Multiple Quadrature Kalman Filtering

Pau Closas, Carles Fernández-Prades, Jordi Vilà-Valls

Multiple Quadrature Kalman Filtering

Pau Closas, *Member, IEEE*, Carles Fernández-Prades, *Senior Member, IEEE*, and Jordi Vilà-Valls

***Abstract*—Bayesian filtering is a statistical approach that naturally appears in many signal processing problems. Ranging from Kalman filter to particle filters, there is a plethora of alternatives depending on model assumptions. With the exception of very few tractable cases, one has to resort to suboptimal methods due to the inability to analytically compute the Bayesian recursion in general dynamical systems. This is why it has attracted the attention of many researchers in order to develop efficient algorithms to implement it. We focus our interest into a recently developed algorithm known as the Quadrature Kalman filter (QKF). Under the Gaussian assumption, the QKF can tackle arbitrary nonlinearities by resorting to the Gauss-Hermite quadrature rules. However, its complexity increases exponentially with the state-space dimension. In this paper we study a complexity reduction technique for the QKF based on the partitioning of the state-space, referred to as the Multiple QKF. We prove that partitioning schemes can effectively be used to reduce the curse of dimensionality in the QKF. Simulation results are also provided to show that a nearly-optimal performance can be attained, while drastically reducing the computational complexity with respect to state-of-the-art algorithms that are able to deal with such nonlinear filtering problems.**

***Index Terms*—Adaptive filters, complexity reduction, high dimensional, Kalman filtering, nonlinear filters, quadrature rules.**

I. Introduction

THIS paper addresses nonlinear filtering problems where both process and measurement noises are additive and normally distributed. The filtering problem involves the recursive estimation of time-varying unknown states of a system using the incoming flow of information (estimation of the states at time k using the measurements up to time k), along some prior statistical knowledge about the variations of such states. The general discrete state-space model can be expressed as

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \boldsymbol{\nu}_{k-1}, \quad (1)$$

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k, \quad (2)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the hidden state of the system at time k, $\mathbf{f}_{k-1}(\cdot)$ is a known, possibly nonlinear, function of the states; and $\boldsymbol{\nu}_k$ is referred to as process noise, modeled with a zero-mean Gaussian distribution of covariance $\mathbf{Q}_k$; $\mathbf{y}_k \in \mathbb{R}^{n_y}$ is the measurement at time k, $\mathbf{h}_k(\cdot)$ is a known, possibly nonlinear, function, which relates measurements with states; and $\mathbf{n}_k$ is referred to as measurement noise, also zero-mean normally distributed with covariance $\mathbf{R}_k$, and independent of $\boldsymbol{\nu}_k$.

We are interested in the marginal distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, which gathers all the information about the system contained in the available observations, with $\mathbf{y}_{1:k} = \{\mathbf{y}_1, \ldots, \mathbf{y}_k\}$. This distribution can be recursively computed in two steps:

1) *Prediction:* we obtain the predictive distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ using prior information, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, and the previous filtering distribution, $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}. \quad (3)$$

2) *Update:* we use the new measurements $\mathbf{y}_k$ (measurement likelihood) and the predictive distribution to obtain the new filtering distribution $p(\mathbf{x}_k|\mathbf{y}_{1:k})$,

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k}. \quad (4)$$

Under the Gaussian assumption, the state transition density and the measurement likelihood function are Gaussian densities, which in turn reverts to a Gaussian posterior density [1]. Therefore, the problem is reduced to the estimation of the mean and covariance of each distribution. The predictive and filtering distributions can be expressed as

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}\right), \quad (5)$$

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), \quad (6)$$

respectively, where $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$ are the state estimates obtained at the prediction and update stages at time k, and $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ are the corresponding covariance matrices of the estimation error.

Depending on the dimensionality of the state-space model, the computational complexity associated to such estimation can be high. Actually, most algorithms are prone to fail in high-dimensional systems, an effect that is typically referred to as the curse of dimensionality. For instance, particle filters (PFs) [2] are known to require a large sample pool to characterize the filtering distribution when n_x increases [3]. In other words, PFs suffer from the curse of dimensionality: it is generally not possible to generate samples with significant weight, as required for importance sampling, in high dimensions. To circumvent this problem, some solutions were reported in the literature in the

The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hing Cheung So. This work was partially supported by the European Commission in the framework of the COST Action IC0803 (RFCSET), by the Spanish Ministry of Economy and Competitiveness project TEC2012-39143 (SOSRAD), the Spanish Science and Technology Commission TEC2010-19171 (MOSAIC), and by project 2009-SGR-1236 of the Catalan Administration (AGAUR).

P. Closas and C. Fernández-Prades are with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Parc Mediterrani de la Tecnologia, 08860 Castelldefels, Barcelona, Spain (e-mail: pclosas@cttc.cat; cfernandez@cttc.cat).

J. Vilà-Valls is with the Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3, 08034 Barcelona, Spain (e-mail: jordi.vila-valls@upc.edu).

Digital Object Identifier 10.1109/TSP.2012.2218811

context of particle filtering. Marginalization of linear states was proposed in [4], where the core idea was to use a Kalman filter (KF) to optimally deal with those states, while reducing the dimension of the state-space that the PF has to explore (that is, the nonlinear part). However, this technique can only be considered when the system has linear substructures and other solutions should be considered for more general cases. The idea of dividing the state-space into subsets and run a PF for each partition was presented in [5], [6]. The resulting method is known as the Multiple PF (MPF), which showed promising results both in terms of computational reduction and estimation performance. Recently the Iterated MPF (iMPF) was proposed in [7], which enhances the performance of the MPF by using tools of game theory to deal with the coupling of the filters and overcome the scarce data exchanged among them.

In this paper, we borrow the idea of the Multiple PF and apply it to the dimension reduction of the Quadrature Kalman filter (QKF) [8], a kind of sigma-point Kalman filter (SPKF) [9]. SPKFs are a family of derivative-free Gaussian filters, which are based on a weighted sum of function evaluations at L specified (i.e., deterministic) points within the domain of integration, as opposite to the stochastic sampling performed by particle filtering methods. Different SPKF-like algorithms involve different sigma-points and weights, which are used to approximate the transformation of means and covariances when applied to a nonlinear system. For instance, the Unscented Kalman filter (UKF) draws points using the Unscented Transform [10], [11], delivering estimates that are exact in mean for monomials up to third degree, while covariance computation is exact only for linear functions; the Cubature Kalman filter (CKF) resorts to spherical-radial cubature rules [12], which are exact in mean for all polynomials in $\mathbb{R}^{n_x}$ up to the third degree and in covariance for linear functions; and the QKF is based on Gauss-Hermite quadrature rules [1], [8], [13], [14]. Whereas UKF and CKF exhibit a linear growth of L with respect to n_x, it is exponential for the QKF method. A single parameter is required for the use of quadrature rules, which is the number of sigma-points per dimension α. The total number of quadrature points is then fixed to $L = \alpha^{n_x}$. Indeed, this parameter can be used to adjust the algorithm, with the quadrature rules being optimal for nonlinearities of degree $2L - 1$. Despite the curse of dimensionality, the remarkable performance improvement exhibited by the QKF with respect to other SPFKs [15], makes it an appealing tool for practitioners. Recently, Smolyak rules (*a.k.a.* sparse grids) have been applied to the QKF in order to avoid tensor products of quadrature rules, further reducing the computational load [16], an approach that can be complementarily applied to the proposed Multiple QKF (MQKF).

Our contribution investigates if partitioning approaches could be used to reduce the computational complexity of the problem and proposes a method based on partitioning which we have named the Multiple Square-root QKF (MSQKF). The assumptions are that the dynamic system is driven by Gaussian processes and that the possible nonlinearities are approximated by the Gauss-Hermite quadrature rules.

The paper is organized as follows. Section II introduces the partitioning of the states and describes the implementation of the MQKF and the corresponding square-root version (MSQKF). Section III discusses some important issues related to the multiple approach. Section IV proves the complexity reduction of partitioning schemes applied to the QKF. Section V supports the approach with computer simulations where two examples are considered: an illustrative academic example and a multiple frequency estimation problem. Section VI draws some conclusions.

II. Multiple Quadrature Kalman Filter

A. State-Space Partitioning

The state-space (1) can be partitioned into S subspaces, possibly with different dimensions. Also, the considered system can be coupled in the observations, where states are combined via $\mathbf{h}_k(\cdot)$. Therefore, state equation can be equivalently expressed as[1]

$$\begin{pmatrix} \mathbf{x}_k^{(1)} \\ \mathbf{x}_k^{(2)} \\ \vdots \\ \mathbf{x}_k^{(S)} \end{pmatrix} = \begin{pmatrix} \mathbf{f}_{k-1}^{(1)}(\mathbf{x}_{k-1}^{(1)}, \mathbf{x}_{k-1}^{(-1)}) \\ \mathbf{f}_{k-1}^{(2)}(\mathbf{x}_{k-1}^{(2)}, \mathbf{x}_{k-1}^{(-2)}) \\ \vdots \\ \mathbf{f}_{k-1}^{(S)}(\mathbf{x}_{k-1}^{(S)}, \mathbf{x}_{k-1}^{(-S)}) \end{pmatrix} + \begin{pmatrix} \boldsymbol{\nu}_{k-1}^{(1)} \\ \boldsymbol{\nu}_{k-1}^{(2)} \\ \vdots \\ \boldsymbol{\nu}_{k-1}^{(S)} \end{pmatrix}, \quad (7)$$

where each function $\mathbf{f}_{k-1}^{(s)}(\mathbf{x}_{k-1}) : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_x^{(s)}}$, with $s \in \mathcal{S} = \{1, \ldots, S\}$, can be different, and we define the dimension of each subspace $n_x^{(s)} = \dim\{\mathbf{x}_k^{(s)}\}$ such that $\sum_{s=1}^{S} n_x^{(s)} = n_x$. Notice that the state-space models which can be considered within the partitioning approach are general. A discussion is provided in Section III on the simplifications and correlations neglected by the multiple approach in general models. In the sequel, we write $\mathbf{f}_{k-1}^{(s)}(\mathbf{x}_{k-1}^{(s)}, \mathbf{x}_{k-1}^{(-s)})$ instead of $\mathbf{f}_{k-1}^{(s)}(\mathbf{x}_{k-1})$ to make the subspace partitioning more explicit.

The s-th process noise is distributed as $\boldsymbol{\nu}_{k-1}^{(s)} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{k-1}^{(s)}\right)$, where $\mathbf{Q}_{k-1}^{(s)}$ is constructed from $\mathbf{Q}_{k-1}$ as the entries corresponding to the s-th subspace. In general, the partitioned process noise vector is equivalent to $\boldsymbol{\nu}_{k-1}$ in (1) taking into account that the cross-covariance matrices among subspace process noise vectors may not be null.

In the proposed approach, each subspace $\mathbf{x}_k^{(s)}$ considers only its own process noise, $\boldsymbol{\nu}_{k-1}^{(s)}$, and thus neglecting the possible cross-correlations among subspace process noise vectors. In practice, the partitioned process noise considered is only equivalent to $\boldsymbol{\nu}_{k-1}$ for uncorrelated partitions. This is also discussed in Section III.

A special case: the state-space partitioning is strongly simplified when dealing with independent subspaces, where the subspace process noise cross-covariances are null and there is no interconnection among subspaces. The s-th subspace process equation is reduced to

$$\mathbf{x}_k^{(s)} = \mathbf{f}_{k-1}^{(s)}(\mathbf{x}_{k-1}^{(s)}) + \boldsymbol{\nu}_{k-1}^{(s)}, \quad (8)$$

with $\mathbf{f}_{k-1}^{(s)}(\mathbf{x}_{k-1}^{(s)}) : \mathbb{R}^{n_x^{(s)}} \mapsto \mathbb{R}^{n_x^{(s)}}$ in this case.

[1] $\mathbf{x}^{(s)}$ denotes the s-th element (possibly a vector) in a vector $\mathbf{x}$ and $\mathbf{x}^{(-s)}$ is the vector of all elements in $\mathbf{x}$ except for $\mathbf{x}^{(s)}$.

B. Bayesian Filtering Formulation in the Gaussian Domain

As we mentioned, the key idea consists of dividing the state-space into S subspaces and applying a QKF to each partition. The Gaussian filter in charge of the s-th subspace, $\mathbf{x}_k^{(s)}$, is mainly interested in the marginal posterior distribution $p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k})$, where the conditioning on $\mathbf{x}_k^{(-s)}$ shows the interconnection among subspaces. The Bayesian recursive solution to obtain this distribution is given in (3) and (4), which in this case reads as

$$\begin{aligned} &p(\mathbf{x}_k^{(s)}|\mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}) \\ &\quad = \int p(\mathbf{x}_k^{(s)}|\mathbf{x}_{k-1}^{(-s)}, \mathbf{x}_{k-1}^{(s)}) p(\mathbf{x}_{k-1}^{(s)}|\mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}^{(s)} \end{aligned} \quad (9)$$

and

$$\begin{aligned} &p(\mathbf{x}_k^{(s)}|\mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}) \\ &\qquad \propto p(\mathbf{y}_k|\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) p(\mathbf{x}_k^{(s)}|\mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}), \end{aligned} \quad (10)$$

respectively.

Notice that the marginal predictive distribution is conditioned to $\mathbf{x}_{k-1}^{(-s)}$ and not to $\mathbf{x}_{k-1}^{(s)}$, because the distribution is obtained by integrating out the latter over the corresponding marginal posterior at $k-1$.

In the Gaussian domain, the expressions of the marginal predictive and filtering distributions are:

$$p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k-1}^{(s)}, \mathbf{P}_{k|k-1}^{(s)}\right), \quad (11)$$

$$p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k}^{(s)}, \mathbf{P}_{k|k}^{(s)}\right), \quad (12)$$

respectively. The problem is then to estimate first the predicted state vector and its covariance matrix, $\{\hat{\mathbf{x}}_{k|k-1}^{(s)}, \mathbf{P}_{k|k-1}^{(s)}\}$, and then the estimates at instant k incorporating the information provided by the new measurement $\mathbf{y}_k$, $\{\hat{\mathbf{x}}_{k|k}^{(s)}, \mathbf{P}_{k|k}^{(s)}\}$.

The equations involved in the k-th prediction and update steps[2] are summarized as follows:

- *Prediction:*

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1}^{(s)} &= \mathbb{E}\left\{\mathbf{x}_k^{(s)}|\ \mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}\right\} \\ &= \int \mathbf{f}(\mathbf{x}_{k-1}^{(s)}, \mathbf{x}_{k-1}^{(-s)}) \\ &\quad \times p(\mathbf{x}_{k-1}^{(s)}|\ \mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}^{(s)} \end{aligned} \quad (13)$$

$$\begin{aligned} \mathbf{P}_{k|k-1}^{(s)} &= \mathbb{E}\left\{\left(\mathbf{x}_k^{(s)} - \hat{\mathbf{x}}_{k|k-1}^{(s)}\right)^2 |\ \mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}\right\} \\ &= \int \mathbf{f}^2(\mathbf{x}_{k-1}^{(s)}, \mathbf{x}_{k-1}^{(-s)}) p(\mathbf{x}_{k-1}^{(s)}|\ \mathbf{x}_{k-1}^{(-s)}, \mathbf{y}_{1:k-1}) \\ &\quad \times d\mathbf{x}_{k-1}^{(s)} - \left(\hat{\mathbf{x}}_{k|k-1}^{(s)}\right)^2 + \mathbf{Q}_{k-1}^{(s)}. \end{aligned} \quad (14)$$

- *Update:*

$$\hat{\mathbf{x}}_{k|k}^{(s)} = \hat{\mathbf{x}}_{k|k-1}^{(s)} + \mathbf{K}_k^{(s)} \left(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}^{(s)}\right) \quad (15)$$

$$\mathbf{P}_{k|k}^{(s)} = \mathbf{P}_{k|k-1}^{(s)} - \mathbf{K}_k^{(s)} \mathbf{\Sigma}_{yy,k|k-1}^{(s)} \left(\mathbf{K}_k^{(s)}\right)^T, \quad (16)$$

where the Kalman gain, measurement prediction, innovation covariance and cross-covariance are obtained as

$$\begin{aligned} \mathbf{K}_k^{(s)} &= \mathbf{\Sigma}_{xy,k|k-1}^{(s)} \left(\mathbf{\Sigma}_{yy,k|k-1}^{(s)}\right)^{-1} \\ \hat{\mathbf{y}}_{k|k-1}^{(s)} &= \int \mathbf{h}(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k-1}) \\ &\quad \times d\mathbf{x}_k^{(s)} \\ \mathbf{\Sigma}_{yy,k|k-1}^{(s)} &= \int \mathbf{h}^2(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k-1}) \\ &\quad \times d\mathbf{x}_k^{(s)} - \left(\hat{\mathbf{y}}_{k|k-1}^{(s)}\right)^2 + \mathbf{R}_k \\ \mathbf{\Sigma}_{xy,k|k-1}^{(s)} &= \int \mathbf{x}_k^{(s)} \mathbf{h}^T(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) \\ &\quad \times p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k-1}) d\mathbf{x}_k^{(s)} \\ &\quad - \hat{\mathbf{x}}_{k|k-1}^{(s)} \left(\hat{\mathbf{y}}_{k|k-1}^{(s)}\right)^T. \end{aligned}$$

Inspecting the prediction and update steps equations, we realize that a number of integrals must be solved by the filter. These integrals can be seen as expectations of some known (nonlinear) function of the states over a distribution, which can be either the predictive or the filtering distribution. Under the Gaussian assumption, these distributions are Gaussian and given by (11) and (12). The assumption done allows us to resort to the Gauss-Hermite quadrature rules to solve for the integral [17]. Those rules were seen to be a powerful tool to approximate integrals of the kind we mentioned [13],

$$I = \int \text{nonlinear function} \cdot \text{Gaussian density},$$

constituting the core concept of the QKF.

The difference with a conventional filter is that each step is performed in parallel by the bank. At each step, the s-th filter is aware of the estimates delivered by the rest of filters. The different filters are interconnected at each step as expressed in (11) and (12) with the conditional distribution upon $\mathbf{x}_{k-1}^{(-s)}$ and $\mathbf{x}_k^{(-s)}$, respectively. Unfortunately, these two quantities are not available due to the inability to marginalize over nonlinear functions, and hence approximations should be performed. Following the approach in [5], [6], here we propose to use

$$\mathbf{x}_{k-1}^{(-s)} = \hat{\mathbf{x}}_{k-1|k-1}^{(-s)} \quad \text{and} \quad \mathbf{x}_k^{(-s)} = \hat{\mathbf{x}}_{k|k-1}^{(-s)}, \quad (17)$$

in prediction and update steps of each parallel filter, respectively. Notice that these values are available at the required step. $\hat{\mathbf{x}}_{k-1|k-1}^{(-s)}$ denotes the filtering solution of all the filters but s at the previous time instant, which is shared among filters to predict at k. The update is performed using the predicted values of each subspace, $\hat{\mathbf{x}}_{k|k-1}^{(-s)}$.

To sum up, the simplifications made here when applying partitioning schemes are twofold: i) the cross-terms among state-partitions in $\hat{\mathbf{Q}}_k$ are neglected and ii) the coupling among filters is performed with the point estimates in (17). The impact

[2]We write $(\mathbf{x})^2$, $(\mathbf{y})^2$, $\mathbf{f}^2(\cdot)$ and $\mathbf{h}^2(\cdot)$ as the shorthand for $\mathbf{x}\mathbf{x}^T$, $\mathbf{y}\mathbf{y}^T$, $\mathbf{f}(\cdot)\mathbf{f}^T(\cdot)$ and $\mathbf{h}(\cdot)\mathbf{h}^T(\cdot)$, respectively. We omitted the dependence with time and the superscript $^{(s)}$ of the process function $\mathbf{f}_{k-1}^{(s)}(\cdot)$, and the dependence with time of the measurement function $\mathbf{h}_k(\cdot)$, for the sake of clarity.

of these simplifications is studied in Sections III-B1 and III-B2, respectively.

C. Multiple Square-Root Quadrature Kalman Filter

In the following, we describe the operation of the proposed MSQKF. Initially, we have to specify the state-space partitioning $\{L_s \triangleq n_x^{(s)}\}_{s=1,\ldots,S}$, the number of sigma-points per dimension α, an estimation of the process and measurement noise covariances $\hat{\mathbf{Q}}_k$ and $\hat{\mathbf{R}}_k$ (assuming stationarity for the sake of simplicity, they will be considered constants over the time of observation), and the initial conditions $\hat{\mathbf{x}}_{0|0}$ and the corresponding covariance $\mathbf{\Sigma}_{x,0|0}$.

This new algorithm is constructed as a bank of S parallel SQKFs, each one tracking a different subspace. Each filter computes sigma-points and weights, $\{\boldsymbol{\xi}_i^{(s)}, \omega_i^{(s)}\}_{i=1,\ldots,L_s}$, resorting to the Gauss-Hermite quadrature rules [15], [17]. The corresponding deterministic points are indeed function of the dimension of the subspace and the number of points per dimension α. Hereinafter, the latter is considered equal for all filters. We define the weighting matrix as $\mathbf{W}^{(s)} = \mathrm{diag}(\sqrt{\omega_i^{(s)}})$.

The prediction step can be done in parallel for each subspace. At time step k, the previous state and covariance estimates ($\hat{\mathbf{x}}_{k-1|k-1}^{(s)}$, $\hat{\mathbf{x}}_{k-1|k-1}^{(-s)}$ and $\mathbf{S}_{x,k-1|k-1}^{(s)}$) and the estimation of the subspace process noise covariance $\hat{\mathbf{Q}}_{k-1}^{(s)}$ are available. First we evaluate the sigma-points

$$\mathbf{x}_{i,k-1|k-1}^{(s)} = \mathbf{S}_{x,k-1|k-1}^{(s)} \boldsymbol{\xi}_i^{(s)} + \hat{\mathbf{x}}_{k-1|k-1}^{(s)},\ i = 1, \ldots, L_s, \tag{18}$$

and then we propagate each one through the corresponding transition function

$$\tilde{\mathbf{x}}_{i,k|k-1}^{(s)} = \mathbf{f}_{k-1}(\mathbf{x}_{i,k-1|k-1}^{(s)}, \hat{\mathbf{x}}_{k-1|k-1}^{(-s)}),\ i = 1, \ldots, L_s. \tag{19}$$

The subspace predicted state is computed as the weighted mean of the propagated sigma-points

$$\hat{\mathbf{x}}_{k|k-1}^{(s)} = \sum_{i=1}^{L_s} \omega_i^{(s)} \tilde{\mathbf{x}}_{i,k|k-1}^{(s)}, \tag{20}$$

The predicted state $\hat{\mathbf{x}}_{k|k-1}$ is then constructed as the concatenation of each subspace predicted state, $\hat{\mathbf{x}}_{k|k-1}^{(1)}, \ldots, \hat{\mathbf{x}}_{k|k-1}^{(S)}$.

The square-root factor of the predicted error covariance is

$$\mathbf{S}_{x,k|k-1}^{(s)} = \mathrm{Tria}\left(\left[\tilde{\mathcal{X}}_{k|k-1}^{(s)} \middle| \mathbf{S}_{\hat{\mathbf{Q}}_{k-1}}^{(s)}\right]\right), \tag{21}$$

where $\mathbf{S}_{\hat{\mathbf{Q}}_{k-1}}^{(s)}$ is a square-root factor of $\hat{\mathbf{Q}}_{k-1}^{(s)}$ such that $\hat{\mathbf{Q}}_{k-1}^{(s)} = \mathbf{S}_{\hat{\mathbf{Q}}_{k-1}}^{(s)} \left(\mathbf{S}_{\hat{\mathbf{Q}}_{k-1}}^{(s)}\right)^T$, and $\tilde{\mathcal{X}}_{k|k-1}^{(s)} = \left[\tilde{\mathbf{x}}_{1,k|k-1}^{(s)} - \hat{\mathbf{x}}_{k|k-1}^{(s)}, \ldots, \tilde{\mathbf{x}}_{L_s,k|k-1}^{(s)} - \hat{\mathbf{x}}_{k|k-1}^{(s)}\right] \mathbf{W}^{(s)}$, and $\mathbf{S} = \mathrm{Tria}(\mathbf{A})$ denotes a general triangulation algorithm (for instance, the QR decomposition), where $\mathbf{A} \in \mathbb{R}^{p \times q}$, $p < q$, and $\mathbf{S}$ is a lower triangular matrix such that $\mathbf{S}^T\mathbf{S} = \mathbf{A}$. For computational reasons, we prefer to keep the square root as a triangular matrix of the dimension $p \times p$. This can be achieved by the thin QR decomposition [18, Section 5.2, Theorem 5.2.2]], which has a computational complexity of $\mathcal{O}(qp^2)$ flops.

As for the prediction step, the update step can also be parallelized. Considering the s-th filter, the sigma-points are first evaluated using the subspace predicted state and its corresponding predicted error covariance matrix

$$\mathbf{x}_{i,k|k-1}^{(s)} = \mathbf{S}_{x,k|k-1}^{(s)} \boldsymbol{\xi}_i^{(s)} + \hat{\mathbf{x}}_{k|k-1}^{(s)},\ i = 1, \ldots, L_s, \tag{22}$$

and then propagated through the measurement function

$$\tilde{\mathbf{y}}_{i,k|k-1}^{(s)} = \mathbf{h}_k(\mathbf{x}_{i,k|k-1}^{(s)}, \hat{\mathbf{x}}_{k|k-1}^{(-s)}), \quad i = 1, \ldots, L_s. \tag{23}$$

The predicted measurement is computed as the weighted mean of the propagated sigma-points

$$\hat{\mathbf{y}}_{k|k-1}^{(s)} = \sum_{i=1}^{L_s} \omega_i^{(s)} \tilde{\mathbf{y}}_{i,k|k-1}^{(s)}. \tag{24}$$

The square-root of the innovation covariance matrix is

$$\mathbf{S}_{y,k|k-1}^{(s)} = \mathrm{Tria}\left(\left[\mathcal{Y}_{k|k-1}^{(s)} \middle| \mathbf{S}_{\hat{\mathbf{R}}_k}\right]\right), \tag{25}$$

where $\mathbf{S}_{\hat{\mathbf{R}}_k}$ denotes a square-root factor of $\hat{\mathbf{R}}_k$ such that $\hat{\mathbf{R}}_k = \mathbf{S}_{\hat{\mathbf{R}}_k} \left(\mathbf{S}_{\hat{\mathbf{R}}_k}\right)^T$, and

$$\mathcal{Y}_{k|k-1}^{(s)} = \left[\tilde{\mathbf{y}}_{1,k|k-1}^{(s)} - \hat{\mathbf{y}}_{k|k-1}^{(s)}, \ldots, \tilde{\mathbf{y}}_{L_s,k|k-1}^{(s)} - \hat{\mathbf{y}}_{k|k-1}^{(s)}\right] \mathbf{W}^{(s)}.$$

The cross-covariance matrix can be obtained from

$$\mathbf{\Sigma}_{xy,k|k-1}^{(s)} = \mathcal{X}_{k|k-1}^{(s)} \left(\mathcal{Y}_{k|k-1}^{(s)}\right)^T, \tag{26}$$

where

$$\mathcal{X}_{k|k-1}^{(s)} = \left[\mathbf{x}_{1,k|k-1}^{(s)} - \hat{\mathbf{x}}_{k|k-1}^{(s)}, \ldots, \mathbf{x}_{L_s,k|k-1}^{(s)} - \hat{\mathbf{x}}_{k|k-1}^{(s)}\right] \mathbf{W}^{(s)}.$$

Using now the standard Kalman solution we obtain the Kalman gain

$$\mathbf{K}_k^{(s)} = \left(\mathbf{\Sigma}_{xy,k|k-1}^{(s)} / \left(\mathbf{S}_{y,kk-1}^{(s)}\right)^T\right) / \mathbf{S}_{y,k|k-1}^{(s)}, \tag{27}$$

the subspace estimated state

$$\hat{\mathbf{x}}_{k|k}^{(s)} = \hat{\mathbf{x}}_{k|k-1}^{(s)} + \mathbf{K}_k^{(s)} \left(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}^{(s)}\right), \tag{28}$$

and the square-root factor of the corresponding error covariance[3]

$$\mathbf{S}_{x,k|k}^{(s)} = \mathrm{Tria}\left(\left[\mathcal{X}_{k|k-1}^{(s)} - \mathbf{K}_k^{(s)} \mathcal{Y}_{k|k-1}^{(s)} \middle| \mathbf{K}_k^{(s)} \mathbf{S}_{\hat{\mathbf{R}}_k}\right]\right). \tag{29}$$

The final state estimate $\hat{\mathbf{x}}_{k|k}$ is obtained as the concatenation of each subspace state estimates, $\hat{\mathbf{x}}_{k|k}^{(1)}, \ldots, \hat{\mathbf{x}}_{k|k}^{(S)}$.

[3]Notice that the SQKF computes the square-root factor of the estimation covariances, $\mathbf{S}_{x,k|k-1}^{(s)}$ and $\mathbf{S}_{x,k|k}^{(s)}$. The distributions in (11) and (12) can be readily obtained using $\mathbf{P}_{k|k-1}^{(s)} = \mathbf{S}_{x,k|k-1}^{(s)} \left(\mathbf{S}_{x,k|k-1}^{(s)}\right)^T$ and $\mathbf{P}_{k|k}^{(s)} = \mathbf{S}_{x,k|k}^{(s)} \left(\mathbf{S}_{x,k|k}^{(s)}\right)^T$.

III. On the Multiple Approach

Considering the partitioned state-space model defined by (2) and (7), the posterior distribution of interest is $p(\mathbf{x}_k|\mathbf{y}_{1:k}) = p(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}|\ \mathbf{y}_{1:k})$. In the proposed method, the s-th filter of the bank tracks the s-th subspace using the subspace marginal predictive and filtering distributions specified in (11) and (12), considering the implementation approximation stated in (17), and the final estimate is constructed from the concatenation of the subspace estimates.

In the following subsections we address some questions which naturally arise when using the multiple approach.

A. Equivalence Between Both Approaches

Intuitively we can see that the estimation using the partitioning and the direct estimation of the full state may not be equivalent in a general case. But, when is the estimation obtained with both approaches equivalent?

The standard estimation can be expressed as

$$\hat{\mathbf{x}}_k = \mathbb{E}\left\{\mathbf{x}_k|\ \mathbf{y}_{1:k}\right\} = \int \mathbf{x}_k p(\mathbf{x}_k|\ \mathbf{y}_{1:k}) d\mathbf{x}_k, \tag{30}$$

and the proposed multiple approach is

$$\begin{pmatrix} \hat{\mathbf{x}}_k^{(1)} \\ \vdots \\ \hat{\mathbf{x}}_k^{(S)} \end{pmatrix} = \begin{pmatrix} \int \mathbf{x}_k^{(1)} p(\mathbf{x}_k^{(1)}|\ \mathbf{x}_k^{(-1)}, \mathbf{y}_{1:k}) d\mathbf{x}_k^{(1)} \\ \vdots \\ \int \mathbf{x}_k^{(S)} p(\mathbf{x}_k^{(S)}|\ \mathbf{x}_k^{(-S)}, \mathbf{y}_{1:k}) d\mathbf{x}_k^{(S)} \end{pmatrix}. \tag{31}$$

Expressions (30) and (31) are equivalent only when the subspaces are independent and one can write the posterior distribution as

$$\begin{aligned} p(\mathbf{x}_k|\ \mathbf{y}_{1:k}) &= p(\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(S)}|\ \mathbf{y}_{1:k}) \\ &= \prod_{s=1}^{S} p(\mathbf{x}_k^{(s)}|\ \mathbf{y}_{1:k}), \end{aligned} \tag{32}$$

and the subspace marginal filtering distributions as

$$p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}) = p(\mathbf{x}_k^{(s)}|\mathbf{y}_{1:k}). \tag{33}$$

In this case, we have that $\hat{\mathbf{x}}_k = \left[\hat{\mathbf{x}}_k^{(1)}, \ldots, \hat{\mathbf{x}}_k^{(S)}\right]^T$, where

$$\hat{\mathbf{x}}_k^{(s)} = \int \mathbf{x}_k^{(s)} p(\mathbf{x}_k^{(s)}|\ \mathbf{y}_{1:k}) d\mathbf{x}_k^{(s)}. \tag{34}$$

Notice that the independence among subspaces does not imply that we deal with S independent estimation problems, because the observations still may combine the full state:

$$\begin{aligned} p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}) &= p(\mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}|\ \mathbf{x}_k^{(s)}) p(\mathbf{x}_k^{(s)}) \\ &= p(\mathbf{y}_{1:k}|\ \mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) p(\mathbf{x}_k^{(-s)}) \\ &\quad \times p(\mathbf{x}_k^{(s)}). \end{aligned} \tag{35}$$

The coupling of all the subspaces can also be seen within the MQKF structure through the innovations. Because the subspaces are independent, the s-th filter subspace prediction is

$$\hat{\mathbf{x}}_{k|k-1}^{(s)} = \int \mathbf{f}(\mathbf{x}_{k-1}^{(s)}) p(\mathbf{x}_{k-1}^{(s)}|\ \mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}^{(s)}, \tag{36}$$

but the predicted measurement (24) still depends on the full state

$$\hat{\mathbf{y}}_{k|k-1}^{(s)} = \int \mathbf{h}(\mathbf{x}_k^{(s)}, \mathbf{x}_k^{(-s)}) p(\mathbf{x}_k^{(s)}|\ \mathbf{y}_{1:k-1}) d\mathbf{x}_k^{(s)}. \tag{37}$$

Moreover, the different subspaces are coupled in (15) through the innovations and the cross-covariance to compute the subspace estimates.

B. QKF Vs MQKF: Theory and Practice

In the previous subsection we stated that from a theoretical point of view the multiple approach is equivalent to the standard filtering method only when the subspaces are independent. But in a general case, which is the difference between both methods?

From a theoretical standpoint, we have to analyze the difference between the standard and the multiple approach. Regarding the implementation of the MSQKF, we need to assess the impact of using the estimates $\hat{\mathbf{x}}_{k-1|k-1}^{(-s)}$ and $\hat{\mathbf{x}}_{k|k-1}^{(-s)}$ instead of $\mathbf{x}_{k-1}^{(-s)}$ and $\mathbf{x}_k^{(-s)}$.

1) Theoretical Analysis: The standard solution uses the posterior density

$$\begin{aligned} p_{opt} &= p(\mathbf{x}_k|\ \mathbf{y}_{1:k}) = p(\mathbf{x}_k^{(1)}, \ldots, \mathbf{x}_k^{(S)}|\ \mathbf{y}_{1:k}) \\ &= p(\mathbf{x}_k^{(1)}|\ \mathbf{x}_k^{(2)}, \ldots, \mathbf{x}_k^{(S)}, \mathbf{y}_{1:k}) \\ &\quad \times p(\mathbf{x}_k^{(2)}|\ \mathbf{x}_k^{(3)}, \ldots, \mathbf{x}_k^{(S)}, \mathbf{y}_{1:k}) \ldots p(\mathbf{x}_k^{(S-1)}|\ \mathbf{x}_k^{(S)}, \mathbf{y}_{1:k}) \\ &\quad \times p(\mathbf{x}_k^{(S)}|\ \mathbf{y}_{1:k}), \end{aligned} \tag{38}$$

and the multiple approach uses the subspace filtering densities

$$p_{ma} = p(\mathbf{x}_k^{(1)}|\ \mathbf{x}_k^{(-1)}, \mathbf{y}_{1:k}) \ldots p(\mathbf{x}_k^{(S)}|\ \mathbf{x}_k^{(-S)}, \mathbf{y}_{1:k}). \tag{39}$$

From a probabilistic point of view, the inability to marginalize the subspace filtering densities makes impossible to obtain a closed form comparison of these two distributions. That is why we resort to the Gaussian expressions of the posterior and the subspace marginal densities:

$$p(\mathbf{x}_k|\ \mathbf{y}_{1:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}\right), \tag{40}$$

$$p(\mathbf{x}_k^{(s)}|\ \mathbf{x}_k^{(-s)}, \mathbf{y}_{1:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k}^{(s)}, \mathbf{P}_{k|k}^{(s)}\right). \tag{41}$$

Using expressions (41) and (39), we can compute the multiple approach log-posterior density

$$\ln(p_{ma}) = \beta - \frac{1}{2} \sum_{s=1}^{S} \left[\mathbf{x}_k^{(s)} - \hat{\mathbf{x}}_{k|k}^{(s)}\right]^T \left(\mathbf{P}_{k|k}^{(s)}\right)^{-1} \left[\mathbf{x}_k^{(s)} - \hat{\mathbf{x}}_{k|k}^{(s)}\right]$$

with $\beta = -\frac{1}{2} \sum_{s=1}^{S} \ln\left((2\pi)^{n_x^{(s)}} \det\left(\mathbf{P}_{k|k}^{(s)}\right)\right)$. This distribution may be reformulated as

$$\ln(p_{ma}) = \beta - \frac{1}{2} \left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{ma}\right]^T \left(\mathbf{P}_{k|k}^{ma}\right)^{-1} \left[\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{ma}\right]$$

where $\hat{\mathbf{x}}_{k|k}^{ma}$ corresponds to a vector containing the subspace estimates and $\mathbf{P}_{k|k}^{ma}$ is a block diagonal matrix

$$\hat{\mathbf{x}}_{k|k}^{ma} = \left[\hat{\mathbf{x}}_{k|k}^{(1)}, \ldots, \hat{\mathbf{x}}_{k|k}^{(S)}\right]^T, \tag{42}$$

$$\mathbf{P}_{k|k}^{ma} = \operatorname{diag}\left(\mathbf{P}_{k|k}^{(1)}, \ldots, \mathbf{P}_{k|k}^{(S)}\right). \tag{43}$$

The difference between both methods and the information that we lose with the partitioning is given by the difference between both Gaussian distributions, p_{opt} and p_{ma}, which are characterized by their mean and covariance matrix: $\left\{\hat{\mathbf{x}}_{k|k}, \hat{\mathbf{x}}^{ma}_{k|k}, \mathbf{P}_{k|k}, \mathbf{P}^{ma}_{k|k}\right\}$. As we are dealing with a recursive estimation method, the update is directly related to the prediction. Notice that the same derivation can be done for the standard and subspace marginal predictive distributions, $p(\mathbf{x}_k|\ \mathbf{y}_{1:k-1})$ and $p(\mathbf{x}^{(s)}_k|\ \mathbf{x}^{(-s)}_{k-1}, \mathbf{y}_{1:k-1})$, respectively, and we also need to compare their mean and covariance matrix: $\left\{\hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{x}}^{ma}_{k|k-1}, \mathbf{P}_{k|k-1}, \mathbf{P}^{ma}_{k|k-1}\right\}$, where $\hat{\mathbf{x}}^{ma}_{k|k-1}$ is also constructed as the concatenation of the subspace predicted states and $\mathbf{P}^{ma}_{k|k-1}$ is block diagonal.

Concerning the prediction step (see Section II-B), note that each subspace predicted error covariance matrix $\mathbf{P}^{(s)}_{k|k-1}$ is computed using the corresponding subspace process noise covariance $\mathbf{Q}^{(s)}_{k-1}$, so the block diagonal matrix $\mathbf{P}^{ma}_{k|k-1}$ considers only the noise affecting each subspace, $\mathbf{Q}^{ma}_{k-1} = \operatorname{diag}\left(\mathbf{Q}^{(1)}_{k-1}, \ldots, \mathbf{Q}^{(S)}_{k-1}\right)$, instead of the full process noise covariance matrix $\mathbf{Q}_{k-1}$. We lose the information about the correlation of the noise among partitions. Also, if there exists any information, because of the interconnection among subspaces, which may contribute to the off-diagonal elements of $\mathbf{P}^{ma}_{k|k-1}$, it is not taken into account by the algorithm.

The errors introduced in the prediction step are propagated to the update step and the computation of $\mathbf{x}^{ma}_{k|k}$ and $\mathbf{P}^{ma}_{k|k}$. In this step, we also note that the innovation's covariance matrix is computed only over the subspace marginal predictive distribution instead of the complete predictive density, what may also imply a loss of information. For numerical results illustrating these effects see Section V-A.

2) Practical Implications: To assess the impact of using the predicted and estimated states instead of the true values for the interconnection among the different subfilters, refer to (17), we perform here a first-order analysis. To that aim we consider the following linear/Gaussian state-space model

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1}, \tag{44}$$

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{n}_k, \tag{45}$$

where $\mathbf{F}_{k-1}$ and $\mathbf{H}_k$ act as the state transition and measurement matrices. The rest of the variables are defined as in the original state-space model, (1) and (2).

Applying the multiple approach, the s-th subspace state equation is

$$\mathbf{x}^{(s)}_k = \mathbf{F}^{(s)}_{k-1}\begin{bmatrix}\mathbf{x}^{(s)}_{k-1}\\ \mathbf{x}^{(-s)}_{k-1}\end{bmatrix} + \boldsymbol{\nu}^{(s)}_{k-1}, \tag{46}$$

where the process noise is distributed as $\boldsymbol{\nu}^{(s)}_{k-1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}^{(s)}_{k-1}\right)$ and $\mathbf{F}^{(s)}_{k-1}$ is the corresponding subspace state transition matrix. The multiple prediction step should be

$$\hat{\mathbf{x}}^{(s)}_{k|k-1} = \mathbf{F}^{(s)}_{k-1}\begin{bmatrix}\hat{\mathbf{x}}^{(s)}_{k-1|k-1}\\ \mathbf{x}^{(-s)}_{k-1}\end{bmatrix}, \tag{47}$$

but instead we use the following approximation

$$\hat{\mathbf{x}}^{(s)}_{k|k-1} = \mathbf{F}^{(s)}_{k-1}\begin{bmatrix}\hat{\mathbf{x}}^{(s)}_{k-1|k-1}\\ \hat{\mathbf{x}}^{(-s)}_{k-1|k-1}\end{bmatrix}. \tag{48}$$

We can separate the matrix into two blocks (the first one affecting the s-th subspace, $\mathbf{x}^{(s)}_{k-1}$, and the other the rest, $\mathbf{x}^{(-s)}_{k-1}$)

$$\mathbf{F}^{(s)}_{k-1} = \left(\mathbf{F}^{(s)}_{k-1,1}\middle|\mathbf{F}^{(s)}_{k-1,2}\right), \tag{49}$$

with dimensions $\left(n^{(s)}_x \times n^{(s)}_x\right)$ and $\left(n^{(s)}_x \times n^{(-s)}_x\right)$, respectively.

The covariance of the prediction error $\mathbf{x}^{(s)}_k - \hat{\mathbf{x}}^{(s)}_{k|k-1}$ is

$$\begin{aligned}\mathbf{P}^{(s)}_{k|k-1} =& \mathbf{F}^{(s)}_{k-1,1}\mathbf{P}^{(s)}_{k-1|k-1}\mathbf{F}^{(s)T}_{k-1,1}\\ &+ \mathbf{F}^{(s)}_{k-1,2}\mathbf{P}^{(-s)}_{k-1|k-1}\mathbf{F}^{(s)T}_{k-1,2}\\ &+ 2\cdot\mathbf{F}^{(s)}_{k-1,1}\mathbf{P}^{(s)}_{cross}\mathbf{F}^{(s)T}_{k-1,2} + \mathbf{Q}^{(s)}_{k-1},\end{aligned} \tag{50}$$

instead of what we would expect

$$\mathbf{P}^{(s)}_{k|k-1} = \mathbf{F}^{(s)}_{k-1,1}\mathbf{P}^{(s)}_{k-1|k-1}\mathbf{F}^{(s)T}_{k-1,1} + \mathbf{Q}^{(s)}_{k-1}, \tag{51}$$

where

$$\mathbf{P}_{k|k-1} = \left(\begin{array}{c|c}\mathbf{P}^{(s)}_{k|k-1} & \mathbf{P}^{(s)}_{cross}\\ \hline \mathbf{P}^{(s)T}_{cross} & \mathbf{P}^{(-s)}_{k|k-1}\end{array}\right), \tag{52}$$

and $\mathbf{P}^{(s)T}_{cross} = \mathbb{E}\left\{\left(\mathbf{x}^{(s)}_k - \hat{\mathbf{x}}^{(s)}_{k|k-1}\right)\left(\mathbf{x}^{(-s)}_k - \hat{\mathbf{x}}^{(-s)}_{k|k-1}\right)^T\right\}$. Thus, the effect of using the estimation $\hat{\mathbf{x}}^{(-s)}_{k-1|k-1}$ instead of the true value $\mathbf{x}^{(-s)}_{k-1}$ is that we introduce noise into the system, as it is evidenced in (50) and (51). The prediction error cross covariance matrix between the s-th subspace and the rest is not available, and this is precisely the information that the algorithm discards when using the partitioning strategy.

Regarding the measurement update step, we can see that the different subspaces are interconnected to compute the innovations,

$$\hat{\mathbf{y}}^{(s)}_{k|k-1} = \mathbf{H}^{(s)}_k\begin{bmatrix}\hat{\mathbf{x}}^{(s)}_{k|k-1}\\ \hat{\mathbf{x}}^{(-s)}_{k|k-1}\end{bmatrix}, \tag{53}$$

and the corresponding innovation covariance is

$$\boldsymbol{\Sigma}^{(s)}_{yy,k|k-1} = \mathbf{H}^{(s)}_k\mathbf{P}_{k|k-1}\mathbf{H}^{(s)T}_k + \mathbf{R}_k, \tag{54}$$

with $\mathbf{P}_{k|k-1}$ defined as in (52). In this case, we are also introducing noise into the innovation, because the terms $\mathbf{P}^{(-s)}_{k|k-1}$ and $\mathbf{P}^{(s)}_{cross}$ should be null for equality.

Regarding the state estimation we can see that this also introduces an error

$$\begin{aligned}\hat{\mathbf{x}}^{(s)}_{k|k} &= \hat{\mathbf{x}}^{(s)}_{k|k-1} + \mathbf{K}^{(s)}_k\left(\mathbf{y}_k - \hat{\mathbf{y}}^{(s)}_{k|k-1}\right)\\ &= \hat{\mathbf{x}}^{(s)}_{k|k-1} + \mathbf{K}^{(s)}_k\left(\mathbf{y}_k - \mathbf{H}^{(s)}_{k,1}\hat{\mathbf{x}}^{(s)}_{k|k-1} - \mathbf{H}^{(s)}_{k,2}\hat{\mathbf{x}}^{(-s)}_{k|k-1}\right)\end{aligned} \tag{55}$$

where we used a similar partition than for the state transition matrix

$$\mathbf{H}_k^{(s)} = \left(\mathbf{H}_{k,1}^{(s)} \middle| \mathbf{H}_{k,2}^{(s)} \right) \tag{56}$$

with dimensions $\left(n_x \times n_x^{(s)}\right)$ and $\left(n_x \times n_x^{(-s)}\right)$, respectively.

C. On the State Partitioning

We have seen in the previous subsections that the MQKF does not take into account possible correlations among noise process subspaces, therefore, there is a possible loss of information regarding interconnections among such subspaces due to the architecture of the algorithm. As a result, the information that is lost using the multiple approach is directly related to the partitioning, which is set by the user and should be determined according to the model. The questions to be considered here are: "How do we specify the state partitioning?" and "Which is the best partitioning?".

For the first question, the most important point to consider is that the partitioning should be based on the model under study. The preferred partitioning design depends on the application, the degree of the nonlinearity and the available computational resources. It is important to keep in mind that the required computational complexity which defines the resources needed, is directly related to the number of sigma-points per dimension, α, and the subspace dimensions, $n_x^{(s)}$. For a fixed complexity, the best partitioning is the one that groups the states which are highly correlated and puts the uncorrelated ones (or with low correlation) into different subspaces.

For example, in a multiple target tracking application, we would group the parameters of each target into a subspace. Otherwise, we could use an adaptive state partitioning method which first computes the correlation among states and then sets the partitions according to the result obtained.

IV. A Complexity Result for the MQKF

The time complexity of an algorithm can be viewed as the number of floating point operations it performs per time step. The asymptotic time complexity of square-root SPKFs is known to be $\mathcal{O}(n^3 + Ln^2)$, where $n = \max(n_x, n_y)$ and L represents the number of sigma points. This computation only considers the cost of the matrix operations, without taking into account the L evaluations of the nonlinear functions, which implementation can range from a look-up table to very complex numerical methods that could represent the actual computational bottleneck of the algorithm. More details of this analysis can be found in [14], [19]. Thus, the number of sigma points L is the dominant parameter in QKF schemes, as L grows exponentially with the state dimension, and can be considered an indicator of the computational complexity of the algorithms.

Whereas the number of points used by a QKF is $L = \alpha^{n_x}$, the total number of points generated by the MQKF is $L_M = \sum_{s=1}^{S} \alpha^{n_x^{(s)}}$. The computational complexity of SPKFs in general is highly related to the number of generated points, and thus reducing this value has a beneficial impact on the implementation cost. In this section, we are interested in identifying whether L is larger than L_M in general, and under which conditions. The result is stated in Proposition 1.

Proposition 1: Let $L = \alpha^{n_x}$ and $L_M = \sum_{s=1}^{S} \alpha^{n_x^{(s)}}$ with $S \leq n_x \in \mathbb{Z}$, $n_x^{(s)} \in \{1, \ldots, n_x\} \subset \mathbb{Z}$, and guaranteeing that $\sum_{s=1}^{S} n_x^{(s)} = n_x$. Then,

$$L > L_M \quad \text{if} \quad \alpha > S^{\frac{1}{S-1}} \tag{57}$$

for $S \geq 2$ and $\alpha > 2$.

Proof: The proof is given in Appendix A. ∎

This result proves that the MQKF is always reducing the number of generated quadrature points, and thus the computational complexity of the overall filter.

As a final remark, it should be pointed out that the multiple filtering architecture might not be suitable in general SPKFs. For instance, the UKF requires $2n_x + 1$ sigma-points, whereas a Multiple UKF would require $\sum_{s=1}^{S}(2n_x^{(s)} + 1) = 2n_x + S$, that is an increase of $S - 1$ points. Similarly, the CKF generates $2n_x$ cubature points, and a Multiple CKF would use exactly the same amount.

V. Computer Simulations

In this Section, in order to provide illustrative numerical results, the performance of the proposed method is shown in an academic example and a practical estimation problem.

A. Illustrative Academic Example

We considered the academic example proposed in [12], where state dimension can be arbitrarily set. States evolved according to $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{\nu}_{k-1}$, with $\boldsymbol{\nu}_{k-1}$ a zero-mean Gaussian noise with covariance matrix $\mathbf{Q}_{k-1}$, and measurements were obtained as

$$y_k = \left(\sqrt{1 + \mathbf{x}_k^T \mathbf{x}_k}\right)^q + n_k, \tag{58}$$

where q is a parameter used to tune the nonlinearity of the function and $n_k \sim \mathcal{N}(0, \sigma_n^2)$.

With this setup, we compared the Root Mean Square Error (RMSE) of 6 nonlinear filters in their square-root version (i.e., with enhanced numerical stability), namely, the standard SQKF and the MSQKF with 5 different state partitions (see Table I). We averaged over 5000 independent Monte Carlo runs, with 100 measurements per run. The standard QKF was used as a benchmark for this problem, and the results were also compared to the CKF and UKF, being the state-of-the-art reference methods for lightweight nonlinear Gaussian filtering.

As the filters not always converged, we define the % of failure: for each realization, a filter is said to fail if its resulting RMSE is larger than the RMSE at initialization. In the figures we only plot the RMSE of the realizations which did not fail, while the % of failure is shown in the corresponding tables. Notice that the results in this section do not include those obtained by an Extended Kalman filter (EKF), the reason is that EKF exhibited a 100% of failure, that is it diverged in this highly nonlinear example.

In order to assess the performance of the proposed method we considered three explanatory cases:

1) Case 1: In this first case, we considered that the state transition and the process noise covariance matrices were diagonal, so the state evolved according to $\mathbf{x}_k \sim \mathcal{N}(0.8 \cdot \mathbf{x}_{k-1}, \mathbf{Q})$, where

TABLE I
DIFFERENT SUBSPACE PARTITIONING AND THE CORRESPONDING NUMBER OF SIGMA-POINTS FOR $\alpha = 5$ AND $n_x = 6$

Filter	State partitioning	Sigma-points
SQKF	full state space	$L = \alpha^{n_x} = 15625$
S1	MSQKF - full state space	$L_M = 15625$
S2	Triples - $\{1,2,3\}\{4,5,6\}$	$L_M = 2\alpha^3 = 250$
S3	Pairs - $\{1,2\}\{3,4\}\{5,6\}$	$L_M = 3\alpha^2 = 75$
S4	{1}{2}{3,4}{5,6}	$L_M = 2\alpha + 2\alpha^2 = 60$
S5	{1}{2}{3}{4}{5}{6}	$L_M = 6\alpha = 30$
CKF	full state space	$L_C = 2n_x = 12$
UKF	full state space	$L_U = 2n_x + 1 = 13$

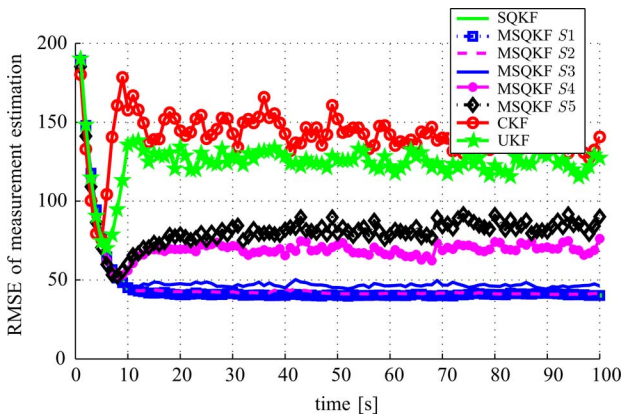

Fig. 1. Case 1: RMSE of state estimation, with $\hat{\mathbf{Q}}_k = \mathbf{Q}$.

$\mathbf{F}_{k-1} = 0.8 \cdot \mathbf{I}$, $\mathbf{Q} = \sigma_\nu^2 \cdot \mathbf{I}$, and $\sigma_\nu = 10$. The measurement noise variance was set to $\sigma_n^2 = 10$, $\alpha = 5$, $n_x = 6$ and $q = 11$. The filter used $\hat{\mathbf{Q}}_k = \mathbf{Q}$ and $\hat{\mathbf{R}}_k = \sigma_n^2$, and was initialized as $\mathcal{N}\left(\mathbf{x}_0, 100^2\mathbf{I}\right)$. The MSQKF with $S = 1$ ($S1$ filter) is equivalent to the conventional SQKF, which is plotted in order to validate our implementation (we will not plot again the results of $S1$ in the following figures). As simulations show, Fig. 1, a remarkable reduction of the computational cost can be achieved by employing the multiple architecture with moderated (depending on the state-space partitioning) accuracy degradation, and always obtaining better results than the UKF and the CKF. The reason is that the third-order rules used by the UKF and CKF are not able to characterize the high nonlinearity of the model. Notice that the standard SQKF and the MSQKF considering the full state use 15625 points, and the MSQKF with more than 2 subspaces use less than 100 points (Table I), with the consequent reduction in the overall computational cost.

The filters which considered partitions with dimension $n_x^{(s)} = 1$, namely $S4$ and $S5$, provided a worse estimation performance because the errors introduced by the MSQKF implementation (see Section III-B2) are larger than in case of considering triples $(S2)$ and pairs $(S3)$. Moreover, the smaller is the subspaces dimension, the larger is the estimation error. To overcome this problem we could to increase the noise covariance matrices, to allow the filters converge to the results obtained by the SQKF, which is known to be optimal in the minimum-mean-square-error (MMSE), the maximum likelihood (ML), and the maximum a posteriori (MAP) senses for

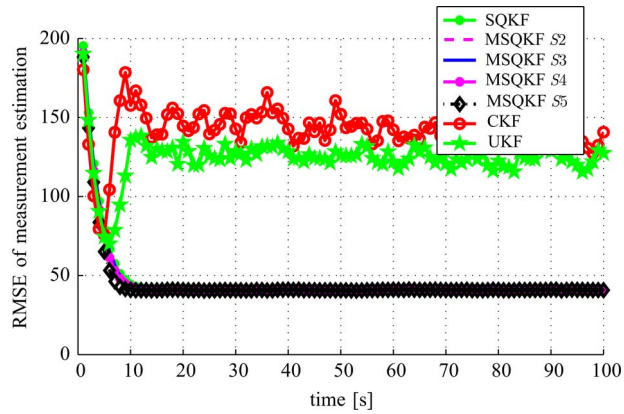

Fig. 2. Case 1: RMSE of state estimation, with an overestimated process noise covariance $\hat{\mathbf{Q}}_k = 100 \cdot \mathbf{Q}$.

TABLE II
% OF FAILURE FOR THE 8 FILTERS FOR CASE 1 WITH TWO CONFIGURATIONS OF $\hat{\mathbf{Q}}_k$

Filter	$\hat{\mathbf{Q}}_k = \mathbf{Q}$	$\hat{\mathbf{Q}}_k = 100 \cdot \mathbf{Q}$
SQKF	0.0%	0.0%
S1	0.0%	0.0%
S2	0.0%	0.0%
S3	0.0%	0.0%
S4	4.0%	0.0%
S5	13.0%	0.0%
CKF	80.0%	80.0%
UKF	50.3%	50.3%

all functions $\mathbf{f}_{k-1}(\cdot)$ and $\mathbf{h}_k(\cdot)$ expressible as a polynomial of degree $\leq 2\alpha^{n_x} - 1$[20]. Fig. 2 shows the results obtained with a modified process covariance matrix: $\hat{\mathbf{Q}}_k = 100 \cdot \mathbf{Q}_k$. With this configuration we can see that all the MSQKFs performed as the SQKF with notable savings in the computational cost.

Table II provides the % of failure for each filter. Considering the setup with $\hat{\mathbf{Q}}_k = \sigma_\nu^2\mathbf{I}$, we can see that filters $S2$ and $S3$ never failed in 5000 realizations. When compared to the CKF and the UKF, the % of failure for the MSQKF is much lower even in the worst partitioning cases ($S4$ and $S5$). When overestimating the process noise covariance, we decrease also the number of fails.

2) Case 2: In this case, we used the same setup as in Case 1 but introducing correlation in the process noise, that is, a non-diagonal covariance matrix $\mathbf{Q}$. We considered a weak correlation matrix $\mathbf{Q}'$, with values σ_ν^2 in the main diagonal and $0.1 \cdot \sigma_\nu^2$ in the off-diagonal entries; and a strong correlation matrix $\mathbf{Q}''$ with values σ_ν^2 in the main diagonal and $0.5 \cdot \sigma_\nu^2$ in the off-diagonal entries.

In the weak correlation case, $\mathbf{Q}'$, we can see (Fig. 3) that all the filters deal correctly with the process noise correlation, because there is few information outside the block matrices considered $\mathbf{Q}_{k-1}^{(s)}$. When dealing with a strong correlation, $\mathbf{Q}''$, there is much more information outside $\mathbf{Q}_{k-1}^{(s)}$ which is lost, so a greater error is introduced. Notice that the smaller the subspace

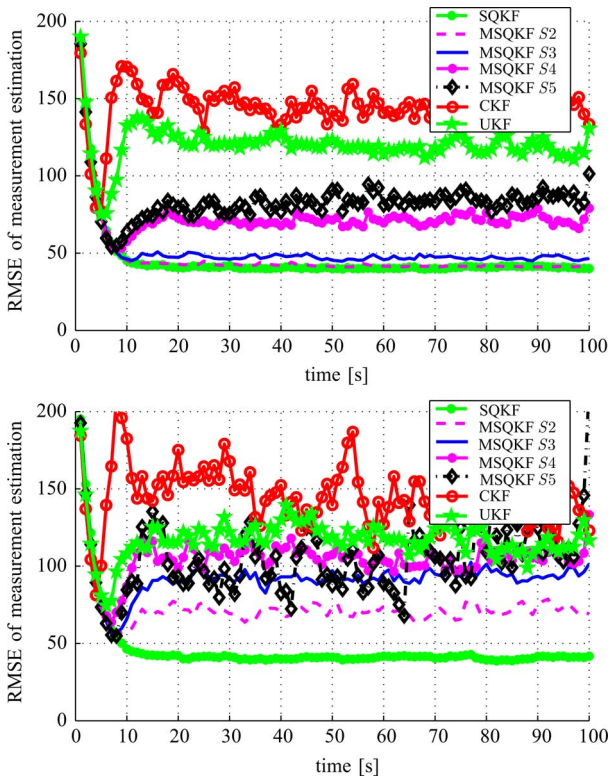

Fig. 3. Case 2: RMSE with two different correlated process noise covariance matrices, namely, $\mathbf{Q}'$ (top) and $\mathbf{Q}''$ (bottom).

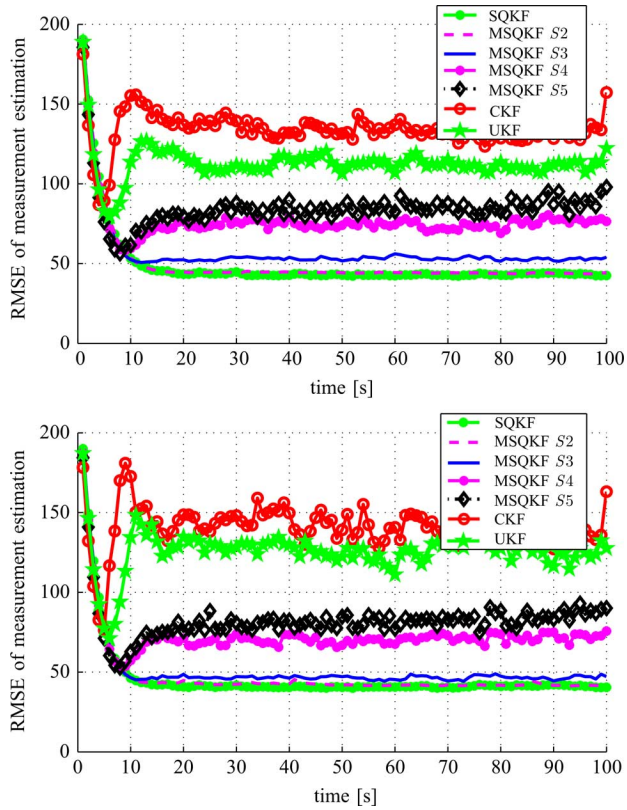

Fig. 4. Case 3: RMSE with two different state transition matrices (3×3 and 2×2 correlations), namely, $\mathbf{F}^1_{k-1}$ (top) and $\mathbf{F}^2_{k-1}$ (bottom).

TABLE III
% OF FAILURE FOR THE 8 FILTERS IN BOTH SCENARIOS FOR CASE 2

Filter	$\hat{\mathbf{Q}}_k = \mathbf{Q}'$	$\hat{\mathbf{Q}}_k = \mathbf{Q}''$	$\hat{\mathbf{Q}}_k = 100 \cdot \mathbf{Q}''$
SQKF	0.0%	0.0%	0.0%
S2	0.0%	10%	0.0%
S3	0.3%	33.1%	0.0%
S4	5.9%	69.9%	0.0%
S5	21%	96.8%	0.0%
CKF	83.4%	95.2%	95.2
UKF	44.4%	60.2%	60.2

dimensions are, the larger the amount of information lost and the bigger the estimation errors are.

Table III shows that with a weak correlation, the % of failure for the MSQKFs is slightly higher than in the reference case $\hat{\mathbf{Q}}_k = \mathbf{Q}$, but in the strong correlation scenario only $S2$ and $S3$ resist to fail most of the time. Thus, in such strongly correlated noise scenarios, the partitions with higher dimension capture more information, and this constitutes the recommended partition strategy. With adequate tuning of the process noise covariance matrix, even in the strong correlation case, the MSQKFs perform as the SQKF with a drastically reduced computational cost, obtaining the same results as plotted in Fig. 2.

3) Case 3: In the third case we introduced an interconnection among states in their evolution, that is, a non-diagonal state transition matrix $\mathbf{F}_{k-1}$. We considered a first scenario where states were correlated in triples:

$$\mathbf{F}^1_{k-1} = \begin{pmatrix} 1 & 0.1 & -0.05 & 0 & 0 & 0 \\ -0.1 & 1 & 0.1 & 0 & 0 & 0 \\ 0.05 & 0.1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0.1 & -0.05 \\ 0 & 0 & 0 & -0.1 & 1 & 0.1 \\ 0 & 0 & 0 & 0.05 & 0.1 & 1 \end{pmatrix},$$

and a second scenario where the states were correlated in pairs:

$$\mathbf{F}^2_{k-1} = \begin{pmatrix} 1 & -0.1 & 0 & 0 & 0 & 0 \\ 0.05 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -0.1 & 0 & 0 \\ 0 & 0 & 0.05 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -0.1 \\ 0 & 0 & 0 & 0 & 0.05 & 1 \end{pmatrix}.$$

Simulation results are in Fig. 4 and Table IV. Notice that, as expected, the error for $S3$ (which considers the partition in pairs) was larger with $\mathbf{F}^1_{k-1}$ than with $\mathbf{F}^2_{k-1}$. It can be explained by the fact that partition strategy is optimal when it fits the interconnection among states. So the partitioning should be related to the model, that is, considering correlated states in the same subspace provides better results.

When overestimating the process noise covariance, as done in Case 1 and Case 2, we obtained optimal results with all the MSQKFs, because a larger process noise covariance matrix allowed the filter to converge and implicitly introduced a higher

TABLE IV
% OF FAILURE FOR THE 8 FILTERS IN BOTH SCENARIOS FOR CASE 3

Filter	$\mathbf{F}^1_{k-1}$, $\hat{\mathbf{Q}}_k = \mathbf{Q}$	$\mathbf{F}^2_{k-1}$, $\hat{\mathbf{Q}}_k = \mathbf{Q}$	$\mathbf{F}^2_{k-1}$, $\hat{\mathbf{Q}} = 100 \cdot \mathbf{Q}$
SQKF	0.0%	0.0%	0.0%
S2	0.0%	0.0%	0.0%
S3	0.3%	0.2%	0.0%
S4	5.3%	4.5%	0.0%
S5	17.9%	16.6%	0.0%
CKF	65.7%	83.9%	83.9%
UKF	28.6%	52.5%	52.5%

uncertainty on the system model, that is, trusting the observation much more than the model itself. This is not a general statement and an overestimated process noise covariance may give worse performances depending on the application and the overestimation ratio.

B. Multiple Superimposed Sinusoids Estimation

In the presented application, we considered the problem of estimating the parameters of P superimposed sinuosoids [21], which is a problem appearing in many engineering fields, such as power systems [22]–[24], nuclear science [25] or communications [26].

The complex measurements were obtained as

$$y_k = \sum_{i=1}^{P} a_{i,k} \exp(j(2\pi f_{i,k} k)) + n_k, \tag{59}$$

which can be rewritten as

$$\mathbf{y}_k = \begin{bmatrix} \sum_{i=1}^{P} a_{i,k} \cos(2\pi f_{i,k} k) \\ \sum_{i=1}^{P} a_{i,k} \sin(2\pi f_{i,k} k) \end{bmatrix} + \mathbf{n}_k, \tag{60}$$

where $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R} = \sigma_n^2 \mathbf{I})$. The state consisted of the amplitudes $a_{i,k}$ and the normalized frequencies $f_{i,k}$ per sinusoid,

$$\mathbf{x}_k = \{a_{i,k}, f_{i,k}\}_{i=1}^{P}, \tag{61}$$

that is, $2 \times P$ parameters were estimated. For each parameter we considered a random walk evolution, so the state evolved according to $\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \boldsymbol{\nu}_k$ with $\mathbf{F}_{k-1} = \mathbf{I}$, $\boldsymbol{\nu}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{Q} = \text{diag}\left(\left\{\sigma^2_{a_i}, \sigma^2_{f_i}\right\}_{i=1}^{P}\right)$.

We considered 3 sinusoids with the following initial parameters: $f_1 = 100$ Hz, $f_2 = 1$ KHz, $f_3 = 2$ KHz, $a_1 = 5$ V, $a_2 = 4$ V, $a_3 = 3$ V, so the initial state was $\mathbf{x}_0 = [0.02, 0.2, 0.4, 5, 4, 3]^T$. The signal was sampled at 5 KHz, the variances of the Gaussian process noise driving the states were $\sigma^2_{a_i} = 0.5\ \mu\frac{\text{V}}{\text{ms}}$ and $\sigma^2_{f_i} = 0.1\ \mu\frac{\text{Hz}}{\text{ms}}$, and the measurement noise variance was set to $\sigma^2_n = 0.09\ \text{V}^2$. The filter was set with $\hat{\mathbf{Q}}_k = \mathbf{Q}$ and $\hat{\mathbf{R}}_k = \sigma^2_n$, and initialized with a random variable with distribution $\mathcal{N}(\mathbf{x}_0, \mathbf{C})$, where $\mathbf{C} = \text{diag}(0.05, 0.05, 0.05, 0.5, 0.5, 0.5)$.

With this setup, we compared the RMSE of the estimation of the frequencies, $\text{RMSE}_f = \sqrt{\frac{(\text{MSE}_{f_1} + \text{MSE}_{f_2} + \text{MSE}_{f_3})}{3}}$, using

TABLE V
DIFFERENT SUBSPACE PARTITIONING AND THE CORRESPONDING NUMBER OF SIGMA-POINTS FOR THE MULTIPLE SUPERIMPOSED SINUSOIDS ESTIMATION PROBLEM

Filter	State partitioning	Sigma-points
SQKF	full state space	$L = \alpha^{n_x} = 729$
S1	$\{f_1, f_2, f_3\}\{a_1, a_2, a_3\}$	$L_M = 2\alpha^3 = 54$
S2	$\{f_1, a_1\}\{f_2, a_2\}\{f_3, a_3\}$	$L_M = 3\alpha^2 = 27$
S3	$\{f_1\}\{f_2\}\{f_3\}\{a_1\}\{a_2\}\{a_3\}$	$L_M = 6\alpha = 18$
CKF	full state space	$L_C = 2n_x = 12$

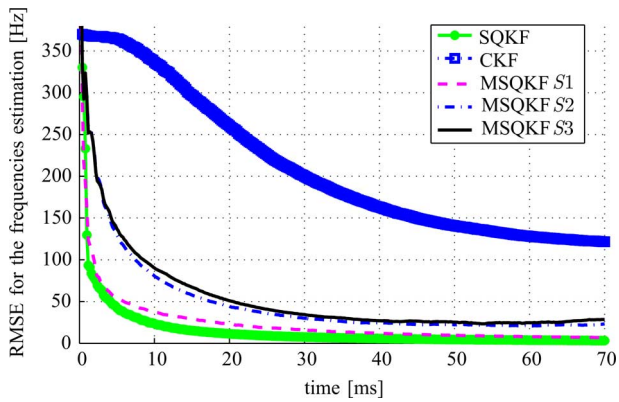

Fig. 5. Results of the estimation of the frequencies for the multiple superimposed sinusoids problem.

TABLE VI
% OF FAILURE FOR THE MULTIPLE SINUSOIDS ESTIMATION PROBLEM

Filter	SQKF	CKF	S1	S2	S3
Failure	1%	57%	2.6%	7%	9%

the filters in Table V. As in the case of the academic example, the EKF diverged for all realizations. We made 400 independent Monte Carlo runs with 350 measurements per run (70 ms), and we set the number of sigma-point per dimension to $\alpha = 3$.

The % of failure was defined as follows: a realization was said to fail if the error on the estimation of the frequencies was larger than 300 Hz. Considering this threshold we obtained the results plotted in Fig. 5 and the % of failure given in Table VI. The simulations show the good performance of the MSQKF, corroborating the results obtained in the previous example. The MSQKF improves the CKF with a similar number of sigma-points ($S3$ uses 18 points) while being more robust, and with a suitable partitioning ($S1$) gives almost optimal performance compared to the SQKF with a notable complexity reduction (729 vs 54 sigma-points).

VI. CONCLUSIONS

The QKF constitutes a powerful tool for dealing with nonlinear dynamic systems. However, in order to estimate accurately the posterior distribution, it requires a number of sigma-points that increases exponentially with the dimension of the state-space model.

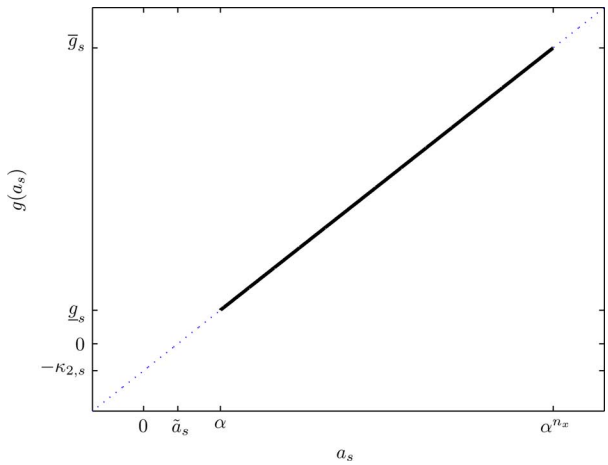

Fig. 6. Plot of function $g(a_s)$ (dotted line) with the evaluation over the valid support $[\alpha, \alpha^{n_x}]$ in solid thick line.

In this paper, a new method based on the partitioning of the state-space, called Multiple QKF, was proposed in order to alleviate this issue. It was proved in Proposition 1 that the state-partitioning approach reduces the computational complexity of a QKF operating on the complete system.

The performance of the filter was assessed by computer simulations using an academic example, and compared to the standard QKF, CKF and UKF, with a remarkable reduction of the computational cost with a relatively low performance degradation compared to the QKF, while improving the performance of the other state-of-the-art methods. The new method's performance was also shown for a practical estimation problem (where the standard QKF is not an option because of its high computational cost), corroborating the good results obtained in the first example.

The assumptions made are that the dynamic system is driven by Gaussian processes and that the possible nonlinearities are approximated by the Gauss-Hermite quadrature rules. Separability of process function is not required, although it is desirable to partition the state-space in such a way that one groups the states which are highly correlated and puts the uncorrelated ones (or with low correlation) into different subspaces. The simplifications made are that the subspace cross-terms of the noise covariance are neglected and that the coupling among filters is performed by point estimates. Depending on the partitioning, these simplifications may degrade the performance of the filter. We have observed that overestimation of the model covariance might result in more robust implementations of the MQKF when subspaces are correlated.

To conclude, the MQKF can be considered to be a versatile tool for practitioners with an applicability covering a large variety of high-dimensional problems.

APPENDIX
PROOF OF PROPOSITION 1

Proof: Let us operate with the difference $\Delta = L - L_M$:

$$\Delta = \alpha^{\sum_{s=1}^{S} n_x^{(s)}} - \sum_{s=1}^{S} \alpha^{n_x^{(s)}} \tag{62}$$

$$= \prod_{s=1}^{S} \alpha^{n_x^{(s)}} - \sum_{s=1}^{S} \alpha^{n_x^{(s)}}, \tag{63}$$

we would like to see if $\Delta > 0$, that is under which conditions a MQKF scheme provides computational cost reduction with respect to conventional QKF. We resort to the procedure used to proof some set of Weierstrass' inequalities [27], which tackle a similar problem[4].

For the sake of clarity, let us define $a_s \triangleq \alpha^{n_x^{(s)}}$, then we can express Δ as the following function of interest

$$g(a_1, \ldots, a_S) = \prod_{s=1}^{S} a_s - \sum_{s=1}^{S} a_s, \tag{64}$$

which is multivariate and depending on the different dimensions of each subspace s.

Since we seek to proof that $g(a_1, \ldots, a_S) > 0$, it suffices to show that the function is lower bounded by a positive constant. Let us thus analyze and characterize the function as follows, with Fig. 6 being a supporting illustration for the mathematical derivations. Suppose now that all but one variable are fixed, then we obtain a linear function of a single variable

$$g(a_s) = (\kappa_{1,s} - 1)a_s - \kappa_{2,s}, \tag{65}$$

with constants

$$\kappa_{1,s} = \prod_{\substack{s'=1 \\ s' \neq s}}^{S} a_{s'} \quad \text{and} \quad \kappa_{2,s} = \sum_{\substack{s'=1 \\ s' \neq s}}^{S} a_{s'}. \tag{66}$$

The support in which function (65) can be evaluated is in the closed set $\alpha \leq a_s \leq \alpha^{n_x}$ in the real line, with the limit points corresponding to the cases $S = n_x$ (i.e., each state has an associated filter) and $S = 1$ (i.e., the conventional QKF without state partitioning), respectively. Since (65) is linear on a_s, its lower bound is given by evaluation of one of its limit points depending on the sign of the slope $m_s = \kappa_{1,s} - 1$.

We can easily see that $m_s > 0, \ \forall s \in \mathcal{S}$, if we identify that the dimension of subspaces is at least equal to 1 and that the minimum number of quadrature points[5] is 2. Then,

$$\begin{array}{rcl} n_x^{(s)} & \geq & 1 \\ \alpha & \geq & 2 \end{array} \Rightarrow a_s = \alpha^{n_x^{(s)}} > 1 \tag{67}$$

and thus

$$\kappa_{1,s} > 1 \Rightarrow m_s > 0, \ \forall s \in \mathcal{S}. \tag{68}$$

Therefore, the supremum and infimum of $g(a_s)$ are

$$\bar{g}_s = \sup\{g(a_s)\} = g(\alpha^{n_x}) \tag{69}$$

$$\underline{g}_s = \inf\{g(a_s)\} = g(\alpha) \tag{70}$$

[4]More precisely, we recall the Weierstrass' product inequality that states that

$$\prod_{i=1}^{n}(1 \pm \beta_i) \geq 1 \pm \sum_{i=1}^{n} \beta_i,$$

where $\beta_i \in [0, 1] \in \mathbb{R}$ and $i = 1, \ldots, n$.

[5]Note that for $\alpha = 1$, Gauss-Hermite rules are exact for polynomials of degree 1. The case of linear systems is not of interest in the context of QKF as these could be tackled optimally by other means, namely the KF. Also, first-order approximations could be obtained by the Extended KF.

for any subspace $s \in \mathcal{S}$, respectively. Notice that the root for which $g(\tilde{a}_s) = 0$ is positive

$$\tilde{a}_s = \frac{\kappa_{2,s}}{\kappa_{1,s} - 1} > 0. \tag{71}$$

The analysis holds $\forall s \in \mathcal{S}$, and thus it can be extended to the rest of variables in (64) to see that the lower bound is

$$\inf\{g(a_1, \ldots, a_S)\} = g(\alpha, \ldots, \alpha) \tag{72}$$

$$= \prod_{s=1}^{S} \alpha - \sum_{s=1}^{S} \alpha = \alpha^S - S\alpha \tag{73}$$

and then $\Delta \geq \alpha^S - S\alpha$. To show that $\Delta > 0$ it suffices to prove that $\alpha^S - S\alpha > 0$. After straightforward mathematical manipulation of the latter, we found the following condition to ensure strict positiveness of Δ:

$$\alpha > S^{\frac{1}{S-1}} \tag{74}$$

which is easy to attain if $S \geq 2$, *q.e.d.* ■

REFERENCES

[1] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Autom. Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[2] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.

[3] F. Daum and J. Huang, "Curse of dimensionality and particle filters," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, Mar. 2003, vol. 4, pp. 1979–1993.

[4] T. Schön, F. Gustafsson, and P. Nordlund, "Marginalized particle filters for mixed linear/nonlinear state-space models," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2279–2289, Jul. 2005.

[5] M. F. Bugallo, T. Lu, and P. M. Djuric, "Target tracking by multiple particle filtering," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, Mar. 2007.

[6] P. M. Djuric, T. Lu, and M. F. Bugallo, "Multiple particle filtering," presented at the ICASSP, Honolulu, HI, Apr. 2007.

[7] P. Closas and M. F. Bugallo, "Improving accuracy by iterated multiple particle filtering," *IEEE Signal Process. Lett.*, vol. 19, no. 8, pp. 531–534, Aug. 2012.

[8] W. I. Tam and D. Hatzinakos, "An efficient radar tracking algorithm using multidimensional Gauss-Hermite quadratures," presented at the ICASSP, Munich, Germany, Apr. 1997.

[9] R. Van der Merwe and E. Wan, "Sigma-point Kalman filters for probabilistic inference in dynamic state-space models," presented at the Workshop on Adv. Mach. Learn., Montreal, QC, Canada, Jun. 2003.

[10] S. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," in *Proc. Int. Symp. Aerosp./Defense Sens., Simul. Contr.*, Orlando, FL, 1997, vol. 3, pp. 182–193.

[11] R. V. d. Merwe and E. A. Wan, "The square-root unscented Kalman filter for state and parameter estimation," in *Proc. ICASSP*, Salt Lake City, UT, May 2001, vol. 6, pp. 3461–3464.

[12] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Trans. Autom. Control*, vol. 54, no. 6, pp. 1254–1269, Jun. 2009.

[13] G. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, pp. 413–432, 1973.

[14] I. Arasaratnam and S. Haykin, "Square-root quadrature Kalman filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2589–2593, Jun. 2008.

[15] C. Fernández-Prades and J. Vilà-Valls, "Bayesian nonlinear filtering using quadrature and cubature rules applied to sensor data fusion for positioning," presented at the IEEE Int. Conf. Commun., Cape Town, South Africa, May 2010.

[16] B. Jia, M. Xin, and Y. Cheng, "Sparse Gauss-Hermite quadrature filter with application to spacecraft attitude estimation," *J. Guid., Contr., Dyn.*, vol. 34, no. 2, pp. 367–379, Mar./Apr. 2011.

[17] I. Arasaratnam, S. Haykin, and R. J. Elliot, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proc. IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[18] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The John Hopkins Univ. Press, 1996.

[19] P. Closas and C. Fernández-Prades, "Bayesian nonlinear filters for direct position estimation," presented at the IEEE Aerosp. Conf., Big Sky, MT, Mar. 2010.

[20] G. H. Golub and J. H. Welsch, "Calculation of Gauss quadrature rules," *Math. Computat.*, vol. 23, no. 106, pp. 221–230, Apr. 1969.

[21] T. Hilands and S. Thomopoulos, "Nonlinear filtering methods for harmonic retrieval and model selection in Gaussian and non-Gaussian noise," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 982–995, Apr. 1997.

[22] P. K. Dash and R. K. Mallik, "Accurate tracking of harmonic signals in VSC-HVDC systems using PSO based unscented transformation," *Elect. Power Energy Syst.*, vol. 33, pp. 1315–1325, 2011.

[23] P. K. Dash, S. Hasan, and B. K. Panigrahi, "A hybrid unscented filtering and particle swarm optimization technique for harmonic analysis of nonstationary signals," *Measurement*, vol. 43, pp. 1447–1457, 2010.

[24] J. B. V. Reddy, P. K. Dash, R. Samantaray, and A. K. Moharana, "Fast tracking of power quality disturbance signals using an optimized unscented filter," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 12, pp. 3943–3952, Dec. 2009.

[25] D. Alves and R. Coelho, "An adaptive algorithm for real-time multi-tone estimation and frequency tracking of non-stationary signals," *IEEE Trans. Nucl. Sci.*, vol. 58, no. 4, pp. 1582–1587, Aug. 2011.

[26] M. Niedzwiecki and P. Kaczmarek, "Estimation and tracking of complex-valued quasi-periodically varying systems," *Automatica*, vol. 41, pp. 1503–1516, 2005.

[27] D. S. Mitrinovic and P. M. Vasic, *Analytic Inequalities*. New York: Springer-Verlag, 1970.

Pau Closas (S'04–M'10) received the M.Sc. and Ph.D. degrees (*cum-laude*) in electrical engineering from the Technical University of Catalonia (UPC) in 2003 and 2009, respectively. In 2005, he received a Spanish Ministry of Education Ph.D. grant.

In 2003, he joined the Department of Signal Theory and Communications, UPC, as a Research Assistant. During 2008, he was Research Visitor at the Stony Brook University (SBU), New York. In September 2009, he joined the CTTC, where he currently holds a position as a Research Associate in the Communications Subsystems Area and coordinates the Positioning Systems research line. He has many years of experience in projects funded by the European Commission, Spanish and Catalan Governments, as well as the European Space Agency (ESA) in both technical and managerial duties. He has collaborated in projects of the USA's National Science Foundation (NSF) and the National Natural Science Foundation (NNSF) of China. He has numerous contributions in his primary areas of interest, which include statistical and array signal processing, estimation and detection theory, Bayesian filtering, robustness analysis, and game theory, with applications to positioning systems, wireless sensor networks, and mathematical biology.

Dr. Closas is a member of the European Association for Signal Processing (EURASIP) and the Institute of Navigation (ION) since 2005, 2006, and 2009, respectively. In 2011, he was involved in the organizing committees of EUSIPCO, IEEE IMWS, and IEEE RFID-TA conferences.

Carles Fernández-Prades (S'02–M'06–SM'12) received the M.S. and Ph.D. (*cum laude*) degrees in electrical engineering from the Technical University of Catalonia (UPC) in 2001 and 2006, respectively.

In 2001, he joined the Department of Signal Theory and Communications, UPC, as a Research Assistant, getting involved in research projects at European and National levels. He served as occasional lecturer (2001–2004) and as Associate Professor (2005) at UPC in the field of analog and digital communications. In 2006, he joined the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) as a Research Associate, although during the second semester of 2007, he was a Visiting Lecturer at the Universidad Tecnológica Metropolitana del Estado

de Chile (UTEM), Santiago. At CTTC, he serves as the coordinator of the Communications Subsystems Area. In his research activities, he has made contributions in synchronization techniques for antenna arrays, nonlinear Bayesian filtering, positioning techniques, design of radio frequency front-ends for GNSS receivers and software defined radio, authoring more than 80 papers in international, peer-reviewed journals and conferences, several book chapters and public research reports. He has been advisor of 2 Ph.D. degree theses and several M.S. degree theses in electrical engineering at UPC and in computer engineering at the Open University of Catalonia (UOC).

Jordi Vilà-Valls received the M.Sc. degree in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, and the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, respectively, in 2006, under a double diploma program. He received the Ph.D. degree in electrical engineering from Grenoble INP in March 2010.

In October 2006, he joined the GIPSA-Lab, Grenoble, France, as a Research Assistant with a French Ministry of Science Ph.D. grant. In 2009, he was a research visitor with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). In October 2010, he joined the Department of Signal Theory and Communications (TSC), UPC, where he currently holds a position as a Postdoctoral Research Associate in the Signal Processing and Communications group (SPCOM). His primary areas of interest include statistical signal processing, Bayesian estimation theory, and adaptive/robust filtering methods, with applications to positioning and tracking systems, wireless sensor networks, and communications.