Iterated statistical linear regression for Bayesian updates

Ángel F. García-Fernández⋆, Lennart Svensson†, Mark R. Morelande◦
⋆Dept. of Electrical and Computer Engineering, Curtin University, Australia
†Dept. of Signals and Systems, Chalmers University of Technology, Sweden
◦Dept. of Electrical and Electronic Engineering, The University of Melbourne, Australia
Emails: angel.garciafernandez@curtin.edu.au, lennart.svensson@chalmers.se, mrmore@unimelb.edu.au

***Abstract*—This paper deals with Gaussian approximations to the posterior probability density function (PDF) in Bayesian nonlinear filtering. In this setting, using sigma-point based approximations to the Kalman filter (KF) recursion is a prominent approach. In the update step, the sigma-point KF approximations are equivalent to performing the statistical linear regression (SLR) of the (nonlinear) measurement function with respect to the prior PDF. In this paper, we indicate that the SLR of the measurement function with respect to the posterior is expected to provide better results than the SLR with respect to the prior. The resulting filter is referred to as the posterior linearisation filter (PLF). In practice, the exact PLF update is intractable but can be efficiently approximated by carrying out iterated SLRs based on sigma-point approximations. On the whole, the resulting filter, the iterated PLF (IPLF), is expected to outperform all sigma-point KF approximations as demonstrated by numerical simulations.**

***Index Terms*—Bayes' rule, Kalman filter, nonlinear filtering, sigma-points, statistical linear regression**

I. Introduction

A lot of problems in science and engineering deal with the estimation of the state of a dynamic process from noisy observations [1]. In the Bayesian framework, the probability density function (PDF) of the current state given the available measurements is necessary in order to estimate the current state in an optimal manner. This PDF is referred to as the posterior PDF and can be calculated recursively in two phases: prediction and update.

If the measurement equation is non-linear/non-Gaussian, the posterior PDF cannot be calculated analytically so approximations must be used in practice. If the posterior is unimodal, Gaussian approximations usually provide sufficient accuracy. Therefore, it is of interest to develop computationally efficient Gaussian approximations. In this case, the prediction step typically consists of calculating/approximating the first two moments of a random variable that undergoes a possibly nonlinear transformation. In the update step, we use the prior PDF, the current measurement and Bayes' rule to obtain the posterior. For nonlinear measurements, the update step is more difficult and is the focus of this paper.

In the Gaussian filtering update step, one possibility is to set the maximum a posteriori (MAP) estimator as the updated mean and obtain the updated covariance matrix by linearising the measurement function around the MAP, as in the iterated extended Kalman filter (IEKF) [2]. The posterior approximation based on the MAP is asymptotically optimal as the measurement noise tends to zero [3]. Nevertheless, the most widely used Gaussian filtering update step algorithm consists of approximating the updated mean by the linear minimum mean square error (LMMSE) estimator and the updated covariance by its mean square error matrix. The LMMSE-based approximation is usually more accurate than the MAP one as the metric to assess the filter performance is usually the square error.

The LMMSE-based approximation is sometimes referred to as (nonlinear) Kalman filter (KF) update [4, Sec. II.A] [5] or Gaussian filter update [6]. This algorithm is equivalent to performing statistical linear regression (SLR) of the measurement function with respect to the prior PDF and calculating Bayes' rule with the resulting linearised measurement model [7]. As the KF is usually known in the literature as the solution to the linear/Gaussian filtering recursion and the term Gaussian filter can also refer to many other types of approximations [8], we find it useful to refer to this algorithm as prior linearisation filter (PrLF) in the context of this paper. The PrLF update requires the calculation of some moments: the mean and covariance matrix of the current measurement and cross-covariance between the current state and the current measurement [4]. In practice, these moments (PrLF moments) cannot be calculated in closed-form so we require approximations. In this paper, we refer to these approximations as approximations to the PrLF. For instance, the extended Kalman filter (EKF) approximates the PrLF moments using analytical linearisation while the unscented KF (UKF) [4], cubature KF (CKF) [9] or linear regression KF [10] use sigma-points. The main drawback of the PrLF and all its approximations is that they often perform poorly if the measurement noise is low enough [5]. Therefore, more accurate computationally efficient approximations must be sought.

In this paper, we argue that, rather than performing SLR of the measurement function with respect to the prior, SLR should be performed with respect to the posterior. The intuition behind this idea is that the approximation of the measurement function should be accurate in the region of interest, which is indicated by the posterior, not the prior. We later confirm and formalise our intuition by showing that the mean square error (MSE) of the linearisation is minimised if it is performed with respect to the posterior. The resulting filter is referred to as

the posterior linearisation filter (PLF). The PLF is intractable as we would need to know the posterior to approximate the posterior. Nevertheless, we propose an approximation of the PLF by performing iterated SLRs based on sigma-points. This filter is referred to as the iterated PLF (IPLF) and is expected to perform better than any PrLF-type algorithm, such as EKF, UKF or CKF. This is due to the fact that the IPLF approximates the PLF, which is expected to outperform the PrLF.

The rest of the paper is organised as follows. In Section II, we formulate the problem. The PLF is introduced in Section III. The mathematical justification behind the PLF is given in Section IV. We provide the IPLF is Section V. Numerical simulations for assessing the filter performance are given in Section VI. Finally, conclusions are drawn in Section VII.

II. Problem statement

In this section, we explain the approximations we consider in the update step of Bayesian filtering. As we focus on the update phase, the time index of the filtering recursion is removed for the sake of notational simplicity. The state $\mathbf{x} \in \mathbb{R}^{n_x}$ has a Gaussian prior PDF $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \overline{\mathbf{x}}, \mathbf{P})$ with mean $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{P}$. The measurement equation is

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\eta} \tag{1}$$

where $\mathbf{z} \in \mathbb{R}^{n_z}$ is the measurement, $\mathbf{h}(\cdot)$ is the measurement function and $\boldsymbol{\eta}$ is a zero-mean Gaussian measurement noise with covariance matrix $\mathbf{R}$.

The posterior PDF $p(\mathbf{x}|\mathbf{z})$ of the state after observing measurement $\mathbf{z}$ is obtained by Bayes' rule

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})\, p(\mathbf{x}) \tag{2}$$

where $\propto$ means "is proportional to" and the likelihood $p(\mathbf{z}|\mathbf{x})$ can be obtained using (1)

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mathbf{h}(\mathbf{x}), \mathbf{R}) \tag{3}$$

In practice, the posterior does not admit a closed-form expression so it must be approximated. In this paper, we seek a Gaussian posterior approximation

$$q(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}; \overline{\mathbf{x}}_u, \mathbf{P}_u) \tag{4}$$

that is obtained using the enabling approximation

$$\mathbf{h}(\mathbf{x}) \approx \tilde{\mathbf{h}}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b} + \mathbf{e} \tag{5}$$

where $\tilde{\mathbf{h}}(\mathbf{x})$ is the approximation of $\mathbf{h}(\mathbf{x})$, $\mathbf{A} \in \mathbb{R}^{n_z \times n_x}$, $\mathbf{b} \in \mathbb{R}^{n_z}$ and $\mathbf{e} \in \mathbb{R}^{n_z}$ is a zero-mean Gaussian distributed random variable with covariance matrix $\boldsymbol{\Omega}$. The variable $\mathbf{e}$ is uncorrelated with $\mathbf{x}$ and $\boldsymbol{\eta}$. It should be noted that affine measurement functions with additive Gaussian noise are the only functions for which the posterior is exactly Gaussian. The class of functions in (5) is therefore of particular interest since it allows us to represent all such measurement functions of importance.

Once the enabling approximation in (5) is employed, $p(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \mathbf{A}\mathbf{x} + \mathbf{b}, \boldsymbol{\Omega} + \mathbf{R})$ and the posterior moments become

$$\overline{\mathbf{x}}_u = \overline{\mathbf{x}} + \mathbf{P}\mathbf{A}^T \left(\mathbf{A}\mathbf{P}\mathbf{A}^T + \boldsymbol{\Omega} + \mathbf{R}\right)^{-1} (\mathbf{z} - \mathbf{A}\overline{\mathbf{x}} - \mathbf{b}) \tag{6}$$

$$\mathbf{P}_u = \mathbf{P} - \mathbf{P}\mathbf{A}^T \left(\mathbf{A}\mathbf{P}\mathbf{A}^T + \boldsymbol{\Omega} + \mathbf{R}\right)^{-1} \mathbf{A}\mathbf{P} \tag{7}$$

A. Relations with previous work

In the Gaussian filtering literature, there are three important kinds of linearisations that are used in (5): analytical linearisation at the prior mean, analytical linearisation at the MAP estimate and SLR w.r.t. the prior PDF. As finding the MAP estimate or performing SLR w.r.t. the prior is not always tractable, different approximations to these linearisations have been proposed. These give rise to different filters but their foundation is the use of one of these linearisations.

More specifically, if we select $\boldsymbol{\Omega} = \mathbf{0}$, and $\mathbf{A}$ and $\mathbf{b}$ by analytical linearisation at the prior mean, the resulting algorithm is the EKF. If we select $\boldsymbol{\Omega} = \mathbf{0}$, and $\mathbf{A}$ and $\mathbf{b}$ by analytical linearisation at the MAP estimate, obtained by a Gauss-Newton search, the resulting algorithm is the IEKF [2]. If we select $\mathbf{A}$, $\mathbf{b}$ and $\boldsymbol{\Omega}$ using SLR with respect to the prior, the resulting algorithm is the PrLF. If the PrLF moments are approximated using sigma-points drawn from the prior, the resulting algorithms are the widely used sigma-point KFs, e.g., UKF or CKF.

In Section III, we propose a fourth type of linearisation to be used in (5), for which a practical implementation is developed in Section V.

III. Posterior linearisation filter

As indicated in Section II-A, the PrLF selects the parameters of approximation (5) using SLR with respect to the prior. In this section, we first review the concept of SLR of a function with respect to a PDF, which indicates the region of the state space where the linearisation is accurate. We then motivate why the SLR of the measurement function should be done with respect to the posterior, instead of the prior, to obtain a suitable approximation $\tilde{\mathbf{h}}(\mathbf{x})$ in (5). The mathematical justification of this step is deferred until Section IV.

A. Review of SLR

In this section, we explain the statistical linear regression (SLR) of a function $\mathbf{h}(\cdot)$ with respect to a PDF $p(\cdot)$, whose first two moments are $\overline{\mathbf{x}}$ and $\mathbf{P}$. The SLR of $\mathbf{h}(\cdot)$ is an approximation of the form (5) in which the values of $\mathbf{A}$, $\mathbf{b}$ and $\boldsymbol{\Omega}$ are chosen such that the first two moments of the approximated joint variable $\left[\mathbf{x}^T, \tilde{\mathbf{h}}^T(\mathbf{x})\right]^T$ match those of the joint variable $\left[\mathbf{x}^T, \mathbf{h}^T(\mathbf{x})\right]^T$. We get that [7]

$$\mathbf{A}^+ = \boldsymbol{\Psi}^T \mathbf{P}^{-1} \tag{8}$$

$$\mathbf{b}^+ = \overline{\mathbf{z}} - \mathbf{A}^+ \overline{\mathbf{x}} \tag{9}$$

$$\boldsymbol{\Omega}^+ = \boldsymbol{\Phi} - \mathbf{A}^+ \mathbf{P} \left(\mathbf{A}^+\right)^T \tag{10}$$

where

$$\overline{\mathbf{z}} = \int \mathbf{h}(\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \tag{11}$$

$$\mathbf{\Psi} = \int (\mathbf{x} - \overline{\mathbf{x}}) (\mathbf{h}(\mathbf{x}) - \overline{\mathbf{z}})^T p(\mathbf{x})\, d\mathbf{x} \tag{12}$$

$$\mathbf{\Phi} = \int (\mathbf{h}(\mathbf{x}) - \overline{\mathbf{z}}) (\mathbf{h}(\mathbf{x}) - \overline{\mathbf{z}})^T p(\mathbf{x})\, d\mathbf{x} \tag{13}$$

Another interesting property is that the SLR of a function $\mathbf{h}(\cdot)$ provides the best affine approximation of $\mathbf{h}(\cdot)$ in the mean square error (MSE) sense and its MSE w.r.t. $p(\cdot)$ [7]. That is,

$$\begin{aligned}&\left(\mathbf{A}^+, \mathbf{b}^+\right)\\ &\quad = \underset{(\mathbf{A},\mathbf{b})}{\arg\min}\, \mathrm{E}\left[(\mathbf{h}(\mathbf{x}) - \mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{h}(\mathbf{x}) - \mathbf{A}\mathbf{x} - \mathbf{b})\right]\end{aligned}$$

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

$$\mathbf{\Omega}^+ = \mathrm{E}\left[\left(\mathbf{h}(\mathbf{x}) - \mathbf{A}^+\mathbf{x} - \mathbf{b}^+\right)\left(\mathbf{h}(\mathbf{x}) - \mathbf{A}^+\mathbf{x} - \mathbf{b}^+\right)^T\right]$$

so $\mathrm{tr}\left(\mathbf{\Omega}^+\right)$ is the MSE.

B. SLR in the update step

As we mentioned in the previous section, the SLR of a function with respect to a PDF provides us with the best linear approximation of the function in the region where the PDF lies. This fact is widely used in the update step of Bayesian filtering to get an enabling approximation of the form (5). The conventional way to apply SLR in the update step is to approximate the nonlinear measurement function as in (5) using SLR with respect to the prior. The resulting algorithm is the PrLF, which can be easily approximated using sigma-point methods as in the UKF or CKF. In Appendix A, we prove for completeness that the LMMSE approximation to the posterior (which is approximated by UKF and CKF) is the same as the PrLF.

It is known that the PrLF and therefore all of its approximations do not work well with nonlinear measurement functions if the measurement noise is low enough [5]. In the following, we provide one possible interpretation of this drawback of the PrLF in terms of the accuracy of the approximation (5). This interpretation motivates the introduction of the posterior linearisation filter (PLF).

If the measurement noise is low enough, the posterior PDF is considerably narrower than the prior PDF. Then, if the measurement function is nonlinear and we have performed SLR with respect to the prior, chances are that the linear approximation of the measurement function is not accurate in the region where the posterior actually lies, which is our region of interest. That is, before we process the measurement, the PrLF provides us with the best linear approximation of the measurement function in our region of interest, which is indicated by the prior. However, when we receive the measurement, the region of interest changes according to the posterior, and the linear approximation given by SLR with respect to the prior is not necessarily accurate in the new region.

Intuition tells us that we should approximate the measurement function in our region of interest. That is, the enabling approximation (5) should be chosen by the SLR w.r.t. the posterior PDF, not w.r.t. the prior. The algorithm that uses the SLR of the measurement function with respect to the posterior in the enabling approximation (5) is referred to as PLF. A mathematical justification of using the PLF instead of the PrLF is provided in Section IV. It should also be noted that the PLF is intractable because it requires knowledge of the posterior to approximate the posterior. Nevertheless, we can design an iterative procedure to approximate the PLF. This is addressed in Section V.

1) Illustrative example: In order to clarify the concepts of the previous discussion, we find it convenient to use the following illustrative example. The prior PDF is Gaussian with mean $\overline{x} = 3$ and variance $P = 4$. The measurement equation is

$$z = ax^3 + \eta \tag{14}$$

where η is the measurement noise with variance $R = 0.1$ and $a = 0.01$. In this example, the required moments, i.e., (11)-(13) can be calculated analytically [11] so we can use the exact PrLF instead of an approximation, such as the UKF or CKF.

We analyse the case where we measure $z = 1.5$. The prior and posterior PDFs are shown in Figure 1. In this figure, the posterior has been obtained by using a dense grid of points. This method is not generally practical because of its high computational burden so it is convenient to find computationally efficient approximations to the posterior. The PrLF approximates the posterior by the enabling approximation (5) using SLR of the measurement function with respect to the prior. The measurement function and its PrLF approximation are shown in Figure 2. The linearisation used in the PrLF would be the best linearisation of $h(\cdot)$ if our region of interest were given by the prior, which would be the case if we did not know the measurement. The fact is that we know that the measurement is $z = 1.5$ and we therefore argue that the region of interest is now given by the posterior. The linearisation of $h(\cdot)$ with respect to the posterior is also plotted in Figure 2. It can be clearly seen that the PrLF linearisation is quite different from the linearisation we would like to use in the enabling approximation (5). As the linearisation of the PrLF is not accurate in our region of interest, it is not surprising that the resulting PrLF posterior approximation is poor, see Figure 3. On the contrary, if we use the SLR of the measurement function in the region of interest, given the current measurement, i.e., SLR with respect to the posterior, the resulting posterior approximation is rather accurate.

IV. Mathematical justification of the PLF

In the previous section we provided the intuition behind the PLF. In this section, we provide a mathematical justification as to why the PLF is expected to outperform the PrLF. In the PrLF, the values of $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{\Omega}$ in (5) do not depend on the measurement. Nevertheless, in the update step, the measurement $\mathbf{z}$ is known and we can allow $\mathbf{A}$, $\mathbf{b}$ and $\mathbf{\Omega}$ to depend on $\mathbf{z}$, which means that we can find the values that minimise the MSE of the measurement function approximation for any given $\mathbf{z}$. This implies that the resulting MSE of the measurement function approximation averaged over the state

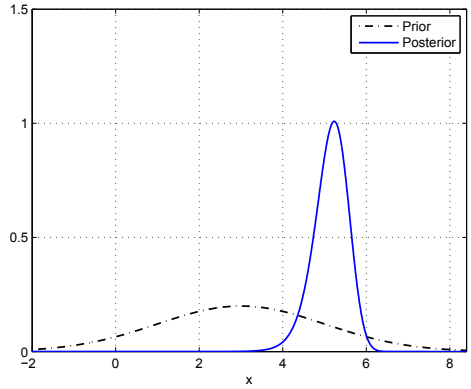


Figure 1: Prior and posterior for $z = 1.5$. The posterior is markedly narrower than the prior.

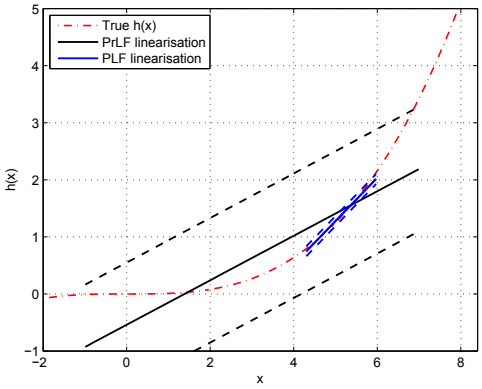


Figure 2: Measurement function and the PrLF and PLF approximations. We plot the linearisations in the 2σ-regions of the PDF that was used in SLR. The dashed lines indicate 2σ-regions of the error term in (5). The PrLF approximation is quite different from the PLF approximation, which gives us the best approximation of $h(\cdot)$ for the current measurement.

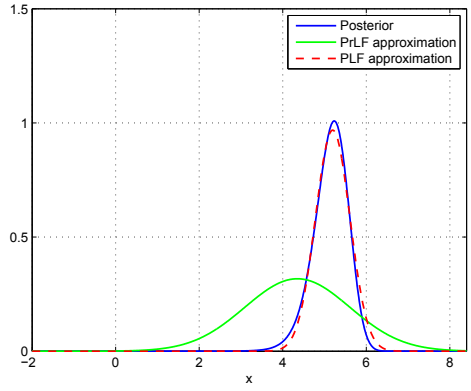


Figure 3: Posterior and the PrLF and PLF approximations. As the PrLF linearisation of $h(\cdot)$ is not accurate, the PrLF approximation is inaccurate. On the contrary, the PLF provides a quite accurate approximation.

and the measurement is lower than for the PrLF. This is explained in the following.

We can write the MSE averaged over the state and the measurement as

$$\begin{aligned}&\mathrm{E}\left[\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\mathbf{x}-\mathbf{b}\right)^{T}\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\mathbf{x}-\mathbf{b}\right)\right]\\&\quad=\int p\left(\mathbf{z}\right)\int p\left(\mathbf{x}\left|\mathbf{z}\right.\right)\\&\qquad\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\mathbf{x}-\mathbf{b}\right)^{T}\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\mathbf{x}-\mathbf{b}\right)d\mathbf{x}d\mathbf{z}\end{aligned}\tag{15}$$

The values of $\mathbf{A},\mathbf{b}$ that minimise (15) are given by SLR with respect to the prior $p\left(\mathbf{x}\right)$. If $\mathbf{A}$ and $\mathbf{b}$ are instead functions of $\mathbf{z}$, rather than minimising (15) globally, we can minimise the inner integral of (15) per each value of $\mathbf{z}$

$$\begin{aligned}&\left(\mathbf{A}^{\star}\left(\mathbf{z}\right),\mathbf{b}^{\star}\left(\mathbf{z}\right)\right)=\\&\underset{\left(\mathbf{A}\left(\mathbf{z}\right),\mathbf{b}\left(\mathbf{z}\right)\right)}{\arg\min}\int p\left(\mathbf{x}\left|\mathbf{z}\right.\right)\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\left(\mathbf{z}\right)\mathbf{x}-\mathbf{b}\left(\mathbf{z}\right)\right)^{T}\\&\qquad\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}\left(\mathbf{z}\right)\mathbf{x}-\mathbf{b}\left(\mathbf{z}\right)\right)d\mathbf{x}\end{aligned}\tag{16}$$

Taking into account the information we know from Section III-A, $\mathbf{A}^{\star}\left(\mathbf{z}\right),\mathbf{b}^{\star}\left(\mathbf{z}\right)$ are given by the SLR of $\mathbf{h}\left(\cdot\right)$ with respect to the posterior.

The MSE $\mathrm{tr}\left(\mathbf{\Omega}^{\star}\left(\mathbf{z}\right)\right)$ of the measurement function conditioned on the measurement is

$$\begin{aligned}\mathrm{tr}\left(\mathbf{\Omega}^{\star}\left(\mathbf{z}\right)\right)=&\int p\left(\mathbf{x}\left|\mathbf{z}\right.\right)\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}^{\star}\left(\mathbf{z}\right)\mathbf{x}-\mathbf{b}^{\star}\left(\mathbf{z}\right)\right)^{T}\\&\left(\mathbf{h}\left(\mathbf{x}\right)-\mathbf{A}^{\star}\left(\mathbf{z}\right)\mathbf{x}-\mathbf{b}^{\star}\left(\mathbf{z}\right)\right)d\mathbf{x}\end{aligned}$$

As proved in Appendix B, the MSE averaged over the state and the measurement of the measurement function approximation provided by any other linearisation, including that of the PrLF and MAP (as in the IEKF), is equal or larger than the MSE provided by the PLF. Therefore, the PLF approximation to the measurement function is optimal in the MSE sense. As the performance of these filters basically depends on the quality of the approximation (5), the PLF is expected to outperform the PrLF and MAP filters.

V. Iterated Posterior Linearisation Filter

In this section, we provide a practical algorithm to approximate the PLF, called the iterated PLF (IPLF). The PLF selects the enabling approximation (5) using SLR with respect to the posterior. The problem is that the posterior is what we aim to approximate, so it is unknown. The PLF approximation provided by the IPLF is based on carrying out iterated SLRs. That is, since we do not have access to the posterior, we perform SLR with respect to the best available approximation of the posterior. At the end of each iteration, we expect to obtain an improved approximation of the posterior which means that we can use it to obtain a better SLR that we compute at the next iteration.

More specifically, we build a sequence $\left(\overline{\mathbf{x}}_{u}^{i},\mathbf{P}_{u}^{i},\mathbf{A}^{i},\mathbf{b}^{i},\mathbf{\Omega}^{i}\right)$ $i\in\mathbb{N}$ of Gaussian posterior and measurement function approximations in the following way. We start with the prior moments $\overline{\mathbf{x}}_{u}^{1}=\overline{\mathbf{x}}$, $\mathbf{P}_{u}^{1}=\mathbf{P}$ and calculate $\mathbf{A}^{1},\mathbf{b}^{1},\mathbf{\Omega}^{1}$ using the SLR of $\mathbf{h}\left(\cdot\right)$ with respect to $\overline{\mathbf{x}}_{u}^{1}$, $\mathbf{P}_{u}^{1}$. Based on $\mathbf{A}^{1},\mathbf{b}^{1}$ and $\mathbf{\Omega}^{1}$, we obtain $\overline{\mathbf{x}}_{u}^{2}$ and $\mathbf{P}_{u}^{2}$, which characterise the posterior

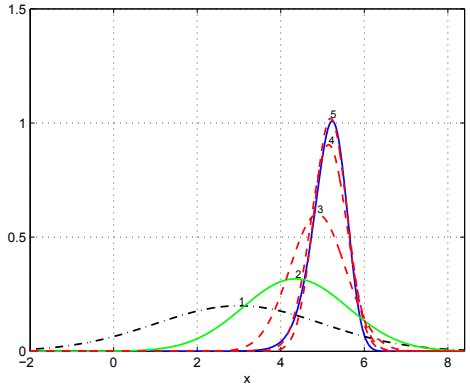


Figure 4: Illustration of IPLF. The true posterior is shown in blue, the prior in black, the PrLF posterior approximation in green and 3 more iterations of the IPLF in red. A number on the maximum of each PDF represents variable i in the IPLF recursion, see Algorithm 1. The PrLF is simply the first step of the recursion and is not a good approximation of the posterior. If we continue with the IPLF iteration, we attain an accurate approximation.

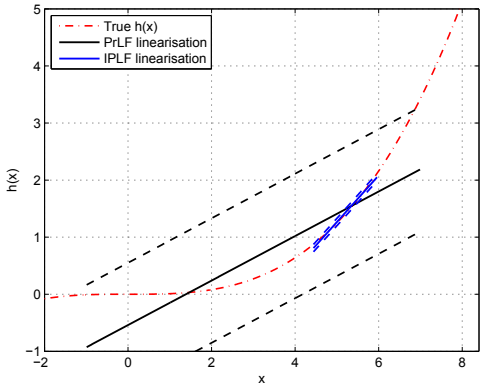


Figure 5: Comparison between the true measurement function and the linearisations used by the PrLF and the IPLF. We plot the linearisations in the 2σ-regions of the PDF that was used in SLR. The dashed lines indicate 2σ-regions of the error term in (5). The SLR of the IPLF is plotted with respect to the last posterior approximation, see Figure 4.

approximation for $i = 2$, from (6) and (7). An interesting observation is that this is the posterior approximation provided by the PrLF. However, as we want to perform SLR with respect to the posterior, we can continue the iteration until convergence. At every step, we obtain $\mathbf{A}^i, \mathbf{b}^i, \mathbf{\Omega}^i$ based on $\overline{\mathbf{x}}_u^i$ and $\mathbf{P}_u^i$. Then, we calculate $\overline{\mathbf{x}}_u^{i+1}$ and $\mathbf{P}_u^{i+1}$ from (6) and (7) using $\mathbf{A}^i, \mathbf{b}^i, \mathbf{\Omega}^i$. In practice, the required integrals of SLR, which are (11)-(13), can be approximated using any sigma-point method, e.g., the unscented transform (UT) [4]. The steps of the IPLF are summarised in Algorithm 1.

Algorithm 1 The update step of IPLF

Input: Prior moments $\overline{\mathbf{x}}_u^1 = \overline{\mathbf{x}}$, $\mathbf{P}_u^1 = \mathbf{P}$.
Output: Posterior moments $\overline{\mathbf{x}}_u^i$, $\mathbf{P}_u^i$.
- Initialise by setting $i = 1$.
repeat
- Calculate the SLR $\mathbf{A}^i, \mathbf{b}^i, \mathbf{\Omega}^i$:
  - Approximate (11)-(13) for $\overline{\mathbf{x}}_u^i$, $\mathbf{P}_u^i$, e.g., using the UT.
  - Obtain $\mathbf{A}^i$, $\mathbf{b}^i$ and $\mathbf{\Omega}^i$ from (8)-(10) using $\overline{\mathbf{x}}_u^i$, $\mathbf{P}_u^i$ instead of $\overline{\mathbf{x}}$, $\mathbf{P}$.
- Compute posterior approximation moments $\overline{\mathbf{x}}_u^{i+1}$, $\mathbf{P}_u^{i+1}$:
  - Use $\mathbf{A}^i$, $\mathbf{b}^i$ and $\mathbf{\Omega}^i$ in (6) and (7).
- $i \leftarrow i + 1$

until convergence, see Section V-A.

Before addressing the convergence rule, let us first analyse how the IPLF works in the illustrative example of Section III-B1. The integrals required for the SLRs are calculated analytically and the results of the IPLF recursion for the illustrative example are shown in Figure 4. The PrLF corresponds to the first step of the IPLF and is not a good approximation of the posterior. As we continue with the IPLF iteration, the posterior approximation becomes closer and closer to the true posterior. It is appealing that the improvement in performance of the IPLF with respect to the PrLF is quite significant with just a few more iterations.

We also show the measurement function and the SLR used at the final step of the IPLF in Figure 5. By comparing this figure with Figure 2, it is remarkable the similarity between the measurement function approximations of the PLF and the IPLF. Once we obtain the measurement, the PrLF linearisation is not a good approximation of $h(\cdot)$ because the state is expected to lie in a smaller area, which is given by the posterior, and the SLR in this area is quite different from the SLR with respect to the prior. On the contrary, IPLF seeks to linearise the measurement function with respect to the posterior. This is done by performing repeated SLRs with respect to the best available approximation to the posterior, which is given by PDF of the last iteration. The resulting posterior approximation is much more accurate than for the PrLF as can be seen in Figure 4.

Finally, we want to show that the improvement indicated by Figures 4 and 5 results in a lower error in state estimation. We estimate the root mean square error (RMSE) of several estimators averaged over the state and the measurement using Monte Carlo simulation. We use 10^5 samples of the state and the measurement and the results are: posterior mean 1.12, IPLF 1.18, PrLF 1.80. As expected, the IPLF outperforms the PrLF. The posterior mean produces the lowest error as it is the minimum MSE estimator.

A. Convergence rule

In this section, we propose a converge rule to determine when we should stop the IPLF iteration. The idea is that the IPLF recursion should finish if the change in the posterior approximation at a given iteration is negligible. In this section, we denote the ith Gaussian approximation as $\mathcal{N}_i(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \overline{\mathbf{x}}_u^i, \mathbf{P}_u^i)$. A usual method to evaluate the similarity between PDFs is the Kullback-Leibler divergence (KLD). Therefore, we stop the recursion if

$$D(\mathcal{N}_i \| \mathcal{N}_{i+1}) < \Gamma \tag{17}$$

where Γ is a threshold and the KLD of $\mathcal{N}_{i+1}(\cdot)$ from $\mathcal{N}_i(\cdot)$ is

$$\begin{aligned} D(\mathcal{N}_i \| \mathcal{N}_{i+1}) = & \int \mathcal{N}_i(\mathbf{x}) \log \frac{\mathcal{N}_i(\mathbf{x})}{\mathcal{N}_{i+1}(\mathbf{x})} d\mathbf{x} \\ = & \frac{1}{2}\left[\operatorname{tr}\left(\left(\mathbf{P}_u^{i+1}\right)^{-1}\mathbf{P}_u^i\right)\right. \\ & + \left(\overline{\mathbf{x}}_u^{i+1} - \overline{\mathbf{x}}_u^i\right)^T \left(\mathbf{P}_u^{i+1}\right)^{-1}\left(\overline{\mathbf{x}}_u^{i+1} - \overline{\mathbf{x}}_u^i\right) \\ & \left. - \ln\left(\frac{\left|\mathbf{P}_u^i\right|}{\left|\mathbf{P}_u^{i+1}\right|}\right) - n_x\right] \end{aligned}$$

In principle, we could have also chosen the KLD $D(\mathcal{N}_{i+1} \| \mathcal{N}_i)$ rather than (17). However, in the examples of Section VI, (17) works better. The main reason why this happens is that the KLD in (17) is large if $\mathcal{N}_{i+1}(\cdot)$ is small in the region where $\mathcal{N}_i(\cdot)$ is large [12]. As exemplified in Figure 4, $\mathcal{N}_{i+1}(\cdot)$ is expected to be more concentrated than $\mathcal{N}_i(\cdot)$ until the algorithm converges. Therefore, in order to increment the value of the KLD before the algorithm converges, it is convenient to use the KLD (17) rather than $D(\mathcal{N}_{i+1} \| \mathcal{N}_i)$.

B. A comparison with other iterated filters

Iterated sigma-point filters have previously been proposed in [13] and [14] but with ad-hoc approaches. In [13], only one set of sigma-points is generated and used to approximate the prior moments. Using these prior moment approximations the iteration proceeds similarly to the IEKF [2]. However, the analytical linearisation of the IEKF is replaced by an ad-hoc linearisation that mixes SLR w.r.t. the prior and analytical linearisation at the current MAP estimate. In [14], the iteration requires several ad-hoc parameters and conditions and the overall effect of the iteration is that several corrections are performed with the same measurement even though we observe it only once. Furthermore, the underlying philosophy of the filters in [13] and [14] is also different from ours as they attempt to find the MAP estimate while our objective is to approximate the PLF.

VI. Numerical example

In this section, we compare the performance of the IPLF with other filters commonly used in the literature. Specifically, we compare it with the following approximations of the PrLF: EKF, UKF and CKF. The UT of the UKF and IPLF has been implemented with $N_s = 2n_x + 1$ sigma-points and the weight of the sigma-point located on the mean is 1/3. The threshold $\Gamma = 10^{-1}$. The prediction step of the IPLF is performed as in the UKF. In addition, we have also tested the IEKF, which is a MAP filter, with 50 iterations.

We consider an air-traffic control scenario, where an aircraft executes a maneuvering turn in a horizontal plane. The state vector at time k is $\mathbf{x}^k = \left[p_x^k, \dot{p}_x^k, p_y^k, \dot{p}_y^k, \Omega^k\right]^T$ where Ω^k is the turn rate at time k and, $\left[p_x^k, p_y^k\right]^T$ and $\left[\dot{p}_x^k, \dot{p}_y^k\right]^T$ are the position and velocity vector in the x and y coordinates at time k respectively. The kinematics of the turning motion are modeled by

$$\mathbf{x}^{k+1} = \mathbf{F}\left(\Omega^k\right)\mathbf{x}^k + \mathbf{v}^k \tag{18}$$

where

$$\mathbf{F}(\Omega) = \begin{bmatrix} 1 & \frac{\sin\Omega\tau}{\Omega} & 0 & -\left(\frac{1-\cos\Omega\tau}{\Omega}\right) & 0 \\ 0 & \cos\Omega\tau & 0 & -\sin\Omega\tau & 0 \\ 0 & \left(\frac{1-\cos\Omega\tau}{\Omega}\right) & 1 & \frac{\sin\Omega\tau}{\Omega} & 0 \\ 0 & \sin\Omega\tau & 0 & \cos\Omega\tau & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{19}$$

and τ is the sampling period and $\mathbf{v}^k$ is the process noise at time k. We assume that $\mathbf{v}^k$ is zero-mean Gaussian distributed with covariance matrix

$$\mathbf{Q} = \begin{bmatrix} q_1\frac{\tau^3}{3} & q_1\frac{\tau^2}{2} & 0 & 0 & 0 \\ q_1\frac{\tau^2}{2} & q_1\tau & 0 & 0 & 0 \\ 0 & 0 & q_1\frac{\tau^3}{3} & q_1\frac{\tau^2}{2} & 0 \\ 0 & 0 & q_1\frac{\tau^2}{2} & q_1\tau & 0 \\ 0 & 0 & 0 & 0 & q_2 \end{bmatrix} \tag{20}$$

where q_1 and q_2 are parameters of the motion model. As is usually assumed in tracking, $\mathbf{v}^k$ is independent of $\mathbf{v}^m$ if $m \neq k$.

The sensor produces range, bearings and range rate measurements modelled by [15]

$$\mathbf{z}^k = \begin{bmatrix} \sqrt{\left(p_x^k\right)^2 + \left(p_y^k\right)^2} \\ \operatorname{atan2}\left(p_y^k, p_x^k\right) \\ \frac{p_x^k \dot{p}_x^k + p_y^k \dot{p}_y^k}{\sqrt{\left(p_x^k\right)^2 + \left(p_y^k\right)^2}} \end{bmatrix} + \mathbf{w}^k \tag{21}$$

where $\operatorname{atan2}(\cdot,\cdot)$ is the four-quadrant inverse tangent and $\mathbf{w}^k$ is the zero-mean Gaussian measurement noise at time k such that $\mathbf{w}^k$ is independent of $\mathbf{w}^m$ if $m \neq k$.

In order to illustrate how the filter performances vary with the accuracy of the measurements, we consider a scenario where the tracking system has two measurement modes with different accuracies. Sensors with different accuracies are usually used in sensor management applications. In the first type of measurement, the covariance matrix of $\mathbf{w}^k$ is $\mathbf{R}_1 = \operatorname{diag}\left(\left[\sigma_{r,1}^2, \sigma_{\theta,1}^2, \sigma_{\dot{r},1}^2\right]\right)$ and, in the second, $\mathbf{R}_2 = \operatorname{diag}\left(\left[\sigma_{r,2}^2, \sigma_{\theta,2}^2, \sigma_{\dot{r},2}^2\right]\right)$. In this example, measurements of the second type are performed every M time steps. The target trajectory used to evaluate the filter performance is shown in Figure 6. The total number of time steps in the simulation is 160.

The prior at time 0 is

$$p(\mathbf{x}^0) = \mathcal{N}\left(\mathbf{x}^0; \bar{\mathbf{x}}^0, \mathbf{\Sigma}^0\right) \tag{22}$$

where $\mathbf{\Sigma}^0 = \operatorname{diag}\left(\left[\sigma_{p_x}^2, \sigma_{\dot{p}_x}^2, \sigma_{p_y}^2, \sigma_{\dot{p}_y}^2, \sigma_\Omega^2\right]\right)$, with $\sigma_{p_x}^2 = \sigma_{p_y}^2 = 200\,\mathrm{m}^2$, $\sigma_{\dot{p}_x}^2 = 10\,\mathrm{m}^2/\mathrm{s}^2$, $\sigma_{\dot{p}_y}^2 = 50\,\mathrm{m}^2/\mathrm{s}^2$, $\sigma_\Omega^2 = 10^{-3}\,\mathrm{rad}^2/\mathrm{s}^2$, and the prior mean $\bar{\mathbf{x}}^0$ is chosen randomly from a Gaussian PDF with mean identical to the true initial state and covariance matrix $\mathbf{\Sigma}^0$. The scenario parameters are shown in Table I. Every 20 steps, the accuracy of the measurement is much higher and the approximations to the PrLF (EKF, UKF, CKF) are expected to perform worse than the IPLF [5].

We assess the filters by Monte Carlo simulation with 1000 runs. The performances of EKF and IEKF are quite poor. EKF diverges in all runs and IEKF diverges in 35.1% of the runs so they are not further considered. The rest of the filters do not

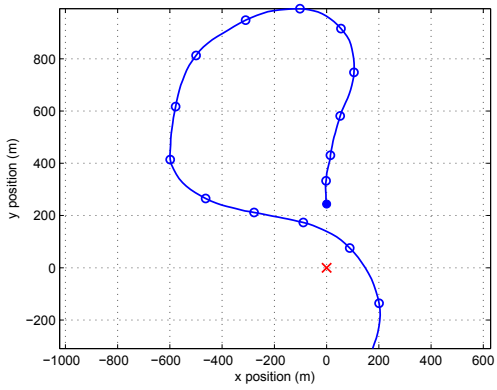


Figure 6: Turning target tracking scenario. The trajectory of the target is represented in blue. The target position every ten time steps is represented by a blue circumference and its initial position by a filled blue circle. The sensor position is represented by a red cross.

Table I: Parameters for the turning target example

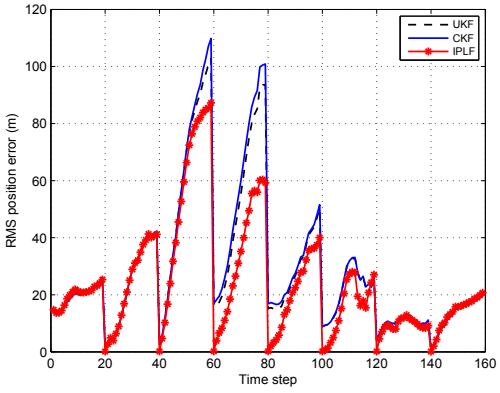
Parameter	Value
τ	1 s
q_1	$1\ \mathrm{m}^2/\mathrm{s}^3$
q_2	$1.75 \cdot 10^{-4}\mathrm{rad}/\mathrm{s}^3$
$\sigma^2_{r,1}$	$50\ \mathrm{m}^2$
$\sigma^2_{\theta,1}$	$(8\pi/180)^2\ \mathrm{rad}^2$
$\sigma^2_{\dot{r},1}$	$0.1\ \mathrm{m}^2/\mathrm{s}^2$
$\sigma^2_{r,2}$	$0.01\ \mathrm{m}^2$
$\sigma^2_{\theta,2}$	$(0.01\pi/180)^2\ \mathrm{rad}^2$
$\sigma^2_{\dot{r},2}$	$10^{-3}\ \mathrm{m}^2/\mathrm{s}^2$
M	20

diverge. The root mean square (RMS) error for the position, velocity and turn rate against time for the rest of the algorithms are shown in Figure 7. The IPLF error at time steps multiple of 20 is much lower than the error of PrLF algorithms due to the high accuracy of the measurement. The consequence of this is that at other time steps, especially between time steps 40 and 120, the RMS error of the IPLF is much lower than the error of PrLF algorithms. As a result, on the whole, IPLF clearly outperforms PrLF algorithms.

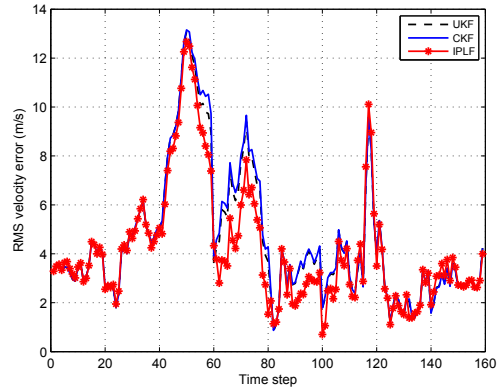
We show the average number of iterations of the IPLF against time in Figure 8. For the given converge rule, the IPLF does not require a high number of iterations to converge, which increase for accurate measurements. The execution times in milliseconds of our non-optimised Matlab implementation of the algorithms are: CKF and UKF (40) and IPLF (80). The computational burden of IPLF can be adjusted with the parameter Γ that controls the stopping rule of the recursion. A higher value of Γ implies a lower computational burden. If Γ is high enough, the IPLF reduces to the UKF.

VII. Conclusions

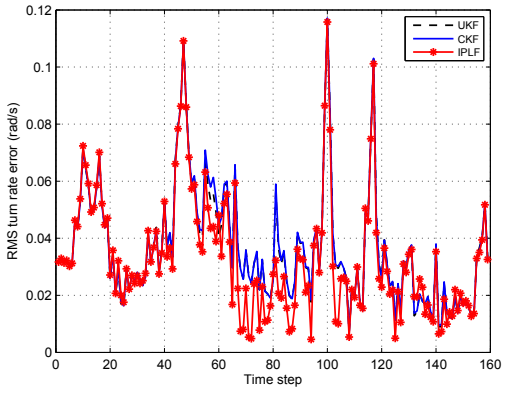
In this paper, we have developed the PLF and its practical approximation, the IPLF. The PLF is based on performing SLR of the measurement function with respect to the posterior instead of the prior. This is motivated by the fact that the approximation of the measurement function should be accurate in the region of interest, which is indicated by the posterior. The PLF is intractable but can be efficiently approximated by



(a)



(b)



(c)

Figure 7: RMS errors against time for turning target example (a) position (b) velocity (c) turn rate. On the whole, IPLF provides a lower error, especially between time steps 40 and 120.

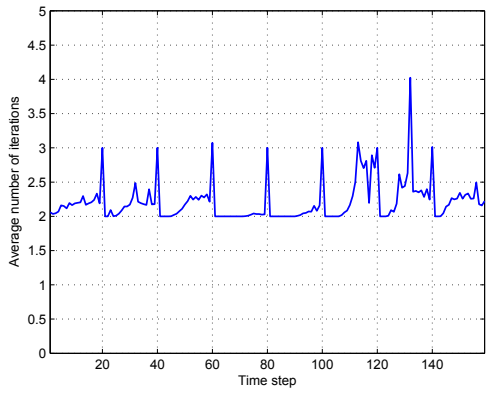


Figure 8: Average number of iterations of the IPLF against time. For accurate measurements, more steps are carried out in the recursion.

the IPLF. The IPLF is based on performing iterated SLR of the measurement function with respect to the best available approximation of the posterior. On the whole, as the IPLF is an approximation of the PLF, it is expected to outperform PrLF and its approximations, e.g., EKF, UKF and CKF.

Future work will address the convergence of the IPLF and its extension to smoothing problems and Bayesian graphical models.

VIII. Acknowledgements

Ángel F. García-Fernández is supported by the Australian Research Council under Discovery Project DP130104404.

Appendix A

In this appendix, we show that the PrLF update is equivalent to the LMMSE approximation to the posterior, which is sometimes referred to as (non-linear) Kalman filter. The updated mean and covariance matrix of the PrLF are

$$\overline{\mathbf{x}}_u = \overline{\mathbf{x}} + \mathbf{P}\left(\mathbf{A}^+\right)^T \left(\mathbf{A}^+\mathbf{P}\left(\mathbf{A}^+\right)^T + \mathbf{\Omega}^+ + \mathbf{R}\right)^{-1} \left(\mathbf{z} - \mathbf{A}^+\overline{\mathbf{x}} - \mathbf{b}^+\right) \tag{23}$$

$$\mathbf{P}_u = \mathbf{P} - \mathbf{P}\left(\mathbf{A}^+\right)^T \left(\mathbf{A}^+\mathbf{P}\left(\mathbf{A}^+\right)^T + \mathbf{\Omega}^+ + \mathbf{R}\right)^{-1} \mathbf{A}^+\mathbf{P} \tag{24}$$

where $\mathbf{A}^+$, $\mathbf{b}^+$ and $\mathbf{\Omega}^+$ are given by (8)-(10). From (8), we get $\mathbf{P}\left(\mathbf{A}^+\right)^T = \mathbf{\Psi}$. If we also substitute (9), (10) into (23) and (24), we complete the proof

$$\overline{\mathbf{x}}_u = \overline{\mathbf{x}} + \mathbf{\Psi}\left(\mathbf{\Phi} + \mathbf{R}\right)^{-1}\left(\mathbf{z} - \overline{\mathbf{z}}\right) \tag{25}$$

$$\mathbf{P}_u = \mathbf{P} - \mathbf{\Psi}\left(\mathbf{\Phi} + \mathbf{R}\right)^{-1}\mathbf{\Psi}^T \tag{26}$$

where $\overline{\mathbf{z}}$, $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are provided by (11)-(13). Equations (25) and (26) correspond to the LMMSE posterior moments, which are approximated by widely known algorithms such as UKF or CKF [4], [9].

Appendix B

In this appendix, we prove that the MSE averaged over the state and the measurement of the measurement function approximation provided by PLF is lower or equal than for any other linearisation $\mathbf{A}(\mathbf{z})$, $\mathbf{b}(\mathbf{z})$, including that of the PrLF and MAP (as in the IEKF). The MSE $\mathrm{tr}(\mathbf{\Omega})$ of linearisation $\mathbf{A}(\mathbf{z}), \mathbf{b}(\mathbf{z})$ can be written as

$$\mathrm{tr}(\mathbf{\Omega}) = \int p(\mathbf{z})\, d\mathbf{z} \int p(\mathbf{x}|\mathbf{z}) \left(\mathbf{h}(\mathbf{x}) - \mathbf{A}(\mathbf{z})\mathbf{x} - \mathbf{b}(\mathbf{z})\right)^T \left(\mathbf{h}(\mathbf{x}) - \mathbf{A}(\mathbf{z})\mathbf{x} - \mathbf{b}(\mathbf{z})\right) d\mathbf{x}$$

Because of (16), we get

$$\begin{aligned}\mathrm{tr}(\mathbf{\Omega}) &\geq \int p(\mathbf{z})\, d\mathbf{z} \int p(\mathbf{x}|\mathbf{z}) \left(\mathbf{h}(\mathbf{x}) - \mathbf{A}^\star(\mathbf{z})\mathbf{x} - \mathbf{b}^\star(\mathbf{z})\right)^T \left(\mathbf{h}(\mathbf{x}) - \mathbf{A}^\star(\mathbf{z})\mathbf{x} - \mathbf{b}^\star(\mathbf{z})\right) d\mathbf{x} \\ &= \int p(\mathbf{z})\, \mathrm{tr}\left(\mathbf{\Omega}^\star(\mathbf{z})\right) d\mathbf{z}\end{aligned}$$

where the last integral is the MSE averaged over the state and the measurement of the PLF measurement function approximation.

References

[1] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.

[2] B. Bell and F. Cathey, "The iterated Kalman filter update as a Gauss-Newton method," *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 294–297, Feb. 1993.

[3] A. M. Walker, "On the asymptotic behaviour of posterior distributions," *Journal of the Royal Statistical Society. Series B (Methodological).*, vol. 31, no. 1, pp. 80–88, 1969.

[4] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[5] M. R. Morelande and A. F. García-Fernández, "Analysis of Kalman filter approximations for nonlinear measurements," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5477–5484, Nov. 2013.

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

[7] I. Arasaratnam, S. Haykin, and R. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[8] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, Oct. 2003.

[9] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.

[10] T. Lefebvre, H. Bruyninckx, and J. De Schuller, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [and authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, Aug. 2002.

[11] A. F. García-Fernández, "Detection and tracking of multiple targets using wireless sensor networks," Ph.D. dissertation, Universidad Politécnica de Madrid, 2011. [Online]. Available: http://oa.upm.es/9823/

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science + Business Media, 2006.

[13] G. Sibley, G. Sukhatme, and L. Matthies, "The iterated sigma point Kalman filter with applications to long range stereo," in *Proceedings of Robotics: Science and Systems*, 2006.

[14] R. Zhan and J. Wan, "Iterated unscented Kalman filter for passive target tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 1155–1163, July 2007.

[15] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.