C}^{(k)}\mathbf{P}_0 + \mathbf{D}^{(k)})^{-1} \tag{A.9}$$

where the $\mathbf{A}^{(k)}$, $\mathbf{B}^{(k)}$, $\mathbf{C}^{(k)}$, $\mathbf{D}^{(k)}$ only depend on $\boldsymbol{\Omega}_l$, $\mathbf{Q}_l$ and $\mathbf{M}_l$, $1 \leq l \leq k$. Our purpose now is to compute $\mathbf{P}_k$ for any t_k *in the perfect model* case using the linear representation Eq. (A.3). This is the focus of the rest of this appendix.

A.2. Solution in the autonomous case. We consider first the autonomous case, where $\mathbf{M}_k$, $\boldsymbol{\Omega}_k$ and $\mathbf{Z}_k$ are all independent of time, and we can suppress the time index from the notation. Hence, we would compute the power iterates $\mathbf{Z}^k$ of $\mathbf{Z}$ (not to be confused with the $\mathbf{Z}_k$ defined in Eq. (A.3). Let us assume that $\mathbf{Z}^k$ has the form

$$\mathbf{Z}^k \triangleq \begin{pmatrix} \mathbf{M}^k & \mathbf{0} \\ \left(\mathbf{M}^k\right)^{-\mathrm{T}}\boldsymbol{\Theta}'_k & \left(\mathbf{M}^k\right)^{-\mathrm{T}} \end{pmatrix}. \tag{A.10}$$

Note that we want $\mathbf{Z}^0 = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$, so that $\boldsymbol{\Theta}'_0 = \mathbf{0}$. Then the recurrence on $\mathbf{Z}^k$ imposes the recurrence on the $\boldsymbol{\Theta}'_k$

$$\boldsymbol{\Theta}'_{k+1} = \mathbf{M}^{\mathrm{T}}\boldsymbol{\Theta}'_k\mathbf{M} + \boldsymbol{\Omega}, \tag{A.11}$$

which identifies $\boldsymbol{\Theta}'_k$ with $\boldsymbol{\Theta}_k$ as defined by Eq. (2.14). Because $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Omega}$ are symmetric, all $\boldsymbol{\Theta}_k$ for $k \geq 1$ are also symmetric. We can see it as an arithmetico-geometrico recurrence and define by $\boldsymbol{\Psi}$ the solution of

$$\boldsymbol{\Psi} = \mathbf{M}^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{M} + \boldsymbol{\Omega}. \tag{A.12}$$

This is the so-called discrete algebraic Lyapunov equation. The existence of a solution (and its uniqueness) depends on $\sigma(\mathbf{M})$ the set of eigenvalues of $\mathbf{M}$ [10]. Let us assume there is at least one such solution $\boldsymbol{\Psi}$. Then, we obtain by subtracting Eq. (A.12) from Eq. (A.11) and by iterating

$$\boldsymbol{\Theta}_k = \boldsymbol{\Psi} - (\mathbf{M}^k)^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{M}^k, \tag{A.13}$$

and

$$\mathbf{Z}^k = \begin{pmatrix} \mathbf{M}^k & \mathbf{0} \\ \left(\mathbf{M}^k\right)^{-\mathrm{T}}\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{M}^k & \left(\mathbf{M}^k\right)^{-\mathrm{T}} \end{pmatrix}. \tag{A.14}$$

Using the linear representation leads to

$$\begin{pmatrix} \mathbf{X}_k \\ \mathbf{Y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{M}^k & \mathbf{0} \\ \left(\mathbf{M}^k\right)^{-\mathrm{T}}\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{M}^k & \left(\mathbf{M}^k\right)^{-\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{P}_0 \\ \mathbf{I} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{M}^k\mathbf{P}_0 \\ \left\{ \left(\mathbf{M}^k\right)^{-\mathrm{T}}\boldsymbol{\Psi} - \boldsymbol{\Psi}\mathbf{M}^k \right\}\mathbf{P}_0 + \left(\mathbf{M}^k\right)^{-\mathrm{T}} \end{pmatrix}. \tag{A.15}$$

Using $\mathbf{P}_k = \mathbf{X}_k\mathbf{Y}_k^{-1}$, we conclude

$$\mathbf{P}_k = \mathbf{M}^k\mathbf{P}_0 \left[\left\{ \boldsymbol{\Psi} - \left(\mathbf{M}^k\right)^{\mathrm{T}}\boldsymbol{\Psi}\mathbf{M}^k \right\}\mathbf{P}_0 + \mathbf{I} \right]^{-1} \left(\mathbf{M}^k\right)^{\mathrm{T}}. \tag{A.16}$$

A.3. Solution in the non-autonomous case. In the non-autonomous case, we need to define

$$\mathbf{Z}^{(k)} \triangleq \prod_{i=0}^{k} \mathbf{Z}_k. \tag{A.17}$$

The product is still in the symplectic group and of the form

$$\mathbf{Z}^{(k)} \triangleq \begin{pmatrix} \mathbf{M}_{k:0} & \mathbf{0} \\ \mathbf{\Gamma}'_k \mathbf{M}_{k:0} & \mathbf{M}_{k:0}^{-\mathrm{T}} \end{pmatrix}, \tag{A.18}$$

which leads to the following recurrence on $\mathbf{\Gamma}'_k$

$$\mathbf{\Gamma}'_{k+1} = \mathbf{M}_{k+1}^{-\mathrm{T}} \left(\mathbf{\Gamma}'_k + \mathbf{\Omega}_k \right) \mathbf{M}_{k+1}^{-1}. \tag{A.19}$$

The finite-time solution to this recurrence is

$$\mathbf{\Gamma}'_k = \sum_{l=0}^{k-1} \mathbf{M}_{k:l}^{-\mathrm{T}} \mathbf{\Omega}_l \mathbf{M}_{k:l}^{-1} \tag{A.20}$$

which coincides with the definition of $\mathbf{\Gamma}_k$ in Eq. (2.7). Hence, we have an expression for $\mathbf{Z}^{(k)}$. We can use it to obtain a solution for the recurrence on $\mathbf{P}_k$ using the linear representation

$$\mathbf{Z}^{(k)} \begin{pmatrix} \mathbf{P}_0 \\ \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{M}_{k:0} \mathbf{P}_0 \\ \mathbf{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0 + \mathbf{M}_{k:0}^{-\mathrm{T}} \end{pmatrix} \tag{A.21}$$

from which we obtain

$$\begin{aligned}
\mathbf{P}_k &= \mathbf{M}_{k:0} \mathbf{P}_0 \left[\mathbf{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0 + \mathbf{M}_{k:0}^{-\mathrm{T}} \right]^{-1} \\
&= \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^{\mathrm{T}} \left[\mathbf{I} + \mathbf{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0 \mathbf{M}_{k:0}^{\mathrm{T}} \right]^{-1} \\
&= \mathbf{M}_{k:0} \mathbf{P}_0 \left[\mathbf{I} + \mathbf{M}_{k:0}^{\mathrm{T}} \mathbf{\Gamma}_k \mathbf{M}_{k:0} \mathbf{P}_0 \right]^{-1} \mathbf{M}_{k:0}^{\mathrm{T}}
\end{aligned} \tag{A.22}$$

which coincides with Eq. (2.11) and Eq. (2.12).

Appendix B. A few useful properties of the symmetric positive (semi-) definite matrices.

Here we provide a selection of definitions and results about the symmetric positive (semi-)definite matrices that we use in this paper. An introduction and detailed proofs of several of these results can be found in [29, chapter 6].

1. The partial ordering on $\mathcal{C}^n$ is defined by: for any $\mathbf{A}$ and $\mathbf{B}$ in $\mathcal{C}^n$, $\mathbf{A} \leq \mathbf{B}$ if and only if for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} \leq \mathbf{x}^{\mathrm{T}} \mathbf{B} \mathbf{x}$.
2. If $\mathbf{A}$ and $\mathbf{B}$ are in $\mathcal{C}^n$ and $\mathbf{G}$ is in $\mathbb{R}^n$, we have that $\mathbf{A} \leq \mathbf{B}$ implies $\mathbf{G} \mathbf{A} \mathbf{G}^{\mathrm{T}} \leq \mathbf{G} \mathbf{B} \mathbf{G}^{\mathrm{T}}$ which is immediate from the previous definition of the partial ordering.
3. If $\mathbf{A}$ and $\mathbf{B}$ are in $\mathcal{C}_+^n$, $\mathbf{A} \leq \mathbf{B}$ is equivalent to $\mathbf{A}^{-1} \geq \mathbf{B}^{-1}$. This can be shown using the double diagonalization theorem which states that there exists an invertible matrix $\mathbf{G}$ such that $\mathbf{G} \mathbf{A} \mathbf{G}^{\mathrm{T}}$ and $\mathbf{G} \mathbf{B} \mathbf{G}^{\mathrm{T}}$ are both diagonal.
4. If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are in $\mathcal{C}^n$, by $\mathbf{A} \leq \min\{\mathbf{B}, \mathbf{C}\}$ we mean that for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} \leq \min\{\mathbf{x}^{\mathrm{T}} \mathbf{B} \mathbf{x}, \mathbf{x}^{\mathrm{T}} \mathbf{C} \mathbf{x}\}$.

5. If $\mathbf{A}$ is in $\mathcal{C}^n$, it has the eigendecomposition $\mathbf{A} = \sum_{i=1}^{n} \sigma_i \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$, with $\sigma_i \geq 0$ and $\{\mathbf{v}_i\}_{1 \leq i \leq n}$ an orthonormal basis. Let $\sigma_{\max} = \max_{1 \leq i \leq n} \sigma_i$ and $\sigma_{\min} = \min_{1 \leq i \leq n} \sigma_i$ For any $\mathbf{x}$ in $\mathbb{R}^n$, we have the decomposition $\mathbf{x} = \sum_{i=1}^{n} \left(\mathbf{v}_i^{\mathrm{T}} \mathbf{x} \right) \mathbf{v}_i$ on the eigenvectors of $\mathbf{A}$. As a consequence, we have

$$\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = \sum_{i=1}^{n} \sigma_i \left(\mathbf{v}_i^{\mathrm{T}} \mathbf{x} \right)^2 \leq \sigma_{\max} \sum_{i=1}^{n} \left(\mathbf{v}_i^{\mathrm{T}} \mathbf{x} \right)^2 = \sigma_{\max} \mathbf{x}^{\mathrm{T}} \mathbf{x}$$

$$\geq \sigma_{\min} \sum_{i=1}^{n} \left(\mathbf{v}_i^{\mathrm{T}} \mathbf{x} \right)^2 = \sigma_{\min} \mathbf{x}^{\mathrm{T}} \mathbf{x} \qquad \text{(B.1)}$$

which leads to $\sigma_{\min} \mathbf{I} \leq \mathbf{A} \leq \sigma_{\max} \mathbf{I}$. Further, as

$$\|\mathbf{A} \mathbf{x}\|^2 = \sum_{i=1}^{n} \sigma_i^2 \left(\mathbf{v}_i^{\mathrm{T}} \mathbf{x} \right)^2, \qquad \text{(B.2)}$$

it follows that $\mathbf{x}^{\mathrm{T}} \mathbf{A} \mathbf{x} = 0 \iff \mathbf{A} \mathbf{x} = \mathbf{0}$.

Now, assume $\{\mathbf{A}_k\}_{k \in \mathbb{N}}$ is a uniformly bounded sequence in $\mathcal{C}^n$ and $\{\mathbf{x}_k\}_{k \in \mathbb{N}}$ is a uniformly bounded sequence in $\mathbb{R}^n$. Then $\lim_{k \to \infty} \mathbf{A}_k \mathbf{x}_k = \mathbf{0}$ implies that $\lim_{k \to \infty} \mathbf{x}_k^{\mathrm{T}} \mathbf{A}_k \mathbf{x}_k = 0$ by virtue of the boundedness of $\mathbf{x}_k$. Owing to the uniform boundedness of $\mathbf{A}_k$, we introduce $\sigma = \sup_{k \in \mathbb{N}, 1 \leq i \leq n} \sigma_{k,i} < \infty$ and obtain

$$\|\mathbf{A}_k \mathbf{x}_k\|^2 = \sum_{i=1}^{n} \sigma_{k,i}^2 \left(\mathbf{v}_{k,i}^{\mathrm{T}} \mathbf{x}_k \right)^2$$

$$\leq \sigma \sum_{i=1}^{n} \sigma_{k,i} \left(\mathbf{v}_{k,i}^{\mathrm{T}} \mathbf{x}_k \right)^2 \leq \sigma \mathbf{x}_k^{\mathrm{T}} \mathbf{A}_k \mathbf{x}_k. \qquad \text{(B.3)}$$

Hence, $\lim_{k \to \infty} \mathbf{x}_k^{\mathrm{T}} \mathbf{A}_k \mathbf{x}_k = 0$ implies that $\lim_{k \to \infty} \mathbf{A}_k \mathbf{x}_k = \mathbf{0}$. It follows that $\mathbf{x}_k^{\mathrm{T}} \mathbf{A}_k \mathbf{x}_k = 0 \iff \mathbf{A}_k \mathbf{x}_k = \mathbf{0}$.

6. Let $\mathbf{A} \in \mathcal{C}^n$ and $\alpha \geq 0$ be a constant. If there is a subspace $\mathcal{W} \subseteq \mathbb{R}^n$ of dimension $s \geq 1$ such that for all unit vectors $\mathbf{h} \in \mathcal{W}$, $\mathbf{h}^{\mathrm{T}} \mathbf{A} \mathbf{h} \leq \alpha$, then $\mathbf{A}$ has at least s of its eigenvalues less than or equal to α.

To see this, decompose $\mathbf{A} = \sum_{i=1}^{n} \sigma_i \mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}$ in its orthonormal eigenbasis where $\sigma_i \geq 0$ and ordered as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. Consider $\mathcal{V}$ the $s-1$-dimensional subspace span of $\{\mathbf{v}_{n-s+2}, \ldots, \mathbf{v}_n\}$, which we take to be the null space if $s = 2$. The orthogonal subspace $\mathcal{V}^\perp$ of $\mathcal{V}$ is of dimension $n - s + 1$. The intersection $\mathcal{W} \cup \mathcal{V}^\perp$ is of dimension at least 1. Let us pick $\mathbf{h}$ of Euclidean norm 1 in this intersection. We have

$$\alpha \geq \mathbf{h}^{\mathrm{T}} \mathbf{A} \mathbf{h} = \sum_{i=1}^{n} \sigma_i (\mathbf{h}^{\mathrm{T}} \mathbf{v}_i)^2 = \sum_{i=1}^{n-s+1} \sigma_i (\mathbf{h}^{\mathrm{T}} \mathbf{v}_i)^2$$

$$\geq \sigma_{n-s+1} \sum_{i=1}^{n-s+1} (\mathbf{h}^{\mathrm{T}} \mathbf{v}_i)^2 = \sigma_{n-s+1}. \qquad \text{(B.4)}$$

Hence $\alpha \geq \sigma_{n-s+1} \geq \cdots \geq \sigma_n$.

7. Let $\mathbf{P} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^{\mathrm{T}} & \mathbf{C} \end{pmatrix}$ in $\mathcal{C}^n$ with $\mathbf{A}$ of size $n_0 \times n_0$ ($n_0 \leq n$) and full-rank. As a consequence of $\mathbf{P} \in \mathcal{C}^n$ and $\mathrm{rank}(\mathbf{A}) = n_0$, we have $\mathbf{A} \in \mathcal{C}_+^n$ and $\mathbf{C} \in \mathcal{C}^{n-n_0}$.

23

We can write $\mathbf{C} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$ where $\mathbf{\Lambda} \in \mathcal{C}_+^q$ is diagonal with $q \leq n - n_0$ and $\mathbf{U} \in \mathbb{R}^{(n-n_0) \times q}$ is such that $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$. We define $\mathbf{\Pi_C} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ as the orthogonal projector onto the column space of $\mathbf{C}$. Moreover, if $\mathbf{u} \in \mathbb{R}^{n-n_0}$ is in the null space of $\mathbf{C}$, $\mathbf{v} = (\mathbf{0}, \mathbf{u}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^n$ is in the null space of $\mathbf{P}$ by virtue of point 5 of this appendix, so that $\mathbf{u}$ is also in the null space of $\mathbf{B}$. It follows that $\mathbf{B}\,(\mathbf{I} - \mathbf{\Pi_C}) = \mathbf{0}$ since $\mathbf{I} - \mathbf{\Pi_C}$ is the orthogonal projector onto the null space of $\mathbf{C}$. Hence, we can write

$$\mathbf{P} = \begin{pmatrix} \mathbf{A} & \mathbf{B}\mathbf{\Pi_C} \\ (\mathbf{B}\mathbf{\Pi_C})^{\mathrm{T}} & \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{U} \end{pmatrix} \widetilde{\mathbf{P}} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}^{\mathrm{T}} \end{pmatrix} \tag{B.5}$$

where

$$\widetilde{\mathbf{P}} \triangleq \begin{pmatrix} \mathbf{A} & \mathbf{B}\mathbf{U} \\ (\mathbf{B}\mathbf{U})^{\mathrm{T}} & \mathbf{\Lambda} \end{pmatrix} \tag{B.6}$$

which is in $\mathcal{C}_+^{n_0+q}$ since $\widetilde{\mathbf{P}} \in \mathcal{C}^{n_0+q}$, $\mathbf{A}$ is in $\mathcal{C}_+^{n_0}$ and $\mathbf{\Lambda}$ is in $\mathcal{C}_+^q$. Denoting σ the smallest eigenvalue of $\widetilde{\mathbf{P}}$, which is positive, we have $\widetilde{\mathbf{P}} \geq \sigma\mathbf{I}$ and, from Eq. (B.5),

$$\mathbf{P} \geq \sigma \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Pi_C} \end{pmatrix} \geq \sigma \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{B.7}$$

If $\mathbf{\Pi}_+ \in \mathbb{R}^{n \times n}$ is the orthogonal projector onto the row space of $(\mathbf{A}, \mathbf{0})$, we have in particular $\mathbf{P} \geq \sigma\mathbf{\Pi}_+$.

Appendix C. Matrix shift lemma. Assuming $x \mapsto f(x)$ can be written as a formal power series, i.e. $f(x) = \sum_{i=0}^{\infty} a_i x^i$, one has $\mathbf{A}f(\mathbf{B}\mathbf{A}) = \sum_{i=0}^{\infty} a_i \mathbf{A}(\mathbf{B}\mathbf{A})^i = \sum_{i=0}^{\infty} a_i (\mathbf{A}\mathbf{B})^i \mathbf{A} = f(\mathbf{A}\mathbf{B})\mathbf{A}$. This proves the matrix shift lemma, i.e. $\mathbf{A}f(\mathbf{B}\mathbf{A}) = f(\mathbf{A}\mathbf{B})\mathbf{A}$.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1979.
[2] M. Bocquet, C. A. Pires, and L. Wu, *Beyond Gaussian statistical modeling in geophysical data assimilation*, Mon. Wea. Rev., 138 (2010), pp. 2997–3023.
[3] M. Bocquet, P. N. Raanes, and A. Hannart, *Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation*, Nonlin. Processes Geophys., 22 (2015), pp. 645–662.
[4] M. Bocquet and P. Sakov, *An iterative ensemble Kalman smoother*, Q. J. R. Meteorol. Soc., 140 (2014), pp. 1521–1535.
[5] P. Bougerol, *Kalman filtering with random coefficients and contractions*, SIAM J. Control Optim., 31 (1993), pp. 942–959.
[6] A. Carrassi, M. Ghil, A. Trevisan, and F. Uboldi, *Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system*, Chaos, 18 (2008), p. 023112.
[7] 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*, Nonlin. Processes Geophys., 15 (2008), pp. 503–521.
[8] A. Carrassi, S. Vannitsem, D. Zupanski, and M. Zupanski, *The maximum likelihood ensemble filter performances in chaotic systems*, Tellus A, 61 (2009), pp. 587–600.
[9] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*, Springer-Verlag Berlin Heidelberg, second ed., 2009.
[10] Z. Gajić and M. T. J. Qureshi, *Lyapunov matrix equation in system stability and control*, Academic Press, San Diego, California, 1995.

[11] K. Gurumoorthy, C. Grudzien, A. Apte, A. Carrassi, and C. Jones, *Rank deficiency of Kalman error covariance matrices in linear perfect model*, arXiv preprint arXiv:1503.05029, 0 (2015), pp. 0–0.

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

[13] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Fluids Engineering, 82 (1960), pp. 35–45.

[14] B. Legras and R. Vautard, *A guide to lyapunov vectors*, in ECMWF Workshop on Predictability, Reading, United-Kingdom, 1996, ECMWF, pp. 135–146.

[15] E. N. Lorenz and K. A. Emanuel, *Optimal sites for supplementary weather observations: simulation with a small model*, J. Atmos. Sci., 55 (1998), pp. 399–414.

[16] G.-H. C. Ng, D. McLaughlin, D. Entekhabi, and A. Ahanin, *The role of model dynamics in ensemble Kalman filter performance for chaotic systems*, Tellus A, 63 (2011), pp. 958–977.

[17] L. Palatella, A. Carrassi, and A. Trevisan, *Lyapunov vectors and assimilation in the unstable subspace: theory and applications*, J. Phys. A: Math. Theor., 46 (2013), p. 254020.

[18] L. Palatella and A. Trevisan, *Interaction of Lyapunov vectors in the formulation of the nonlinear extension of the Kalman filter*, Phys. Rev. E, 91 (2015), p. 042905.

[19] 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), p. 022901.

[20] T. S. Parker and L. O. Chua, *Practical numerical algorithms for chaotic systems*, Springer-Verlag New York, 1989.

[21] P. Sakov and L. Bertino, *Relation between two common localisation methods for the EnKF*, Comput. Geosci., 15 (2011), pp. 225–237.

[22] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*, John Wiley & Sons, Inc., 2006.

[23] A. Trevisan, M. D'Isidoro, and O. Talagrand, *Four-dimensional variational assimilation in the unstable subspace and the optimal subspace dimension*, Q. J. R. Meteorol. Soc., 136 (2010), pp. 487–496.

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

[25] A. Trevisan and F. Pancotti, *Periodic orbits, Lyapunov vectors, and singular vectors in the Lorenz system*, J. Atmos. Sci., 55 (1998), pp. 390–398.

[26] 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.

[27] 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*, Nonlin. Processes Geophys., 16 (2006), pp. 67–81.

[28] M. P. Wojtowski, *Geometry of Kalman fiters*, J. Geom. and Symmetry in Physics, 9 (2007), pp. 83–95.

[29] F. Zhang, *Matrix theory: basic results and techniques*, Springer-Verlag New-York Inc., 1999.