ON THE STABILITY AND THE UNIFORM PROPAGATION OF CHAOS PROPERTIES OF ENSEMBLE KALMAN–BUCY FILTERS

BY P. DEL MORAL AND J. TUGAUT

Center INRIA Bordeaux Sud-Ouest & Institut de Mathématiques de Bordeaux and Université de Lyon, Université Jean Monnet

The ensemble Kalman filter is a sophisticated and powerful data assimilation method for filtering high dimensional problems arising in fluid mechanics and geophysical sciences. This Monte Carlo method can be interpreted as a mean-field McKean–Vlasov-type particle interpretation of the Kalman–Bucy diffusions. In contrast to more conventional particle filters and nonlinear Markov processes, these models are designed in terms of a diffusion process with a diffusion matrix that depends on particle covariance matrices.

Besides some recent advances on the stability of nonlinear Langevin-type diffusions with drift interactions, the long-time behaviour of models with interacting diffusion matrices and conditional distribution interaction functions has never been discussed in the literature. One of the main contributions of the article is to initiate the study of this new class of models. The article presents a series of new functional inequalities to quantify the stability of these nonlinear diffusion processes.

In the same vein, despite some recent contributions on the convergence of the ensemble Kalman filter when the number of sample tends to infinity very little is known on stability and the long-time behaviour of these mean-field interacting type particle filters. The second contribution of this article is to provide uniform propagation of chaos properties as well as $\mathbb{L}_n$-mean error estimates w.r.t. to the time horizon. Our regularity condition is also shown to be sufficient and necessary for the uniform convergence of the ensemble Kalman filter.

The stochastic analysis developed in this article is based on an original combination of functional inequalities and Foster–Lyapunov techniques with coupling, martingale techniques, random matrices and spectral analysis theory.

CONTENTS

1. Introduction.

1.1. *The Ensemble Kalman filter.* The Ensemble Kalman filter (*abbreviated EnKF*) has been introduced by G. Evensen in the seminal article [26] published in 1994. In the last two decades, the EnKF has become one of the main numerical techniques for solving high dimensional forecasting and data assimilation problems, particularly in ocean and atmosphere sciences [2, 35, 44, 46, 54], weather forecasting [4, 5, 16, 32], environmental and ecological statistics [25, 34], as well as in oil reservoir simulations [27, 52, 57, 58, 67], and many others. We also refer the reader to [10, 24] for recent reviews on Riccati equations, estimation and linear filtering techniques.

The mathematical foundations and the convergence of the EnKF are more recent. They have started in 2011 with the independent pioneering works of F. Le Gland, V. Monbet and V. D. Tran [43], and the one by J. Mandel, L. Cobb, J. D. Beezley [48]. These articles provide $\mathbb{L}_n$-mean error estimates for discrete time EnKF and show that the EnKF converges towards the Kalman filter as the number of samples tends to infinity. In a more recent study by X. T. Tong, A. J. Majda and D. Kelly the authors analyze the long-time behaviour and the ergodicity of discrete generation EnKF using Foster–Lyapunov techniques ensuring that the filter is asymptotically stable w.r.t. any erroneous initial condition [64]. These important properties ensure that the EnKF has a single invariant measure and initialization errors of the EnKF will dissipate w.r.t. the time parameter.

Beside the importance of these properties, the only ergodicity of the particle process does not give any information of the convergence and the accuracy of the EnKF towards the optimal filter as the number of samples tends to infinity.

One of the main objective of this article is to analyze this convergence and quantify the fluctuation of errors on large-time horizon. We provide uniform $\mathbb{L}_n$-mean error estimates w.r.t. the time parameter for the sample mean as well as for the sample covariance matrices. Incidentally, the stochastic analysis we have developed also allows to quantify the stability properties of the Kalman–Bucy filter and the corresponding matrix valued Riccati equations. These estimates are deduced from the stability properties of a nonlinear diffusion interpretation of the Kalman–Bucy filter equations.

To better connect this work with existing literature on nonlinear Markov processes and particle methods, we emphasize that the EnKF can be seen as a mean-field particle interpretation of a nonlinear McKean–Vlasov-type diffusion. These probabilistic models were introduced in the end of the 1960s by H. P. McKean [49]. For a detailed discussion on these models and their application domains, we refer the reader to the lecture notes of A. S. Sznitman [63], the ones by S. Méléard [50], and the research monograph [21].

The refined convergence as well as the long-time behaviour of nonlinear diffusion processes is still an active research area. When the interaction function only enters in the drift part of the diffusion several results including uniform estimates w.r.t., the time horizon are available [11, 17, 22, 23, 47]. Most of these works are based on powerful and sophisticated coupling methods, nonlinear semigroup analysis, as well as Gamma-two type techniques and optimal transport theory.

In our context, the Kalman–Bucy filter and the Riccati equation represent the evolution equations of the mean and the covariance matrices of the random states of a nonlinear diffusion process. We shall call this process the Kalman–Bucy diffusion. The diffusion part of this class of processes depends on the covariance matrix of its random states. The recent techniques developed in nonlinear Markov processes theory are not suited to analyze the stability of these complex nonlinear processes. To the best of our knowledge, the long-time behaviour of such nonlinear diffusions with covariances matrices depending on the distribution of the random states remains an open and important research question.

In the present article, we initiate the study of the stability of this class of non-linear diffusion models. We present a series of functional inequalities to quantify the stability of these nonlinear diffusion processes. We also analyze the exponential stability of these processes w.r.t. Wasserstein distances and relative entropy inequalities. The stability properties of the Kalman–Bucy filter are deduced by a direct application of Jensen-type inequalities.

At the level of the particle population model, the EnKF also belongs to the class of mean-field type particle filters. The stochastic analysis of particle filters and related diffusions Monte Carlo schemes is rather well understood; see, for instance, [20, 21] and the references therein. Nevertheless, the EnKF strongly differs from particle filters or sequential Monte Carlo methods currently used in nonlinear filtering theory, Bayesian inference and computational physics. Roughly speaking, the EnKF is designed to approximate the Kalman filter (as well as the extended Kalman filter) for high dimensional problems. In the reverse angle, particle filters are designed to estimate the nonlinear filtering equation, and to sample sequentially according to the flow of conditional distributions. In continuous time settings, the EnKF is an interacting diffusion while particle filters are interacting jump particle systems. As a result, none of the techniques developed in particle filtering theory applies to analyze the fluctuations of the EnKF uniformly w.r.t. the time horizon. It is clearly not the scope of this article to compare in full details these two particle filtering methods. For a more thorough discussion on particle filtering techniques, we refer the reader to [20, 21], and the references therein.

In the same vein, the stochastic analysis developed so far in the literature on more general classes of mean-field particle methods cannot be used to analyze the uniform convergence of particle approximating schemes involving interacting covariances matrices. As mentioned above, the EnKF belongs to this class of nonlinear diffusions with a mean-field particle interpretation based on covariance matrices of interacting multidimensional particles. To the best of our knowledge, the uniform propagation of chaos estimates developed in the present article seems to be the first result of this type for this class of nonlinear diffusions.

To derive these uniform estimates, we develop a novel stochastic fluctuation analysis which combines Foster–Lyapunov techniques with matrix valued martingale methods, as well as random matrices and spectral analysis theory. The central idea is to take advantage of the linear-Gaussian structure of the filtering problem to enter the stability properties of the signal process and the (nonlinear) Riccati matrix-valued equation into the fluctuation analysis of the EnKF. We also prove that the stability property of the signal is a sufficient and necessary condition to obtain uniform propagation of chaos estimates.

1.2. *Organization of the article*. The article is organized as follows:

Section 2 is dedicated to the description of the Kalman–Bucy filter, the nonlinear diffusion process interpretation of the filter, as well as the mean-field EnKF particle algorithm. In Section 3, we state some of the main theorems of the article.

The first one shows that the sample mean and the random interacting covariance matrices of the EnKF satisfy the same equation as the EnKF and the Riccati equation up to some fluctuation martingales whose angle brackets only depend on the sample covariance matrices. These diffusion equations in matrix spaces are pivotal as they allow to analyze the fluctuations of the EnKF using Foster–Lyapunov and martingale techniques combined with trace and spectral-type inequalities.

The second theorem provides uniform convergence and propagations of chaos estimates w.r.t. the time parameters.

Section 4 provides a detailed discussion on our regularity conditions. In Section 4.1, we analyze the stability properties and the catastrophic divergence issues of EnFK filters in terms of global divergence regions and ill-conditioned filtering problems. We analyze the propagations of the fluctuations induced by the sample covariance matrices in terms of observer-type filters and stochastic Ornstein–Ulhenbeck diffusions. In control theory, the terminology "observer" is often restricted to deterministic models.

As its name indicates a stochastic observer is a stochastic process that uses sensory history to estimate the true signal; the randomness comes from the fact that the perturbations of the sensor are random. We design and we analyze the long time behaviour of a class of stochastic observer driven by stochastic covariance matrices. We also discuss some pivotal semigroup contraction properties in terms of log-norms of matrices. Several illustrations are provided in Section 4.2.

Section 5 discusses the stability properties of Kalman–Bucy diffusions. Section 5.2 is dedicated to uniform contraction inequalities for the nonlinear semigroups associated with the Riccati equation and Kalman–Bucy diffusions. Section 5.3 presents some local functional inequalities to estimate the fluctuations of the models around their steady state version w.r.t. the Wasserstein distance and the relative entropy.

The remainder of the article is mainly concerned with the proof of the main theorems presented in Section 3 and Section 5.

Section 6 presents some technical preliminary results used in the further development of the article. Section 6.1 shows that our regularity conditions that ensure the uniform convergence of the EnKF are sharp and cannot be relaxed. The section also provides some uniform convergence estimates on the filter, the signal states, and the Riccati equation. It also presents some semigroup estimates and related trace inequalities of current use in this study. Section 6.2 is dedicated to the Riccati equation. We analyze the explicit solution in the one-dimensional case and we present a trace-type comparison lemma to analyze multivariate models.

Section 7 is concerned with the stochastic analysis of the EnKF. Section 7.1 is dedicated to the proof of the stochastic differential equations satisfied by the EnKF sample mean and the particle covariance matrices. Section 7.2 is dedicated to uniform moments estimates for the trace of the particle covariance matrices and the random states of the EnKF. These results are deduced from a technical lemma, of

its own interest; combining Foster–Lyapunov with martingale techniques to control the moments of Riccati-type stochastic differential equations uniformly w.r.t. the time horizon.

Section 8 is mainly concerned with the detailed proofs of the uniform propagation of chaos theorem presented in Section 3.

The final section, Section 12, presents a brief summary of the contributions of the article and proposes an avenue of open research projects.

1.3. *Some basic notation and preliminary results.* This section provides with some notation and terminology used in several places in the article. Given some random variable Z with some probability measure μ and some function f on some product space $\mathbb{R}^r$, we let $\mu(f) = \mathbb{E}(f(Z)) = \int f(x)\mu(dx)$ be the integral of f w.r.t. μ or the expectation of $f(X)$. As a rule any multivariate random variable, say Z, is represented by a column vector and we use the transposition operator Z' to denote the row vector. Given a distribution η on some product space $\mathbb{R}^r$ and some measurable function f from $\mathbb{R}^r$ into $\mathbb{R}^r$, we set $\eta(f) = (\eta(f_i))_{1 \leq i \leq r}$ the column vector with entries $\eta(f_i)$ where f_i stands for the ith coordinate mapping from $\mathbb{R}^r$ into $\mathbb{R}$. We also denote by $a_+ = \max(a, 0)$ the real part of a number $a \in \mathbb{R}$.

We let $\| \cdot \|$ be the Euclidean norm on $\mathbb{R}^r$, for some $r \geq 1$. We denote by $\mathbb{S}_r$ the set of $(r \times r)$ symmetric matrices with real entries, and by $\mathbb{S}_r^+$ the subset of positive definite matrices. We let $\mathrm{Spec}(A)$ be the set of eigenvalues of a square matrix A. With a slight abuse of notation, we denote by Id the $(r \times r)$ identity matrix, for any $r \geq 1$.

We often denote by $\lambda_i(A)$, with $1 \leq i \leq r$, the nonincreasing sequence of eigenvalues of a symmetric $(r \times r)$-matrix A. We also often denote by $\lambda_{\min}(A) = \lambda_r(A)$ and $\lambda_{\max}(A) = \lambda_1(A)$ the minimal and the maximal eigenvalue. We also set $A_{\mathrm{sym}} := (A + A')/2$ for any $(r \times r)$-square matrix A. We recall that the norm $\|A\|$ and logarithmic norm $\mu(A)$ of an $(r_1 \times r_1)$-square matrix A are defined by $\|A\| := \sup_{\|x\|=1} \|Ax\|$ and

$$
\begin{aligned}
\mu(A) &:= \inf \left\{ \alpha : \forall x \, \langle x, Ax \rangle \leq \alpha \|x\|^2 \right\} = \lambda_{\max}(A_{\mathrm{sym}}) \\
&= \inf \left\{ \alpha : \forall t \geq 0 \, \|\exp(At)\|_2 \leq \exp(\alpha t) \right\}.
\end{aligned}
\tag{1}
$$

The above equivalent formulations show that

$$
\mu(A) \geq \varsigma(A) := \max \left\{ \mathrm{Re}(\lambda) : \lambda \in \mathrm{Spec}(A) \right\},
$$

where $\mathrm{Re}(\lambda)$ stands for the real part of the eigenvalues λ. The parameter $\varsigma(A)$ is often called the spectral abscissa of A. Also notice that A_{sym} is negative semidefinite as soon as $\mu(A) < 0$. The Frobenius matrix norm of a given $(r_1 \times r_2)$ matrix A is defined by

$$
\|A\|_F^2 = \mathrm{tr}(A'A) \qquad \text{with the trace operator } \mathrm{tr}(\cdot).
$$

If A is a matrix $r \times r$, we have $\|A\|_F^2 = \sum_{1 \leq i,j \leq r} A(i,j)^2 \geq \|A\|^2$.

We also need to consider the nth Wasserstein distance between two probability measures ν_1 and ν_2 on $\mathbb{R}^r$ defined by

$$\mathbb{W}_n(\nu_1, \nu_2) = \inf \left\{ \mathbb{E}\left(\|Z_1 - Z_2\|^n\right)^{\frac{1}{n}} \right\}.$$

The infimum in the above displayed formula is taken of all pairs of random variables (Z_1, Z_2) such that $\mathrm{Law}(Z_i) = \nu_i$, with $i = 1, 2$. We denote by $\mathrm{Ent}(\nu_1 | \nu_2)$ the Boltzmann-relative entropy

$$\mathrm{Ent}(\nu_1 | \nu_2) := \int \log\left(\frac{d\nu_1}{d\nu_2}\right) d\nu_1 \quad \text{if } \nu_1 \ll \nu_2, \text{ and } +\infty \text{ otherwise.}$$

The state transition matrix associated with a smooth flow of $(r \times r)$-matrices $A : u \mapsto A_u$ is denoted by

$$\mathcal{E}_{s,t}(A) = \exp\left[\oint_s^t A_u \, du\right]$$

$$\iff \quad \frac{d}{dt}\mathcal{E}_{s,t}(A) = A_t \mathcal{E}_{s,t}(A) \quad \text{and} \quad \partial_s \mathcal{E}_{s,t}(A) = -\mathcal{E}_{s,t}(A) A_s$$

for any $s \leq t$, with $\mathcal{E}_{s,s} = \mathrm{Id}$, the identity matrix. Equivalently, in terms of the fundamental solution matrices $\mathcal{E}_t(A) := \mathcal{E}_{0,t}(A)$ we have $\mathcal{E}_{s,t}(A) = \mathcal{E}_t(A)\mathcal{E}_s(A)^{-1}$. Observe that for any $s \leq r \leq t$ the exponential semigroup property

$$\exp\left[\oint_s^t A_u \, du\right] = \exp\left[\oint_r^t A_u \, du\right] \exp\left[\oint_s^r A_u \, du\right].$$

The following technical lemma provides a pair of semigroup estimates of the state transition matrices associated with a sum of drif-type matrices.

LEMMA 1.1 (Perturbation lemma). *Let $A : u \mapsto A_u$ and $B : u \mapsto B_u$ be some smooth flows of $(r \times r)$-matrices. For any $s \leq t$, we have*

$$\left\|\mathcal{E}_{s,t}(A + B)\right\|_2 \leq \exp\left(\int_s^t \mu(A_u) \, du + \int_s^t \|B_u\|_2 \, du\right).$$

In addition, for any matrix norm $\|\cdot\|$ we have

$$\left\|\mathcal{E}_{s,t}(A + B)\right\| \leq \alpha_A \exp\left[-\beta_A(t - s) + \alpha_A \int_s^t \|B_u\| \, du\right]$$

as soon as

$$\forall 0 \leq s \leq t \qquad \left\|\mathcal{E}_{s,t}(A)\right\| \leq \alpha_A \exp\left(-\beta_A(t - s)\right).$$

These estimates are probably well known but we have not found a precise reference. For the convenience of the reader, the detailed proof of this lemma is housed in the Appendix, on page 840. For time homogeneous matrices $A_t = A$,

the state transition matrix reduces to the conventional matrix exponential $\mathcal{E}_{s,t}(A) = e^{(t-s)A} = \mathcal{E}_{t-s}(A)$.

The norm of $\mathcal{E}_t(A)$ can be estimated in various ways: The first one is based on the Jordan decomposition $T^{-1}AT = J$ decomposition of the matrix A in terms of k Jordan blocks associated with the eigenvalues with multiplicities m_i, with $1 \le i \le k$. In this situation, we have the Jordan-type estimate

$$(2) \qquad e^{\varsigma(A)t} \le \|\mathcal{E}_t(A)\|_2 \le \kappa_{\mathrm{Jor},t}(T)e^{\varsigma(A)t}$$

with

$$\kappa_{\mathrm{Jor},t}(T) = \left(\bigvee_{0 \le j < n} \frac{t^j}{j!}\right)\|T\|_2\|T^{-1}\|_2 \quad \text{and} \quad n := \bigvee_{1 \le i \le k} m_i.$$

Observe that $\kappa_{\mathrm{Jor},t}(T)$ depends on the time horizon t as soon as A is not of full rank. In addition, whenever A is close to singular, the conditioning number $\mathrm{cond}(T) := \|T\|_2\|T^{-1}\|_2$ tends to be very large.

A second strategy is based on Schur decomposition $U'AU = D + T$ in terms of an unitary matrix U, with $D = \mathrm{diag}(\lambda_1(A), \ldots; \lambda_r(A))$ and a strictly triangular matrix T s.t. $T_{i,j} = 0$ for any $i \ge j$. In this case, we have the Schur-type estimate

$$(3) \qquad \|\mathcal{E}_t(A)\|_2 \le \kappa_{\mathrm{Sch},t}(T)e^{\varsigma(A)t} \qquad \text{with } \kappa_{\mathrm{Sch},t}(T) := \sum_{0 \le i \le r} \frac{(\|T\|t)^i}{i!}.$$

The proof of these estimates can be found in [51, 65]. In both cases for any $\varepsilon \in]0, 1]$ and any $t \ge 0$, we have

$$(4) \qquad e^{\varsigma(A)t} \le \|\mathcal{E}_t(A)\|_2 \le \kappa(\varepsilon)e^{(1-\varepsilon)\varsigma(A)t}$$

for some constants $\kappa(\varepsilon)$ whose values only depend on the parameters ε. When A is asymptotically stable; that is all its eigenvalues have negative real parts, for any positive definite matrix B we have

$$e^{\varsigma(A)t} \le \|\mathcal{E}_t(A)\|_2 \le \mathrm{cond}(T_B)\exp\left[-t/\|B^{-1/2}T_B B^{-1/2}\|\right]$$

with the positive definite matrix

$$T_B = \int_0^\infty e^{A't}Be^{At}\,dt \qquad \Longleftrightarrow \qquad A'T + TA = -B.$$

In this case, we have $-1/\|B^{-1/2}TB^{-1/2}\| \ge 2\varsigma(A)$. The proof of these estimates can be found in [66] (Theorem 13.6 and exercise 13.11).

Recalling the norm equivalence formulae

$$\|A\|_2^2 = \lambda_{\max}(A'A) \le \mathrm{tr}(A'A) = \|A\|_F^2 \le r\|A\|_2^2$$

for any $(r \times r)$-matrix A, the above estimates are valid if we replace the $\mathbb{L}_2$-norm by the Frobenius norm.

Most of the semigroup analysis and the contraction inequalities developed in this article are based on the logarithm norms instead of the spectral abscissa given by the top (real part of the) eigenvalues. The reasons are two-fold:

First, as its name indicates, the logarithmic norm represents the logarithmic decays of semigroups w.r.t. the $\mathbb{L}_2$-norm [cf. (1)]. These norms facilitate the stability analysis of exponential semigroups.

On the other hand, most of the matrix exponential estimates expressed in terms of spectral abscissas involve numerical constants that depends on the norm of the diagonalization matrix and its inverse, but also on polynomial functions w.r.t. the time parameter. When the matrix has an ill conditioned eigen-system, these constants are generally too large to obtain an effective useful estimate. We refer the reader to the formulae (2) and (3).

For a more thorough discussion on these norms and their used in the stability analysis of homogeneous semigroups of the form e^{tA}, we refer to [73].

We end this section with a couple of rather well-known estimates in matrix theory. For any $(r \times r)$-square matrices (P, Q) by a direct application of Cauchy–Schwarz inequality, we have

$$(5) \qquad \qquad \left| \text{tr}(PQ) \right| \leq \|P\|_F \|Q\|_F.$$

For any (symmetric and) positive semidefinite $(r \times r)$-square matrices P and Q, we have

$$(6) \qquad \begin{aligned} \text{tr}(P^2) \leq (\text{tr}(P))^2 \leq r \, \text{tr}(P^2) \quad \text{and} \\ \lambda_{\min}(P) \, \text{tr}(Q) \leq \text{tr}(PQ) \leq \lambda_{\max}(P) \, \text{tr}(Q). \end{aligned}$$

The above inequality is also valid when Q is positive semidefinite and P is symmetric. We check this claim using an orthogonal diagonalization of P and recalling that Q remains positive semidefinite (thus with nonnegative diagonal entries). When both matrices P and Q are negative semidefinite, the right-hand side inequality is still valid if we replace $(\lambda_{\min}(P), \lambda_{\max}(P))$ by $(\lambda_{\max}(P), \lambda_{\min}(P))$.

The tensor product, respectively the symmetric tensor product, of $(r \times r)$-matrices A, B are the $(r^2 \times r^2)$-matrix $A \otimes B$, respectively, $A \otimes_s B$, given by the formulae:

$$(A \otimes B)_{(i,j),(k,l)} = A_{i,k} B_{j,l} = (A \otimes B)^{\sharp}_{(i,k),(j,l)},$$
$$4(A \otimes_s B)_{(i,j),(k,l)} = A_{i,k} B_{j,l} + A_{i,l} B_{j,k} + A_{j,l} B_{i,k} + A_{j,k} B_{i,l}.$$

The angle bracket $\langle M \rangle$ of an r-column-vector continuous martingale M is the $(r \times r)$ matrix $\langle M \rangle$ such that

$$MM' - \langle M \rangle$$

is a martingale. More generally, the angle bracket $\langle M \rangle$ of an $(r \times r)$-matrix valued continuous martingale is the $(r^2 \times r^2)$-matrix $\langle M \rangle$ such that

$$M \otimes M' - \langle M \rangle \qquad \text{is a martingale.}$$

2. Description of the models.

2.1. *The Kalman–Bucy filter.* Consider a time homogeneous linear-Gaussian filtering model of the following form:

$$
(7) \qquad \begin{cases} dX_t = (AX_t + a)\, dt + R_1^{1/2}\, dW_t \\ dY_t = (CX_t + c)\, dt + R_2^{1/2}\, dV_t. \end{cases}
$$

In the above display, (W_t, V_t) is an $(r_1 + r_2)$-dimensional Brownian motion, X_0 is a r_1-valued Gaussian random vector with mean and covariance matrix $(\mathbb{E}(X_0), P_0)$ [independent of (W_t, V_t)], the symmetric matrices $R_1^{1/2}$ and $R_2^{1/2}$ are invertible, A is a square $(r_1 \times r_1)$-matrix, C is an $(r_2 \times r_1)$-matrix, a is a given r_1-dimensional column vector and c is an r_2-dimensional column vector, and $Y_0 = 0$. We also let $\mathcal{F}_t = \sigma(Y_s, s \le t)$ be the filtration generated by the observation process.

It is well known that the conditional distribution η_t of the signal state X_t given $\mathcal{F}_t$ is a r_1-dimensional Gaussian distribution with mean and covariance matrix

$$
\widehat{X}_t := \mathbb{E}(X_t | \mathcal{F}_t) \quad \text{and} \quad P_t := \mathbb{E}\big((X_t - \mathbb{E}(X_t|\mathcal{F}_t))(X_t - \mathbb{E}(X_t|\mathcal{F}_t))'\big)
$$

given by the Kalman–Bucy filter

$$
(8) \qquad d\widehat{X}_t = (A\widehat{X}_t + a)\, dt + P_t C' R_2^{-1}\big(dY_t - (C\widehat{X}_t + c)\, dt\big)
$$

and the Riccati equation

$$
(9) \qquad \frac{d}{dt} P_t = \mathrm{Ricc}(P_t)
$$

defined in terms of the quadratic drift function

$$
\mathrm{Ricc} : Q \in \mathbb{S}_{r_1} \mapsto \mathrm{Ricc}(Q) = AQ + QA' - QSQ + R \in \mathbb{S}_{r_1}
$$

with $R = R_1$ and $S := C' R_2^{-1} C$. Note that the covariance matrix is just instrumental. More precisely, the Riccati equation (9) does not depend on the observations, and can be solved offline. Nevertheless, when the dimension of the state is too large, as in most ocean and atmosphere stochastic models, the solving of the Riccati matrix evolution equation is untractable. Besides the problem of storing high dimensional matrices, we often need to resort to spectral technique and change of vector basis to solve analytically the Riccati equation. For high dimensional problems, these spectral techniques cannot be applied and another level of approximation need to be added. The idea of the EnKF is to replace the covariance matrices by sample covariance matrices associated with a well-chosen mean-field particle model. These probabilistic models are defined in more details in the next section.

2.2. *A nonlinear Kalman–Bucy diffusion.* Let $e(x) = x$ denote the identity mapping from $\mathbb{R}^{r_1}$ into itself. For any probability measure η on $\mathbb{R}^{r_1}$, let $m(\eta) =$

$\eta(e)$ and $\mathcal{P}_\eta = \eta[(e - \eta(e))(e - \eta(e))']$ denote the mean vector and the covariance matrix, respectively.

We consider the conditional nonlinear McKean–Vlasov-type diffusion process

$$
\begin{aligned}
(10) \quad d\overline{X}_t ={}& (A\overline{X}_t + a)\, dt + R_1^{1/2} d\overline{W}_t \\
&+ \mathcal{P}_{\eta_t} C' R_2^{-1}\big[dY_t - \big((C\overline{X}_t + c)\, dt + R_2^{1/2}\, d\overline{V}_t\big)\big],
\end{aligned}
$$

where $(\overline{W}_t, \overline{V}_t, \overline{X}_0)$ are independent copies of (W_t, V_t, X_0) (thus independent of the signal and the observation path). In the above displayed formula $\eta_t := \mathrm{Law}(\overline{X}_t | \mathcal{F}_t)$ stands for the conditional distribution of the random state $\overline{X}_t$ given the σ-field $\mathcal{F}_t$. We shall call this probabilistic model the Kalman–Bucy (nonlinear) diffusion process.

In contrast to conventional nonlinear diffusions, the interaction does not take place only on the drift part but also on the diffusion matrix functional. In addition, the nonlinearity does not depend on the distribution of the random states $\pi_t = \mathrm{Law}(\overline{X}_t)$ but on their conditional distributions $\eta_t := \mathrm{Law}(\overline{X}_t | \mathcal{F}_t)$.

Section 5.1 discusses in some details the mathematical foundations of this conditional nonlinear diffusion process.

We will also check that the conditional expectations of the random states $\overline{X}_t$ and their conditional covariance matrices $\mathcal{P}_{\eta_t}$ w.r.t. $\mathcal{F}_t$ satisfy the Kalman–Bucy filter and the Riccati equations (8) and (9), *even when the initial variable is not Gaussian*; that is, we have that

$$
(11) \qquad\qquad m_t := m(\eta_t) = \widehat{X}_t \quad \text{and} \quad P_t = \mathcal{P}_{\eta_t}.
$$

In other words, the flow of matrices $\mathcal{P}_{\eta_t}$ only depends on the covariance matrix of the initial state $\overline{X}_0$. This property comes from the structure of the nonlinear process equation which ensures that the mean and the covariance matrices satisfy the Kalman–Bucy filter and the Riccati equation. This property simplifies the stability analysis of this process.

Given $\mathcal{P}_{\eta_0}$, the Kalman–Bucy diffusion (10) can be interpreted as a nonhomogeneous Ornstein–Uhlenbeck-type diffusion with a conditional covariance matrix $P_t = \mathcal{P}_{\eta_t}$ that satisfies the Riccati equation (9) starting from $P_0 = \mathcal{P}_{\eta_0}$. In this interpretation, the nonlinearity of the process is encapsulated in the Riccati equation (9).

A more detailed description of the nonlinear semigroup of (10) is provided in Section 5 dedicated to the stability properties of the Kalman–Bucy diffusions (see, for instance, Lemma 5.2). If in addition $\overline{X}_0 \overset{\mathrm{law}}{=} X_0$ (which is Gaussian), then given $\mathcal{F}_t$ the random states $\overline{X}_t$ of the nonlinear diffusion (10) are r_1-valued Gaussian random variables with mean $\widehat{X}_t$ and covariance matrix P_t. Notice that deterministic initial states $\overline{X}_0$ can also be seen as Gaussian with a null covariance matrix.

By (11), the stability properties of the Kalman–Bucy filter resume to the stability of the conditional expectations of the Kalman–Bucy diffusion, the reverse is clearly not true. These questions are developed in some details in Section 5.2 and Section 5.3.

2.3. *The Ensemble Kalman–Bucy filter.* The ensemble Kalman–Bucy filter co-incides with the mean-field particle interpretation of the nonlinear diffusion process (10). To be more precise, we let $(\overline{W}_t^i, \overline{V}_t^i, \xi_0^i)_{1 \leq i \leq N}$ be N independent copies of $(\overline{W}_t, \overline{V}_t, \overline{X}_0)$. In this notation, the EnKF is given by the Mckean–Vlasov-type interacting diffusion process

$$(12) \quad \begin{cases} d\xi_t^i = (A\xi_t^i + a)\,dt + R_1^{1/2}\,d\overline{W}_t^i \\ \qquad + P_t^N C' R_2^{-1}[dY_t - ((C\xi_t^i + c)\,dt + R_2^{1/2}\,d\overline{V}_t^i)] \\ i = 1, \dots, N \end{cases}$$

with the rescaled particle covariance matrices

$$(13) \qquad P_t^N := \left(1 - \frac{1}{N}\right)^{-1} \mathcal{P}_{\eta_t^N} = \frac{1}{N-1} \sum_{1 \leq i \leq N} (\xi_t^i - m_t^N)(\xi_t^i - m_t^N)',$$

the empirical measures

$$\eta_t^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{\xi_t^i} \quad \text{and the sample mean } m_t^N := m(\eta_t^N) = \frac{1}{N} \sum_{1 \leq i \leq N} \xi_t^i.$$

Note that (12) is a set of N stochastic differential equations coupled through the empirical covariance matrix P_t^N.

We also consider the N-particle model $\zeta_t = (\zeta_t^i)_{1 \leq i \leq N}$ defined as $\xi_t = (\xi_t^i)_{1 \leq i \leq N}$ by replacing the sample covariance matrix P_t^N by the true covariance matrix P_t (in particular we have $\xi_0 = \zeta_0$).

We end this section with some comments on these particle/ensemble filtering processes.

When $C = 0$, the EnKF reduce to N independent copies of the Ornstein–Uhlenbeck diffusive signal. In the same vein, for a single particle the covariance matrix is null so that the EnKF reduce to a single independent copy of the signal. In the case $r_1 = 1$, we have

$$(14) \quad \begin{aligned} \mathbb{E}(\|m_t^N - X_t\|^2) &= 2\,\mathrm{Var}(X_t) \quad \text{and} \\ \mathbb{E}(\|m_t^N - \widehat{X}_t\|^2) &= \mathbb{E}(\|m_t^N - X_t\|^2) + P_t. \end{aligned}$$

In these rather elementary situations, *the stability property of the signal drift matrix A is crucial to design some useful uniform estimates w.r.t. the time parameter.* The stability of the signal is a necessary condition to derive uniform estimates for any type of particle filters [20, 21] w.r.t. the time parameter.

On the other hand by the rank nullity theorem, when $N < r_1$ the sample covariance matrix P_t^N is the sample mean of N matrices of unit rank so that it has null eigenvalues. As a result, in some principal directions the EnKF is only driven by the signal diffusion. For unstable drift matrices, the EnKF experiences divergence as it is not corrected by the innovation process.

It should be clear from the above discussion that the stability of the signal is a necessary and sufficient condition to design useful uniform estimates w.r.t. the time horizon.

We shall return to this question in Section 6.1.

We also recall that evolution of the conditional distributions of a nonlinear signal given the observations satisfy a complex nonlinear and stochastic measure valued equation. In probability and statistics literature, this equation is called the Kushner–Stratonovitch nonlinear filtering equation [39, 61]. The unnormalized version of the conditional probability distribution follows a linear measure-valued equation, called the Zakai equation [71].

For linear-Gaussian models, the conditional distributions of the states are Gaussian distributions. The evolution of the conditional expectations and covariance error matrices resume to the Kalman–Bucy filter [14]. When the signal process (7) is nonlinear or when its perturbations are non-Gaussian, it is tempting to design a mean-field approximation of a nonlinear diffusion defined as in (10) by replacing the linear drift $x \mapsto Ax + a$ by a nonlinear one, say $x \mapsto A(x)$. Of course, under some appropriate regularity conditions the sample means converge to the first moment of the nonlinear process at hand.

Unfortunately, it is well known that this nonlinear diffusion process cannot capture any of the statistics of the optimal filter.

We easily check this assertion by showing that the nonlinear Fokker–Planck equation associated with the nonlinear process differs from the Kushner–Stratonovitch optimal filter equation.

Up to the best of our knowledge, there does not exist a single result that quantifies the error between the conditional distributions nor some statistics of the random states of the particle model with the ones of the optimal filter.

Besides these drawbacks and these open important questions, the EnKF is of current use in nonlinear settings. This type of stochastic model can be thought of as an extended EnKF approximation scheme.

The first step to understand these probabilistic models is to ensure the convergence of the mean-field approximation to the distribution of the nonlinear diffusion. We also refer the reader to the pioneering article [43] for a discussion on the convergence of these nonoptimal mean-field models in the discrete time case. The uniform convergence of these nonlinear EnKF is still an open important questions for continuous as well as for discrete time filtering problems. We plan to investigate these questions in a forthcoming article.

3. Statement of the main results.

3.1. *A stochastic perturbation theorem.* Our first main result shows that the stochastic processes (m_t^N, P_t^N) satisfy the same equation as (m_t, P_t) up to some local fluctuation orthogonal martingales with angle brackets that only depends on the sample covariance matrix P_t^N.

THEOREM 3.1 (Perturbation theorem). *The stochastic processes (m_t^N, P_t^N) defined in* (13) *satisfy the diffusion equations*

$$(15) \quad dm_t^N = \left(A m_t^N + a\right) dt + P_t^N C' R_2^{-1}\left(dY_t - \left(C m_t^N + c\right) dt\right) + \frac{1}{\sqrt{N}} d\overline{M}_t$$

with the vector-valued martingale $\overline{M}_t = (\overline{M}_t(k))_{1 \le k \le r_1}$ and the angle-brackets

$$(16) \qquad\qquad \frac{d}{dt} \langle \overline{M} \rangle_t = R + P_t^N S P_t^N.$$

We also have the matrix-valued diffusion

$$(17) \qquad d P_t^N = \left(A P_t^N + P_t^N A' - P_t^N S P_t^N + R\right) dt + \frac{1}{\sqrt{N-1}} dM_t$$

with a symmetric matrix-valued martingale $M_t = (M_t(k,l))_{1 \le k, l \le r_1}$ and the angle brackets

$$(18) \qquad\qquad \frac{d}{dt} \langle M \rangle_t^{\sharp} = 4 P_t^N \otimes_s \left(R + P_t^N S P_t^N\right).$$

In addition, we have the orthogonality property

$$\forall 1 \le k, l, l' \le r_1 \quad \langle M(k,l), \overline{M}(l') \rangle_t = 0.$$

This fluctuation-type theorem shows that the EnKF and the interacting sample covariance matrices satisfy a Kalman–Bucy recursion and a stochastic-type Riccati equation. The extra level of randomness comes from the fact the particles mimic the random evolution of the nonlinear McKean–Vlasov-type diffusion (10). The amplitude of these fluctuation-martingales is of order $1/\sqrt{N}$, as any mean field-type particle process.

In contrast with the optimal Kalman–Bucy filter, the fluctuations of the covariance matrices may corrupt the natural stabilizing effects of the innovation process defined by the centered observations increments $(dY_t - (C m_t^N + c) dt)$.

For partially observed signals, we cannot expect any stability properties of Kalman–Bucy filter and the EnKF without introducing some structural conditions of observability and controllability on the signal-observation equation (7). When $a = 0 = c$, observe that the Kalman–Bucy equation (8) implies that

$$(19) \qquad d(\widehat{X}_t - X_t) = (A - P_t S)(\widehat{X}_t - X_t) dt + P_t C' R_2^{-1/2} dV_t + R_1^{1/2} dW_t.$$

The EnKF evolution equation (15) stated in Theorem 3.1 shows that the evolution of the error-vector $(m_t^N - X_t)$ has the same form as above by replacing P_t by P_t^N, up to some fluctuation martingale due to the interacting sample covariance matrices P_t^N and the internal perturbations of the particles.

This equation shows that the stability properties of these processes depends on the nature of the eigenvalues of the matrices $A - P_t S$ and $A - P_t^N S$.

3.2. *Stability of Kalman–Bucy nonlinear diffusions.* We further assume that $(A, R_1^{1/2})$ is a controllable pair and (A, C) is observable, that is the matrices

(20) $$\left[R_1^{1/2}, A(R_1^{1/2}), \ldots, A^{r_1-1}R_1^{1/2}\right] \quad \text{and} \quad \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{r_1-1} \end{bmatrix}$$

have rank r_1. By Bucy's theorem (cf. Theorem 5.6 on page 816), under these conditions P_t converge exponentially fast to P as $t \uparrow \infty$. This unique fixed point is called the steady state error covariance matrix and it satisfies the algebraic Riccati equation:

(21) $$\mathrm{Ricc}(P) = AP + PA' - PSP + R = 0.$$

In addition, the matrix difference $A - PS$ is asymptotically stable even when the signal drift matrix A is unstable. Under the above condition, there exists some parameters $\upsilon, \varpi_\pm > 0$ such that

(22)
$$\varpi_- \, \mathrm{Id} \leq \int_0^\upsilon e^{As} R e^{A's} \, ds \leq \varpi_+ \, \mathrm{Id} \quad \text{and}$$
$$\varpi_- \, \mathrm{Id} \leq \int_0^\upsilon e^{-A's} S e^{-As} \, ds \leq \varpi_+ \, \mathrm{Id}.$$

The parameter υ is often called the interval of observability-controllability.

For a more detailed discussion on these fixed-point Riccati equations (a.k.a. algebraic Riccati equation) and the connexions with optimal control theory, we refer the reader to [1, 15, 55, 68, 72] and the references therein.

In reference to signal processing and control theory literature, we adopt the following terminology.

DEFINITION 3.2. The Kalman–Bucy filter $\widehat{X}_t$ associated with the initial covariance matrix $P_0 = P$ is called the steady state Kalman–Bucy filter. The Kalman–Bucy diffusion $\overline{X}_t$ starting from an initial random state $\overline{X}_0$ with covariance matrix $\overline{P}_0 = P$ is called the steady state Kalman–Bucy diffusion.

It is important to observe that this stationary type filter does not require to solve the Riccati equation (9). For a more thorough discussion on the stability properties of Kalman–Bucy filters and Riccati equations, we refer the reader to [1, 3, 41, 55, 60, 68, 72]. We also refer the reader to Section 5.3 dedicated to contraction estimates of Riccati semigroups and related fundamental matrices.

Our main objective is to extend these stability properties at the level of the Kalman–Bucy nonlinear diffusion.

The conditional distributions η_t of the random states $\overline{X}_t$ given $\mathcal{F}_t$ is a stochastic measure valued process. In contrast to more conventional nonlinear Markov diffusions, the nonlinearity depends on the $\mathcal{F}_t$-conditional covariance matrices. Thus

the computation of the distribution $\pi_t = \text{Law}(\overline{X}_t)$ of the random states $\overline{X}_t$ requires to compute the conditional covariance matrices $\mathcal{P}_{\eta_t}$.

To study the stability properties of the flow π_t, it is natural to introduce a copy $\overline{X}_t^\star$ of $\overline{X}_t$ coupled to $\overline{X}_t$ by choosing the same observation process, and by only changing the random perturbations $(\overline{W}_t, \overline{V}_t)$. In other words, $(\overline{X}_t, \overline{X}_t^\star)$ are $\mathcal{F}_t$-conditionally independent.

To describe our main results with some precision, we introduce some terminology.

DEFINITION 3.3. Let $(P_t, P_t^\star)$ be a couple of solutions of the Riccati equation (9) starting at two possibly different values $(P_0, P_0^\star)$. We also denote by $(\overline{X}_t, \overline{X}_t^\star)$ a couple of Kalman–Bucy diffusions (10) starting from two random states with covariances matrices $(P_0, P_0^\star)$. We denote by $(\pi_t, \pi_t^\star)$ and $(\eta_t, \eta_t^\star)$ the distributions and the $\mathcal{F}_t$-conditional distributions of $(\overline{X}_t, \overline{X}_t^\star)$.

Let $(\widehat{X}_t, \widehat{X}_t^\star)$ denote the solutions of the Kalman–Bucy equation (8) using the covariance matrices $(P_t, P_t^\star)$ [and starting from a random initial condition with covariance matrices $(P_0, P_0^\star)$]. Equivalently (cf. Lemma 5.2), we have

$$(23) \qquad (\widehat{X}_t, P_t) = (m(\eta_t), P(\eta_t)) \quad \text{and} \quad (\widehat{X}_t^\star, P_t^\star) = (m(\eta_t^\star), P(\eta_t^\star)).$$

Since $Y_0 = 0$, the signal state is not observed at the origin. This clearly implies that

$$(\pi_0, \pi_0^\star) = (\eta_0, \eta_0^\star) \quad \text{and} \quad (\mathcal{P}_{\pi_0}, P_{\pi_0^\star}) = (\mathcal{P}_{\eta_0}, \mathcal{P}_{\eta_0^\star}) = (P_0, \widehat{P}_0).$$

Using (23), it should be clear that most of the stability properties of the Kalman–Bucy diffusions (expressed in terms of some convex criteria) can be used to deduce the ones of their conditional averages but the reverse is clearly not true. For instance, using (23), for any $n \geq 1$ and $t \geq 0$ we readily check that

$$\mathbb{W}_n\big(\text{Law}(\widehat{X}_t), \text{Law}(\widehat{X}_t^\star)\big) \leq \mathbb{W}_n\big(\eta_t, \eta_t^\star\big).$$

In this context, one of our main results can basically be stated as follows.

THEOREM 3.4. *Assume the existence of a positive semidefinite fixed point P of (21) s.t. $\mu(A - PS) < 0$. For any $n \geq 2$, $\varepsilon \in \,]0, 1/2]$ and $t \geq 0$, we have*

$$(24) \quad \mathbb{W}_n\big(\pi_t, \pi_t^\star\big) \leq c(\varepsilon) \exp\big[(1 - \varepsilon)\mu(A - PS)t\big]\big(\mathbb{W}_n\big(\pi_0, \pi_0^\star\big) + \|P_0 - P_0^\star\|_F\big)$$

for some constant $c(\varepsilon) < \infty$ whose values depend on the parameter ε and on the initial matrices $(P_0, P_0^\star)$.

The proof of this theorem is provided in Section 10.

When $(\overline{X}_0, \overline{X}_0^\star)$ are Gaussian variables, the probability distributions $(\eta_t, \eta_t^\star)$ are also Gaussian and, therefore, it is no surprise that any distance between these Gaussian probability distributions could be controlled by some distance between their respective mean vectors and covariance matrices; see, for instance, the article [28].

THEOREM 3.5. *Assume that $(\overline{X}_0, \overline{X}_0^\star)$ are two Gaussian random variables with covariance matrices $(P_0, P_0^\star)$. Also assume that the algebraic Riccati equation* (21) *has a positive definite fixed-point P and $\mu(A - PS) < 0$. In this situation, there exists some t_0 and some finite constant c that depends on (P_0, P) such that for any $t \geq t_0$ and any $\varepsilon \in {]0, 2]}$ we have the quenched almost estimate*

$$(25) \quad \mathrm{Ent}(\eta_t | \eta_t^\star) \leq c(\varepsilon) \big[\exp\big((1 - \varepsilon)\mu(A - PS)t\big) \| P_0 - P_0^\star \|_F + \| \widehat{X}_t^\star - \widehat{X}_t \|^2 \big].$$

In addition, for any $t \geq t_0$ and any $\varepsilon \in {]0, 1/2]}$ we have the annealed estimate

$$(26) \quad \mathrm{Ent}(\pi_t | \pi_t^\star) \leq c(\varepsilon) \exp\big((1 - \varepsilon)\mu(A - PS)t\big)\big(\mathbb{W}_2^2(\pi_0, \pi_0^\star) + \| P_0 - P_0^\star \|_F\big).$$

In the above display, $c(\varepsilon)$ stands for some finite constant whose values depend on the parameter ε and on the initial matrices $(P_0, P_0^\star)$

We easily check (26) using (24), (25) and the convexity property of the relative entropy

$$\mathrm{Ent}(\nu K_1 | \nu K_2) \leq \int \nu(dy) \, \mathrm{Ent}\big(K_1(y, \cdot) | K_2(y, \cdot)\big)$$

which is valid for any Markov transitions $K_i(y, dx)$ from $\mathbb{R}^{r_2}$ into $\mathbb{R}^{r_1}$ and any probability distribution ν on $\mathbb{R}^{r_2}$. In the above display, (νK_i) stands for the probability measures on $\mathbb{R}^{r_1}$ defined by the transport formula

$$(\nu K_i)(dx) := \int \nu(dy) K_i(y, dx).$$

The proof of the almost sure estimate (25) is provided in Section 11. For a more thorough discussion on the stability properties of Kalman–Bucy diffusions under weaker observability and controllability conditions, we refer the reader to the recent review article [9].

3.3. *An uniform propagation of chaos theorem.* The EnKF avoid the numerical solving of the Riccati equation (12) or the use of the steady state P by using interacting sample covariance matrices. Even when $(\xi_0^i)_{1 \leq i \leq N}$ are independent copies of a Gaussian random variable $\overline{X}_0$ with the steady covariance matrix, the initial sample covariance matrices P_t^N defined in (13) fluctuates around the limiting value P.

These random fluctuations of the sample covariance matrices P_t^N eventually corrupt the stability in the EnKF, even if the filtering problem is observable and controllable in the conventional sense. For instance, in practical situations the empirical covariance matrices may not be invertible for small sample sizes. This simple observation shows that the Lyapunov theory based on inverse of covariance matrices developed in [3, 6, 45] cannot be applied in the context of EnKF.

In practice, it has also been observed that fluctuations of the sample-covariances of the EnKF eventually mislead the natural stabilizing effect of the observation

process in the Kalman–Bucy filter evolution. These fluctuations induce an underestimation of the true error covariances. As a result, the EnKF ignores the important information delivered by the sensors. This lack of observation also leads to the divergence of the filter.

Last but not least, from the numerical viewpoint the Kalman–Bucy filter and the EnKF are also know to be nonrobust, in the sense that arithmetic errors may accumulate even if the exact filter is stable.

All of these instability properties of the EnKF are well known. They are often referred as the catastrophic filter divergence in data assimilation literature. For a more thorough discussion on these issues, we refer the reader to the articles [29, 31, 36], and the references therein. As mentioned by the authors in [36], "catastrophic filter divergence is a well documented but mechanistically mysterious phenomenon whereby ensemble-state estimates explode to machine infinity despite the true state remaining in a bounded region." The second main result of this article is to provide uniform propagations of chaos properties w.r.t. the time horizon. To this end, we assume that the following observability condition is satisfied:

$$(27) \qquad (S) \quad S = \rho(S)\,\mathrm{Id} \quad \text{for some } \rho(S) > 0.$$

Under this condition, we have

$$(28) \quad \mu(A) - \rho(S)\lambda_{\max}(P) \leq \mu(A - PS) \leq \mu(A) - \rho(S)\lambda_{\min}(P) \leq \mu(A).$$

This shows that $\mu(A - PS) < 0$ as soon as $\mu(A) < 0$. When $\mu(A) > 0$, it is also met as soon as $\mu(A) \leq \rho(S)\lambda_{\min}(P)$.

Several examples of sensor models satisfying this condition are discussed in Section 4.2.

THEOREM 3.6. *Assume that the observability condition* (S) *is met and* $\mu(A) < 0$. *In this situation, for any $n \geq 1$ and any sufficiently large N we have the uniform estimates*

$$(29) \qquad \sup_{t \geq 0} \mathbb{E}\big[\|P_t^N - P_t\|_F^n\big]^{\frac{1}{n}} \leq c(n)/\sqrt{N} \quad and$$

$$\sup_{t \geq 0} \mathbb{E}\big[\|\xi_t^1 - \zeta_t^1\|^n\big]^{\frac{1}{n}} \leq c'(n)/\sqrt{N}.$$

When condition (S) is not necessarily met, but the signal-drift matrix A is stable we have the uniform estimate

$$\mu(A) < 0 \quad \Longrightarrow \quad \forall 1 \leq n < 1 + \frac{N-1}{2r_1}\frac{\lambda_{\min}(S)}{\lambda_{\max}(S)} \qquad \sup_{t \geq 0} \mathbb{E}\big([\mathrm{tr}(P_t^N)]^n\big) \leq c(n)$$

(cf. Proposition 7.2). The proof of the estimates (29) relies on the observability condition (S). This condition is satisfied for any filtering problem with $r_1 = r_2$ with an invertible matrix C, up to a change of basis; see, for instance, (38).

We conjecture that condition (S) is purely technical and can be relaxed for stable signal-drift matrices. For unstable drift matrices, the form of the matrix S may also corrupt the regularity of the sample covariance matrices. We already know that for $r_1 > N$ the matrices P_t^N have necessarily at least one null eigenvalue, so that the number of null eigenvalues of $P_t^N S$ is more likely to be larger when S is not of full rank.

The following corollary is a consequence of (29).

COROLLARY 3.7. *Under the assumptions stated in Theorem* 3.6, *for any* $n \geq 1$ *we have the uniform estimates*

$$\sup_{t \geq 0} \mathbb{W}_n\big(\mathrm{Law}(\xi_t^1), \eta_t\big) \leq c(n)/\sqrt{N} \quad and$$

(30)

$$\sup_{t \geq 0} \mathbb{E}\big(\big|\eta_t^N(f) - \eta_t(f)\big|^n\big)^{\frac{1}{n}} \leq c'(n)/\sqrt{N}$$

for any 1-*Lipschitz function* f *on* $\mathbb{R}^{r_1}$. *In particular, this implies that*

(31)
$$\sup_{t \geq 0} \mathbb{E}\big[\big\|m_t^N - \widehat{X}_t\big\|^n\big]^{\frac{1}{n}} \leq c(n)/\sqrt{N}.$$

Inversely, the uniform estimates (31) *implies that* $\mu(A) < 0$ *as soon as* $R \neq 0 = C$.

The detailed proof of Theorem 3.1 is postponed to Section 7.1. The proof of Theorem 3.6 is based on uniform moments estimates developed in Section 7.2. The detailed proof of the uniform estimates (29) is presented in full details in Section 8.

It is important to observe that all the $\mathbb{L}_n$-mean error estimates between the sample covariance p_0 and P_0, as well as the ones between the sample mean m_0 and $\widehat{X}_0$, are immediate for $t = 0$ (as a direct application of the law of large numbers for independent random sequences).

A discussion on the stability condition $\mu(A) < 0$ is provided in Section 6. In Section 6.1, we will show that this condition cannot be relaxed to derive uniform estimates (31) as soon as $R \neq 0$ and $C = 0$. The strong observability condition (S) is discussed in some details in Section 4. Section 4.1 is dedicated to the stability of stochastic observer processes defined as the Kalman–Bucy diffusion (8) by replacing $\mathcal{P}_{\eta_t}$ by a flow of stochastic covariance matrices. Several illustrations are presented in Section 4.2.

We also emphasize that the conditions (S) and $\mu(A) < 0$ are only used to derive uniform estimates w.r.t. the time horizon. Without these conditions, the statements of Theorem 3.6 and Corollary 3.7 remain valid without the supremum operations w.r.t. the time parameter. In this general situation, the constants $c(n)$ and $c'(n)$ are replaced by some constants $c_t(n)$ and $c_t'(n)$ whose values depend on the time horizon.

4. Some comments on our regularity conditions. The stability analysis of diffusion processes is always much more documented than the ones on their possible divergence. For instance, in contrast with conventional Kalman–Bucy filters, the stability properties of the EnKF are not induced by some kind of observability or controllability condition. The only known results for discrete generation EnKF is the recent work by X. T. Tong, A. J. Majda and D. Kelly [64]. One of the main assumptions of the article is that the sensor-matrix is of full rank. The authors also provide a concrete numerical example of filtering problem with sparse observations for which the EnKF experiences a catastrophic divergence.

The stability of linear dynamical systems with time varying drift-matrices A_t is much complex than the one of time homogeneous models. In our context, the drift-matrix are also random since they encapsulate the fluctuations of the stochastic covariance matrices. As shown in [33], the fact that the real part of the spectrum of A_t is negative is neither necessary nor sufficient for the exponential stability of the system. It may even happen that the semigroup of the system is unstable even if the real part of the eigenvalues of A_t remain negative for all times. Inversely, the system may be stable even when one of the eigenvalues is positive for all times.

The analysis of the exponential stability of time varying linear system requires to estimate the variations of the matrices A_t w.r.t. the time parameter. All types of sufficient conditions stated in [13, 18, 33, 37, 38, 56] are restricted to deterministic systems, and thus cannot be applied to interacting diffusion processes. They are also based on continuity/tightness-type properties, as well as on piecewise smoothness properties or on the uniform boundedness of the velocity field $\frac{d}{dt} A_t$. These regularity properties does not hold for stochastic diffusions. In our context, the fluctuation matrices are given by

$$(32) \qquad Q_t := \sqrt{N}(P_t^N - P_t) \in \mathbb{S}_{r_1} \quad \Longleftrightarrow \quad P_t^N = P_t + \frac{1}{\sqrt{N}} Q_t \in \mathbb{S}_{r_1}^+.$$

Our next objective is to initiate a more refined mathematical analysis to understand these divergence properties in terms of global divergence properties, locally ill-conditioned filtering models and stochastic type observers.

4.1. *Stable and divergence regions.* To clarify the presentation, we further assume that the Kalman–Bucy filter and the Riccati equation start at the steady state $P_0 = P$.

Let Z_t be some observer type process defined as (8) by replacing $\mathcal{P}_{\eta_t}$ by some covariance matrix of the form $(P + Q_t) \in \mathbb{S}_{r_1}^+$, for some flow of symmetric matrices $Q_t \in \mathbb{S}_{r_1}$ mimicking the fluctuation matrices (32). In other words, the matrices Q_t reflects the fluctuations of the empirical covariance matrices P_t^N around the steady state P.

In this case, we have

$$(33) \qquad dZ_t = \big[AZ_t - (P + Q_t)S(Z_t - X_t)\big] dt + (P + Q_t)C'R_2^{-1/2} dV_t.$$

When $Q_t = 0$, the null matrix, the process Z_t resumes to the steady state Kalman–Bucy filter. We set

$$A - (P + Q_t)S = \overline{A} - Q_t S \quad \text{with } \overline{A} := A - PS$$

and

$$E_t := \exp\left[\oint_0^t (\overline{A} - Q_s S)\, ds\right].$$

The vector error process $\mathcal{Z}_t := Z_t - X_t$ is given by the stochastic Ornstein–Ulhenbeck process:

$$d\mathcal{Z}_t = (\overline{A} - Q_t S)\mathcal{Z}_t\, dt + \sqrt{2}\, d\mathcal{W}_t \quad \Rightarrow \quad \mathcal{Z}_t = E_t \mathcal{Z}_0 + \sqrt{2}\int_0^t E_t E_s^{-1}\, d\mathcal{W}_s.$$

In the above display, $\mathcal{W}_t$ stands for diffusion

$$\mathcal{W}_t := (\overline{W}_t - W_t)/\sqrt{2} + (P + Q_t)C'R_2^{-1/2}(V_t - \overline{V}_t)/\sqrt{2}$$

$$\stackrel{\text{law}}{=} W_t + (P + Q_t)C'R_2^{-1/2}V_t$$

with covariance matrices $\mathcal{R}_t = I + (P + Q_t)S(P + Q_t)$. When the flow of matrices Q_t enter into the set

$$(34) \qquad \mathcal{Q}_{\text{div}} = \{Q \in \mathbb{S}_{r_1} : \varsigma(\overline{A} - QS) > 0\}$$

the observer experiences a divergence in at least one of the principal directions. Notice that

$$(35) \qquad \|E_t\|_2 \leq \exp\left[\int_0^t \mu(\overline{A} - Q_s S)\, ds\right].$$

This semigroup estimate allows to quantify the stability of the process $\mathcal{Z}_t$ as soon as $\mu(\overline{A} - Q_s S) < -\delta$ for some $\delta > 0$ for sufficiently large time horizons.

One natural strategy is to analyze the contraction properties of the stochastic flow E_t generated by the stochastic matrices $\overline{A} - Q_t S$ and their logarithmic norms $\mu(\overline{A} - Q_t S)$. More precisely, under the strong observability condition (S) stated in (27) we have

$$(36) \qquad \mu(\overline{A} - Q_s S) \leq \mu(A) + \rho(S)\mu\big(-(P + Q_s)\big) \leq \mu(A) < 0$$

as soon as $\mu(A) < 0$, for *any* possible symmetric fluctuations Q_s s.t. $P + Q_s \geq 0$.

This shows that for stable signal-drift matrices $\mu(A) < 0$ the condition (S) ensures that the stochastic observer is both theoretically and numerically stable for any type of fluctuations Q_s. The same reasoning will be used to show that the stability of the signal is transferred to the EnKF filter.

Without condition (S) is easy to work out several examples of 2-dimensional filtering problems with a stable-drift matrix $\mu(A) < 0$ and such that $\mu(\overline{A} - Q_s S) > 0 > \lambda_{\min}((\overline{A} - Q_s S)_{\text{sym}})$ for some flow of symmetric matrices Q_s s.t. $P + Q_s \geq 0$.

In this context, even if the EnKF is numerically stable it is difficult to analyze theoretically this class of locally ill conditioned models using spectral and semigroup techniques.

In the reverse angle, in practical situations the EnKF generally experiences severe divergence when $\mu(A) > 0 > \mu(\overline{A})$. In this situation, we already know from (14) that we cannot expect to have uniform propagation of chaos estimates for any fluctuation matrices. Also observe that

$$((S) \text{ and } \mu(\overline{A}) < 0) \implies \mu(\overline{A} - Q_s S) < \mu(\overline{A}) + \rho(S)\mu(-Q_s) < \mu(\overline{A}) < 0$$

for any *positive* semidefinite fluctuations $Q_s \geq 0$ around the steady state P. Unfortunately, we cannot ensure that the fluctuations of the sample covariance matrices are always positive.

As mentioned above, the pivotal semigroup estimate (35) requires to estimate the logarithmic norm of the stochastic flow of matrices $\overline{A} - Q_t S$. Several technical difficulties arise.

The first one comes from the fact that $\varsigma(\overline{A}) < 0$ does not implies that $\mu(\overline{A}) < 0$, since $\overline{A} \neq \overline{A}'$. When $\mu(\overline{A}) > 0$, the steady state Kalman–Bucy is locally ill-conditioned, in the sense that the worst fluctuation around the true signal behave like $\exp[\mu(\overline{A})\Delta t]$ is a short transient time Δt. This local divergence property may occur even when all the eigenvalues of $\overline{A}$ or even the ones of A are negative. This indicates that it is hopeless to analyze the stability of the Kalman–Bucy filters estimates based on the semigroup inequality (35) for such ill-conditioned systems. We are faced to the same issues if we try to quantify the propagations of the fluctuation Q_t in the system. Some illustrations are discussed in the Appendix, on page 842.

This discussion indicates that the stability property $\mu(A) < 0$ and the observability condition (27) seem to be essential to control the fluctuations of the EnKF sample covariance matrices *for any number of samples*. These conditions also ensure the semigroup contraction properties needed to derive uniform $\mathbb{L}_n$-mean error estimates of the EnKF particle filter.

4.2. *Full observation sensors.* When the observation variables are the same as the ones of the signal, the signal observation has the same dimension as the signal and resumes to some equation of the form

$$(37) \qquad dY_t = bX_t\,dt + \sigma_2\,dV_t$$

for some parameters $b \in \mathbb{R}$ and $\sigma_2 > 0$. These sensors are used in data grid-type assimilation problems when measurements can be evaluated at each cell. These fully observed models are discussed in Section 4 in [30] in the context of the Lorentz-96 filtering problems. These observation processes are also used to the article [8] for application to nonlinear and multiscale filtering problem. In this context, the observed variables represents the slow components of the signal. When the fast components are represented by a some Brownian motion with a prescribed

covariance matrix, the filtering of the slow components with full observations take the form (37).

The sensor model discussed in (37) clearly satisfies condition (27) with the parameter $\rho(S) = (b/\sigma_2)^2$. This rather strong condition (27) ensures that these fluctuations does not propagate w.r.t. the time parameter, regardless of the initial data. Under this condition, we shall prove that the particle EnKF has uniformly bounded $\mathbb{L}_n$-moments for any $n \geq 1$ (see for Proposition 7.2).

We emphasize that the observability condition (27) is satisfied when the filtering problem are similar to a fully observed sensor model; that is, up to a change of basis functions. More precisely, any filtering problem (7) with $r_1 = r_2$ and s.t. $\mathcal{C} := (R_2^{-1/2}C)$ is invertible can be turned into a filtering problem equipped with an identity sensor matrix; even when the original matrix $S = C'R_2^{-1}C = \mathcal{C}'\mathcal{C}$ does not satisfies (27). To check this claim, we observe that

$$(38) \quad \mathcal{Y}_t := R_2^{-1/2}Y_t \quad \text{and} \quad \mathcal{X}_t := \mathcal{C}X_t \Longrightarrow \begin{cases} d\mathcal{X}_t = \mathcal{A}\mathcal{X}_t \, dt + \mathcal{R}_1^{1/2} \, dW_t, \\ d\mathcal{Y}_t = \mathcal{X}_t \, dt + dV_t \end{cases}$$

with the signal drift matrix $\mathcal{A} := \mathcal{C}A\mathcal{C}^{-1}$ and the diffusion covariance matrix $\mathcal{R}_1 := \mathcal{C}R_1\mathcal{C}'$. In this situation, the filtering model $(\mathcal{X}_t, \mathcal{Y}_t)$ satisfies (27) with $\mathcal{S} = \mathrm{Id} \Rightarrow \rho(\mathcal{S}) = 1$. The link between the logarithmic norm of $\mathcal{A}$ and the original signal drift matrix A is given by the formula

$$\mu(\mathcal{A}) = \frac{1}{2}\lambda_{\max}(\mathcal{C}A\mathcal{C}^{-1} + (\mathcal{C}A\mathcal{C}^{-1})').$$

For orthogonal matrices $\mathcal{C}$, we have $\mu(\mathcal{A}) = \mu(\mathcal{C}A\mathcal{C}') = \mu(A)$. Otherwise, the condition $\mu(\mathcal{A}) < 0$ depends on the triplet of matrices (A, C, R_2) associated with the original filtering problem. For instance, when $r_1 = r_2 = 2$, $R_2 = \mathrm{Id}$ and a symmetric negative definite drift matrix A, the condition $\mu(A) < 0$ is equivalent to the fact that

$$(39) \qquad\qquad A_{1,1} < 0 \quad \text{and} \quad A_{1,1}A_{2,2} > A_{1,2}^2.$$

For sensor matrices of the form $C = \begin{pmatrix} 1 & 0 \\ 0 & \beta \end{pmatrix}$, for some $\beta \neq 0$, condition $\mu(\mathcal{A}) < 0$ takes the strongest form

$$A_{1,1} < 0 \quad \text{and} \quad A_{1,1}A_{2,2} > \left(\frac{\beta + \beta^{-1}}{2}\right)^2 A_{1,2}^2.$$

To better understand the importance of the matrix S introduced in (9) observe that

$$(40) \qquad\qquad d\widehat{X}_t = (A\widehat{X}_t + a) \, dt + P_t(d\overline{Y}_t - S\widehat{X}_t \, dt)$$

with the r_1-dimensional observation process $\overline{Y}_t$ given by

$$\overline{Y}_t := C'R_2^{-1}Y_t \quad \Longrightarrow \quad d\overline{Y}_t = SX_t \, dt + C'R_2^{-1/2} \, dV_t.$$

Let us assume that $r_1 = 1 < r_2 = 2$, and the (2×1)-sensor matrix is given by

$$C = [1, \alpha] \quad \text{for some } \alpha \in \mathbb{R}.$$

For unit diffusion covariance matrices, the partially observed filtering systems (X_t, Y_t) and $(X_t, \overline{Y}_t)$ observable as soon as

(41) $$A_{1,2} + \alpha A_{2,2} \neq \alpha(A_{1,1} + \alpha A_{2,1}).$$

Nevertheless, in this situation the diffusion matrix $S = \left[\begin{smallmatrix} 1 & \alpha \\ \alpha & \alpha^2 \end{smallmatrix}\right]$ of this new 2-dimensional $\overline{Y}_t$ is no more invertible. This shows that these 1-dimensional partial observations models cannot be turned into regular 2-dimensional sensors. These 2-dimensional filtering problems equipped with a 1-dimensional sensor are one of the simplest examples of controllable and observable filtering problems that does not satisfy the observability condition (27) even if the signal drift is stable.

5. Stability properties of Kalman–Bucy diffusions.

5.1. *Kalman–Bucy diffusions.* As noticed in the Introduction, the Kalman–Bucy diffusion (10) strongly differs from conventional nonlinear diffusion processes. The evolution of this new class of probabilistic models depends on the $\mathcal{F}_t$-conditional distribution of the random states.

This section provides a more detailed discussion on this new class of nonlinear McKean–Vlasov-type diffusions with $\mathcal{F}_t$-conditional distribution interactions.

DEFINITION 5.1. Let $\varphi_{s,t}$ be the dynamical semigroup of the Riccati equation (9) given for any $s \leq t$ by

$$\varphi_{s,t}(P_s) = P_t \quad \text{and set } \varphi_{0,t} = \varphi_t.$$

Assuming that the nonlinear diffusion (10) is well-posed, it is almost immediate to check that the conditional mean and covariance satisfy the Kalman–Bucy equations (8) and (9). The next lemma ensures that the Kalman–Bucy diffusion (10) is well-posed.

LEMMA 5.2. *Consider the nonhomogeneous diffusion given by*

(42)
$$\begin{aligned}
d\overline{X}_t &= (A\overline{X}_t + a)\, dt + R_1^{1/2}\, d\overline{W}_t \\
&\quad + \varphi_t(P_0)C'R_2^{-1}\big(dY_t - ((C\overline{X}_t + c)\, dt + R_2^{1/2}\, d\overline{V}_t)\big),
\end{aligned}$$

where $(\overline{W}_t, \overline{V}_t, \overline{X}_0)$ are independent copies of (W_t, V_t, X_0) and $P_0 = \mathcal{P}_{\eta_0}$. In this situation, we have

$$\eta_t := \operatorname{Law}(\overline{X}_t | \mathcal{F}_t) \implies \mathcal{P}_{\eta_t} = \varphi_t(P_0) \quad \text{and} \quad \mathbb{E}(\overline{X}_t | \mathcal{F}_t) = \widehat{X}_t,$$

where $\widehat{X}_t$ stands for the solution of the Kalman-filter (8) driven by the solution $P_t = \mathcal{P}_{\eta_t}$ of Riccati equation (9) starting at $P_0 = \mathcal{P}_{\eta_0}$.

PROOF.   Observe that (42) is a nonhomogeneous linear diffusion process. This implies that

$$(43) \qquad d\mathbb{E}(\overline{X}_t|\mathcal{F}_t) = \big(A\mathbb{E}(\overline{X}_t|\mathcal{F}_t) + a\big)\,dt$$
$$+ \varphi_t(P_0)C'R_2^{-1}\big(dY_t - (C_t\mathbb{E}(\overline{X}_t|\mathcal{F}_t) + c)\,dt\big).$$

We set $\widetilde{X}_t := \overline{X}_t - \mathbb{E}(\overline{X}_t|\mathcal{F}_t)$. In this notation, we have

$$d\widetilde{X}_t = \big[A - \varphi_t(P_0)C'R_2^{-1}C\big]\widetilde{X}_t\,dt + R_1^{1/2}\,d\overline{W}_t - \varphi_t(P_0)C'R_2^{-1/2}\,d\overline{V}_t.$$

This implies that

$$d(\widetilde{X}_t\widetilde{X}_t') = \big[A - \varphi_t(P_0)C'R_2^{-1}C\big]\widetilde{X}_t\widetilde{X}_t'\,dt$$
$$+ \big[R_1^{1/2}\,d\overline{W}_t - \varphi_t(P_0)C'R_2^{-1/2}\,d\overline{V}_t\big]\widetilde{X}_t'$$
$$+ \widetilde{X}_t\widetilde{X}_t'\big[A - \varphi_t(P_0)C'R_2^{-1}C\big]'\,dt$$
$$+ \widetilde{X}_t\big[R_1^{1/2}\,d\overline{W}_t - \varphi_t(P_0)C'R_2^{-1/2}\,d\overline{V}_t\big]'$$
$$+ \big[R_1 + \varphi_t(P_0)C'R_2^{-1}C\varphi_t(P_0)\big]\,dt.$$

This shows that the covariance matrix

$$Q_t := \mathcal{P}_{\eta_t} := \mathbb{E}\big(\widetilde{X}_t\widetilde{X}_t'|\mathcal{F}_t\big) = \mathbb{E}\big(\widetilde{X}_t\widetilde{X}_t'\big)$$

does not depend on the observation process. In addition, taking the expectations in the above displayed formula,

$$\frac{d}{dt}Q_t = AQ_t + Q_tA' - \varphi_t(P_0)SQ_t - Q_tS\varphi_t(P_0) + \varphi_t(P_0)S\varphi_t(P_0) + R.$$

We set

$$U_t := Q_t - \varphi_t(P_0).$$

In this notation, we find that

$$\frac{d}{dt}U_t = AU_t + U_tA' - \varphi_t(P_0)SU_t - U_tS\varphi_t(P_0)$$
$$= \big(A - \varphi_t(P_0)S\big)U_t + U_t\big(A - \varphi_t(P_0)S\big)'.$$

The solution is given by

$$U_t = \exp\Big(\oint_0^t (A - \mathcal{P}_{\varphi_s(\eta_0)}S)\,ds\Big)U_0\Big(\exp\Big(\oint_0^t (A - \mathcal{P}_{\varphi_s(\eta_0)}S)\,ds\Big)\Big)'.$$

This shows that

$$Q_0 = \mathcal{P}_{\eta_0} \quad\Longleftrightarrow\quad \forall t \geq 0 \qquad Q_t = \varphi_t(P_0).$$

This completes the proof of the lemma. $\square$

5.2. *Stable signal processes.* This short section provides some rather elementary contraction inequalities when the drif-matrix of the signal process is stable w.r.t. the log-norm. The next proposition presents some *global* Lipschitz property.

PROPOSITION 5.3. *For any time horizon $t \geq 0$, we have the Lipschitz properties*:

$$(44) \quad (S) \quad and \quad \mu(A) < 0 \quad \Longrightarrow \quad \|P_t - P_t^\star\|_F \leq \exp(2\mu(A)t)\|P_0 - P_0^\star\|_F.$$

In addition for any $n \geq 2$, we have

$$(45) \qquad \mathbb{W}_n(\pi_t, \pi_t^\star) \leq \exp(\mu(A)t/2)[\mathbb{W}_n(\pi_0, \pi_0^\star) + c\|P_0 - P_0^\star\|_F]$$

for some finite constant c.

The detailed proofs of (44) and (45) are provided in Section 9.
Rewritten in terms of the Riccati semigroup, by (23) we have

$$(44) \quad \Leftrightarrow \quad \|P_t - P_t^\star\|_F = \|\mathcal{P}_{\eta_t} - \mathcal{P}_{\eta_t^\star}\|_F = \|\varphi_{0,t}(\mathcal{P}_{\pi_0}) - \varphi_{0,t}(\mathcal{P}_{\pi_0^\star})\|_F$$

$$\leq e^{2\mu(A)t}\|\mathcal{P}_{\pi_0} - \mathcal{P}_{\pi_0^\star}\|_F.$$

Of course there exist many distributions with a prescribed covariance matrix. The next lemma provides some Lipschitz properties of the trace and the Frobenius norm w.r.t. the Wasserstein metric. These properties allow to quantify the continuity property of the covariation matrices w.r.t. a given distribution.

LEMMA 5.4. *For any probability distributions $(\pi_0, \pi_0^\star)$ on $\mathbb{R}^{r_1}$, we have the regularity property*

$$(4^{-1}|\mathrm{tr}(\mathcal{P}_{\pi_0} - \mathcal{P}_{\pi_0^\star})|) \vee \|\mathcal{P}_{\pi_0} - \mathcal{P}_{\pi_0^\star}\|_F$$

$$\leq \mathbb{W}_2(\pi_0, \pi_0^\star)\|\pi_0^\star(e_2)\|^{1/2} + 2^{-1/2}\mathbb{W}_2(\pi_0, \pi_0^\star)^2$$

with the function $x = (x_i)_{1 \leq i \leq r_1} \in \mathbb{R}^{r_1} \mapsto e_2(x) := (x_i^2)_{1 \leq i \leq r_1}\mathbb{R}^{r_1}$.

The proof of this lemma is rather technical and lengthy, thus it its housed in the Appendix on page 841.
Lemma 5.4 can be used to deduce several functional contraction inequalities w.r.t the Wasserstein distance between the initial distributions of the Kalman–Bucy diffusion. For instance, combining Lemma 5.4 with (45) we readily obtain the following proposition.

PROPOSITION 5.5. *Assume that $\mu(A) < 0$ and (S) is satisfied. In this case, for any $t \geq 0$, the following nonlinear functional inequality holds*:

$$\mathbb{W}_2(\pi_t, \pi_t^\star) \leq ce^{\mu(A)t/2}\mathbb{W}_2(\pi_0, \pi_0^\star)[1 + \|\pi_0^\star(e_2)\|^{1/2} + \mathbb{W}_2(\pi_0, \pi_0^\star)]$$

for some finite constant c.

5.3. *Unstable signal processes.* The two main theorems stated in Section 3.2 (Theorems 3.4 and 3.5) show that the nonlinear Kalman–Bucy diffusions can be

stable even when the drift-matrix of the signal is unstable. The proof of these stability properties rely on Bucy's analysis of the Riccati equation.

The following theorem is a direct consequence of the Lyapunov inequalities and the uniform spectral estimates stated in Lemmas 4 and 5 and Theorem 4, in the pioneering article by R. S. Bucy [14].

THEOREM 5.6 (Bucy [14]). *When the filtering problem is uniformly observable and controllable, for any $t \geq s \geq \upsilon$, we have the uniform estimates*

$$\sup_{P_0 \in \mathbb{S}_{r_1}^+} \left\| \exp\left[\oint_s^t (A - P_u S) \, du \right] \right\|_2 \leq \alpha_\upsilon \exp\{-\beta_\upsilon (t - s)\}$$

for some parameters $\alpha_\upsilon < \infty$ and $\beta_\upsilon > 0$. In addition, for any $t \geq 0$ we have

$$(46) \qquad \left\| P_t - P_t^\star \right\|_2 \leq \alpha_\upsilon(P_0, P_0^\star) \exp\{-2\beta_\upsilon t\} \left\| P_0 - P_0^\star \right\|_2$$

for some constant $\alpha_\upsilon(P_0, P_0^\star)$ whose values only depend on $(\upsilon, P_0, P_0^\star)$.

These important contributions were published in 1967 by R. S. Bucy in [15].

Combining Theorem 5.6 with the perturbation Lemma 1.1, we find the following corollary.

COROLLARY 5.7. *Under the assumptions of Theorem 5.6 for any $\varepsilon \in \,]0, 1]$, any $P_0 \in \mathbb{S}_{r_1}^+$, and any $s \leq t$ we have the exponential semigroup estimates*

$$(47) \quad \left\| \exp\left[\oint_s^t (A - P_u S) \, du \right] \right\|_2 \leq \overline{\kappa}_{\varepsilon, \varsigma}(P_0, \upsilon) \exp\left((1 - \varepsilon)\varsigma(A - P S)(t - s)\right)$$

and

$$(48) \qquad \left\| \exp\left[\oint_s^t (A - P_u S) \, du \right] \right\|_2 \leq \overline{\kappa}_\mu(P_0, \upsilon) \exp\left(\mu(A - P S)(t - s)\right).$$

In the above displayed formulae, the finite constants $\overline{\kappa}_\mu(P_0, \upsilon)$ and $\overline{\kappa}_{\varepsilon, \varsigma}(P_0, \upsilon)$ defined by

$$\log \overline{\kappa}_\mu(P_0, \upsilon) = \kappa(\varepsilon)^{-1} \log\left[\overline{\kappa}_{\varepsilon, \varsigma}(P_0, \upsilon) / \kappa(\varepsilon) \right]$$

$$= \|P_0 - P\|_2 \|S\|_2 \alpha_\upsilon(P_0, P) / (2\beta_\upsilon)$$

with the parameters $(\kappa(\varepsilon), \alpha_\upsilon(P_0, P), \beta_\upsilon)$ presented in (4) and (46).

We also have

$$\frac{d}{dt}(P_t - P_t^\star) = (A - P_t^\star S)(P_t - P_t^\star) + (P_t - P_t^\star)(A - P_t S)'$$

$$(49) \qquad \Rightarrow \quad (P_t - P_t^\star) = \exp\left(\oint_s^t (A - P_u^\star S) \, du \right)(P_s - P_s^\star)$$

$$\times \left[\exp\left(\oint_s^t (A - P_u S) \, du \right) \right]'.$$

This readily implies the following result.

COROLLARY 5.8. *Under the assumptions of Theorem* 5.6 *for any* $\varepsilon \in {]0, 1]}$ *and any* $t \geq 0$, *we have the exponential semigroup estimates*

$$(50) \qquad \begin{aligned} \| P_t - P_t^\star \|_2 &\leq \overline{\kappa}_{\varepsilon,\varsigma}(P_0, \upsilon) \overline{\kappa}_{\varepsilon,\varsigma}(P_0^\star, \upsilon) \\ &\quad \times \exp\left(2(1-\varepsilon)\varsigma(A - PS)t\right) \| P_0 - P_0^\star \|_2 \end{aligned}$$

as well as

$$(51) \qquad \| P_t - P_t^\star \|_2 \leq \overline{\kappa}_\mu(P_0, \upsilon) \overline{\kappa}_\mu(P_0^\star, \upsilon) \exp\left(2\mu(A - PS)t\right) \| P_0 - P_0^\star \|_2$$

with functions $Q \mapsto \overline{\kappa}_\mu(Q, \upsilon)$ *and* $\overline{\kappa}_\varsigma(Q, \upsilon)$ *defined in Corollary* 5.7.

The article [53] also provides a similar exponential decay when $P_0^\star = P$, without the Lipschitz property w.r.t. the initial covariance matrix, and with half of the order of the rate of decays to equilibrium stated above.

For completeness and to better connect our work with existing literature on Riccati differential matrix equations, we end this section with some comments on the contraction theory of Riccati flows w.r.t the Thompson metric.

We recall that the Thompson's metric (a.k.a. part metric) on the space of definite positive matrices P_1, P_2 is defined by

$$d_T(P_1, P_2) = \log \max \left(M(P_1/P_2), M(P_2/P_1) \right)$$

with

$$M(P_1/P_2) := \inf \{ u \geq 0 : P_1 \leq u P_2 \}.$$

By a recent article by D. A. Snyder [59], we have

$$\| P_1 - P_2 \|_F \leq \left(\exp\left[d_T(P_1, P_2) \right] - 1 \right) \sqrt{ \frac{ \| P_1 \|_F^2 + \| P_2 \|_F^2 }{ 1 + \exp\left(2 d_T(P_1, P_2) \right) } }$$

$$\leq e d_T(P_1, P_2) \sqrt{ \| P_1 \|_F^2 + \| P_2 \|_F^2 }.$$

The last assertion is valid as soon as $d_T(P_1, P_2) \leq 1$.

Let $(P_t, P_t^\star)$ be two solutions of the Riccati equation starting at some possibly different states $(P_0, P_0^\star)$ such that

$$\alpha_1^{-1} P_0^\star \leq P_0 \leq \alpha_2 P_0^\star \quad \Longrightarrow \quad M(P_0/P_0^\star) \leq \alpha_2 \quad \text{and} \quad M(P_0^\star/P_0) \leq \alpha_1$$

for some $\alpha_1, \alpha_2 > 0$. By Theorem 8.5 in [42], for any $\beta > 0$ we have the contraction inequality

$$R \geq \beta S \quad \Longrightarrow \quad d_T(P_t, P_t^\star) \leq e \exp\left(-2\sqrt{\beta} t\right) d_T(P_0, P_0^\star)$$

$$(52) \qquad\qquad\qquad\qquad \leq e \exp\left(-2\sqrt{\beta} t\right)(\alpha_1 \vee \alpha_2)$$

$$\Longrightarrow \quad \| P_t - P_t^\star \|_F \leq e \exp\left(-2\sqrt{\beta} t\right)(\alpha_1 \vee \alpha_2) \sqrt{ \| P_t \|_F^2 + \| P_t^\star \|_F^2 }$$

as soon as $t \geq (2\sqrt{\beta})^{-1} \log(\alpha_1 \vee \alpha_2)$. Choosing $P_0^\star = P$ and $\alpha_1 = 1 \leq \alpha := \alpha_2$, we conclude that

(53)
$$P \leq P_0 \leq \alpha P \quad \text{and} \quad R \geq \beta S$$
$$\implies \quad \|P_t - P\|_F \leq \alpha e e^{-2\sqrt{\beta}t} (2\|P\|_F + e^{2\mu(A - PS)t} \|P_0 - P\|_F).$$

The estimates (52) and (53) are useful as soon $\|P_t\|_F$ is uniformly bounded. This property is ensured when the filtering problem is uniformly observable and controllable. In this situation, the exponential rate to equilibrium given in (53) is related to a signal to noise ratio associated with the pair of matrices (R, S).

We have derived a series of quantitative estimates for Kalman–Bucy diffusions. These estimates can be used to analyze the stability properties of Kalman–Bucy filters. For a more thorough discussion and a more recent account on the stability of discrete generation Kalman filters, we refer to [12, 19, 62] and the references therein. See also the pioneering article of Anderson [3], the one by Ocone and Pardoux [53] on the stability of continuous time Kalman–Bucy filters, as well as the book by H. Kwakernaak, R. Sivan [40].

6. A brief review on Ornstein–Ulhenbeck processes and Riccati equations.

6.1. *Some uniform moment estimates.* Our analysis on the convergence of the EnKF requires that $\mu(A) < 0$. This is not really surprising. The EnKF is designed in terms of interacting covariance matrices and interacting Monte Carlo samples based on the signal evolution. When $\mu(A) \geq 0$, the signal contains an unstable component. In this case, the fluctuations induced by the Monte Carlo samples may increase dramatically the global error variances. To analyze these interacting filters based on an extra level of randomness, we need to strengthen the usual condition $\mu(A - PS) < 0$ discussed in Section 5 to ensure that the signal itself is stable.

Next, we show that the condition $\mu(A) < 0$ cannot be relaxed. In the one-dimensional case when $C = 0$, the EnKF resumes to N independent copies of the signal and $\widehat{X}_t = \mathbb{E}(X_t)$. In this case, we have

$$\mu(A) \geq 0 \quad \Longleftrightarrow \quad N\mathbb{E}[(m_t^N - \widehat{X}_t)^2] = R(2\mu(A))^{-1}(e^{2\mu(A)t} - 1) \longrightarrow_{t \uparrow \infty} \infty.$$

When $\mu(A) = 0$, we use the convention $(2\mu(A))^{-1}(e^{2\mu(A)t} - 1) = t$. This shows that, even for one-dimensional Brownian signal motions, it is hopeless to try to find some uniform estimates for the sample mean.

Next, we provide a brief discussion on the stability of multidimensional signal processes. For multidimensional filtering problems, the solution of the signal stochastic differential equation is given by the Ornstein–Uhlenbeck formula

$$X_t = e^{At} X_0 - (\text{Id} - e^{At}) A^{-1} a + \int_0^t e^{A(t-s)} R^{1/2} \, dW_s.$$

The mean vector and the covariance matrix are given by

$$(\mathbb{E}(X_t) + A^{-1}a) = e^{tA} (\mathbb{E}(X_0) + A^{-1}a)$$

and

$$P_t^X = e^{tA} P_0 e^{A't} + \int_0^t e^{As} R e^{A's}\, ds.$$

In signal processing and control theory, the integral in the right-hand side term is called the controllability Grammian. Recalling that $\|\exp(tA)\| \leq \exp(\mu(A)t)$ we find that

$$\|\mathbb{E}(X_t) + A^{-1}a\| \leq \exp(\mu(A)t)\|\mathbb{E}(X_0) + A^{-1}a\|$$

and

$$\left\|\int_t^\infty e^{As} R e^{A's}\, ds\right\| \leq \exp(2t\mu(A))\frac{\|R\|}{2|\mu(A)|}.$$

Recall that $-A^{-1} = \int_0^\infty e^{tA}\, dt \Rightarrow \|A^{-1}\| \leq 1/|\mu(A)|$. Thus, the condition $\mu(A) < 0$ (and of course course $\|a\| < \infty$) ensures that

$$\lim_{t\to\infty}\|\mathbb{E}(X_t) + A^{-1}a\| = 0 = \lim_{t\to\infty}\left\|P_t^X - \int_0^\infty e^{As} R e^{A's}\, ds\right\|.$$

It also yields the uniform moment estimates

(54)
$$\sup_{t\geq 0}\mathbb{E}\big(\|\widehat{X}_t\|^n\big) < \infty \quad \text{and} \quad \sup_{t\geq 0}\mathbb{E}\big(\|X_t\|^n\big) < \infty$$
$$\text{as well as } \sup_{t\geq 0}\|P_t^X\|_F < \infty$$

for any $n \geq 1$. The last assertion is easily checked using Bernstein's inequality

$$\text{tr}\big(e^A e^{A'}\big) \leq \text{tr}\big(e^{A+A'}\big) \qquad \big(\leq r_1 e^{2\mu(A)}\big).$$

A proof of this inequality result can be found in [7]; see also [69, 70] for a more thorough discussion on trace inequalities.

6.2. *Riccati equations.* The Riccati equation (9) can be solved analytically for nonobserved or noise-free signals (i.e., $C = 0$ or $R = 0$). The situation $C = 0$ has already been discussed above. In this case, $P_t = P_t^X$ resumes to the covariance matrix P_t^X of the signal process.

When $R = 0$, the solution is given by

$$P_t = e^{tA}\left(P_0^{-1} + \int_0^t e^{sA'} S e^{sA}\, ds\right)^{-1} e^{tA'}$$
$$\big(\Rightarrow P_t = (P_0^{-1} + St)^{-1} \text{ when } A = 0\big).$$

In more general situations, we need to resort to some numerical scheme or to some additional algebraic development such as the Bernoulli substitution approach to reduce the problem to an ordinary linear differential equation in $2r_1$-dimensions. For

one-dimensional signal processes ($r_1 = 1$), P_t coincides with the variance between $\widehat{X}_t$ and X_t. When $S \neq 0$ and $R \neq 0$, the Riccati equation takes the form

$$\frac{d}{dt} P_t = -S(P_t - z_1)(P_t - z_2)$$

with the couple of roots

$$z_1 = \frac{A - \sqrt{A^2 + SR}}{S} < 0 < z_2 = \frac{A + \sqrt{A^2 + SR}}{S}.$$

The solution is given by the formula:

$$(55) \qquad P_t - z_2 = (P_0 - z_2)\frac{(z_2 - z_1)e^{-2t\sqrt{A^2+SR}}}{(z_2 - P_0)e^{-2t\sqrt{A^2+SR}} + (P_0 - z_1)} \longrightarrow_{t\to\infty} 0.$$

The above formula underlines the fact that the Riccati equation is stable even for unstable signals, that is, when $A > 0$. Also observe that

$$(56) \qquad 0 \le P_t \le z_2 + (P_0 - z_2)_+ e^{-2t\sqrt{A^2+SR}}$$

as soon as $P_0 \ge 0$, where $a_+ := \max(a, 0)$, for any $a \in \mathbb{R}$. We check this claim using the decomposition

$$P_t - z_2 = (P_0 - z_2)\frac{(z_2 - z_1)e^{-2t\sqrt{A^2+SR}}}{(P_0 - z_2)(1 - e^{-2t\sqrt{A^2+SR}}) + (z_2 - z_1)}.$$

Assume that $S \neq 0$ and let $P = z_2$ the unique positive fixed point. In this case, we notice that $-\sqrt{A^2 + SR} = A - PS < 0$ iff $A^2 \wedge R > 0$; and $A - PS/2 = z_1 S/2 < 0$ iff $R > 0$. Our regularity assumption $A - PS/2 < 0$ cannot capture the case $R = 0$ and $A \neq 0$.

Another direct consequence of this result is that the minimum variance function $t \mapsto P_t$ is uniformly bounded w.r.t. the time parameter. For matrix valued Riccati equation, we can use the following comparison lemma.

LEMMA 6.1.   *We assume that $\mu(A) < 0$. In this situation, we have* $\mathrm{tr}(P_t) \le g_t$ *where g_t stands for the solution of the Riccati equation*

$$\frac{d}{dt} g_t = 2\alpha g_t - \beta g_t^2 + r \qquad starting\ at\ g_0 = \mathrm{tr}(P_0)$$

with the parameters $(\alpha, \beta, r) = (\mu(A), r_1^{-1}\lambda_{\min}(S), \mathrm{tr}(R))$.

PROOF.   The key idea is to use the commutation inequality

$$(57) \quad \begin{aligned} &\mathrm{tr}(AP_t + P_t A' - P_t S P_t + R) \\ &\qquad = 2\,\mathrm{tr}(A_{\mathrm{sym}} P_t) - \mathrm{tr}(S P_t^2) + \mathrm{tr}(R) \le 2\alpha\,\mathrm{tr}(P_t) - \beta\big(\mathrm{tr}(P_t)\big)^2 + r. \end{aligned}$$

In the last display, we have used (6). This yields the Riccati differential inequality

$$\frac{d}{dt} \operatorname{tr}(P_t) \leq 2\alpha \operatorname{tr}(P_t) - \beta \big(\operatorname{tr}(P_t)\big)^2 + r$$

from which we conclude that

$$\operatorname{tr}(P_t) \leq g_t + e^{2\alpha t}\big(\operatorname{tr}(P_0) - g_0\big) \qquad \text{with } \partial g_t = 2\alpha g_t - \beta g_t^2 + r.$$

This completes the proof of the lemma. $\square$

Using Theorem 3.4 or Lemma 6.1, we readily deduce the following uniform estimates:

$$(58) \quad \mu(A) < 0 \quad \Longrightarrow \quad \sup_{t \geq 0} \operatorname{tr}(P_t) < \infty \quad \text{or} \quad \text{equivalently } \sup_{t \geq 0} \|P_t\|_F < \infty.$$

The left-hand side inequality in (58) is proven using the same analysis as the one of the scalar Riccati equation (55). The equivalence property in (58) is a direct consequence of (6). The estimate (56) also implies that

$$0 \leq \operatorname{tr}(P_t) \leq r_1 \mu(A)\frac{1+\delta}{\lambda_{\min}(S)} + \left(\operatorname{tr}(P_0) - r_1 \mu(A)\frac{1+\delta}{\lambda_{\min}(S)}\right)_+ \exp\big[2t\mu(A)\delta\big]$$

with the parameter $\delta := \sqrt{1 + r_1^{-1}[\operatorname{tr}(R)\lambda_{\min}(S)]/\mu(A)^2}$. These estimates are useful as soon as $P_0 \geq P$. When $P_0 \leq P$, we clearly have $\operatorname{tr}(P_t) \leq \operatorname{tr}(P)$.

7. The Ensemble Kalman–Bucy filter equations.

7.1. *Sample mean and covariance diffusions.* This section is mainly concerned with the proof of the stochastic differential equations (15) and (16).

The stochastic diffusion equation of the EnKF sample mean (15) is easily checked using (12) with the r_1-multidimensional martingale defined by

$$d\overline{M}_t = \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} R^{1/2}\, d\overline{W}_t^i - \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq N} P_t^N C' R_2^{-1/2}\, d\overline{V}_t^i.$$

This clearly implies (16). Using (12), we also readily check that

$$d(\xi_t^i - m_t^N) = (A - P_t^N S)(\xi_t^i - m_t^N)\, dt + dM_t^i$$

with the r_1-dimensional martingale

$$dM_t^i := R^{1/2}\, d\widetilde{W}_t^i - P_t^N C' R_2^{-1/2}\, d\widetilde{V}_t^i$$

defined in terms of the diffusion processes

$$\widetilde{W}_t^i = \overline{W}_t^i - N^{-1} \sum_{1 \leq j \leq N} \overline{W}_t^j \quad \text{and} \quad \widetilde{V}_t^i = \overline{V}_t^i - N^{-1} \sum_{1 \leq j \leq N} \overline{V}_t^j.$$

The angle-brackets $\langle M^i(k), M^j(k')\rangle_t$ associated with the collection of vector valued martingales M^i_t are given by the formulae

$$(59) \qquad \frac{d}{dt}\langle M^i(k), M^i(k')\rangle_t = \left(1 - \frac{1}{N}\right)(R + P^N_t S P^N_t)(k, k')$$

and for $i \neq j$

$$(60) \qquad \frac{d}{dt}\langle M^i(k), M^j(k')\rangle_t = -\frac{1}{N}(R + P^N_t S P^N_t)(k, k').$$

To check this claim, we observe that

$$dM^i_t := \left(1 - \frac{1}{N}\right)[R^{1/2} d\overline{W}^i_t - P^N_t C' R_2^{-1/2} d\overline{V}^i_t]$$

$$- \frac{1}{N} \sum_{j \neq i}[R^{1/2} d\overline{W}^j_t - P^N_t C' R_2^{-1/2} d\overline{V}^j_t]$$

$$= \sum_j \varepsilon^i_j [R^{1/2} d\overline{W}^j_t - P^N_t C' R_2^{-1/2} d\overline{V}^j_t]$$

with

$$\varepsilon^i_j = 1_{i=j}\left(1 - \frac{1}{N}\right) - \frac{1}{N} 1_{i \neq j}.$$

Therefore,

$$dM^i_t(k) = \sum_{j,l} \varepsilon^i_j R^{1/2}(k, l) d\overline{W}^j_t(l) - \sum_{j,l} \varepsilon^i_j (P^N_t C' R_2^{-1/2})(k, l) d\overline{V}^j_t(l)$$

from which we conclude that

$$\frac{d}{dt}\langle M^i(k), M^i(k')\rangle_t$$

$$= \sum_{j,l} \sum_{j',l'} \varepsilon^i_j \varepsilon^i_{j'} R^{1/2}(k, l) R^{1/2}(k', l') 1_{j=j'} 1_{l=l'}$$

$$+ \sum_{j,l} \sum_{j',l'} \varepsilon^i_j \varepsilon^i_{j'} (P^N_t C' R_2^{-1/2})(k, l)(P^N_t C' R_2^{-1/2})(k', l') 1_{j=j'} 1_{l=l'}$$

$$= \sum_j (\varepsilon^i_j)^2 \Big[\sum_l R^{1/2}(k, l) R^{1/2}(k', l)$$

$$+ \sum_{j,l} (P^N_t C' R_2^{-1/2})(k, l)(P^N_t C' R_2^{-1/2})(k', l)\Big]$$

$$= \sum_j (\varepsilon^i_j)^2 (R - P^N_t S P^N_t)(k, k').$$

In the last assertion, we have used the symmetry of the matrices $R^{1/2}$ and

$$(P_t^N C' R_2^{-1/2})(k', l) = (P_t^N C' R_2^{-1/2})'(l, k') = (R_2^{-1/2} C P_t^N)(l, k').$$

Observe that

$$\sum_j (\varepsilon_j^i)^2 = \left(1 - \frac{1}{N}\right)^2 + \sum_{j \neq i} \frac{1}{N^2} = \left(1 - \frac{1}{N}\right)\left(\left(1 - \frac{1}{N}\right) + \frac{1}{N}\right) = \left(1 - \frac{1}{N}\right).$$

When $i \neq i'$ we have

$$\frac{d}{dt} \langle M^i(k), M^{i'}(k') \rangle_t = \sum_j \varepsilon_j^i \varepsilon_j^{i'} (R - P_t^N S P_t^N)(k, k')$$

and

$$\sum_j \varepsilon_j^i \varepsilon_j^{i'} = \varepsilon_i^i \varepsilon_i^{i'} + \varepsilon_{i'}^i \varepsilon_{i'}^{i'} + \sum_{j \notin \{i, i'\}} \varepsilon_j^i \varepsilon_j^{i'}$$

$$= -2 \frac{1}{N} \left(1 - \frac{1}{N}\right) + (N - 2) \frac{1}{N^2}$$

$$= \frac{1}{N} \left(\left(\frac{2}{N} - 2\right) + \left(1 - \frac{2}{N}\right)\right) = -\frac{1}{N}.$$

This completes the proof of the angle-bracket formulae (59) and (60).

This implies that

$$d[(\xi_t^i - m_t^N)(\xi_t^i - m_t^N)']$$

$$= (A - P_t^N S)(\xi_t^i - m_t^N)(\xi_t^i - m_t^N)' \, dt$$

$$+ (\xi_t^i - m_t^N)(\xi_t^i - m_t^N)'(A' - S P_t^N) \, dt$$

$$+ (1 - N^{-1})(R + P_t^N S P_t^N) \, dt + (\xi_t^i - m_t^N)(dM_t^i)' + dM_t^i (\xi_t^i - m_t^N)'.$$

Summing the indices, we find that

$$dP_t^N = [(A - P_t^N S) P_t^N + P_t^N (A' - S P_t^N) + (R + P_t^N S P_t^N)] \, dt$$

$$+ \frac{1}{\sqrt{N - 1}} dM_t$$

with

$$dM_t := \frac{1}{\sqrt{N - 1}} \sum_{1 \leq i \leq N} [(\xi_t^i - m_t^N)(dM_t^i)' + dM_t^i (\xi_t^i - m_t^N)'].$$

This completes the proof of (17). To check (18), we set $\varepsilon_t^i := \xi_t^i - m_t^N$. In this notation, we have

$$dM_t(k, l) := \frac{1}{\sqrt{N - 1}} \sum_{1 \leq i \leq N} [\varepsilon_t^i(k) dM_t^i(l) + dM_t^i(k) \varepsilon_t^i(l)].$$

This implies that

$$(N - 1) \frac{d}{dt} \langle M(k, l), M(k', l') \rangle$$

$$= \sum_{1 \le i, i' \le N} \left[\varepsilon_t^i(k) \varepsilon_t^{i'}(k') \frac{d}{dt} \langle M^i(l), M^{i'}(l') \rangle_t \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^{i'}(l') \frac{d}{dt} \langle M^i(k), M^{i'}(k') \rangle_t \right]$$

$$+ \sum_{1 \le i, i' \le N} \left[\varepsilon_t^i(k) \varepsilon_t^{i'}(l') \frac{d}{dt} \langle M^i(l), M^{i'}(k') \rangle_t \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^{i'}(k') \frac{d}{dt} \langle M^i(k), M^{i'}(l') \rangle_t \right].$$

By (59) and (60), we have

$$(N - 1) \frac{d}{dt} \langle M(k, l), M(k', l') \rangle$$

$$= \left(1 - \frac{1}{N} \right) \sum_{1 \le i \le N} \left[\varepsilon_t^i(k) \varepsilon_t^i(k') (R + P_t^N S P_t^N)(l, l') \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^i(l') (R + P_t^N S P_t^N)(k, k') \right]$$

$$+ \left(1 - \frac{1}{N} \right) \sum_{1 \le i \le N} \left[\varepsilon_t^i(k) \varepsilon_t^i(l') (R + P_t^N S P_t^N)(l, k') \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^i(k') (R + P_t^N S P_t^N)(k, l') \right]$$

$$- \frac{1}{N} \sum_{1 \le i \ne i' \le N} \left[\varepsilon_t^i(k) \varepsilon_t^{i'}(k') (R + P_t^N S P_t^N)(l, l') \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^{i'}(l') (R + P_t^N S P_t^N)(k, k') \rangle_t \right]$$

$$- \frac{1}{N} \sum_{1 \le i \ne i' \le N} \left[\varepsilon_t^i(k) \varepsilon_t^{i'}(l') (R + P_t^N S P_t^N)(l, k') \right.$$

$$+ \varepsilon_t^i(l) \varepsilon_t^{i'}(k') (R + P_t^N S P_t^N)(k, l') \right].$$

Recalling that

$$\frac{1}{N - 1} \sum_{1 \le i \le N} \varepsilon_t^i(k) \varepsilon_t^i(k') = P_t^N(k, k')$$

and

$$\sum_{1 \leq i \leq N} \varepsilon_t^i(k) \sum_{1 \leq i' \leq N} \varepsilon_t^{i'}(k') = 0$$

$$\Rightarrow \quad \frac{1}{N-1} \sum_{1 \leq i \neq i' \leq N} \varepsilon_t^i(k) \varepsilon_t^{i'}(k') = -\frac{1}{N-1} \sum_{1 \leq i \leq N} \varepsilon_t^i(k) \varepsilon_t^i(k')$$

$$= -P_t^N(k, k')$$

we find that

$$\frac{d}{dt} \langle M(k, l), M(k', l') \rangle$$

$$= \left(1 - \frac{1}{N}\right)$$

$$\times \left[P_t^N(k, k')(R + P_t^N S P_t^N)(l, l') + P_t^N(l, l')(R + P_t^N S P_t^N)(k, k') \right]$$

$$+ \left(1 - \frac{1}{N}\right)$$

$$\times \left[P_t^N(k, l')(R + P_t^N S P_t^N)(l, k') + P_t^N(l, k')(R + P_t^N S P_t^N)(k, l') \right]$$

$$+ \frac{1}{N} \left[P_t^N(k, k')(R + P_t^N S P_t^N)(l, l') + P_t^N(l, l')(R + P_t^N S P_t^N)(k, k')\rangle_t \right]$$

$$+ \frac{1}{N} \left[p(k, l')(R + P_t^N S P_t^N)(l, k') + P_t^N(l, k')(R + P_t^N S P_t^N)(k, l') \right].$$

This completes the proof of (18). The last assertion can be checked easily using the fact that $\langle \overline{M}(n), M^i \rangle_t$ does not depend on the index i.

This completes the proof of the EnKF differential equations (15) and (16).

7.2. *Uniform moments estimates.* The following technical lemma combines a Foster–Lyapunov approach with martingale techniques to control the moments of Riccati-type stochastic differential equations uniformly w.r.t. the time horizon.

LEMMA 7.1. *Let Z_t be some stochastic processes adapted to some filtration $\mathcal{F}_t$ and taking values in some measurable state space $(E, \mathcal{E})$. Let H be some nonnegative measurable function on $(E, \mathcal{E})$ such that*

(61) $$dH(Z_t) = \mathcal{L}_t H(Z_t)\, dt + d\mathcal{M}_t(H)$$

with an $\mathcal{F}_t$-martingale $\mathcal{M}_t(H)$ and some $\mathcal{F}_t$-adapted process $\mathcal{L}_t H(Z_t)$.

• *Assume that*

$$\mathcal{L}_t H(Z_t) \leq 2\gamma \sqrt{H(Z_t)} + 3\alpha H(Z_t) - \beta H(Z_t)^2 + r,$$

$$\frac{d}{dt} \langle \mathcal{M}(H) \rangle_t \leq H(Z_t)\left(\tau_0 + \tau_1 H(Z_t) + \tau_2 H(Z_t)^2\right)$$

for some parameters $\alpha < 0$ and $\gamma, \beta, r, \tau_0, \tau_1, \tau_2 \geq 0$. In this situation, we have the uniform moment estimates

$$1 \leq n < 1 + 2\min\left(\beta/\tau_2, |\alpha|/\tau_1\right) \quad \Longrightarrow \quad \sup_{t \geq 0}\mathbb{E}\left(H(Z_t)^n\right) < \infty$$

with the convention $\beta/0 = \infty = |\alpha|/0$ when $\tau_2 = 0$ or when $\tau_1 = 0$.

- *Assume that*

$$\mathcal{L}_t H(Z_t) \leq 2\tau_t(Z_t)\sqrt{H(Z_t)} + 2\alpha H(Z_t) + \beta_t(Z_t)$$

$$\frac{d}{dt}\langle \mathcal{M}(H)\rangle_t \leq H(Z_t)\gamma_t(Z_t)$$

for some $\alpha < 0$ and some nonnegative functions $(\tau_t, \beta_t, \gamma_t)$ s.t.

$$\delta_{\tau,t}(n) := \mathbb{E}\left(\tau_t(Z_t)^n\right)^{\frac{1}{n}} < \infty$$

$$\delta_{\beta,t}(n) := \mathbb{E}\left(\beta_t(Z_t)^n\right)^{\frac{1}{n}} < \infty \quad \text{and} \quad \delta_{\gamma,t}(n) := \mathbb{E}\left(\gamma_t(Z_t)^n\right)^{\frac{1}{n}} < \infty$$

for any $n \geq 1$. In this situation, we have the estimate

$$\mathbb{E}\left(H(Z_t)^n\right)^{\frac{1}{n}}$$

$$\leq e^{\alpha t}\mathbb{E}\left(H(Z_0)^n\right)^{\frac{1}{n}}$$

$$+ \int_0^t e^{\alpha(t-s)}\left[\left(\delta_{\tau,s}(2n)^2/|\alpha| + \delta_{\beta,s}(n) + (n-1)\delta_{\gamma,s}(n)/2\right)\right]ds.$$

PROOF.   First, we observe that

$$\forall \varepsilon > 0 : 2\gamma\sqrt{H(Z_t)} \leq \varepsilon H(Z_t) + \frac{1}{\varepsilon}\gamma^2.$$

Choosing $\varepsilon = |\alpha| > 0$, we find that

$$\mathcal{L}_t H(Z_t) \leq 2\alpha H(Z_t) - \beta H(Z_t)^2 + r + \gamma^2/|\alpha|.$$

For any $n \geq 1$, we have

$$dH(Z_t)^n = \mathcal{L}_{n,t}H(Z_t)\,dt + d\mathcal{M}_{n,t}(H)$$

with the martingale $d\mathcal{M}_{n,t}(H) := nH(Z_t)^{n-1}d\mathcal{M}_t(H)$ and the drift

$$\mathcal{L}_{n,t}H(Z_t) = n\left[H(Z_t)^{n-1}\mathcal{L}_t H(Z_t) + \frac{(n-1)}{2}H(Z_t)^{n-2}\frac{d}{dt}\langle \mathcal{M}(H)\rangle_t\right]$$

$$\leq 2n\alpha_n H(Z_t)^n - n\beta_n H(Z_t)^{n+1} + n\rho_n H(Z_t)^{n-1}$$

with

$$\alpha_n = \alpha + \tau_1\frac{(n-1)}{2} < 0,$$

$$\beta_n = \left(\beta - \frac{(n-1)}{2}\tau_2\right) > 0$$

and
$$\rho_n = \left(r + \frac{(n-1)}{2}\tau_0\right) > 0$$

as soon as $1 \le n < 1 + 2|\alpha|\tau_1^{-1}$ and $1 \le n < 1 + 2\beta\tau_2^{-1}$. This implies that

$$\frac{1}{n}\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big) \le 2\alpha_n\mathbb{E}\big(H(Z_t)^n\big) - \beta_n\mathbb{E}\big(H(Z_t)^{n+1}\big) + \rho_n\mathbb{E}\big(H(Z_t)^{n-1}\big)$$

$$\le 2\alpha_n\mathbb{E}\big(H(Z_t)^n\big) - \beta_n\mathbb{E}\big(H(Z_t)^n\big)^{1+\frac{1}{n}} + \rho_n\mathbb{E}\big(H(Z_t)^n\big)^{1-\frac{1}{n}}$$

and, therefore,

$$\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}} = \frac{1}{n}\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}-1}\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big)$$

$$\le 2\alpha_n\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}} - \beta_n\mathbb{E}\big(H(Z_t)^n\big)^{\frac{2}{n}} + \rho_n.$$

This shows that

$$\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}} \le g_{n,t} + e^{2\alpha_n t}\big(\mathbb{E}\big(H_0(Z_0)^n\big)^{\frac{1}{n}} - g_{n,0}\big)$$

with

$$\frac{d}{dt}g_{n,t} = 2\alpha_n g_{n,t} - \beta_n g_{n,t}^2 + \rho_n.$$

The end of the proof of the first assertion follows the same lines of arguments as the ones of Lemma 6.1.

Now we come to the proof of the second assertion.

Arguing as above, we have

$$\forall \varepsilon > 0 : 2\tau_t(Z_t)\sqrt{H(Z_t)} \le \varepsilon H(Z_t) + \tau_t(Z_t)^2/\varepsilon.$$

Choosing $\varepsilon = |\alpha|$, we find that

$$\mathcal{L}_t H(Z_t) \le \alpha H(Z_t) + \beta_t(Z_t) + \tau_t(Z_t)^2/|\alpha|.$$

Therefore, there is no loss of generality to assume that $\tau_t(Z_t) = 0$ by changing 2α by α and $\beta_t(Z_t)$ by $\beta_t(Z_t) + \frac{\tau_t(Z_t)^2}{|\alpha|}$. By (61), we have

$$dH(Z_t)^n = nH(Z_t)^{n-1}\mathcal{L}_t H(Z_t)\,dt$$
$$+ \frac{n(n-1)}{2}H(Z_t)^{n-2}\frac{d}{dt}\langle\mathcal{M}(H)\rangle_t\,dt + nH(Z_t)^{n-1}\,d\mathcal{M}_t(H).$$

This implies that

$$\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big) \le n\alpha\mathbb{E}\big(H(Z_t)^n\big)$$
$$+ n\bigg[\mathbb{E}\big(\beta_t(Z_t)H(Z_t)^{n-1}\big) + \frac{(n-1)}{2}\mathbb{E}\big(\gamma_t(Z_t)H(Z_t)^{n-1}\big)\bigg].$$

Using Hölder inequality, we have

$$\mathbb{E}\big(\beta_t(Z_t)H(Z_t)^{n-1}\big) \leq \mathbb{E}\big(H(Z_t)^n\big)^{1-\frac{1}{n}}\mathbb{E}\big(\beta_t^n(Z_t)\big)^{\frac{1}{n}} \leq \delta_{\beta,t}(n)\mathbb{E}\big(H(Z_t)^n\big)^{1-\frac{1}{n}}.$$

In much the same way, we have

$$\mathbb{E}\big(\gamma_t(Z_t)H(Z_t)^{n-1}\big) \leq \delta_{\gamma,t}(n)\mathbb{E}\big(H(Z_t)^n\big)^{1-\frac{1}{n}}.$$

This yields the estimate:

$$\partial \mathbb{E}\big(H(Z_t)^n\big) \leq n\alpha\mathbb{E}\big(H(Z_t)^n\big) + n\delta_t(n)\mathbb{E}\big(H(Z_t)^n\big)^{1-\frac{1}{n}}$$

with $\delta_t(n) = \delta_{\beta,t}(n) + (n-1)\delta_{\gamma,t}(n)/2$. We conclude that

$$\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}} = \frac{1}{n}\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}-1}\frac{d}{dt}\mathbb{E}\big(H(Z_t)^n\big) \leq \alpha\mathbb{E}\big(H(Z_t)^n\big)^{\frac{1}{n}} + \delta_t(n).$$

The last assertion is a direct application of Grönwall inequality.

The proof of the lemma is now complete. □

PROPOSITION 7.2. *Assume that $\mu(A) < 0$. In this situation, we have the uniform trace moment estimates*

$$1 \leq n < 1 + \frac{N-1}{2r_1}\frac{\lambda_{\min}(S)}{\lambda_{\max}(S)} \quad \Longrightarrow \quad \sup_{t \geq 0}\mathbb{E}\big([\mathrm{tr}(P_t^N)]^n\big) \leq c(n)$$

with the convention $\frac{\lambda_{\min}(S)}{\lambda_{\max}(S)} = 0$ when $S = 0$. In addition, when condition (S) *is met we have*

$$\sup_{t \geq 0}\mathbb{E}\big(\|\zeta_t^1\|^{2n}\big) \leq c(n) \quad and \quad \sup_{t \geq 0}\mathbb{E}\big(\|\xi_t^1\|^{2n}\big) \leq c(n).$$

The left-hand side estimates are valid for any $n \geq 1$, while the right-hand side ones are valid for $3n < 1 + \frac{N-1}{2r_1}\frac{\lambda_{\min}(S)}{\lambda_{\max}(S)}$.

PROOF. We set $H(P_t^N) := \mathrm{tr}(P_t^N)$ the trace function of the random sample covariance matrices P_t^N. Using (17), we prove evolution equation

$$dH\big(P_t^N\big) = LH\big(P_t^N\big)dt + \frac{1}{\sqrt{N-1}}d\mathcal{M}_t \qquad \text{with } \mathcal{M}_t := \mathrm{tr}(M_t)$$

and the drift

$$LH\big(P_t^N\big) := 2\mathrm{tr}\big(A_{\mathrm{sym}}P_t^N\big) - \mathrm{tr}\big(Sp_t^2\big) + \mathrm{tr}(R).$$

Following the proof of Lemma 6.1, we also have the estimates

$$0 \leq \frac{d}{dt}\langle \mathcal{M}\rangle_t = 4\mathrm{tr}\big((R + P_t^N S P_t^N)P_t^N\big) \leq 4H\big(P_t^N\big)\big(\mu(R) + \mu(S)H\big(P_t^N\big)^2\big)$$

and

$$LH(P_t^N) \leq 2\alpha H(P_t^N) - \beta H(P_t^N)^2 + r$$

with

$$\alpha := 2\mu(A) < 0, \qquad \beta := r_1^{-1}\lambda_{\min}(S) \quad \text{and} \quad r := \text{tr}(R).$$

The moment estimates of the trace of the sample covariance matrices are now easily checked using Lemma 7.1.

Now we come to the proof of the moments estimates of the norm of the samples. Notice that

$$d\xi_t^i = ((A - P_t^N S)\xi_t^i + a + P_t^N SX_t) dt + dM_t^{\xi,i}$$

with r_1-dimensional martingale

$$dM_t^{\xi,i} := R^{1/2} d\overline{W}_t^i + P_t^N C' R_2^{-1/2} d(V_t - \overline{V}_t^i).$$

We set $H(X_t, \xi_t) := \|\xi_t^1\|^2$. In this notation, we have

$$dH(X_t, \xi_t) = \mathcal{L}_t H(X_t, \xi_t) dt + d\mathcal{M}_t(H) \quad \text{with } d\mathcal{M}_t(H) = 2\langle \xi_t^1, dM_t^{\xi,1}\rangle$$

and

$$\begin{aligned}\mathcal{L}_t H(X_t, \xi_t) = &\langle \xi_t^1, (A + A' - (P_t^N S + SP_t^N))\xi_t^1\rangle + 2\langle \xi_t^1, a + P_t^N SX_t\rangle \\ &+ \text{tr}(R + 2P_t^N SCP_t^N).\end{aligned}$$

On the other hand,

$$(S) \implies \langle \xi_t^1, (A + A' - (P_t^N S + SP_t^N))\xi_t^1\rangle \leq 2\mu(A)\|\xi_t^1\|^2.$$

Using the inequality,

$$(62) \qquad\qquad 2\langle x, y\rangle \leq \frac{1}{\varepsilon}\|x\|^2 + \varepsilon\|y\|^2$$

with $0 < \varepsilon = -\mu(A) < -2\mu(A)$ [recall that $\mu(A) < 0$] $y = \xi_t^i$ and $x = a + P_t^N SX_t$, we find that

$$\mathcal{L}_t H(X_t, \xi_t) \leq \mu(A)\|\xi_t^i\|^2 + \beta_t(X_t, \xi_t) = \mu(A)H(X_t, \xi_t) + \beta_t(X_t, \xi_t)$$

with

$$\beta_t(X_t, \xi_t) = |\mu(A)|^{-1}\|a + P_t^N SX_t\|^2 + \text{tr}(R + 2P_t^N SCP_t^N).$$

On the other hand, we have

$$\|a + P_t^N SX_t\| \leq \|a\| + \|S\|\|X_t\|\|P_t^N\| \qquad \text{with } \|P_t^N\| \leq \|P_t^N\|_F \leq \text{tr}(P_t^N)$$

and

$$\text{tr}((R + P_t^N SP_t^N)P_t^N) \leq \text{tr}(P_t^N)(\mu(R) + \mu(S)(\text{tr}(P_t^N))^2).$$

Combining these estimates with the trace estimates, we have just proven and the signal state uniform moment estimates stated in (54) a direct application of the Lemma 7.1 yields

$$\sup_{t \geq 0} \mathbb{E}(\|\xi_t^1\|^{2n}) \leq c(n).$$

Notice that the control of the nth moment of $\beta_t(X_t, \xi_t)$ involves the control of the $(3n)$th moment of the trace of P_t^N. The same analysis applies to ζ_t^1 with P_t^N replaced by P_t.

This ends the proof of the proposition. $\square$

8. Uniform estimates. This section is mainly concerned with the proof of the uniform estimates presented in Theorem 3.6. The first step is to control and to estimate the fluctuations of the particle covariance matrices involved in the EnKF filter uniformly w.r.t. the time horizon. In Section 8.1, we present a key uniform control of the Frobenius norm between the particle covariance matrices and their limiting values. These estimates are used in Section 8.2 to derive the uniform propagation of chaos properties of the EnKF.

8.1. *Uniform convergence of empirical covariance matrices.* The next theorem is pivotal. It describes the evolution of the Frobenius norm of the "centered" sample covariance matrices in terms of a nonlinear diffusion and provides some key uniform convergence results.

THEOREM 8.1. *The Frobenius norm of the sample covariance matrix fluctuations satisfies the diffusion equation*

$$(63) \qquad d\|P_t^N - P_t\|_F^2 = \left(\alpha_t(P_t^N) + \frac{1}{N-1}\beta_t(P_t^N)\right) dt + \frac{2}{\sqrt{N-1}} d\mathcal{M}_t$$

with the drift functions

$$\alpha_t(P_t^N) := 2\operatorname{tr}\left(\left[(A + A') - \frac{1}{2}\{(P_t^N + P_t)S + S(P_t^N + P_t)\}\right](P_t^N - P_t)^2\right)$$

$$\beta_t(P_t^N) := 2[\gamma_t(P_t^N) + \operatorname{tr}(R + P_t^N S P_t^N)\operatorname{tr}(P_t^N)]$$

$$\gamma_t(P_t^N) := \operatorname{tr}((R + P_t^N S P_t^N)P_t^N)$$

and a martingale $\mathcal{M}_t$ with angle bracket

$$\frac{d}{dt}\langle\mathcal{M}\rangle_t = 4\operatorname{tr}(P_t^N(P_t^N - P_t)(R + P_t^N S P_t^N)(P_t^N - P_t)).$$

In addition, when $\mu(A) < 0$ and (S) is satisfied we have the uniform mean error estimates

$$2n < 1 + \frac{N-1}{2r_1} \quad \Longrightarrow \quad \sup_{t \geq 0} \mathbb{E}(\|P_t^N - P_t\|_F^n)^{\frac{1}{n}} \leq c(n)/\sqrt{N}.$$

PROOF. By (9) and (17), we have

$$d(P_t^N - P_t) = [A(P_t^N - P_t) + (P_t^N - P_t)A' - (P_t^N S P_t^N - P_t S P_t)] dt$$
$$+ \frac{1}{\sqrt{N-1}} dM_t$$

with the martingale M_t with angle brackets defined in (18). Using the decomposition,

$$P_t^N S P_t^N - P_t S P_t = (P_t^N - P_t)S(P_t^N - P_t) + P_t S(P_t^N - P_t) + (P_t^N - P_t)S P_t,$$

we readily check that

$$A(P_t^N - P_t) + (P_t^N - P_t)A' - (P_t^N S P_t^N - P_t S P_t)$$
$$= \left[A - \frac{1}{2}(P_t^N + P_t)S\right](P_t^N - P_t) + (P_t^N - P_t)\left[A' - \frac{1}{2}S(P_t^N + P_t)\right].$$

This implies that

$$d((P_t^N - P_t)^2)$$
$$= 2(P_t^N - P_t)\left\{\left[A - \frac{1}{2}(P_t^N + P_t)S\right](P_t^N - P_t)\right.$$
$$\left. + (P_t^N - P_t)\left[A' - \frac{1}{2}S(P_t^N + P_t)\right]\right\} dt$$
$$+ \frac{1}{N-1}[(R + P_t^N S P_t^N)P_t^N + P_t^N(R + P_t^N S P_t^N)$$
$$+ \operatorname{tr}(P_t^N)(R + P_t^N S P_t^N) + (\operatorname{tr}(R + P_t^N S P_t^N))P_t^N] dt$$
$$+ \frac{2}{\sqrt{N-1}}(P_t^N - P_t) dM_t.$$

Taking the trace, we find that

$$d\|P_t^N - P_t\|_F^2$$
$$= 2\operatorname{tr}\left(\left[(A + A') - \frac{1}{2}\{(P_t^N + P_t)S + S(P_t^N + P_t)\}\right](P_t^N - P_t)^2\right) dt$$
$$+ \frac{2}{N-1}[\operatorname{tr}((R + P_t^N S P_t^N)p) + \operatorname{tr}((R + P_t^N S P_t^N))\operatorname{tr}(P_t^N)] dt$$
$$+ \frac{2}{\sqrt{N-1}} d\mathcal{M}_t$$

with the martingale

$$d\mathcal{M}_t = \operatorname{tr}((P_t^N - P_t) dM_t) = \sum_{1 \le k,l \le r_1} (P_t^N - P_t)(l,k) dM_t(k,l).$$

The angle bracket of $\mathcal{M}_t$ is computed using (18). More precisely, we have

$$\frac{d}{dt}\langle\mathcal{M}\rangle_t = 4\operatorname{tr}((R + P_t^N S P_t^N)(P_t^N - P_t)P_t^N(P_t^N - P_t)).$$

We check this claim using the decomposition

$$\frac{d}{dt}\langle\mathcal{M}\rangle_t = \sum_{1\le k,l,k',l'\le r_1}(P_t^N - P_t)(l,k)(P_t^N - P_t)(l',k')\frac{d}{dt}\langle M(k,l), M(k',l')\rangle_t.$$

Recalling that

$$\frac{d}{dt}\langle M(k,l), M(k',l')\rangle_t = (R + P_t^N S P_t^N)(k,k')P_t^N(l,l')$$

$$+ (R + P_t^N S P_t^N)(l,l')P_t^N(k,k')$$

$$+ (R + P_t^N S P_t^N)(l',k)P_t^N(k',l)$$

$$+ (R + P_t^N S P_t^N)(l,k')P_t^N(k,l')$$

and using the symmetry of the matrices (P_t^N, P_t) and $(R + P_t^N S P_t^N)$ we find that

$$\frac{d}{dt}\langle\mathcal{M}\rangle_t = \sum_{1\le k,l,k',l'\le r_1}P_t^N(l',l)(P_t^N - P_t)(l,k)(R + P_t^N S P_t^N)(k,k')$$

$$\times (P_t^N - P_t)(k',l')$$

$$+ \sum_{1\le k,l,k',l'\le r_1}(P_t^N - P_t)(k,l)(R + P_t^N S P_t^N)(l,l')(P_t^N - P_t)(l',k')$$

$$\times P_t^N(k',k)$$

$$+ \sum_{1\le k,l,k',l'\le r_1}P_t^N(k',l)(P_t^N - P_t)(l,k)(R + P_t^N S P_t^N)(k,l')$$

$$\times (P_t^N - P_t)(l',k')$$

$$+ \sum_{1\le k,l,k',l'\le r_1}(P_t^N - P_t)(l,k)P_t^N(k,l')(P_t^N - P_t)(l',k')$$

$$\times (R + P_t^N S P_t^N)(k',l).$$

This shows that

$$\frac{d}{dt}\langle\mathcal{M}\rangle_t = 4\operatorname{tr}(P_t^N(P_t^N - P_t)(R + P_t^N S P_t^N)(P_t^N - P_t)).$$

This completes the proof of the first assertion.

We set $H(P_t^N, P_t) = \|P_t^N - P_t\|_F^2$. In this notation, equation (63) takes the form

$$dH(P_t^N, P_t) = \mathcal{L}_t H(P_t^N, P_t)\,dt + d\mathcal{M}_t(H) \qquad \text{with } \mathcal{M}_t(H) := \frac{2}{\sqrt{N-1}}\mathcal{M}_t$$

and

$$\mathcal{L}_t H\big(P_t^N, P_t\big) = \alpha_t\big(P_t^N\big) + \frac{1}{N-1}\beta_t\big(P_t^N\big).$$

Under condition (S), we have

$$\operatorname{tr}\Big(\Big[(A + A') - \frac{1}{2}\{(P_t^N + P_t)S + S(P_t^N + P_t)\}\Big](P_t^N - P_t)^2\Big)$$
$$= \operatorname{tr}\big((A + A')(P_t^N - P_t)^2\big) - \rho(S)\operatorname{tr}\big((P_t^N + P_t)(P_t^N - P_t)^2\big).$$

Using (6), this implies that

$$\alpha_t\big(P_t^N\big) \le 2\lambda_{\max}\big(A + A'\big)\|P_t^N - P_t\|_F^2 \le 4\mu(A)H\big(P_t^N, P_t\big)$$

and by (5) we have

$$\frac{d}{dt}\langle\mathcal{M}(H)\rangle_t \le 4^2(N-1)^{-1}\operatorname{tr}\big(R + P_t^N S P_t^N\big)\operatorname{tr}\big(P_t^N\big)H\big(P_t^N, P_t\big).$$

This implies that

$$\mathcal{L}_t H\big(P_t^N, P_t\big) \le 4\mu(A)H\big(P_t^N, P_t\big) + (N-1)^{-1}\beta_t\big(P_t^N\big).$$

Arguing as in the end of the proof of Proposition 7.2, we also have

$$\sup_{t\ge 0}\mathbb{E}\big(\beta_t\big(P_t^N\big)^n\big) \le c(n) \quad\text{and}\quad \sup_{t\ge 0}\mathbb{E}\big(\gamma_t\big(P_t^N\big)^n\big) \le c'(n)$$

for any n s.t. $3n < 1 + \frac{N-1}{2r_1}$. Using Lemma 7.1, we conclude that

$$\mathbb{E}\big(\|P_t^N - P_t\|_F^{2n}\big)^{\frac{1}{n}} \le c(n)/N.$$

For odd numbers, we use Hölder's inequality

$$\mathbb{E}\big(\|P_t^N - P_t\|_F^{2n+1}\big) \le \mathbb{E}\big(\|P_t^N - P_t\|_F^{2n}\big)^{\frac{1}{2}}\mathbb{E}\big(\|P_t^N - P_t\|_F^{2(n+1)}\big)^{\frac{1}{2}}.$$

The end of the proof of the uniform estimates is now easily completed. The proof of the theorem is now complete. $\square$

8.2. *Uniform propagation of chaos.* This section is mainly concerned with the proof of the second uniform estimate stated in the right-hand side of (29). We set

$$Z_t = (P_t, X_t, \zeta_t, \xi_t)$$
$$\chi_t := \xi_t^1 - \zeta_t^1$$
$$H(Z_t) := \|\chi_t\|^2$$

and

$$q_t := P_t^N - P_t.$$

We have

$$d\chi_t = \left(A\chi_t - (P_t^N S\xi_t^1 - P_t S\zeta_t^1) + q_t SX_t\right)dt + q_t C' R_2^{-1/2} d(V_t - \overline{V}_t^1).$$

Observe that

$$P_t^N S\xi_t^1 - P_t S\zeta_t^1 = P_t^N SX_t + q_t S\zeta_t^1.$$

This implies that

(64) $$d\chi_t = \left((A - P_t^N S)\chi_t - q_t S(\zeta_t^1 - X_t)\right)dt + dM_t$$

with

$$dM_t = q_t C' R_2^{-1/2} d(V_t - \overline{V}_t^1).$$

The angle bracket matrix $\langle M \rangle_t = ((\langle M(k), M(l) \rangle_t)_{1 \le k, l \le r_1}$ is given by the formula

(65) $$\frac{d}{dt}\langle M \rangle_t = 2q_t Sq_t.$$

On the other hand, we have

$$d\|\chi_t\|^2 = 2\langle \chi_t, d\chi_t \rangle + \mathrm{tr}\left(\frac{d}{dt}\langle M \rangle_t\right)dt$$
$$= 2\left[\langle \chi_t, (A - P_t^N S)\chi_t \rangle - \langle \chi_t, q_t S(\zeta_t^1 - X_t) \rangle + \mathrm{tr}(q_t^2 S)\right] + 2\langle \chi_t, dM_t \rangle.$$

This yields

$$dH(Z_t) = \mathcal{L}_t H(Z_t)\,dt + d\mathcal{M}_t(H) \qquad \text{with } d\mathcal{M}_t(H) := 2\langle \chi_t, dM_t \rangle$$

with

$$\mathcal{L}_t H(Z_t) = 2\left[\langle \chi_t, (A - P_t^N S)\chi_t \rangle - \langle \chi_t, q_t S(\zeta_t^1 - X_t) \rangle + \mathrm{tr}(q_t^2 S)\right]$$

and

$$\frac{d}{dt}\langle \mathcal{M}(H) \rangle_t := 8\langle \chi_t, q_t Sq_t \chi_t \rangle.$$

Using (27), (6) and (62), we check that

$$\mathcal{L}_t H(Z_t) \le \mu(A)\|\chi_t\|^2 + |\mu(A)|^{-1}\|q_t S(\zeta_t^1 - X_t)\|^2 + 2\mu(S)\|q_t\|_F^2.$$

On the other hand, we have

$$\|q_t S(\zeta_t^1 - X_t)\| \le \|q_t\|_F \|S\|\|\zeta_t^1 - X_t\|.$$

This implies that

$$\mathcal{L}_t H(Z_t) \le \mu(A)H(Z_t) + \beta_t(Z_t)$$

with

$$\beta_t(Z_t) = \left[|\mu(A)|^{-1}\|S\|^2\|\zeta_t^1 - X_t\|^2 + 2\mu(S)\right]\|q_t\|_F^2.$$

In much the same way, we have

$$\frac{d}{dt}\langle \mathcal{M}(H)\rangle_t \le H(Z_t)\gamma_t(Z_t) \qquad \text{with } \gamma_t(Z_t) := 8\|S\|\|q_t\|_F^2.$$

By Theorem 8.1, and using the uniform estimates stated in (54) and in Proposition 7.2 for any $2n < 1 + \frac{N-1}{2r_1}$ we have

$$\sup_{t\ge 0} \mathbb{E}\big(\beta_t(Z_t)^n\big)^{\frac{1}{n}} \le c(n)/\sqrt{N} \quad \text{and} \quad \sup_{t\ge 0} \mathbb{E}\big(\gamma_t(Z_t)^n\big)^{\frac{1}{n}} \le c(n)/\sqrt{N}.$$

The proof of the uniform estimates in (29) is now a direct consequence of Lemma 7.1. This completes the proof of Theorem 3.6.

9. Proof of Proposition 5.3. The estimate (44) is a direct consequence of the perturbation Lemma 1.1 and the triangle inequality (28).

Now we come to the proof of (45). We set

$$\mathcal{X}_t = \overline{X}_t^\star - \overline{X}_t \quad \text{and} \quad Q_t = P_t^\star - P_t.$$

Arguing as in (64), we have

$$d\mathcal{X}_t = \big((A - P_t^\star S)\mathcal{X}_t - Q_t S(\overline{X}_t - X_t)\big)\,dt + dM_t$$

(66)

$$\text{with } \frac{d}{dt}\langle M\rangle_t = 2Q_t S Q_t.$$

This implies that

$$d\|\mathcal{X}_t\|^2 = 2\big[\langle \mathcal{X}_t, (A - P_t^\star S)\mathcal{X}_t\rangle - \langle \mathcal{X}_t, Q_t S(\overline{X}_t - X_t)\rangle + \mathrm{tr}(Q_t^2 S)/2\big] + d\mathcal{M}_t$$

with a real valued martingale with angle bracket

$$\frac{d}{dt}\langle \mathcal{M}\rangle_t = 8\langle \mathcal{X}_t, Q_t S Q_t \mathcal{X}_t\rangle \le 8\|S\|\|\mathcal{X}_t\|^2\|Q_t\|^2.$$

Using the same arguments as in the proof of (29) given in the end of Section 8, we conclude that

$$\mathbb{E}(\|\mathcal{X}_t\|^{2n})^{\frac{1}{n}} \le e^{\mu(A)t}\mathbb{E}(\|\mathcal{X}_0\|^{2n})^{\frac{1}{n}} + c\int_0^t e^{\mu(A)(t-s)}\|Q_s\|_F^2\,ds$$

for any n and some finite constant c. Using (44), we arrive at the estimate

$$\mathbb{E}\big(\|\overline{X}_t - \overline{X}_t^\star\|^{2n}\big)^{\frac{1}{2n}} \le e^{\mu(A)t/2}\mathbb{E}\big(\|\overline{X}_0 - \overline{X}_0^\star\|^{2n}\big)^{\frac{1}{2n}}$$

$$+ c\bigg[\int_0^t e^{\mu(A)(t-s)}e^{2\mu(A)s}\,ds\bigg]^{1/2}\|P_0 - P_0^\star\|_F$$

$$\le e^{\mu(A)t/2}\big[\mathbb{E}\big(\|\overline{X}_0 - \overline{X}_0^\star\|^{2n}\big)^{\frac{1}{2n}} + c'\|P_0 - P_0^\star\|_F\big]$$

for some finite constant c'. Also notice that

(23) and (45)

$$\Rightarrow \quad \mathbb{E}\big(\|\widehat{X}_t - \widehat{X}_t^\star\|^{2n}\big)^{\frac{1}{2n}} \le e^{\mu(A)t/2}\big[\mathbb{E}\big(\|\widehat{X}_0 - \widehat{X}_0^\star\|^{2n}\big)^{\frac{1}{2n}} + c\|P_0 - P_0^\star\|_F\big].$$

This completes the proof of Proposition 5.3.

10. Proof of Theorem 3.4.   We have

$$d(\overline{X}_t - X_t) = (A - P_t S)(\overline{X}_t - X_t) + R_1^{1/2}(d\overline{W}_t - W_t) + P_t C' R_2^{-1/2}(dV_t - \overline{V}_t).$$

This implies that

$$d\|\overline{X}_t - X_t\|^2 = \left[2\langle \overline{X}_t - X_t, (A - P_t S)(\overline{X}_t - X_t)\rangle + \mathrm{tr}(R_1 + P_t S P_t)\right]dt + dM_t$$

with some martingale M_t s.t.

$$\frac{d}{dt}\langle M_t\rangle \leq \|\overline{X}_t - X_t\|^2 \mathrm{tr}(R_1 + P_t S P_t).$$

We set $A_P := A - P S$. For any $u \in]0, 1]$, there exists some time horizon $\tau_u(P_0) \geq 0$ such that for any $t \geq \tau_u(P_0)$

$$\sup_{t \geq \tau_u(P_0)} \mu(A - P_t S) \leq (1 - u)\mu(A_P).$$

This yields

$$d\|\overline{X}_t - X_t\|^2 = \left[2(1 - u)\mu(A_P)\|\overline{X}_t - X_t\|^2 + c(u)\right]dt + dM_t$$

and

$$\frac{d}{dt}\langle M_t\rangle \leq \|\overline{X}_t - X_t\|^2 c(u)$$

for some finite constant $c(u)$ whose values only depend on u. By Lemma 7.1, we conclude that

$$\sup_{t \geq 0}\mathbb{E}(\|\overline{X}_t - X_t\|^n) < \infty$$

for any $n \geq 1$.

When $\overline{X}_t^\star$ is the steady state Kalman–Bucy diffusion, we have $P_t^\star = P$, for any $t \geq 0$. In this situation, (66) takes the form

$$d\mathcal{X}_t = (A_P \mathcal{X}_t - Q_t S(\overline{X}_t - X_t))dt + dM_t$$

with $Q_t = P - P_t$ and a martingale M_t with $\frac{d}{dt}\langle M\rangle_t = 2Q_t S Q_t$. This implies that

$$d\|\mathcal{X}_t\|^2 = 2\left[\langle \mathcal{X}_t, A_P \mathcal{X}_t\rangle - \langle \mathcal{X}_t, Q_t S(\overline{X}_t - X_t)\rangle + \mathrm{tr}(Q_t^2 S)/2\right] + d\mathcal{M}_t$$

with a real valued martingale with angle bracket

$$\frac{d}{dt}\langle \mathcal{M}\rangle_t = 8\langle \mathcal{X}_t, Q_t S Q_t \mathcal{X}_t\rangle \leq 8\|S\|\|\mathcal{X}_t\|^2\|Q_t\|^2.$$

Using the log-norm Lipschitz estimate (51) in Corollary 5.8, we have

$$\|Q_t S\|_F \leq \|S\|_F\|Q_t\|_F \leq c(u)\exp\left(2(1 - u)\mu(A_P)t\right)\|S\|_F\|P_0 - P\|_F,$$

$$\mathrm{tr}(Q_t^2 S) \vee \|Q_t S Q_t\|_F \leq c(u)^2 \exp\left(4(1 - u)\mu(A_P)t\right)\|S\|_F\|P_0 - P\|_F^2$$

for any $u \in]0, 1]$, and $t \geq 0$, and for some constant $c(u)$ that depends on u.

We set

$$Z_t = (\mathcal{X}_t, P_t), \qquad H(Z_t) = \|\mathcal{X}_t\|^2 \quad \text{and} \quad M_t(H) := M_t$$

and we consider the parameter $\alpha := \mu(A_P)$ and the function

$$\tau_t(Z_t) := c(u) \exp\big(2(1-u)\alpha t\big) \|S\|_F \|P_0 - P\|_F \|\overline{X}_t - X_t\|,$$

$$\beta_t(Z_t) := c(u)^2 \exp\big(4(1-u)\alpha t\big) \|S\|_F \|P_0 - P\|_F^2,$$

$$\gamma_t(Z_t) := \big(4\|S\|\big) c(u)^2 \exp\big(4(1-u)\alpha t\big) \|P_0 - P\|_F^2.$$

In this notation, we have

$$dH(Z_t) \leq \mathcal{L}_t H(Z_t)\, dt + M_t(H)$$

with

$$\mathcal{L}_t H(Z_t) \leq 2\alpha H(Z_t) + 2\tau_t(Z_t)\sqrt{H(Z_t)} + \beta_t(Z_t)$$

and

$$\frac{d}{dt}\langle M(H)\rangle_t \leq H(Z_t)\gamma_t(Z_t).$$

We also have

$$\mathbb{E}\big(\tau_t(Z_t)^n\big)^{\frac{1}{n}} \leq c_1(u) \exp\big(2(1-u)\alpha t\big)\|P_0 - P\|_F,$$

$$\mathbb{E}\big(\beta_t(Z_t)^n\big)^{\frac{1}{n}} \vee \mathbb{E}\big(\gamma_t(Z_t)^n\big)^{\frac{1}{n}} \leq c_2(u) \exp\big(4(1-u)\alpha t\big)\|P_0 - P\|_F^2$$

for any $n \geq 1$, for some constants $c_i(u)$ that depends on u, with $i \in \{1, 2\}$. Using Lemma 7.1, we check that for any $u \in\,]0, 1]$ and any $t \geq 0$ we have

$$\mathbb{E}\big(\|\mathcal{X}_t\|^{2n}\big)^{\frac{1}{n}} \leq e^{\alpha t}\mathbb{E}\big(\|\mathcal{X}_0\|^{2n}\big)^{\frac{1}{n}} + c(u)e^{(1-u)\alpha t}\|P_0 - P\|_F^2$$

for some constants $c(u)$ that depends on u. Using (23), we also check that

$$\mathbb{E}[\|\widehat{X}_t - \widehat{X}_t^\star\|^{2n}]^{\frac{1}{2n}} \leq c(u)e^{(1-u)\alpha t/2}\big(\mathbb{E}(\|\widehat{X}_0 - \widehat{X}_0^\star\|^{2n})^{\frac{1}{2n}} + \|P_0 - P\|_F\big),$$

where $\widehat{X}_t^\star$ is the steady state Kalman–Bucy filter. This completes the proof of the proposition.

11. Proof of Theorem 3.5. To simplify the presentation, we assume that $P_0^\star = P$.

We further assume that the algebraic Riccati equation (21) has a positive definite fixed point P (so that P is invertible). We also assume that $\mu(A - PS) < 0$.

Let $(\overline{X}_t, \overline{X}_t^\star)$ be a couple of Kalman–Bucy diffusions (10) starting from two possibly different *Gaussian* random variables with covariance matrices (P_0, P). We recall that

$$\widehat{X}_t = \mathbb{E}(\overline{X}_t | \mathcal{F}_t) \quad \text{and} \quad \widehat{X}_t^\star = \mathbb{E}(\overline{X}_t^\star | \mathcal{F}_t)$$

satisfy the Kalman–Bucy filter recursion (8) associated with the covariance matri-
ces (P_t, P), with P_t given by the Riccati equation (9).

Let $(\eta_t, \eta_t^\star)$ be the (Gaussian) conditional distributions of $(\overline{X}_t, \overline{X}_t^\star)$ given the
σ-field $\mathcal{F}_t$ generated by the observation process. The conditional Boltzmann–
Kullback Leibler relative entropy of η_t w.r.t. $\eta_t^\star$ is given by the formula:

$$\mathrm{Ent}(\eta_t | \eta_t^\star) = -\frac{1}{2}\big(\mathrm{tr}(I - P^{-1}P_t) + \log\det(P_t P^{-1}) - \langle(\widehat{X}_t^\star - \widehat{X}_t), P^{-1}(\widehat{X}_t^\star - \widehat{X}_t)\rangle\big).$$

To estimate the logarithm of the determinant of the matrices $P_t P^{-1}$ as $t \uparrow \infty$, we
use the following technical lemma.

LEMMA 11.1. *For any $(r \times r)$-matrix A, we have*

$$\|A\|_F < \frac{1}{2} \quad \Longrightarrow \quad \big|\log\det(I - A)\big| \leq \frac{3}{2}\|A\|_F.$$

PROOF. By (5), for any $n \geq 1$ we have

$$\big|\mathrm{tr}(A^n)\big| \leq \|A\|_F^n.$$

Using the well-known trace formulae,

$$\log\det(I - A) = \mathrm{tr}(\log(I - A)) = -\sum_{n \geq 1} n^{-1}\,\mathrm{tr}(A^n)$$

we conclude that

$$\big|\log\det(I - A)\big| \leq -\log(1 - \|A\|_F).$$

The last assertion comes from the inequality

$$0 \leq -\log(1 - u) \leq u + \frac{1}{2}\frac{u^2}{1 - u} = u\left(1 + \frac{1}{2}\frac{u}{1 - u}\right) \leq 3u/2$$

which is valid for any $u \in [0, 1/2[$.
This completes the proof of the lemma. □

For any $u \in]0, 1]$, and $t \geq 0$, there exists some $c(u)$ that depends on u s.t.

$$\|I - P^{-1}P_t\|_F \leq \|P^{-1}\|_F \|P - P_t\|_F$$
$$\leq c(u)\exp\big(2(1 - u)\mu(A - PS)t\big)\|P_0 - P\|_F.$$

Applying Lemma 11.1 to $A = (P - P_t)P^{-1}$, there exists some t_0 that depends
on (P_0, P) and some finite constant c such that

$$\big|\log\det(P_t P^{-1})\big| \leq \frac{3}{2}\|P - P_t\|_F \|P^{-1}\|_F$$

for any $t \geq t_0$. On the other hand, we have

$$\big|\mathrm{tr}(I - P^{-1}P_t)\big| = \big|\mathrm{tr}(P^{-1}(P - P_t))\big| \leq \|P^{-1}\|_F \|P - P_t\|_F.$$

Finally, we notice that

$$\left|\langle(\widehat{X}_t^\star - \widehat{X}_t), P^{-1}(\widehat{X}_t^\star - \widehat{X}_t)\rangle\right| \leq \|P^{-1}\|_2 \|\widehat{X}_t^\star - \widehat{X}_t\|^2.$$

We conclude that

$$0 \leq \text{Ent}(\eta_t | \eta_t^\star) \leq \frac{3}{2} \|P^{-1}\|_F \left(\|\widehat{X}_t^\star - \widehat{X}_t\|^2 + \|P - P_t\|_F\right)$$

for any $t \geq t_0$. The end of the proof of (25) is now clear. This completes the proof of Theorem 3.5.

12. Conclusion. We have designed and analyzed a new class of conditional nonlinear diffusion processes arising in filtering theory. In contrast with conventional nonlinear Markov models, these Kalman–Bucy diffusion type models depends on the conditional covariance matrices of the internal random states. To analyze the stability properties of these models, a series of functional contraction inequalities have been developed w.r.t. the Wasserstein distance, Frobenius norms on random matrices and relative entropy criteria.

In this framework, the traditional Kalman–Bucy filter resumes to the time evolution of the conditional averages of these nonlinear diffusions. The stability properties of the filter are now deduced directly from the ones of the nonlinear model.

The second important contribution of the article concerns the long-time behaviour and the refined convergence analysis of ensemble Kalman filters. The EnKF is interpreted as a natural mean-field particle approximation of nonlinear Kalman–Bucy diffusions. The performance of the EnKF is measured in terms of uniform $\mathbb{L}_n$-mean error estimates and uniform propagation of chaos properties w.r.t. the time horizon.

We end this article with an avenue of open research problems.

The first project is to extend the analysis to nonlinear diffusions with an interacting function that depends on the covariance matrices of the random states. A toy model of that form is given by the one-dimensional diffusion

$$dX_t = \text{Var}(X_t) \qquad (dW_t - X_t\, dt),$$

where W_t stands for a Brownian motion. It is readily check that this nonlinear diffusion is well-posed. In addition, the variance $P_t = \text{Var}(X_t) = P_0/(1 + P_0 t)$ satisfies the Riccati equation:

$$\frac{d}{dt} P_t = -P_t^2 \quad \text{and} \quad X_t = (X_0 + P_0 W_t)/(1 + P_0 t).$$

Besides the fact that the convergence rate of P_t towards 0 is not exponential, following the stochastic analysis developed in the present article several uniform propagation of chaos properties can be developed for this toy model is a rather simple way. The extension of these results to more general multidimensional diffusions with drift remains an open research question.

The second open question is to analyze the long-time behaviour of the extended EnKF commonly used in nonlinear filtering theory and signal processing. In this connection, we emphasize that these extended Kalman–Bucy filters require computing the derivative of the drift coefficient w.r.t. the state variable, which is a complex task and subject to potential error in high dimension. For nonlinear systems, the EnFK is commonly used in data assimilation literature with nonlinear drift coefficients. The analysis of the long-time behaviour of these derivative-free EnKF filters also remains an open important research problem.

Last but not least, another important problem is clearly to develop uniform propagations of chaos properties of the EnKF in discrete time settings. Last, but not least a series of research projects can be developed around the fluctuations and the large deviations of this new class of mean-field-type particle models.

APPENDIX

Proof of Lemma 1.1. The first assertion is a direct consequence of the inequality:

$$\left\| \mathcal{E}_{s,t}(A+B) \right\|_2 \le \exp\left(\int_s^t \mu(A_u)\, du + \int_s^t \|B_u\|_2\, du \right).$$

The above estimate is a direct consequence of the matrix log-norm inequality

$$\mu(A_t + B_t) < \mu(A_t) + \mu(B_t) \quad \text{and the fact that } \mu(B_t) \le \|B_t\|_2.$$

This completes the proof of the first assertion. To check the second assertion, we observe that

$$\partial_t \mathcal{E}_{s,t}(A+B) = \left(\partial_t \mathcal{E}_t(A+B) \right) \mathcal{E}_s(A+B)^{-1}$$
$$= A_t \mathcal{E}_{s,t}(A+B) + B_t \mathcal{E}_{s,t}(A+B).$$

This implies that

$$\mathcal{E}_{s,t}(A+B) = \mathcal{E}_{s,t}(A) + \int_s^t \mathcal{E}_{u,t}(A) B_u \mathcal{E}_{s,u}(A+B)\, du$$

for any $s \le t$ from which we prove that

$$e^{\beta_A(t-s)} \left\| \mathcal{E}_{s,t}(A+B) \right\| \le \alpha_A + \alpha_A \int_s^t \|B_u\| e^{\beta_A(u-s)} \left\| \mathcal{E}_{s,u}(A+B) \right\|\, du.$$

By Grönwall's lemma, this implies that

$$e^{\beta_A(t-s)} \left\| \mathcal{E}_{s,t}(A+B) \right\| \le \alpha_A \exp\left[\int_s^t \alpha_A \|B_u\|\, du \right].$$

This completes the proof of the lemma.

Proof of Lemma 5.4. This section is mainly concerned with the proof of Lemma 5.4. Let P_1 and P_2 be the covariance matrices of some r-valued random variables Z_1 and Z_2. Let $\overline{Z}_i$ be an independent copie of Z_i, with $i = 1, 2$, and set

$$Z_{1,2} := Z_1 - Z_2 \quad \text{and} \quad \overline{Z}_{1,2} := \overline{Z}_1 - \overline{Z}_2.$$

Observe that

$$Z_1 - \overline{Z}_1 = (Z_2 - \overline{Z}_2) + (Z_{1,2} - \overline{Z}_{1,2})$$

and for any $i = 1, 2$ we have

$$2P_i = \mathbb{E}(\mathcal{Z}_i) \quad \text{with the random matrix } \mathcal{Z}_i := (Z_i - \overline{Z}_i)(Z_i - \overline{Z}_i)'.$$

This yields the decomposition

$$\begin{aligned}
\mathcal{Z}_1 - \mathcal{Z}_2 = {} & (Z_2 - \overline{Z}_2)(Z_{1,2} - \overline{Z}_{1,2})' \\
& + (Z_{1,2} - \overline{Z}_{1,2})(Z_2 - \overline{Z}_2)' \\
& + (Z_{1,2} - \overline{Z}_{1,2})(Z_{1,2} - \overline{Z}_{1,2})'.
\end{aligned}$$

This shows that

$$\begin{aligned}
2(P_1 - P_2) = {} & \mathbb{E}\big((Z_2 - \overline{Z}_2)(Z_{1,2} - \overline{Z}_{1,2})'\big) \\
& + \mathbb{E}\big((Z_{1,2} - \overline{Z}_{1,2})(Z_2 - \overline{Z}_2)'\big) \\
& + \mathbb{E}\big((Z_{1,2} - \overline{Z}_{1,2})(Z_{1,2} - \overline{Z}_{1,2})'\big)
\end{aligned}$$

from which we prove the trace formula

$$\begin{aligned}
2\operatorname{tr}(P_1 - P_2) = {} & 2\mathbb{E}\big(\langle Z_{1,2} - \overline{Z}_{1,2}, Z_2 - \overline{Z}_2 \rangle\big) \\
& + \mathbb{E}\big(\|Z_{1,2} - \overline{Z}_{1,2}\|^2\big).
\end{aligned}$$

By the Cauchy–Schwarz inequality, we find that

$$\begin{aligned}
2\big|\operatorname{tr}(P_1 - P_2)\big| \leq {} & 8\mathbb{E}\big(\|Z_{1,2}\|^2\big)^{1/2}\mathbb{E}\big(\|Z_2\|^2\big)^{1/2} \\
& + 4\mathbb{E}\big(\|Z_{1,2}\|^2\big).
\end{aligned}$$

This yields

$$\begin{aligned}
\frac{1}{2}\big|\operatorname{tr}(P_1 - P_2)\big| \leq {} & 2\mathbb{E}\big(\|Z_1 - Z_2\|^2\big)^{1/2}\mathbb{E}\big(\|Z_2\|^2\big)^{1/2} \\
& + \mathbb{E}\big(\|Z_1 - Z_2\|^2\big)
\end{aligned}$$

from which we find that

$$\frac{1}{2}\big|\operatorname{tr}(\mathcal{P}_{\eta_0}) - \operatorname{tr}(\mathcal{P}_{\eta_0^\star})\big| \leq 2\mathbb{W}_2(\eta_0, \eta_0^\star)\big\|\eta_0^\star(e_2)\big\|^{1/2} + \mathbb{W}_2(\eta_0, \eta_0^\star)^2.$$

In much the same way, we have

$$4\|P_1 - P_2\|_F^2 = \sum_{1 \le i,j \le r} \left[\mathbb{E}((Z_{1,2} - \overline{Z}_{1,2})(i)(Z_{1,2} - \overline{Z}_{1,2})(j)) \right.$$
$$+ \mathbb{E}((Z_2 - \overline{Z}_2)(i)(Z_{1,2} - \overline{Z}_{1,2})(j))$$
$$\left. + \mathbb{E}((Z_2 - \overline{Z}_2)(j)(Z_{1,2} - \overline{Z}_{1,2})(i)) \right]^2.$$

This implies that

$$2\|P_1 - P_2\|_F^2 = \sum_{1 \le i,j \le r} \left[\{ \mathbb{E}((Z_{1,2} - \overline{Z}_{1,2})(i)(Z_{1,2} - \overline{Z}_{1,2})(j)) \}^2 \right.$$
$$+ \{ \mathbb{E}((Z_2 - \overline{Z}_2)(i)(Z_{1,2} - \overline{Z}_{1,2})(j)) \}^2$$
$$\left. + \{ \mathbb{E}((Z_2 - \overline{Z}_2)(j)(Z_{1,2} - \overline{Z}_{1,2})(i)) \}^2 \right].$$

Using the Cauchy–Schwarz inequality, we prove that

$$2\|P_1 - P_2\|_F^2 \le 2 \sum_{1 \le i \le r} \mathbb{E}((Z_{1,2} - \overline{Z}_{1,2})(i)^2) \sum_{1 \le j \le r} \mathbb{E}((Z_2 - \overline{Z}_2)(j)^2)$$
$$+ \left[\sum_{1 \le i \le r} \mathbb{E}((Z_{1,2} - \overline{Z}_{1,2})(i)^2) \right]^2.$$

This implies that

$$\sqrt{2}\|P_1 - P_2\|_F \le 4\sqrt{2} \mathbb{E}(\|Z_{1,2}\|^2)^{1/2} \mathbb{E}(\|Z_2\|^2)^{1/2} + 4\mathbb{E}(\|Z_{1,2}\|^2).$$

We find that

$$\|\mathcal{P}_{\eta_0} - \mathcal{P}_{\eta_0^\star}\|_F \le \mathbb{W}_2(\eta_0, \eta_0^\star) \|\eta_0^\star(e_2)\|^{1/2} + \frac{1}{\sqrt{2}} \mathbb{W}_2(\eta_0, \eta_0^\star)^2.$$

This completes the proof of the lemma.

DIVERGENCE REGIONS—2D OBSERVERS

We illustrate the spectral analysis discussion given in Section 4.1 with 2-dimensional partially observed filtering problems associated with the parameters

$$(r_1, r_2) = (2, 1), \qquad C = [1, 0], \qquad (R_1, R_2) = (\mathrm{Id}, 1)$$

and some unstable drift matrix A with a saddle equilibrium, that is,

$$-\det(A) = A_2 \wedge A_1 > 0$$

with the column vectors $A_1 = \binom{A_{1,1}}{A_{1,2}}$ and $A_2 = \binom{A_{2,1}}{A_{2,2}}$. In the above display, $A_2 \wedge A_1$ stands for the cross product of the vectors A_2 and A_1. Whenever $A_{1,2} \ne 0$, the

system is observable and controllable; thus there exists some unique steady state P and $\text{Re}(\lambda_{\max}(\overline{A})) < 0$, or equivalently

$$\text{tr}(A) < P_{1,1} \quad \text{and} \quad A_2 \wedge P_1 > A_2 \wedge A_1.$$

The set of admissible fluctuation matrices Q (i.e., s.t. $P + Q \geq 0$) is defined by

$$\mathcal{Q}(P) := \left\{ Q \in \mathbb{S}_{r_1} : \left[P_1 - \frac{1}{2} Q_1 \right] \wedge Q_2 + \left[P_2 + \frac{1}{2} Q_2 \right] \wedge Q_1 < P_1 \wedge P_2 \right\}$$

$$\cap \left\{ Q \in \mathbb{S}_{r_1} : Q_{1,1} \in [-P_{1,1}, \infty[\text{ and } Q_{2,2} \in [-P_{2,2}, \infty[\right\}.$$

Given some $Q \in \mathcal{Q}(P)$ several cases can happen. In the most favorable case, we have

$$\left(P_{1,1} - \text{tr}(A) \right) > -Q_{1,1} \quad \text{and} \quad \det([(A_1 - P_1) - Q_1, A_2]) > 0$$

$$\iff \quad \text{Re}(\lambda_{\max}(\overline{A} - QS)) < 0.$$

The determinant condition is equivalent to

$$Q_1 \wedge A_2 < A_2 \wedge P_1 - A_2 \wedge A_1.$$

To be more precise, we have two negative eigenvalues when $\text{tr}(\overline{A} - QS)^2 > 4\det(\overline{A} - QS)$, otherwise we have a spiral phase portrait with complex eigenvalues with negative real parts. In both cases, the matrix $\overline{A} - QS$ remains stable. Skipping the discussion on borderline cases, the other situations that may arise is that

$$Q_1 \wedge A_2 > A_2 \wedge P_1 - A_2 \wedge A_1 \quad \text{or} \quad 0 < P_{1,1} - \text{tr}(A) < -Q_{1,1}.$$

When the left-hand side condition is met, the eigenvalues have opposite signs and the stochastic observer experience a catastrophic divergence in the direction of the eigenvector associated with the positive one. When the right-hand side condition is met, both eigenvalues are negative real numbers if $\text{tr}(\overline{A} - QS)^2 > 4\det(\overline{A} - QS)$, otherwise they are both complex with the same negative real part. In both situations, the observer diverges. The divergence set (34) is given by

$$\mathcal{Q}_{\text{div}} = \left\{ Q \in \mathcal{Q}(P) : Q_1 \wedge A_2 > A_2 \wedge P_1 - A_2 \wedge A_1 \right\}$$

$$\cup \left\{ Q \in \mathcal{Q}(P) : Q_{1,1} < -(P_{1,1} - \text{tr}(A)) \right\}.$$

For instance, for

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \implies P \simeq \begin{bmatrix} 8.7 & 14.5 \\ 14.5 & 30 \end{bmatrix} \quad \text{and} \quad \overline{A} \simeq \begin{bmatrix} -7.4 & 2 \\ -13.5 & 3 \end{bmatrix}.$$

Notice that the system is locally ill-conditioned as $\lambda_{\max}(\overline{A}) = 5.4913 > 0$. The set of admissible fluctuations is given by

$$\mathcal{Q}(P) := \left\{ Q \in \mathbb{S}_{r_1} : Q_{1,2} Q_{2,1} + 14.5(Q_{1,2} + Q_{2,1}) \right.$$

$$< 52.2 + 8.7 Q_{2,2} + 30 Q_{1,1} + Q_{1,1} Q_{2,2} \right\}$$

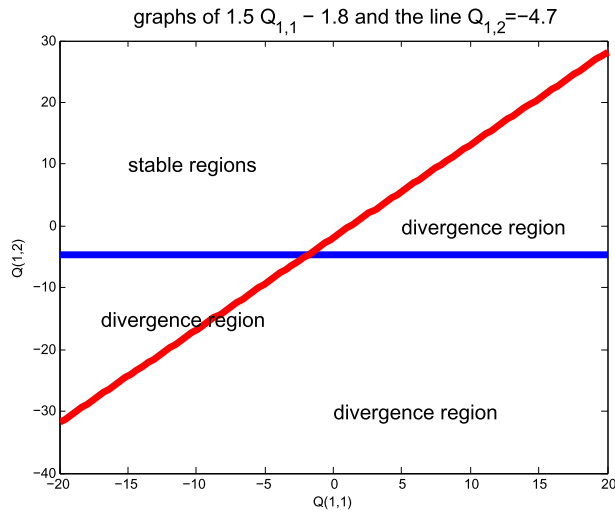$$\cap \left\{ Q \in \mathbb{S}_{r_1} : Q_{1,1} \in [-8.7, \infty[\text{ and } Q_{2,2} \in [-30, \infty[\right\}.$$

FIG. 1.   *Fluctuation-divergence effects.*

In this situation, we have

$$\mathcal{Q}_{\text{div}} \simeq \{Q \in \mathcal{Q}(P) : Q_{1,2} < 1.5Q_{1,1} - 1.8\}$$
$$\cup \{Q \in \mathcal{Q}(P) : Q_{1,1} \in \,]{-8.7}, -4.7[\}.$$

In this situation, the stable subset of $\mathcal{Q}(P)$ is given by

$$Q \in \mathcal{Q}(P) : Q_{1,2} > 1.5Q_{1,1} - 1.8 \quad \text{and} \quad Q_{1,1} \in \,]{-4.7}, \infty[.$$

These convergence and divergence sets without the admissible conditions are illustrated in Figure 1.

The trace of the divergence domain with diagonal matrices (i.e., $Q_{1,2} = 0 = Q_{2,1}$) resume to diagonal matrices s.t.

$$Q_{1,1} \in \,]1.2, \infty[\quad \text{and} \quad 0 < 52.2 + 8.7Q_{2,2} + 30Q_{1,1} + Q_{1,1}Q_{2,2}.$$

The trace with the stable domain is

$$Q_{1,1} \in \,]{-4.7}, 1.2[\quad \text{and} \quad 0 < 52.2 + 8.7Q_{2,2} + 30Q_{1,1} + Q_{1,1}Q_{2,2}.$$

An illustration of this set is given in Figure 2.

The fluctuation/divergence effects we can expect when the observer is driven by fluctuations entering into the divergence domain are illustrated in Figure 3. A series of realization of the stochastic observer driven by fluctuation matrices in the stable domain are presented in Figure 4(a), (b); the ones driven by fluctuation matrices in the divergence set are presented in Figure 4(c), (d). The entries $Q_{2,2}$ are not seen by the observer so we assume that $Q_{2,2} = 0$.
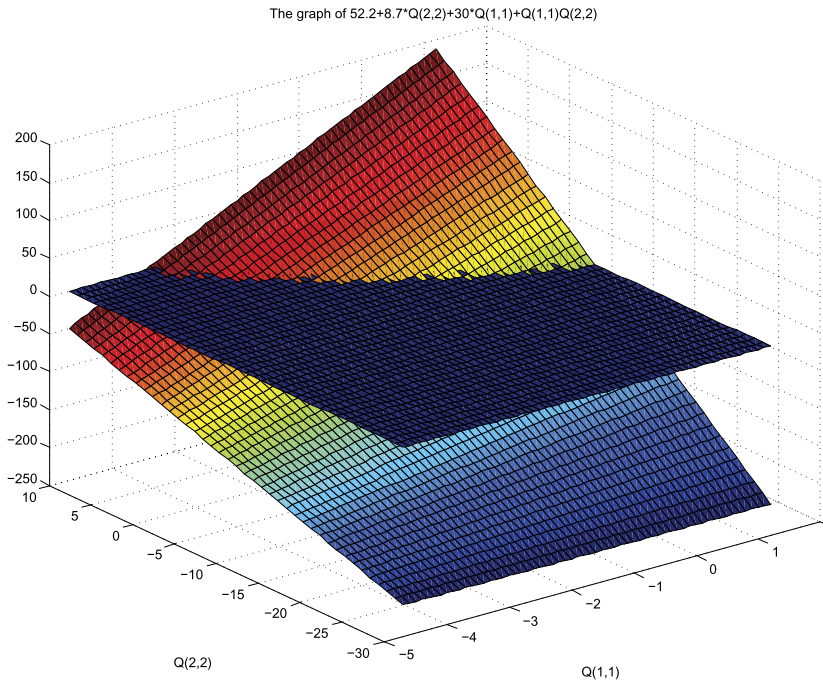
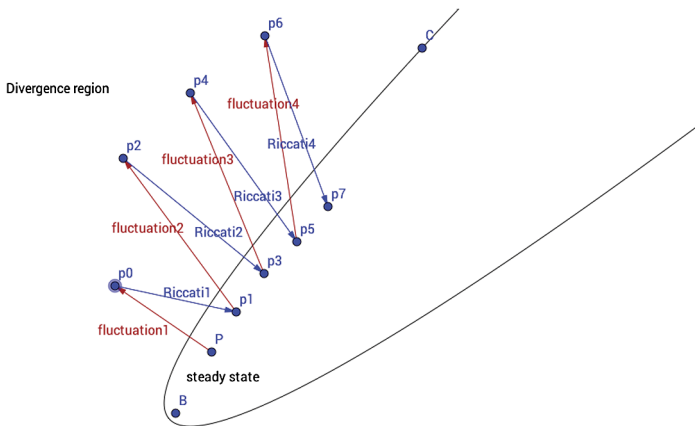FIG. 2.   *Trace of the stable domain with diagonal matrices.*

FIG. 3.   *Fluctuation-divergence effects.*

(a) Observer [red] /Steady state [blue]
$Q_{1,1} = 1.1$ with $t = 100$

(b) Observer [red] /Steady state [blue]
$Q_{1,1} = 1.2$ with $t = 100$

(c) Observer [red] /Steady state [blue]
$Q_{1,1} = 1.3$ with $t = 125$

(d) Observer [red] /Steady state [blue]
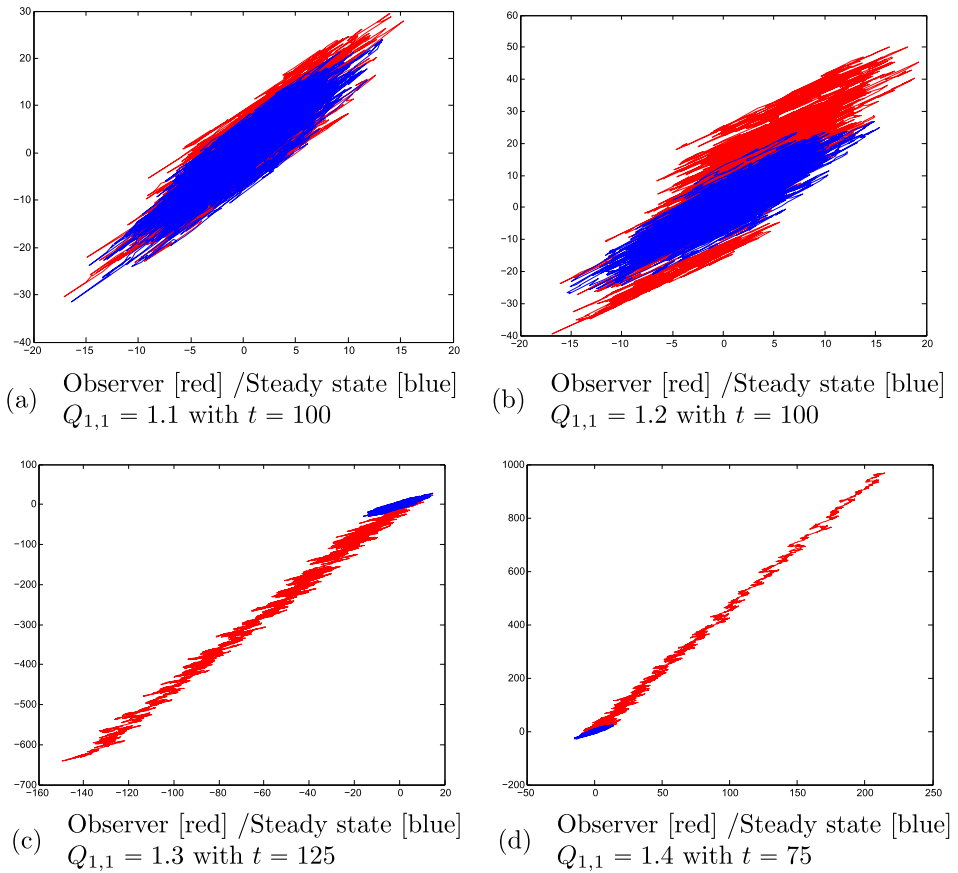$Q_{1,1} = 1.4$ with $t = 75$

FIG. 4. *Stochastic observer and steady state in the stable and in the divergence domains.*

REFERENCES

[1] ABOU-KANDIL, H., FREILING, G., IONESCU, V. and JANK, G. (2003). *Matrix Riccati Equations in Control and Systems Theory. Systems & Control*: *Foundations & Applications*. Birkhäuser, Basel. MR1997753

[2] ALLEN, J. I., EKNES, M. and EVENSEN, G. (2002). An Ensemble Kalman filter with a complex marine ecosystem model: Hindcasting phytoplankton in the Cretan Sea. *Ann. Geophys.* **20** 1–13.

[3] ANDERSON, B. D. O. (1971). Stability properties of Kalman–Bucy filters. *J. Franklin Inst.* **291** 137–144. MR0281515

[4] ANDERSON, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129** 2884–2903.

[5] ANDERSON, J. L. (2003). A local least squares framework for ensemble filtering. *Mon. Weather Rev.* **131** 634–642.

[6] BARAS, J. S., BENSOUSSAN, A. and JAMES, M. R. (1988). Dynamic observers as asymptotic limits of recursive filters: Special cases. *SIAM J. Appl. Math.* **48** 1147–1158. MR0960476

[7] BERNSTEIN, D. S. (1988). Inequalities for the trace of matrix exponentials. *SIAM J. Matrix Anal. Appl.* **9** 156–158. MR0938494

[8] BERRY, T. and HARLIM, J. (2014). Linear theory for filtering nonlinear multiscale systems with model error. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **470** 20140168, 25. MR3209215

[9] BISHOP, A. N. and DEL MORAL, P. (2016). On the stability of Kalman–Bucy diffusion processes. Available at arXiv:1610.04686.

[10] BITTANTI, S., LAUB, A. J. and WILLEMS, J. C. (1991). *The Riccati Equation. Communications and Control Engineering Series*. Springer, Berlin.

[11] BOLLEY, F., GUILLIN, A. and MALRIEU, F. (2010). Trend to equilibrium and particle approximation for a weakly selfconsistent Vlasov–Fokker–Planck equation. *M2AN Math. Model. Numer. Anal.* **44** 867–884. MR2731396

[12] BOUGEROL, P. and FAKHFAKH, S. (1996). A note on the stability of the Kalman–Bucy filter with randomly time-varying parameters. *J. Math. Sci.* **78** 28–33. MR1381032

[13] BROCKETT, R. W. (2015). *Finite Dimensional Linear Systems. Classics in Applied Mathematics* **74**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1970 original. MR3486166

[14] BUCY, R. S. (1965). Nonlinear filtering theory. *IEEE Trans. Automat. Control* **10** 198–198.

[15] BUCY, R. S. (1967). Global theory of the Riccati equation. *J. Comput. System Sci.* **1** 349–361. MR0229916

[16] BURGERS, G., VAN LEEUWEN, P. J. and EVENSEN, G. (1998). Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126** 1719–1724.

[17] CATTIAUX, P., GUILLIN, A. and MALRIEU, F. (2008). Probabilistic approach for granular media equations in the non-uniformly convex case. *Probab. Theory Related Fields* **140** 19–40. MR2357669

[18] COPPEL, W. A. (1978). *Dichotomies in Stability Theory. Lecture Notes in Mathematics* **629**. Springer, Berlin–New York. MR0481196

[19] COSTA, E. F. (2008). On the stability of the recursive Kalman filter for linear time-invariant systems. In *Proceedings of the IEEE* 2008 *American Control Conference* 1286–1291, Seattle, WA.

[20] DEL MORAL, P. (2004). *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications. Probability and Its Applications* (*New York*). Springer, New York. MR2044973

[21] DEL MORAL, P. (2013). *Mean Field Simulation for Monte Carlo Integration. Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

[22] DEL MORAL, P. and TUGAUT, J. (2013). Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions. https://hal.archives-ouvertes.fr/hal-00798813.

[23] DYDA, B. and TUGAUT, J. (2017). Exponential rate of convergence independent of the dimension in a mean-field system of particles. *Probab. Math. Statist.* **37** 145–161. MR3652205

[24] EINICKE, G. (2012). Continuous-time minimum-variance filtering. In *Smoothing, Filtering and Prediction—Estimating the Past, Present and Future*. InTech, London.

[25] EKNES, M. and EVENSEN, G. (2002). An Ensemble Kalman filter with a 1-D marine ecosystem model. *JMS* **36** 75–100.

[26] EVENSEN, G. (1994). Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99** 10143–10162.

[27] EVENSEN, G., HOVE, J., MEISINGSET, H. C., REISO, E. and SEIM, K. S. (2007). Using the EnKF for assisted history matching of a North Sea reservoir model. *SPE* **106184**.

[28] GIVENS, C. R. and SHORTT, R. M. (1984). A class of Wasserstein metrics for probability distributions. *Michigan Math. J.* **31** 231–240. MR0752258

[29] GOTTWALD, G. and MAJDA, A. J. (2013). A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Process. Geophys.* **20** 705–712.

[30] HARLIM, J. and HUNT, B. (2005). Local Ensemble Transform Kalman Filter: An Efficient Scheme for Assimilating Atmospheric Data. Preprint. Available at https://www.atmos.umd.edu/ekalnay/pubs/harlim_hunt05.pdf.

[31] HARLIM, J. and MAJDA, A. J. (2010). Catastrophic filter divergence in filtering nonlinear dissipative systems. *Commun. Math. Sci.* **8** 27–43. MR2655912

[32] HOUTEKAMER, P. and MITCHELL, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **126** 796–811.

[33] ILCHMANN, A., OWENS, D. H. and PRÄTZEL-WOLTERS, D. (1987). Sufficient conditions for stability of linear time-varying systems. *Systems Control Lett.* **9** 157–163. MR0906235

[34] JOHNS, C. J. and MANDEL, J. (2008). A two-stage ensemble Kalman filter for smooth data assimilation. *Environ. Ecol. Stat.* **15** 101–110. MR2412683

[35] KALNAY, E. (2003). *Atmospheric Modeling*, *Data Assimilation*, *and Predictability*. Cambridge Univ. Press, Cambridge.

[36] KELLY, D., MAJDA, A. J. and TONG, X. T. (2015). Concrete ensemble Kalman filters with rigorous catastrophic filter divergence. *Proc. Natl. Acad. Sci. USA* **112** 10589–10594.

[37] KRAUSE, J. M. and KUMAR, K. S. P. (1986). An alternate stability analysis framework for adaptive control. *Systems Control Lett.* **7** 19–24. MR0833061

[38] KREISSELMEIER, G. (1985). An approach to stable indirect adaptive control. *Automatica J. IFAC* **21** 425–431. MR0798187

[39] KUSHNER, H. J. (1964). On the differential equations satisfied by conditional probablitity densities of Markov processes, with applications. *J. Soc. Indust. Appl. Math. Ser. A Control* **2** 106–119. MR0180407

[40] KWAKERNAAK, H. and SIVAN, R. (1972). *Linear Optimal Control Systems*. Wiley, New York. MR0406607

[41] LANCASTER, P. and RODMAN, L. (1995). *Algebraic Riccati Equations*. *Oxford Science Publications*. The Clarendon Press, New York. MR1367089

[42] LAWSON, J. and LIM, Y. (2007). A Birkhoff contraction formula with applications to Riccati equations. *SIAM J. Control Optim.* **46** 930–951. MR2338433

[43] LE GLAND, F., MONBET, V. and TRAN, V.-D. (2011). Large sample asymptotics for the ensemble Kalman filter. In *The Oxford Handbook of Nonlinear Filtering* 598–631. Oxford Univ. Press, Oxford. MR2884610

[44] LISAETER, K. A., ROSANOVA, J. and EVENSEN, G. (2003). Assimilation of ice concentration in a coupled ice-ocean model, using the Ensemble Kalman filter. *Ocean Dyn.* **53** 368–388.

[45] LUENBERGER, D. G. (1966). Observers for multivariable systems. *IEEE Trans. Automat. Control* **AC-11** 190–199.

[46] MAJDA, A. J. and HARLIM, J. (2012). *Filtering Complex Turbulent Systems*. Cambridge Univ. Press, Cambridge. MR2934167

[47] MALRIEU, F. (2001). Logarithmic Sobolev inequalities for some nonlinear PDE's. *Stochastic Process. Appl.* **95** 109–132. MR1847094

[48] MANDEL, J., COBB, L. and BEEZLEY, J. D. (2011). On the convergence of the ensemble Kalman filter. *Appl. Math.* **56** 533–541. MR2886236

[49] MCKEAN, H. P. JR. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595

[50] MÉLÉARD, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean–Vlasov and Boltzmann models. In *Probabilistic Models for Nonlinear Partial Differential Equations* (*Montecatini Terme*, 1995). *Lecture Notes in Math.* **1627** 42–95. Springer, Berlin. MR1431299

[51] MOLER, C. and VAN LOAN, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Rev.* **45** 3–49. MR1981253

[52] NAEVDAL, G., JOHNSEN, L. M., AANONSEN, S. I. and VEFRING, E. H. (2005). Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE Journal* **10** 66–74.

[53] OCONE, D. and PARDOUX, E. (1996). Asymptotic stability of the optimal filter with respect to its initial condition. *SIAM J. Control Optim.* **34** 226–243. MR1372912

[54] OTT, E., HUNT, B. R., SZUNYOGH, I., ZIMIN, A. V., KOSTELICH, E. J., CORAZZA, M., KALNAY, E., PATIL, D. and YORKE, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, *Ser. A Dyn. Meteorol. Oceanogr.* **56** 415–428.

[55] POUBELLE, M.-A., PETERSEN, I. R., GEVERS, M. R. and BITMEAD, R. R. (1986). A miscellany of results on an equation of Count J. F. Riccati. *IEEE Trans*. *Automat*. *Control* **31** 651–654. MR0844922

[56] ROSENBROCK, H. H. (1963). The stability of linear time-dependent control systems. *J. Electron*. *Control* **15** 73–80. MR0159085

[57] SEILER, A., EVENSEN, G., SKJERVHEIM, J.-A., HOVE, J. and VABØ, J. G. (2011). Using the ensemble Kalman filter for history matching and uncertainty quantification of complex reservoir models. In *Large-Scale Inverse Problems and Quantification of Uncertainty*. *Wiley Ser. Comput*. *Stat*. 247–271. Wiley, Chichester. MR2856659

[58] SKJERVHEIM, J.-A., EVENSEN, G., AANONSEN, S. I., RUUD, B. O. and JOHANSEN, T. A. (2005). Incorporating 4D seismic data in reservoir simulation models using ensemble Kalman filter. *SPE* **12** 95789.

[59] SNYDER, D. A. (2016). On the Relation of Schatten Norms and the Thompson Metric. Available at ArXiv:1608.03301.

[60] SONTAG, E. D. (1998). *Mathematical Control Theory*: *Deterministic Finite-Dimensional Systems*, 2nd ed. *Texts in Applied Mathematics* **6**. Springer, New York. MR1640001

[61] STRATONOVIČ, R. L. (1960). Conditional Markov processes. *Teor. Verojatnost. i Primenen.* **5** 172–195. MR0137157

[62] SU, J., LI, B. and CHEN, W.-H. (2015). On existence, optimality and asymptotic stability of the Kalman filter with partially observed inputs. *Automatica J. IFAC* **53** 149–154. MR3318583

[63] SZNITMAN, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX—1989*. *Lecture Notes in Math*. **1464** 165–251. Springer, Berlin. MR1108185

[64] TONG, X. T., MAJDA, A. J. and KELLY, D. (2016). Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity* **29** 657–691. MR3461611

[65] VAN LOAN, C. (1977). The sensitivity of the matrix exponential. *SIAM J. Numer. Anal.* **14** 971–981. MR0468137

[66] VESELIĆ, K. (2011). *Damped Oscillations of Linear Systems*: *A Mathematical Introduction*. *Lecture Notes in Math*. **2023**. Springer, Heidelberg. MR2798348

[67] WEN, X.-H. and CHEN, W. H. (2005). Real-time reservoir model updating using ensemble Kalman filter SPE-92991-MS. In *SPE Reservoir Simulation Symposium*, *the Woodlands*, *Texas*.

[68] WONHAM, W. M. (1968). On a matrix Riccati equation of stochastic control. *SIAM J. Control* **6** 681–697. MR0239161

[69] YANG, X. (2000). A matrix trace inequality. *J. Math. Anal. Appl.* **250** 372–374. MR1893896

[70] YANG, X. M., YANG, X. Q. and TEO, K. L. (2001). A matrix trace inequality. *J. Math. Anal. Appl.* **263** 327–331. MR1865285

[71] ZAKAI, M. (1969). On the optimal filtering of diffusion processes. *Z. Wahrsch. Verw. Gebiete* **11** 230.

[72] ZELIKIN, M. I. (1991). On the theory of the matrix Riccati equation *Mat. Sb.* **182** 970–984. Available at http://www.mathnet.ru/links/5052aa2d9c921f105d61c94a204912c1/sm1336.pdf.

[73] ZIEBUR, A. D. (1970). On determining the structure of A by analyzing e^{At}. *SIAM Rev.* **12** 98–102. MR0254074

P. DEL MORAL
UNIVERSITÉ DE BORDEAUX I
33405 BORDEAUX CEDEX
FRANCE
E-MAIL: p.del-moral@unsw.edu.au

J. TUGAUT
10 RUE TRÉFILERIE
CS 82301
42023 SAINT-ÉTIENNE CEDEX 2
FRANCE