M. R. Morelande and Á. F. García-Fernández, "Analysis of Kalman Filter Approximations for Nonlinear Measurements," in IEEE Transactions on Signal Processing, vol. 61, no. 22, pp. 5477-5484, Nov.15, 2013, doi: 10.1109/TSP.2013.2279367

Analysis of Kalman filter approximations for nonlinear measurements

Mark R. Morelande and Ángel F. García-Fernández

Abstract

A theoretical analysis is presented of the correction step of the Kalman filter (KF) and its various approximations for the case of a nonlinear measurement equation with additive Gaussian noise. The KF is based on a Gaussian approximation to the joint density of the state and the measurement. The analysis metric is the Kullback-Leibler divergence of this approximation from the true joint density. The purpose of the analysis is to provide a quantitative tool for understanding and assessing the performance of the KF and its variants in nonlinear scenarios. This is illustrated using a numerical example.

EDICS Category: SSP-TRAC, SSP-FILT.

M. R. Morelande is with the Melbourne Systems Laboratory, the University of Melbourne. M.R. Morelande is the corresponding author.

Á.F. García-Fernández is with Departamento de Señales, Sistemas y Radiocomunicaciones, Universidad Politécnica de Madrid, Spain.

I. Introduction

The goal of filtering is to recursively estimate a vector-valued Markov process from noisy, partial observations. In a Bayesian framework, optimal estimation of the state requires the posterior density, i.e., the density of the state conditional on measurements taken up to the current time. Since exact computation of the state is rarely possible, approximations are usually required. A large number of approximations have been developed including grid-based [6], Monte Carlo [3] and Gaussian [12] approximations. The focus here is on the class of Gaussian approximations based on the Kalman filter (KF). Although the KF provides the minimum mean square error (MMSE) only for linear/Gaussian systems, it produces accurate estimates for many nonlinear/non-Gaussian systems. This, along with its computational efficiency, has led to the widespread popularity of the KF and its variants.

The KF recursion can be divided into prediction and correction steps. In our analysis we consider the correction step of the filtering recursion. Our emphasis on the correction step is motivated by the observation that this is generally where errors leading to divergence of the KF approximation of the posterior arise [17]. The KF approximates the posterior mean and covariance matrix by the linear MMSE estimator and its mean square error matrix, respectively. This requires the computation of a number of moments involving the state and the measurements. In many cases these moments cannot be computed exactly. Well-known approximations include analytical linearisation, which leads to the extended KF (EKF) [13], sigma point methods, as in the unscented KF (UKF) [14] and cubature KF (CKF) [1] and quadrature rules such as the Gauss-Hermite filter (GHF) [12]. These filters, which we refer to as KF approximations, are the most common ways of implementing the KF recursion in practice.

An often unspoken assumption is that the KF will provide good performance and is therefore well worth approximating. While this may often be the case it is certainly not always so. Despite this, there seems to be little understanding of when the KF itself will provide an accurate approximation of the posterior. As a result, when a KF approximation fails it may not be clear if this is because the approximation to the KF is poor or the KF itself is a poor approximation to the posterior. In the latter case, the posterior will be poorly approximated regardless of which KF approximation, e.g. EKF, UKF or CKF, is used.

The aim of this paper is to provide some understanding of when the KF and its various approximations can be expected to perform well. We do so by performing an error analysis of the KF and some well-known KF approximations, the EKF, UKF and CKF, for the case of a Gaussian prior and a nonlinear measurement equation with additive Gaussian noise. The performance metric used in the analysis is the Kullback-Leibler divergence (KLD) of the KF approximation of the joint density of the state and measurement from the true joint density. Our analysis provides general results which show the effects of system parameters on the accuracy of KF approximations. No analyses given in the literature provide results of this nature. There have been many simulation analyses comparing KF variants for particular problems [1], [7], [14] but these do not provide general results. Most theoretical work on KF approximations has focussed on stability analyses, e.g., stability of the EKF was analysed in [16], [18] and the UKF was studied in [19]. These analyses are certainly useful for determining the applicability of a particular filter but do not provide specific information about the quality of the approximation it provides compared to alternatives. Several interesting insights into the relationship between various KF approximations were given in [11].

The paper is organised as follows. In Section II the KLD for a general KF approximation is derived, expressions for specific KFs are developed and some pertinent points are discussed. A numerical example is presented in Section III to demonstrate the theoretical results. Conclusions are drawn in Section IV.

II. Analysis of a general Kalman filter

Assume the state vector $\mathbf{x} \in \mathbb{R}^n$ has a prior density $\pi_0(\mathbf{x}) = \mathsf{N}(\mathbf{x}; \mathbf{x}_0, \mathbf{P}_0)$ where $\mathsf{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Conditional on $\mathbf{x}$ the measurement $\mathbf{y} \in \mathbb{R}^m$ is distributed as $g(\mathbf{y}|\mathbf{x}) = \mathsf{N}(\mathbf{y}; \mathbf{h}(\mathbf{x}), \mathbf{R})$ where $\mathbf{h}(\cdot)$ is the measurement function and $\mathbf{R}$ is the measurement noise covariance matrix. The Bayes optimal approach to estimating the state $\mathbf{x}$ given a measurement $\mathbf{y}$ is based on the true joint density

$$
\begin{aligned}
p(\mathbf{x}, \mathbf{y}) &= g(\mathbf{y}|\mathbf{x})\pi_0(\mathbf{x}) && (1)\\
&= \frac{\exp\left(-(\mathbf{x}-\mathbf{x}_0)'\mathbf{P}_0^{-1}(\mathbf{x}-\mathbf{x}_0)/2 - (\mathbf{y}-\mathbf{h}(\mathbf{x}))'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x}))/2\right)}{\sqrt{|2\pi\mathbf{P}_0|\,|2\pi\mathbf{R}|}} && (2)
\end{aligned}
$$

where $'$ is the matrix transpose. The class of KF approximations use a joint density of the form

$$
q(\mathbf{x}, \mathbf{y}) = \frac{\exp\left(-\left[(\mathbf{x}-\mathbf{x}_0)'\mathbf{A}(\mathbf{x}-\mathbf{x}_0) + 2(\mathbf{x}-\mathbf{x}_0)'\mathbf{C}(\mathbf{y}-\widetilde{\mathbf{y}}) + (\mathbf{y}-\widetilde{\mathbf{y}})'\mathbf{B}(\mathbf{y}-\widetilde{\mathbf{y}})\right]/2\right)}{\sqrt{|2\pi\boldsymbol{\Omega}|}} \tag{3}
$$

where $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are given by

$$
\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}' & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0 & \widetilde{\boldsymbol{\Psi}} \\ \widetilde{\boldsymbol{\Psi}}' & \mathbf{R} + \widetilde{\boldsymbol{\Phi}} \end{bmatrix}^{-1} = \boldsymbol{\Omega}^{-1} \tag{4}
$$

The quantities $\widetilde{\mathbf{y}}$, $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Phi}}$ are approximations of the moments:

$$
\begin{aligned}
\hat{\mathbf{y}} &= \int \mathbf{h}(\mathbf{x})\pi_0(\mathbf{x})\,\mathrm{d}\mathbf{x} && (5)\\
\boldsymbol{\Psi} &= \int (\mathbf{x}-\mathbf{x}_0)(\mathbf{h}(\mathbf{x})-\hat{\mathbf{y}})'\pi_0(\mathbf{x})\,\mathrm{d}\mathbf{x} && (6)
\end{aligned}
$$

$$\mathbf{\Phi} = \int (\mathbf{h}(\mathbf{x}) - \hat{\mathbf{y}})(\mathbf{h}(\mathbf{x}) - \hat{\mathbf{y}})' \pi_0(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{7}$$

The exact KF uses $\widetilde{\mathbf{y}} = \hat{\mathbf{y}}$, $\widetilde{\mathbf{\Psi}} = \mathbf{\Psi}$ and $\widetilde{\mathbf{\Phi}} = \mathbf{\Phi}$. If these moments are intractable, as they often are, approximations are required, as in, for example, the EKF and UKF.

We consider the KLD of the KF approximation $q(\cdot)$ from the true joint density $p(\cdot)$. The divergence of $q(\cdot)$ from $p(\cdot)$ is considered rather than the divergence of $p(\cdot)$ from $q(\cdot)$ because it allows the derivation of closed-form expressions in terms of the prior moments (5)-(7). As is noted in [5], the KLD of $q(\cdot)$ from $p(\cdot)$ will be small if $q(\mathbf{x}, \mathbf{y})$ is significantly non-zero when $p(\mathbf{x}, \mathbf{y})$ is. In fact, the KLD of $q(\cdot)$ from $p(\cdot)$ may still be quite small even if the support of $q(\cdot)$ is much greater than that of $p(\cdot)$. This should be kept in mind when interpreting the quantitative results.

To derive an expression for the KLD which can easily be applied to different KF variants it is useful to consider the moment approximations $\widetilde{\mathbf{y}}$, $\widetilde{\mathbf{\Psi}}$ and $\widetilde{\mathbf{\Phi}}$ as perturbations from the linearised moment approximations. These are based on the Taylor series expansion of the measurement function $\mathbf{h}(\cdot)$ about the prior mean $\mathbf{x}_0$:

$$\mathbf{h}(\mathbf{x}) = \mathbf{y}_0 + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{n}(\mathbf{x}) \tag{8}$$

where $\mathbf{y}_0 = \mathbf{h}(\mathbf{x}_0)$, $\mathbf{H} = \nabla \mathbf{h}(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_0}$ and $\mathbf{n}(\mathbf{x})$ contains all higher-order terms. Substituting (8) into (5)-(7) gives the following expressions for the prior moments used in the KF correction:

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{y}_0 + \mathsf{E}(\mathbf{n}(\mathbf{x})) && (9)\\
\mathbf{\Psi} &= \int (\mathbf{x} - \mathbf{x}_0)(\mathbf{y}_0 + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{n}(\mathbf{x}) - \hat{\mathbf{y}})' \pi_0(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= \mathbf{\Psi}_0 + \mathsf{E}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})') && (10)\\
\mathbf{\Phi} &= \int (\mathbf{y}_0 + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{n}(\mathbf{x}) - \hat{\mathbf{y}})(\mathbf{y}_0 + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{n}(\mathbf{x}) - \hat{\mathbf{y}})' \pi_0(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= \mathbf{\Phi}_0 + \mathsf{cov}(\mathbf{n}(\mathbf{x})) + \mathbf{H}\mathsf{E}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})') + \mathsf{E}(\mathbf{n}(\mathbf{x})(\mathbf{x} - \mathbf{x}_0))\mathbf{H}' && (11)
\end{aligned}$$

where $\mathbf{\Psi}_0 = \mathbf{P}_0\mathbf{H}'$ and $\mathbf{\Phi}_0 = \mathbf{H}\mathbf{P}_0\mathbf{H}'$. The moment approximations used in a particular KF variant can then be considered to be of the form:

$$\begin{aligned}
\widetilde{\mathbf{y}} &= \mathbf{y}_0 + \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) && (12)\\
\widetilde{\mathbf{\Psi}} &= \mathbf{\Psi}_0 + \widehat{\mathsf{E}}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})') && (13)\\
\widetilde{\mathbf{\Phi}} &= \mathbf{\Phi}_0 + \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) + \mathbf{H}\widehat{\mathsf{E}}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})') + \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})(\mathbf{x} - \mathbf{x}_0)')\mathbf{H}' && (14)
\end{aligned}$$

where $\widehat{\mathsf{E}}(\cdot)$ and $\widehat{\mathsf{cov}}(\cdot)$ denote approximations of the mean and covariance matrix, respectively.

We then have the following:

Theorem 1. *The KLD of the density $q(\cdot)$, given in (3), from $p(\cdot)$, given in (2), is*

$$I(p,q) = \left[\log\left(|\mathbf{R}^{-1}\mathbf{\Lambda}|\right) + \mathrm{tr}((\mathbf{G}+\mathbf{e}\mathbf{e}')\mathbf{\Lambda}^{-1})\right] / 2 \tag{15}$$

where

$$\mathbf{\Lambda} = \mathbf{R} + \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')\mathbf{P}_0^{-1}\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{16}$$

$$\mathbf{G} = \mathsf{cov}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) - 2[\mathsf{E}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)') - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')]\mathbf{P}_0^{-1}\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{17}$$

$$\mathbf{e} = \mathsf{E}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) \tag{18}$$

Proof: See the appendix. ∎

Evaluation of the KLD for a given KF variant then requires substituting the approximations for the moments of the remainder term $\mathbf{n}(\cdot)$ into (16)-(18) and then substituting the resulting expressions into (15). In the following subsections Theorem 1 is applied to the KF, EKF and UKF, respectively.

A. Kalman filter

The KF uses the exact prior moments so that $\widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) = \mathsf{E}(\mathbf{n}(\mathbf{x}))$, $\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})) = \mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x}))$ and $\widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) = \mathsf{cov}(\mathbf{n}(\mathbf{x}))$. We then have that $\mathbf{e} = \mathbf{0}$ and $\mathbf{G} = \mathbf{0}$ so the KLD, given in (15), reduces to

$$I(p,q) = \log(|\mathbf{R}^{-1}\mathbf{\Lambda}|)/2 \tag{19}$$

where

$$\mathbf{R}^{-1}\mathbf{\Lambda} = \mathbf{I} + \mathbf{R}^{-1}[\mathsf{cov}(\mathbf{n}(\mathbf{x})) - \mathsf{E}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')\mathbf{P}_0^{-1}\mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})')] \tag{20}$$

A readier interpretation of (19)-(20), as well as an expression more suited to computation, can be obtained by replacing the moments of the remainder term $\mathbf{n}(\cdot)$ with the prior moments (5)-(7) of the KF recursion. We have from (10) and (11) that

$$\mathsf{cov}(\mathbf{n}(\mathbf{x})) = \mathbf{\Phi} + \mathbf{\Phi}_0 - \mathbf{H}\mathbf{\Psi} - \mathbf{\Psi}'\mathbf{H}' \tag{21}$$

$$\mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') = \mathbf{\Psi} - \mathbf{\Psi}_0 \tag{22}$$

Substituting (21) and (22) into (20) enables the KLD of the KF to be written as

$$I(p,q) = \log(|\mathbf{I} + \mathbf{R}^{-1}(\mathbf{\Phi} - \mathbf{\Psi}'\mathbf{P}_0^{-1}\mathbf{\Psi})|) \tag{23}$$

Eq. (23) reveals the direct dependence of the KLD of the KF joint density approximation on the prior moments. If these prior moments can be calculated then the KLD can be obtained exactly, otherwise approximation is necessary. The quantity $\mathbf{\Phi} - \mathbf{\Psi}'\mathbf{P}_0^{-1}\mathbf{\Psi}$ is the covariance matrix of the residual resulting from statistical linearisation of the measurement function with respect to the prior [2]. It can therefore be interpreted as the effective nonlinearity of the measurement function, taking into account the prior covariance matrix. In (23) this quantity, scaled by the inverse of the measurement noise covariance matrix, determines the effect of measurement function nonlinearity on the accuracy of the KF approximation.

B. Extended Kalman filter

The EKF makes no attempt to approximate the moments of the remainder term $\mathbf{n}(\cdot)$. Thus, $\widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) = \mathbf{0}$, $\widehat{\mathsf{E}}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})) = \mathbf{0}$ and $\widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) = \mathbf{0}$. The terms required for the KLD are then found, from (16)-(18), as $\mathbf{\Lambda} = \mathbf{R}$, $\mathbf{G} = \mathsf{cov}(\mathbf{n}(\mathbf{x}))$ and $\mathbf{e} = \mathsf{E}(\mathbf{n}(\mathbf{x}))$. The KLD is then found as

$$\begin{aligned} I(p,q) &= \mathrm{tr}([\mathsf{cov}(\mathbf{n}(\mathbf{x})) + \mathsf{E}(\mathbf{n}(\mathbf{x}))\mathsf{E}(\mathbf{n}(\mathbf{x}))']\mathbf{R}^{-1})/2 \\ &= \mathrm{tr}(\mathsf{E}(\mathbf{n}(\mathbf{x})\mathbf{n}(\mathbf{x})')\mathbf{R}^{-1})/2 \end{aligned} \tag{24}$$

As with the KLD of the KF, the moments of the remainder term $\mathbf{n}(\cdot)$ can be replaced by the prior moments (5)-(7). Using $\mathsf{E}(\mathbf{n}(\mathbf{x})) = \hat{\mathbf{y}} - \mathbf{y}_0$, from (9), and (21) gives

$$I(p,q) = \mathrm{tr}(\mathbf{R}^{-1}[\mathbf{\Phi} + \mathbf{\Phi}_0 - \mathbf{H}\mathbf{\Psi} - \mathbf{\Psi}'\mathbf{H}' + (\hat{\mathbf{y}} - \mathbf{y}_0)(\hat{\mathbf{y}} - \mathbf{y}_0)'])/2. \tag{25}$$

Thus, as with the KF, the KLD for the EKF joint density approximation can be found in closed-form provided the KF moments can be computed. The effect of nonlinearity on the accuracy of the EKF approximation is measured by the term in square brackets in (25). This term differs from the corresponding term in the KF but it has the same basic property of depending on the prior covariance matrix as well as the measurement function and is also scaled by the inverse of the measurement noise covariance matrix.

C. Unscented Kalman filter

The UKF uses numerical integration to approximate the prior moments (5)-(7). In a procedure akin to Gaussian quadrature, the integrals (5)-(7) are approximated based on evaluating the integrands at a number of specially selected values of the state, referred to as sigma points. A variety of schemes have been proposed which use different numbers and constellations of sigma points [15]. We consider a generic scheme using s sigma points $\mathcal{X}_1, \ldots, \mathcal{X}_s$ with weights $w_1, \ldots, w_s$. The moments (5)-(7) are approximated by

$$\widetilde{\mathbf{y}} = \sum_{i=1}^{s} w_i \mathcal{Y}_i \tag{26}$$

$$\widetilde{\boldsymbol{\Psi}} = \sum_{i=1}^{s} w_i (\mathcal{X}_i - \mathbf{x}_0)\mathcal{Y}_i' \tag{27}$$

$$\widetilde{\boldsymbol{\Phi}} = \sum_{i=1}^{s} w_i (\mathcal{Y}_i - \widetilde{\mathbf{y}})(\mathcal{Y}_i - \widetilde{\mathbf{y}})' \tag{28}$$

where $\mathcal{Y}_i = \mathbf{h}(\mathcal{X}_i)$, $i = 1, \ldots, s$. In terms of (12)-(14), this amounts to using the moment approximations:

$$\widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) = \sum_{i=1}^{s} w_i \mathbf{n}(\mathcal{X}_i) \tag{29}$$

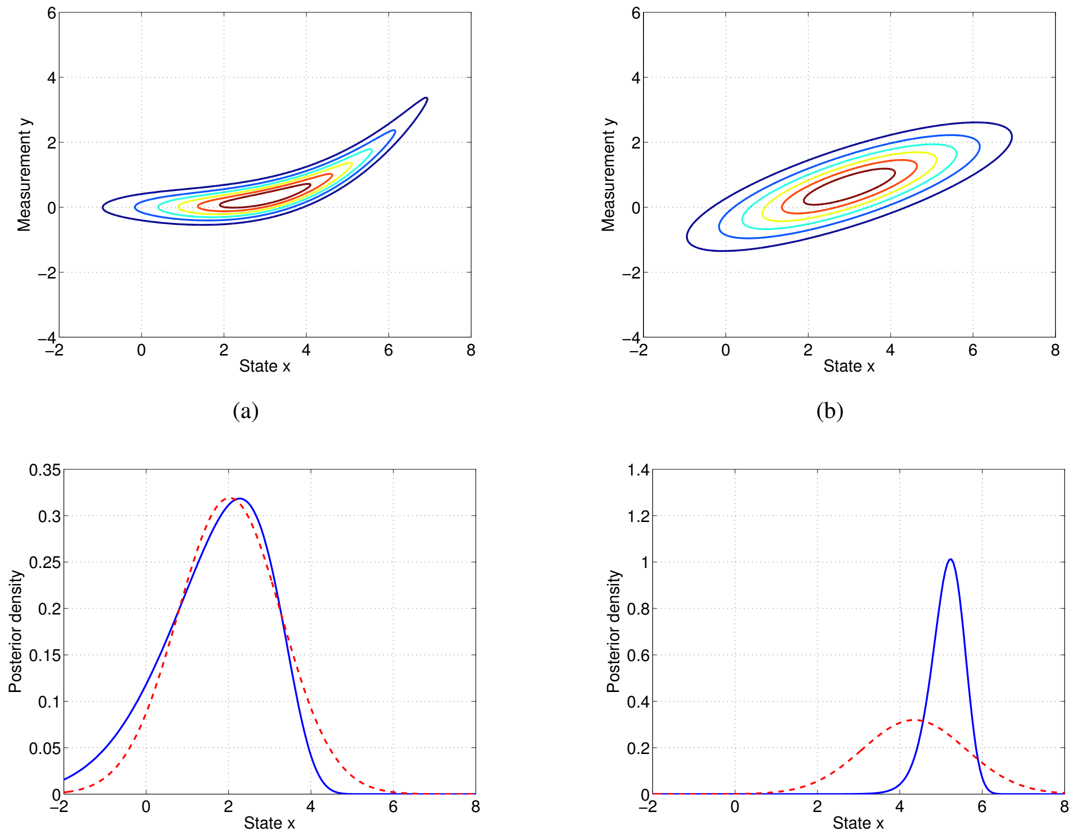
$$\widehat{\mathsf{E}}((\mathbf{x} - \mathbf{x}_0)\mathbf{n}(\mathbf{x})') = \sum_{i=1}^{s} w_i (\mathcal{X}_i - \mathbf{x}_0)\mathbf{n}(\mathcal{X}_i)' \tag{30}$$

$$\widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) = \sum_{i=1}^{s} w_i [\mathbf{n}(\mathcal{X}_i) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x}))][\mathbf{n}(\mathcal{X}_i) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x}))]' \tag{31}$$

The KLD for the UKF can be found by substituting (29)-(31) into (16)-(18) and then substituting the resulting expressions into (15). The moments of the remainder term $\mathbf{n}(\cdot)$ appearing in (16)-(18) can be computed exactly from the prior moments (5)-(7) when available but otherwise must be approximated. Replacing the remainder term moments with prior moments, as was done in the KLD expressions for the KF and EKF, does not aid particularly in the interpretation or computation of the KLD for the UKF. The CKF [1] and GHF [12] also use moment approximations of the form (26)-(28) so the results obtained here for the UKF also apply to these filters.

D. Discussion

In estimation we are interested in the posterior density, i.e., the conditional density of the state $\mathbf{x}$ given the measurement $\mathbf{y}$. In the analysis we have considered the joint density of $\mathbf{x}$ and $\mathbf{y}$ for reasons of tractability. While a small KLD for the joint density approximation implies a good posterior approximation for all measurements, a large KLD only means that the posterior density approximation is poor for some, but not necessarily all, measurement values. This can be seen by considering a simple example. Let $y = h(x) + \eta = x^3/100 + \eta$ where $\eta \sim \mathsf{N}(0, 0.1)$. The prior density for the state is $\mathsf{N}(3, 4)$. The KLD for this example can be found from (23) as 0.69. Figs. 1(a) and 1(b) show the true joint density and its KF approximation, respectively. In this example the KF moments (5)-(7) can be found in closed-form [8]. As expected from (19), given the small value of the measurement noise variance compared to the prior variance, the KF approximation to the joint density is quite poor. As a result, it is clear that there are many measurements values for which the KF posterior density approximation is poor although there are some for which the KF approximates the posterior reasonably well. This is illustrated in Figs. 1(c) and 1(d) which show that the KF provides a good posterior density approximation for $y = 0$ but not for $y = 1.5$.



(a) (b) (c) (d)

Fig. 1: An example of posterior density approximation using the KF: (a) joint density of the state and measurement and (b) KF approximation to the joint density. Posterior density (blue) and its KF approximation (red dashed) for (c) $y = 0$ and (d) $y = 1.5$.

A common element of the KF variants considered here is that, for a fixed prior density $\pi_0(\cdot)$ and measurement function $\mathbf{h}(\cdot)$, the KLD of the joint density increases as $\mathbf{R}$ decreases, i.e., as the measurements become more precise. This means that, for sufficiently precise measurements, none of the KF variants, including the KF itself, are based on an accurate approximation of the joint density. Conversely, for imprecise measurements all KF variants accurately approximate the joint density, and therefore the posterior density. Differences between the KF variants become evident for moderately precise measurements. The KLD for the KF, which uses the exact prior moments, increases logarithmically with

$\mathbf{R}^{-1}$ while the increase for the EKF is linear. Thus the KF should provide better performance than the EKF for moderately precise measurements or, equivalently, provide good performance down to smaller values of $\mathbf{R}$. The KLD for the UKF is composed of both logarithmic and linear terms with the size of these terms depending on the accuracy of its prior moment approximations. These properties of the KF variants will be demonstrated in a numerical example in the following section. It should be noted that in this discussion the notion of precision depends not just on the measurement noise covariance matrix $\mathbf{R}$ but also on the prior density and measurement function, e.g., even a measurement with large $\mathbf{R}$ could be considered precise if the prior density is vague.

III. NUMERICAL EXAMPLE

In this section the KLD analysis is applied to a numerical example in which measurements of an object's bearing are obtained by several spatially distributed sensors. A similar example appeared in [11]. Note that this is a static estimation problem, rather than the filtering problem to which KFs are usually applied, because our focus is on the correction step. In a filtering context, we are considering measurements collected at a particular time instant.

The object position $\mathbf{x} = [x, y]' \in \mathbb{R}^2$ has prior density $\mathsf{N}(\mathbf{x}_0, \mathbf{P}_0)$ where $\mathbf{x}_0 = [4, 4]'$ and $\mathbf{P}_0 = \mathbf{I}_2$. Bearings measurements are obtained by sensors with positions $\mathbf{p}_i = [\xi_i, \zeta_i]'$, $i = 1, 2$. The measurements are collected into the vector $\mathbf{y} = [\beta_1, \beta_2]'$ with the ith measurement given by

$$\beta_i = \text{atan2}(y - \zeta_i, x - \xi_i) + \eta_i \tag{32}$$

where $\text{atan2}(\cdot)$ is the four-quadrant arctangent and η_i are independent, zero-mean Gaussian random variables with variance τ^2. The sensor positions are $\mathbf{p}_1 = [0, 0]'$, $\mathbf{p}_2 = [5, 0]'$. Several values of the measurement noise variance τ^2 are considered.

The KF, EKF, UKF and CKF are applied to the problem of estimating the object position $\mathbf{x}$ given the measurements $\mathbf{y}$. The UKF is implemented as suggested in [14] with $s = 5$ sigma points. The weights $w_1, \ldots, w_s$ and sigma points $\mathcal{X}_1, \ldots, \mathcal{X}_s$ are given by

$$w_i = \begin{cases} 1/3, & i = 1, \\ 1/6, & i = 2, \ldots, 5, \end{cases} \tag{33}$$

$$\mathcal{X}_i = \begin{cases} \mathbf{x}_0, & i = 1, \\ \mathbf{x}_0 + (-1)^i \sqrt{3}\, \mathbf{a}_{\lfloor i/2 \rfloor}, & i = 2, \ldots, 5. \end{cases} \tag{34}$$

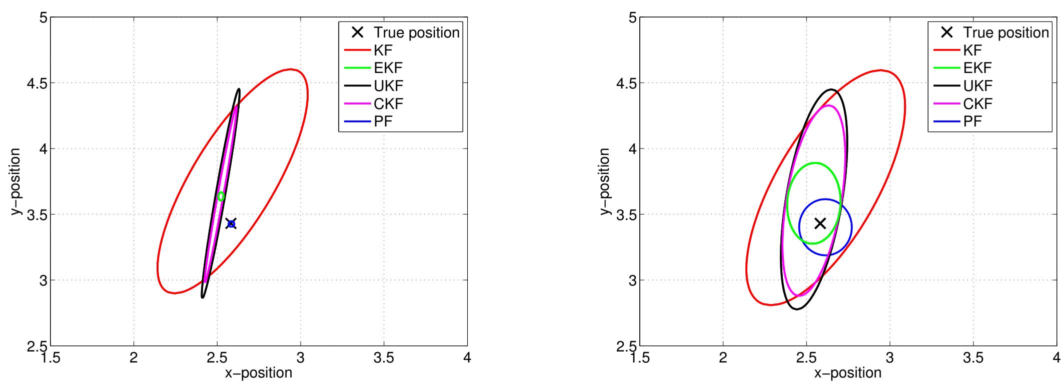
where $\lfloor \cdot \rfloor$ is the floor operation and $\mathbf{a}_i$ is the ith column of the matrix $\mathbf{A}$ satisfying $\mathbf{A}\mathbf{A}' = \mathbf{P}_0$. A particle filter (PF) is also implemented to provide a benchmark. The particular PF implemented here is the bootstrap filter in which samples are drawn from the prior and weighted by the likelihood. A sample size of 500 000 is used. With this sample size the PF approximation may be regarded as almost exact. The moments required in the KF are approximated using importance sampling with samples drawn from the prior. A sample size of 10^6 is used so that the moment approximations may be considered to be almost exact. Filter performance is measured by the root mean square error (RMSE) of the position estimates and a consistency measure defined as follows. As in [4], consistency is used here to mean the ability of a filter to provide a reliable indication of its accuracy. Let $\mathsf{N}(\hat{\mathbf{x}}, \hat{\mathbf{P}})$ denote a posterior density approximation obtained from one of the filters. Then we consider the probability that $(\hat{\mathbf{x}} - \mathbf{x})'\hat{\mathbf{P}}^{-1}(\hat{\mathbf{x}} - \mathbf{x}) \leq \Gamma$ where Γ is the 95th percentile of the chi-squared distribution with two degrees-of-freedom. Assuming that the posterior is roughly Gaussian, a consistency measure close to 95% indicates that the filter is providing a realistic measure of its uncertainty, a lower value indicates that the filter is optimistic and a higher value indicates that it is pessimistic.

First, we consider the KLDs for the KF and its approximations. These are given in Table I for measurement noise standard deviations of $\tau = 0.1$, 1 and 10 degrees. The moments required for the KLDs, all of which can be found from the prior moments (5)-(7), are approximated using importance sampling with samples drawn from the prior. The sample size is 10^6. The KF has by far the lowest KLD for $\tau = 0.1$ and 1 degrees while all four algorithms have approximately the same KLD for $\tau = 10$ degrees. The limitations of striving to approximate the KF should be evident since even the KF does not provide accurate posterior density approximation, at least not for all measurement values, for the smaller values of τ.

TABLE I: KLDs for the KF, EKF, UKF and CKF for position estimation using two bearings measurements with various measurement noise standard deviations.

	Measurement noise standard deviation (°)		
	0.1	1	10
KF	6.41	2.01	0.08
EKF	877	8.76	0.09
UKF	113	4.25	0.09
CKF	299	4.37	0.09

To give an idea of how the KLDs of Table I translate into posterior density approximations the results for a particular measurement realisation are given in Fig. 2 for $\tau = 0.1$ and 1 degrees. The plots show the 95% ellipses for the KF, EKF, UKF, CKF and a Gaussian approximation obtained from the PF by moment matching. As noted earlier, the KLD of $q(\cdot)$ from $p(\cdot)$ can be quite small provided the support of $q(\cdot)$ covers that of $p(\cdot)$. The KF's posterior approximation clearly has this property although its region of support not only covers that of the PF approximation but is much larger, particularly for $\tau = 0.1$ degrees. This explains why the KLD of the KF is quite low although its posterior density approximation is quite different to the true posterior. The 95% ellipses of the EKF, UKF and CKF do not overlap at all with that of the PF for $\tau = 0.1$ degrees, as would be expected given their extremely large KLDs. For $\tau = 0.1$ degrees, only the KF and PF provide consistent estimates since their 95% ellipsoids include the true value. The moderate KLDs obtained by the EKF, UKF and CKF for $\tau = 1$ degree are reflected in posterior density approximations which overlap significantly with the PF's approximation. All filters provide consistent estimates for $\tau = 1$ degree.



(a) (b)

Fig. 2: Posterior density approximations for position estimation using two bearings measurements with measurement noise standard deviations of (a) 0.1 degrees and (b) 1 degree.

The RMSE and consistency of the KF, EKF, UKF, CKF and PF are shown in Table II for measurement noise standard deviations of $\tau = 0.1$, 1 and 10 degrees. Relative to the PF, all four KF variants deteriorate as τ decreases. In fact, the RMSEs of the KF variants are only slightly smaller for $\tau = 0.1$ degrees than for $\tau = 1$ degree while the RMSE of the PF decreases by almost an order of magnitude. This behaviour is in accordance with the KLDs shown in Table I, all of which increase as τ decreases. Interestingly, the much lower KLD of the KF compared to the UKF, CKF and EKF does not result in a significantly smaller RMSE. However, the consistency measure is much closer to the nominal value of 95%. Similar comments apply to the relative performances of the UKF, CKF and EKF: a lower KLD seems to produce better consistency rather than a lower RMSE.

The value of the improved consistency of the KF for estimation becomes apparent if we receive a further bearings measurement from a sensor located at $\mathbf{p}_3 = [0, 5]'$. The measurements are processed sequentially to model a situation in which filtering estimation of the position of a stationary object is to be performed using measurements taken first at positions $\mathbf{p}_1$ and $\mathbf{p}_2$ and then later at position $\mathbf{p}_3$. The results are given in Table III. The estimation accuracy of the KF is now clearly better than that of the EKF, UKF and CKF, reflecting its superior consistency after processing the first two measurements. The RMSEs of the UKF and CKF are similar and significantly lower than that of the EKF.

TABLE II: Performance of the KF, EKF, UKF, CKF and PF for position estimation using two bearings measurements with various measurement noise standard deviations. The set value for consistency is 95%.

(a) RMSE

	Measurement noise standard deviation (°)		
	0.1	1	10
KF	0.393	0.423	0.986
EKF	0.423	0.461	0.986
UKF	0.409	0.444	0.987
CKF	0.405	0.442	0.987
PF	0.017	0.161	0.966

(b) Consistency

	Measurement noise standard deviation (°)		
	0.1	1	10
KF	74.4	87.3	94.4
EKF	10.2	59.5	94.0
UKF	33.1	75.1	93.9
CKF	22.4	74.0	94.0
PF	92.5	94.7	94.3

In summary, this numerical example suggests that the KLD of the joint density approximation on which the KF is based is a reliable indicator of performance. It is also evident that seeking increasingly accurate approximations of the KF can be of limited benefit, particularly when measurements are accurate. The applicability of this conclusion to object tracking is demonstrated in [9] in several scenarios. There it is shown that, for low measurement noise variances, alternatives such as the truncated KF should be considered.

TABLE III: Performance of the KF, EKF, UKF, CKF and PF for position estimation using three bearings measurements with various measurement noise standard deviations. The set value for consistency is 95%.

(a) RMSE

	Measurement noise standard deviation (°)		
	0.1	1	10
KF	0.138	0.179	0.791
EKF	0.230	0.270	0.792
UKF	0.197	0.212	0.792
CKF	0.206	0.208	0.792
PF	0.013	0.096	0.775

(b) Consistency

	Measurement noise standard deviation (°)		
	0.1	1	10
KF	75.0	87.3	94.2
EKF	13.0	68.3	94.2
UKF	32.6	76.4	94.5
CKF	22.6	76.3	94.4
PF	89.6	94.7	94.7

IV. Conclusions

We have performed an analysis of the KF and its various approximations for the case of a nonlinear measurement equation with additive Gaussian noise. The performance metric used in the analysis was the Kullback Leibler divergence of the KF approximation of the joint density of the state and measurements from the true joint density. The analysis provides a means of comparing the performances of the various KF approximations and determining the conditions under which KF approximations can be expected to perform well. The main points of interest are that the accuracy of any KF variant, including the KF itself, deteriorates as the measurement noise variance decreases and that the deterioration is much slower for the KF than for KF approximations such as the EKF and UKF, particularly the EKF. This was demonstrated in a numerical example involving position estimation using bearings measurements.

An interesting extension would be to consider the case of non-Gaussian measurement noise.

Appendix

We begin by re-writing the exponent of the true joint density, given in (2). Substituting the Taylor series expansion (8) of the measurement function $\mathbf{h}(\cdot)$ about the prior mean $\mathbf{x}_0$ gives

$$(\mathbf{y}-\mathbf{h}(\mathbf{x}))'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x})) = (\mathbf{y}-\mathbf{y}_0-\mathbf{H}(\mathbf{x}-\mathbf{x}_0)-\mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{y}_0-\mathbf{H}(\mathbf{x}-\mathbf{x}_0)-\mathbf{n}(\mathbf{x})) \quad (35)$$

Adding and subtracting $\widetilde{\mathbf{y}}$ in each term gives

$$(\mathbf{y}-\mathbf{h}(\mathbf{x}))'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x})) = (\mathbf{y}-\widetilde{\mathbf{y}})'\mathbf{R}^{-1}(\mathbf{y}-\widetilde{\mathbf{y}}) - (\mathbf{x}-\mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}(\mathbf{y}-\widetilde{\mathbf{y}}) - (\mathbf{y}-\widetilde{\mathbf{y}})'\mathbf{R}^{-1}\mathbf{H}(\mathbf{x}-\mathbf{x}_0)$$

$$+ (\mathbf{x}-\mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}(\mathbf{x}-\mathbf{x}_0) + u(\mathbf{x},\mathbf{y}) \tag{36}$$

where

$$\begin{aligned} u(\mathbf{x},\mathbf{y}) &= -2(\mathbf{y}-\widetilde{\mathbf{y}})'\mathbf{R}^{-1}\mathbf{v} - 2(\mathbf{y}-\widetilde{\mathbf{y}})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}) + \mathbf{v}'\mathbf{R}^{-1}\mathbf{v} + 2(\mathbf{x}-\mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}\mathbf{v} + 2\mathbf{v}'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}) \\ &\quad + 2(\mathbf{x}-\mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}) + \mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}) \end{aligned} \tag{37}$$

with $\mathbf{v} = \mathbf{y}_0 - \widetilde{\mathbf{y}}$. After substituting (36) into (2) the KLD of $q(\cdot)$ from $p(\cdot)$ can be written as

$$I(p,q) = \mathsf{E}_{x,y}\left(\log(p(\mathbf{x},\mathbf{y})/q(\mathbf{x},\mathbf{y}))\right) \tag{38}$$

$$= \left[\log\left(\frac{|\boldsymbol{\Omega}|}{|\mathbf{P}_0|\,|\mathbf{R}|}\right) + \mathsf{E}_{x,y}(d(\mathbf{x},\mathbf{y}))\right] \Big/ 2 \tag{39}$$

where the expectation is with respect to the density $p(\cdot)$ and

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &= (\mathbf{x}-\mathbf{x}_0)'(\mathbf{A}-\mathbf{P}_0^{-1}-\mathbf{H}'\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}-\mathbf{x}_0) + 2(\mathbf{x}-\mathbf{x}_0)'(\mathbf{C}+\mathbf{H}'\mathbf{R}^{-1})(\mathbf{y}-\widetilde{\mathbf{y}}) \\ &\quad + (\mathbf{y}-\widetilde{\mathbf{y}})'(\mathbf{B}-\mathbf{R}^{-1})(\mathbf{y}-\widetilde{\mathbf{y}}) - u(\mathbf{x},\mathbf{y}) \end{aligned} \tag{40}$$

It follows from (4) that

$$\frac{|\boldsymbol{\Omega}|}{|\mathbf{P}_0|\,|\mathbf{R}|} = |\mathbf{I} + \mathbf{R}^{-1}(\widetilde{\boldsymbol{\Phi}} - \widetilde{\boldsymbol{\Psi}}'\mathbf{P}_0^{-1}\widetilde{\boldsymbol{\Psi}})| \tag{41}$$

The expectation of $d(\cdot)$ with respect to the joint density $p(\cdot)$ is

$$\mathsf{E}_{x,y}(d(\mathbf{x},\mathbf{y})) = t_1 + 2t_2 + t_3 - \mathsf{E}_{x,y}(u(\mathbf{x},\mathbf{y})) \tag{42}$$

where

$$\begin{aligned} t_1 &= \mathsf{E}_{x,y}((\mathbf{x}-\mathbf{x}_0)'(\mathbf{A}-\mathbf{P}_0^{-1}-\mathbf{H}'\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}-\mathbf{x}_0)) \\ &= \mathrm{tr}((\mathbf{A}-\mathbf{P}_0^{-1}-\mathbf{H}'\mathbf{R}^{-1}\mathbf{H})\mathsf{E}((\mathbf{x}-\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)')) \\ &= \mathrm{tr}((\mathbf{A}-\mathbf{P}_0^{-1}-\mathbf{H}'\mathbf{R}^{-1}\mathbf{H})\mathbf{P}_0) \end{aligned} \tag{43}$$

$$\begin{aligned} t_2 &= \mathsf{E}_{x,y}((\mathbf{x}-\mathbf{x}_0)'(\mathbf{C}+\mathbf{H}'\mathbf{R}^{-1})(\mathbf{y}-\widetilde{\mathbf{y}})) \\ &= \mathrm{tr}((\mathbf{C}+\mathbf{H}'\mathbf{R}^{-1})\mathsf{E}_{x,y}((\mathbf{x}-\mathbf{x}_0)(\mathbf{y}-\widetilde{\mathbf{y}})')) \\ &= \mathrm{tr}((\mathbf{C}+\mathbf{H}'\mathbf{R}^{-1})\boldsymbol{\Psi}') \end{aligned} \tag{44}$$

$$\begin{aligned} t_3 &= \mathsf{E}_{x,y}((\mathbf{y}-\widetilde{\mathbf{y}})'(\mathbf{B}-\mathbf{R}^{-1})(\mathbf{y}-\widetilde{\mathbf{y}})) \\ &= \mathrm{tr}((\mathbf{B}-\mathbf{R}^{-1})\mathsf{E}_{x,y}((\mathbf{y}-\widetilde{\mathbf{y}})(\mathbf{y}-\widetilde{\mathbf{y}})')) \\ &= \mathrm{tr}((\mathbf{B}-\mathbf{R}^{-1})\mathsf{E}_{x,y}((\mathbf{y}-\hat{\mathbf{y}})(\mathbf{y}-\hat{\mathbf{y}})')) + \mathrm{tr}((\mathbf{B}-\mathbf{R}^{-1})(\hat{\mathbf{y}}-\widetilde{\mathbf{y}})(\hat{\mathbf{y}}-\widetilde{\mathbf{y}})') \\ &= \mathrm{tr}((\mathbf{B}-\mathbf{R}^{-1})\mathbf{S}) + \mathbf{e}'(\mathbf{B}-\mathbf{R}^{-1})\mathbf{e} \end{aligned} \tag{45}$$

with $\mathbf{S} = \mathbf{R} + \mathbf{\Phi}$ and $\mathbf{e} = \hat{\mathbf{y}} - \widetilde{\mathbf{y}}$. It follows from (4) and the Woodbury formula [10, (6.26)] that

$$\mathbf{A} = \mathbf{P}_0^{-1} + \mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Psi}}'\mathbf{P}_0^{-1} \tag{46}$$

$$\mathbf{C} = -\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}\mathbf{\Lambda}^{-1} \tag{47}$$

$$\mathbf{B} = \mathbf{\Lambda}^{-1} \tag{48}$$

where $\widetilde{\mathbf{S}} = \mathbf{R} + \widetilde{\mathbf{\Phi}}$ and $\mathbf{\Lambda} = \widetilde{\mathbf{S}} - \widetilde{\mathbf{\Psi}}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}$. Each term in (42) is considered in turn. The first term can be written as

$$\begin{aligned} t_1 &= \mathrm{tr}(\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}\mathbf{\Lambda}^{-1}\widetilde{\mathbf{\Psi}}') - \mathrm{tr}(\mathbf{H}'\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_0) && (49)\\ &= \mathrm{tr}(\widetilde{\mathbf{\Psi}}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}\mathbf{\Lambda}^{-1}) - \mathrm{tr}(\mathbf{\Phi}_0\mathbf{R}^{-1}) && (50)\\ &= \mathrm{tr}(\widetilde{\mathbf{S}}\mathbf{\Lambda}^{-1}) - \mathrm{tr}(\mathbf{S}_0\mathbf{R}^{-1}) && (51) \end{aligned}$$

where $\mathbf{\Phi}_0 = \mathbf{H}\mathbf{P}_0\mathbf{H}'$ and $\mathbf{S}_0 = \mathbf{R} + \mathbf{\Phi}_0$. The second and third terms can be written as

$$t_2 = -\mathrm{tr}(\mathbf{\Psi}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}\mathbf{\Lambda}^{-1}) + \mathrm{tr}(\mathbf{\Psi}'\mathbf{P}_0^{-1}\mathbf{\Psi}_0\mathbf{R}^{-1}) \tag{52}$$

$$t_3 = \mathrm{tr}(\mathbf{S}\mathbf{\Lambda}^{-1}) - \mathrm{tr}(\mathbf{S}\mathbf{R}^{-1}) + \mathbf{e}'(\mathbf{\Lambda}^{-1} - \mathbf{R}^{-1})\mathbf{e} \tag{53}$$

where $\mathbf{\Psi}_0 = \mathbf{P}_0\mathbf{H}'$. Now consider the expectation of $u(\cdot)$. We have that

$$\begin{aligned} \mathsf{E}_{x,y}((\mathbf{y} - \widetilde{\mathbf{y}})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) &= \mathsf{E}((\mathbf{h}(\mathbf{x}) - \widetilde{\mathbf{y}})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}))\\ &= \mathsf{E}((\mathbf{v} + \mathbf{H}(\mathbf{x} - \mathbf{x}_0) + \mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}))\\ &= \mathbf{v}'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) + \mathsf{E}((\mathbf{x} - \mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) + \mathsf{E}(\mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}))\\ &= \mathbf{e}'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) - \mathsf{E}(\mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) + \mathsf{E}((\mathbf{x} - \mathbf{x}_0)'\mathbf{H}'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x}))\\ &\quad + \mathsf{E}(\mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) && (54) \end{aligned}$$

where we have used $\mathbf{e} - \mathbf{v} = \hat{\mathbf{y}} - \mathbf{y}_0 = \mathsf{E}(\mathbf{n}(\mathbf{x}))$. Taking the expected value of (37) and using (54) gives

$$\mathsf{E}_{x,y}(u(\mathbf{x}, \mathbf{y})) = \mathsf{E}(\mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) - \mathsf{E}(\mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) - \mathbf{e}'\mathbf{R}^{-1}\mathbf{e} \tag{55}$$

Substituting (51), (52), (53) and (55) into (42) gives

$$\begin{aligned} \mathsf{E}_{x,y}(d(\mathbf{x}, \mathbf{y})) &= \mathrm{tr}((\widetilde{\mathbf{S}} + \mathbf{S} - 2\mathbf{\Psi}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}})\mathbf{\Lambda}^{-1}) - \mathrm{tr}((\mathbf{S} + \mathbf{S}_0 - 2\mathbf{\Psi}'\mathbf{P}_0^{-1}\mathbf{\Psi}_0)\mathbf{R}^{-1})\\ &\quad + \mathbf{e}'\mathbf{\Lambda}^{-1}\mathbf{e} - \mathsf{E}(\mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) + \mathsf{E}(\mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) && (56)\\ &= \mathrm{tr}(\mathbf{W}\mathbf{\Lambda}^{-1}) - \mathrm{tr}(\mathbf{\Delta}_2\mathbf{R}^{-1}) + 2\mathrm{tr}(\mathbf{\Delta}_1'\mathbf{H}'\mathbf{R}^{-1})\\ &\quad + \mathbf{e}'\mathbf{\Lambda}^{-1}\mathbf{e} - \mathsf{E}(\mathbf{n}(\mathbf{x}))'\mathbf{R}^{-1}\mathsf{E}(\mathbf{n}(\mathbf{x})) + \mathsf{E}(\mathbf{n}(\mathbf{x})'\mathbf{R}^{-1}\mathbf{n}(\mathbf{x})) && (57) \end{aligned}$$

where $\mathbf{\Delta}_1 = \mathbf{\Psi} - \mathbf{\Psi}_0$, $\mathbf{\Delta}_2 = \mathbf{\Phi} - \mathbf{\Phi}_0$ and

$$\mathbf{W} = \mathbf{E}_2 - 2\mathbf{E}_1'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}} \tag{58}$$

with $\mathbf{E}_1 = \mathbf{\Psi} - \widetilde{\mathbf{\Psi}}$ and $\mathbf{E}_2 = \mathbf{\Phi} - \widetilde{\mathbf{\Phi}}$. It follows from (10) and (11) that

$$\mathbf{\Delta}_1 = \mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{59}$$

$$\begin{aligned}\mathbf{\Delta}_2 &= \mathbf{H}\mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') + \mathsf{E}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')\mathbf{H}' + \mathsf{cov}(\mathbf{n}(\mathbf{x})) \\ &= \mathbf{H}\mathbf{\Delta}_1 + \mathbf{\Delta}_1'\mathbf{H}' + \mathsf{cov}(\mathbf{n}(\mathbf{x}))\end{aligned} \tag{60}$$

Using (59) and (60) gives

$$\mathrm{tr}(\mathbf{\Delta}_2\mathbf{R}^{-1}) - 2\mathrm{tr}(\mathbf{\Delta}_1'\mathbf{H}'\mathbf{R}^{-1}) = \mathrm{tr}(\mathsf{cov}(\mathbf{n}(\mathbf{x}))\mathbf{R}^{-1}) \tag{61}$$

Substituting (61) and $\mathsf{cov}(\mathbf{n}(\mathbf{x})) = \mathsf{E}(\mathbf{n}(\mathbf{x})\mathbf{n}(\mathbf{x})') - \mathsf{E}(\mathbf{n}(\mathbf{x}))\mathsf{E}(\mathbf{n}(\mathbf{x}))'$ into (57) and using $\mathrm{tr}(\mathbf{a}\mathbf{a}'\mathbf{Z}) = \mathbf{a}\mathbf{Z}\mathbf{a}'$ for a vector $\mathbf{a}$ and matrix $\mathbf{Z}$ gives

$$\mathsf{E}_{x,y}(d(\mathbf{x},\mathbf{y})) = \mathrm{tr}(\mathbf{W}\mathbf{\Lambda}^{-1}) + \mathbf{e}'\mathbf{\Lambda}^{-1}\mathbf{e} \tag{62}$$

To obtain (62) we have used $\mathsf{cov}(\mathbf{n}(\mathbf{x})) = \mathsf{E}(\mathbf{n}(\mathbf{x})\mathbf{n}(\mathbf{x})') - \mathsf{E}(\mathbf{n}(\mathbf{x}))\mathsf{E}(\mathbf{n}(\mathbf{x}))'$. The KLD can then be found from (39), (41) and (62) as

$$I(p,q) = \left[\log\left(|\mathbf{R}^{-1}\mathbf{\Lambda}|\right) + \mathrm{tr}(\mathbf{W}\mathbf{\Lambda}^{-1}) + \mathbf{e}'\mathbf{\Lambda}^{-1}\mathbf{e}\right]/\,2 \tag{63}$$

where we have used $\mathbf{\Lambda} = \mathbf{R} + \widetilde{\mathbf{\Phi}} - \widetilde{\mathbf{\Psi}}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}$.

It follows from (9)-(14) that the errors in the prior moment approximations can then be found as

$$\mathbf{e} = \mathsf{E}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})) \tag{64}$$

$$\mathbf{E}_1 = \mathsf{E}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') - \widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{65}$$

$$\mathbf{E}_2 = \mathbf{H}\mathbf{E}_1 + \mathbf{E}_1'\mathbf{H}' + \mathsf{cov}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) \tag{66}$$

Substituting $\widetilde{\mathbf{\Psi}} = \mathbf{\Psi}_0 + \widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})')$ and (66) into (58) gives

$$\mathbf{W} = \mathsf{cov}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) + \mathbf{H}\mathbf{E}_1 - \mathbf{E}_1'\mathbf{H}' - 2\mathbf{E}_1'\mathbf{P}_0^{-1}\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{67}$$

Since $\mathrm{tr}(\mathbf{Z}') = \mathrm{tr}(\mathbf{Z})$ for a generic $d \times$d matrix $\mathbf{Z}$, the terms $\mathbf{H}\mathbf{E}_1$ and $\mathbf{E}_1'\mathbf{H}'$ cancel in $\mathrm{tr}(\mathbf{W}\mathbf{\Lambda}^{-1})$ so that $\mathrm{tr}(\mathbf{W}\mathbf{\Lambda}^{-1}) = \mathrm{tr}(\mathbf{G}\mathbf{\Lambda}^{-1})$ for the matrix $\mathbf{G}$ given in (17).

Substituting (13) and (14) into $\mathbf{\Lambda} = \mathbf{R} + \widetilde{\mathbf{\Phi}} - \widetilde{\mathbf{\Psi}}'\mathbf{P}_0^{-1}\widetilde{\mathbf{\Psi}}$ gives

$$\mathbf{\Lambda} = \mathbf{R} + \mathbf{H}\mathbf{P}_0\mathbf{H}' + \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) + \mathbf{H}\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') + \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')\mathbf{H}'$$

$$- (\mathbf{P}_0^{-1}\mathbf{H}' + \widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})'))'\mathbf{P}_0^{-1}(\mathbf{P}_0^{-1}\mathbf{H}' + \widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})')) \tag{68}$$

$$= \mathbf{R} + \widehat{\mathsf{cov}}(\mathbf{n}(\mathbf{x})) - \widehat{\mathsf{E}}(\mathbf{n}(\mathbf{x})(\mathbf{x}-\mathbf{x}_0)')\mathbf{P}_0^{-1}\widehat{\mathsf{E}}((\mathbf{x}-\mathbf{x}_0)\mathbf{n}(\mathbf{x})') \tag{69}$$

as required.

REFERENCES

[1] I. Arasaratnam and S. Haykin. Cubature Kalman filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269, 2009.

[2] I. Arasaratnam, S. Haykin, and R.J. Elliott. Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature. *Proceedings of the IEEE*, 95(5):953–977, 2007.

[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*, 50(2):174–188, 2002.

[4] Y. Bar-Shalom and X-R. Li. *Estimation and Tracking: Principles, Techniques and Software*. Artech House, 1993.

[5] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[6] R.S. Bucy and K.D. Senne. Digital synthesis of non-linear filters. *Automatica*, 7:287–298, 1971.

[7] A. Farina, B. Ristic, and D. Benvenuti. Tracking a ballistic target: Comparison of several nonlinear filters. *IEEE Transactions on Aerospace and Electronic Systems*, 38(3):854–867, 2002.

[8] Á. García-Fernández. *Detection and Tracking of Multiple Targets Using Wireless Sensor Networks*. PhD thesis, Universidad Politécnica de Madrid, Madrid, Spain, 2011. [Online] http://oa.upm.es/9823/.

[9] Á. García-Fernández, M.R. Morelande, and J. Grajal. Truncated unscented Kalman filtering. *IEEE Transactions on Signal Processing*, 60(7):3372–3386, 2012.

[10] J. Gentle. *Matrix Algebra: Theory, Computations and Applications in Statistics*. Springer, 2007.

[11] F. Gustafsson and G. Hendeby. Some relations between extended and unscented Kalman filters. *IEEE Transactions on Signal Processing*, 60(2):545–555, 2012.

[12] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910–927, 2000.

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

[14] S. Julier, J. Uhlmann, and H.F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, 2000.

[15] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[16] B.F. La Scala, R.R. Bitmead, and M.R. James. Conditions for stability of the extended Kalman filter and their application to the frequency tracking problem. *Mathematics and Control of Signals and Systems*, 8:1–26, 1995.

[17] C. Musso, N. Oudjane, and F. Le Gland. Improving regularised particle filters. In A. Doucet, N. de Freitas, and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York, 2001.

[18] K. Reif, S. Günther, E. Yaz, and R. Unbehauen. Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transaction on Automatic Control*, 44(4):714–728, 1999.

[19] K. Xiong, H.Y. Zhang, and C.W. Chan. Performance evaluation of UKF-based nonlinear filtering. *Automatica*, 42:261–270, 2006.