Fast Kalman-like filtering for large-dimensional linear and Gaussian state-space models

Boujemaa Ait-El-Fquih and Ibrahim Hoteit

Abstract

This paper considers the filtering problem for linear and Gaussian state-space models with large dimensions, a setup in which the optimal Kalman Filter (KF) might not be applicable owing to the excessive cost of manipulating huge covariance matrices. Among the most popular alternatives that enable cheaper and reasonable computation is the Ensemble KF (EnKF), a Monte Carlo-based approximation. In this paper, we consider a class of *a posteriori* distributions with *diagonal* covariance matrices and propose fast approximate deterministic-based algorithms based on the Variational Bayesian (VB) approach. More specifically, we derive two iterative KF-like algorithms that differ in the way they operate between two successive filtering estimates; one involves a smoothing estimate and the other involves a prediction estimate. Despite its iterative nature, the prediction-based algorithm provides a computational cost that is, on the one hand, independent of the number of iterations in the limit of very large state dimensions, and on the other hand, always much smaller than the cost of the EnKF. The cost of the smoothing-based algorithm depends on the number of iterations that may, in some situations, make this algorithm slower than the EnKF. The performances of the proposed filters are studied and compared to those of the KF and EnKF through a numerical example.

Index Terms

Kalman filtering; Bayesian filtering; High dimension; Variational Bayes; Fast estimation algorithms.

B. Ait-El-Fquih and I. Hoteit are with the Division of Applied Mathematics and Computational Science, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia (e-mails: boujemaa.aitelfquih@kaust.edu.sa, ibrahim.hoteit@kaust.edu.sa).

I. INTRODUCTION

A. Background

The estimation problem of a process, $\mathbf{x} = \{\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_N\}$, from a set of observations, $\mathbf{y} = \{\mathbf{y}_0, \mathbf{y}_1, \cdots, \mathbf{y}_N\}$, is of particular interest in many areas, including target tracking [1] [2] [3] [4], navigation [2] [5] [6], wireless communications [7] [8], and geophysical fluid applications [4] [9] [10] [11] [12]. Let $\mathbf{x}_n \in \mathbb{R}^{n_\mathbf{x}}$ and $\mathbf{y}_n \in \mathbb{R}^{n_\mathbf{y}}$. This problem is generally formulated in the framework of a state-space model:

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{F}_n\mathbf{x}_n + \mathbf{u}_n, \\ \mathbf{y}_n = \mathbf{H}_n\mathbf{x}_n + \mathbf{v}_n, \end{cases} \tag{1}$$

for which $\mathbf{F}_n$ and $\mathbf{H}_n$ respectively denote the state transition matrix and the measurement matrix at time n; the input noise, $\mathbf{u} = \{\mathbf{u}_n\}_{n\in\mathbb{N}}$, and the measurement noise, $\mathbf{v} = \{\mathbf{v}_n\}_{n\in\mathbb{N}}$, are assumed to be independent, jointly independent and independent of the initial state, $\mathbf{x}_0$; and $\mathbf{x}_0$, $\mathbf{u}_n$ and $\mathbf{v}_n$ are Gaussian. Let $\mathbf{x}_0 \sim \mathcal{N}(\widehat{\mathbf{x}}_0, \mathbf{P}_0)$, $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_n)$ with $\mathbf{Q}_n$ diagonal and all diagonal entries are non-zero, $\mathbf{v}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_n)$ with $\mathbf{R}_n$ positive-definite, $\mathbf{x}_{0:n} = \{\mathbf{x}_i\}_{i=0}^n$ and $\mathbf{y}_{0:n} = \{\mathbf{y}_i\}_{i=0}^n$. Let also $p(\mathbf{x}_n)$ and $p(\mathbf{x}_n|\mathbf{y}_{0:m})$ denote the probability density function (pdf) (with respect to (w.r.t.) the Lebesgue measure) of $\mathbf{x}_n$ and the pdf of $\mathbf{x}_n$ conditional on $\mathbf{y}_{0:m}$, respectively. A fundamental problem, so-called *filtering*, consists of estimating, at each time n, the state $\mathbf{x}_n$ from the measurements $\mathbf{y}_{0:n}$. The classical solution is given by the *a posteriori* mean (AM)

$$\widehat{\mathbf{x}}_{n|n} = \mathbb{E}_{p(\mathbf{x}_n|\mathbf{y}_{0:n})}[\mathbf{x}_n] = \int \mathbf{x}_n p(\mathbf{x}_n|\mathbf{y}_{0:n})d\mathbf{x}_n, \tag{2}$$

which minimizes the *a posteriori* mean square error (MSE). Similarly to (2), throughout this paper, $\mathbb{E}_{p(\mathbf{x})}[f(\mathbf{x})]$ denotes the expected value of the function $f(\mathbf{x})$ w.r.t. the distribution $p(\mathbf{x})$. With regard to the computation of (2), the Kalman Filter (KF) has been introduced as an indispensible tool owing to its optimality and recursive character [13] [14] [15]. Consequently, different algorithms have been proposed based on the KF. To name just a few, robust filters have been proposed [16]; smoothing algorithms have been developed [14] [17] [18] [19] [20] [21]; the independence assumptions on $\mathbf{u}$ and/or $\mathbf{v}$ have been dropped [1] [14] [22] [23] [24] [25]; the relaxation of some conditional independence assumptions on $\mathbf{x}$ and/or $\mathbf{y}$ has been considered [26] [27]; and more recently, a new KF-like algorithm has been derived in [28] that does not require the specification of the initial errors and noise covariances. In very large dimensional state-space models, *i.e.*, when the state dimension, $n_\mathbf{x}$, is very large, the KF becomes impractical because of the prohibitive computational cost, $\mathcal{O}(n_\mathbf{x}^3)$, required for the calculations of the error covariance matrices.

In the particular case for which (1) is a time-invariant model, exact computation of the AM estimate (2) as well as the associated error covariance matrix is possible with only a squared cost (see e.g., Chandrasekhar's algorithm [29] [30]). This, however, is no longer true when relaxing the time-invariant property, a situation for which a number of approximate methods have been proposed (see e.g., [10] [31] [32] [33] [34] [35] and references therein). Up to now, the most popular scheme for filtering high-dimensional systems is the Ensemble Kalman Filter (EnKF) [9] [10] [36], which could be used in more general situations including both linear and nonlinear state-space models. The Monte Carlo-based approximation scheme of (2) and the resulting error covariance matrices have been demonstrated to be efficient in many different applications, even when implemented with a very limited number of samples (see § II-B below).

B. Contribution of the paper

In this paper, we introduce a new fast approach for filtering high-dimensional linear state-space models. The basic idea is to compute an estimate of the system state based on a *diagonal* covariance matrix. This allows us to significantly reduce the computational cost and storage requirements, since the computation of huge full covariance matrices would reduce to the computation of scalar variances. From a probabilistic point of view, this amounts to remove the conditional dependence between the state variables, which amounts to approximating the *a posteriori* pdf of the state vector by a product of independent marginal pdfs of its components. Concretely, approximate marginal pdfs are computed from the joint state pdf following the variational Bayesian (VB) optimization criteria, in the sense of minimization of the Kullback-Leibler Divergence (KLD) [37] [38] [39] (see also [40] [41] [42] [43] [44] [45] [46] [47] [48] for a more recent literature about the VB approach for Bayesian inferences). Once the VB-type marginal pdfs are computed, an estimate of each state component can then be obtained by its expected value w.r.t. the associated marginal pdf. More precisely, we apply in this work the VB approach to two joint *a posteriori* pdfs, as follows:

- Starting from the joint pdf of filtering and smoothing $p(\mathbf{x}_n, \mathbf{x}_{n-1}|\mathbf{y}_{0:n})$, we derive an iterative KF-like algorithm to propagate for each kth component of the state vector, an approximation of the filtering and smoothing estimates, $\widehat{x}_{n|n}^k$ and $\widehat{x}_{n-1|n}^k$, respectively. The use of $p(\mathbf{x}_n, \mathbf{x}_{n-1}|\mathbf{y}_{0:n})$ is inspired by [39]. However, in [39], the conditional independence is "enforced", via the VB approach, only between the state vectors $\mathbf{x}_n$ and $\mathbf{x}_{n-1}$; the components of each of these vectors remain dependent conditionally on $\mathbf{y}_{0:n}$. The complexity of the introduced algorithm is square in $n_{\mathbf{x}}$ and depends on the number of iterations, which may make this algorithm more expensive than the EnKF in certain

situations.

- We then rely on the joint pdf of filtering and prediction $p(\mathbf{x}_n, \mathbf{x}_{n+1}|\mathbf{y}_{0:n})$ and derive another iterative KF-like algorithm to propagate an approximation of the filtering estimate, $\widehat{x}_{n|n}^k$, and the prediction estimate, $\widehat{x}_{n+1|n}^k$ (instead of the smoothing estimate). The complexity of this iterative algorithm, which is square in $n_{\mathbf{x}}$, is independent of the number of iterations and, furthermore, is always $\frac{2}{5}M$ times smaller than that of the EnKF, with M being the ensemble size.

The VB approach has been already used in the context of the KF considering (reasonable) low-dimensional state-space systems (see for instance [40] [43] [44] [45] [48] and references therein). More precisely, the work in [40] aimed at estimating the system state, $\mathbf{x}_n$, and the measurement noise covariance, $\mathbf{R}_n$. The VB-like decoupling is used to decouple the *a posteriori* dependence of $\mathbf{x}_n$ and $\mathbf{R}_n$. [43] addressed the filtering problem in an augmented state-space system, $(\mathbf{x}, \mathbf{r}, \mathbf{y})$, for which the covariance of the noise of the (auxiliary) hidden process $\mathbf{r}$, $\mathbf{Q}_n^r$, is unknown. The VB approach was used to insert the *a posteriori* independence between $\mathbf{x}_n$, $\mathbf{r}_n$ and $\mathbf{Q}_n^r$. In [44], a VB-like KF robust to outliers has been introduced. A similar algorithm has then been derived in [45] in the context of backward state-space systems, *i.e.,* systems that evolve in decreasing time direction, and for which a robust VB-like fixed-interval Kalman smoother has been also proposed. More recently, an adaptive VB-like KF has been introduced in [48] to update the covariance of the (assumed Gaussian) proposal distribution of the Markov Chain Monte Carlo-based Metropolis algorithm. Unlike these works, we focus here on large-dimensional state-space systems and we adopt the idea of splitting the system state using the VB approach, which, in the best of our knowledge, is original. This paper is organized as follows. Section II briefly reviews the KF and EnKF algorithms in the classical Bayesian framework. The VB approach of the filtering problem is then described in Section III, highlighting the fact that the application of the VB criteria on the sole filtering pdf is generally not enough, especially when the state transition matrix, $\mathbf{F}_n$, is not diagonal. We thus apply in Section IV the VB criteria on $p(\mathbf{x}_n, \mathbf{x}_{n-1}|\mathbf{y}_{0:n})$ and $p(\mathbf{x}_n, \mathbf{x}_{n+1}|\mathbf{y}_{0:n})$, to derive two approximate KF-like algorithms with squared complexity. Numerical simulations demonstrating the relevance of the proposed algorithms, and comparing them with the KF and EnKF are presented in Section V. The paper concludes with a general discussion, in Section VI.

II. BAYESIAN FILTERING

Consider a general state-space model,

$$\begin{cases} \mathbf{x}_{n+1} = \mathbf{f}_n(\mathbf{x}_n, \mathbf{u}_n), \\ \mathbf{y}_n = \mathbf{h}_n(\mathbf{x}_n, \mathbf{v}_n), \end{cases} \tag{3}$$

for which $\mathbf{f}_n$ and $\mathbf{h}_n$ are not necessarily linear, and $\mathbf{u}_n$ and $\mathbf{v}_n$ are not necessarily Gaussian. In such a model, one can use the independence properties of $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{x}_0$, stated in Section I-A above, to obtain,

$$p(\mathbf{x}_n|\mathbf{x}_{0:n-1}) = p(\mathbf{x}_n|\mathbf{x}_{n-1}), \tag{4}$$

$$p(\mathbf{y}_{0:n}|\mathbf{x}_{0:n}) = \prod_{i=0}^{n} p(\mathbf{y}_i|\mathbf{x}_{0:n}), \tag{5}$$

$$p(\mathbf{y}_i|\mathbf{x}_{0:n}) = p(\mathbf{y}_i|\mathbf{x}_i), \ i = 0, 1, \cdots, n. \tag{6}$$

In other words, the unknown state process, $\mathbf{x}$, is a Markov chain (eq. (4)), the observed process, $\mathbf{y}$, is conditionally independent of $\mathbf{x}$ (eq. (5)), and each observation, $\mathbf{y}_i$, depends on $\mathbf{x}$ through the state at the same time instant, $\mathbf{x}_i$ (eq. (6)). These properties are indeed those of a Hidden Markov Chain (HMC) of the transition pdf $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ and the likelihood $p(\mathbf{y}_n|\mathbf{x}_n)$; these key tools enable recursive efficient computation of the filtering pdf. More precisely, each recursion $(n-1) \rightarrow n$ consists of two steps:

- *The Markovian step* (or *prediction*), in which the transition pdf of the Markov chain, $\mathbf{x}$, is used to compute the prediction pdf:

$$p(\mathbf{x}_n|\mathbf{y}_{0:n-1}) = \int p(\mathbf{x}_n|\mathbf{x}_{n-1})p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1})d\mathbf{x}_{n-1}. \tag{7}$$

- *The Bayesian step* (or *filtering*), in which the likelihood is used to update the prediction pdf via the Bayes' rule:

$$p(\mathbf{x}_n|\mathbf{y}_{0:n}) = \frac{p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{y}_{0:n-1})}{\int p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{y}_{0:n-1})d\mathbf{x}_n}. \tag{8}$$

Once $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ is computed, the AM estimate $\widehat{\mathbf{x}}_{n|n}$ can thus be obtained from (2). In practice, however, the explicit forms of $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ and $\widehat{\mathbf{x}}_{n|n}$ are often intractable as this requires analytical computation of integrals in (7), (8) and (2). Indeed, analytical computation of such integrals is often impossible due to the possible nonlinear character of $\mathbf{f}_n$ and $\mathbf{h}_n$, as well as the fact that $\mathbf{u}_n$ and/or $\mathbf{v}_n$ may be non-Gaussian. Accordingly, a number of deterministic and Monte Carlo-based methods have been derived to approximate (2); see for instance [3] [14] [49] [50] [51] and references therein. Here, we are interested in linear and Gaussian state-space models (1), *i.e.*, $\mathbf{f}_n$ and $\mathbf{h}_n$ in (3) are linear, and $\mathbf{u}_n$ and $\mathbf{v}_n$ are Gaussian. In this case, exact computation of the AM estimate (2) is possible by the KF algorithm [13] [14] [15], which we briefly review below.

A. Kalman filter

In the state-space model (1), one can easily check that

$$p(\mathbf{x}_n|\mathbf{x}_{n-1}) = \mathcal{N}_{\mathbf{x}_n}(\mathbf{F}_{n-1}\mathbf{x}_{n-1}, \mathbf{Q}_{n-1}), \tag{9}$$

$$p(\mathbf{y}_n|\mathbf{x}_n) = \mathcal{N}_{\mathbf{y}_n}(\mathbf{H}_n\mathbf{x}_n, \mathbf{R}_n), \tag{10}$$

where, except when stated otherwise, $\mathcal{N}_{\mathbf{u}}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian pdf with argument $\mathbf{u}$ and parameters $(\mathbf{m}, \mathbf{C})$. Accordingly, the filtering and prediction pdfs are Gaussian and are thus described by their first two moments. Propagating $p(\mathbf{x}_n|\mathbf{y}_{0:m})$, $m = n - 1, n$, amounts to propagating their means, $\widehat{\mathbf{x}}_{n|m}$, and covariance matrices, $\mathbf{P}_{n|m}$. More precisely, (7), which computes $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ from $p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1})$ using $p(\mathbf{x}_n|\mathbf{x}_{n-1})$ (which is parameterized by $\mathbf{F}_{n-1}$ and $\mathbf{Q}_{n-1}$), reduces to the prediction step of the KF relating $(\widehat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1})$ with $(\widehat{\mathbf{x}}_{n-1|n-1}, \mathbf{P}_{n-1|n-1})$ through $\mathbf{F}_{n-1}$ and $\mathbf{Q}_{n-1}$:

$$\widehat{\mathbf{x}}_{n|n-1} = \mathbf{F}_{n-1}\widehat{\mathbf{x}}_{n-1|n-1}, \tag{11}$$

$$\mathbf{P}_{n|n-1} = \mathbf{F}_{n-1}\mathbf{P}_{n-1|n-1}\mathbf{F}_{n-1}^T + \mathbf{Q}_{n-1}. \tag{12}$$

Moreover, (8), which computes $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ from $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ using $p(\mathbf{y}_n|\mathbf{x}_n)$ (as parameterized by $\mathbf{H}_n$ and $\mathbf{R}_n$), reduces to the filtering step of the KF relating $(\widehat{\mathbf{x}}_{n|n}, \mathbf{P}_{n|n})$ to $(\widehat{\mathbf{x}}_{n|n-1}, \mathbf{P}_{n|n-1})$ via $\mathbf{H}_n$ and $\mathbf{R}_n$:

$$\mathbf{K}_n = \mathbf{P}_{n|n-1}\mathbf{H}_n^T \left[\mathbf{H}_n\mathbf{P}_{n|n-1}\mathbf{H}_n^T + \mathbf{R}_n\right]^{-1}, \tag{13}$$

$$\widehat{\mathbf{x}}_{n|n} = \widehat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n \left(\mathbf{y}_n - \mathbf{H}_n\widehat{\mathbf{x}}_{n|n-1}\right), \tag{14}$$

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n\mathbf{H}_n\mathbf{P}_{n|n-1}. \tag{15}$$

The number of floating operations (flops) required to implement the KF equations at each recursion, $(n - 1) \to n$, is approximately,

$$\mathcal{C}_{KF} = 3n_{\mathbf{x}}^3 + \left(3n_{\mathbf{y}} + \frac{1}{2}\right) n_{\mathbf{x}}^2 + 3\left(n_{\mathbf{y}}^2 + n_{\mathbf{y}}\right) n_{\mathbf{x}} + \frac{2}{3}n_{\mathbf{y}}^3 + \frac{3}{2}n_{\mathbf{y}}^2 + \frac{5}{6}n_{\mathbf{y}}. \tag{16}$$

When the state dimension, $n_{\mathbf{x}}$, is very large, direct implementation of the KF is not possible because of the cubic term $3n_{\mathbf{x}}^3$, which originates from the computation of $\mathbf{P}_{n|n-1}$ in (12). We briefly review below the EnKF algorithm, one of the most popular alternatives used to circumvent this problem.

B. Ensemble Kalman filter

The basic idea behind the EnKF is to use Monte Carlo Gaussian-based approximations of the prediction and filtering distributions. Starting from an ensemble of independent samples, $\{\mathbf{x}_{n-1}^{f,(m)}\}_{m=1}^{M}$, drawn from $p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1})$, the prediction step of the EnKF uses (7) to compute an ensemble of independent samples, $\{\mathbf{x}_n^{p,(m)}\}_{m=1}^{M}$, approximating $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ as

$$\mathbf{x}_n^{p,(m)} = \mathbf{F}_{n-1}\mathbf{x}_{n-1}^{f,(m)} + \mathbf{u}_{n-1}^{(m)}, \text{ with } \mathbf{u}_{n-1}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{n-1}). \tag{17}$$

The AM estimate (11) is then approximated by the empirical mean $\widetilde{\mathbf{x}}_{n|n-1} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{x}_n^{p,(m)}$. An approximation of the covariance matrix (12) is also given by

$$\widetilde{\mathbf{P}}_{n|n-1} = \mathbf{A}_n\mathbf{A}_n^T, \tag{18}$$

with

$$\mathbf{A}_n = \frac{1}{\sqrt{M-1}}\left[\widetilde{\mathbf{x}}_{n|n-1}-\mathbf{x}_n^{p,(1)}, \widetilde{\mathbf{x}}_{n|n-1}-\mathbf{x}_n^{p,(2)}, \cdots, \widetilde{\mathbf{x}}_{n|n-1}-\mathbf{x}_n^{p,(M)}\right].$$

In the filtering step, an approximation $\{\mathbf{x}_n^{f,(m)}\}_{m=1}^{M}$ of $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ is computed, based on (8), by correcting each sample, $\mathbf{x}_n^{p,(m)}$, using the KF update step:

$$\mathbf{x}_n^{f,(m)} = \mathbf{x}_n^{p,(m)} + \widetilde{\mathbf{K}}_n(\mathbf{y}_n^{(m)} - \mathbf{H}_n\mathbf{x}_n^{p,(m)}), \tag{19}$$

where $\mathbf{y}_n^{(m)} \sim \mathcal{N}(\mathbf{y}_n, \mathbf{R}_n)$ is a stochastic perturbation of the observation, and $\widetilde{\mathbf{K}}_n = \mathbf{A}_n\mathbf{G}_n^T[\mathbf{G}_n\mathbf{G}_n^T + \mathbf{R}_n]^{-1}$ with $\mathbf{G}_n = \mathbf{H}_n\mathbf{A}_n$, is the Kalman gain. An approximation of the filtering estimate of the state (14) is computed by the empirical mean and that of $\mathbf{P}_{n|n}$ as in (18). The number of flops required by the EnKF to compute the prediction and filtering estimates at each time n is approximately,

$$\mathcal{C}_{EnKF} = 2Mn_{\mathbf{x}}^2 + \left(2n_{\mathbf{y}}^2 + (8M-1)n_{\mathbf{y}} + 6M + 1\right)n_{\mathbf{x}} + n_{\mathbf{y}}^3 + \left(2M + \frac{5}{2}\right)n_{\mathbf{y}}^2 + \left(M + \frac{3}{2}\right)n_{\mathbf{y}}. \tag{20}$$

The EnKF was originally introduced in [9] for large-scale nonlinear state-space models and was successfully used in many atmospheric / oceanic data assimilation applications for which $n_{\mathbf{x}}$ is in the order of millions. The first encouraging results have opened the way for a large number of works based on this filter, including derivation of other EnKF variants [32] [52] [53] [54], relaxation of the Gaussian assumption on filtering and prediction pdfs by assuming Gaussian mixture representations [55] [56] [57] [58] [59], or development of smoothing algorithms [60] [61].

III. VARIATIONAL BAYESIAN FILTERING

A VB-like approximation of the filtering pdf of interest by a separable product of *a posteriori* marginal pdfs can be obtained by minimizing the KLD, *i.e.*,

$$p(\mathbf{x}_n|\mathbf{y}_{0:n}) \approx \prod_{k=1}^{n_{\mathbf{x}}} q(x_n^k|\mathbf{y}_{0:n}), \tag{21}$$

with

$$\prod_{k=1}^{n_{\mathbf{x}}} q(x_n^k|\mathbf{y}_{0:n}) = \underset{\prod_k \widetilde{p}(x_n^k|\mathbf{y}_{0:n})}{\arg\min} \text{KLD}\left(\prod_{k=1}^{n_{\mathbf{x}}}\widetilde{p}(x_n^k|\mathbf{y}_{0:n})||p(\mathbf{x}_n|\mathbf{y}_{0:n})\right),$$

$$= \underset{\prod_k \widetilde{p}(x_n^k|\mathbf{y}_{0:n})}{\arg\min} \int \prod_{k=1}^{n_{\mathbf{x}}}\widetilde{p}(x_n^k|\mathbf{y}_{0:n})\ln\left(\frac{\prod_{k=1}^{n_{\mathbf{x}}}\widetilde{p}(x_n^k|\mathbf{y}_{0:n})}{p(\mathbf{x}_n|\mathbf{y}_{0:n})}\right)d\mathbf{x}_n; \tag{22}$$

DRAFT

u^k denotes the kth component of a vector $\mathbf{u}$, and $\mathbf{u}^{k^-}$ its complementary part in $\mathbf{u}$. Each marginal $q(x_n^k|\mathbf{y}_{0:n})$ is given by [39] [42]:

$$q(x_n^k|\mathbf{y}_{0:n}) \propto \exp\left(\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{x}_n,\mathbf{y}_{0:n}))\right]\right). \tag{23}$$

According to (23), although the approximated marginal pdf of x_n^k, $q(x_n^k|\mathbf{y}_{0:n})$, is independent of those of the other components $\mathbf{x}_n^{k^-}$, $q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})$, it nevertheless remains dependent of the expected value of $\ln(p(\mathbf{x}_n,\mathbf{y}_{0:n}))$ w.r.t. $q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})$. In the particular case of Gaussian pdfs, one can easily check that for each k, $q(x_n^k|\mathbf{y}_{0:n})$ remains dependent of the first and second moments of $q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})$.

Once $q(x_n^k|\mathbf{y}_{0:n})$ is computed, an estimate of x_n^k can then be obtained by the AM estimate $\mathbb{E}_{q(x_n^k|\mathbf{y}_{0:n})}[x_n^k]$. What is left now is to perform a *recursive* computation of these estimates based on (23) and the dynamical structure of model (1). Indeed, using the factorization

$$p(\mathbf{x}_n,\mathbf{y}_{0:n}) \propto p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{y}_{0:n-1}), \tag{24}$$

the computation of the right-hand side (r.h.s) of (23) requires that of the prediction pdf. For that purpose, two cases should be distinguished depending on whether or not $\mathbf{F}_n$ is diagonal.

A. A diagonal state transition matrix

Considering model (1) with a diagonal state transition matrix, we have

$$p(x_n^k|\mathbf{x}_{0:n-1}) = p(x_n^k|x_{0:n-1}^k) = p(x_n^k|x_{n-1}^k). \tag{25}$$

In other words, each marginal process, $x^k = \{x_n^k\}_{n\in\mathbb{N}}$, forms a Markov chain regardless of those of the other components, x^l, $l \neq k$. This property is important in practice since it leads to a separable approximation of the prediction pdf as

$$p(\mathbf{x}_n|\mathbf{y}_{0:n-1}) \approx \prod_{k=1}^{n_{\mathbf{x}}} q(x_n^k|\mathbf{y}_{0:n-1}), \tag{26}$$

with

$$q(x_n^k|\mathbf{y}_{0:n-1}) = \int p(x_n^k|x_{n-1}^k)q(x_{n-1}^k|\mathbf{y}_{0:n-1})dx_{n-1}^k, \tag{27}$$

which, in turn, yields a separable approximation of the filtering pdf (21) by inserting (26) first in (24) and next in (23). Now, using the linearity and Gaussian structure of (1), an approximate KF-like algorithm propagating the state estimates with associated diagonal covariance matrices can be derived. This algorithm has been recently introduced in [42].

B. An off-diagonal state transition matrix

In the general case in which $\mathbf{F}_n$ is not diagonal, the Markov chain property (25) is no longer a valid and separable approximation of the prediction pdf, (26)-(27), is no longer possible. More precisely, one rather has

$$p(\mathbf{x}_n|\mathbf{y}_{0:n-1}) \approx \int \prod_{k=1}^{n_\mathbf{x}} p(x_n^k|\mathbf{x}_{n-1}) q(x_{n-1}^k|\mathbf{y}_{0:n-1}) d\mathbf{x}_{n-1}. \tag{28}$$

In practice, starting from a Gaussian pdf, $\prod_{k=1}^{n_\mathbf{x}} q(x_{n-1}^k|\mathbf{y}_{0:n-1})$, with a diagonal covariance matrix, (28) provides a Gaussian pdf but with an off-diagonal covariance matrix, and thereby the initial problem of the huge computational burden and storage capacity persists. Accordingly, (21) should be replaced by an alternative VB approximation involving diagonal matrices in both the Markovian and Bayesian steps. Two schemes are proposed below.

- The first approximation is applied on the joint pdf of smoothing and filtering as

$$p(\mathbf{x}_{n-1}, \mathbf{x}_n|\mathbf{y}_{0:n}) \quad \approx \quad q(\mathbf{x}_{n-1}|\mathbf{y}_{0:n}) q(\mathbf{x}_n|\mathbf{y}_{0:n}), \tag{29}$$

$$\approx \quad \prod_{k=1}^{n_\mathbf{x}} q(x_{n-1}^k|\mathbf{y}_{0:n}) q(x_n^k|\mathbf{y}_{0:n}). \tag{30}$$

  By doing so, we impose independence between the filtering and smoothing pdfs of the whole state vector (cf. (29)) and between their components (cf. (30)). Approximation (29) has already been used in [39] in the context of particle filtering. We exploit it here for reducing the computational burden of the KF by splitting the state vector as in (30). This leads to an approximate KF-like algorithm propagating the filtering and smoothing estimates with diagonal covariance matrices (cf. § IV-A).

- The second approximation relies on the joint pdf of prediction and filtering:

$$p(\mathbf{x}_{n-1}, \mathbf{x}_n|\mathbf{y}_{0:n-1}) \approx \prod_{k=1}^{n_\mathbf{x}} q(x_{n-1}^k, x_n^k|\mathbf{y}_{0:n-1}). \tag{31}$$

  Unlike (29), the conditional dependence of $\mathbf{x}_{n-1}$ and $\mathbf{x}_n$ is conserved in (31), while imposing independence between their components only. Furthermore, one can see that (31) follows from (21) combined with the approximation $p(x_n^k|\mathbf{x}_{n-1}) \approx q(x_n^k|x_{n-1}^k, \mathbf{y}_{0:n-1})$. Accordingly, the prediction approximation (28) reduces to (26)-(27) by replacing $p(x_n^k|x_{n-1}^k)$ with $q(x_n^k|x_{n-1}^k, \mathbf{y}_{0:n-1})$. This leads to another approximate KF-like algorithm involving the prediction and filtering estimates with diagonal covariance matrices (cf. § IV-B).

IV. VB Smoothing- and Prediction-Based KF

This section presents the VB approach based on the approximations (30) and (31). Before we proceed, we define the notations, for a matrix $\mathbf{M}$, used hereafter.

- $M^{i,j}$ denotes the $(i,j)^{\text{th}}$ entry.

- $\mathbf{M}(i,:)$ denotes the ith row and $\mathbf{M}(:,j)$ the jth column.

- $\mathbf{M}(i,j^-)$ (resp. $\mathbf{M}(i^-,j)$) denotes the complementary part of $M^{i,j}$ in $\mathbf{M}(i,:)$ (resp. in $\mathbf{M}(:,j)$).

- $\mathbf{M}(:,j^-)$ (resp. $\mathbf{M}(i^-,:)$) denotes the complementary part of $\mathbf{M}(:,j)$ (resp. of $\mathbf{M}(i,:)$) in $\mathbf{M}$.

A. VB Smoothing-Based KF (VBSKF)

Similarly to (21)-(23), the VB-marginal pdfs, solutions of (30), are given by

$$q(x_{n-1}^k|\mathbf{y}_{0:n}) \quad \propto \quad \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1}^{k-},\mathbf{x}_n|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{x}_{n-1},\mathbf{x}_n,\mathbf{y}_{0:n}))\right]\right), \tag{32}$$

$$q(x_n^k|\mathbf{y}_{0:n}) \quad \propto \quad \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1},\mathbf{x}_n^{k-}|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{x}_{n-1},\mathbf{x}_n,\mathbf{y}_{0:n}))\right]\right). \tag{33}$$

Solutions (32) and (33) share the joint pdf $p(\mathbf{x}_{n-1},\mathbf{x}_n,\mathbf{y}_{0:n})$, which, based on the HMC properties (4)-(6), can be factorized as

$$p(\mathbf{x}_{n-1},\mathbf{x}_n,\mathbf{y}_{0:n}) \propto p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{x}_{n-1})p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1}). \tag{34}$$

1) Explicit derivation: Replacing in (34) $p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1})$ with its VB approximation, (32) and (33) respectively become[1],

$$q(x_{n-1}^k|\mathbf{y}_{0:n}) \quad \propto \quad \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1}^{k-},\mathbf{x}_n|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{x}_n|\mathbf{x}_{n-1}))\right]\right)q(x_{n-1}^k|\mathbf{y}_{0:n-1}), \tag{35}$$

$$q(x_n^k|\mathbf{y}_{0:n}) \quad \propto \quad \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1}|\mathbf{y}_{0:n})}\left[\ln(p(x_n^k|\mathbf{x}_{n-1}))\right] + \mathbb{E}_{q(\mathbf{x}_n^{k-}|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{y}_n|\mathbf{x}_n))\right]\right). \tag{36}$$

Let $\overline{x}_{l|n}^k$ and $v_{l|n}^k$ denote respectively the mean and variance of $q(x_l^k|\mathbf{y}_{0:n})$ for $l,n \in \mathbb{N}$; $\overline{\mathbf{x}}_{l|n}$ and $\overline{\mathbf{P}}_{l|n} = \text{diag}(\mathbf{v}_{l|n})$ will thus stand for the mean and diagonal covariance matrix of $\prod_{k=1}^{n_\mathbf{x}} q(x_l^k|\mathbf{y}_{0:n})$. According to the Gaussian property with linear mean in $\mathbf{x}_n$ of the HMC laws $p(\mathbf{y}_n|\mathbf{x}_n)$ and $p(\mathbf{x}_n|\mathbf{x}_{n-1})$, it appears that $q(x_{n-1}^k|\mathbf{y}_{0:n})$ and $q(x_n^k|\mathbf{y}_{0:n})$ are Gaussian with moments propagated as (cf. Appendix):

$$v_{n-1|n}^k = \left[\left(v_{n-1|n-1}^k\right)^{-1} + \mathbf{F}_{n-1}^T(:,k)\mathbf{Q}_{n-1}^{-1}\mathbf{F}_{n-1}(:,k)\right]^{-1}, \tag{37}$$

$$\overline{x}_{n-1|n}^k = v_{n-1|n}^k\left[\left(v_{n-1|n-1}^k\right)^{-1}\overline{x}_{n-1|n-1}^k + \mathbf{F}_{n-1}^T(:,k)\mathbf{Q}_{n-1}^{-1}(\overline{\mathbf{x}}_{n|n} - \mathbf{F}_{n-1}(:,k^-)\overline{\mathbf{x}}_{n-1|n}^{k-})\right], \tag{38}$$

$$v_{n|n}^k = \left[\left(Q_{n-1}^{k,k}\right)^{-1} + \mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}\mathbf{H}_n(:,k)\right]^{-1}, \tag{39}$$

$$\overline{x}_{n|n}^k = v_{n|n}^k\left[\left(Q_{n-1}^{k,k}\right)^{-1}\mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n} + \mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}(\mathbf{y}_n - \mathbf{H}_n(:,k^-)\overline{\mathbf{x}}_{n|n}^{k-})\right]. \tag{40}$$

A particular case of (37)-(40) has been introduced in [38, chapter 7, § 7.4][2] in which the system state

---

[1] Without abuse of language, the r.h.s. of (35) (resp. (36)) is rather proportional to an approximation of (32) (resp. (33)).

[2] A typo has been noticed in eqs. (7.37) and (7.38) in [38, chapter 7, § 7.4]. Indeed, $\mathbf{A}'\mathbf{R}_\theta^{-1}\mathbf{A}$ and $\mathbf{R}_\theta^{-1}\mathbf{A}\widehat{\theta}_{t-1}$ should be replaced by $\mathbf{R}_\theta^{-1}$ and $\mathbf{A}\widehat{\theta}_{t-1}$, respectively.

is not split, *i.e.*, only the approximation (29) is considered. In such a case, the terms $\mathbf{F}_{n-1}(:,k^-)\overline{\mathbf{x}}_{n-1|n}^{k^-}$ in (38) and $\mathbf{H}_n(:,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}$ in (40) vanish, and $\mathbf{F}_{n-1}(:,k)$, $Q_{n-1}^{k,k}$ and $\mathbf{H}_n(:,k)$ are replaced with $\mathbf{F}_{n-1}$, $\mathbf{Q}_{n-1}$ and $\mathbf{H}_n$, respectively.

Remark 1: Parameters $(\overline{x}_{n-1|n}^k, v_{n-1|n}^k)$ in (37)-(38) and $(\overline{x}_{n|n}^k, v_{n|n}^k)$ in (39)-(40) look like those of a Gaussian *a posteriori* pdf computed within a classical fully Bayesian framework. Indeed, according to Prop. 1 (cf. Appendix), equations (37)-(38) come down to computing the Gaussian posterior, $q(x_{n-1}^k|\mathbf{y}_{0:n}) = q(x_{n-1}^k|\mathbf{x}_{n-1}^{k^-} = \overline{\mathbf{x}}_{n-1|n}^{k^-}, \mathbf{y}_{0:n})$, from the Gaussian prior, $q(x_{n-1}^k|\mathbf{y}_{0:n-1}) = q(x_{n-1}^k|\mathbf{x}_{n-1}^{k^-} = \overline{\mathbf{x}}_{n-1|n}^{k^-}, \mathbf{y}_{0:n-1})$, using the likelihood[3]

$$l(\mathbf{y}_n|x_{n-1}^k, \mathbf{x}_{n-1}^{k^-} = \overline{\mathbf{x}}_{n-1|n}^{k^-}, \mathbf{y}_{0:n-1}) \overset{(35)}{\propto} \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1}^{k^-}, \mathbf{x}_n|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{x}_n|\mathbf{x}_{n-1}))\right]\right), \tag{41}$$

$$= \mathcal{N}_{\overline{\mathbf{x}}_{n|n}}(\mathbf{F}_{n-1}(:,k)x_{n-1}^k + \mathbf{F}_{n-1}(:,k^-)\overline{\mathbf{x}}_{n-1|n}^{k^-}, \mathbf{Q}_{n-1}). \tag{42}$$

Similarly, (39)-(40) point out to the relationship between the Gaussian prior, $q(x_n^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n-1}) = \mathcal{N}_{x_n^k}(\mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n}, Q_{n-1}^{k,k})$, and the Gaussian posterior, $q(x_n^k|\mathbf{y}_{0:n}) = q(x_n^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n})$, via the likelihood,

$$p(\mathbf{y}_n|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}) = \mathcal{N}_{\mathbf{y}_n}(\mathbf{H}_n(:,k)x_n^k + \mathbf{H}_n(:,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{R}_n). \tag{43}$$

Making the connection with (36), one obtains

$$q(x_n^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n-1}) \propto \exp\left(\mathbb{E}_{q(\mathbf{x}_{n-1}|\mathbf{y}_{0:n})}\left[\ln(p(x_n^k|\mathbf{x}_{n-1}))\right]\right),$$

$$p(\mathbf{y}_n|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}) \propto \exp\left(\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})}\left[\ln(p(\mathbf{y}_n|\mathbf{x}_n))\right]\right).$$

To summarize, the VB approach in a Gaussian framework can be seen as a fully Bayesian approach by assigning the marginal variables their *a posteriori* expectation.

2) Practical implementation: Let us turn back to eqs. (37)-(40). A problem may arise in (38) and (40) from the fact that $\overline{x}_{n-1|n}^k$ in (38) requires knowledge of $\overline{\mathbf{x}}_{n-1|n}^{k^-}$ and $\overline{\mathbf{x}}_{n|n}$, and $\overline{x}_{n|n}^k$ in (40) requires knowledge of $\overline{\mathbf{x}}_{n|n}^{k^-}$ and $\overline{\mathbf{x}}_{n-1|n}$. This makes it impossible to evaluate $\overline{x}_{n-1|n}^k$ and $\overline{x}_{n|n}^k$ exactly and approximations should therefore be performed. A classical idea would be to proceed with iterations by evaluating one variable while the others are kept fixed [38] (see e.g., [62] in which the convergence of marginal VB-based solutions has been proven). At each iteration i, this consists of evaluating $\overline{x}_{n-1|n}^k$ with (38) by using $\{\overline{x}_{n|n}^j\}_{j=k}^{n_\mathbf{x}}$ and $\{\overline{x}_{n-1|n}^j\}_{j=k+1}^{n_\mathbf{x}}$ computed at iteration $i-1$, and $\{\overline{x}_{n|n}^j\}_{j=1}^{k-1}$ and $\{\overline{x}_{n-1|n}^j\}_{j=1}^{k-1}$ obtained at iteration i; the update of $\overline{x}_{n|n}^k$ in (40) is performed similarly. Furthermore, it should be

---

[3]Without abuse of notation, $\mathcal{N}_{\overline{\mathbf{x}}_{n|n}}(.,.)$ in (42) does not represent a Gaussian pdf, but the value of this pdf at point $\overline{\mathbf{x}}_{n|n}$; this remark holds for $\mathcal{N}_{\mathbf{y}_n}(.,.)$ in (43) and $\mathcal{N}_{\overline{\mathbf{x}}_{n+1|n}^{k^-}}(.,.)$ in (52) as well.

noted that for each component k and iteration i, the updates of $\xi_n^{\mathbf{x}} = \mathbf{F}_{n-1}(:, k^-)\overline{\mathbf{x}}_{n-1|n}^{k^-}$ in (38) and $\xi_n^{\mathbf{y}} = \mathbf{H}_n(:, k^-)\overline{\mathbf{x}}_{n|n}^{k^-}$ in (40) are computationally demanding since they require $n_{\mathbf{x}}(n_{\mathbf{x}}-1)$ and $n_{\mathbf{y}}(n_{\mathbf{x}}-1)$ flops operations, respectively. Based on the linear combination property[4] of $\xi_n^{\mathbf{x}}$ and $\xi_n^{\mathbf{y}}$, one can reduce this cost by updating only the term associated with the $(k-1)^{\text{th}}$ component. More specifically, in (38) and (40), only $\mathbf{F}_{n-1}(:, k-1)\overline{x}_{n-1|n}^{k-1}$ and $\mathbf{H}_n(:, k-1)\overline{x}_{n|n}^{k-1}$ are updated, which reduces the initial computational cost by a factor of $(n_{\mathbf{x}} - 1)$. This iterative KF-like scheme, which involves the smoothing estimate of the system state with diagonal covariance, is summarized in Algorithm 1.

Algorithm 1 VB Smoothing-based KF (VBSKF)

Let $\widetilde{\mathbf{H}}_n = \mathbf{H}_n^T \mathbf{R}_n^{-1}$, $\mathbf{d}_{h,n} = \text{diag}(\widetilde{\mathbf{H}}_n \mathbf{H}_n)$, $\widetilde{\mathbf{F}}_{n-1} = \mathbf{F}_{n-1}^T \mathbf{Q}_{n-1}^{-1}$, $\mathbf{d}_{f,n-1} = \text{diag}(\widetilde{\mathbf{F}}_{n-1}\mathbf{F}_{n-1})$.

- Variances. For $k = 1, 2, \cdots, n_{\mathbf{x}}$,

$$v_{n-1|n}^k = \left[\left(v_{n-1|n-1}^k\right)^{-1} + d_{f,n-1}^k \right]^{-1},$$

$$v_{n|n}^k = \left[\left(Q_{n-1}^{k,k}\right)^{-1} + d_{h,n}^k \right]^{-1}.$$

- Means. We proceed with iterations by initializing as $\overline{\mathbf{x}}_{n-1|n} = \overline{\mathbf{x}}_{n-1|n-1}$, $\overline{\mathbf{x}}_{n|n} = \mathbf{F}_{n-1}\overline{\mathbf{x}}_{n-1|n-1}$. Let $\mathbf{z}_n^{\mathbf{x}} = \mathbf{F}_{n-1}\overline{\mathbf{x}}_{n-1|n}$, $\mathbf{z}_n^{\mathbf{y}} = \mathbf{H}_n\overline{\mathbf{x}}_{n|n}$, $\mathbf{c}_{n-1|n} = \overline{\mathbf{P}}_{n-1|n}\overline{\mathbf{P}}_{n-1|n-1}^{-1}\overline{\mathbf{x}}_{n-1|n-1}$. For $iter = 1, 2, \cdots, I \gg 1$; for $k = 1, 2, \cdots, n_{\mathbf{x}}$,

  - *Computation of means.*

$$\xi_n^{\mathbf{x}} = \mathbf{z}_n^{\mathbf{x}} - \mathbf{F}_{n-1}(:, k)\overline{x}_{n-1|n}^k,$$

$$\overline{x}_{n-1|n}^k = c_{n-1|n}^k + v_{n-1|n}^k \widetilde{\mathbf{F}}_{n-1}(k,:)(\overline{\mathbf{x}}_{n|n} - \xi_n^{\mathbf{x}}),$$

$$\xi_n^{\mathbf{y}} = \mathbf{z}_n^{\mathbf{y}} - \mathbf{H}_n(:, k)\overline{x}_{n|n}^k,$$

$$\overline{x}_{n|n}^k = v_{n|n}^k \left[\left(Q_{n-1}^{k,k}\right)^{-1} \mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n} + \widetilde{\mathbf{H}}_n(k,:)(\mathbf{y}_n - \xi_n^{\mathbf{y}}) \right].$$

  - *Update of $\mathbf{z}_n^{\mathbf{x}}$ and $\mathbf{z}_n^{\mathbf{y}}$.*

$$\mathbf{z}_n^{\mathbf{x}} = \xi_n^{\mathbf{x}} + \mathbf{F}_{n-1}(:, k)\overline{x}_{n-1|n}^k,$$

$$\mathbf{z}_n^{\mathbf{y}} = \xi_n^{\mathbf{y}} + \mathbf{H}_n(:, k)\overline{x}_{n|n}^k.$$

Note that instead of presetting a number of iterations, I, an alternative stopping criteria based, for

---

[4]Note that $\xi_n^{\mathbf{x}} = \sum_{\substack{j=1 \\ j \neq k}}^{n_{\mathbf{x}}} \mathbf{F}_{n-1}(:, j)\overline{x}_{n-1|n}^j$ and $\xi_n^{\mathbf{y}} = \sum_{\substack{j=1 \\ j \neq k}}^{n_{\mathbf{x}}} \mathbf{H}_n(:, j)\overline{x}_{n|n}^j$.

instance, on the relative squared error norm, $\frac{||\overline{\mathbf{x}}_{n|n}^{(iter)} - \overline{\mathbf{x}}_{n|n}^{(iter-1)}||_2}{||\overline{\mathbf{x}}_{n|n}^{(iter-1)}||_2}$, or on KLD, could be used. The number of flops required at each time instant is approximately

$$\mathcal{C}_{VBSKF} = (9I + 5)\, n_{\mathbf{x}}^2 + \left(2n_{\mathbf{y}}^2 + (7I+4)n_{\mathbf{y}} + 2I + 4\right) n_{\mathbf{x}} + \frac{2}{3}n_{\mathbf{y}}^3 + \frac{3}{2}n_{\mathbf{y}}^2 - \frac{n_{\mathbf{y}}}{6}. \qquad (44)$$

B. VB Prediction-Based KF (VBPKF)

As above, the VB-marginal pdfs associated with the approximation (31) at time n are given by

$$q(x_n^k, x_{n+1}^k | \mathbf{y}_{0:n}) \propto \exp\left(\mathbb{E}_{q(\mathbf{x}_n^{k-}, \mathbf{x}_{n+1}^{k-} | \mathbf{y}_{0:n})} \left[\ln(p(\mathbf{x}_n, \mathbf{x}_{n+1}, \mathbf{y}_{0:n})) \right] \right), \qquad (45)$$

with

$$p(\mathbf{x}_n, \mathbf{x}_{n+1}, \mathbf{y}_{0:n}) \propto p(\mathbf{x}_{n+1}|\mathbf{x}_n)p(\mathbf{y}_n|\mathbf{x}_n)p(\mathbf{x}_n|\mathbf{y}_{0:n-1}). \qquad (46)$$

1) Explicit derivation: Replacing $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ in (46) by its VB-approximation, (45) can be approximated by,

$$q(x_n^k, x_{n+1}^k|\mathbf{y}_{0:n}) \propto \exp\!\Big(\mathbb{E}_{q(\mathbf{x}_n^{k-}, \mathbf{x}_{n+1}^{k-}|\mathbf{y}_{0:n})}[\ln(p(\mathbf{x}_{n+1}|\mathbf{x}_n))] + \mathbb{E}_{q(\mathbf{x}_n^{k-}|\mathbf{y}_{0:n})}[\ln(p(\mathbf{y}_n|\mathbf{x}_n))] \Big) q(x_n^k|\mathbf{y}_{0:n-1}). \qquad (47)$$

Based on the Gaussian property with linear mean in $\mathbf{x}_n$ of $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$ and $p(\mathbf{y}_n|\mathbf{x}_n)$, $q(x_n^k|\mathbf{y}_{0:n})$ and $q(x_{n+1}^k|\mathbf{y}_{0:n})$ are Gaussian whose moments satisfy (cf. Appendix):

$$v_{n|n}^k = \left[\mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}\mathbf{H}_n(:,k) + \frac{1}{v_{n|n-1}^k} + \mathbf{F}_n^T(k^-,k)\mathbf{Q}_n^{-1}(k^-,k^-)\mathbf{F}_n(k^-,k) \right]^{-1}, \qquad (48)$$

$$\overline{x}_{n|n}^k = \left[\mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}\mathbf{H}_n(:,k) + \frac{1}{v_{n|n-1}^k} \right]^{-1} \left[\mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}(\mathbf{y}_n - \mathbf{H}_n(:,k^-)\overline{\mathbf{x}}_{n|n}^{k-}) + \frac{\overline{x}_{n|n-1}^k}{v_{n|n-1}^k} \right], \qquad (49)$$

$$\overline{x}_{n+1|n}^k = \mathbf{F}_n(k,:)\overline{\mathbf{x}}_{n|n}, \qquad (50)$$

$$v_{n+1|n}^k = (F_n^{k,k})^2 v_{n|n}^k + Q_n^{k,k}. \qquad (51)$$

Equations (48)-(51) can be seen as a generalization of an algorithm recently introduced in [42] in the particular case of $\mathbf{F}_n = \epsilon \times \mathbb{I}_{n_{\mathbf{x}}}$, $\mathbf{Q}_n = \sigma_u^2 \times \mathbb{I}_{n_{\mathbf{x}}}$ and $\mathbf{R}_n = \sigma_v^2 \times \mathbb{I}_{n_{\mathbf{y}}}$, and without the third term in the r.h.s. of (48).

Remark 2: Similarly to Remark 1 above, (48)-(51) can be also obtained from a classical Bayesian framework by keeping the marginal variables equal to their *a posteriori* expectation. Indeed, (48) and (49) (which coincides with (63)), translate the connection of the Gaussian prior, $q(x_n^k|\mathbf{y}_{0:n-1})$, with the Gaussian posterior, $q(x_n^k|\mathbf{y}_{0:n})$. This can be shown using Prop. 2 (cf. Appendix) in the following Bayes'

rule:

$$
\begin{aligned}
q(x_n^k|\mathbf{y}_{0:n}) &= q(x_n^k|\mathbf{x}_{n+1}^{k^-} = \overline{\mathbf{x}}_{n+1|n}^{k^-}, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n}), \\
&\propto l(\overline{\mathbf{x}}_{n+1|n}^{k^-}, \mathbf{y}_n|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-})q(x_n^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n-1}), \\
&= \underbrace{p(\mathbf{y}_n|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-})}_{(43)} \underbrace{l(\overline{\mathbf{x}}_{n+1|n}^{k^-}|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-})}_{\mathcal{N}_{\overline{\mathbf{x}}_{n+1|n}^{k^-}}(\mathbf{F}_n(k^-,k)x_n^k+\mathbf{F}_n(k^-,k^-)\overline{\mathbf{x}}_{n|n}^{k^-},\mathbf{Q}_n(k^-,k^-))} q(x_n^k|\mathbf{y}_{0:n-1}). \quad (52)
\end{aligned}
$$

Furthermore, one can easily show that (50)-(51) reflect the transition from the Gaussian pdf, $q(x_n^k|\mathbf{y}_{0:n}) = q(x_n^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n})$, to the Gaussian pdf, $q(x_{n+1}^k|\mathbf{y}_{0:n}) = q(x_{n+1}^k|\mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-}, \mathbf{y}_{0:n})$, by the following classical marginalization formula (Markovian step):

$$
q(x_{n+1}^k|\mathbf{y}_{0:n}) = \int \underbrace{p(x_{n+1}^k|x_n^k, \mathbf{x}_n^{k^-} = \overline{\mathbf{x}}_{n|n}^{k^-})}_{\mathcal{N}_{x_{n+1}^k}(F_n^{k,k}x_n^k+\mathbf{F}_n(k,k^-)\overline{\mathbf{x}}_{n|n}^{k^-},Q_n^{k,k})} q(x_n^k|\mathbf{y}_{0:n})dx_n^k.
$$

2) Practical implementation: Unlike the VBSKF algorithm above, only (49) needs to be iterated to provide an approximation of $\overline{x}_{n|n}^k$, since it involves $\overline{\mathbf{x}}_{n|n}^{k^-}$. Nonetheless, $\overline{x}_{n+1|n}^k$ in (50) is exactly known once $\overline{\mathbf{x}}_{n|n}$ is available. Furthermore, as mentioned in § IV-A2 above, the computation of $\overline{x}_{n|n}^k$ at iteration i by (49) requires updating only $\mathbf{H}_n(:, k-1)\overline{x}_{n|n}^{k-1}$ rather than all the terms of $\xi_n^{\mathbf{Y}} = \sum_{\substack{j=1 \\ j \neq k}}^{n_{\mathbf{x}}} \mathbf{H}_n(:, j)\overline{x}_{n|n}^j$. We thus obtain another iterative KF-like scheme that involves the prediction estimate of the system state with a diagonal covariance (cf. Algorithm 2).

The number of flops required at each time instant is approximately

$$
\mathcal{C}_{VBPKF} = 5n_{\mathbf{x}}^2 + \left(2n_{\mathbf{y}}^2 + (7I+4)n_{\mathbf{y}} + I + 8\right)n_{\mathbf{x}} + \frac{2}{3}n_{\mathbf{y}}^3 + \frac{3}{2}n_{\mathbf{y}}^2 - \frac{n_{\mathbf{y}}}{6}. \quad (53)
$$

C. Comments and remarks

The introduced algorithms provide online deterministic approximations of the filtering estimate (2) of the system state with diagonal covariance matrices. VBSKF involves the smoothing estimate between two successive filtering estimates and VBPKF involves the prediction estimate. The loss of correlations between the components of the system state, is inherently partially compensated by their functional dependence through the equations of the means. Furthermore, splitting the system state, $\mathbf{x}_n$, to $n_{\mathbf{x}}$ components by making them conditionally independent can be easily extended to dividing it into $\frac{n_{\mathbf{x}}}{n_{\mathbf{s}}} \in \mathbb{N}$ parts $\mathbf{s}_n^k$ of size $n_{\mathbf{s}}$, by introducing conditional independence only between these parts. In such a case, the diagonal condition under $\mathbf{Q}_n$ can be relaxed by assuming that $\mathbf{Q}_n$ is block diagonal only with $n_{\mathbf{s}} \times n_{\mathbf{s}}$ blocks. Below are some comparative remarks about the proposed algorithms:

Algorithm 2 VB Prediction-based KF (VBPKF)

Let $\widetilde{\mathbf{H}}_n = \mathbf{H}_n^T \mathbf{R}_n^{-1}$, $\mathbf{d}_{h,n} = \mathrm{diag}(\widetilde{\mathbf{H}}_n \mathbf{H}_n)$, $\mathbf{d}_{f,n} = \mathrm{diag}(\mathbf{F}_n^T \mathbf{Q}_n^{-1} \mathbf{F}_n)$.

- Prediction. For $k = 1, 2, \cdots, n_{\mathbf{x}}$,

$$\overline{x}_{n|n-1}^k = \mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n-1},$$

$$v_{n|n-1}^k = (F_{n-1}^{k,k})^2 v_{n-1|n-1}^k + Q_{n-1}^{k,k}.$$

- Filtering.

  - *Variances.* For $k = 1, 2, \cdots, n_{\mathbf{x}}$,

$$\eta_n^k = d_{h,n}^k + \left(v_{n|n-1}^k\right)^{-1},$$

$$v_{n|n}^k = \left[\eta_n^k + d_{f,n}^k - (F_n^{k,k})^2 (Q_n^{k,k})^{-1}\right]^{-1}.$$

  - *Means.* We proceed with iterations starting from $\overline{\mathbf{x}}_{n|n} = \overline{\mathbf{x}}_{n|n-1}$. Let $\mathbf{z}_n^{\mathbf{y}} = \mathbf{H}_n \overline{\mathbf{x}}_{n|n}$ and $\mathbf{c}_{n|n-1} = (\mathrm{Diag}(\boldsymbol{\eta}_n))^{-1} \overline{\mathbf{P}}_{n|n-1}^{-1} \overline{\mathbf{x}}_{n|n-1}$. For $iter = 1, 2, \cdots, I; k = 1, 2, \cdots, n_{\mathbf{x}}$,

$$\xi_n^{\mathbf{y}} = \mathbf{z}_n^{\mathbf{y}} - \mathbf{H}_n(:,k)\overline{x}_{n|n}^k,$$

$$\overline{x}_{n|n}^k = c_{n|n-1}^k + \left(\eta_n^k\right)^{-1} \widetilde{\mathbf{H}}_n(k,:)(\mathbf{y}_n - \xi_n^{\mathbf{y}}),$$

$$\mathbf{z}_n^{\mathbf{y}} = \xi_n^{\mathbf{y}} + \mathbf{H}_n(:,k)\overline{x}_{n|n}^k.$$

- In VBSKF the effect of enforced independence (29) between the smoothing pdf $p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n})$ and the filtering pdf $p(\mathbf{x}_n|\mathbf{y}_{0:n})$ arises from the lack of connection between their covariance matrices calculated by (37) and (39), respectively. This independence is however compensated by the link between their means through (38) and (40). Furthermore, (37) and (39) emphasize the fact that the covariances depend only on the model parameters, $\mathbf{F}_n$, $\mathbf{Q}_n$, $\mathbf{H}_n$ and $\mathbf{R}_n$. The covariances are thus unchanged, $\forall n \in \mathbb{N}$, for fixed model parameters; such a property has been suggested by [38, chapter 7, § 7.4] as an inconsistency of approximation (29) with the KF. However, the approximation (31) that leads to VBPKF conserves dependence between the prediction pdf $p(\mathbf{x}_n|\mathbf{y}_{0:n-1})$ and the filtering pdf and is thus more consistent with the KF. This can be seen from (48) and (51) in which the covariance changes depending on whether or not the model parameters are time-dependent.

- The computational costs of the VBSKF, VBPKF and EnKF are of the same order, $\mathcal{O}(n_{\mathbf{x}}^2)$, and are much smaller than that of the KF, $\mathcal{O}(n_{\mathbf{x}}^3)$. On the other hand, for very large state dimension problems

(compared to $n_{\mathbf{y}}$, M and I), one has

$$\mathcal{C}_{EnKF} \overset{(20)}{\approx} 2Mn_{\mathbf{x}}^2,$$

$$\mathcal{C}_{VBSKF} \overset{(44)}{\approx} (9I+5)\,n_{\mathbf{x}}^2,$$

$$\mathcal{C}_{VBPKF} \overset{(53)}{\approx} 5n_{\mathbf{x}}^2.$$

This means that despite its iterative nature, the VBPKF has almost a fixed computational cost, which is lower than those of the VBSKF and EnKF. One can see that the VBPKF cost is approximately $\frac{2}{5}M$ less than that of the EnKF. However, the VBSKF, whose cost depends on I, may be more expensive than that of the EnKF when I approximately exceeds $\frac{2M-5}{9}$.

The gain in computational efficiency of using the VB approach generally comes at the cost of loss of estimation accuracy (see e.g. [63] [37]). When the partitions ($\mathbf{s}_n^k$ or x_n^k) of the state $\mathbf{x}_n$ are weakly correlated *a posteriori*, one can intuitively expect a good accuracy of the VB-like approximation (and nearly perfect if these partitions are almost independent[5]). In contrast, when these partitions become strongly correlated, "the assumption" of conditional independence becomes too strong and the approximation accuracy is expected to degrade. Based on this, the number of partitions, $n_{\mathbf{s}}$, should be chosen judiciously to ensure as good as possible a tradeoff between accuracy and computational efficiency. The larger $n_{\mathbf{s}}$ is, the more important computational efficiency is, but in the same time, the more dependence between the state partitions will be lost which leads to a poorer approximation. Thereby, the guideline principle should be to choose $n_{\mathbf{s}}$ large enough to only achieve an acceptable computational complexity.

D. Extension to the case of an off-diagonal covariance $\mathbf{Q}_n$

The derivation of our algorithms relies on the assumption that the covariance, $\mathbf{Q}_n$, of the input noise is diagonal, for all n, which is common in large scales applications. In our Gaussian case, this ensures that at any time n, conditionally on the previous state, $\mathbf{x}_{n-1}$, the components of the current state, $\{x_n^k\}_{k=1}^{n_{\mathbf{x}}}$, are independent (see (58) below). However, unless the state transition matrix, $\mathbf{F}_n$, is diagonal, this assumption does not "absolutely" break the link between components, $\{x_n^k\}_{k=1}^{n_{\mathbf{x}}}$, since $p(\mathbf{x}_n) \neq \prod_{k=1}^{n_{\mathbf{x}}} p(x_n^k)$ still hold. The assumption that $\mathbf{Q}_n$ is diagonal has been considered, for instance, in an application of dynamic tomography [42] for which pixels of the image at any time are assumed independent conditionally on the previous image.

---

[5]This can happen in the case for which the state transition matrix $\mathbf{F}_n$ and the measurement matrix $\mathbf{H}_n$ are nearly orthogonal, and the input noise transition covariance $\mathbf{Q}_n$ and the measurement noise covariance $\mathbf{R}_n$ are diagonal.

In the more general case of an off-diagonal covariance, $\mathbf{Q}_n$, the conditional independence property (58) is no longer valid. One way to tackle this issue[6] is to approximate, using the VB approach, the joint transition pdf, $p(\mathbf{x}_n|\mathbf{x}_{n-1})$, with a separable product of marginal pdfs, $\prod_{k=1}^{n_{\mathbf{x}}} q(x_n^k|\mathbf{x}_{n-1})$. Indeed, similarly to (23) one obtains,

$$q(x_n^k|\mathbf{x}_{n-1}) \overset{(9)}{\propto} \exp\left(\Lambda_{n-1}^{k,k}(x_n^k)^2 - 2x_n^k\left[\Lambda_{n-1}(k,:)\mathbf{F}_{n-1}\mathbf{x}_{n-1} - \Lambda_{n-1}(k,k^-)\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{x}_{n-1})}[\mathbf{x}_n^{k^-}]\right]\right), \quad (54)$$

with $\Lambda_{n-1} = \mathbf{Q}_{n-1}^{-1}$ ($\mathbf{Q}_n$ is assumed to be positive-definite, $\forall n$). Approximating $\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{x}_{n-1})}[\mathbf{x}_n^{k^-}]$ with $\mathbb{E}_{p(\mathbf{x}_n^{k^-}|\mathbf{x}_{n-1})}[\mathbf{x}_n^{k^-}] = \mathbf{F}_{n-1}(k^-,:)\mathbf{x}_{n-1}$, then using Prop. 3 (cf. Appendix) one obtains,

$$q(x_n^k|\mathbf{x}_{n-1}) \approx \mathcal{N}_{x_n^k}\left(\mathbf{F}_{n-1}(k,:)\mathbf{x}_{n-1}, \left(\Lambda_{n-1}^{k,k}\right)^{-1}\right), \ k = 1, 2, \cdots, n_{\mathbf{x}}. \quad (55)$$

In (55), the variances $\left(\Lambda_{n-1}^{k,k}\right)^{-1}$, which are not equal to $Q_{n-1}^{k,k}$, are computed from all entries of $\mathbf{Q}_{n-1}$. This is important since it emphasizes the fact that although the covariance of approximated transition pdf, $\prod_{k=1}^{n_{\mathbf{x}}} q(x_n^k|\mathbf{x}_{n-1})$, is diagonal, it neverthless involves both diagonal and off-diagonal entries of the original covariance, $\mathbf{Q}_{n-1}$.

Now, using (55), the VBSKF and VBPKF remain valid if one replaces $Q_{n-1}^{k,k}$ by $\left(\Lambda_{n-1}^{k,k}\right)^{-1}$. Neverthless, this requires inverting $\mathbf{Q}_{n-1}$, and therefore the computational costs of these algorithms are no longer of order $\mathcal{O}(n_{\mathbf{x}}^2)$, as highlighted above, but $\mathcal{O}(n_{\mathbf{x}}^3)$, just like the KF and EnKF. One situation in which the cubic costs of the proposed algorithms could be reduced is the particular case of a time-invariant covariance, $\mathbf{Q}$. Indeed, in such a situation, $\mathbf{Q}$ needs to be inverted at the initial time only to compute $\{\Lambda^{k,k}\}_{k=1}^{n_{\mathbf{x}}}$; these are then used at any time $n \geq 0$. Accordingly, at any time $n \geq 1$, the costs of VBSKF and VBPKF become almost equal to the quadratic costs $\mathcal{C}_{VBSKF}$ and $\mathcal{C}_{VBPKF}$, respectively.

V. NUMERICAL RESULTS

We present results from numerical experiments that we performed to evaluate the performance of the proposed filters. We consider a time-invariant, linear and Gaussian state-space system for which:

- $n_{\mathbf{x}} = 1000$, $n_{\mathbf{y}} = 100$, $\widehat{\mathbf{x}}_0 = \mathbf{0}_{n_{\mathbf{x}} \times 1}$ and $\mathbf{P}_0 = \mathbb{I}_{n_{\mathbf{x}}}$.

---

[6]One can also proceed, similarly to [14, chapter 11], with a state-space transformation by building a new model for which the input noise is with a diagonal covariance. However, this leads to a cubic cost, $\mathcal{O}(n_{\mathbf{x}}^3)$, since it requires, among other matrices operations, the use of Cholesky decomposition to factorize $\mathbf{Q}_n$.

- The state transition matrix, $\mathbf{F}$, which is assumed to be a Toeplitz and circulant matrix, is defined from its first row and column as (cf. Figure 1):

$$\mathbf{F}(1,:) = [\mathbf{f}^T, \mathbf{f}^T, \mathbf{f}^T, \mathbf{f}^T, \mathbf{0}_{1\times600}],$$

$$\mathbf{F}(:,1) = [0.1, \mathbf{0}_{1\times697}, \widetilde{\mathbf{f}}^T, \widetilde{\mathbf{f}}^T, \widetilde{\mathbf{f}}^T, 0.02, 0.05]^T,$$

where $\mathbf{f} = [0.1, 0.05, 0.02, \mathbf{0}_{1\times97}]^T$ and $\widetilde{\mathbf{f}} = [0.02, 0.05, 0.1, \mathbf{0}_{1\times97}]^T$.

- The input noise covariance matrix, $\mathbf{Q}$, is assumed to be diagonal equal $\mathbb{I}_{n_\mathbf{x}}$. The off-diagonal case will be considered later.

- Each row, $\mathbf{H}(k,:)$, $k = 1, 2, \cdots, n_\mathbf{y}$, of the measurement matrix is defined as (cf. Figure 2),

$$\mathbf{H}(k,:)\mathbf{x}_n = \sum_{j=0}^{9}(1 - \frac{j}{10})x_n^{k+10j} - 0.1\sum_{\ell=k+896}^{k+900} x_n^\ell.$$

- The measurement noise covariance matrix, $\mathbf{R}$, is assumed to be diagonal with, $\mathbf{R} = \sigma^2 \times \mathbb{I}_{n_\mathbf{y}}$. Instead of testing the sensibility using the variance, σ^2, we will use the Signal-to-Noise Rate (SNR), which can be defined as,

$$\text{SNR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1}||\mathbf{H}\mathbf{x}_n||_2}{\sum_{n=0}^{N-1}||\mathbf{v}_n||_2}\right),$$

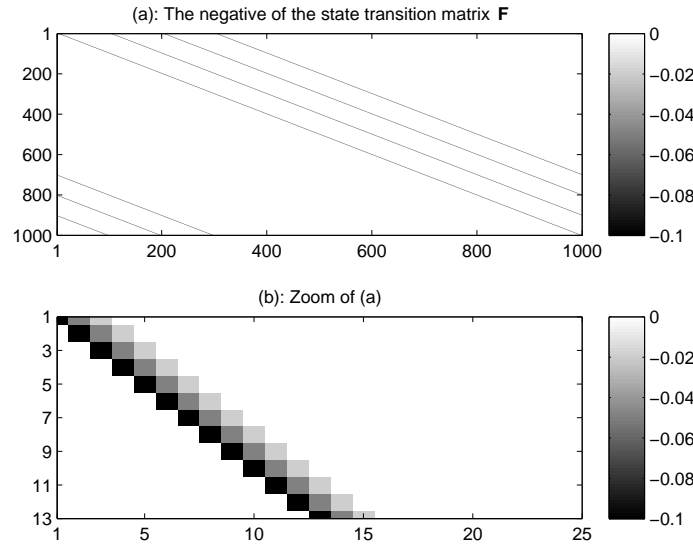where $||\mathbf{v}_n||_2 = n_\mathbf{y}\sigma^2$ (on average).

Figure 1: The negative of the state transition matrix $\mathbf{F}$. The values of non-zero (main and secondary) diagonal bands are identical. Sub-Fig. (b) shows the (main) band associated to $-\mathbf{F}(1:13,1:25)$.
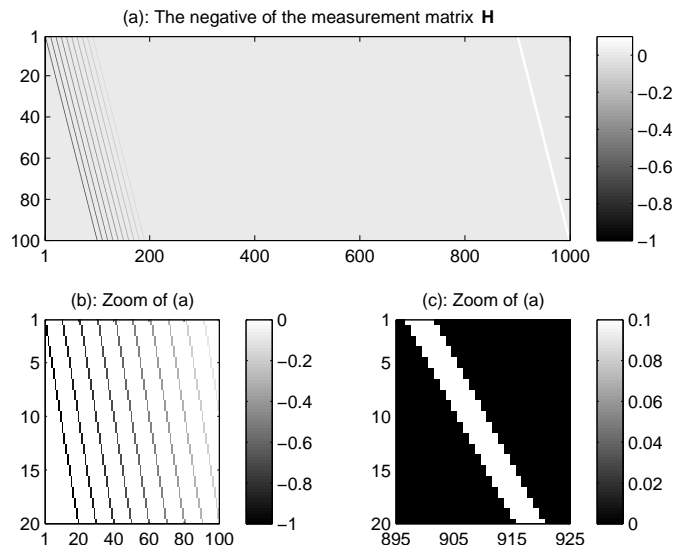
Figure 2: The negative of the measurement matrix $\mathbf{H}$. Sub-Fig. (b) and (c) plot the (non-zero) parts $-\mathbf{H}(1{:}20,1{:}100)$ and $-\mathbf{H}(1{:}20,895{:}925)$, respectively.

To avoid any favourable situation for the proposed algorithms, the matrices $\mathbf{F}$ and $\mathbf{H}$ have been chosen to be well structured and clearly non-orthogonal (see Section IV-C). In our experiments, we proceed as follows:

- we independently simulate, using the state-space system described above, $S = 30$ state processes, $\{\mathbf{x}_0(s), \mathbf{x}_1(s), \cdots, \mathbf{x}_{N-1}(s)\}_s$, and observation processes, $\{\mathbf{y}_0(s), \mathbf{y}_1(s), \cdots, \mathbf{y}_{N-1}(s)\}_s$, with $N = 50$ (we consider 50 observations $\mathbf{y}_n$);

- for each simulation, $s = 1, 2, \cdots, S$, we estimate the system state, $\mathbf{x}_n(s)$, from the observations, $\mathbf{y}_{0:n}(s)$, $n = 0, 1, \cdots, N-1$, using the VBPKF, VBSKF, EnKF and the KF, which serves as a benchmark solution.

All the results are then averaged over the number of simulations, S.

A. Convergence illustration

We consider SNR = 20 dB. Figure 3 plots the evolution of the Kullback-Leibler divergence, KLD, as function of iterations, of the filtering pdf computed by the KF from its approximations computed by the VBPKF and VBSKF for instances $n = 0, 15, 30, 49$. Note that KLD can be exactly evaluated in the Gaussian case (see e.g. [64]). Apart from the initial time instant in which the VBSKF is slightly less

accurate than the VBPKF, the KLDs resulting from these algorithms are almost equal. Furthermore, the minimum of KLD at any time is often reached after about 10 iterations only.
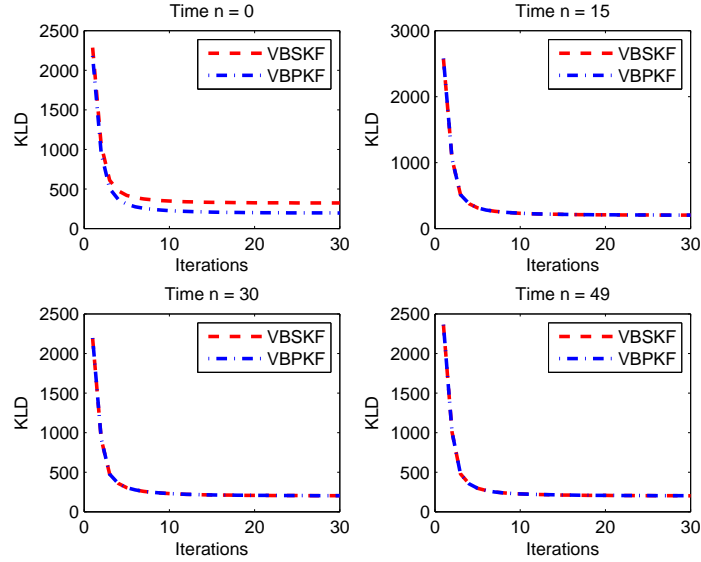

Figure 3: KLD of the filtering pdf given by the KF from its approximations given by VBPKF and VBSKF.

B. Performance Study

We assess here the performance of the proposed algorithms w.r.t. the KF and the EnKF (with perturbed observations), by setting $I = 10$ iterations based on the above results. The performances are measured using the empirical MSE of the filtering estimates, and/or the empirical standard deviation which is defined by,

$$\mathcal{D} = \frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{1}{S} \sum_{s=1}^{S} \| \mathbf{x}_n(s) - \mathbf{x}_{n|n}(s) \|^2 \right]^{\frac{1}{2}},$$

where $\mathbf{x}_{n|n}(s)$ denotes the filtering estimate of $\mathbf{x}_n(s)$ by one of the algorithms. Unlike the optimal KF, the performance of the sub-optimal EnKF depends on some parameters (localization and inflation parameters, among others), as we will see in the section below in which we investigate the effect of these parameters on the EnKF before we compare it with our schemes.

1) Sensitivity of the EnKF to localization and inflation parameters: The EnKF runs with $M = 500$ members. Although M seems to be small compared to the state dimension, $n_{\mathbf{x}}$, it has been proven in a wide range of geophysical applications that M of the order $\mathcal{O}(100)$ provides good performances, even

when the state-space system is nonlinear [9] [52] [32]. The success of the EnKF in the situations for which $M < n_{\mathbf{x}}$, relies, in part, on two very popular strategies commonly used to overcome the undesirable effects that are usually encountered in these situations: i) the covariance localization strategy which tackles the rank-deficiency and spuriously large cross-correlations between distant state variables in the ensemble covariance matrix [65], and ii) the covariance inflation strategy which tackles the underestimation of the sample error variances associated with the use of small ensemble size, among other neglected uncertainties [66].

Let us now turn back to our system, and use the localization and inflation techniques in the EnKF with $M = 500$. The localization is performed using the fifth-order correlation function given in [67, Eq. (4.10)]. Figure 4 shows the contour map of the empirical standard deviation, $\mathcal{D}$, as a function of the localization and inflation parameters. Overall, the values of $\mathcal{D}$ over the whole range of localization and inflation parameters vary within a small interval, [31.07,31.14], which suggests that the performance of the EnKF is not very sensitive to these parameters. The minimum value of $\mathcal{D}$ corresponds to localization and inflation parameters equal 12 and 1.5, respectively; such values will be used in the next section to compare the EnKF w.r.t. the KF and the proposed schemes.
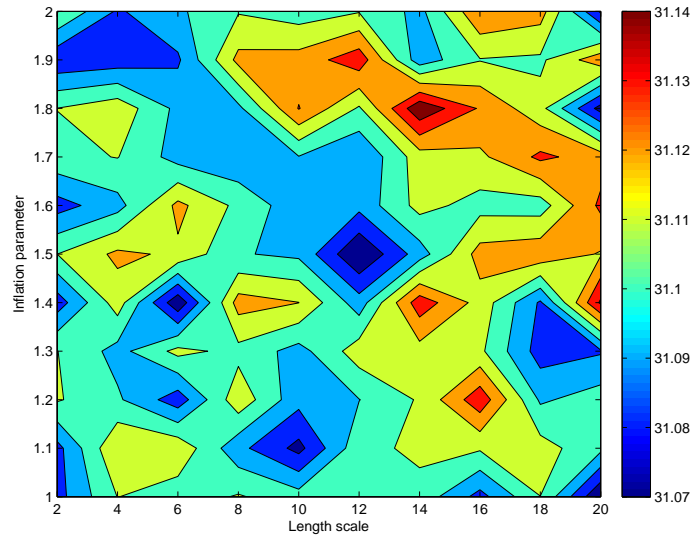

Figure 4: Standard deviation for the filtering estimates resulting from the EnKF with $M = 500$ for different localization and inflation parameters, in the case of $\mathbf{Q}$ diagonal. The minimum occurs at the point (12,1.5).

2) Performance comparison: Figure 5 plots the first four components of the true state and their estimates as they result from the KF, EnKF, VBPKF and VBSKF. Overall, all the algorithms provide very good estimates. To distinguish their performances, the empirical MSE for the filtering estimates is plotted in Figure 6. The proposed filters are slightly less accurate than the EnKF, which provides MSEs closer to those of the KF. Our filters are however much faster than the EnKF; the VBPKF is about 43 times faster and the VBSKF is about 12 times faster.

To study the sensitivity of the proposed algorithms to the SNR, we display in Table I the standard deviation, $\mathcal{D}$, for different values of SNR (recall that so far, SNR = 20 dB). As expected, the values of $\mathcal{D}$ resulting from the four algorithms are inversely proportional to SNR, and the KF always outperforms EnKF, VBPKF and VBSKF. On the other hand, the proposed filters become slightly more accurate than the EnKF for low SNR values (starting from 15 dB).

We finally consider the case of an off-diagonal input noise covariance matrix, $\mathbf{Q}$ (Section IV-D above). For that purpose, the entries of $\mathbf{Q}$ are defined as,

$$Q^{k,\ell} = \exp\left(-\frac{|k-\ell|}{10}\right), \quad k,\ell = 1, 2, \cdots, n_{\mathbf{x}}.$$

Similarly to Table I, Table II plots $\mathcal{D}$ for different values of SNR. The results of the proposed filters were obtained using $I = 10$ iterations, while for the EnKF, $M = 500$ members, and the localization and inflation parameters were set to 16 and 1.2, respectively (these "optimal" parameters have been chosen by trial and error as in Section V-B1 above). Once again, the values of $\mathcal{D}$ as resulting from all the filters are inversely proportional to SNR, and the KF outperforms the other filters, as expected. The proposed filters are always slightly less accurate than the EnKF, with the performances becoming very close for low SNR values. The lower accuracy of the proposed filters w.r.t. the EnKF may originate from the fact that the EnKF uses the true (exact) covariance, $\mathbf{Q}$, while the proposed filters use an approximate diagonal covariance matrix, that is estimated from $\mathbf{Q}$ using the strategy presented in Section IV-D above.

Table I: Empirical standard deviations for the filtering estimates ($\mathbf{Q}$ diagonal).

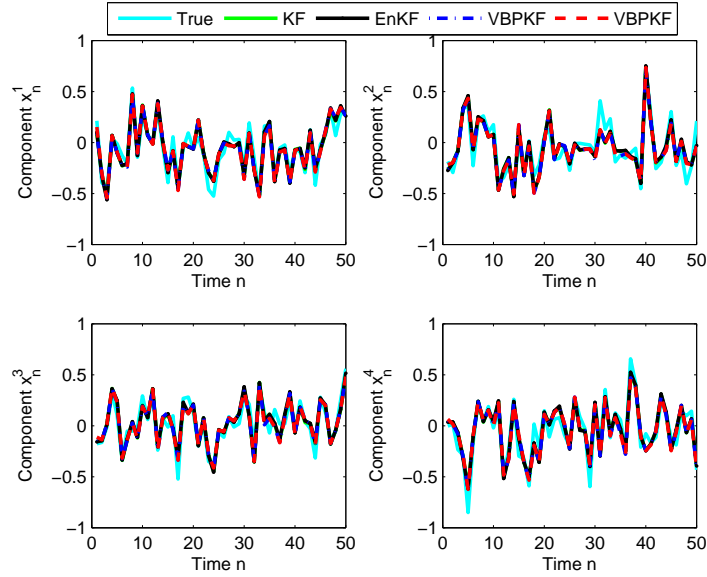| SNR (dB)	25	20	15	10	5	1
KF	30.9220	31.0204	31.1130	31.4251	31.8386	32.0699
EnKF	30.9716	31.0732	31.1708	31.4927	31.9241	32.1556
VBPKF	31.0559	31.1021	31.1641	31.4383	31.8462	32.0786
VBSKF	31.1033	31.1705	31.1902	31.4536	31.8539	32.0815

Figure 5: Tracking of the true state and KF estimate by those obtained by VBPKF, VBSKF and the EnKF with $M = 500$.
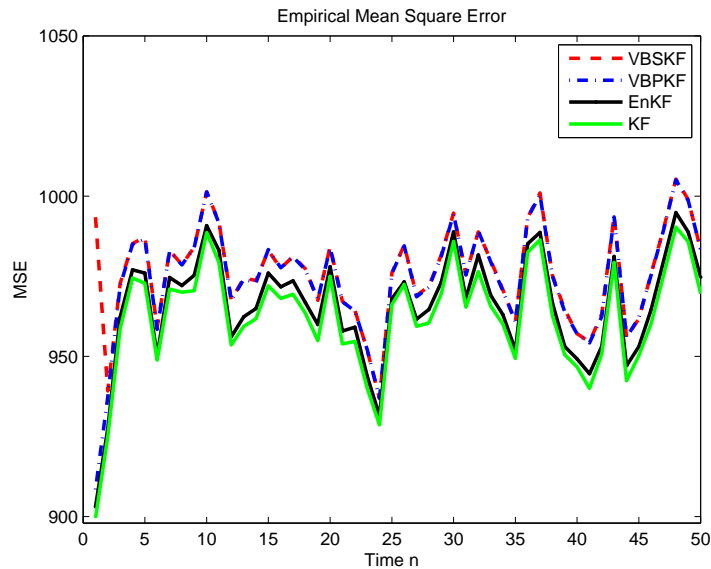

Figure 6: Empirical MSE associated with KF, VBPKF, VBSKF and EnKF with $M = 500$.

Table II: Empirical standard deviations for the filtering estimates ($\mathbf{Q}$ off-diagonal).

SNR (dB)	25	20	15	10	5	1
KF	31.8510	31.8825	31.9735	32.1537	32.2641	32.3662
EnKF	32.9819	32.0256	32.0985	32.2776	32.4269	32.5274
VBPKF	32.1165	32.2125	32.2602	32.2979	32.4305	32.5286
VBSKF	32.1169	32.1950	32.2713	32.2857	32.4289	32.5281

VI. CONCLUSION

We proposed two new approximate iterative Kalman Filters (KF) based on the Variational Bayesian (VB) approach for linear and Gaussian state-space models with very large state dimensions, $n_{\mathbf{x}}$. While sharing the particularity of propagating state estimates with *diagonal* covariance matrices, they differ in the way the successive filtering estimates are linked. The VB Smoothing-based KF (VBSKF) involves a smoothing estimate while the VB Prediction-based KF (VBPKF) involves a prediction estimate. The diagonal property of the covariance matrices enables a squared computational cost in $n_{\mathbf{x}}$ compared to a cubic cost in the KF. Moreover, notwithstanding the iterative character of VBPKF, its implementation cost is always around $\frac{2}{5}M$ times less than that of the Ensemble KF (EnKF) for a given ensemble size M; the comparison between the cost of the VBSKF, which depends on the number of iterations, I, and that of the EnKF, is a function of the value of I w.r.t. M. We studied the performances of our filters and compared them with those of the KF and the EnKF with localization and inflation, through a numerical example. For this example, the proposed filters show performances that are comparable to those of the KF and EnKF, in both cases of diagonal and off-diagonal input noise covariance. Furthermore, the proposed filters slightly outperform the EnKF in the diagonal case for low Signal-to-Noise Rate (SNR) values, while the EnKF becomes slightly more accurate in the off-diagonal case for all SNR values. Applying the proposed algorithms to real-world data and establishing theoretic convergence results are important directions for future work. We will also consider this framework in the context of Kalman-like smoothing focusing on computational cost and storage capacity. Finally, we are currently exploring the same idea of splitting the state vector using the VB approach, in the nonlinear state-space model case following the framework of particle filters.

APPENDIX A

SOME USEFUL PROPERTIES OF GAUSSIAN PDFS

The derivations in this paper rely on the following properties of Gaussian random variables, which are recalled for convenience.

Proposition 1 ([21, Prop. 7]): Let $p(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{P})$ and $p(\mathbf{y}|\mathbf{x}) = \mathcal{N}_{\mathbf{y}}(\mathbf{A}\mathbf{x} + \mathbf{b}, \Sigma)$. Further define the information matrix Γ and information vector ν associated with $p(\mathbf{y}|\mathbf{x})$ as

$$\nu = \mathbf{A}^T \Sigma^{-1} (\mathbf{y} - \mathbf{b}),$$

$$\Gamma = \mathbf{A}^T \Sigma^{-1} \mathbf{A}.$$

Then, $p(\mathbf{x}|\mathbf{y}) = \mathcal{N}_{\mathbf{x}}(\widehat{\mathbf{x}}_{|\mathbf{y}}, \mathbf{P}_{|\mathbf{y}})$ with

$$\mathbf{P}_{|\mathbf{y}}^{-1} = \mathbf{P}^{-1} + \Gamma,$$

$$\mathbf{P}_{|\mathbf{y}}^{-1} \widehat{\mathbf{x}}_{|\mathbf{y}} = \mathbf{P}^{-1} \mu + \nu.$$

Proposition 2 ([21, Prop. 11]): Let $(\mathbf{x}, \underbrace{(\mathbf{y}_1, \mathbf{y}_2)}_{\mathbf{y}})$ be Gaussian in which, conditionally on $\mathbf{x}$, $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent. Let $p(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{P})$ and let ν_i and Γ_i be the information parameters of $p(\mathbf{y}_i|\mathbf{x})$. Then,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}_1|\mathbf{x})p(\mathbf{y}_2|\mathbf{x})p(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\widehat{\mathbf{x}}_{|\mathbf{y}}, \mathbf{P}_{|\mathbf{y}});$$

$$\mathbf{P}_{|\mathbf{y}}^{-1} \widehat{\mathbf{x}}_{|\mathbf{y}} = \nu_1 + \nu_2 + \mathbf{P}^{-1} \mu,$$

$$\mathbf{P}_{|\mathbf{y}}^{-1} = \Gamma_1 + \Gamma_2 + \mathbf{P}^{-1}.$$

Proposition 3: Let $p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\mathcal{J}\right)$ with $\mathcal{J} = \mathbf{x}^T \mathbf{P}^{-1} \mathbf{x} - 2\mathbf{x}^T \mathbf{P}^{-1} \mu$. Then, $p(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{P})$.

Proposition 4: Let $p(\mathbf{x}, \mathbf{y}) \propto \exp\left(-\frac{1}{2}\mathcal{J}\right)$ with

$$\mathcal{J} = \mathbf{x}^T[\mathbf{P}^{-1} + \mathbf{A}^T \Sigma^{-1} \mathbf{A}]\mathbf{x} + \mathbf{y}^T \Sigma^{-1} \mathbf{y} - 2\mathbf{x}^T \mathbf{A}^T \Sigma^{-1} \mathbf{y} - 2\mathbf{x}^T[\mathbf{P}^{-1}\mu - \mathbf{A}^T \Sigma^{-1}\mathbf{b}] - 2\mathbf{y}^T \Sigma^{-1}\mathbf{b}.$$

Then, $p(\mathbf{x}, \mathbf{y})$ is Gaussian with $p(\mathbf{x}) = \mathcal{N}_{\mathbf{x}}(\mu, \mathbf{P})$ and $p(\mathbf{y}) = \mathcal{N}_{\mathbf{y}}(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\mathbf{P}\mathbf{A}^T + \Sigma)$.

APPENDIX B

PROOF OF EQUATIONS (37)-(40)

We show here that (37)-(40) are obtained from (35) and (36). Let us start with an explicit derivation of (36). From (9)-(10), we have

$$p(x_n^k|\mathbf{x}_{n-1}) = \mathcal{N}_{x_n^k}(\mathbf{F}_{n-1}(k,:)\mathbf{x}_{n-1}, Q_{n-1}^{k,k}), \tag{56}$$

$$p(\mathbf{y}_n|\mathbf{x}_n) = \mathcal{N}_{\mathbf{y}_n}\left(\underbrace{\mathbf{H}_n(:,k)x_n^k + \mathbf{H}_n(:,k^-)\mathbf{x}_n^{k^-}}_{\mathbf{H}_n\mathbf{x}_n}, \mathbf{R}_n\right). \tag{57}$$

Taking the expected value of $\ln\left[p(x_n^k|\mathbf{x}_{n-1})\right]$ w.r.t $q(\mathbf{x}_{n-1}|\mathbf{y}_{0:n})$ and of $\ln\left[p(\mathbf{y}_n|\mathbf{x}_n)\right]$ w.r.t. $q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})$, the r.h.s. of (36) becomes proportional to

$$\exp\left(-\frac{1}{2}\left[\left(\left(Q_{n-1}^{k,k}\right)^{-1}+\Gamma_{n,k}\right)\left(x_n^k\right)^2-2x_n^k\left(\mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n}+\mu_{n,k}\right)\right]\right),$$

with $\Gamma_{n,k}=\mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}\mathbf{H}_n(:,k)$ and $\mu_{n,k}=\mathbf{H}_n^T(:,k)\mathbf{R}_n^{-1}\left(\mathbf{y}_n-\mathbf{H}_n(:,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}\right)$. Therefore, using Prop. 3, we obtain a Gaussian pdf, $q(x_n^k|\mathbf{y}_{0:n})$, whose parameters are given in (39) and (40).

Concerning the derivation of (35), we start with the following factorization that directly arises from model (1):

$$p(\mathbf{x}_n|\mathbf{x}_{n-1})=\prod_{k=1}^{n_\mathbf{x}}p(x_n^k|\mathbf{x}_{n-1}), \tag{58}$$

in which we insert (56) to obtain,

$$\ln\left[p(\mathbf{x}_n|\mathbf{x}_{n-1})\right]=-\frac{1}{2}\left[\sum_{j=1}^{n_\mathbf{x}}\frac{\left(F_{n-1}^{j,k}\right)^2}{Q_{n-1}^{j,j}}\left(x_{n-1}^k\right)^2-2x_{n-1}^k\sum_{j=1}^{n_\mathbf{x}}\frac{F_{n-1}^{j,k}}{Q_{n-1}^{j,j}}\left(x_n^j-\mathbf{F}_{n-1}(j,k^-)\mathbf{x}_{n-1}^{k^-}\right)\right]+C_1, \tag{59}$$

with C_1 independent of x_{n-1}^k. Now, taking the expected value of (59) w.r.t. $q(\mathbf{x}_{n-1}^{k^-},\mathbf{x}_n|\mathbf{y}_{0:n})$, the r.h.s. of (35) becomes proportional to

$$\exp\left(-\frac{1}{2}\left[\left(\sum_{j=1}^{n_\mathbf{x}}\frac{\left(F_{n-1}^{j,k}\right)^2}{Q_{n-1}^{j,j}}+\frac{1}{v_{n-1|n-1}^k}\right)\left(x_{n-1}^k\right)^2-2x_{n-1}^k\left(\sum_{j=1}^{n_\mathbf{x}}\frac{F_{n-1}^{j,k}}{Q_{n-1}^{j,j}}\left(\overline{x}_{n|n}^j-\mathbf{F}_{n-1}(j,k^-)\overline{\mathbf{x}}_{n-1|n}^{k^-}\right)+\frac{\overline{x}_{n-1|n-1}^k}{v_{n-1|n-1}^k}\right)\right]\right).$$

This eventually shows, using Prop. 3, that $q(x_{n-1}^k|\mathbf{y}_{0:n})$ is Gaussian with moments given in (37) and (38). ∎

APPENDIX C

PROOF OF EQUATIONS (48)-(51)

We show here that (48)-(51) arise from (47). Using (56) and (58), we obtain

$$\mathbb{E}_{q(\mathbf{x}_n^{k^-},\mathbf{x}_{n+1}^{k^-}|\mathbf{y}_{0:n})}\left[\ln\left(p(\mathbf{x}_{n+1}|\mathbf{x}_n)\right)\right]=\underbrace{\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})}\left[\ln\left(p(x_{n+1}^k|\mathbf{x}_n)\right)\right]}_{\mathcal{E}^k}$$

$$+\sum_{\substack{j=1\\j\neq k}}^{n_\mathbf{x}}\underbrace{\mathbb{E}_{q(\mathbf{x}_n^{k^-},x_{n+1}^j|\mathbf{y}_{0:n})}\left[\ln\left(p(x_{n+1}^j|\mathbf{x}_n)\right)\right]}_{\mathcal{E}^j}, \tag{60}$$

with

$$\mathcal{E}^k=-\frac{1}{2Q_n^{k,k}}\left[\left(x_{n+1}^k\right)^2+\left(F_n^{k,k}\right)^2\left(x_n^k\right)^2-2x_{n+1}^kF_n^{k,k}x_n^k-2\left(x_{n+1}^k-F_n^{k,k}x_n^k\right)\mathbf{F}_n(k,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}\right]+C_2,$$

$$\mathcal{E}^j=-\frac{1}{2Q_n^{j,j}}\left[\left(F_n^{j,k}\right)^2\left(x_n^k\right)^2-2\left(x_{n+1}^j-\mathbf{F}_n(j,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}\right)F_n^{j,k}x_n^k\right]+C_3,$$

C_2 and C_3 are independent of x_{n+1}^k and x_n^k. On the other hand, from (57), we have

$$\mathbb{E}_{q(\mathbf{x}_n^{k^-}|\mathbf{y}_{0:n})}[\ln(p(\mathbf{y}_n|\mathbf{x}_n))] = -\frac{1}{2}\left[\Gamma_{n,k}\left(x_n^k\right)^2 - 2x_n^k\mu_{n,k}\right] + C_4, \tag{61}$$

with C_4 independent of x_n^k, and $\Gamma_{n,k}$ and $\mu_{n,k}$ are defined in Appendix B. Accordingly, using (60) and (61) and assuming that $q(x_n^k|\mathbf{y}_{0:n-1})$ is Gaussian, the r.h.s. of (47) is proportional to

$$\exp\left(-\frac{1}{2}\left[\frac{(x_{n+1}^k)^2}{Q_n^{k,k}} + \left(\Gamma_{n,k} + \frac{1}{v_{n|n-1}^k} + \sum_{i=1}^{n_{\mathbf{x}}}\frac{(F_n^{i,k})^2}{Q_n^{i,i}}\right)\left(x_n^k\right)^2 - 2\frac{F_n^{k,k}}{Q_n^{k,k}}x_{n+1}^k x_n^k - \frac{2}{Q_n^{k,k}}\mathbf{F}_n(k,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}x_{n+1}^k\right.$$
$$\left. -2\left(\mu_{n,k} + \frac{\overline{x}_{n|n-1}}{v_{n|n-1}^k} - \frac{F_n^{k,k}}{Q_n^{k,k}}\mathbf{F}_n(k,k^-)\overline{\mathbf{x}}_{n|n}^{k^-} + \sum_{\substack{i=1\\i\neq k}}^{n_{\mathbf{x}}}\frac{F_n^{i,k}}{Q_n^{i,i}}\left(\overline{x}_{n+1|n}^i - \mathbf{F}_n(i,k^-)\overline{\mathbf{x}}_{n|n}^{k^-}\right)\right)x_n^k\right]\right). \tag{62}$$

Therefore, using Prop. 4, we can easily verify that $q(x_{n+1}^k|\mathbf{y}_{0:n})$ is Gaussian with parameters given by (50) and (51), and that $q(x_n^k|\mathbf{y}_{0:n})$ is Gaussian with a variance $v_{n|n}^k$ given by (48) and a mean equal to

$$\overline{x}_{n|n}^k = v_{n|n}^k\left[\mu_{n,k} + \frac{\overline{x}_{n|n-1}^k}{v_{n|n-1}^k} + \overbrace{\sum_{\substack{i=1\\i\neq k}}^{n_{\mathbf{x}}}\frac{F_n^{i,k}}{Q_n^{i,i}}(\overline{x}_{n+1|n}^i - \mathbf{F}_n(i,k^-)\overline{\mathbf{x}}_{n|n}^{k^-})}^{\lambda}\right]. \tag{63}$$

Now, using (50), we have $\lambda = \sum_{\substack{i=1\\i\neq k}}^{n_{\mathbf{x}}}\frac{(F_n^{i,k})^2}{Q_n^{i,i}}\overline{x}_{n|n}^k$; moving $v_{n|n}^k \times \lambda$ into the l. h. s. of (63), then multiplying both sides of (63) by $\left(v_{n|n}^k\right)^{-1}$, we eventually obtain (49). The proof is achieved by noticing that the Gaussianity of $q(x_n^k|\mathbf{y}_{0:n-1})$, assumed above, is guaranteed for all n when starting at the initial time from $\prod_{k=1}^{n_{\mathbf{x}}} q(x_0^k|\mathbf{y}_{-1}) = p(\mathbf{x}_0)$ by assuming that $p(\mathbf{x}_0)$ is Gaussian with a diagonal covariance matrix. ∎

APPENDIX D
DETAILS OF COMPUTATIONAL COMPLEXITY FORMULAS

The computational complexities provided in this paper consider each addition, subtraction, multiplication, division, or square-root operation as one flop [68]. They are based on:

- $2d - 1$ flops to compute the inner product, $\mathbf{u}^T.\mathbf{v}$, of two $d \times 1$ vectors, $\mathbf{u}$ and $\mathbf{v}$;

- $2d_1d_2d_3 - d_1d_3$ flops to compute a non-symmetric matrix, $\mathbf{M}_1\mathbf{M}_2$, with $\mathbf{M}_1$ and $\mathbf{M}_2$ are $d_1 \times d_2$ and $d_2 \times d_3$ matrices, respectively. The number of flops reduces (approximately) to the half, say, $d_1d_2d_3 - \frac{d_1d_3}{2}$ when $\mathbf{M}_1\mathbf{M}_2$ is symmetric;

- $\frac{d^3}{3} + d^2 + \frac{2}{3}d$ flops to compute a square-root (and lower triangular or upper triangular) matrix of a $d \times d$ matrix using the Cholesky decomposition (see e.g. [60, page 175] in which only the leading d term is considered);

- $\frac{d^3}{3} + \frac{d^2}{2} + \frac{d}{6}$ flops to invert a lower (or upper) triangular matrix.

The approximate numbers of flops of the proposed algorithms are detailed in Tables III and IV, keeping in mind that $\mathbf{Q}_n$ is assumed diagonal. Those of KF (16) and EnKF (20) can be similarly obtained (see also [69] and [70]). Finally, note that in the EnKF we assumed that the sampling cost of drawing one sample from $\mathcal{N}(0,1)$, which is very small, is equivalent to one flop; such an assumption intervenes only in the secondary terms of $\mathcal{C}_{EnKF}$ (20), which are ignored in the high-dimensional state case as it was discussed in section IV-C above.

Table III: Details of $\mathcal{C}_{VBSKF}$ (44). The inverse $\mathbf{Q}_{n-1}^{-1}$ is explicitly computed in the step S_8 and stored to be used in S_{15}. In S_5 however, the entries of $\widetilde{\mathbf{F}}_{n-1}$ are computed by a division, $\widetilde{F}_{n-1}^{k,l} = \frac{F_{n-1}^{l,k}}{Q_{n-1}^{l,l}}$. The inverses $\left(v_{n-1|n-1}^k\right)^{-1}$ are available from S_8 at previous time recursion, $(n-2) \to (n-1)$, and thus their computation in S_7 and S_{11} is not required. Indeed, in S_8 the addition $\left(Q_{n-2}^{k,k}\right)^{-1} + d_{h,n-1}^k$ is first performed leading to $\left(v_{n-1|n-1}^k\right)^{-1}$, then inverted to obtain $v_{n-1|n-1}^k$.

		Operation	Cost
		$\mathsf{S}_1.\ \mathbf{U}_n = \mathrm{Cholesky}(\mathbf{R}_n)$, *i.e.*, $\mathbf{R}_n = \mathbf{U}_n^T \mathbf{U}_n$	$\frac{n_{\mathbf{y}}^3}{3} + n_{\mathbf{y}}^2 + \frac{2}{3}n_{\mathbf{y}}$
		$\mathsf{S}_2.\ \mathbf{U}_n^{-1}$	$\frac{n_{\mathbf{y}}^3}{3} + \frac{n_{\mathbf{y}}^2}{2} + \frac{n_{\mathbf{y}}}{6}$
		$\mathsf{S}_3.\ \widetilde{\mathbf{H}}_n = \mathbf{H}_n^T \mathbf{R}_n^{-1} = \mathbf{H}_n^T \mathbf{U}_n^{-1}(\mathbf{U}_n^{-1})^T$	$2n_{\mathbf{y}}^2 n_{\mathbf{x}}$
		$\mathsf{S}_4.\ \mathbf{d}_{h,n} = \mathrm{diag}\left(\widetilde{\mathbf{H}}_n \mathbf{H}_n\right)$	$(2n_{\mathbf{y}} - 1)n_{\mathbf{x}}$
		$\mathsf{S}_5.\ \widetilde{\mathbf{F}}_{n-1} = \mathbf{F}_{n-1}^T \mathbf{Q}_{n-1}^{-1}$	$n_{\mathbf{x}}^2$
		$\mathsf{S}_6.\ \mathbf{d}_{f,n-1} = \mathrm{diag}\left(\widetilde{\mathbf{F}}_{n-1} \mathbf{F}_{n-1}\right)$	$2n_{\mathbf{x}}^2 - n_{\mathbf{x}}$
Variances		$\mathsf{S}_7.\ v_{n-1|n}^k = \left[\left(v_{n-1|n-1}^k\right)^{-1} + d_{f,n-1}^k\right]^{-1}, k=1,\cdots,n_{\mathbf{x}}$	$2n_{\mathbf{x}}$
		$\mathsf{S}_8.\ v_{n|n}^k = \left[\left(Q_{n-1}^{k,k}\right)^{-1} + d_{h,n}^k\right]^{-1}, k=1,\cdots,n_{\mathbf{x}}$	$3n_{\mathbf{x}}$
Means		$\mathsf{S}_9.\ \overline{\mathbf{x}}_{n|n} = \mathbf{F}_{n-1}\overline{\mathbf{x}}_{n-1|n-1}$	$2n_{\mathbf{x}}^2 - n_{\mathbf{x}}$
		$\mathsf{S}_{10}.\ \mathbf{z}_n^{\mathbf{y}} = \mathbf{H}_n\overline{\mathbf{x}}_{n|n}$	$2n_{\mathbf{y}}n_{\mathbf{x}} - n_{\mathbf{y}}$
		$\mathsf{S}_{11}.\ \mathbf{c}_{n-1|n} = \overline{\mathbf{P}}_{n-1|n}\overline{\mathbf{P}}_{n-1|n-1}^{-1}\overline{\mathbf{x}}_{n-1|n-1}$	$2n_{\mathbf{x}}$
		For $k=1,\cdots,n_{\mathbf{x}}, iter=1,\cdots,I$,	
		$\mathsf{S}_{12}.\ \quad \boldsymbol{\xi}_n^{\mathbf{x}} = \mathbf{z}_n^{\mathbf{x}} - \mathbf{F}_{n-1}(:,k)\overline{x}_{n-1|n}^k$	$2n_{\mathbf{x}}$
		$\mathsf{S}_{13}.\ \quad \overline{x}_{n-1|n}^k = c_{n-1|n}^k + v_{n-1|n}^k \widetilde{\mathbf{F}}_{n-1}(k,:)\left(\overline{\mathbf{x}}_{n|n} - \boldsymbol{\xi}_n^{\mathbf{x}}\right)$	$3n_{\mathbf{x}} + 1$
		$\mathsf{S}_{14}.\ \quad \boldsymbol{\xi}_n^{\mathbf{y}} = \mathbf{z}_n^{\mathbf{y}} - \mathbf{H}_n(:,k)\overline{x}_{n|n}^k$	$2n_{\mathbf{y}}$
		$\mathsf{S}_{15}.\ \quad \overline{x}_{n|n}^k = v_{n|n}^k \left[\left(Q_{n-1}^{k,k}\right)^{-1}\mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n} + \widetilde{\mathbf{H}}_n(k,:)(\mathbf{y}_n - \boldsymbol{\xi}_n^{\mathbf{y}})\right]$	$2n_{\mathbf{x}} + 3n_{\mathbf{y}} + 1$
		$\mathsf{S}_{16}.\ \quad \mathbf{z}_n^{\mathbf{x}} = \boldsymbol{\xi}_n^{\mathbf{x}} + \mathbf{F}_{n-1}(:,k)\overline{x}_{n-1|n}^k$	$2n_{\mathbf{x}}$
		$\mathsf{S}_{17}.\ \quad \mathbf{z}_n^{\mathbf{y}} = \boldsymbol{\xi}_n^{\mathbf{y}} + \mathbf{H}_n(:,k)\overline{x}_{n|n}^k$	$2n_{\mathbf{y}}$

Table IV: Details of $\mathcal{C}_{VBPKF}$ (53). The inverse $\mathbf{Q}_n^{-1}$ is not explicitly computed in the steps S_5 and S_{10} since in these steps, a division by the entries $Q_n^{k,k}$ is sufficient. The squares $\left(F_n^{k,k}\right)^2$ computed in S_{10} are stored and used at the next time recursion in S_8. The multiplication in S_{12} and S_{14} by the inverses $\left(\eta_n^k\right)^{-1}$ amounts to directly dividing by η_n^k.

		Operation	Cost
		S_1. $\mathbf{U}_n = \text{Cholesky}(\mathbf{R}_n)$, *i.e.*, $\mathbf{R}_n = \mathbf{U}_n^T\mathbf{U}_n$	$\frac{n_{\mathbf{y}}^3}{3} + n_{\mathbf{y}}^2 + \frac{2}{3}n_{\mathbf{y}}$
		S_2. $\mathbf{U}_n^{-1}$	$\frac{n_{\mathbf{y}}^3}{3} + \frac{n_{\mathbf{y}}^2}{2} + \frac{n_{\mathbf{y}}}{6}$
		S_3. $\widetilde{\mathbf{H}}_n = \mathbf{H}_n^T\mathbf{R}_n^{-1} = \mathbf{H}_n^T\mathbf{U}_n^{-1}(\mathbf{U}_n^{-1})^T$	$2n_{\mathbf{y}}^2 n_{\mathbf{x}}$
		S_4. $\mathbf{d}_{h,n} = \text{diag}\left(\widetilde{\mathbf{H}}_n\mathbf{H}_n\right)$	$(2n_{\mathbf{y}} - 1)n_{\mathbf{x}}$
		S_5. $\widetilde{\mathbf{F}}_n = \mathbf{F}_n^T\mathbf{Q}_n^{-1}$	$n_{\mathbf{x}}^2$
		S_6. $\mathbf{d}_{f,n} = \text{diag}\left(\widetilde{\mathbf{F}}_n\mathbf{F}_n\right)$	$2n_{\mathbf{x}}^2 - n_{\mathbf{x}}$
Prediction		S_7. $\overline{x}_{n|n-1}^k = \mathbf{F}_{n-1}(k,:)\overline{\mathbf{x}}_{n-1|n-1}, \ k = 1, \cdots, n_{\mathbf{x}}$	$2n_{\mathbf{x}}^2 - n_{\mathbf{x}}$
		S_8. $v_{n|n-1}^k = \left(F_{n-1}^{k,k}\right)^2 v_{n-1|n-1}^k + Q_{n-1}^{k,k}, \ k = 1, \cdots, n_{\mathbf{x}}$	$2n_{\mathbf{x}}$
Filtering		S_9. $\eta_n^k = d_{h,n}^k + \left(v_{n|n-1}^k\right)^{-1}, \ k = 1, \cdots, n_{\mathbf{x}}$	$2n_{\mathbf{x}}$
		S_{10}. $v_{n|n}^k = \left[\eta_n^k + d_{f,n}^k - \left(F_n^{k,k}\right)^2\left(Q_n^{k,k}\right)^{-1}\right]^{-1}, \ k = 1, \cdots, n_{\mathbf{x}}$	$5n_{\mathbf{x}}$
		S_{11}. $\mathbf{z}_n^{\mathbf{y}} = \mathbf{H}_n\overline{\mathbf{x}}_{n|n}$	$2n_{\mathbf{y}}n_{\mathbf{x}} - n_{\mathbf{y}}$
		S_{12}. $\mathbf{c}_{n|n-1} = (\text{Diag}(\boldsymbol{\eta}_n))^{-1}\overline{\mathbf{P}}_{n|n-1}^{-1}\overline{\mathbf{x}}_{n|n-1}$	$2n_{\mathbf{x}}$
		$\quad$ For $k = 1, \cdots, n_{\mathbf{x}}, \ iter = 1, \cdots, I,$	
		S_{13}. $\quad \boldsymbol{\xi}_n^{\mathbf{y}} = \mathbf{z}_n^{\mathbf{y}} - \mathbf{H}_n(:,k)\overline{x}_{n|n}^k$	$2n_{\mathbf{y}}$
		S_{14}. $\quad \overline{x}_{n|n}^k = c_{n|n-1}^k + \left(\eta_n^k\right)^{-1}\widetilde{\mathbf{H}}_n(k,:)(\mathbf{y}_n - \boldsymbol{\xi}_n^{\mathbf{y}})$	$3n_{\mathbf{y}} + 1$
		S_{15}. $\quad \mathbf{z}_n^{\mathbf{y}} = \boldsymbol{\xi}_n^{\mathbf{y}} + \mathbf{H}_n(:,k)\overline{x}_{n|n}^k$	$2n_{\mathbf{y}}$

REFERENCES

[1] C. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*. Berlin, DE: Springer, 1999.

[2] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: Wiley, 2001.

[3] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, February 2002.

[4] C. Yardim, Z.-H. Michalopoulou, and P. Gerstoft, "An Overview of Sequential Bayesian Filtering in Ocean Acoustics," *IEEE Journal of Oceanic Engineering*, vol. 36, no. 1, pp. 71–89, January 2011.

[5] H. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms," *IEEE Robotics & Automation Magazine*, vol. 13, pp. 99–110, 2006.

[6] F. Gustafsson, "Particle filter theory and practice with positioning applications," *IEEE Transactions on Aerospace and Electronic Systems Magazine*, vol. 25, no. 7, pp. 53–82, 2010.

[7] P. M. Djuric, J. Zhang, T. Ghirmai, Y. Huang, and J. H. Kotecha, "Applications of particle filtering to communications : A review," in *Proceedings of the EUropean SIgnal Processing COnference (EUSIPCO)*, Toulouse, France, September 2002.

[8] S. Haykin, K. Huber, and Z. Chen, "Bayesian Sequential State Estimation for MIMO Wireless Communications," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 439–54, 2004.

[9] G. Evensen, "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99, no. C5, pp. 143–62, 1994.

[10] ——, *Data Assimilation: The Ensemble Kalman Filter*. New York: Springer, 2006.

[11] J. L. Anderson, "Ensemble Kalman filters for large geophysical applications," *IEEE Control Systems Magazine*, vol. 29, no. 3, pp. 66–82, June 2009.

[12] J. Mandel, J. D. Beezley, J. L. Coen, and K. Minjeong, "Data assimilation for wildland fires," *IEEE Control Systems Magazine*, vol. 29, no. 3, pp. 47–65, June 2009.

[13] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering, Series D*, vol. 82, no. 1, pp. 35–45, 1960.

[14] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[15] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*, ser. Prentice Hall Information and System Sciences Series. Upper Saddle River, NJ: Prentice Hall, 2000.

[16] P. Kaminski, A. Bryson, and S. Schmidt, "Discrete square root filtering:A survey of current techniques," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 727–36, 1971.

[17] H. Rauch, "Solutions to smoothing problem," *IEEE Transactions on Automatic Control*, vol. 8, pp. 371–372, October 1963.

[18] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *Journal of the American Institute of Aeronautics and Astronautics*, vol. 3, no. 8, pp. 1445–1450, August 1965.

[19] D. Q. Mayne, "A solution to the smoothing problem for linear dynamical systems," *Automatica*, vol. 4, pp. 73–92, 1966.

[20] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Transactions on Automatic Control*, vol. 7, pp. 387–90, August 1969.

[21] B. Ait-El-Fquih and F. Desbouvries, "On Bayesian Fixed-Interval Smoothing Algorithms," *IEEE Transactions on Automatic Control*, vol. 53, no. 10, pp. 2437–42, November 2008.

[22] J. A. E. Bryson and D. E. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," *IEEE Transactions on Automatic Control*, vol. AC-10, pp. 4–10, January 1965.

[23] H. W. Sorenson, "Kalman filtering techniques," in *Advances in Control Systems Theory and Applications*, C. T. Leondes, Ed. Academic Press, 1966, vol. 3, pp. 219–92.

[24] A. Bryson and L. Henrikson, "Estimation using sampled data containing sequentially correlated noise," *Journal of Spacecraft and Rockets*, vol. 5, pp. 662–65, June 1968.

[25] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. San Diego: Academic Press, 1970, vol. 64.

[26] B. Ait-El-Fquih and F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2957–63, Aug. 2006.

[27] V. Némesin and S. Derrode, "Robust Blind Pairwise Kalman Algorithms Using QR Decompositions," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 5–9, January 2013.

[28] Y. Shmaliy, "An Iterative Kalman-Like Algorithm Ignoring Noise and Initial Conditions," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2465–73, 2011.

[29] T. Kailath, "Some new algorithms for recursive estimation in constant linear systems," *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 750–60, 1973.

[30] M. Morf, G. Sidhu, and T. Kailath, "Some new algorithms for recursive estimation in constant, linear, discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 19, no. 4, pp. 315–23, 1974.

[31] D. Pham, J. Verron, and M. Rouband, "Singular evolutive Kalman filter with EOF initialization for data assimilation in oceanography," *Journal of Marine Systems*, vol. 16, no. 3-4, pp. 323–40, 1998.

[32] I. Hoteit, D. Pham, and J. Blum, "A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific," *Journal of Marine Systems*, vol. 36, no. 1–2, pp. 101–27, 2002.

[33] I. Hoteit and D. Pham, "Evolution of the reduced state space and data assimilation schemes based on the Kalman filter," *Journal of Meteorological Society of Japan*, vol. 81, no. 1, pp. 21–39, 2003.

[34] L. Paninski, "Fast Kalman filtering on quasilinear dendritic trees," *Journal of Computational Neuroscience*, vol. 28, pp. 211–28, 2010.

[35] R. J. Lorentzen and G. Nævdal, "An Iterative Ensemble Kalman Filter," *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1990–95, August 2011.

[36] S. Lakshmivarahan and D. J. Stensrud, "Ensemble Kalman Filter," *IEEE Control Systems Magazine*, vol. 29, no. 3, pp. 34–46, June 2009.

[37] T. Jaakkola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods - Theory and Practice*, M. Opper and D. Saad, Eds.   MIT Press, 2000, pp. 129–59.

[38] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*.   Springer, 2006.

[39] ——, "Variational Bayesian Filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, 2008.

[40] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, March 2009.

[41] M. A. Chappell, A. R. Groves, B. Whitcher, and M. W. Woolrich, "Variational Bayesian Inference for a Nonlinear Forward Model," *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 223–36, January 2009.

[42] B. Ait-El-Fquih and T. Rodet, "Variational Bayesian Kalman Filtering in Dynamical Tomography," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 22-27 2011.

[43] J. Sun, J. Zhou, and X. Gu, "Variational Bayesian Two-Stage Kalman Fliter for Systems with Unknown Inputs," in *Proceedings of the International Workshop on Information and Electronics Engineering (IWIEE)*, Harbin, China, 2012.

[44] J. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *Proceedings of IEEE Int. Conf. Intell. Robots Syst.*, 2007.

[45] B. Ait-El-Fquih and C. Gouy-Pailler, "Backward hidden Markov chain for outlier-robust filtering and fixed-interval smoothing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 26-31 2013.

[46] M. Mansouri, H. Nounou, and M. Nounou, "State Estimation of a Chemical Reactor Process Model - A Comparative Study," in *Proceedings of the IEEE International Multi-Conference on Systems, Signals & Devices (SSD)*, Hammamet, Tunisia, March 18-21 2013.

[47] Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian Methods For Multimedia Problems," *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 1000–17, June 2014.

[48] I. Mbalawata, S. Sarkka, M. Vihola, and H. Haario, "Adaptive Metropolis Algorithm Using Variational Bayesian Adaptive Kalman Filter," *Computational Statistics & Data Analysis*, vol. 83, pp. 101—15, 2015.

[49] H. Tanizaki, *Nonlinear Filters, Estimation and Applications*, 2nd ed. Berlin: Springer, 1996.

[50] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Statistics for Engineering and Information Science. New York: Springer Verlag, 2001.

[51] P. Stano, Z. Lendek, J. Braaksma, R. Babuska, C. de Keizer, and A. J. den Dekker, "Parametric Bayesian Filters for Nonlinear Stochastic Dynamical Systems: A Survey," *IEEE Transactions on Sybernetics*, vol. 43, no. 6, pp. 1607–24, December 2013.

[52] D. Pham, "Stochastic methods for sequential data assimilation in strongly nonlinear systems," *Monthly Weather Review*, vol. 129, pp. 1194–1207, 2001.

[53] C. Bishop, B. Etherton, and S. Majumdar, "Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects," *Monthly Weather Review*, vol. 129, pp. 420–36, 2001.

[54] B. Hunt, E. Kostelich, and I. Szunyogh, "Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter," *Physica D*, vol. 230, pp. 112–26, 2007.

[55] I. Hoteit, D.-T. Pham, G. Triantafyllou, and G. Korres, "A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography," *Monthly Weather Review*, vol. 136, no. 1, pp. 317–334, 2008.

[56] A. S. Stordal, H. A. Karlsen, G. Naevdal, H. J. Skaug, and B. Valles, "Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter," *Computational Geosciences*, vol. 15, no. 2, pp. 293–305, March 2011.

[57] I. Hoteit, X. Luo, and D. Pham, "Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters," *Monthly Weather Review*, vol. 140, pp. 528–42, 2012.

[58] M. Frei and H. R. Künsch, "Mixture ensemble Kalman filters," *Computational Statistics and Data Analysis*, vol. 58, pp. 127–38, February 2013.

[59] B. Liu, B. Ait-El-Fquih, and I. Hoteit, "Efficient Kernel-Based Ensemble Gaussian Mixture Filtering," *Monthly Weather Review (under review)*, 2015.

[60] G. Evensen and P. van Leeuwen, "An ensemble Kalman smoother for nonlinear dynamics," *Monthly Weather Review*, vol. 128, no. 6, pp. 1852–67, 2000.

[61] S. Dunne, D. Entekhabi, and E. Njoku, "Impact of Multiresolution Active and Passive Microwave Measurements on Soil Moisture Estimation Using the Ensemble Kalman Smoother," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 4, pp. 1016–28, 2007.

[62] M. Sato, "Online model selection based on the variational Bayes," *Neural Computation*, vol. 13, no. 7, pp. 1649–81, July 2001.

[63] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, "An introduction to variational methods for graphical models," *Machine Learning*, vol. 37, pp. 183—233, 1999.

[64] S. Robert and W. Penny, "Variational Bayes for generalized autoregressive models," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2245–57, 2002.

[65] T. M. Hamill and J. S. Whitaker, "What constrains spread growth in forecasts initialized from ensemble Kalman filters?" *Monthly Weather Review*, vol. 139, no. 1, pp. 117–31, 2010.

[66] J. L. Anderson and S. L. Anderson, "A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," *Monthly Weather Review*, vol. 127, no. 12, pp. 2741–58, 1999.

[67] G. Gaspari and S. E. Cohn, "Construction of correlation functions in two and three dimensions," *Quarterly Journal of the Royal Meteorological Society*, vol. 554, no. 125, pp. 723–57, 1999.

[68] L. N. Trefethen and D. Bau, *Numerical linear algebra.* Philadelphia: Society for Industrial and Applied Mathematics, 1997.

[69] G. J. Bierman, "A comparison of discrete linear filtering algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 9, no. 1, pp. 28–37, 1973.

[70] J. Mandel, "Efficient Implementation of the Ensemble Kalman Filter," University of Colorado at Denver and Health Sciences Center, Tech. Rep., 2006.

Dr. Boujemaa Ait-El-Fquih received the Ph.D. degree in signal processing from Télécom SudParis (formerly Institut national des télécommunications), Evry, France in 2007. He worked at the Institut Français du Pétrole (IFP), Lyon, as a postdoctoral fellow (2007), the University of Paris XI, Orsay, as a teaching and research assistant (2008), the University of Bordeaux I, jointly with Electricité de France (EDF), Chatou, as a research associate (2010), and the Commissariat à l'Energie Atomique (CEA), Saclay, as a research engineer (2011). In September 2013, he joined the King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, in which he is currently working as a research scientist. His research interests are in Bayesian estimation methods and their applications in geophysics and signal processing.

Dr. Ibrahim Hoteit is an associate professor in the Earth Sciences and Engineering program and the Applied Mathematics and Computational Sciences program at KAUST. Prior to joining KAUST in 2009, he was a research scientist at Scripps Institution of Oceanography in University of California. Dr. Hoteit's research interests are in modeling and simulation of environmental fluid systems on supercomputers, with specific interest in data assimilation and inverse problems and associated uncertainty quantification for large-scale systems. Dr. Hoteit co-authored more than 90 scientific papers and is a co-recipient of five best conference paper awards. He is a member of the American and European Geophysical Unions, the Society of Industrial and Applied Mathematics, and an elected member of the UNESCO Center of Pure and Applied Mathematics. Dr. Hoteit earned his B.S. (1997), M.S. (1998), and Ph.D. (2002) from the University of Joseph Fourier, Grenoble, France.