

Iterated Filters for Nonlinear Transition Models

Anton Kullberg, Isaac Skog and Gustaf Hendeby

The self-archived postprint version of this conference paper is available at Linköping University Institutional Repository (DiVA):

<https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-197794>

N.B.: When citing this work, cite the original publication.

Kullberg, A., Skog, I., Hendeby, G., (2023), Iterated Filters for Nonlinear Transition Models, 2023 26th International Conference on Information Fusion (FUSION 2023).

<https://doi.org/10.23919/FUSION52260.2023.10224211>

Original publication available at:

<https://doi.org/10.23919/FUSION52260.2023.10224211>

Copyright: Institute of Electrical and Electronics Engineers (IEEE)

<http://www.ieee.org/>

©2023 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Iterated Filters for Nonlinear Transition Models

Anton Kullberg*, Isaac Skog†, *Senior Member, IEEE*, and Gustaf Hendeby*, *Senior Member, IEEE*.

*Dept. Electrical Engineering, Linköping University, Linköping, Sweden

†Dept. Electrical Engineering, Uppsala University, Uppsala, Sweden

Email: {anton.kullberg, gustaf.hendeby}@liu.se, isaac.skog@angstrom.uu.se

Abstract—A new class of iterated linearization-based nonlinear filters, dubbed dynamically iterated filters, is presented. Contrary to regular iterated filters such as the iterated extended Kalman filter (IEKF), iterated unscented Kalman filter (IUKF) and iterated posterior linearization filter (IPLF), dynamically iterated filters also take nonlinearities in the transition model into account. The general filtering algorithm is shown to essentially be a (locally over one time step) iterated Rauch-Tung-Striebel smoother. Three distinct versions of the dynamically iterated filters are especially investigated: analogues to the IEKF, IUKF and IPLF. The developed algorithms are evaluated on 25 different noise configurations of a tracking problem with a nonlinear transition model and linear measurement model, a scenario where conventional iterated filters are not useful. Even in this “simple” scenario, the dynamically iterated filters are shown to have superior root mean-squared error performance as compared with their respective baselines, the EKF and UKF. Particularly, even though the EKF diverges in 22 out of 25 configurations, the dynamically iterated EKF remains stable in 20 out of 25 scenarios, only diverging under high noise.

I. INTRODUCTION

State estimation in dynamical systems is a universal problem occurring in the fields of engineering, robotics, economics, etc. State estimation requires a system model describing the dynamical evolution of the system and a measurement model relating the measured quantities to the state of the system. If the model is affine with additive Gaussian noise, the most well-known state estimation algorithm is the analytically tractable Kalman filter, which is the optimal estimator in the *mean-squared error* (MSE) sense [1].

In many practical problems, a nonlinear system model is necessary to accurately describe the system. This means that the state estimation problem is no longer analytically tractable and approximate inference techniques must be used. Approximate inference in state-space models is a well-studied field in signal processing, machine learning, etc. Here, we shall focus on linearization-based approximate inference techniques. These inference techniques linearize the nonlinear model locally (in each time instance) and then employ the Kalman filter. Analytical linearization leads to the *extended Kalman filter* (EKF), while sigma-point filters, such as the *unscented Kalman filter* (UKF) and the *cubature Kalman filter* (CKF), can be thought of as statistical linearization filters [1]–[3].

General (Gaussian) state-space models, in the form of a transition model and a measurement model, may equiva-

lently be probabilistically interpreted as a transition density and a measurement density. Under this interpretation, the linearization-based approximate inference techniques can be thought of as approximating the transition and measurement densities, e.g.,

$$\begin{aligned} \mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{w}_k) & \rightarrow p(\mathbf{x}_{k+1}|\mathbf{x}_k) \stackrel{a}{\approx} q(\mathbf{x}_{k+1}|\mathbf{x}_k) \\ \mathbf{y}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{e}_k) & \rightarrow p(\mathbf{y}_k|\mathbf{x}_k) \stackrel{a}{\approx} q(\mathbf{y}_k|\mathbf{x}_k), \end{aligned}$$

where $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and $p(\mathbf{y}_k|\mathbf{x}_k)$ are the transition and measurement density and $q(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and $q(\mathbf{y}_k|\mathbf{x}_k)$ the corresponding approximations. Particularly, the linearization-based filters assume affine Gaussian densities for $q(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and $q(\mathbf{y}_k|\mathbf{x}_k)$ and the Kalman filter is then applied to this “auxiliary” model. The quality of the auxiliary model, and in extension the estimation performance of linearization-based filters, is thus highly dependent on the point (distribution in the statistical case) about which the models are linearized. Typically, the linearization point (distribution) is chosen to be the mean (distribution) of the current state estimate. However, a large error in the state estimate can lead to significant linearization errors that may cause even larger estimation errors in the next time step. This may, in the worst case, cause the filter to diverge. To alleviate such issues, several variants of iterated filters have been developed, such as the *iterated extended Kalman filter* (IEKF), the *iterated unscented Kalman filter* (IUKF) and the *iterated posterior linearization filter* (IPLF) [4]–[8]. These filters essentially iterate the measurement update, where each iteration the measurement model is re-linearized with the “latest” iterate. The research efforts within the field of iterated filters have particularly focused on finding a better linearization point for the measurement model, which is motivated by the fact that nonlinearities in the measurement model (likelihood) affect the resulting state estimate to a greater extent than nonlinearities in the transition model (prior). Nevertheless, these methods are for instance not useful in the case of a nonlinear transition model but linear measurement model.

In this paper, we seek to fill this gap by developing a class of iterated filters encompassing both the transition model and the measurement model in the iterative process, which we dub dynamically iterated filters. Note that a dynamically iterated filter based on posterior linearization was first derived in [9] for models with non-additive state transition noise. Further, the L-scan IPLF in [10] is somewhat similar to the dynamical IPLF developed here, but requires access to past observations

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

and is thus not strictly a filter. In this paper, we particularly focus on additive noise models and treat both analytical as well as statistical linearization in a common framework. The algorithms developed here are essentially dynamically iterated analogues of the IEKF, IUKF and IPLF, as well as other iterated sigma-point filters and does thus not require access to past observations. These new iterative algorithms encompass both the transition model as well as the measurement model. Thereby, the proposed algorithms constitute a generalization of conventional iterated filters. To illustrate the benefits of the proposed algorithms, it is empirically shown that iterating over the transition linearization improves the estimation performance even in the case of a linear measurement model. Thus, the contributions are twofold:

- A detailed derivation of dynamically iterated filters
- An extensive numerical evaluation of the developed algorithms as compared to standard nonlinear filters

The paper is organized as follows. In Section II, analytical and statistical linearization as well as the (affine) Kalman smoother equations are restated for completeness. In Section III, the state estimation problem is formulated in terms of approximate transition and measurement densities. Section IV derives the dynamically iterated filters and connects the final solution to iterated (affine) smoothers. Lastly, Section V provides a numerical example of the developed algorithm in a tracking scenario where conventional iterated filters are not useful.

II. BACKGROUND

For clarity, we here present analytical and statistical linearization in a common framework, as well as restate the well-known Kalman smoother equations.

A. (Affine) Kalman Smoother

The well-known Kalman filter and *Rauch-Tung-Striebel* (RTS) smoother equations are repeated here for clarity in terms of a time update, measurement update, and a smoothing step. These can for instance be found in [11]. Assume an affine state-space model with additive Gaussian noise, such as

$$\mathbf{x}_{k+1} = \mathbf{A}_f \mathbf{x}_k + \mathbf{b}_f + \tilde{\mathbf{w}}_k \quad (1a)$$

$$\mathbf{y}_k = \mathbf{A}_h \mathbf{x}_k + \mathbf{b}_h + \tilde{\mathbf{e}}_k. \quad (1b)$$

Here, $\tilde{\mathbf{w}}_k \sim \mathcal{N}(\tilde{\mathbf{w}}_k; \mathbf{0}, \mathbf{Q} + \mathbf{\Omega}_f)$ and $\tilde{\mathbf{e}}_k \sim \mathcal{N}(\tilde{\mathbf{e}}_k; \mathbf{0}, \mathbf{R} + \mathbf{\Omega}_h)$ are assumed to be mutually independent. Note that usually, $\mathbf{\Omega}_f = \mathbf{\Omega}_h = \mathbf{0}$. For this model, the (affine) Kalman smoother update equations are given by Algorithm 1.

B. Analytical and Statistical Linearization

Given a nonlinear model

$$\mathbf{z} = \mathbf{g}(\mathbf{x}),$$

we wish to find an affine representation

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{A}\mathbf{x} + \mathbf{b} + \eta, \quad (5)$$

Algorithm 1 (Affine) Kalman smoother

1) Time update

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}_f \hat{\mathbf{x}}_{k|k} + \mathbf{b}_f \quad (2a)$$

$$\mathbf{P}_{k+1|k} = \mathbf{A}_f \mathbf{P}_{k|k} \mathbf{A}_f^\top + \mathbf{Q} + \mathbf{\Omega}_f. \quad (2b)$$

2) Measurement update

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \mathbf{A}_h \hat{\mathbf{x}}_{k|k-1} - \mathbf{b}_h) \quad (3a)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{A}_h \mathbf{P}_{k|k-1} \quad (3b)$$

$$\mathbf{K}_k \triangleq \mathbf{P}_{k|k-1} \mathbf{A}_h^\top (\mathbf{A}_h \mathbf{P}_{k|k-1} \mathbf{A}_h^\top + \mathbf{R} + \mathbf{\Omega}_h)^{-1}. \quad (3c)$$

3) Smoothing step

$$\hat{\mathbf{x}}_{k|K}^s = \hat{\mathbf{x}}_{k|k} + \mathbf{G}_k (\hat{\mathbf{x}}_{k+1|K}^s - \hat{\mathbf{x}}_{k+1|k}) \quad (4a)$$

$$\mathbf{P}_{k|K}^s = \mathbf{P}_{k|k} + \mathbf{G}_k (\mathbf{P}_{k+1|K}^s - \mathbf{A}_f \mathbf{P}_{k|k} \mathbf{A}_f^\top - \mathbf{Q} - \mathbf{\Omega}_f) \mathbf{G}_k^\top \quad (4b)$$

$$\mathbf{G}_k \triangleq \mathbf{P}_{k|k} \mathbf{A}_f^\top (\mathbf{A}_f \mathbf{P}_{k|k} \mathbf{A}_f^\top + \mathbf{Q} + \mathbf{\Omega}_f)^{-1} \quad (4c)$$

with $\eta \sim \mathcal{N}(\eta; \mathbf{0}, \mathbf{\Omega})$. In this affine representation, there are three free parameters: \mathbf{A} , \mathbf{b} and $\mathbf{\Omega}$. Analytical linearization through first-order Taylor expansion selects the parameters as

$$\mathbf{A} = \frac{d}{d\mathbf{x}} \mathbf{g}(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}}, \quad \mathbf{b} = \mathbf{g}(\mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}} - \mathbf{A}\bar{\mathbf{x}}, \quad \mathbf{\Omega} = \mathbf{0}, \quad (6)$$

where $\bar{\mathbf{x}}$ is the point about which the function $\mathbf{g}(\mathbf{x})$ is linearized. Note that $\mathbf{\Omega} = \mathbf{0}$ essentially implies that the linearization is assumed to be error free.

Statistical linearization instead linearizes w.r.t. a distribution $p(\mathbf{x})$. Assuming that such a distribution $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \mathbf{P})$ is given, statistical linearization selects the affine parameters as

$$\mathbf{A} = \Psi^\top \mathbf{P}^{-1} \quad (7a)$$

$$\mathbf{b} = \bar{\mathbf{z}} - \mathbf{A}\hat{\mathbf{x}} \quad (7b)$$

$$\mathbf{\Omega} = \Phi - \mathbf{A}\mathbf{P}\mathbf{A}^\top \quad (7c)$$

$$\bar{\mathbf{z}} = \mathbb{E}[\mathbf{g}(\mathbf{x})] \quad (7d)$$

$$\Psi = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{z}})^\top] \quad (7e)$$

$$\Phi = \mathbb{E}[(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{z}})(\mathbf{g}(\mathbf{x}) - \bar{\mathbf{z}})^\top], \quad (7f)$$

where the expectations are taken w.r.t. $p(\mathbf{x})$. The major difference from analytical linearization is that $\mathbf{\Omega} \neq \mathbf{0}$, implying that the error in the linearization is captured.

III. PROBLEM FORMULATION

To set the stage for the algorithm development, the general state estimation problem is described here with a probabilistic viewpoint. To that end, consider a discrete-time state-space model (omitting a possible input \mathbf{u}_k for notational brevity) given by

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k \quad (8a)$$

$$\mathbf{y}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{e}_k \quad (8b)$$

$$p(\mathbf{w}_k) = \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}), \quad p(\mathbf{e}_k) = \mathcal{N}(\mathbf{e}_k; \mathbf{0}, \mathbf{R}). \quad (8c)$$

Here, \mathbf{x}_k , \mathbf{y}_k , \mathbf{w}_k and \mathbf{e}_k denote the state, the measurement, the process noise and the measurement noise at time k , respectively. It is further assumed that $\mathbf{x}_k \in \mathcal{X}, \forall k$ and that \mathbf{w}_k and \mathbf{e}_k are mutually independent. Note that (8a) and (8b) can equivalently be written as a *transition density* and a *measurement density* as

$$p(\mathbf{x}_{k+1}|\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_{k+1}; \mathbf{f}(\mathbf{x}_k), \mathbf{Q}) \quad (9a)$$

$$p(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{h}(\mathbf{x}_k), \mathbf{R}). \quad (9b)$$

Further, the initial state distribution is assumed to be given by

$$p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}). \quad (10)$$

Given the transition and measurement densities and a sequence of measurements $\mathbf{y}_{1:k} = [\mathbf{y}_1^\top \dots \mathbf{y}_k^\top]^\top$, the state estimation problem consists of computing the posterior of the state sequence (trajectory), i.e., computing

$$p(\mathbf{x}_{0:k}|\mathbf{y}_{1:k}) = \frac{1}{\mathbf{Z}_{1:k}} p(\mathbf{x}_0) \prod_{i=1}^k p(\mathbf{y}_i|\mathbf{x}_i) p(\mathbf{x}_i|\mathbf{x}_{i-1}), \quad (11)$$

where

$$\mathbf{Z}_{1:k} = \int_{\mathcal{X}} p(\mathbf{x}_0) \prod_{i=1}^k p(\mathbf{y}_i|\mathbf{x}_i) p(\mathbf{x}_i|\mathbf{x}_{i-1}) d\mathbf{x}_0 \dots d\mathbf{x}_k,$$

is the marginal likelihood of $\mathbf{y}_{1:k}$. The posterior (11) is commonly referred to as the joint *smoothing* distribution which, in the case of linear \mathbf{f} and \mathbf{h} , can be analytically found through the Kalman smoother, e.g., the RTS smoother [11].

In the setting considered here, i.e., in *filtering* applications, the densities of interest are rather the *marginal* posteriors

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k) \int_{\mathcal{X}} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}}{\mathbf{Z}_k}, \quad (12)$$

for all times k , where

$$\mathbf{Z}_k = \int_{\mathcal{X}} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1} d\mathbf{x}_k.$$

Again, in the case of linear \mathbf{f} and \mathbf{h} , the (analytical) solution is given by the Kalman filter [1].

In the general case, the marginal posteriors can not be computed analytically. Inspecting (12), there are two integrals that require attention. We turn first to the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int_{\mathcal{X}} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) d\mathbf{x}_{k-1}. \quad (13)$$

Assuming that $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ is Gaussian, (13) has a closed form solution given by (2), if $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is Gaussian and (8a) is affine. Therefore, as (9a) is Gaussian, we seek an affine approximation of the transition function \mathbf{f} as

$$\mathbf{f}(\mathbf{x}_{k-1}) \approx \mathbf{A}_f \mathbf{x}_{k-1} + \mathbf{b}_f + \eta_f, \quad (14)$$

with $p(\eta_f) = \mathcal{N}(\eta_f; \mathbf{0}, \mathbf{\Omega}_f)$. Hence, the transition density $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is approximated by $q(\mathbf{x}_k|\mathbf{x}_{k-1})$ as

$$q(\mathbf{x}_k|\mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{A}_f \mathbf{x}_{k-1} + \mathbf{b}_f, \mathbf{Q} + \mathbf{\Omega}_f). \quad (15)$$

If \mathbf{A}_f , \mathbf{b}_f and $\mathbf{\Omega}_f$ are chosen to be the analytical linearization of \mathbf{f} around the mean of the posterior $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$, the EKF time update is recovered through (2). Similarly, statistical linearization around the posterior at time $k-1$ recovers the sigma-point filter time updates. This yields an approximate predictive distribution $q(\mathbf{x}_k|\mathbf{y}_{1:k-1})$, which can then be used to approximate the second integral of interest (and subsequently, the posterior at time k). Explicitly, the second integral is given by

$$\mathbf{Z}_k \approx \int_{\mathcal{X}} p(\mathbf{y}_k|\mathbf{x}_k) q(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k. \quad (16)$$

Similarly to (14), (16) has a closed form solution if $p(\mathbf{y}_k|\mathbf{x}_k)$ is Gaussian and (8b) is affine. Thus, as (9b) is Gaussian, we seek an affine approximation of the measurement function \mathbf{h} as

$$\mathbf{h}(\mathbf{x}_k) \approx \mathbf{A}_h \mathbf{x}_k + \mathbf{b}_h + \eta_h, \quad (17)$$

with $p(\eta_h) = \mathcal{N}(\eta_h; \mathbf{0}, \mathbf{\Omega}_h)$. Hence, the measurement density $p(\mathbf{y}_k|\mathbf{x}_k)$ is approximated by $q(\mathbf{y}_k|\mathbf{x}_k)$ as

$$q(\mathbf{y}_k|\mathbf{x}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{A}_h \mathbf{x}_k + \mathbf{b}_h, \mathbf{R} + \mathbf{\Omega}_h), \quad (18)$$

which leads to an analytically tractable integral. With (15) and (18), the (approximate) marginal posterior (12) is now given by

$$q(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{q(\mathbf{y}_k|\mathbf{x}_k) q(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{\int_{\mathcal{X}} q(\mathbf{y}_k|\mathbf{x}_k) q(\mathbf{x}_k|\mathbf{y}_{1:k-1}) d\mathbf{x}_k}, \quad (19)$$

which is analytically tractable and given by (3). Note that analytical linearization of (17) around the mean of $q(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ recovers the EKF measurement update whereas statistical linearization recovers the sigma-point measurement update(s).

The quality of the approximate marginal posterior (19) thus directly depends on the quality of the approximations (15) and (18). The quality of (15) and (18) in turn directly depends on the choice of linearization points or densities, which is typically chosen to be the approximate predictive and previous approximate posterior distributions. This choice is of course free and iterative filters such as the IEKF, IUKF or IPLF have been proposed to improve these approximations [4]–[6], [12]. These filters can be thought of as finding an approximate posterior $q^i(\mathbf{x}_k|\mathbf{y}_{1:k})$ which is then used to re-linearize the function \mathbf{h} , producing a new approximation $q^{i+1}(\mathbf{x}_k|\mathbf{y}_{1:k})$. This is then iterated until some convergence criterion is satisfied; typically until a fixed point is reached or a maximum number of iterations has been reached.

However, none of these algorithms, except [9], encompass the approximate density (15), even though this approximation directly affects the approximate marginal posterior as well. This is motivated by the fact that nonlinearities in the likelihood affect the posterior approximation to a greater extent than the prior. Nevertheless, standard iterated filters are for instance not useful in the case of a nonlinear transition function \mathbf{f} but linear measurement function \mathbf{h} , even though the linearization of \mathbf{f} also affects the quality of the approximate posterior. Next, a general linearization-based algorithm encompassing both the transition density as well as the measurement density approximations is developed.

IV. DYNAMICALLY ITERATED FILTER

To derive an algorithm encompassing both the transition density (15), as well as the observation density (18), at time k , we naturally need to seek an approximate posterior over both \mathbf{x}_{k-1} as well as \mathbf{x}_k . To do so, we generalize the derivation in [8] to extend backwards one step. Define two auxiliary variables, \mathbf{g}_k , \mathbf{g}_{k-1} as

$$\mathbf{g}_{k-1} = \mathbf{f}(\mathbf{x}_{k-1}) + \psi \quad (20a)$$

$$\mathbf{g}_k = \mathbf{h}(\mathbf{x}_k) + \phi \quad (20b)$$

$$p(\psi) = \mathcal{N}(\mathbf{0}, \alpha \mathbf{I}), \quad p(\phi) = \mathcal{N}(\mathbf{0}, \beta \mathbf{I}), \quad (20c)$$

where ψ and ϕ are independent of each other as well as the process noise \mathbf{w} and the measurement noise \mathbf{e} . Note that as $\alpha, \beta \rightarrow 0$, $\mathbf{g}_{k-1} \rightarrow \mathbf{f}(\mathbf{x}_{k-1})$ and $\mathbf{g}_k \rightarrow \mathbf{h}(\mathbf{x}_k)$. Now, the true joint posterior of \mathbf{x}_{k-1} , \mathbf{x}_k , \mathbf{g}_{k-1} and \mathbf{g}_k is given by

$$\begin{aligned} p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) &\propto \\ p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}). \end{aligned} \quad (21)$$

Following [8], we assume that the approximate posterior can be decomposed in the same manner, i.e.,

$$\begin{aligned} q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) &\approx \\ q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) q_\theta(\mathbf{g}_k | \mathbf{x}_k) q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}), \end{aligned} \quad (22)$$

where θ are the parameters of the affine approximation of the transition model and measurement model, i.e., $\theta = [\mathbf{A}_f, \mathbf{b}_f, \Omega_f, \mathbf{A}_h, \mathbf{b}_h, \Omega_h]$.

We now seek a θ such that $q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})$ is close to $p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})$, in some sense. Formally, the optimal parameters θ^* , and hence the optimal affine approximations of \mathbf{f} and \mathbf{h} , are found through

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\theta). \quad (23)$$

The loss $\mathcal{L}(\theta)$ is free to choose, but a natural choice of dissimilarity measure between distributions is the Kullback–Leibler (KL) divergence, which we pursue here. The KL divergence between the true joint posterior and the approximate joint posterior is given by

$$\begin{aligned} \text{KL}(p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})) &= \\ \text{KL}(p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})) &+ \\ \mathbb{E}[\text{KL}(p(\mathbf{g}_k | \mathbf{x}_k) || q_\theta(\mathbf{g}_k | \mathbf{x}_k))] &+ \\ \mathbb{E}[\text{KL}(p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) || q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}))] &\triangleq \mathcal{L}(\theta). \end{aligned} \quad (24)$$

See Appendix A for the derivation. Note that the expectations in $\mathcal{L}(\theta)$ are taken with respect to the true joint posterior $p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$. It is noteworthy that $\mathcal{L}(\theta)$ can be decomposed into three distinct terms, each dealing with each respective factor of (22). The first term is simply the KL divergence between the true and approximate joint posterior of the states at time k and $k-1$. The second and third terms are the expected KL divergences of the affine approximation of the measurement model and transition model, respectively, where the expectation is taken with respect to the true joint posterior $p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$.

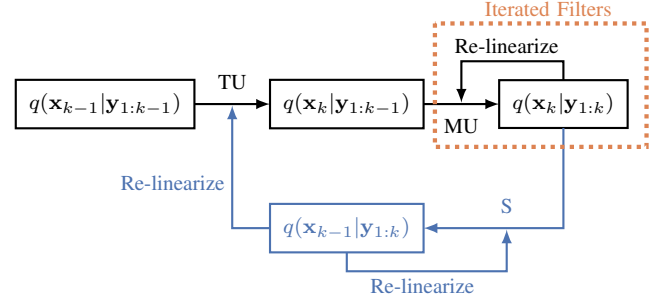


Fig. 1: Schematic illustration of a dynamically iterated filter. Ordinary iterated filters, marked in dotted orange, only re-linearize the measurement update. Dynamically iterated filters also re-linearize the time update through a smoothing step, marked in blue. The time update (TU) and the smoothing step (S) are linearized w.r.t. the smoothed distribution $q(\mathbf{x}_{k-1} | \mathbf{y}_{1:k})$. The measurement update (MU) is linearized w.r.t. the current posterior $q(\mathbf{x}_k | \mathbf{y}_{1:k})$. The steps are iterated until some convergence criterion is met.

It is impractical to minimize (24), seeing as the expectations are taken w.r.t. the true joint posterior $p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$. Nevertheless, an iterative procedure may be used to approximately solve this minimization problem.

A. Iterative Solution

To practically optimize (23), we assume access to an i :th approximation to the state joint posterior $p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) \approx q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$. We then use $q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$ in place of $p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$ in (24) and thus optimize an approximate loss, i.e., the approximate optimization problem is given by

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \text{KL}(q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta^{i+1}(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})) + \\ &\mathbb{E}_{q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_k | \mathbf{x}_k) || q_\theta^{i+1}(\mathbf{g}_k | \mathbf{x}_k))] + \\ &\mathbb{E}_{q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) || q_\theta^{i+1}(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}))], \end{aligned}$$

where the expectations are now over $q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$. Sufficiently close to a fixed point, the first KL term is approximately 0 and the final optimization problem is thus given by

$$\begin{aligned} \theta^* &= \arg \min_{\theta} \mathbb{E}_{q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} \left[\text{KL}(p(\mathbf{g}_k | \mathbf{x}_k) || q_\theta^{i+1}(\mathbf{g}_k | \mathbf{x}_k)) \right. \\ &\quad \left. + \text{KL}(p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) || q_\theta^{i+1}(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})) \right]. \end{aligned} \quad (25)$$

Technically, the optimal θ^* is given by statistical linearization of \mathbf{f} and \mathbf{h} w.r.t. the current approximation $q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$, see e.g., [8]. Note that statistical linearization of \mathbf{f} w.r.t. $q_\theta^i(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$ only requires the marginal $q_\theta^i(\mathbf{x}_{k-1} | \mathbf{y}_{1:k})$. Similarly, statistical linearization of \mathbf{h} only requires the marginal $q_\theta^i(\mathbf{x}_k | \mathbf{y}_{1:k})$. Thus, the algorithm conceptually amounts to predicting forward in time, performing a measurement update and smoothing backwards in time in order to provide new linearization points (densities) for both the transition density as well as the measurement density simultaneously. These steps are then iterated until fixed point convergence, finally providing an approximate posterior $q(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})$. The general algorithm is summarized in Algorithm 2 and schematically depicted in Fig. 1.

Algorithm 2 Dynamically iterated filter

Require: $q(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})$

Compute $\hat{\mathbf{x}}_{k|k-1}^0, \mathbf{P}_{k|k-1}^0$ by (2)

Compute $\hat{\mathbf{x}}_{k|k}^0, \mathbf{P}_{k|k}^0$ by (3)

Compute $\hat{\mathbf{x}}_{k-1|k}^0, \mathbf{P}_{k-1|k}^0$ by (4)

$i \leftarrow 0$

while not converged **do**

Calculate $(\mathbf{A}_f, \mathbf{b}_f, \mathbf{\Omega}_f)$ by analytical or statistical linearization of \mathbf{f} about $\hat{\mathbf{x}}_{k-1|k}^i, \mathbf{P}_{k-1|k}^i$

Compute $\hat{\mathbf{x}}_{k|k-1}^{i+1}, \mathbf{P}_{k|k-1}^{i+1}$ by (2)

Calculate $(\mathbf{A}_h, \mathbf{b}_h, \mathbf{\Omega}_h)$ by analytical or statistical linearization of \mathbf{h} about $\hat{\mathbf{x}}_{k|k}^i, \mathbf{P}_{k|k}^i$

Compute $\hat{\mathbf{x}}_{k|k}^{i+1}, \mathbf{P}_{k|k}^{i+1}$ by (3)

Compute $\hat{\mathbf{x}}_{k-1|k}^{i+1}, \mathbf{P}_{k-1|k}^{i+1}$ by (4)

$i \leftarrow i + 1$

end while

return $\hat{\mathbf{x}}_{k|k}^i, \mathbf{P}_{k|k}^i, \hat{\mathbf{x}}_{k-1|k}^i, \mathbf{P}_{k-1|k}^i$

Note that the algorithm is applicable to all possible combinations of models with linear and nonlinear \mathbf{f} and \mathbf{h} . Further, even though the developed solution is essentially an IPLF also encompassing the transition density, by changing the linearization method from statistical to analytical, an “extended” version is recovered in similar spirit to the IEKF. Furthermore, an IUKF version, similar to [5], may also be recovered by “freezing” the covariance matrices $\mathbf{P}_{k-1|k}^i = \mathbf{P}_{k-1|k-1}$ and $\mathbf{P}_{k|k}^i = \mathbf{P}_{k|k-1}$ and only updating these during the last iteration. It is also worthwhile to point out that the dynamically iterated filters are essentially “local” iterated smoothers, analogous to the *iterated extended Kalman smoother* (IEKS) [13] and the *iterated posterior linearization smoother* (IPLS) [10], operating on just one time instance and observation. Therefore, as noted in [9], a byproduct of the algorithm is a one-step smoothed state estimate and the method can thus be thought of as an iterated one-step fixed-lag smoother as well.

All that is left is to determine a stopping criterion for the iterations. Similarly to [8], a stopping criterion for the iterative updates may be formed on the basis of the KL divergence between two successive approximations of the posterior, i.e.,

$$\text{KL}(q^i(\mathbf{x}_k|\mathbf{y}_{1:k})||q^{i+1}(\mathbf{x}_k|\mathbf{y}_{1:k})) < \gamma.$$

Another possibility to check for fixed-point convergence is to instead use the smoothed density $q(\mathbf{x}_{k-1}|\mathbf{y}_{1:k})$ in a similar manner as the posterior. This is not investigated in detail here. Instead, in the numerical example in Section V, a fixed number of iterations are used for simplicity.

V. NUMERICAL EXAMPLES

To demonstrate the application of the dynamically iterated filters, we provide an illustrative example demonstrating the iterative procedure of the algorithm. We also provide a numerical example of maneuvering target tracking with a nonlinear transition model but a *linear* measurement model.

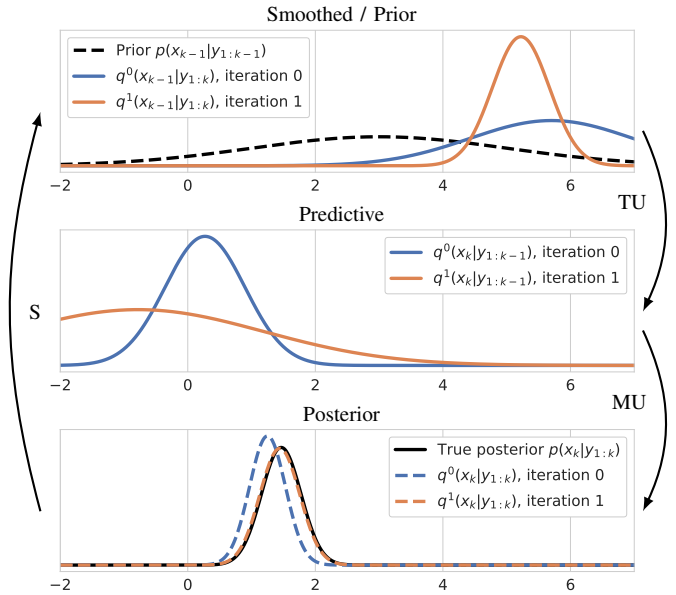


Fig. 2: Illustration of the (extended) dynamically iterated filter. The black curves in the top and bottom plots are the prior $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ and true posterior $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, respectively. The blue curves from top to bottom illustrate the approximate smoothed, predictive and posterior densities at iteration 0, respectively. The orange curves illustrate the same densities during the second iteration of the filter. The filter thus moves from the prior (top) to the predictive (middle) to the posterior (bottom) and back up to the smoothed (top). The time update, measurement update and smoothing step are indicated similarly to Fig. 1. Notice that iteration 0 exactly corresponds to an EKF.

A. Illustrative example

To illustrate the iterative procedure of the algorithm, we use an example similar to that in [12] but alter it to include a dynamical model. Therefore, let the model be given by

$$\begin{aligned} \mathbf{x}_{k+1} &\sim \mathcal{N}(\mathbf{x}_{k+1}; a\mathbf{x}_k^3, Q) \\ \mathbf{y}_k &\sim \mathcal{N}(\mathbf{y}_k; \mathbf{x}_k, R), \end{aligned}$$

with $a = 0.01, Q = 0.1$ and $R = 0.1$. We assume that a prior is given at time $k-1$ as $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_{k-1}; 3, 4)$. We then apply an analytically linearized version of the dynamically iterated filter to this model and plot the intermediary and final approximate predictive, posterior, and smoothed densities. The true posterior is found simply through evaluating the posterior density over a dense grid. The example is illustrated in Fig. 2, where two iterations are enough for the posterior approximation to be accurate.

B. Maneuvering Target Tracking

We consider a numerical example of maneuvering target tracking with a nonlinear transition model but a *linear* measurement model. This is a typically “easy” tracking scenario where standard filters generally do well.

Three versions of the dynamically iterated filters are evaluated, an extended version (DIEKF), an unscented version (DIUKF), and a posterior linearization version (DIPLF) based on unscented transform. These are compared to their respective non-iterated counterparts, i.e., the EKF and the UKF. For the unscented filters, we use the tuning parameters $\alpha =$

$\sqrt{3/n_x}, \kappa = \frac{n_x(3/2 - \alpha^2)}{\alpha^2}$ and $\beta = 2$, where n_x is the dimension of \mathbf{x} . This tuning corresponds to a weighting of 1/3 on the central sigma point.

We consider a target maneuvering in a plane and describe the target using the state vector $\mathbf{x}_k^\top = [p_k^x \ v_k^x \ p_k^y \ v_k^y \ \omega_k]$. Here, $p_k^x, p_k^y, v_k^x, v_k^y$ are the Cartesian coordinates and velocities of the target, respectively, and ω_k is the turn rate at time k . The transition model is thus given by

$$\mathbf{x}_{k+1} = \mathbf{F}(\omega_k)\mathbf{x}_k + \mathbf{w}_k, \quad (26)$$

where

$$\mathbf{F}(\omega_k) = \begin{bmatrix} 1 & \frac{\sin(T\omega_k)}{\omega_k} & 0 & -\frac{(1-\cos(T\omega_k))}{\omega_k} & 0 \\ 0 & \cos(T\omega_k) & 0 & -\sin(T\omega_k) & 0 \\ 0 & \frac{(1-\cos(T\omega_k))}{\omega_k} & 1 & \frac{\sin(T\omega_k)}{\omega_k} & 0 \\ 0 & \sin(T\omega_k) & 0 & \cos(T\omega_k) & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

T is the sampling period and $\mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q})$ is the process noise at time k , with

$$\mathbf{Q} = \begin{bmatrix} q_1 \frac{T^3}{3} & q_1 \frac{T^2}{2} & 0 & 0 & 0 \\ q_1 \frac{T^2}{2} & q_1 T & 0 & 0 & 0 \\ 0 & 0 & q_1 \frac{T^3}{3} & q_1 \frac{T^2}{2} & 0 \\ 0 & 0 & q_1 \frac{T^2}{2} & q_1 T & 0 \\ 0 & 0 & 0 & 0 & q_2 \end{bmatrix},$$

where q_1 and q_2 are parameters of the model.

In order to isolate the benefits of iterating over the time update, a linear positional measurement model is used, i.e.,

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \mathbf{e}_k, \quad (27)$$

with $\mathbf{H} = \text{diag}[1 \ 0 \ 1 \ 0 \ 0]$ and $\mathbf{e}_k \sim \mathcal{N}(\mathbf{e}_k; \mathbf{0}, \sigma^2 \mathbf{I})$.

The prior at time 0 is given by

$$p(\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}),$$

with $\hat{\mathbf{x}}_{0|0}^\top = [130 \ 35 \ -20 \ -20 \ -4\frac{\pi}{180}]$ and $\mathbf{P}_{0|0} = \text{diag}[\sigma_{p_x}^2 \ \sigma_{v_x}^2 \ \sigma_{p_y}^2 \ \sigma_{v_y}^2 \ \sigma_\omega^2]$, with $\sigma_{p_x}^2 = \sigma_{v_x}^2 = \sigma_{p_y}^2 = \sigma_{v_y}^2 = 5$ and $\sigma_\omega^2 = 10^{-2}$. The initial state for the ground truth trajectories are drawn from this prior.

We fix $q_2 = 10^{-2}, T = 1$ and sweep over all pairs of

$$q_1 = \{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$$

$$\sigma^2 = \{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\},$$

i.e., 25 different noise configurations. For each noise configuration, we simulate 10 individual targets along 20 different trajectories of length $K = 130$ time steps, for a total of 200 simulations per configuration. Note that the 20 trajectories are different for each noise configuration and that the 10 targets for each trajectory differ only in their measurement noise realization. However, the trajectories and measurement noise realizations are exactly the same for each algorithm. Five example trajectories along with one measurement sample from each trajectory for a specific noise configuration is depicted in Fig. 3.

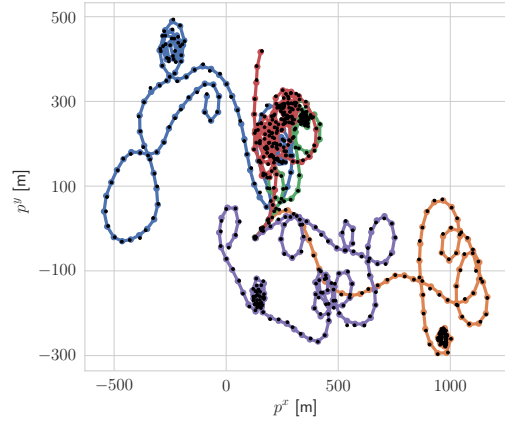


Fig. 3: Five example trajectories from one noise configuration of the considered tracking problem. Each trajectory is depicted as a separate color. The black smaller dots are a specific measurement realization along each trajectory.

To evaluate the performance of each dynamically iterated filter, we calculate the average position and velocity RMSE (separately) over the simulations for each of the filters and their corresponding baselines. We also compute a “relative” RMSE, relative the non-iterated counterpart, i.e.,

$$V = \frac{\text{RMSE}_{\text{iter}}}{\text{RMSE}_{\text{base}}}, \quad (28)$$

where clearly, $V \in [0, \infty]$ and lower is better. A relative score of 0.9 thus translates to a 10% lower RMSE as compared to the baseline. This yields a “quick glance” picture of the expected RMSE performance improvement in each particular noise configuration for each respective algorithm. For the DIEKF the non-iterated baseline is the EKF whereas for both the DIUKF and DIPLF the baseline is the UKF.

The results are presented as 5×5 matrices where each cell corresponds to a particular noise configuration for a particular pair of algorithms, e.g., the results for the DIEKF and EKF are summarized in one matrix. The results can be found in Fig. 4 where the position and velocity RMSEs are presented in Fig. 4(a) and Fig. 4(b), respectively. The leftmost matrix in each of the figures corresponds to the DIEKF and EKF. The middle matrix contains the results for the DIUKF and UKF and the rightmost matrix for the DIPLF and UKF. The top number in each cell is the RMSE for the dynamically iterated filter whereas the bottom number corresponds to the baseline. The color of each cell represents the RMSE of the dynamically iterated filter relative its baseline, according to (28). A deeper green color thus indicates a more substantial improvement than a lighter green. Lastly, an algorithm is considered to have diverged if its position RMSE is approximately larger than σ , where σ is the measurement noise standard deviation, as a position RMSE of σ can be expected by just using the raw measurements. Divergence is illustrated by a “-” in the corresponding cell in the matrices.

From Fig. 4(a), it is clear that even though all of the dynamically iterated filters improve upon their baselines, the

analytically linearized DIEKF benefits the most from the iterative procedure. Astonishingly, the EKF diverges for 22 out of 25 configurations whereas the DIEKF manages to lower that to 5 out of 25 and only diverges in the high noise scenario ($\sigma^2 = 10^2$). The performance increase in position RMSE is more modest for the DIUKF and DIPLF but still sees improvement, particularly for low process noise regimes. For the velocity RMSE in Fig. 4(b), the improvement for all of the three dynamically iterated filters is substantial. For low process noise regimes the improvement is up to 10-fold for the DIEKF and 5-fold for the DIUKF and DIPLF. Even for modest noise levels, the DIUKF and DIPLF roughly manage a 2-fold performance improvement. For the high noise scenario ($\sigma^2 = 10^2$), the DIUKF and DIPLF show a 10-fold performance improvement and bring the velocity RMSE down to reasonable levels where the RMSE for the UKF is very high.

VI. CONCLUSION

Dynamically iterated filters, a new class of iterated nonlinear filters, has been presented. The dynamically iterated filters, as opposed to previous iterated filters, are applicable to all possible combinations of (Gaussian) linear and nonlinear transition and measurement models. The filters were evaluated against their respective non-iterated baselines in a numerical example

$$\begin{aligned}
& \text{KL}(p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})) = \int p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) \log \frac{p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})}{q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})} d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} = \\
& \int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \log \frac{p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})}{q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) q_\theta(\mathbf{g}_k | \mathbf{x}_k) q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})} d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} = \\
& \int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \left[\log \frac{p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})}{q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} + \log \frac{p(\mathbf{g}_k | \mathbf{x}_k)}{q_\theta(\mathbf{g}_k | \mathbf{x}_k)} + \log \frac{p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})}{q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})} \right] d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} = \\
& \int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \log \frac{p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})}{q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} + \int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \log \frac{p(\mathbf{g}_k | \mathbf{x}_k)}{q_\theta(\mathbf{g}_k | \mathbf{x}_k)} d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} + \\
& \int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \log \frac{p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})}{q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})} d\mathbf{x}_{k-1:k} d\mathbf{g}_{k-1:k} = \\
& \underbrace{\int p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) \log \frac{p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})}{q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k})} d\mathbf{x}_{k-1:k}}_{\text{KL}(p(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta(\mathbf{x}_{k-1:k} | \mathbf{y}_{1:k}))} + \underbrace{\int p(\mathbf{x}_k | \mathbf{y}_{1:k}) p(\mathbf{g}_k | \mathbf{x}_k) \log \frac{p(\mathbf{g}_k | \mathbf{x}_k)}{q_\theta(\mathbf{g}_k | \mathbf{x}_k)} d\mathbf{x}_k d\mathbf{g}_k}_{\mathbb{E}_{p(\mathbf{x}_k | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_k | \mathbf{x}_k) || q_\theta(\mathbf{g}_k | \mathbf{x}_k))]} + \underbrace{\int p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k}) p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) \log \frac{p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})}{q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1})} d\mathbf{x}_{k-1} d\mathbf{g}_{k-1}}_{\mathbb{E}_{p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) || q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}))]} \triangleq \mathcal{L}(\theta). \\
& \text{KL}(p(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k}) || q_\theta(\mathbf{x}_{k-1:k}, \mathbf{g}_{k-1:k} | \mathbf{y}_{1:k})) + \mathbb{E}_{p(\mathbf{x}_k | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_k | \mathbf{x}_k) || q_\theta(\mathbf{g}_k | \mathbf{x}_k))] + \mathbb{E}_{p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k})} [\text{KL}(p(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}) || q_\theta(\mathbf{g}_{k-1} | \mathbf{x}_{k-1}))] \triangleq \mathcal{L}(\theta).
\end{aligned}$$

REFERENCES

- [1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
- [2] S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A New Approach for Filtering Nonlinear Systems," in *Proceedings of 1995 American Control Conference - ACC'95*, vol. 3. Seattle, WA, USA: American Autom Control Council, 1995, pp. 1628–1632.
- [3] 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" [with authors' reply]," *IEEE Trans. Automat. Contr.*, vol. 47, no. 8, pp. 1406–1409, Aug. 2002.
- [4] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [5] M. A. Skoglund, F. Gustafsson, and G. Hendeby, "On Iterative Unscented Kalman Filter using Optimization," in *2019 22th International Conference on Information Fusion (FUSION)*, Ottawa, ON, Canada, Jul. 2019, pp. 1–8.
- [6] R. Zhan and J. Wan, "Iterated Unscented Kalman Filter for Passive Target Tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 1155–1163, Jul. 2007.
- [7] G. Sibley, G. Sukhatme, and L. Matthies, "The Iterated Sigma Point Kalman Filter with Applications to Long Range Stereo," in *Robotics: Science and Systems II*. Robotics: Science and Systems Foundation, Aug. 2006.

with a nonlinear transition model and a linear measurement model. Even in this "simple" case, where standard filters typically perform well, the dynamically iterated filters had improved RMSE performance, especially for non-measurable states. Further, even though the EKF diverged in 22 out of 25 configurations considered, the dynamically iterated EKF was empirically shown to be stable for 20 out of 25 noise configurations, only diverging for high noise ($\sigma^2 = 10^2$).

Future work includes more extensive testing on other models as well as determining in what particular scenarios the statistically linearized versions perform better than the analytically and vice versa.

VII. ACKNOWLEDGMENTS

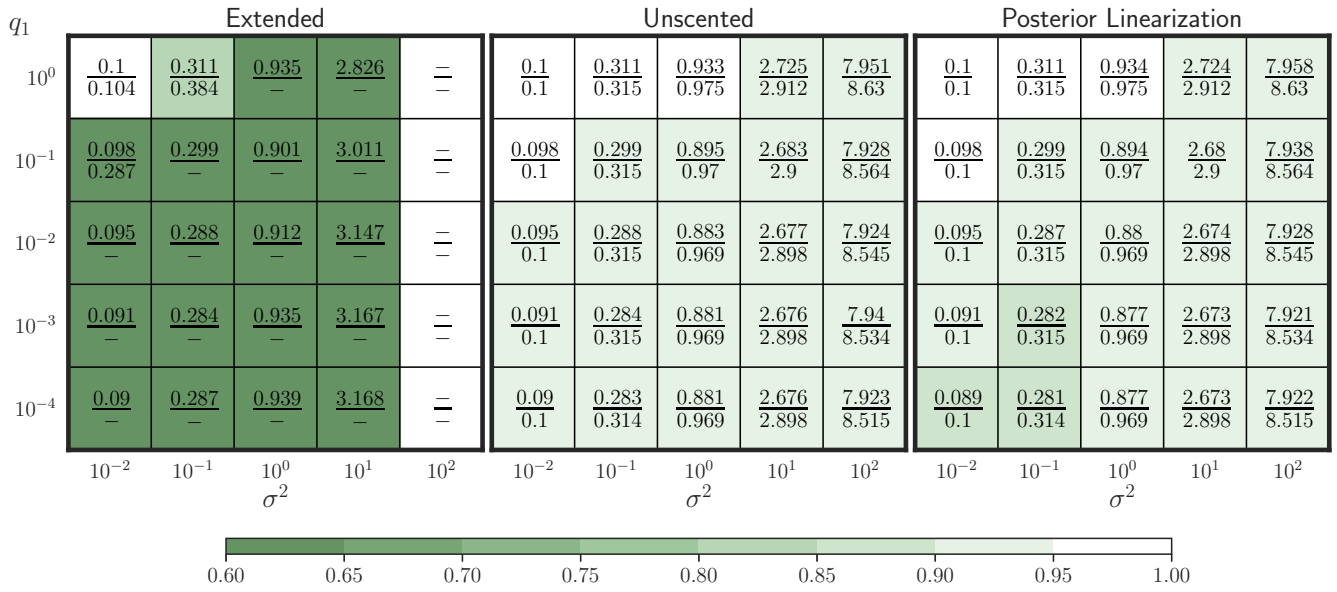
The authors would like to extend sincere gratitude to Martin Skoglund for excellent tips for the experimental evaluation.

APPENDIX A LOSS DERIVATION

The KL divergence between the true joint (21) and the approximate (22) is given by

- [8] Á. F. García-Fernández, L. Svensson, M. R. Morelande, and S. Särkkä, "Posterior Linearization Filter: Principles and Implementation Using Sigma Points," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5561–5573, Oct. 2015.
- [9] M. Raitoharju, R. Hostettler, and S. Sarkka, "Posterior Linearisation Filter for Non-linear State Transformation Noises," in *2022 25th International Conference on Information Fusion (FUSION)*. Linköping, Sweden: IEEE, Jul. 2022, pp. 1–6.
- [10] Á. F. García-Fernández, L. Svensson, and S. Särkkä, "Iterated Posterior Linearization Smoother," *IEEE Trans. Automat. Contr.*, vol. 62, no. 4, pp. 2056–2063, Apr. 2017.
- [11] S. Särkkä, *Bayesian Filtering and Smoothing*, 1st ed. Cambridge University Press, Sep. 2013.
- [12] Á. F. García-Fernández, L. Svensson, and M. R. Morelande, "Iterated Statistical Linear Regression for Bayesian Updates," in *17th International Conference on Information Fusion (FUSION)*, Salamanca, Spain, Jul. 2014, pp. 1–8.
- [13] B. M. Bell, "The Iterated Kalman Smoother as a Gauss–Newton Method," *SIAM J. Optim.*, vol. 4, no. 3, pp. 626–636, Aug. 1994.

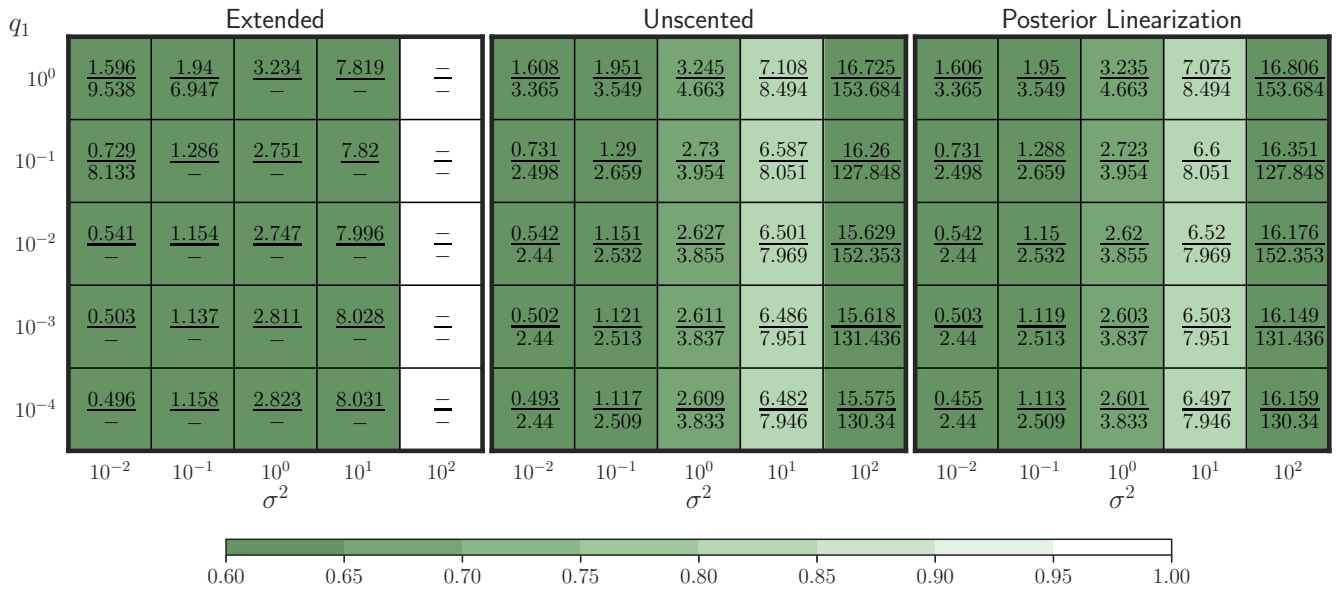
Position RMSE [m]



Relative RMSE Iterated/Baseline

(a) Position RMSE. Clearly, the analytically linearized DIEKF improves the most over its baseline the EKF which diverges in 22/25 configurations. The DIUKF (middle) and DIPLF (right) share similar performance where the DIPLF is slightly better for some configurations.

Velocity RMSE [m/s]



Relative RMSE Iterated/Baseline

(b) Velocity RMSE. Clearly, all three dynamically iterated filters improve substantially over their corresponding baselines. For low noise regimes, up to 10-fold improvements are seen for the DIEKF whereas the DIUKF and DIPLF have approximately a 5-fold improvement. For high noise regimes the results for the DIUKF and DIPLF are even better with approximately 10-fold improvements.

Fig. 4: RMSE for the dynamically iterated filters as compared to their respective baselines for 25 noise configurations, where each cell corresponds to one noise configuration given by the figure axes. The top number in each cell is the RMSE for the dynamically iterated filter whereas the bottom number is the RMSE for the corresponding baseline. A “-” indicates that the positional RMSE of the filter is larger than σ and it is thus considered to have diverged, since an RMSE of σ corresponds to just using the raw measurements. The left plot in (a) and (b) shows the RMSE for the DIEKF and EKF. The middle figure shows the DIUKF and UKF and finally, the right most figure shows the DIPLF and UKF. Each cell is colored according to the RMSE of the dynamically iterated filter relative to the baseline. A relative RMSE of 0.9 corresponds to a 10% reduction of the RMSE.