Aalto University

Tronarp, Filip; Garcia Fernandez, Angel; Särkkä, Simo

Iterative Filtering and Smoothing In Non-Linear and Non-Gaussian Systems Using Conditional Moments

Iterative Filtering and Smoothing In Non-Linear and Non-Gaussian Systems Using Conditional Moments

Filip Tronarp, Ángel F. García-Fernández, and Simo Särkkä

DOI:

Copyright:

Iterative Filtering and Smoothing In Non-Linear and Non-Gaussian Systems Using Conditional Moments

Filip Tronarp, Ángel F. García-Fernández, and Simo Särkkä, *Senior Member, IEEE*

Abstract—This letter presents the development of novel iterated filters and smoothers that only require specification of the conditional moments of the dynamic and measurement models. This leads to generalisations of the iterated extended Kalman filter, the iterated extended Kalman smoother, the iterated posterior linearisation filter, and the iterated posterior linearisation smoother. The connections to the previous algorithms are clarified and a convergence analysis is provided. Furthermore, the merits of the proposed algorithms are demonstrated in simulations of the stochastic Ricker map where they are shown to have similar or superior performance to competing algorithms.

Index Terms—State estimation, non-linear/non-Gaussian systems, statistical linear regression, iterative methods.

I. INTRODUCTION

STATE estimation is a frequently occurring problem in science and engineering. Its applications include tracking, audio signal processing, and time series modelling [1]–[4]. Formally, the system is described by a partially observed Markov process (POMP) for which inference is conducted, in the Bayesian sense, by computing conditional distributions given the measurements resulting in either *filtering* or *smoothing* distributions depending on what measurements are included in the conditioning (see [2]). In linear Gaussian models, these problems are efficiently solved by the Kalman filter (KF) and the Rauch-Tung-Striebel (RTS) smoother [5], [6]. However, exact inference is intractable in general.

In the case of a non-linear system excited by Gaussian white noise, an early approach was to make local affine approximations by means of Taylor series which results in *extended Kalman filter/smoother* (EKF/EKS) [1], [2]. This approach belongs to the *assumed density* framework [7], where a Gaussian assumption is enforced on the state marginals. More recently, other methods for *assumed Gaussian* estimation, based on numerical integration, were suggested in [8], [9] (see [10] for a thorough discussion) which can be seen as another approach to linearising the system called *statistical linear regression* (SLR) [11] (see also [12]). More general Gaussian filters can be obtained if the conditional mean and variance of the dynamic and measurement models are tractable [13], [14].

Although the aforementioned approaches are often adequate, their performance can deteriorate severely when the system

F. Tronarp and S. Särkkä are with the Department of Electrical Engineering and Automation, Aalto University, 02150 Espoo, Finland (email: filip.tronarp@aalto.fi). Á. F. García-Fernández is with the Department of Electrical Engineering and Electronics, University of Liverpool L69 3GJ, United Kingdom. This work was partly supported by Academy of Finland grants #295080 and #266940.

description departs too much from affine and Gaussian. In such a case, sequential Monte Carlo offers an arbitrarily low approximation error, though at an arbitrarily large computational cost [15]–[17]. This prompts investigations into lightweight algorithms for approximate and accurate inference. Examples that improve the update step are the recursive update filter (RUF) [18] and the progressive Gaussian filters (PGF42/PGFL) [19], [20]. However, RUF only operates under the additive Gaussian noise assumption, PGF42 only operates under the Gaussian excitation assumption, and the PGFL requires a tractable likelihood, which is not always available. Moreover, neither of the aforementioned methods offers improvement to the smoothing recursion. On the other hand, Expectation propagation [21] targets the smoothing solution. However, it also requires a tractable likelihood and, in general, the required integrals cannot be approximated with lightweight methods.

Another approach, which can improve filtering and smoothing, though only applicable to the additive Gaussian noise case, is the Gauss-Newton method resulting in the iterated extended Kalman filter/smoother (IEKF/IEKS) [22], [23]; this enables the use of traditional optimisation methods (see [24] for an overview). This approach was recently extended on the basis that it ought to be better to perform the SLR with respect to the posterior distribution rather than the prior distribution. While this is intractable, it leads to a scheme where the SLR is iteratively computed using the current best approximation to the posterior, why it is called *iterated posterior linearisation filtering/smoothing* (IPLF/IPLS) [25], [26]. However, these methods assume additive Gaussian noise.

In this letter, new iterative filters and smoothers are developed by exploiting tractable conditional moments in the dynamic and measurement models. The developed algorithms offer accurate state estimation in a wide class of non-linear/non-Gaussian models at a low computational cost and IEKF/IEKS and IPLF/IPLS come out as special cases. The algorithms are verified on the stochastic Ricker map [27], where iterations turn out to be essential.

II. PROBLEM FORMULATION

Herein the problem of estimating the state in a POMP is considered. More specifically, consider the latent Markov process $\{X_t \in \mathbb{R}^{D_X}\}_{t=0}^{T}$ for which a series of imperfect and noisy measurements are available, $\{Y_t \in \mathbb{Y}\}_{t=1}^{T}$, where $\mathbb{Y} \subset \mathbb{R}^{D_Y}$. The system is then specified as follows:

$$X_t \mid X_{t-1} = x \sim f_{X_t \mid X_{t-1}}(x_t \mid x), \tag{1a}$$

$$Y_t \mid X_t = x \sim f_{Y_t \mid X_t}(y_t \mid x), \tag{1b}$$

where $X_0 \sim f_{X_0}(x_0)$, $\sim$ denotes drawn from or distributed as, $Y \mid X = x$ the random variable Y conditioned on $X = x$, and $f_{Y \mid X}(y \mid x)$ its probability density function (pdf)[1]. The subscripts of the pdf f shall be omitted unless needed for clarity. Furthermore, let $\mathbb{E}[X]$, $\mathbb{C}[X, Y]$, and $\mathbb{V}[X]$ denote the expected value of X, the cross-covariance matrix of X and Y, and the covariance matrix of X, respectively.

In state estimation, the goal is to obtain a series of densities $f_{X_t \mid Y_{1:\tau}}(x \mid y_{1:\tau}) := f(x_t \mid y_{1:\tau})$; where $t = \tau$ and $\tau > t$ correspond to filtering and smoothing, respectively [1], [2]. As exact inference is generally intractable for the system in Eq. (1), the filtering and smoothing densities are approximated by Gaussians. Previous approaches are reviewed in Section II-A and the present contributions are summarised in Section II-B.

A. Prior Work

Typically, the densities in Eq. (1) are implicitly defined by a transformation of Gaussian variables according to

$$X_t = a(X_{t-1}, W_t), \tag{2a}$$
$$Y_t = c(X_t, V_t), \tag{2b}$$

where $W_t \sim \mathcal{N}(0, \Sigma_{W_t})$ and $V_t \sim \mathcal{N}(0, \Sigma_{V_t})$ are mutually independent white noise sequences. When a and c are non-affine, a typical strategy is to make an implicit linear Gaussian approximations according to [11]

$$X_t \approx A_t X_{t-1} + b_t + Q_t, \tag{3a}$$
$$Y_t \approx C_t X_t + d_t + R_t, \tag{3b}$$

where $A_t \in \mathbb{R}^{D_X \times D_X}$, $b_t \in \mathbb{R}^{D_X}$, $C_t \in \mathbb{R}^{D_Y \times D_X}$, $d_t \in \mathbb{R}^{D_Y}$ and Q_t, R_t are white noise sequences with covariance matrices Σ_{Q_t} and Σ_{R_t}, respectively. Once the approximation in Eq. (3) has been made, filtering and smoothing can be done by the linear methods [5], [6]. If the parameters are chosen by Taylor series linearisation around the filtered/predicted mean then EKF/EKS are retrieved [1], [2] and sigma-point approximations of the SLR with respect to the filtering/predictive distribution [11] results in the sigma-point filters/smoothers [8], [9]. Hence, a series of linearisations of a and c are constructed based on the sequence of approximate filtering and predictive densities, $\{f(x_t \mid y_{1:t})\}_{t=1}^T$, $\{f(x_t \mid y_{1:t-1})\}_{t=1}^T$. This results in an approximation of Eq. (2) by a linear time-varying system of the form given in Eq. (3) for which inference can be carried out based on linear estimation theory [5], [6]. The IEKF/IEKS iteratively linearise the system by Taylor series around the current mean of the approximate filtering/smoothing distribution [22], [23]. On the other hand, IPLF/IPLS iteratively linearise the system using SLR with respect to the current approximate filtering/smoothing density [25], [26]. The RTS smoother is given in Alg. 1 for future reference.

B. The Contribution

In this letter, linear methods for estimation in Eq. (1) are developed under the assumptions that (i) $\mathbb{E}[X_t \mid X_{t-1}]$, $\mathbb{E}[Y_t \mid X_t]$, $\mathbb{V}[X_t \mid X_{t-1}]$, and $\mathbb{V}[Y_t \mid X_t]$ are tractable and (ii)

[1]Here pdf is used for both discrete and continuous random variables.

Algorithm 1 Affine (RTS) Smoother

Input: prior moments $\mu_{0|0}, \Sigma_{0|0}$ and measurement sequence $\{y_t\}_{t=1}^T$
Output: smoothing means $\{\mu_{t|T}\}_{t=0}^T$ and covariances $\{\Sigma_{t|T}\}_{t=0}^T$
  for $t = 1$ **to** $t = T$ {Forward pass}
    $\mu_{t|t-1} \leftarrow A_t \mu_{t-1|t-1} + b_t$ {Prediction}
    $\Sigma_{t|t-1} \leftarrow A_t \Sigma_{t-1|t-1} A_t^\mathsf{T} + \Sigma_{Q_t}$
    $G_t \leftarrow \Sigma_{t-1|t-1} A_t^\mathsf{T} \Sigma_{t|t-1}^{-1}$ {Used in backward pass}
    $\Sigma_{Y_t|t-1} \leftarrow C_t \Sigma_{t|t-1} C_t^\mathsf{T} + \Sigma_{R_t}$ {Update}
    $K_t \leftarrow \Sigma_{t|t-1} C_t^\mathsf{T} \Sigma_{Y_t|t-1}^{-1}$
    $\mu_{t|t} \leftarrow \mu_{t|t-1} + K_t(y_t - C_t \mu_{t|t-1} - d_t)$
    $\Sigma_{t|t} \leftarrow \Sigma_{t|t-1} - K_t \Sigma_{Y_t|t-1} K_t^\mathsf{T}$
  end
  for $t = T-1$ **to** $t = 0$ {Backward pass}
    $\mu_{t|T} \leftarrow \mu_{t|t} + G_t(\mu_{t+1|T} - \mu_{t+1|t})$
    $\Sigma_{t|T} \leftarrow \Sigma_{t|t} + G_t(\Sigma_{t+1|T} - \Sigma_{t+1|t})G_t^\mathsf{T}$
  end

$\mathbb{C}[X_t, X_{t-1}] \neq 0$, $\mathbb{C}[Y_t, X_t] \neq 0$, $\forall t$. This leads to extensions of IEKF/IEKS and IPLF/IPLS since if $Y_t = c(X_t, V_t)$ then

$$\mathbb{E}[Y_t \mid X_t] = \mathbb{E}[c(X_t, V_t) \mid X_t] \tag{4a}$$
$$\mathbb{V}[Y_t \mid X_t] = \mathbb{E}[c(X_t, V_t)c(X_t, V_t)^\mathsf{T} \mid X_t]$$
$$- \mathbb{E}[c(X_t, V_t) \mid X_t]\mathbb{E}[c(X_t, V_t) \mid X_t]^\mathsf{T}. \tag{4b}$$

These expectations are readily approximated by sigma-point integration over V_t. More importantly, the present development may be applied to systems that are not explicitly generated by Gaussian excitations; consider, for example, Poisson measurements, $Y_t \mid X_t = x \sim \mathrm{Po}(c \exp(x))$. Then

$$\mathbb{E}[Y_t \mid X_t] = \mathbb{V}[Y_t \mid X_t] = c \exp(X_t), \tag{5}$$

which does not fit in the frameworks [1]–[9] that assume Y_t of the form in Eq. (2), which would require an inversion of the cumulative probability function, lacking closed form.

III. GENERALISED STATISTICAL LINEAR REGRESSION

Previous presentations of SLR have considered linearisation of deterministic functions with respect to the distribution of its random inputs [11], [12], [28]. Here, a more general presentation is required. Suppose a linear relationship is sought between the random variables X and Y, $Y = CX + d + R$, with C and d being linearisation parameters and R is a random variable accounting for the error. Choosing $(C, d) = \arg\min \mathbb{E}[\|R\|^2]$ gives the SLR formulae that can be computed only using the conditional moments. This can be seen by using laws of total expectation/covariance and the principle of orthogonality (c.f [29]). This results in Thm. 1 below.

Theorem 1. *Let X and Y be random variables with finite variance. The affine function, $CX + d$, that minimises the mean square norm of the residual, $R = Y - CX - d$, and the resulting moments of R are given by*

$$C = \mathbb{C}[Y, X]\mathbb{V}[X]^{-1}, \ d = \mathbb{E}[Y] - C\mathbb{E}[X],$$
$$\mathbb{E}[R] = 0, \ \Sigma_R = \mathbb{V}[Y] - C\mathbb{V}[X]C^\mathsf{T}. \tag{6}$$

Furthermore, the parameters C, d, and Σ_R can be computed using the conditional moments, as follows

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]], \tag{7a}$$
$$\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y \mid X]] + \mathbb{V}[\mathbb{E}[Y \mid X]], \tag{7b}$$
$$\mathbb{C}[Y, X] = \mathbb{C}[\mathbb{E}[Y \mid X], X]. \tag{7c}$$

The consequence of Thm. 1 is that the POMP in Eq. (1) can be linearised in a mean square optimal manner by just computing expectations of the moments of $Y \mid X$, which are possibly non-linear mappings of X, for example, see Eq. (5). This means that applying Thm. 1 is in general intractable. Fortunately, the usual approximation methods, Taylor series and sigma-points, can still be used. The latter approximation is obtained by expanding $\mathbb{E}[Y \mid X]$ and $\mathbb{V}[Y \mid X]$ around $\hat{\mu}$ according to

$$\mathbb{E}[Y \mid X] \approx \mathbb{E}[Y \mid \hat{\mu}] + (\nabla_X \mathbb{E}[Y \mid X])(\hat{\mu})(X - \hat{\mu}), \quad (8a)$$

$$\mathbb{V}[Y \mid X]_{i,j} \approx \mathbb{V}[Y \mid \hat{\mu}]_{i,j}, \quad (8b)$$

where ∇_X denotes the Jacobian operator, and then substituting Eq. (8) into Eq. (7). On the other hand, the sigma-point approach is to produce a set of points $\{\mathcal{X}^{(n)}\}_{n=1}^{N}$ with corresponding weights $\{w^{(n)}\}_{n=1}^{N}$ and then for any function, $\gamma(X)$, approximate its expectation, $\mathbb{E}[\gamma(X)]$, according to (see [2], [8], [10])

$$\mathbb{E}[\gamma(X)] \approx \sum_{n=1}^{N} w^{(n)} \gamma(\mathcal{X}^{(n)}).$$

Therefore, with appropriate choices for γ (see Eq. (7)), the joint moments of X and Y can be approximated.

IV. ITERATIVE FILTERS AND SMOOTHERS BASED ON CONDITIONAL MOMENTS

On the basis of re-linearising the system around the current posterior approximation, this section presents the development of novel iterative filters and smoothers by repeated applications of Thm. 1. The filter is presented in Section IV-A, where emphasis is on the update as it is the only stage of the algorithm that utilises iterations [25]. The iterative smoother is presented in Section IV-B and a convergence theorem is provided in Section IV-C. In Section IV-D, connections between the present algorithms and IEKF/IEKS and IPLF/IPLS are explored.

A. Filter

Suppose a Gaussian approximation to the filtering pdf is available, $f(x_{t-1} \mid y_{1:t-1}) \approx \mathcal{N}(x_{t-1}; \mu_{t-1|t-1}, \Sigma_{t-1|t-1})$. In order to form an approximation to the predictive pdf, $f(x_t \mid y_{1:t-1})$, Thm. 1 is used to compute the first two moments $\mathbb{E}[X_t] = \mathbb{E}[\mathbb{E}[X_t \mid X_{t-1}]]$ and $\mathbb{V}[X_t] = \mathbb{E}[\mathbb{V}[X_t \mid X_{t-1}]] + \mathbb{V}[\mathbb{E}[X_t \mid X_{t-1}]]$, the outer expectations being taken with respect to $f(x_{t-1} \mid y_{1:t-1})$ and the approximations described in Section III are used if necessary. In the update, the linearisation parameters, C_t, d_t, and Σ_{R_t} (see Eq. (3)) can be acquired by using Thm. 1. The approximate posterior moments of X_t are then computed with the usual linear estimator (see Update step of Alg. 1).

The key insight of [25] was that posterior moment approximations can be improved by re-computing the C_t, d_t and Σ_{R_t} using the current approximate posterior together with SLR, while keeping the original predictive moments in the Update step of Alg. 1. This yields a family of approximate posteriors $\{f_{X|Y}^{(j)}(x)\}_{j=0}^{J}$ with $f_{X|Y}^{(0)} := f_X(x)$, see Alg. 2. An illustrative example of the update is provided in the supplement.

Algorithm 2 Conditional Moments Based Iterative Update

Input: prior moments μ_X, Σ_X, evaluators of $\mathbb{E}[Y \mid X]$, $\mathbb{V}[Y \mid X]$, measurement y, and maximum number of iterations J.
Output: posterior moment approximations $\mu_{X|Y}^{(J)}$, $\Sigma_{X|Y}^{(J)}$.

Set $\mu_{X|Y}^{(0)} \leftarrow \mu_X$ and $\Sigma_{X|Y}^{(0)} \leftarrow \Sigma_X$
for $j = 1$ **to** $j = J$
    compute $C^{(j)}$, $d^{(j)}$, $\Sigma_R^{(j)}$ using $f_{X|Y}^{(j-1)}$ and Thm. 1.
    $K^{(j)} \leftarrow \Sigma_X (C^{(j)})^\mathsf{T} (C^{(j)} \Sigma_X (C^{(j)})^\mathsf{T} + \Sigma_R^{(j)})^{-1}$
    $\mu_{X|Y}^{(j)} \leftarrow \mu_X + K^{(j)}(y - C^{(j)} \mu_X - d^{(j)})$
    $\Sigma_{X|Y}^{(j)} \leftarrow \Sigma_X - K^{(j)} (C^{(j)} \Sigma_X (C^{(j)})^\mathsf{T} + \Sigma_R^{(j)})(K^{(j)})^\mathsf{T}$
end

Algorithm 3 Conditional Moments Based Iterative Smoother

Input: prior moments $\mu_{0|0}, \Sigma_{0|0}$, evaluators of $\mathbb{E}[X_t \mid X_{t-1}]$, $\mathbb{V}[X_t \mid X_{t-1}]$, $\mathbb{E}[Y_t \mid X_t]$ and $\mathbb{V}[Y_t \mid X_t]$, measurement sequence $\{y_t\}_{t=1}^{T}$, and maximum number of iterations J.
Output: posterior moment approximations $(\mu_{0:T|T}^{(J)}, \Sigma_{0:T|T}^{(J)})$.

Compute the smoothing moments $(\mu_{0:T|T}^{(0)}, \Sigma_{0:T|T}^{(0)})$ by using Thm. 1 to linearise with respect to the filtering distribution and Alg. 1.
Compute $\Theta^{(0)} = (A_{1:T}^{(0)}, b_{1:T}^{(0)}, \Sigma_{Q_{1:T}}^{(0)}, C_{1:T}^{(0)}, d_{1:T}^{(0)}, \Sigma_{R_{1:T}}^{(0)})$, using $(\mu_{t|T}^{(0)}, \Sigma_{t|T}^{(0)})$, and Thm. 1 to compute $(A_{t+1}^{(0)}, b_{t+1}^{(0)}, \Sigma_{Q_{t+1}}^{(0)})$ and $(C_t^{(0)}, d_t^{(0)}, \Sigma_{R_t}^{(0)})$.
for $j = 1$ **to** $j = J$
    Compute the smoothing moments $(\mu_{0:T|T}^{(j)}, \Sigma_{0:T|T}^{(j)})$ using $\Theta^{(j-1)}$ and Alg. 1.
    Compute $\Theta^{(j)} = (A_{1:T}^{(j)}, b_{1:T}^{(j)}, \Sigma_{Q_{1:T}}^{(j)}, C_{1:T}^{(j)}, d_{1:T}^{(j)}, \Sigma_{R_{1:T}}^{(j)})$, using $(\mu_{t|T}^{(j)}, \Sigma_{t|T}^{(j)})$, and Thm. 1 to compute $(A_{t+1}^{(j)}, b_{t+1}^{(j)}, \Sigma_{Q_{t+1}}^{(j)})$ and $(C_t^{(j)}, d_t^{(j)}, \Sigma_{R_t}^{(j)})$.
end

B. Smoother

Building on the previous section, an iterative fixed interval smoother can be obtained by starting from a collection of smoothing marginals, $\{f^{(0)}(x_t \mid y_{1:T})\}_{t=1}^{T}$, obtained by for instance the usual sigma-point smoother [2] and perhaps using Alg. 2 in the filter update. Now, let $(A_t^{(0)}, b_t^{(0)}, \Sigma_{Q_t}^{(0)})$ be the parameters of the linearisation $X_t = A_t^{(0)} X_{t-1} + b_t^{(0)} + Q_t^{(0)}$ and $(C_t^{(0)}, d_t^{(0)}, \Sigma_{R_t}^{(0)})$ be the linearisation $Y_t = C_t^{(0)} X_t + d_t^{(0)} + R_t^{(0)}$ using the smoothing densities $f^{(0)}(x_{t-1} \mid y_{1:T})$ and $f^{(0)}(x_t \mid y_{1:T})$ together with Thm. 1. Denote the collection of linearisation parameters after the first smoother pass as $\Theta^{(0)} = (A_{1:T}^{(0)}, b_{1:T}^{(0)}, \Sigma_{Q_{1:T}}^{(0)}, C_{1:T}^{(0)}, d_{1:T}^{(0)}, \Sigma_{R_{1:T}}^{(0)})$. Now, by the same rationale as for the development of Alg. 2, the smoothing solution can be iteratively improved by alternating between running Alg. 1 and re-linearising using the newly obtained smoothing marginals. This yields a family of smoothing marginals $\{f^{(j)}(x_t \mid y_{1:T})\}_{t=1,j=0}^{T,J}$ and a series of affine systems $\{\Theta^{(j)}\}_{j=1}^{J}$, approximating the POMP in Eq. (1) as

$$X_{t+1} \approx A_{t+1}^{(j)} X_t + b_{t+1}^{(j)} + Q_{t+1}^{(j)}, \quad \mathbb{V}[Q_{t+1}^{(j)}] = \Sigma_{Q_{t+1}}^{(j)}, \quad (9a)$$

$$Y_t \approx C_t^{(j)} X_t + d_t^{(j)} + R_t^{(j)}, \quad \mathbb{V}[R_t^{(j)}] = \Sigma_{R_t}^{(j)}, \quad (9b)$$

whereby the marginals $\{f^{(j+1)}(x_t \mid y_{1:T})\}_{t=1}^{T}$ are retrieved by running Alg. 1 on the system in Eq. (9). The procedure is outlined in Alg. 3.

C. Convergence Analysis

In a similar manner to [25], [26] a local convergence analysis can be carried out, resulting in Thm. 2.

Theorem 2. *Alg. 3 converges if it is initialised sufficiently close to a fixed point and the matrix Ξ, as defined in Eq. (11) in the supplement, has a spectral radius less than unity.*

Proof. See supplementary material. □

D. Discussion

Further clarification on the connections between Algs. 2 and 3 and the existing iterative estimators [22], [23], [25], [26] is given as follows. When $X_{0:T}$ and $Y_{1:T}$ are governed by the system, $X_t = a(X_{t-1}) + W_t$, $Y_t = c(X_t) + V_t$, with W_t and V_t mutually independent white noise sequences, then the conditional moments are given by $\mathbb{E}[X_t \mid X_{t-1}] = a(X_{t-1})$, $\mathbb{V}[X_t \mid X_{t-1}] = \Sigma_Q$, $\mathbb{E}[Y_t \mid X_t] = c(X_t)$ and $\mathbb{V}[Y_t \mid X_t] = \Sigma_R$. The Taylor series and sigma-point implementations of Algs. 2 and 3 then correspond to IEKF/IEKS and IPLF/IPLS, respectively.

Now, let a and c admit Taylor series expansions up to first order at any point in $\mathbb{R}^{D_X} \times \mathbb{R}^{D_W}$ and $\mathbb{R}^{D_X} \times \mathbb{R}^{D_V}$, respectively. Furthermore, let the sequences $X_{0:T}$ and $Y_{1:T}$ be governed by Eq. (2), then the conditional moments are given by Eq. (4). If the conditional moments are approximated by replacing a and c in Eq. (4) by their first order Taylor series in W_t and V_t around 0, respectively, then one iteration of Algs. 2 and 3 corresponds to the non-additive EKF/EKS [2, Algorithm 5.5] and several iterations correspond to the extensions of IEKF/IEKS to the non-additive case. Furthermore, when sigma-points are used, augmenting the noises to the state (see [2, Algorithm 5.13]), generalisations of the IPLF/IPLS for non-additive noise are obtained. Supporting calculations are given in the supplement. Note that it is not trivial to implement a particle filter for this case since the likelihood is intractable in general.

Finally, if the system is not described in terms of non-linear transforms of Gaussian variables, a new class of iterative algorithms emerges. As mentioned, for additive Gaussian noise the present algorithm reduces to previous iterative algorithms, hence it has the same computational complexity. In general, only one and $2N + 1$ additional evaluations of $\mathbb{V}[Y_t \mid X_t]$ and $\mathbb{V}[X_t \mid X_{t-1}]$ are required per iteration for the Taylor series and sigma-point implementations, respectively.

V. EXPERIMENTAL RESULTS

The proposed algorithms are evaluated on the stochastic Ricker map, which is a population model, given by

$$X_{t+1} = \log(44.7) + X_t - \exp(X_t) + W_{t+1}, \tag{10a}$$

$$Y_t \mid X_t = x \sim \text{Po}(10\exp(x)), \ X_0 \sim \delta(x_0 - \log 7) \tag{10b}$$

where X_t is the log-population and $W_t \sim \mathcal{N}(0, 0.3^2)$ is environmental noise [27]. There are several factors which make the state space model in Eq. (10) challenging for Kalman filters; (i) the measurements are made in a discrete space, (ii) it is very non-linear. Note that the conventional EKF/UKF cannot deal with the measurement model in Eq. (10) [2].

A Monte Carlo experiment was carried out by simulating the system in Eq. (10) 250 times with trajectories of length $T = 1 + 2^7$. Algs. 2 and 3 are evaluated using the Taylor series

Table I
PERCENTILES OF THE RMSE FOR TS, SP, PGFL, PF, AND FPFBS . LEFT COLUMN: FILTERS, RIGHT COLUMN: SMOOTHERS.

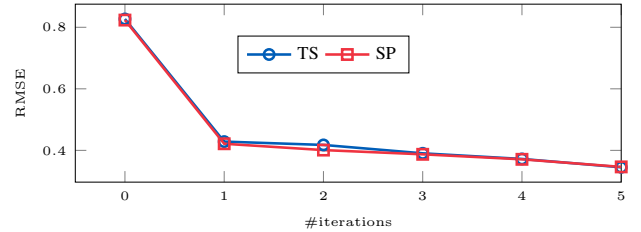| RMSE	TS	SP	PGFL	PF	TS	SP	FPFBS
2.5%	0.542	0.540	0.537	**0.536**	0.243	**0.241**	0.250
50%	0.748	**0.746**	0.749	0.749	**0.328**	**0.328**	0.329
97.5%	1.084	**1.082**	1.091	1.083	0.466	**0.464**	0.484

Figure 1. The RMSE over iterations for TS and SP implementations of Alg. 3. 0 corresponds to the filters. Most of the improvement comes from the first smoother pass, while subsequent iterations make smaller improvements.

(TS) and sigma-point (SP) integration, the latter using a UT [8] with $\kappa = 2$. The number of iterations per filter update and initialisation of the smoothers was set to 15 to reduce the frequency of divergence for TS. *Both filter implementations diverge when using only one iteration*, due to the severe non-linearity of the system, justifying the proposed iterative scheme. The number of smoothing iterations was 5. The suggested estimators are compared to the progressive Gaussian filter (PGFL) using explicit likelihoods [20], using 20 deterministic points (using more had negligible impact on performance). It should be noted there is no smoother counterpart to PGFL.

For reference, a bootstrap filter (PF) [15] was run using $N = 5000$ particles and resampling at an effective number of samples $N/3$. In order to obtain a smoothing reference 100 samples of the smoothing pdf was obtained using a bootstrap filter with backward simulation (FPFBS) [17]. All algorithms used $\mathcal{N}(\log 7, 0.1)$ as initial distribution. It should be noted that the Taylor series implementation diverged 7 times out of the 250 trials. These trials were discounted from the calculation of the RMSE. As indicated in Table I, the sigma-point implementation is, in general, superior to the alternatives.

A new trajectory was simulated to visualise the improvements over iterations. Both implementations of the smoother were run for 0–5 iterations, where 0 corresponds to the filter, which used 15 iterations for the update; see Fig. 1.

VI. CONCLUSION

The IPLF/S methods were generalised to the case where only the conditional moments in the dynamics and the measurement models need to be evaluated, this makes the methods applicable to strictly non-Gaussian state-space models. The resulting algorithms were shown to give the IEKF/S and IPLF/S as special cases for non-linear models with additive Gaussian noise, and provide their natural extensions in the case of non-additive noise. Furthermore, the developed algorithms were evaluated on the stochastic Ricker map, where they were found to be comparable or superior to other state-of-the-art estimators.

REFERENCES

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

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

[3] S. J. Godsill and P. J. Rayner, *Digital Audio Restoration: A Statistical Model based Approach*. Springer, 1998.

[4] E. Lindström, H. Madsen, and J. N. Nielsen, *Statistics for Finance*. Chapman and Hall/CRC, 2015.

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

[6] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic system," *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, Aug 1965.

[7] P. S. Maybeck, *Stochastic Models, Estimation and Control*. Academic Press, 1982, vol. 2.

[8] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new approach for filtering nonlinear systems," in *Proceedings of the 1995 American Control Conference*, 1995, pp. 1628–1632.

[9] S. Särkkä, "Unscented Rauch–Tung–Striebel Smoother," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 845–849, April 2008.

[10] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, Aug 2006.

[11] 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]," *Automatic Control, IEEE Transactions on*, vol. 47, no. 8, pp. 1406–1409, 2002.

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

[13] M. R. Morelande, "Tracking multiple targets with a sensor network," in *9th International Conference on Information Fusion (FUSION)*, 10-13 July 2006.

[14] S. Särkkä, "Recursive Bayesian inference on stochastic differential equations," Doctoral dissertation, Helsinki University of Technology, 2006.

[15] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, pp. 107–113, 1993.

[16] A. Doucet, S. J. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[17] ——, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004.

[18] R. Zanetti, "Recursive update filtering for nonlinear estimation," *IEEE Transactions on Automatic Control*, vol. 57, pp. 1481–1490, 2012.

[19] U. D. Hanebeck, "PGF 42: Progressive Gaussian filtering with a twist," in *16th International Conference on Information Fusion (FUSION)*, 2013.

[20] J. Steinbring and U. D. Hanebeck, "Progressive Gaussian filtering using explicit likelihoods," in *Proceedings of the International Conference on Information Fusion*, 7-10 July 2014.

[21] T. P. Minka, "Expectation propagation for approximate Bayesian inference." in *In proceedings of the seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publisher Inc., 10-13 July 2006, pp. 362–369.

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

[23] B. M. Bell, "The iterated Kalman smoother as a Gauss–Newton method," *SIAM Journal on Optimization*, vol. 4, no. 3, pp. 626–636, 1994.

[24] M. Skoglund, G. Hendeby, and D. Axehill, "Extended Kalman filter modifications based on an optimization view point," in *18th International Conference on Information Fusion (FUSION)*, 6-9 July 2015.

[25] A. F. Garcia-Fernandez, L. Svensson, M. R. Morelande, and S. Särkkä, "Posterior linearization filter: Principles and implementation using sigma points," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5561–5573, 2015.

[26] Á. F. García-Fernández, L. Svensson, and S. Särkkä, "Iterated posterior linearisation smoother," *IEEE Transactions on Automatic Control*, vol. 62, no. 4, pp. 2056–2063, 2017.

[27] M. Fasiolo, N. Pya, and S. N. Wood, "A comparison of inferential methods for highly nonlinear state space models in ecology and epidemiology," *Statistical Science*, vol. 31, no. 1, pp. 96–118, 2016.

[28] F. Tronarp, R. Hostettler, and S. Särkkä, "Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise," in *Proceedings of the International Conference on Information Fusion*, 5-8 July 2016.

[29] A. N. Shiryaev, *Probability*. Springer-Verlag Berlin Heidelberg, 1996.