Cramér-Rao Lower Bound for Linear Filtering with t-Distributed Measurement Noise

Robert Piché
Tampere University of Technology

Abstract—**The Cramér-Rao lower bound (CRLB) on the achievable mean square error (MSE) can be used to evaluate approximate estimation algorithms. For linear filtering problems with non-Gaussian noises, the CRLB can be easily computed using the Kalman filter state covariance recursion with the Fisher information in place of the noise covariance term. This work studies a linear filtering problem with t-distributed measurement noise. It is found that for a t distribution with heavy tails, the CRLB significantly underestimates the optimal MSE, the Kalman filter has significantly larger MSE, and a computationally light variational-Bayes algorithm achieves nearly optimal MSE.**

I. Introduction

The Student t distribution is often used as an alternative to the normal distribution to model observations that have large deviations more frequently than do samples from a normal distribution. The t distribution's degrees of freedom parameter ν controls the heaviness of the tail, and ν values between 3 and 5 are typically used as default values for measurements of physical phenomena. The t distribution with $\nu = 1$ is the Cauchy distribution; its tails are so heavy that the mean and variance do not exist. At the other extreme, the t distribution converges to a normal distribution in the limit as $\nu \to \infty$.

In the standard Kalman filtering model, measurement noise is modelled as a sequence of independent normally-distributed random vectors. If a t distribution is used instead, then approximate methods are needed to compute the estimation posterior. Three such methods are the following.

Kalman filter (KF): A Kalman filter whose noise covariance parameter is chosen to match the noise covariance is a computationally cheap method. This filter is optimal in the sense that no other linear filter gives a smaller mean square error. However, because it is a linear filter, the estimate tends to overreact to occasional outlier measurements, and the filter may need several time steps to recover. Another drawback is that t noise covariance is undefined for $\nu \in (0,2]$, so the filter cannot be used for such heavy-tailed distributions.

Particle Filter (PF): In a bootstrap particle filter [1], the replacement of a normal measurement noise model by t noise is trivial: It suffices to replace the call to the normal density function, in the computation of the particle weights, by a call to the t density function. PF can approximate the Bayesian estimate as closely as desired by using a sufficiently large number of particles, but approximation accuracy comes at the cost of considerably more computational effort than KF, especially when the state dimension is large.

Variational Bayes (VB): Filtering and smoothing algorithms for nonlinear systems with t distributed measurements are presented in [2]. These algorithms are derived using a variational Bayes approximation based on the Gaussian scale mixture representation of the t distribution. The filtering algorithm resembles KF with EM-like iterations that adjust the gain according to the size of the innovation; the algorithm's computational complexity is roughly that of KF multiplied by the number of iterations. In simulations, a small number of iterations (2–4) has been observed to give accuracy similar to that of PF.

A popular tool for evaluating a filtering system design is the Cramér-Rao lower bound (CRLB), a lower bound on the achievable mean square error (MSE). The CRLB can also help evaluate the quality of a filter approximation. In particular, for linear filtering problems, closeness of the CRLB to the MSE matrix of the optimal KF indicates that there would be little advantage in using a computationally heavier nonlinear estimation method in place of KF.

The Bayesian CRLB for discrete-time filtering problems can be computed using a recursive formula [3]. For linear filtering problems with non-Gaussian additive noise, the computation is very light: the formula is identical to the KF covariance evolution formula, except that the term containing the measurement noise covariance is replaced by the inverse of the Fisher information matrix.

CRLBs for linear filtering problems with non-Gaussian measurement and process noises were studied in [4] and [5]. Skewed heavy-tailed noise distributions were modelled by Gaussian mixtures, the Fisher informations were computed numerically, and examples were presented comparing the CRLB with MSE obtained with KF and of PF with a large number of particles. Examples were presented where the CRLB is significantly smaller than the MSE achieved by PF, indicating that the CRLB is a poor approximation of the optimal MSE for these systems. In some of the examples the PF error was also significantly smaller than the KF error; in others it was about the same.

This work revisits the studies of [4], [5], with two modifications. First, the study is based on a non-Gaussian noise distribution for which the Fisher information can be easily computed in closed form. This helps to clarify the analysis and makes the study more attractive as a pedagogical example. Secondly, the study includes results for the VB filter.

19th International Conference on Information Fusion, Heidelberg, Germany, July 5-8, 2016
978-0-9964527-4-8 ©2016 ISIF

II. Theory

In the following, symbols for random variables and vectors are underlined to distinguish them from symbols of real variables and vectors. The notation $\underline{x} \sim \mathrm{MVN}(m,P)$ means that $\underline{x}$ has a multivariate normal distribution with mean m and covariance P; its density function is

$$p_{\underline{x}}(x) \propto \mathrm{e}^{-\frac{1}{2}(x-m)P^{-1}(x-m)}. \tag{1}$$

The notation $\underline{y} \sim \mathrm{MVT}(m,T,\nu)$ means that $\underline{y}$ has a t distribution with location parameter m, shape matrix T and degrees of freedom parameter ν; its density function is

$$p_{\underline{y}}(y) \propto \left(1+\tfrac{1}{\nu}(y-m)'T^{-1}(y-m)\right)^{-\frac{\nu+n}{2}}. \tag{2}$$

where n is the dimension of y. For $\nu > 1$ the distribution's mean is m; for $\nu > 2$ its covariance is $\frac{\nu}{\nu-2}T$. Random samples of $\underline{y} \sim \mathrm{MVT}(m,T,\nu)$ can be obtained using the Gaussian scale mixture representation

$$\underline{z} \sim \chi^2_\nu, \quad \underline{y} \,|\, z \sim \mathrm{MVN}(m, \frac{\nu}{z}T). \tag{3}$$

An n-variate measurement y that is a linear function of x with additive t distributed noise is modelled as

$$\underline{y}|x \sim \mathrm{MVT}(Hx,T,\nu),$$

where $\underline{y}|x$ denotes the conditional random variable $\underline{y}|(\underline{x}=x)$. The measurement's Fisher information is

$$\mathscr{I}(x) = \tfrac{\nu+n}{\nu+n+2}H'T^{-1}H. \tag{4}$$

This formula is derived in [6, Proposition 4]; an alternative derivation is given in appendix A.

Consider a state sequence $\underline{x}_0, \underline{x}_1, \ldots$ that evolves according to a standard linear-Gaussian state-space model:

$$\underline{x}_0 \sim \mathrm{MVN}(m_0,P_0), \tag{5a}$$

$$\underline{x}_{k+1} \,|\, x_k \sim \mathrm{MVN}(F_k x_k, Q_k), \tag{5b}$$

where $k \in \{0,1,2,\ldots\}$ is the time index. Assume that the measurement at each time is a linear function of the current state, with additive noise that is not necessarily Gaussian:

$$\underline{y}_k \,|\, x_k = H_k x_k + \underline{e}_k, \qquad k \in \{1,2,\ldots\}. \tag{6}$$

The initial state, process noises, and measurement noises are assumed independent.

Let $\hat{x}_k$ denote a filter's estimate of the current state $\underline{x}_k$; it is a deterministic function of realised past and current measurements $y_1,\ldots,y_k$. The same function applied to the random variables $\underline{y}_1,\ldots,\underline{y}_k$ is denoted $\underline{\hat{x}}_k$. A measure of the quality of the estimation function is the MSE matrix, defined as

$$\Sigma_k = \mathsf{E}\left((\underline{x}_k - \underline{\hat{x}}_k))(\underline{x}_k - \underline{\hat{x}}_k)'\right). \tag{7}$$

The expectation in (7) is over the joint distribution of $(\underline{x}_k, \underline{y}_{1:k})$.

The filter whose value is the posterior mean is denoted $\hat{x}_k^{\mathrm{pm}}$ and is defined by

$$\hat{x}_k^{\mathrm{pm}} = \mathsf{E}(\underline{x}_k \,|\, y_{1:k}). \tag{8}$$

It is optimal in the sense that its MSE matrix Σ_k^{pm} is no larger in the Loewner partial ordering than the MSE matrix Σ_k of any filter; that is, $\Sigma_k - \Sigma_k^{\mathrm{pm}}$ is non-negative definite. This fact, proved for example in [7], is denoted

$$\Sigma_k \succeq \Sigma_k^{\mathrm{pm}}. \tag{9}$$

For systems with non-Gaussian noise, approximate methods are needed to compute the MSE-optimal estimate. In principle, the PF estimate's MSE can be made as close as desired to the Bayesian filtering distribution, so that $\Sigma_k^{\mathrm{pf}} \approx \Sigma_k^{\mathrm{pm}}$. However, achieving nearly optimal MSE may require an impractically large number of particles N_{pf}, and so the computationally lighter alternatives KF and VB are considered. Details of the KF, PF, and VB approximate filter algorithms for the linear system with t distributed noise are given in appendix B.

The Cramér-Rao lower bound (CRLB) provides a formula for matrices B_k such that $\Sigma_k \succeq B_k$. The recursive formula for computing the CRLB is derived in [3]. For the linear model with Gaussian process noise and non-Gaussian measurement noise, the recursive formula reduces to

$$B_{k+1} \leftarrow \left((F_k B_k F_k' + Q_k)^{-1} + \mathsf{E}(\mathscr{I}_k(\underline{x}_k))\right)^{-1}, \tag{10}$$

where $\mathscr{I}_k(x_k)$ is the Fisher information matrix of the measurement $\underline{y}_k \,|\, x_k$, and the recursion is initialised with $B_0 = P_0$. In (10) the expectation is with respect to the distribution of $\underline{x}_k$.

For the t distributed measurement noise model

$$\underline{y}_k \,|\, x_k \sim \mathrm{MVT}(H_k x_k, T_k, \nu), \tag{11}$$

(4) is substituted into (10) to obtain

$$B_{k+1} \leftarrow \left((F_k B_k F_k' + Q_k)^{-1} + \tfrac{\nu+n}{\nu+n+2}H_k' T_k^{-1} H_k\right)^{-1}. \tag{12}$$

Applying the matrix inversion lemma, this can be written as

$$B_{k|k-1} \leftarrow F_{k-1} B_{k-1} F_{k-1}' + Q_k \tag{13a}$$

$$S_k \leftarrow H_k B_{k|k-1} H_k' + \tfrac{\nu+n_k+2}{\nu+n_k} T_k \tag{13b}$$

$$K_k \leftarrow B_{k|k-1} H_k' S_k^{-1} \tag{13c}$$

$$B_k \leftarrow B_{k|k-1} - K_k S_k K_k'. \tag{13d}$$

This is identical to the Kalman filter's covariance propagation formula, except that the measurement covariance is replaced by $\frac{\nu+n_k+2}{\nu+n_k}T_k$, a multiple of the t noise's shape matrix.

III. Example

Consider a one-dimensional motion described by (5) and (6) with

$$m_0 = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right],\ P_0 = \left[\begin{smallmatrix}40 & 0\\ 0 & 4\end{smallmatrix}\right], \tag{14a}$$

$$F_k = \left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right],\ Q_k = \left[\begin{smallmatrix}0 & 0\\ 0 & 1\end{smallmatrix}\right], \tag{14b}$$

$$H_k = [1\ 0],\ T_k = \tfrac{100}{3}, \quad \nu = 3 \tag{14c}$$

This system is similar to the example in [4], and is an approximation of an integrated Wiener process. The state components are position and velocity; the measurement is position.

Because the measurement noise is not Gaussian, the optimal Bayesian filter for this problem is not a Kalman filter.

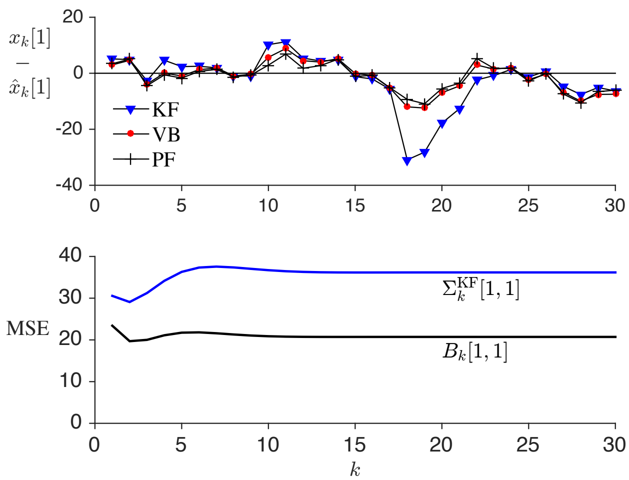

Fig. 1. Error in position estimates for tracking example

However, the measurement noise variance exists (it is 100), and the Kalman filter with this parameter value is the MSE-optimal *linear* filter for this problem. Also, the Kalman filter's computed state covariance matrix is the filter's MSE, because the filter's noise covariance parameter matches the measurement noise variance.

Simulations are computed with $N_{\rm pf} = 1000$ particles and $N_{\rm vb} = 2$ variational Bayes iterations. This VB filter's computational complexity is roughly twice that of KF. Figure 1 shows the results for one of the simulations. It can be seen that the three filters have similar errors most of the time, except most notably when a measurement outlier is encountered at time $k = 18$. The Kalman filter's error makes a large jump because of the thin tail of its underlying Gaussian noise model. Also, because of the large value of the noise covariance parameter, KF needs several time steps to recover. In contrast, the large outlier has no visible effect on the VB and PF estimate.

The lower part of Figure 1 shows the evolution of the position variance term in the Kalman filter's MSE matrix and of the corresponding CRLB. The values at the end of the simulation are

$$B_{30}[1,1] = 20.7, \quad \Sigma_{30}^{\rm kf}[1,1] = 36.2.$$

These agree, to all shown digits, with the values that are found by solving (using `dare` in the MATLAB Control Systems Toolbox) the discrete Riccati equation [7, Eqn (4.4)] that governs the steady state of the recursion. The large gap between the MSE values imply the possibility that a nonlinear filter could be significantly more accurate.

Repeating the simulation 10 000 times with different random numbers, 90% confidence intervals for the position variance of VB and PF at time $k = 30$ are found to be

$$\Sigma_{30}^{\rm vb}[1,1] \approx 25.4 \pm 0.6, \quad \Sigma_{30}^{\rm pf}[1,1] \approx 24.9 \pm 0.6 \qquad (15)$$

Increasing the number of particles to $N_{\rm pf} = 5000$ gives nearly the same value, $\Sigma_{30}^{\rm pf}[1,1] = 23.6$, so the PF value in (15) can be considered to be a good approximation of the optimal value. The VB and PF error variances are nearly equal, which indicates that the VB filter is giving nearly optimal accuracy. Both VB and PF give significant improvement in accuracy relative to the Kalman filter. There is also still a clear gap relative to the CRLB, which indicates that the CRLB is not a tight bound for this system.

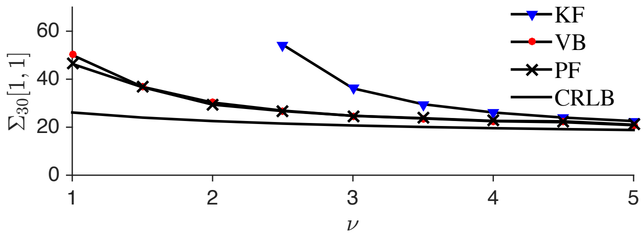

Fig. 2. Position error variances vs. measurement noise kurtosis parameter

Repeating the simulation set with different values of the degrees of freedom parameter ν, keeping the shape parameter T_k fixed, gives the MSE values shown in Figure 2. The gap between CRLB and the MSE of PF increases as ν decreases, that is, the underestimation of CRLB worsens as kurtosis increases and the distribution becomes less Gaussian; for $\nu = 1$ (Cauchy distribution), the CRLB is about half the MSE of PF. The gap between CRLB and KF also increases as ν decreases. VB and PF results are very close for most values of ν, except at $\nu = 1$ (Cauchy noise) where a small difference is visible: the values and 90% intervals are

$$\Sigma_{30}^{\rm vb}[1,1] \approx 50.1 \pm 1.2, \quad \Sigma_{30}^{\rm pf}[1,1] \approx 46.4 \pm 1.1$$

IV. Conclusion

Section III presented a simple example of a linear filtering problem with heavy-tailed non-Gaussian noise in which a VB filter is clearly more accurate than the Kalman filter. Because VB has an MSE close to that of PF with a large number of particles, VB can be considered to be close to optimal, even though the CRLB is considerably smaller when the noise is very heavy-tailed. Because VB is an algorithm that is a relatively simple modification of KF, and is computationally light, it can be recommended for use in filtering problems with heavy-tailed measurement data.

Appendix A
Fisher information of multivariate-t

The Fisher information matrix of a measurement y is defined as

$$\mathscr{I}_k(x) = -\mathsf{E}\big(\mathscr{H}(\underline{y}|x, x)\big), \qquad (16)$$

where $\mathscr{H}(y,x) = \Delta_x\big(\ln p_{\underline{y}|\underline{x}}(y\,|\,x)\big)$ denotes the Hessian matrix (i.e. matrix of second-order partial derivatives) with respect to x of the log-likelihood, and the expectation in (16) is with respect to the distribution of $\underline{y}|x$.

The log likelihood of $\underline{y}\,|\,x \sim \mathrm{MVT}(Hx, T, \nu)$ is

$$\ln p_{\underline{y}|\underline{x}}(y\,|\,x) = \tfrac{\nu+n}{2} \ln\big(1 + \tfrac{1}{\nu}(y - Hx)' T^{-1} (y - Hx)\big) + \text{const}.$$

Denoting $\phi = 1+\frac{1}{\nu}(y-Hx)'T^{-1}(y-Hx)$, one has

$$\nabla\phi = \frac{2}{\nu}H'T^{-1}(Hx-y), \quad \Delta\phi = \frac{2}{\nu}H'T^{-1}H.$$

Then, applying the chain rule

$$\Delta\ln(\phi) = \frac{1}{\phi}\Delta\phi - \frac{1}{\phi^2}(\nabla\phi)(\nabla\phi)',$$

one obtains

$$\begin{aligned}
-\mathscr{H}(y,x) &= \tfrac{\nu+n}{2}\Delta\ln\big(1+\tfrac{1}{\nu}(y-Hx)'T^{-1}(y-Hx)\big) \\
&= \tfrac{\nu+n}{2}\Big(\frac{(2/\nu)H'T^{-1}H}{1+\frac{1}{\nu}(y-Hx)'T^{-1}(y-Hx)} \\
&\quad - \frac{(4/\nu^2)H'T^{-1}(y-Hx)(y-Hx)'T^{-1}H}{\big(1+\frac{1}{\nu}(y-Hx)'T^{-1}(y-Hx)\big)^2}\Big).
\end{aligned}$$

Substituting this into (16) gives

$$\mathscr{I}(x) = H'T^{-\frac{1}{2}}\,\mathsf{E}\Big(\frac{(\nu+n)/\nu}{1+\frac{1}{\nu}\|\underline{u}\|^2}I_n - \frac{2(\nu+n)/\nu^2}{(1+\frac{1}{\nu}\|\underline{u}\|^2)^2}\underline{u}\underline{u}'\Big)T^{-\frac{1}{2}}H, \tag{17}$$

where the expectation is with respect to

$$\underline{u} = T^{-1/2}(\underline{y}|x-Hx) \sim \mathrm{MVT}(0,I_n,\nu),$$

whose density function is

$$p_{\underline{u}}(u) = \frac{1}{(\pi\nu)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}(1+\tfrac{\|u\|^2}{\nu})^{-(\nu+n)/2}.$$

The ith diagonal element of the expectation in (17) is

$$\begin{aligned}
&\int\Big(\frac{(\nu+n)/\nu}{1+\frac{\|u\|^2}{\nu}} - \frac{2(\nu+n)/\nu^2}{(1+\frac{\|u\|^2}{\nu})^2}u_i^2\Big)p_{\underline{u}}(u)\,\mathrm{d}u \\
&= \frac{1}{(\pi\nu)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\Big(\tfrac{\nu+n}{\nu}\int(1+\tfrac{\|u\|^2}{\nu})^{-(\nu+n)/2-1}\,\mathrm{d}u \\
&\qquad - \tfrac{2(\nu+n)}{\nu^2}\int u_i^2(1+\tfrac{\|u\|^2}{\nu})^{-(\nu+n)/2-2}\,\mathrm{d}u\Big) \\
&= \frac{1}{(\pi\nu)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\Big(\tfrac{\nu+n}{\nu}\big(\tfrac{\nu}{\nu+2}\big)^{n/2}\int(1+\tfrac{\|u\|^2}{\nu+2})^{-(\nu+2+n)/2}\,\mathrm{d}u \\
&\qquad - \tfrac{2(\nu+n)}{\nu^2}\big(\tfrac{\nu}{\nu+4}\big)^{(n+2)/2}\int u_i^2(1+\tfrac{\|u\|^2}{\nu+4})^{-(\nu+4+n)/2}\,\mathrm{d}u\Big) \\
&= \frac{1}{(\pi\nu)^{n/2}}\frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})}\Big(\tfrac{\nu+n}{\nu}\big(\tfrac{\nu}{\nu+2}\big)^{n/2}(\pi(\nu+2))^{n/2}\frac{\Gamma(\frac{\nu+2}{2})}{\Gamma(\frac{\nu+2+n}{2})} \\
&\qquad - \tfrac{2(\nu+n)}{\nu^2}\big(\tfrac{\nu}{\nu+4}\big)^{(n+2)/2}\tfrac{\nu+4}{\nu+4-2}(\pi(\nu+4))^{n/2}\frac{\Gamma(\frac{\nu+4}{2})}{\Gamma(\frac{\nu+4+n}{2})}\Big) \\
&= \frac{\nu+n}{\nu+n+2}.
\end{aligned}$$

Also, because the standard multivariate t distribution's density is radially symmetric about the origin, the off-diagonal terms of $\mathscr{I}(x)$ are zero. Putting these results together, (4) is obtained.

Appendix B
Approximate filters

The following approximate filters are used to estimate the states of the system specified by (5) and (6).

A. Kalman filter (KF)

The Kalman filter equations are

$$\hat{x}_{k|k-1} \leftarrow F_{k-1}\hat{x}^{\mathrm{kf}}_{k-1} \tag{18a}$$
$$\Sigma^{\mathrm{kf}}_{k|k-1} \leftarrow F_{k-1}\Sigma^{\mathrm{kf}}_{k-1}F'_{k-1} + Q_{k-1} \tag{18b}$$
$$S_k \leftarrow H_k\Sigma_{k|k-1}H'_k + \tfrac{\nu}{\nu-2}T_k \tag{18c}$$
$$K_k \leftarrow \Sigma_{k|k-1}H'_kS_k^{-1} \tag{18d}$$
$$\hat{x}^{\mathrm{kf}}_k \leftarrow \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}) \tag{18e}$$
$$\Sigma^{\mathrm{kf}}_k \leftarrow \Sigma_{k|k-1} - K_kS_kK'_k \tag{18f}$$

for $k = 1,2\ldots$; the recursion is initialised with $\hat{x}^{\mathrm{kf}}_0 = m_0$, $\Sigma^{\mathrm{kf}}_0 = P_0$. This recursion uses the measurement covariance in step (18c), and so the filter can only be used if $\nu > 2$.

B. Variational-Bayes Filter (VB)

For the given model, the variational Bayes algorithm presented in [2] reduces to

$$\hat{x}_{k|k-1} \leftarrow F_{k-1}\hat{x}^{\mathrm{vb}}_{k-1} \tag{19a}$$
$$P_{k|k-1} \leftarrow F_{k-1}P_{k-1}F'_{k-1} + Q_{k-1} \tag{19b}$$
$$\ell_k \leftarrow 1 \tag{19c}$$
do N_{vb} times (19d)
$$S_k \leftarrow HP_{k|k-1}H'_k + \tfrac{1}{\ell_k}T_k \tag{19e}$$
$$K_k \leftarrow P_{k|k-1}H'_kS_k^{-1} \tag{19f}$$
$$P_k \leftarrow P_{k|k-1} - K_kS_kK'_k \tag{19g}$$
$$\hat{x}^{\mathrm{vb}}_k \leftarrow \hat{x}_{k|k-1} + K_k(y_k - H_k\hat{x}_{k|k-1}) \tag{19h}$$
$$\ell_k \leftarrow \frac{\nu+n}{\nu + (y_k - H_k\hat{x}^{\mathrm{vb}}_k)'T_k^{-1}(y - H_k\hat{x}^{\mathrm{vb}}_k) + \mathrm{trace}(T_k^{-1}H_kP_kH'_k)} \tag{19i}$$
end do (19j)

for $k = 1,2\ldots$; the recursion is initialised with $\hat{x}^{\mathrm{vb}}_0 = m_0$.

The parameter ℓ_k introduces a data-dependent scaling to the Kalman gain. Large values of the posterior residual $y_k - H_k\hat{x}^{\mathrm{vb}}_k$ produce small values of ℓ_k, which in turn reduces K_k and thereby reduces the influence of the measurement.

C. Particle filter (PF)

After initialisation of N_{pf} particles $x_0^{(i)} \sim \mathrm{MVN}(m_0,P_0)$, the bootstrap particle filter equations for $k \in 1,2,\ldots$ are

for $i \in 1{:}N_{\mathrm{pf}}$ (20a)
$$x_k^{(i)} \sim \mathrm{MVN}(F_{k-1}x_{k-1}^{(i)}, Q_{k-1}) \tag{20b}$$
$$w_k^{(i)} \leftarrow \mathrm{MVTPDF}\big(y_k;\, H_kx_k^{(i)}, T_k, \nu\big) \tag{20c}$$
end for (20d)

for $i \in 1{:}N_{\mathrm{pf}},\ j_i \sim \mathrm{categ}\big(w_k^{(1:N_{\mathrm{pf}})}/\mathrm{sum}(w_k^{(1:N_{\mathrm{pf}})})\big)$, end for (20e)

$$x_k^{(1:N_{\mathrm{pf}})} \leftarrow x_k^{(j_{1:N_{\mathrm{pf}}})} \tag{20f}$$

Steps (20e– 20f) are multinomial resampling, and $\mathrm{categ}(\cdot)$ denotes the discrete distribution on the values $1:N$.

REFERENCES

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

[2] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *IEEE Intl. Workshop on Machine Learning for Signal Processing*, 2012.

[3] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1386–1396, 1998.

[4] G. Hendeby and F. Gustafsson, "Fundamental filtering limitations in linear non-Gaussian systems," in *IFAC World Congress*, 2005.

[5] G. Hendeby, "Fundamental estimation and detection limits in linear non-Gaussian systems," Ph.D. dissertation, Linköpings universitet, 2005.

[6] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice-Hall, 1979.