18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

Extended Kalman Filter Modifications Based on an Optimization View Point

Martin A. Skoglund*, Gustaf Hendeby*† Daniel Axehill *
* Division of Automatic Control, Linköping University, SE-581 83 Linköping, Sweden,
Email: {ms,hendeby,daniel}@isy.liu.se
† Dept. of Sensor & EW Systems, Swedish Defence Research Agency (FOI), Linköping, Sweden

***Abstract*—The *extended Kalman filter* (EKF) has been an important tool for state estimation of nonlinear systems since its introduction. However, the EKF does not possess the same optimality properties as the Kalman filter, and may perform poorly. By viewing the EKF as an optimization problem it is possible to, in many cases, improve its performance and robustness. The paper derives three variations of the EKF by applying different optimisation algorithms to the EKF cost function and relate these to the iterated EKF. The derived filters are evaluated in two simulation studies which exemplify the presented filters.**

I. Introduction

The *Kalman filter* (KF, [1]) is a minimum variance estimator if the involved models are linear and the noise is additive Gaussian. A common situation is however that dynamic models and/or measurement models are nonlinear. The *extended Kalman filter* (EKF, [2]) approximates these nonlinearities using a first-order Taylor expansion around the current estimate and then the time and the measurement update formulas as in the KF can be used. Higher-order terms can be included in the Taylor expansions if the degree of nonlinearity is high and thus not well approximated using only first-order information. Many suggestions have been made to improve the EKF. In the modified EKF [3] higher-order moments are included, [4, 5] use sigma-point approximations which is related to including second-order moments [6], while [7] proposes a robust EKF using H_∞ theory.

Another viewpoint is that the EKF is a special case of *Gauss-Newton* (GN) optimization without iterations, see *e.g.*, [8, 9] for longer discussions. Thus, iterations may improve the performance with just a little added effort. The idea explored here is that it is not only linearisation errors that causes filter inconsistencies but rather the non-iterative way of including new information as it may not be sufficient for staying close to the true state. There is of course nothing that prevents the inclusion of higher-order terms when re-linearising in the iterations. Notice that the *Fisher scoring algorithm* is also known as *Gauss-Newton* when applied to nonlinear least-squares see, *e.g.*, [10].

An iterative version of EKF is the *iterative extended Kalman filter* (IEKF, [2]) which re-linearises the measurement Jacobian at the updated state and applies the measurement update repeatedly until some stopping criterion is met. It should be noted that the iterations might not reduce the state error or reduce the associated cost function. It is actually not guaranteed even that a single update in the EKF improves the estimate since the full step may be to long. In [11] a bistatic ranging example illustrates when the IEKF converges to the true state as measurements become more accurate whereas the EKF is biased even though the state error covariance converges.

Yet the EKF often works satisfactorily which can be understood in the light of being a GN method in which the first correction often gives a large improvement compared to the following, often smaller, corrections. Furthermore, the cost is often decreased for the full step.

Hence, both IEKF and EKF could easily be improved using step control using *e.g.*, backtracking line search. That is, with step control, or at least checking for cost decrease, unnecessary divergent behaviour of the EKF can be avoided.

This paper shows that line search can be used in EKF and IEKF improve the overall performance. Furthermore, a quasi-Newton approximation for EKF update with a Hessian correction term is derived. This matrix can be successively built during the iterations using a symmetric rank-2 update. A slight modification of the *Levenberg-Marquardt-IEKF* (LM-IEKF, [12]) is made to include a diagonal damping matrix which could further speed up convergence.

The paper is organized as follows. In Sec. II the cost function for the measurement update of the EKF is derived. Sec. III describes the EKF as an optimization problem via Gauss-Newton and introduces line search as a way to improve the convergence rate. The IEKF is further modified to include an explicit step size. A quasi-Newton EKF is the introduced and a slight modification of the LM-IEKF is then derived. Sec. IV contains two simulations exemplifying the derived filters and the papper ends with conclusions and future directions in Sec. V.

II. The Measurement Update and its Cost Function

Consider systems having linear dynamics and nonlinear measurement models according to

$$x_t = F x_{t-1} + w_t \tag{1a}$$
$$y_t = h(x_t) + e_t \tag{1b}$$

where x_t is the state, y_t the measurement, $w_t \sim \mathcal{N}(0, Q_t)$ and $e_t \sim \mathcal{N}(0, R_t)$ all at time t. In a Bayesian setting, the

Algorithm 1 Kalman Filter

Available measurements are $Y = \{y_1, \ldots, y_N\}$. Require an initial state, $\hat{x}_{0|0}$, and an initial state covariance, $P_{0|0}$, Q_t and R_t are the covariance matrices of w_t and e_t, respectively.

1) Time Update

$$\hat{x}_{t|t-1} = F\hat{x}_{t-1|t-1} \tag{3a}$$

$$P_{t|t-1} = FP_{t-1|t-1}F^T + Q_t \tag{3b}$$

2) Measurement Update

$$K = P_{t|t-1}H^T(HP_{t|t-1}H^T + R_t)^{-1}, \tag{4a}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K(y_t - H\hat{x}_{t|t-1}), \tag{4b}$$

$$P_{t|t} = (I - KH)P_{t|t-1}, \tag{4c}$$

nonlinear Gaussian filtering problem propagate the following densities

$$p(x_{t-1}|Y_{t-1}) \approx \mathcal{N}\left(x_{t-1}; \hat{x}_{t-1|t-1}, P_{t-1|t-1}\right) \tag{2a}$$

$$p(x_t|Y_{t-1}) \approx \mathcal{N}\left(x_t; \hat{x}_{t|t-1}, P_{t|t-1}\right) \tag{2b}$$

$$p(x_t|Y_t) \approx \mathcal{N}\left(x_t; \hat{x}_{t|t}, P_{t|t}\right) \tag{2c}$$

where $p(x_{t-1}|Y_{t-1})$ is the initial prior density, $p(x_t|Y_{t-1})$ is the prediction density, $p(x_t|Y_t)$ is the posterior density and $Y_t = \{y_i\}_{i=1}^t$ are the measurements. The dynamic model (1b) is linear which means that the prediction density is exactly propagated in the KF time update equations as given in Algorithm 1. The measurement update is however nonlinear and is given as the *maximum a posteriori* (MAP) estimate according to

$$\hat{x}_{t|t} = \arg\max_{x_t} p(x_t|Y_t). \tag{5}$$

The posterior is proportional to the product of the likelihood and the prior

$$\begin{aligned} p(x_t|Y_t) &\propto p(y_t|x_t)p(x_t|Y_{t-1}) \\ &\propto \exp -\frac{1}{2}\Big(\big(y_t - h(x_t)\big)^T R_t^{-1}\big(y_t - h(x_t)\big) \\ &\quad + (\hat{x}_{t|t-1} - x_t)^T P_{t|t-1}^{-1}(\hat{x}_{t|t-1} - x_t)\Big), \end{aligned}$$

where terms not depending on x_t have been dropped. Since the exponential function is monotone and increasing we can minimise the negative log instead which gives

$$\begin{aligned} \hat{x}_{t|t} &= \arg\min_x \frac{1}{2}\big(y_t - h(x)\big)^T R_t^{-1}\big(y_t - h(x)\big) \\ &\quad + \frac{1}{2}(\hat{x}_{t|t-1} - x)^T P_{t|t-1}^{-1}(\hat{x}_{t|t-1} - x) \\ &= \arg\min_x \frac{1}{2} r^T(x) r(x) \\ &= \arg\min_x V(x) \end{aligned} \tag{6}$$

where

$$r(x) = \begin{bmatrix} R_t^{-1/2}\big(y - h(x)\big) \\ P_{t|t-1}^{-1/2}(\hat{x}_{t|t-1} - x) \end{bmatrix}, \tag{7}$$

and we have dropped the time dependence on the variable x. The first term in the objective function is corresponding to the unknown sensor noise and the second term corresponds to the importance of the prior to the estimate. Furthermore, the optimization problem should return a covariance approximation $\hat{P}_{t|t}$. We simply base it on the assumption that the $\hat{x}_{t|t}$ is optimal resulting in

$$\hat{P}_{t|t} = (I - K_i H_i)\hat{P}_{t|t-1}, \tag{8}$$

since it should reflect the state uncertainty when (5) has converged. A general formulation for the MAP estimation problem is then

$$\{\hat{x}_{t|t}, \hat{P}_{t|t}\} = \arg\min_x V(x) \tag{9}$$

where $\hat{P}_{t|t}$ is computed using (8) while K_i and H_i depends on the optimization method used.

III. EKF Viewed as an Optimization Problem

There is no general method for solving (6) and the nonlinear nature implies that iterative methods are needed to obtain an approximate estimate [11]. The Newton step p is the solution to the equation

$$\nabla^2 V(x) p = -\nabla V(x). \tag{10}$$

Starting from an initial guess x_0, Newton's method iterates the following equations

$$x_{i+1} = x_i - \big(\nabla^2 V(x_i)\big)^{-1} \nabla V(x_i), \tag{11}$$

where $\nabla^2 V(x_i)$ and $\nabla V(x_i)$ are the Hessian and the gradient of the cost function, respectively. Note that if $V(x)$ is quadratic then (11) gives the minimum in a single step. Gauss-Newton can be applied when the minimisation problem is on nonlinear least-squares form as in (6). Then the gradient has a simple form

$$\nabla V(x) = J^T(x) r(x) \quad \text{where} \quad J(x) = \left.\frac{\partial r(s)}{\partial s}\right|_{s=x}, \tag{12}$$

and the Hessian is approximated as

$$\nabla^2 V(x) \approx J(x)^T J(x), \tag{13}$$

avoiding the need for second-order derivatives. The GN step then becomes

$$x_{i+1} = x_i - (J_i^T J_i)^{-1} J_i^T r_i. \tag{14}$$

where $J_i = J(x_i)$ and $r_i = r(x_i)$ are introduced to ease notation. Furthermore, the Jacobian evaluates to

$$J_i = -\begin{bmatrix} R_t^{-1/2} H_i \\ P_{t|t-1}^{-1/2} \end{bmatrix} \quad \text{where} \quad H_i = \left.\frac{\partial h(s)}{\partial s}\right|_{s=x_i}, \tag{15}$$

is the measurement Jacobian.

A. Line Search

The EKF measurement update should ideally give a cost decrease

$$V(x_{i+1}) < V(x_i), \tag{16}$$

at each iteration. This will not always the case and this is why a full step in the EKF, and consequently in the IEKF, may cause problems. A remedy to this behaviour is to introduce a line search such that the estimate actually results in cost decrease. Without an explicit step-size rule the EKF should be expected to have sub-linearly convergence [13] even though it should be linear [9]. A step-size condition was proposed in [8] for smoothing passes over the data batch $0 \leq 1 - (\alpha_i)^t \leq c/t$ where $c > 0$ and $(\alpha_i)^t$ is initially small but should approach unity. There are many strategies for defining a sufficient decrease for the iteration procedure. The general iteration for Gauss-Newton with step-size parameter α is

$$x_{i+1} = x_i - \alpha_i (J_i^T J_i)^{-1} J_i^T r_i, \tag{17}$$

where each α_i is chosen as to satisfy some criterion *e.g.*, the cost decrease condition (16). For the EKF update the GN step with step-size (17) evaluates to

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + \alpha_i K\big(y_t - h(\hat{x}_{t|t-1})\big) \tag{18}$$
$$= \hat{x}_{t|t-1} + \alpha p. \tag{19}$$

The cost decrease condition is then

$$V(\hat{x}_{t|t-1} + \alpha p) < V(\hat{x}_{t|t-1}), \tag{20}$$

and it is only needed to find a suitable α over $0 < \alpha \leq 1$ starting with $\alpha = 1$. Furthermore, an *exact line search* where

$$\alpha = \arg\min_{0<s\leq 1} V(\hat{x}_{t|t-1} + sp), \tag{21}$$

is rather cheap since the search direction p is fixed. A simple line search strategy can be to multiply the current step length with some factor less than unity until sufficient decrease is obtained. More advanced methods can of course also be used such as the Wolfe conditions, see *e.g.*, [14],

$$V(x_i + \alpha_i p_i) \leq V(x_i) + c_1 \alpha \nabla V(x_i)^T p_i, \tag{22a}$$
$$\nabla V(x_i + \alpha_i p_i)^T p_i \geq c_2 \nabla V(x_i)^T p_i, \tag{22b}$$

with $0 < c_1 < c_2 < 1$, which also require that the slope has been reduced sufficiently by α_i. The *backtracking line search*, see *e.g.*, [15], is another effective inexact line search method which simply reduces the step by a factor $\alpha := \alpha\beta$ until

$$V(x + \alpha p) < V(x) + \alpha\gamma \nabla V(x)^T p, \tag{23}$$

is satisfied, starting with $\alpha = 1$, with $0 < \gamma < 0.5$ and $0 < \beta < 1$.

Algorithm 2 Iterated Extended Kalman Measurement Update

Require an initial state, $\hat{x}_{0|0}$, and an initial state covariance, $\hat{P}_{0|0}$.

1. Measurement update iterations

$$H_i = \left.\frac{\partial h(s)}{\partial s}\right|_{s=x_i} \tag{24a}$$
$$K_i = \hat{P}_{t|t-1} H_i^T (H_i \hat{P}_{t|t-1} H_i^T + R_t)^{-1} \tag{24b}$$
$$x_{i+1} = \hat{x}_{t|t-1} + K_i\big(y_t - h(x_i) - H_i(\hat{x}_{t|t-1} - x_i)\big) \tag{24c}$$

2. Update the state and the covariance

$$\hat{x}_{t|t} = x_{i+1}, \tag{25a}$$
$$\hat{P}_{t|t} = (I - K_i H_i)\hat{P}_{t|t-1}. \tag{25b}$$

B. Iterated EKF

The IEKF was derived in [9, 11, 16] and is summarised in Algorithm 2. The iterated update for the IEKF from (24c) is

$$\begin{aligned} x_{i+1} &= \hat{x} + K_i(y_t - h(x_i) - H_i(\hat{x} - x_i)) \\ &= x_i + \Big(\hat{x} - x_i + K_i\big(y_t - h(x_i) - H_i(\hat{x} - x_i)\big)\Big) \end{aligned} \tag{26}$$

where $\hat{x} = \hat{x}_{t|t-1}$ and the iterations are initialized with $x_0 = \hat{x}$. With (17) the step parameter α can be introduced. The modified measurement update of the IEKF is then on the form

$$\begin{aligned} x_{i+1} &= x_i + \alpha_i\Big(\hat{x} - x_i + K_i\big(y_t - h(x_i) - H_i(\hat{x} - x_i)\big)\Big) \\ &= x_i + \alpha_i p_i, \end{aligned} \tag{27}$$

where the step length $0 < \alpha_i \leq 1$ and the search direction p_i is chosen such that (16), or a more sophisticated convergence criterion, is satisfied in each step. Note that for $\alpha_i = 1$, the first iteration is exactly the same as for the EKF. This step-size should always be attempted first since, if it is allowed, will result in faster convergence. We will refer to this method as IEKF-L.

As a special case, the standard EKF is obtained in the first iteration of the measurement update (24). Note that the state covariance is updated using the last Kalman gain and measurement Jacobian when the iterations have finished (25) as discussed in (8).

C. Quasi-Newton EKF

Quasi-Newton type methods try to avoid exact Hessian and/or Jacobian computation for efficiency. The Hessian approximation described in (13) may be bad far from the optimum resulting in *e.g.*, slow convergence. A quite successful approach to avoid this is to introduce a correction

$$\nabla^2 V(x_i) \approx J_i^T J_i + T_i, \tag{28}$$

where T_i should mimic the dropped second-order terms. The step p with Hessian correction is the solution to the problem

$$(J^T J + T)p = -J^T r \tag{29}$$

which is iteratively solved by

$$x_{i+1} = x_i - (J_i^T J_i + T_i)^{-1} J_i^T r_i. \tag{30}$$

The state update for the *quasi-Newton EKF* (QN-EKF) becomes

$$x_{i+1} = \hat{x} + K_i^q\big((y_t - h_i - H_i\tilde{x}_i\big) - S_i^q T_i \tilde{x}_i, \tag{31a}$$

$$S_i^q = (H_i^T R_i^{-1} H_i + P^{-1} + T_i)^{-1}, \tag{31b}$$

$$K_i^q = S_i^q H_i^T R^{-1}, \tag{31c}$$

where $P = P_{t|t-1}$, $h_i = h(x_i)$, $\tilde{x}_i = \hat{x} - x_i$ have been introduced. The derivation can be found in Appendix A.

The matrix T_i can be built during the iterations [17] using

$$T_i = T_{i-1} + \frac{w_i v_i^T + v_i w_i^T}{v_i^T s_i} - \frac{w_i^T s_i}{(v_i^T s_i)^2} v_i v_i^T, \tag{32}$$

where

$$v_i = J_i^T r_i - J_{i-1}^T r_{i-1}$$
$$= -H_i^T R^{-1}(y - h_i) + H_{i-1}^T R^{-1}(y - h_{i-1}), \tag{33a}$$

$$z_i = (J_i - J_{i-1})^T r_i = (H_{i-1} - H_i)^T R^{-1}(y - h_i), \tag{33b}$$

$$s_i = x_i - x_{i-1}, \tag{33c}$$

$$w_i = z_i - T_{i-1} s_i. \tag{33d}$$

Furthermore, the quasi-Newton relation $T_i s_i = z_i$ is satisfied by construction. The iterations are initialized with $T_0 = 0$ and T_i should be replaced with $\tau_i T_i$ before the update where

$$\tau_i = \min\left(1, \frac{|s_i^T z_i|}{|s_i^T T_i s_i|}\right), \tag{34}$$

is a heuristic to ensure that T_i converges to zero when the residual vanishes.

The covariance update with this approach can be updated using (8) for the last iterate

$$\hat{P}_{t|t} = (I - K_i^q H_i)\hat{P}_{t|t-1}. \tag{35}$$

An explicit line-search parameter can further be introduced

$$x_{i+1} = x_i + \alpha_i\Big(\tilde{x}_i + K_i^q\big((y_t - h_i - H_i\tilde{x}_i\big) - S_i^q T_i \tilde{x}_i\Big). \tag{36}$$

using the same argumentation as in Section III-B.

D. Levenberg-Marquardt Iterated EKF

Trust region methods for GN are similar to the Hessian correction above in the way that the Newton step is computed. The main difference being that only a single parameter is adapted at runtime. The well-known Levenberg-Marquardt method introduced by [18, 19] is such a method and it is obtained by adding a damping parameter μ to the GN problem

$$(J^T J + \mu I)p = -J^T r. \tag{37}$$

Applying this to (14) results in

$$x_{i+1} = \hat{x} + K_i\big((y_t - h_i - H_i\tilde{x}_i\big) - \mu_i(I - K_i H_i)\tilde{P}\tilde{x}_i, \tag{38a}$$

$$\tilde{P} = \left(I - P\Big(P + \frac{1}{\mu_i} I\Big)^{-1}\right)P, \tag{38b}$$

$$K_i = \tilde{P} H_i^T (H_i \tilde{P} H_i^T + R)^{-1}, \tag{38c}$$

which is referred to as *Levenberg-Marquardt IEKF* (LM-IEKF). The damping parameter works as an interpolation between GN and steepest-descent when μ is small, and large, respectively. The parameter also affects the step-size, where large μ means shorter steps reflecting the size of the trust-region and it should be adapted at each iteration. The relations in (38) were derived in [12] but, just as in [20], it was never indicated how μ is selected, in fact, they never seem to perform any iterations. Therefore their solution rather works as Tikhonov regularization for the linearized problem with fixed μ. In [12, Example 2] the damping have to be selected as $\mu = \operatorname{cond}(J^T J) \approx 10^6$ in order to obtain roughly the same update. The update, counterintuitively, has a larger cost, V, than the EKF update but a lower RMSE. This example is extremely ill-conditioned and the Hessian approximation is thus not very useful. Furthermore, the condition number can serve as a simple heuristic for selecting the damping automatically resulting in $\operatorname{cond}(J^T J + \mu^2 I) \approx 1$. Another option, due to [19], is to start with some initial $\mu = \mu_0$ (typically $\mu_0 = 0.01$) and factor $v > 1$. The factor is used to either increase or decrease the damping if the cost is increased or decrease, respectively.

The LM-IEKF can be improved by replacing the diagonal damping with $\mu \operatorname{diag}(J^T J)$ in (37) as suggested by [19]. By doing so, larger steps are made in directions where the gradient is small, further speeding up convergence. The resulting equations for the problem (6) then becomes

$$x_{i+1} = \hat{x} + K_i\big((y_t - h_i - H_i\tilde{x}_i\big)$$
$$- \mu_i(I - K_i H_i)\tilde{P} B_i \tilde{x}_i, \tag{39a}$$

$$\tilde{P} = \left(I - P\Big(P + \frac{1}{\mu_i} B_i^{-1}\Big)^{-1}\right)P, \tag{39b}$$

$$K_i = \tilde{P} H_i^T (H_i \tilde{P} H_i^T + R)^{-1}, \tag{39c}$$

$$B_i = \operatorname{diag}(J_i^T J_i) = \operatorname{diag}(H_i^T R^{-1} H_i + P^{-1}), \tag{39d}$$

which is shown in Appendix B.

IV. Results

The filters are evaluated in two simulations highlighting the effect of line search and the overall better performance obtained with the derived optimisation based methods in a bearings only tracking problem.

A. Line Search

To illustrate the effect of using line search in EKF and IEKF, consider the measurement function

$$h(x) = -x^2. \tag{40}$$

This result in a cost function with two minima in $\pm\sqrt{|y|}$ and a local maximum at $x = 0$ which makes the problem difficult.

Fig. 1 shows the estimates obtained with the EKF and IEKF initialized with $x_0 = 0.1$, $R = 0.1$ and $P = 1$ where $V(x_0) = 4.90$, while the true state is $x^* = 1$, and a perfect measurement $y = -1$ is given. In the IEKF the step length is reduced to $\alpha = 0.5$ for the first iteration resulting in $x_1^{\text{IEKF}} = 0.807$, $V(x_1^{\text{IEKF}}) = 0.860$ while the EKF (equivalent to the IEKF

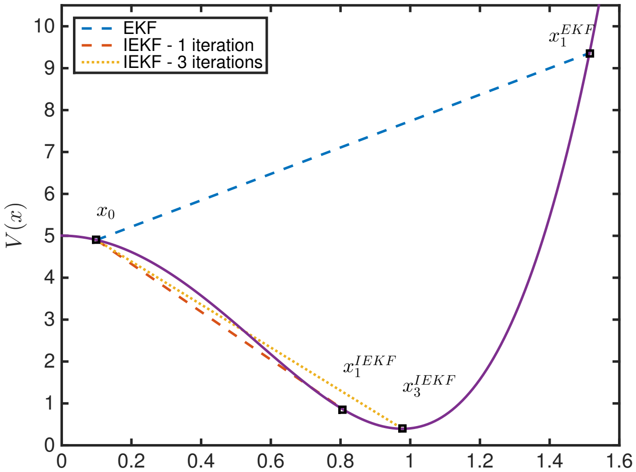

Fig. 1. EKF with full step increases the cost while half the step-size results in a reduced cost. Remember that EKF and IEKF are identical in the first step. Three IEKF iterations with line search significantly reduces the cost and further improves the estimate.

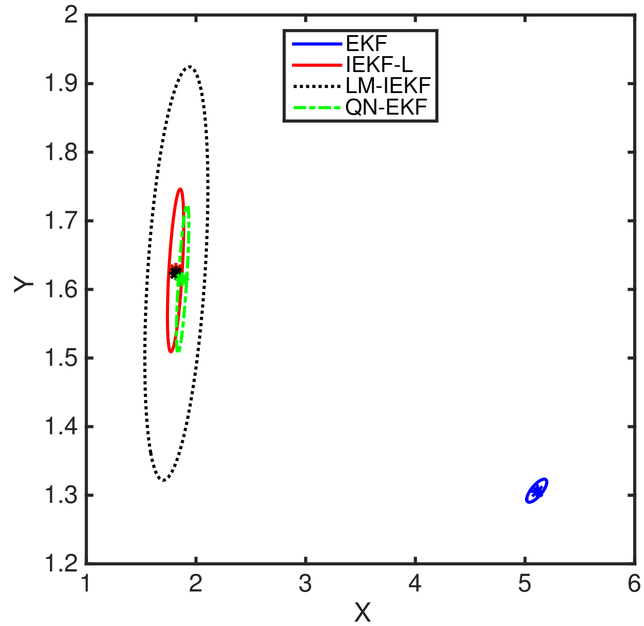

Fig. 2. All filters except for the EKF are iterated 5 times with line search and come closer to the true state with a covariance that captures the uncertainty.

without step-length adaption) yields $x_1^{\text{EKF}} = 1.51$, $V(x_1^{\text{EKF}}) = 9.36$. Another two iterations in the IEKF results in $x_3^{\text{IEKF}} = 0.978$, $V(x_3^{\text{IEKF}}) = 0.395$. Two things can be noted from this simple example. First, the adaptive step-length improves the EKF noticeably at a relatively low cost. Second, the iterations in the IEKF further improves the estimate.

Furthermore, with a smaller measurement covariance *e.g.*, $R = 0.01$ the EKF performance degrades even more giving $x_1^{\text{EKF}} = 4.01$ (not shown in Fig. 1), while the IEKF performs well. This illustrates that EKF struggles with large residuals as the linearisation point differs much from its optimal value.

B. Bearings Only Tracking

The next example, a bearings-only tracking problem involving several sensors, which is also studied in [21, Ch. 5] and [22, Ch. 4], is used to illustrate IEKF-L, QN-EKF and LM-IEKF. In a first stage the target with state $x = [\mathsf{X}, \mathsf{Y}]^T$ is stationary, *i.e.*, $w_t = 0$, at the true position $x^* = [1.5, 1.5]^T$. The bearing measurement function from the j-th sensor S^j at time t is

$$y_t^j = h_j(x_t) + e_t = \arctan2(\mathsf{Y}_t - S_{\mathsf{Y}}^j, \mathsf{X}_t - S_{\mathsf{X}}^j) + e_t, \quad (41)$$

where $\arctan2()$ is the two argument arctangent function, S_{Y} and S_{X} denotes the Y and the X coordinates of the sensors, respectively. The noise is $e_t \sim \mathcal{N}(0, R_t)$. The Jacobians of the dynamics and measurements are

$$F = I, \quad H^j = \frac{1}{(\mathsf{X} - S_{\mathsf{X}}^j)^2 + (\mathsf{Y} - S_{\mathsf{Y}}^j)^2} \begin{bmatrix} -(\mathsf{Y} - S_{\mathsf{Y}}^j) \\ \mathsf{X} - S_{\mathsf{X}}^j \end{bmatrix}. \quad (42)$$

With the two sensors having positions $S^1 = [0, 0]^T$ and $S^2 = [1.5, 0]^T$. The filters are initialised with $\hat{x}_{0|0} = [0.5, 0.1]^T$, $P_{0|0} = 0.1 I_2$, $R = \pi^2 10^{-5} I_2$ and $Q = 0.1 I_2$. Fig. 2 shows the four filters posteriors with $2 - \sigma$ covariances after a single perfect measurement and 5 iterations. All filters, except for the EKF, result in acceptable estimates of the true target state and its uncertainty. After 10 iterations IEKF-L, LM-IEKF and QN-EKF produce nearly identical results.

TABLE I
MONTE CARLO SIMULATION RESULTS FOR THE FOUR FILTERS.

Method	EKF	IEKF-L	LM-IEKF	QN-EKF
Mean-RMSE	0.45	0.19	0.19	0.24

For the final example 10 iterations are used to ensure the all filters have converged (less would have sufficed) and the result are clearly improved estimates compared to the EKF. A small Monte Carlo study is done with 10 trajectories generated by $w_t \sim \mathcal{N}(0, 0.1 I_2)$, $e_t \sim \mathcal{N}\left(0, \pi^2 10^{-5} I_2\right)$ for 20 time steps. The filters are initialized with the true position $\hat{x}_{0|0} = [1.5, 1.5]^T$ with $P_{0|0} = 0.1 I_2$. The results are summarized in Table I where it can be seen again that all filters perform better than EKF with IEKF-L and LM-IEKF being slightly better than QN-EKF.

V. CONCLUSIONS

The cost function used in the *extended Kalman filter* (EKF) has been studied from an optimisation point of view. Applying optimisation techniques to the cost function motivates using an adaptive step length in the EKF and three iterated EKF variations are derived. In simulations it is shown that these filters outperform the standard EKF providing more accurate and consistent results. It remains to investigate the convergence rates obtained this way, as well as how to incorporate other EKF improvements such as the higher order EKF, square-root implementations for improved numerical stability, and how to deal with colored noise.

ACKNOWLEDGMENT

This work was funded by the Vinnova Industry Excellence Center LINK-SIC together with the Swedish Research Council through the grant Scalable Kalman Filters and the Swedish strategic research center Security Link.

APPENDIX A
DERIVATION OF THE QUASI-NEWTON EKF

Using (7), (15) gives a simplified notation

$$r_i = \begin{bmatrix} R^{-1/2}(y - h_i) \\ P^{-1/2}(\hat{x} - x_i) \end{bmatrix}, \tag{43}$$

$$J_i = -\begin{bmatrix} R^{-1/2}H_i \\ P^{-1/2} \end{bmatrix}. \tag{44}$$

The right-hand side of (30) evaluates to

$$\begin{aligned} & x_i - (J_i^T J_i + T_i)^{-1} J_i^T r_i = \\ & = x_i + (H_i^T R^{-1} H_i + P^{-1} + T_i)^{-1} \times \\ & \qquad \big(H_i^T R^{-1}(y - h_i) + P^{-1}(\hat{x} - x_i)\big) \\ & = \hat{x} + (H_i^T R^{-1} H_i + P^{-1} + T_i)^{-1} \times \\ & \qquad \big(H_i^T R^{-1}(y - h_i - H_i(\hat{x} - x_i)) - T_i(\hat{x} - x_i)\big) \\ & = \hat{x} + S_i^q H_i^T R^{-1}\big(y - h_i - H_i(\hat{x} - x_i)\big) - S_i^q T_i(\hat{x} - x_i) \\ & \qquad = \hat{x} + K_i^q(y_t - h_i - H_i\tilde{x}_i) - S_i^q T_i \tilde{x}_i, \end{aligned} \tag{45}$$

which is the expression in (31) with $\tilde{x}_i = \hat{x} - x_i$.

APPENDIX B
DERIVATION OF THE MODIFIED LM-IEKF

The LM-IEKF with modified step is evaluated using (37), (14), (43) and (44) as

$$\begin{aligned} & x_i - (J_i^T J_i + \mu_i B_i)^{-1} J_i^T r_i = \\ & = x_i + (H_i^T R^{-1} H_i + P^{-1} + \mu_i B_i)^{-1} \times \\ & \qquad \big(H_i^T R^{-1}(y - h_i) + P^{-1}(\hat{x} - x_i)\big) \\ & = \hat{x} + (H_i^T R^{-1} H_i + P^{-1} + \mu_i B_i)^{-1} \times \\ & \qquad \big(H_i^T R^{-1}(y - h_i - H_i\tilde{x}_i) - \mu_i B_i \tilde{x}_i\big) \\ & = \hat{x} + K_i(y_t - h_i - H_i\tilde{x}_i) - \mu_i(I - K_i H_i)\tilde{P} B_i \tilde{x}_i, \end{aligned} \tag{46}$$

which is the expression in (39). Where we have used the relations

$$\begin{aligned} K_i &= (H_i^T R^{-1} H_i + \tilde{P}^{-1})^{-1} H_i^T R^{-1} \\ &= \tilde{P} H_i^T (H_i \tilde{P} H_i^T + R)^{-1}, \end{aligned} \tag{47}$$

$$(H_i^T R^{-1} H_i + \tilde{P}^{-1})^{-1} = (I - K_i H_i)\tilde{P}, \tag{48}$$

and

$$\begin{aligned} \tilde{P} &= (P^{-1} + \mu_i B_i)^{-1} \\ &= \left(I - P\Big(P + \frac{1}{\mu_i} B_i^{-1}\Big)^{-1}\right) P. \end{aligned} \tag{49}$$

REFERENCES

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

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, ser. Mathematics in Science and Engineering. Academic Press, Inc, 1970, vol. 64.

[3] N. U. Ahmed and S. M. Radaideh, "Modified extended Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1322–1326, Jun. 1994.

[4] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, Mar 2000.

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

[6] F. Gustafsson and G. Hendeby, "Some relations between extended and unscented Kalman filters," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 545–555, Feb. 2012.

[7] G. A. Einicke and L. B. White, "Robust extended Kalman filtering," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2596–2599, Sep 1999.

[8] D. Bertsekas, "Incremental Least Squares Methods and the Extended Kalman Filter," in *Decision and Control, 1994., Proceedings of the 33rd IEEE Conference on*, vol. 2, Lake Buena Vista, Florida, USA, dec 1994, pp. 1211–1214.

[9] P. J. G. Teunissen, "On the local convergence of the iterated extended kalman filter," in *Proceeding of the XX General Assembly of the IUGG, IAG Section IV*. Vienna: IAG, Aug. 1991, pp. 177–184.

[10] G. K. Smyth, *Encyclopedia of Environmetrics*. John Wiley & Sons, Ltd, 2006. [Online]. Available: http://dx.doi.org/10.1002/9780470057339.vao011

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

[12] R. L. Bellaire, E. W. Kamen, and S. M. Zabin, "New nonlinear iterated filter with applications to target tracking," in *SPIE Signal and Data Processing of Small Targets*, vol. 2561, San Diego, CA, USA, Sep. 1 1995, pp. 240–251.

[13] H. Moriyama, N. Yamashita, and M. Fukushima, "The incremental Gauss-Newton algorithm with adaptive stepsize rule," *Computational Optimization and Applications*, vol. 26, no. 2, pp. 107–141, 2003.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] Y. Bar-Shalom and X.-R. Li, *Estimation and tracking: principles, techniques and software*. Boston: Artech House, 1993.

[17] J. E. Dennis, D. M. Gay, and R. E. Welsh, "Algorithm 573 — NL2SOL, An adaptive nonlinear least-squares algorithm," *ACM Transactions on Mathematical Software*, vol. 7, pp. 348—-368, 1981.

[18] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly Journal of Applied Mathmatics*, vol. II, no. 2, pp. 164–168, 1944.

[19] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[20] M. Fatemi, L. Svensson, L. Hammarstrand, and M. Morlande, "A study of MAP estimation techniques for nonlinear filtering," in *Proceedings of International conference on information fusion (FUSION)*, 2013.

[21] G. Hendeby, "Performance and implementation aspects of nonlinear filtering," Dissertations No 1161, Linköping Studies in Science and Technology, Mar. 2008.

[22] ——, "Fundamental estimation and detection limits in linear non-Gaussian systems," Licentiate Thesis No 1199, Department of Electrical Engineering, Linköpings universitet, Sweden, Nov. 2005.