A Systematic Approach for Kalman-type Filtering with non-Gaussian Noises

Matti Raitoharju, Robert Piché, and Henri Nurminen
Department of Automation Science and Engineering
Tampere University of Technology
Tampere, Finland
Emails: {matti.raitoharju, robert.piche, henri.nurminen}@tut.fi

***Abstract*—For nonlinear systems there exist several Kalman filter extensions that linearize or do moment matching to approximate the nonlinear update. These algorithms usually assume Gaussian measurement noises. The assumption of Gaussian noises degrades the performance when the data contain outliers or are otherwise non-Gaussian. In this paper, we present a new way of treating non-Gaussian noises in a Kalman-type filter. We propose to model non-Gaussian noise as a non-linear transformation of a Gaussian noise, and we develop an algorithm for estimation with this kind of models. Results show that the proposed algorithm can achieve similar estimation accuracy as state-of-the-art methods designed for a specific distribution. However, with some models the estimate diverges and there is still work to do in the development of a suitable Kalman filter extension.**

I. Introduction

Kalman-type filters are commonly used in state estimation of dynamical systems. The original Kalman Filter (KF) [1] can be used with linear-Gaussian state and measurement models. KF has been extended to be able to handle nonlinear measurement and state transition functions. In real-world situations, measurement models are often nonlinear and measurement noises non-Gaussian. Many Kalman-type nonlinear filtering algorithms, such as Unscented Kalman Filter (UKF) [2], can be seen as approximations of the Gaussian filtering paradigm [3]. However, as shown in [4] in Gaussian Filter (GF) a measurement model with non-Gaussian noise will be treated exactly as a Gaussian noise with the same mean and covariance. Thus, the same approach cannot perform well with non-Gaussian noises.

For Kalman-type filtering with non-Gaussian noises a score function approach is proposed in [5]. The score function weights the measurements based on the predicted measurements. The method involves computing a convolution of two probability density functions (PDFs), which can be done analytically only in special cases. The score approach has been further developed in [6], which provides methods for computing approximate scores by expanding the non-Gaussian PDF as a product of a standard normal PDF and Hermite polynomials. The score algorithms are based on computing the residual between realized measurement and expected measurement value. For heavy-tailed noise the update weight is reduced when the residual is large.

In [4] the measurement noise is assumed to be a sum of a Gaussian and a heavy tailed random variable. The algorithm generates a virtual measurement that uses weights based on the PDFs of the Gaussian and heavy tailed distributions.

In [7] an Expectation–Maximization algorithm for robust filtering is presented. For Student's t-distribution [8] and [9] present algorithms based on variational Bayes. These algorithms can be interpreted as iterative algorithms that increase the innovation covariance when measurements are unlikely and they are designed for heavy tailed noise distributions.

The above-presented methods approximate the state distribution after each time step as a Gaussian. Particle Filters (PFs) [10] and Gaussian Mixture Filters (GMFs) [11] can be used in nonlinear and non-Gaussian estimation. PFs use a set of weighted points ("particles") to approximate the state distribution and the application of non-Gaussian noise models is easy; GMFs use sums of Gaussians. Both these approaches have the problem that the number of particles and components has to be set large enough so that the estimation accuracy is good enough, but not so high that the computation time is unacceptable.

In this paper, we present an approach for estimation problems with measurements with additive non-Gaussian noises. The proposed approach can be applied to noise models whose inverse cumulative distribution function (CDF) can be determined or easily approximated. The state distribution is approximated as a Gaussian between the measurement updates and state propagation.

II. Modeling non-Gaussian noises

We consider measurement models of the form

$$y = f(x) + \tilde{\varepsilon}, \tag{1}$$

where x is the d-dimensional state, y is the measurement value, f is the measurement function and $\tilde{\varepsilon}$ is the measurement noise, which is not necessarily normal distributed. The measurements are used in a Bayesian setting to update a Gaussian prior with mean μ^- and covariance P^- to a posterior with mean μ^+ and covariance P^+. We propose to model the noise in (1) as a nonlinear transformation of a standard normal distributed noise ε

$$\tilde{\varepsilon} = g(\varepsilon). \tag{2}$$

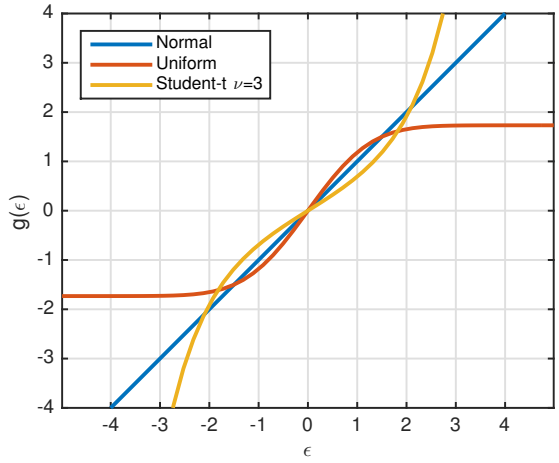

Fig. 1. $g(\varepsilon)$ for normal, uniform and Student-t distributions

Thus, the problem is transformed from filtering with non-Gaussian noise to nonlinear filtering with Gaussian noise.

For a scalar measurement, if the CDF F of the distribution of the noise is invertible, the transforming function $g(\cdot)$ can be obtained using the probability integral transform (aka inversion principle aka quantile transform) as

$$g(\varepsilon) = F^{-1}(\Phi(\varepsilon)), \tag{3}$$

where $\Phi(\cdot)$ is the CDF of a standard normal distribution and the ε has a standard normal distribution. The transformation (3) holds also in the more general case of discontinuous and not strictly increasing CDF by letting F^{-1} denote the quantile function

$$F^{-1}(y) = \inf\{x : F(x) \geq y\}, 0 < y < 1. \tag{4}$$

For independent univariate noise elements, the transformations can be made independently for each noise element. We consider only such measurements in this paper. However, transformations exist also for dependent noises. One possible approach for dependent noises is to use the Knothe–Rosenblatt rearrangement [12], [13].

Figure 1 shows the transformations $g(\varepsilon)$ for normal, uniform and Student-t distributions. All distributions are scaled so that they have zero mean and unit variance. The Student-t distribution has 3 degrees of freedom. The Student-t distribution has heavier tails than the normal distribution. In Figure 1 this can be seen from the steep slope of the function when ε is not close to zero. The uniform distribution does not have any tails and so the function range is limited to $(-\sqrt{3}, \sqrt{3})$.

If a GF is applied to measurement model (1–2), the linearization of $g(\cdot)$ defined by (3) is made at zero because of the standard normal prior for ε, and so the measurement noise could be also replaced in (1) with a normal distributed noise. This naturally applies also to the approximations of the GF, such as UKF.

In [14], the Posterior Linearization Filter (PLF) is presented. Compared to the GF the PLF does the linearization in the posterior and in [14] it is argued that this produces better estimates, because it reduces the joint Kullback-Leibler Divergence (KLD) of the posterior and measurement. The linearization of the nonlinearly transformed noise depends on the posterior and, thus, the approach that uses nonlinear transformations can be used. The computation of the PLF is intractable, but an iterative algorithm is given to approximate it. However, in [14] only the situation with additive Gaussian noise is considered. Using augmented state, i.e. a state where also the noise terms are updated in iterations without an additive Gaussian noise term in the PLF, may produce singular matrices and their inverses, so the algorithm is not feasible for this problem.

The Iterated Extended Kalman Filter (IEKF) [15] would be an option for the filtering, but it requires analytical differentiation of the nonlinear functions and it is known to diverge in certain situations.

To our knowledge the literature does not have an algorithm that would make the linearizations iteratively for the augmented state and that can be used with sigma-point filters, such as UKF. We will extend and use an iterative algorithm called Recursive Update Filter (RUF) [16] for this purpose. The main idea of RUF is to process a measurement in a sequence of update steps using down-weighted Kalman gains. Each down-weighted Kalman gain changes the state estimate less than the full update would. Between each partial measurement update, the filter takes the correlation of the partly updated state and measurement noise into account.

III. RUF and its generalization

The RUF algorithms in [16], [17] use measurement models with additive Gaussian noise

$$y = f(x) + \varepsilon, \tag{5}$$

where ε is assumed to be zero mean Gaussian with covariance $P_{\varepsilon\varepsilon}$. RUF is modified to be used with sigma-point filters (Sigma-Point Recursive Update Filter (SPRUF)) in [17]. The SPRUF algorithm [17] is given in Algorithm 1.

The use of SPRUF is not limited to sigma-point filters. It can be used with any filter that computes the predicted measurement value $\hat{\mu}_y$, innovation covariance $P_{f(x)f(x)}$, and cross covariance between state and measurement $P_{f(x)x}$. In the case of KF or Extended Kalman Filter (EKF) that uses measurements of the form (5) these variables are: $\hat{\mu}_y = f(\mu)$, $P_{f(x)x} = JP_{xx}$ and $P_{f(x)f(x)} = JP_{xx}J^T + P_{\varepsilon\varepsilon}$, where J is the Jacobian $J = \frac{\partial f(x)}{\partial x}\Big|_{x=\mu}$. When these variables are substituted into Algorithm 1 the inverses P_{xx}^{-1} can be eliminated and the algorithm is equivalent to the original RUF algorithm [16]. The estimate of the original RUF algorithm is identical to the estimate obtained with EKF when the number of steps N is 1.

Next we generalize SPRUF for estimation with non-additive noises. Let z denote the augmented state defined as $z = \left[\begin{smallmatrix} x \\ \varepsilon \end{smallmatrix}\right]$.

input : $\mu_{\hat{x}}$ – predicted mean, P_{xx} – predicted covariance, y – measurement value, $f(\cdot)$ – measurement function, $P_{\varepsilon\varepsilon}$ – measurement noise covariance, N – number of steps

output : μ_x^+ – posterior mean, P_{xx}^+ – posterior covariance

$P_{x\varepsilon} \leftarrow \mathbf{0}$ // Correlation
$\mu \leftarrow \mu_{\hat{x}}$ // Assign predicted mean
$P \leftarrow P_{xx}$ // Assign predicted covariance
$i \leftarrow 0$ // Current step index
while $i \leq N$ **do**
  Use a KF extension algorithm to compute:
    Predicted measurement $\hat{\mu}_y$
    Measurement function covariance $P_{f(x)f(x)}$
    Cross covariance of the state and the predicted measurement $P_{f(x)x}$
  $P_{f(x)f(x)} \leftarrow P_{f(x)f(x)} + P_{\varepsilon\varepsilon} + P_{f(x)x}P_{xx}^{-1}P_{x\varepsilon} + \left(P_{f(x)x}P^{-1}P_{x\varepsilon}\right)^T$ // Innovation covariance
  $\gamma \leftarrow \frac{1}{N+1-i}$ // Update weight
  $K \leftarrow \gamma\left(P_{f(x)x}^T + P_{x\varepsilon}\right)P_{f(x)f(x)}^{-1}$ // Kalman gain with reduced weight
  $\mu \leftarrow \mu + K\left(y - \hat{\mu}_y\right)$ // Updated state mean
  $P_{x\varepsilon} \leftarrow P_{x\varepsilon} - K\left(P_{\varepsilon\varepsilon} + P_{f(x)x}P^{-1}P_{x\varepsilon}\right)$ // Updated correlation
  $P \leftarrow P + \left(1 - \frac{2}{\gamma}\right)KP_{f(x)f(x)}K^T$ // Updated covariance
  $i \leftarrow i + 1$
end
$\mu_x^+ \leftarrow \mu$
$P_{xx}^+ \leftarrow P$

Algorithm 1: Update step of the SPRUF

When using the augmented state formulation, the noise term is included in the state and the separate computation of correlation $P_{x\varepsilon}$ used in SPRUF is not required. The augmented measurement model is

$$h(z) = f(x) + g(\varepsilon). \tag{6}$$

One iteration of SPRUF becomes a Generalized Recursive Update Filter (GRUF)-iteration:

$$K \leftarrow \gamma P_{zh(z)}P_{h(z)h(z)}^{-1} \tag{7}$$

$$\mu_z \leftarrow \mu_z + K\left(y - \hat{y}\right) \tag{8}$$

$$P_{zz} \leftarrow P_{zz} + \left(1 - \frac{2}{\gamma}\right)KP_{h(z)h(z)}K^T. \tag{9}$$

After making all iterations of the GRUF-update, the state variable's mean and covariance are extracted from the augmented state's mean and covariance and the noise variables are discarded. Note also that after the last iteration when $\gamma = 1$ the augmented covariance may become singular. If one uses a very large number of iterations the augmented covariance may become nearly singular before the end of the update loop. However, the state part is singular only if the measurement noise covariance is not full rank. So the state posterior covariance may become singular only in same situations where any GF would produce a singular covariance matrix for the state. The use of augmented state allows the predicted state and measurement noise to be correlated i.e. P_{zz} does not need to be block diagonal. The GRUF algorithm is given in Algorithm 2.

input : $\mu_{\hat{x}}$ – prior mean, P_{xx} – prior covariance, y – measurement value, $h(\cdot)$ – measurement function for augmented state, μ_ε – mean of measurement noise, $P_{\varepsilon\varepsilon}$ – measurement noise covariance, $P_{x\varepsilon}$ – cross covariance between state and measurement noise N - Number of steps

output : μ_x^+ – posterior mean, P_{xx}^+ – posterior covariance

$\mu_z \leftarrow \begin{bmatrix}\mu_{\hat{x}} \\ \mu_\varepsilon\end{bmatrix}$ // Set augmented mean
$P_{zz} \leftarrow \begin{bmatrix}P_{xx} & P_{x\varepsilon} \\ P_{x\varepsilon}^T & P_{\varepsilon\varepsilon}\end{bmatrix}$ // Set augmented covariance
$i \leftarrow 0$ // Current progress
while $i < N$ **do**
  Use a KF extension algorithm to compute:
    Predicted measurement $\hat{\mu}_y$
    Innovation covariance $P_{h(z)h(z)}$
    Cross covariance of the augmented state and the predicted measurement $P_{zh(z)}$
  $\gamma \leftarrow \frac{1}{N+1-i}$
  $K \leftarrow \gamma P_{zh(z)}P_{h(z)h(z)}^{-1}$ // Reduced Kalman gain
  $\mu_z \leftarrow \mu_z + K\left(y - \hat{\mu}_y\right)$ // Update the augmented state mean
  $P_{zz} \leftarrow P_{zz} + \left(1 - \frac{2}{\gamma}\right)KP_{h(z)h(z)}K^T$ // Update the augmented state covariance
  $i \leftarrow i + 1$
end
$\mu_x^+ \leftarrow \mu_{z[1:d]}$ // Extract the posterior mean from the augmented mean
$P_{xx}^+ \leftarrow P_{zz[1:d,1:d]}$ // Extract the posterior covariance from the augmented covariance

Algorithm 2: Update step of the GRUF

In addition to the already mentioned EKF and UKF, the covariances in this algorithm can be computed with any appropriate Kalman filter extension. When using only one iteration, the GRUF produces identical results with the KF extension that was used for computing the covariances.

Algorithm 1 contains the inverse of the state covariance, which may become the computationally heaviest part of an iteration. In our proposed filter (Algorithm 2) the inverse is not needed and the computational complexity of the algorithm

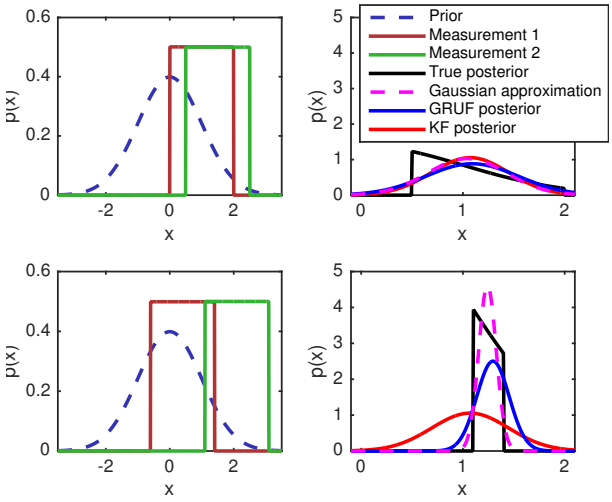

Fig. 2. Update of standard normal prior with 2 uniformly distributed measurements

may be smaller than with the previous algorithm in situations where the state dimension is high. As shown earlier the inverse of the state covariance is not needed when using EKF or other filter that computes the Jacobian.

For RUF an adaptive method for choosing the number of steps is presented in [18]. It can be adapted to be used with GRUF. We use only the pre-defined number of steps in this paper for brevity.

IV. Examples

In this section, we present five examples. We use UKF for computing the moments in GRUF.

A. Uniform noise

In our first example, we consider uniformly distributed noise

$$\tilde{\varepsilon} \sim \mathrm{U}(-1, 1). \tag{10}$$

The inverse CDF is

$$F^{-1}(x) = 2x - 1. \tag{11}$$

We consider a one-dimensional state and two measurements. The measurement model is (1-2) with

$$f(x) = \begin{bmatrix} x \\ x \end{bmatrix} \tag{12}$$

$$g(\varepsilon) = \begin{bmatrix} 2\Phi(\varepsilon_1) - 1 \\ 2\Phi(\varepsilon_2) - 1 \end{bmatrix}, \tag{13}$$

where Φ is the CDF of the standard normal distribution.

Figure 2 shows two cases of a standard normal prior updated using model (12-13). In the top row the measurement vector is $y = \begin{bmatrix} 1 & 1.5 \end{bmatrix}^T$. The left plot shows the prior and the likelihood functions of the two measurements. The right plot shows the true posterior, the posterior obtained with GRUF, the posterior obtained with KF that approximates the noise as a Gaussian, and the Gaussian that has the same mean and covariance as the true posterior. In this example all Gaussian posterior estimates are similar.

In the bottom row the measurement vector is $y = \begin{bmatrix} 0.4 & 2.1 \end{bmatrix}^T$ and the support interval of the joint measurement likelihood is considerably smaller than in the top row case and, thus, the posterior variance is smaller. The KF estimate is exactly the same as in the first case, and its variance is much larger than the true posterior variance. The GRUF estimate's variance is smaller than in first case, and is only slightly larger than the true variance.

B. Range measurements with outliers

In our second example we consider range measurements:

$$f(x) = \|x - r_i\| + \tilde{\varepsilon} \tag{14}$$

Figure 3 shows 4 such measurements one of them having a large error considered to be a Non-Line-of-Sight (NLOS) measurement while 3 other measurements are Line-of-Sight (LOS). The left plot shows the update using UKF, the middle plot uses GRUF with normal noise having variance 3 and the right plot uses Student-t distributed noise with 3 degrees of freedom and the same variance. The inverse CDF for a Student-t does not have a simple analytic form, but efficient numerical algorithms are available. The figure shows how GRUF with Student-t noise has the estimate closest to the true location.

C. 1-D Student-t distributed noise

In our third example, we test the accuracy of the proposed filter with Student-t noise and compare it with other algorithms found in the literature. We consider the measurement model

$$y = x + \tilde{\varepsilon}, \tag{15}$$

where $x \sim \mathrm{N}(0, \sigma^2_{x^-})$ and $\tilde{\varepsilon}$ is Student-t distributed with 3 degrees of freedom. The variance of $\tilde{\varepsilon}_t$ is 3. We chose a linear measurement model for the test, so that the algorithm's ability to handle nonlinear measurements does not affect the result.

We compute the true posterior PDF using a dense grid. For comparison we use a normal distribution with the same mean and covariance as the true posterior and estimates computed with KF and two algorithms designed specifically for Student-t errors: Recursive Outlier Robust Filter (RORF) [8] and Outlier Robust Kalman Filter (ORKF) [9].

The estimate accuracies are evaluated using the Kullback-Leibler (KL) divergence of approximate posteriors with respect to the true posterior computed using the grid. Table I shows the mean KL divergences of the estimates obtained with different methods in 1000 tests of measurement update, with varying prior variances. Column Ref. gives the KL divergence of the Gaussian with same mean and covariance as the true posterior and KF gives the KL divergence when the measurements are updated using KF. The table shows how GRUF is better than the Student-t specific methods when the prior variance is large. When the prior variance is smaller the GRUF is not as good as RORF. ORKF is the worst of the three robust methods.

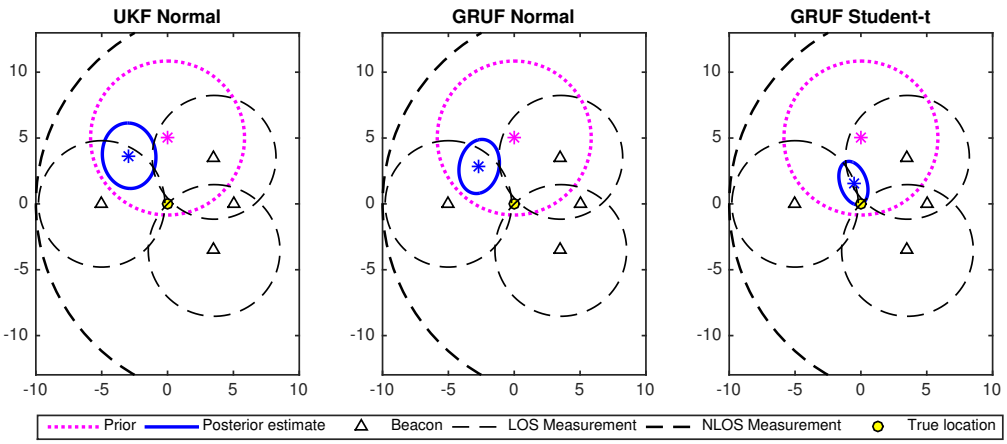

Fig. 3. Examples of updates with range measurements and one outlier. Estimate ellipses contain 68% of the probability.

TABLE I
KULLBACK-LEIBLER DIVERGENCES IN STUDENT-t TEST

	σ^2_{x-}	Ref.	KF	GRUF	RORF	ORKF
KL	0.1	0.0000	0.04	0.04	0.0004	0.02
	1	0.005	0.2	0.08	0.02	0.1
	10	0.06	0.4	0.2	0.3	0.4

TABLE II
MEAN RMS ERRORS IN THE FILTERING TEST

	RORF	GRUF $N = 5$	GRUF $N = 50$	KF
mean RMS x_1	4.8	5.4	21.1	6.6
mean RMS x_x	2.0	2.1	2.8	2.8

D. Filtering example

In our fourth example we consider a simple filtering problem with two-dimensional state with state transition model

$$x_{t+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_t + \varepsilon_Q, \tag{16}$$

where $\varepsilon_Q \sim \mathrm{N}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right]\right)$ and $x_0 \sim \mathrm{N}\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}40&0\\0&4\end{smallmatrix}\right]\right)$. The measurements are noisy observations of the first state variable

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t + \tilde{\varepsilon}_t, \tag{17}$$

where $\tilde{\varepsilon}_t$ is again Student-t distributed with 3 degrees of freedom scaled with $\sqrt{\frac{100}{3}}$. The variance of $\tilde{\varepsilon}_t$ is thus 10. We simulated 50 sequences, each 50 time steps long.

Table II shows mean RMS errors of the filtering tests. The results show that GRUF with 5 iterations is slightly worse in accuracy than RORF, but has better accuracy than KF. However, GRUF with 50 iterations has large errors.

Figure 4 shows the evolution of residuals of the first state variable in one simulation. There are two clear outliers, one at time step 24 and another at 34. In both these situations KF reacts to the outlier more than RORF or GRUF with 5 iterations. GRUF with 50 iterations does not react to any measurements clearly. The reason for poor GRUF performance

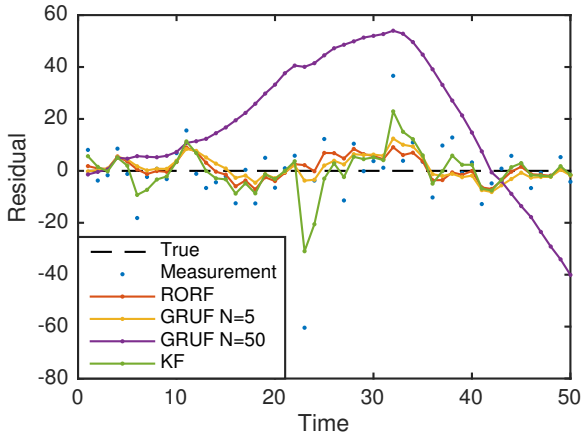

Fig. 4. Residuals of the first state variable on one filtering run

with many iterations seems to be the correlations between variables. The down-weighted updates store the information of the linear dependence of state and noise variables. Because the model is nonlinear, this causes problems in the estimate. We have found similar problems happening with RUF and SPRUF, as illustrated in the next example.

E. Bearings update example

Figure 5 shows an update of a prior using a bearings measurement

$$y = \angle x + \varepsilon. \tag{18}$$

In this example, we used normal distributed ε and, thus, it can be solved with the original EKF-based RUF (red dashed line) and UKF-based SPRUF (blue line).

When the number of iterations is increased, the posterior estimate moves close to the beacon. This is also caused by the correlations of nonlinearly evolving variables that was the cause of problems in the previous example.

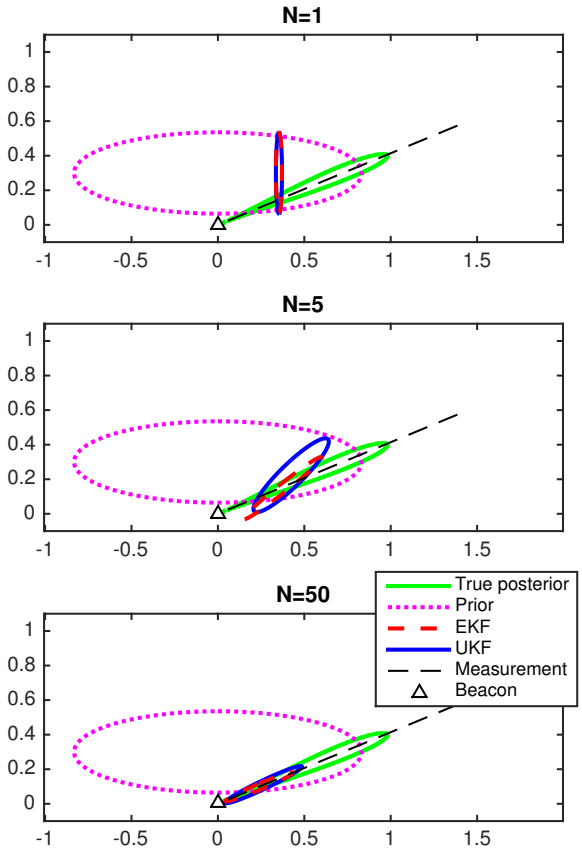

Fig. 5. A bearings measurement update

V. Conclusions and Future Work

In this paper, we presented a systematic approach for estimation with non-Gaussian measurement noises. In our approach, the non-Gaussian noise is represented as a nonlinearly transformed Gaussian noise. The proposed method can be applied to different error distributions.

We extended the RUF algorithm into GRUF to allow filtering with the proposed model. In our tests with Student-t distributed noises, the proposed algorithm had similar accuracy as a method designed specifically for the Student-t noise. However, we found that the GRUF-algorithm used for filtering may have some problems in certain situations when the number of iterations is large. We also showed that a similar problem occurs with the original RUF [16] and SPRUF [17]. Thus, more study is needed to develop a filter that does not exhibit such behavior.

Acknowledgements

M. Raitoharju works in OpenKin project that is funded by the Academy of Finland.

H. Nurminen receives funding from Tampere University of Technology Graduate School, the Foundation of Nokia Corporation, and Tekniikan edistämissäätiö.

References

[1] R. 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] E. Wan and R. Van der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. Adaptive Systems for Signal Process., Commun., and Control Symp.. AS-SPCC.*, 2000, pp. 153–158, doi: 10.1109/ASSPCC.2000.882463.

[3] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000, doi: 10.1109/9.855552.

[4] M. Wüthrich, C. Garcia Cifuentes, S. Trimpe, F. Meier, J. Bohg, J. Issac, and S. Schaal, "Robust Gaussian filtering," in *Proceedings of the American Control Conference*, Boston, MA, USA, Jul. 2016, accepted. [Online]. Available: http://arxiv.org/abs/1509.04072v2

[5] C. Masreliez, "Approximate non-Gaussian filtering with linear state and observation relations," *IEEE Transactions on Automatic Control*, vol. 20, no. 1, pp. 107–110, Feb 1975, doi: 10.1109/TAC.1975.1100882.

[6] W.-R. Wu and A. Kundu, "Recursive filtering with non-Gaussian noises," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1454–1468, Jun 1996, doi: 10.1109/78.506611.

[7] J.-A. Ting, E. Theodorou, and S. Schaal, "A Kalman filter for robust outlier detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*. IEEE, Oct 2007, pp. 1514–1519, doi: 10.1109/IROS.2007.4399158.

[8] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2012, pp. 1–6, doi: 10.1109/MLSP.2012.6349794.

[9] G. Agamennoni, J. Nieto, and E. Nebot, "An outlier-robust Kalman filter," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2011, pp. 1551–1558, doi: 10.1109/ICRA.2011.5979605.

[10] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb 2002, doi: 10.1109/78.978374.

[11] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, Jul. 1971, doi: 10.1016/0005-1098(71)90097-5.

[12] M. Rosenblatt, "Remarks on a multivariate transformation," *The Annals of Mathematical Statistics*, vol. 23, no. 3, pp. 470–472, 1952. [Online]. Available: http://www.jstor.org/stable/2236692

[13] C. Villani, *Optimal transport : old and new*, ser. Grundlehren der mathematischen Wissenschaften. Berlin: Springer, 2009.

[14] A. Garcia-Fernandez, L. Svensson, M. Morelande, and S. Särkka, "Posterior linearization filter: Principles and implementation using sigma points," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5561–5573, Oct 2015, doi: 10.1109/TSP.2015.2454485.

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

[16] R. Zanetti, "Recursive update filtering for nonlinear estimation," *IEEE Transactions on Automatic Control*, vol. 57, no. 6, pp. 1481–1490, June 2012, doi: 10.1109/TAC.2011.2178334.

[17] Y. Huang, Y. Zhang, N. Li, and L. Zhao, "Design of sigma-point Kalman filter with recursive updated measurement," *Circuits, Systems, and Signal Process.*, pp. 1–16, August 2015, doi: 10.1007/s00034-015-0137-y.

[18] R. Zanetti, "Adaptable iterative and recursive Kalman filter schemes," in *AAS Spaceflight Mechanics Meeting*, vol. 24, Santa Fe, NM, January 2014. [Online]. Available: http://hdl.handle.net/2060/20140006041