Efficient Iterated Filtering

Erik Lindström * **Edward Ionides** ** **Jan Frydendall** ***
Henrik Madsen ****

* *Center for Mathematical Sciences, Lund University, Sweden (e-mail: erikl@maths.lth.se).*
** *Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109 USA, (e-mail: ionides@umich.edu)*
*** *DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark (e-mail: jf@imm.dtu.dk)*
**** *DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark, (e-mail: hm@imm.dtu.dk)*

Abstract: Parameter estimation in general state space models is not trivial as the likelihood is unknown. We propose a recursive estimator for general state space models, and show that the estimates converge to the true parameters with probability one. The estimates are also asymptotically Cramer-Rao efficient.
The proposed estimator is easy to implement as it only relies on non-linear filtering. This makes the framework flexible as it is easy to tune the implementation to achieve computational efficiency.
This is done by using the approximation of the score function derived from the theory on Iterative Filtering as a building block within the recursive maximum likelihood estimator.

Keywords: Recursive estimation, maximum likelihood estimator, filtering techniques, stochastic approximation, iterative methods.

1. INTRODUCTION

Iterative filtering, see Ionides et al. (2006, 2011) is a technique for computing the Maximum Likelihood Estimate (MLE) for general state space models when the likelihood is not known in closed form. This is done by augmenting the state vector with the parameter vector, and applying non-linear filtering methods to the augmented system. The parameters in the augmented system are given random walk dynamics. Offline parameter estimates are computed as a linear combination of the filter estimates of the parameter vector. The technique is iterating between non-linear filtering and averaging until convergence.

Augmenting the unobserved state vector is a well known technique, used in the system identification community for decades, see e.g. Ljung (1979); Söderström and Stoica (1989); Lindström et al. (2008). Similar ideas, using Sequential Monte Carlos methods were suggested by Kitagawa (1998); Liu and West (2001). Combined state and parameter estimation is also the standard technique for data assimilation in high-dimensional systems, see Moradkhani et al. (2005); Evensen (2009b,a).

However, introducing random walk dynamics to the parameters with fixed variance leads to a new dynamical stochastic system with properties that may be different from the properties of the original system. That implies that the variance of the random walk should be decreased, when the method is used for offline parameter estimation, cf. Hürzeler and Künsch (2001). We show, by combining results from the Iterated filtering framework with the combined state and parameter estimation technique that the resulting algorithm can be interpreted as a stochastic approximation algorithm, and that the sequence of parameter estimates are converging with probability one to the true parameters under suitable conditions.

The contribution of this paper is twofold. We show that the Iterated filtering algorithm is not optimal in terms of speed of convergence, and we proceed by introducing a new algorithm, called Efficient Iterative Filtering. The new algorithm is shown to be more efficient than the Iterative Filtering algorithm, and it is also shown to be the recursive Maximum Likelihood estimator. A further advantage of proposed algorithm is that it is easier to implement than the ordinary Iterative filtering algorithm.

The paper is organized as follows. Section 2 introduces the concept of iterative filtering, and Section 3 explores possible modifications of this algorithm. The performance of the algorithms are compared in Section 4 and Section 5 concludes.

2. PARAMETER ESTIMATION AND FILTERING

A general state space model consists of an unobserved state process $X_{0:n}$ and observable process $Y_{1:n}$, where the observations are independent conditional on the unobserved state. We write $f_{Y_{1:n}}(y_{1:n})$ for the density of $Y_{1:n}$, using analogous notion for other densities.

The preferred estimation method in statistics is often the maximum likelihood estimator (MLE), defined as

$$\hat{\theta}^{MLE} = \underset{\theta \in R \subseteq \mathbb{R}^p}{\arg\max} \log f_{Y_{1:n}}(y_{1:n}|\theta) = \underset{\theta \in R \subseteq \mathbb{R}^p}{\arg\max} \ell(\theta). \quad (1)$$

The MLE is asymptotically consistent for a huge class of models, and it is often claimed that the MLE is optimal as the covariance of the estimates is the smallest possible among asymptotically unbiased estimators, see Van der Vaart (2000) for conditions.

A practical limitation is that it is difficult or even impossible to compute the log-likelihood and hence the MLE for most general state space models. It is readily shown that the log-likelihood is given by

$$\ell(\theta) = \log \int f_{X_0}(x_0;\theta) \prod_{k=1}^{n} f_{X_k,Y_k|X_{k-1}}(x_k, y_k|x_{k-1};\theta)\mathrm{d}x_{0:n}. \tag{2}$$

Sequential Monte Carlo methods is a popular class of algorithms for inference in general state-space models that solves the integrals using Monte Carlo methods.

2.1 Iterated filtering

Iterated filtering is a method for estimating parameters in general state space models using non-linear filtering. The results in Ionides et al. (2006) holds for deterministic filters while Ionides et al. (2011) extends the framework to stochastic (i.e. sequential Monte Carlo) filters. Their idea is to augment the unobserved state process with the parameters and use the estimated unobserved states to form a sequence of parameter estimates, see also Bretó et al. (2009). We will essentially follow their notation throughout this paper.

A time varying parameter process $\{\breve{\Theta}_n\}$ is introduced. Let κ be a density with compact support, zero mean and covariance matrix Σ, and let ζ_k be an independent draw from κ. The time varying parameter process is then defined as

$$\breve{\Theta}_0 = \theta + \tau\zeta_0, \tag{3}$$
$$\breve{\Theta}_n = \breve{\theta}_{n-1} + \sigma\zeta_n. \tag{4}$$

The stochastically perturbed model is defined conditionally on the time varying parameter process

$$g_{\breve{X}_{0:N},\breve{Y}_{1:N},\breve{\Theta}_{0:N}}(x_{0:N}, y_{1:N}, \breve{\theta}_{0:N};\theta,\tau) \tag{5}$$
$$= f_{X_{0:N},Y_{1:N}}(x_{0:N}, y_{1:N};\breve{\theta}_{0:N}) g_{\breve{\Theta}_{0:N}}(\breve{\theta}_{0:N};\theta,\tau,\Sigma). \tag{6}$$

Define the moments

$$\breve{\theta}_n^F = \mathbb{E}_{\theta,\sigma,\tau}[\breve{\Theta}_n|Y_{1:n} = y_{1:n}], \tag{7}$$
$$\breve{V}_n^P = \mathrm{Var}_{\theta,\sigma,\tau}[\breve{\Theta}_n|Y_{1:n-1} = y_{1:n-1}]. \tag{8}$$

It was shown in Ionides et al. (2011) that the score function can be approximated with these moments.

Theorem 1. (Theorem 3 in Ionides et al. (2011)). Let K_1 be a compact subset of $\mathbb{R}^p$, C_1 is a constant, τ is small enough and $\lim_{\tau\to 0}\sigma(\tau)/\tau = 0$. It then holds that

$$\sup_{\theta\in K_1}\left|\sum_{n=1}^{N}(\breve{V}_n^P)^{-1}\left(\breve{\theta}_n^F - \breve{\theta}_{n-1}^F\right) - \nabla\ell(\theta)\right| \le C_1(\tau + \frac{\sigma^2}{\tau^2}). \tag{9}$$

Moments are often easier to estimate than gradients, which is why this framework is so computationally attractive. A straightforward (plug and play) implementation is algorithm 1, where $Z^T = [X^T\ \breve{\Theta}^T]^T$

Initialize particles $Z_{0,j}^F \sim f_{Z_0}(z_0;\breve{\theta})$
for $n = 1$ to N **do**
  for $j = 1$ to J **do**
    Draw particles $Z_{n,j}^P \sim f_{Z_n|Z_{n-1}}(z_n|Z_{n-1,j}^F;\breve{\theta})$
    Assign $w_{n,j} = f_{Y_n|Z_n}(y_n|Z_{n,j};\breve{\theta})$
    Draw I_j such that $\mathbb{P}(I_j = i) = \frac{w_{n,i}}{\sum_l w_{n,l}}$.
    Assign $Z_{n,j}^F = Z_{n,I_j}^P$
  end for
end for

Algorithm 1. Implementation of a bootstrap filter for estimating the distribution of the unobserved states in the stochastically perturbed system.

The filtered (estimated) states can be combined to form an approximation of the log-likelihood. Define $\tilde{\theta}_n^F$ and $\tilde{V}_n^P$ as the sample versions of (7) and (8) computed from the Monte Carlo filter using J particles.

Theorem 2. (Theorem 4 in Ionides et al. (2011)). Let K_2 be a compact subset of $\mathbb{R}^p$, $\{\tau_m\}$, $\{\sigma_m\}$ and $\{J_m\}$ be sequences such that $\tau_m \to 0$, $\sigma_m\tau_m^{-1} \to 0$ and $\tau_m J_m \to \infty$, and define

$$\widetilde{\nabla}\ell(\theta) = \sum_{n=1}^{N}(\tilde{V}_{n,m}^P)^{-1}(\tilde{\theta}_{n,m}^F - \tilde{\theta}_{n-1,m}^F). \tag{10}$$

It then holds that

$$\lim_{m\to\infty}\sup_{\theta\in K_2}\left|E_{MC}\left[\widetilde{\nabla}\ell(\theta)\right] - \nabla\ell(\theta)\right| = 0, \tag{11}$$
$$\lim_{m\to\infty}\sup_{\theta\in K_2}\left|\tau_m^2 J_m \mathrm{Var}_{MC}\left[\widetilde{\nabla}\ell(\theta)\right]\right| < \infty. \tag{12}$$

This is a remarkable result! We can approximate the score function arbitrarily well when the number of particles J_m is increased. Many other techniques for computing the score function is of order $O(J^2)$, see Poyiadjis et al. (2011). The above theorems were then combined to form a sequence of estimates that converges to the MLE.

Theorem 3. (Theorem 5 in Ionides et al. (2011)). Let $\{a_m\}$, $\{\tau_m\}$, $\{\sigma_m\}$ and $\{J_m\}$ be positive sequences such that $\tau_m \to 0$, $\sigma_m\tau_m^{-1} \to 0, \tau_m J_m \to \infty, \sum_m a_m = \infty$ and $\sum_m a_m^2 J_m \tau_m^{-2} < \infty$ and define $\hat{\theta}_m$ according to:

$$\hat{\theta}_{m+1} = \hat{\theta}_m + a_m\sum_{n=1}^{N}(\tilde{V}_{n,m}^P)^{-1}\left(\tilde{\theta}_{n,m}^F - \tilde{\theta}_{n-1,m}^F\right) \tag{13}$$

The estimate will then converge $\lim_{m\to\infty}\hat{\theta}_m = \hat{\theta}^{MLE}$ with probability one.

Choosing $a_m = m^{-1}$, $\tau_m^2 = m^{-1}, \sigma_m^2 = m^{-(1+\delta)}$ and $J_m = m^{(1/2+\delta)}$ where $\delta > 0$ satisfies the conditions in Theorem 3.

3. RECURSIVE ITERATED FILTERING

We have found several practical issues when using iterated filtering suggested by Ionides et al. (2011).

- The method (as many other stochastic approximation algorithms) is highly sensitive to the choice of a_m, cf. Kantas et al. (2009).
- The Iterated filtering method is essentially a stochastic gradient methods. The rate of convergence can

often be improved by using second order information, cf. the Newton method or second order stochastic approximation, see Spall (2003). Quantity calculus suggests that a_m should be related to V^P. In fact, Ionides et al. (2006), used V_N^P.
- Speed of convergence. a_m will influence the rate of convergence of the offline estimator. The performance would be even better for a recursive estimator, cf. Kantas et al. (2009) as the convergence would be one-dimensional (in n only), not two-dimensional (in n and m).

We use the Iterated filtering framework to derive a recursive estimator.

3.1 *Recursive estimation*

We start with some results on stochastic approximation. The Robbins-Monro stochastic approximation procedure is a class of algorithms solving

$$\theta^\star = \arg\max F(\theta) \tag{14}$$

when only noisy and possibly biased gradient observations are available $f = \nabla F + W + B$, where W is random noise and B is bias.

Theorem 4. (Theorem 1.9 in Ljung et al. (1992)). Let $\mathbb{H}$ be a Hilbert space endowed with the Borel σ-algebra and $f : \mathbb{H} \to \mathbb{H}$ be a measurable function. Let θ_n, W_n and B_n be $\mathbb{H}$-valued random variables and define

$$\theta_{n+1} = \theta_n + a_n(f(\theta_n) + W_n + B_n). \tag{15}$$

Assume that

A.1 $a_n \geq 0, \sum_n a_n = \infty, \sum_n a_n^2 < \infty$.
A.2 $\sum_n a_n E||B_n|| < \infty, \sum_n a_n^2 E||B_n||^2 < \infty$.
A.3 $E[W_n|\theta_0, W_1, \ldots, W_{n-1}, B_1, \ldots, B_{n-1}] = 0$.
A.4 $\sum_n a_n^2 E||V_n||^2 < \infty$.
A.5 $\forall_{\theta\in\mathbb{H}} \ : \ ||f(\theta)|| \leq c(1 + ||\theta||)$.
A.6 $\inf\{(f(\theta), \theta - \theta^\star); \theta \in \mathbb{H}, 1/K \leq ||\theta - \theta^\star|| \leq K\} > 0$.

The sequence of estimates will then converge with probability one to the argument that maximizes the function, $\theta_n \overset{a.s.}{\to} \theta^\star$.

It should be noted the dynamics of the stochastically perturbed system may be different from the original system, cf. (6), but also that the difference is related to the (decreasing) size of the stochastic perturbation.

Additional results on the rate of convergence and asymptotic distribution was derived in Fabian (1978).

Theorem 5. (Fabian (1978)). Assume the conditions in Theorem 4 hold. Let $q(\theta, Y) \in L_2(P)\ \forall\theta \in S \subset \mathbb{R}^p$ be an estimation function with $E[q(\theta^\star, Y)] = 0$, covariance $E[q(\theta^\star, Y)q(\theta^\star, Y)^T] = \Sigma$, and total differential H at $\theta^\star$. Define

$$\theta_{n+1} = \theta_n + \frac{1}{n}\Phi_n q(\theta_n, Y_n) \tag{16}$$

where $\Phi_n \overset{p}{\to} H^{-1}$ when $\theta_n \to \theta^\star$. Assume that $\theta_n \to \theta^\star$. It then follows that

$$\sqrt{n}\,(\theta_n - \theta^\star) \overset{d}{\to} N(0, H^{-1}\Sigma H^{-1}). \tag{17}$$

Theorem 5 proves that the *recursive maximum likelihood estimator* (RMLE), defined as

$$\theta_{n+1} = \theta_n + \frac{1}{n} I(\theta_n)^{-1}\nabla \log p(y_{n+1}|\theta_n, y_{1:n}) \tag{18}$$

converge according to

$$\sqrt{n}\,(\theta_n - \theta^\star) \overset{d}{\to} N(0, I(\theta)^{-1}I(\theta)I(\theta)^{-1}). \tag{19}$$

It is theoretically possible to design Cramer-Rao efficient recursive estimators for general state space models (implementing them is a different story, cf. Holst and Lindgren (1991)).

3.2 *A simple, recursive estimator*

The purpose is to mimic the RMLE while using the Iterated filtering framework to approximate the score function.

Our basis is similar to Ionides et al. (2011), augmenting the unobserved state with the parameters and assign random walk dynamics to the parameters.

Proposition 6. Define the model

$$Y_n = X_n + \eta_n, \eta \sim N(0, \Gamma) \tag{20}$$

$$X_n = X_{n-1} + e_n, e_n \sim N(0, Q_n) \tag{21}$$

and assume that $P_0 = \mathrm{cov}(X_0) = \infty$. Furthermore, let

$$Q_n = P_n q/(n + 1) \tag{22}$$

where q is a scalar and $P_n = \mathrm{cov}(X_n|Y_{1:n})$ is the covariance of the state. P_n is then given by

$$P_n = \frac{(1 + q)\Gamma}{n + 1 + q}, \tag{23}$$

implying that $Q_n = O(n^{-2})$.

Proof. The covariance P_{n+1} is given by

$$P_{n+1} = \frac{\Gamma P_n(1 + q/(n + 1))}{\Gamma + P_n(1 + q/(n + 1))} \tag{24}$$

Introduce $X_n = 1/P_n$ and $S = 1/\Gamma$. Rewriting (24) yields

$$X_{n+1} = \frac{\Gamma + P_n(1 + q/(n + 1))}{\Gamma P_n(1 + q/(n + 1))} \tag{25}$$

$$= \frac{X_n}{1 + q/(n + 1)} + S = \frac{(n + 1)}{n + 1 + q}X_n + S \tag{26}$$

Numerical simulations have indicated that the solution decays approximately as $1/n$. We guess that the solution is given by

$$\bar{P}_n = \frac{c\Gamma}{n + a} \tag{27}$$

Plugging this guess into (26) yields

$$\frac{n + 1}{n + 1 + q}\frac{n + a}{c\Gamma} + \frac{c}{c\Gamma} \tag{28}$$

$$= \frac{1}{c\Gamma}\frac{(n + 1)(n + a) + c(n + 1 + q)}{n + 1 + q} \overset{?}{=} 1/\bar{P}_{n+1} \tag{29}$$

$\bar{P}$ would be a solution to (26) if the second ratio is equal to $(n+a+1)$. We found this to be the case when $c = 1+q$ and $a = 1 + q$, which confirms that $\bar{P}_n$ is a solution to (24). □

Adding noise to the unobserved (parameter)-states prevents particle degeneration, but it also change the dynamics of the model. The side effects are controlled by decreasing the variance of the noise.

Conjecture 7. We assume that the rate of convergence for the covariance of the parameter state estimates is $O(n^{-1})$

(as in (23)) when Q_n is defined as (22) in regardless of the model structure for all identifiable models.

The estimator We will now bring together the ideas and tools from the Iterated filtering framework.

The first modification is to specify the estimator as a recursive estimator, rather than a batch estimator. Recursive estimators are popular for Markov processes due to the conditional independence, but general state space models are not Markov processes. It was suggested in Holst and Lindgren (1991) to combine stochastic approximation with the EM-algorithm to compute parameter estimates recursively. This is computationally expensive as the smoothing distribution is needed. A remedy was proposed in Rydén (1997) where mixing arguments were used to justify that the smoothing distribution can be approximated by a fixed-lag smoother. The resulting estimator is biased, but the bias can be controlled by increasing the number of lags in the smoother. However, the Taylor expansion in Ionides et al. (2006) has provided an elegant alternative. Our algorithm will use their approximation of the score function. A recursive estimator, cf. (18) does only use a single observation per update, meaning that the approximations in Ionides et al. (2011) only will use a single term in the sum in (9).

The second limitation that we would like to eliminate is the difficulty of finding a proper $\{a_m\}$ sequence. The calculus of quantity argument suggested that it should be related to V^P. Proposition 6 and Conjecture 7 indicate that V^P converges at a suitable rate, meaning that V_n^P could be a good candidate. A particularly nice property of V^P is that it is positive definite, cf. Quasi Newton methods.

Proposition 8. Assume that Q_n is defined as in (22) and that Conjecture 7 hold. Furthermore, let $J_n/\sqrt{n} \to \infty$ and define

$$\theta_n = \tilde{\theta}_n^F \ \forall\ n. \tag{30}$$

It then follows that $\theta_n \overset{a.s.}{\to} \theta$.

Proof. It follows from Conjecture 7 that

$$P_n = n^{-1}\Phi_n \tag{31}$$

where $\Phi_n \to H^{-1}$ large n. We also know that $V_n^P = O(P_n)$ since the difference is $o(n^{-1})$. Algebraic manipulations then give

$$\begin{aligned}
\theta_n &= \tilde{\theta}_n^F \pm \tilde{\theta}_{n-1}^F && (32)\\
&= \theta_{n-1} + \left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) && (33)\\
&= \theta_{n-1} + (\tilde{V}_n^P)(\tilde{V}_n^P)^{-1}\left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) && (34)\\
&= \theta_{n-1} + \frac{1}{n}(n\tilde{V}_n^P)[(\tilde{V}_n^P)^{-1}\left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) && (35)\\
&\quad \pm \nabla \log p(y_n|\theta_{n-1}, y_{1:n-1})]
\end{aligned}$$

Now, define $a_n = 1/n$, $\Phi_n = n\tilde{V}_n^P$ and $R_n = (\tilde{V}_n^P)^{-1}\left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) - \nabla \log p(y_n|\theta_{n-1}, y_{1:n-1})$. The algorithm is then given by

$$\theta_n = \theta_{n-1} + a_n\Phi_n(\nabla \log p(y_n|\theta_{n-1}^F, y_{1:n-1}) + R_n) \tag{36}$$

We need to derive expressions for σ_n^2 and τ_n^2. Direct comparisons with the Iterated filter algorithm gives

$$\sigma_n^2 = Q_n = O(n^{-2}) \tag{37}$$
$$\tau_n = P_n = O(n^{-1}). \tag{38}$$

Convergence is shown by using Theorem 4. Introduce:

$$W_n = R_n - E[R_n|\cdot] \tag{39}$$
$$B_n = E[R_n|\cdot] \tag{40}$$

The residual term, R_n, can be decomposed as follows

$$\begin{aligned}
R_n =& (\tilde{V}_n^P)^{-1}\left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) - \nabla \log p(y_n|\theta_{n-1}, y_{1:n-1}) && (41)\\
&\pm (\breve{V}_n^P)^{-1}\left(\breve{\theta}_n^F - \breve{\theta}_{n-1}^F\right)\\
=& \left((\tilde{V}_n^P)^{-1}\left(\tilde{\theta}_n^F - \tilde{\theta}_{n-1}^F\right) - (\breve{V}_n^P)^{-1}\left(\breve{\theta}_n^F - \breve{\theta}_{n-1}^F\right)\right) && (42)\\
&+ \left((\breve{V}_n^P)^{-1}\left(\breve{\theta}_n^F - \breve{\theta}_{n-1}^F\right) - \nabla \log p(y_n|\theta_{n-1}, y_{1:n-1})\right). && (43)
\end{aligned}$$

Only the first term is stochastic. It is possible to show, using the triangle inequality, Theorem 1 and 2 and with the assistance of additional results in Ionides et al. (2011) that

$$E[|R_n|] \le \frac{C_2}{\tau_n J_n} + C_1(\tau_n + \frac{\sigma_n^2}{\tau_n^2}) \tag{44}$$

$$\mathrm{Var}_{MC}(R_n) \le \frac{C_3}{\tau_n^2 J_n} \tag{45}$$

where C_2 and C_3 are constants.

It then follows that the conditions in Theorem 4 are fulfilled when $J_n/\sqrt{n} \to \infty$, e.g. by defining $J_n = O(n^{1/2+\delta})$ for some $\delta > 0$. □

Corollary 9. It has been a standard technique to use a non-linear (e.g. Extended or Unscented) Kalman filters to track parameters. It is well known that non-linear Kalman filters are approximate filters, and that the state estimates are biased if the system is non-linear. The conditions for convergence are therefore not be fulfilled (the bias does not disappear), and the estimate will therefore converge (if the covariances are decreased sufficiently fast) to some value different from the MLE. This result is consistent with Ljung (1979).

Proposition 10. Assume that the conditions in Theorem 5, Conjecture 7 and Proposition 8 are fulfilled. It then follows that

$$\sqrt{n}\,(\theta_n - \theta) \overset{d}{\to} N(0, H^{-1}I(\theta)H^{-1}). \tag{46}$$

Proof. This follows directly from Theorem 5, Conjecture 7 and Proposition 8.

This result holds regardless of q. We can choose q to be arbitrarily close to 0, making H similar to the Fisher information due to the connection between the Hessian of the log-likelihood and the parameter covariance. This means that this version of the Efficient Iterated filter is the most efficient recursive estimator that can be derived.

Batch estimator It is possible to derive a batch version of our algorithm. This algorithm differs from Ionides et al. (2011) not only in terms of how the estimate is updated

$$\theta_m = \tilde{\theta}_{N,m-1}. \tag{47}$$

but also by using the estimated covariance at time N from iteration $m-1$ as the initial covariance for iteration m

$$P_{0,m} = P_{N,m-1}. \tag{48}$$

Practical considerations Our algorithm is a Robbins-Monro type algorithm, where we use the fastest possible cool-down rate. Studies have shows, see Spall (2003) and references therein, that it is often advantageous to replace

$$a_n = \frac{1}{n+1} \tag{49}$$

with

$$a_n = \frac{a}{n+1+A}. \tag{50}$$

where a and A are positive numbers. Taking A as say 5% of the total number of iterations, and adjusting a correspondingly results in an algorithm where the asymptotic rate of convergence still is optimal and the practical performance is improved. We can achieve the same effect by tweaking the Q_n matrix.

Another technique for improving the finite sample performance is to use a large (but decreasing) number of particles for the initial iterations while increasing the number of particles as the number of iteration grows. We have used

$$J_m = J_0 \exp(-(m-1)) + J_1 + J_2 m^{0.6} \tag{51}$$

in our simulations. Adding the J_0 term counteracts particle depletion in the initial iterations, preventing the estimator from getting stuck at a local optima, while controlling the computational cost.

Another advantage with our algorithm is that only $\tilde{\theta}_n^F$ needs to be computed. That means that we can use computationally efficient variance reduction methods to speed up the computations.

4. SIMULATIONS

We have compared the estimates from the proposed algorithm to the MLE, the estimates from the Iterated Filtering algorithm and to the Unscented Kalman Filter, see Julier and Uhlmann (2004) when used for combined state and parameter estimation.

We have only used the batch version of the proposed algorithm as we do not know of any recursive versions of the Iterated filter.

4.1 Constant

We start with the model

$$Y_n = \phi_n + \eta_n, \tag{52}$$
$$\phi_n = \phi_{n-1}. \tag{53}$$

This is the state space formulation for estimating a constant. The MLE is easily computed and both the Iterated Filter and the Efficient Iterated Filter can be computed from a Kalman filter, making comparisons without Monte Carlo noise possible.

The parameters were estimated using the statistical model

$$Y_n = \phi_n + \eta_n, \tag{54}$$
$$\phi_n = \phi_{n-1} + e_n,\ e_n \sim N(0, Q_m), \tag{55}$$

with $Q_m = 5 \cdot 10^{-3}/m^2$, where m is the iteration.

We simulated 1 000 independent copies, each consisting of $N = 200$ observations

$$y_n = 3 + \eta_n,\ \eta_n \sim N(0,1). \tag{56}$$

The distribution for x_0 for the first iteration was chosen as $N(0,1)$. The Iterated filter follows Ionides et al. (2011), while we for comparison use the batch version of the proposed estimator.

The difference between the MLE and the estimated parameters is presented in Figure 1. Both algorithms converge but the proposed algorithm is converging much faster than the Iterated filtering algorithm.

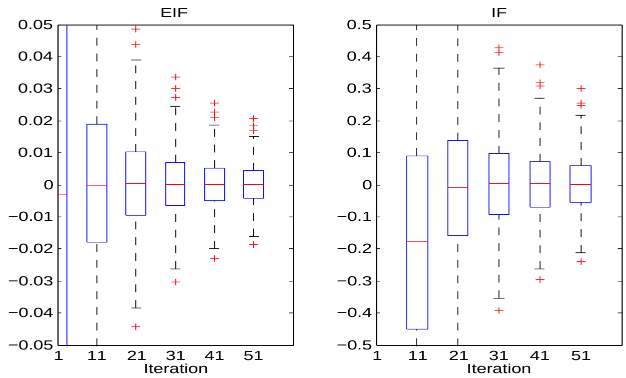

Fig. 1. Convergence to the MLE $\hat{\theta}_m - \hat{\theta}_{MLE}$ of the EIF (left) and IF (right) for 1000 independent realizations of the *iid* (constant) data. Note the different scales.

4.2 Linear model

A somewhat more complex model is linear Gaussian model

$$Y_n = X_n + \eta_n,\ \eta \sim N(0,1), \tag{57}$$
$$X_n = \phi X_{n-1} + e_n^X,\ e^X \sim N(0,1). \tag{58}$$

Augmenting the state vector by introducing ϕ as a state variable leads to the statistical model

$$Y_n = [1\ 0] \begin{bmatrix} X_n \\ \phi_n \end{bmatrix} + \eta_n, \tag{59}$$
$$X_n = X_{n-1}\phi_{n-1} + e_n^X, \tag{60}$$
$$\phi_n = \phi_{n-1} + e_n^\phi,\ e^\phi \sim N(0, Q_m), \tag{61}$$

where $Q_m = 10^{-2}/m^2$.

We can compute the MLE for the unperturbed model, while Monte Carlo filters are used for the proposed algorithm and the Iterated filter algorithm. The simulations used $J_0 = 1000$, $J_1 = 200$ and $J_2 = 100$ particles.

We have compare these with the UKF that is used for combined state and parameter estimation. The results are presented in Figure 2 and 3.

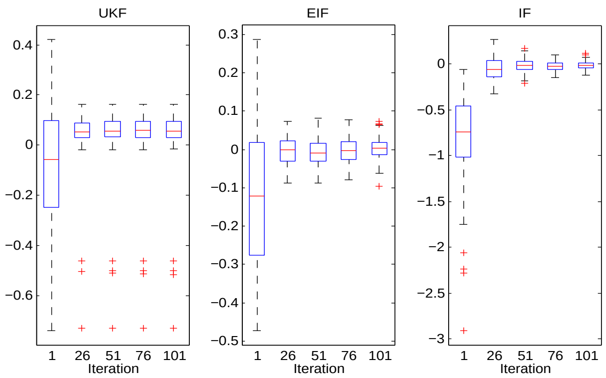

Fig. 2. Convergence to the MLE $\hat{\theta}_m - \hat{\theta}_{MLE}$ of the UKF (left), EIF (middle) and IF(right) for 100 independent realizations from the linear model. Note the different scales.

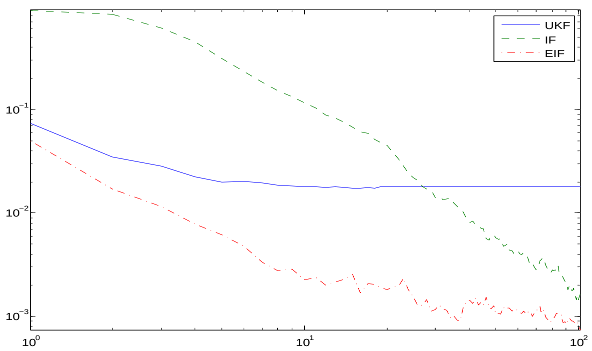

Fig. 3. RMSE between the MLE and the UKF, the IF and EIF for 100 independent realizations from the linear model.

It can be seen that the UKF is biased. The Efficient Iterated filtering is initially converging at about the same speed as the UKF, but the errors continue to decrease as the number of iterations increase. The same hold for the IF, but the RMSE is generally larger than for the EIF.

5. CONCLUSION

We derived a recursive maximum likelihood estimator by combining the Iterated filtering framework with stochastic approximation. The resulting estimator share the theoretical properties of the Iterated filtering while being easier to implement and computationally more efficient.

ACKNOWLEDGEMENTS

We are grateful to prof. Mario Natiello and Björn Stenqvist for stimulating discussions on the topic.

REFERENCES

Bretó, C., He, D., Ionides, E., and King, A. (2009). Time series analysis via mechanistic models. *The Annals of Applied Statistics*, 3(1), 319–348.

Evensen, G. (2009a). *Data assimilation: the ensemble Kalman filter*. Springer Verlag.

Evensen, G. (2009b). The ensemble Kalman filter for combined state and parameter estimation. *Control Systems Magazine, IEEE*, 29(3), 83–104.

Fabian, V. (1978). On asymptotically efficient recursive estimation. *The Annals of Statistics*, 854–866.

Holst, U. and Lindgren, G. (1991). Recursive estimation in mixture models with Markov regime. *Information Theory, IEEE Transactions on*, 37(6), 1683–1690.

Hürzeler, M. and Künsch, H.R. (2001). Approximating and maximising the likelihood for a general state-space model. In *Sequential Monte Carlo Methods in Practice*. Springer Verlag.

Ionides, E.L., Bhadra, A., Atchade, Y.F., and King, A.A. (2011). Iterated filtering. *Annals of Statistics*, 39, 1776–1802.

Ionides, E.L., Breto, C., and King, A.A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103, 18438–18443.

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

Kantas, N., Doucet, A., Singh, S., and Maciejowski, J. (2009). An overview of sequential monte carlo methods for parameter estimation in general state-space models. In *Proceedings IFAC System Identification*.

Kitagawa, G. (1998). A self-organizing state-space model. *Journal of the American Statistical Association*, 1203–1215.

Lindström, E., Ströjby, J., Brodén, M., Wiktorsson, M., and Holst, J. (2008). Sequential calibration of options. *Computational Statistics & Data Analysis*, 52(6), 2877–2891.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice*, 197–217. New York: Springer-Verlag.

Ljung, L. (1979). Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *Automatic Control, IEEE Transactions on*, 24(1), 36–50.

Ljung, L., Pflug, G., and Walk, H. (1992). *Stochastic approximation and optimization of random systems*, volume 17. Birkhäuser.

Moradkhani, H., Sorooshian, S., Gupta, H., and Houser, P. (2005). Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2), 135–147.

Poyiadjis, G., Doucet, A., and Singh, S. (2011). Particle approximations of the score and observed information matrix in state space models with application to parameter estimation. *Biometrika*, 98(1), 65.

Rydén, T. (1997). On recursive estimation for hidden Markov models. *Stochastic Processes and their Applications*, 66(1), 79–96.

Söderström, T. and Stoica, P. (1989). *System Identification*. Prentice-Hall.

Spall, J. (2003). *Introduction to stochastic search and optimization: estimation, simulation, and control*, volume 64. LibreDigital.

Van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge Univ Press.