Student's t-filters for noise scale estimation

Filip Tronarp, Toni Karvonen, and Simo Särkkä

Student's t-Filters for Noise Scale Estimation

Filip Tronarp, *Student Member, IEEE*, Toni Karvonen, and Simo Särkkä, *Senior Member, IEEE*

Abstract—We analyse certain Student's t-filters for linear Gaussian systems with misspecified noise covariances. It is shown that the under appropriate conditions the filter both estimates the state and re-scales the noise covariance matrices in a Kullback–Leibler optimal fashion. If the noise covariances are misscaled by a common scalar, then the re-scaling is asymptotically exact. We also compare the Student's t-filter scale estimates to the maximum likelihood estimates. Simulations demonstrating the results on the Wiener velocity model are provided.

Index Terms—Kalman filtering, Student's t-filtering, model misspecification, noise covariance estimation

I. Introduction

RECENTLY, Student's t-filters, that assume the latent state and system noises have a joint Student's t-distribution, for discrete-time systems have become popular in the signal processing community [1]–[10]. These filters were originally developed in the early 1990s [11] (see also [12], [13]) and later modified and popularised by Roth et al. [1]. The difference between the filters in [11], called here *Student's t-filter* (ST; Alg. 2), and in [1, Sec. 3.1] (ST2; Rmk. 3) is that in the former a joint probability model for complete state and measurement sequences is used while in the latter each time-step gets its own probability model that neglects some of the dependencies present in the model of [11] (see Sec. II-C). This article studies Student's t-filter [11] for linear Gaussian discrete-time systems with misspecified noise covariances.

We (i) establish a relationship (Props. 4 and 5) between Student's t-filter and classical misspecified Kalman filter (MKF); (ii) prove that, under certain assumptions on asymptotics of the scale estimator, Student's t-filter estimates scaling of the noise covariance matrices in a Kullback–Leibler optimal way (Thm. 6); and (iii) show that in the special case of covariance matrices being misspecified by a common scalar factor the re-scaling due to Student's t-filter is asymptotically correct (Cor. 7; in this setting, the scaling provided by the filter is compared to the maximum likelihood estimate in Thm. 8).

The state estimates of the Student's t-filter coincide with those of the misspecified Kalman filter. However, estimating uncertainty matters in applications such as measurement gating [14, Sec. 2.3] and uncertainty quantification for ODE solvers [15]–[17]. For such purposes, Student's t-filter provides computationally attractive tuning of the filter noise covariance matrices. For other, computationally more involved, estimators, see for example [18]–[23].

II. Filtering for Linear Systems

We begin by introducing the classical Kalman filter [24], [25, Ch. 4] and Student's t-filter as it appears in [11], [12]. The difference between ST and ST2 is discussed in Rmk. 3.

A. The System Description

We assume that the latent state $\mathbf{x}_n \in \mathbb{R}^{d_x}$ and partial and corrupted measurements $\mathbf{y}_n \in \mathbb{R}^{d_y}$ of it are, for $n \in \mathbb{N}$, generated by the linear Gaussian time-varying system

$$\mathbf{X}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{P}_0), \tag{1a}$$

$$\mathbf{X}_n \mid \mathbf{X}_{n-1} \sim \mathcal{N}(\mathbf{A}_n \mathbf{X}_{n-1}, \mathbf{Q}), \tag{1b}$$

$$\mathbf{Y}_n \mid \mathbf{X}_n \sim \mathcal{N}(\mathbf{C}_n \mathbf{X}_n, \mathbf{R}), \tag{1c}$$

where $\mathbf{A}_n \in \mathbb{R}^{d_x \times d_x}$ and $\mathbf{C}_n \in \mathbb{R}^{d_y \times d_x}$ are model matrices, $\mathbf{Q}$ and $\mathbf{R}$ are the noise covariances, and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$.

B. The (Misspecified) Kalman Filter

The classical Kalman filter [24], [25, Ch. 4] computes the conditional mean $\mathbb{E}[\mathbf{X}_n \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_n] = \mathbb{E}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$ and variance $\mathbb{V}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$ using simple linear algebraic recursion equations. However, the true noise covariances $\mathbf{Q}$ and $\mathbf{R}$ and the initial error covariance $\mathbf{P}_0$ are often not available, forcing one to employ some other positive-definite matrices $\mathbf{Q}^0$, $\mathbf{R}^0$, and $\mathbf{P}_0^0$ in the filter. The resulting MKF computes the state estimates $\mathbf{X}_{n|n}^{\text{MKF}}$ and associated error covariances $\mathbf{P}_{n|n}^{\text{MKF}}$ as follows.

Algorithm 1 (The misspecified Kalman filter; MKF)**.** *The misspecified Kalman filter consists of the prediction*

$$\mathbf{X}_{n|n-1}^{\textit{MKF}} = \mathbf{A}_n \mathbf{X}_{n-1|n-1}^{\textit{MKF}},$$

$$\mathbf{P}_{n|n-1}^{\textit{MKF}} = \mathbf{A}_n \mathbf{P}_{n-1|n-1}^{\textit{MKF}} \mathbf{A}_n^{\mathsf{T}} + \mathbf{Q}^0,$$

where $\mathbf{X}_{0|0}^{\textit{MKF}} = \boldsymbol{\mu}_0$ and $\mathbf{P}_{0|0}^{\textit{MKF}} = \mathbf{P}_0^0$, and the update

$$\mathbf{Y}_{n|n-1}^{\textit{MKF}} = \mathbf{C}_n \mathbf{X}_{n|n-1}^{\textit{MKF}},$$

$$\mathbf{S}_n^{\textit{MKF}} = \mathbf{C}_n \mathbf{P}_{n|n-1}^{\textit{MKF}} \mathbf{C}_n^{\mathsf{T}} + \mathbf{R}^0,$$

$$\mathbf{K}_n^{\textit{MKF}} = \mathbf{P}_{n|n-1}^{\textit{MKF}} \mathbf{C}_n^{\mathsf{T}} [\mathbf{S}_n^{\textit{MKF}}]^{-1},$$

$$\mathbf{X}_{n|n}^{\textit{MKF}} = \mathbf{X}_{n|n-1}^{\textit{MKF}} + \mathbf{K}_n^{\textit{MKF}} (\mathbf{Y}_n - \mathbf{Y}_{n|n-1}^{\textit{MKF}}),$$

$$\mathbf{P}_{n|n}^{\textit{MKF}} = \mathbf{P}_{n|n-1}^{\textit{MKF}} - \mathbf{K}_n^{\textit{MKF}} \mathbf{S}_n^{\textit{MKF}} [\mathbf{K}_n^{\textit{MKF}}]^{\mathsf{T}}.$$

Due to the use of incorrect noise covariance matrices, $\mathbf{X}_{n|n}^{\text{MKF}} \neq \mathbb{E}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$ and $\mathbf{P}_{n|n}^{\text{MKF}} \neq \mathbb{V}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$ unless $\mathbf{Q}^0 = \mathbf{Q}$, $\mathbf{R}^0 = \mathbf{R}$, and $\mathbf{P}_0^0 = \mathbf{P}_0$. If there is no covariance misspecification, the above filter is the optimal Kalman filter and we will use superscript KF instead of MKF. That is, $\mathbf{X}_{n|n}^{\text{KF}} = \mathbb{E}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$ and $\mathbf{P}_{n|n}^{\text{KF}} = \mathbb{V}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}]$.

Before presenting Student's t-filter, we note that the effect of (partial) model misspecification as above to behaviour and stability of the Kalman filter has been studied for discrete-time systems in [26]–[30] and for continuous-time systems in [31].

C. Student's t-Filter

Let $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ stand for the Gaussian density with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ and $\Gamma^{-1}(s;\nu/2,\nu/2)$ for the reciprocal gamma density with shape and scale parameters $\nu/2$. Girón and Rojano [11] consider the Bayesian model

$$
\begin{aligned}
&p(\mathbf{x}_{0:N},\mathbf{y}_{1:N},s)\\
&= \mathcal{N}(\mathbf{x}_0;\boldsymbol{\mu}_0,s\mathbf{P}_0^0)\prod_{n=1}^{N}\mathcal{N}(\mathbf{x}_n;\mathbf{A}_n\mathbf{x}_{n-1},s\mathbf{Q}^0)\\
&\quad\times\prod_{n=1}^{N}\mathcal{N}(\mathbf{y}_n;\mathbf{C}_n\mathbf{x}_n,s\mathbf{R}^0)\Gamma^{-1}(s;\nu_0/2,\nu_0/2)
\end{aligned}
\tag{2}
$$

for the full latent state and measurement sequences and noise scaling s. They show that by marginalising out s and using properties of Gaussian scale mixtures [13] a Kalman filter recursion for the conditional mean is obtained (essentially our Props. 4 and 5).

In this article, we consider the filter of [11]. In order to relate this filter to the one in [1] and to facilitate subsequent analysis, we sequentially re-parametrise the scale variable according to $s_n := \gamma_n^{-1}s_{n-1}$ and $s_0 := s$. This results in the filter in Alg. 2.

Algorithm 2 (Student's t-filter; ST). *Student's t-filter consists of the prediction*

$$
\begin{aligned}
\mathbf{X}_{n|n-1}^{ST} &= \mathbf{A}_n\mathbf{X}_{n-1|n-1}^{ST},\\
\mathbf{P}_{n|n-1}^{ST} &= \mathbf{A}_n\mathbf{P}_{n-1|n-1}^{ST}\mathbf{A}_n^{\mathsf{T}} + \mathbf{Q}_{n-1}^{ST},
\end{aligned}
$$

where $\mathbf{X}_{0|0}^{ST} = \boldsymbol{\mu}_0$, $\mathbf{P}_{0|0}^{ST} = \mathbf{P}_0^0$, and $\mathbf{Q}_0^{ST} = \mathbf{Q}^0$, and the update

$$
\begin{aligned}
\mathbf{Y}_{n|n-1}^{ST} &= \mathbf{C}_n\mathbf{X}_{n|n-1}^{ST},\\
\mathbf{S}_n^{ST} &= \mathbf{C}_n\mathbf{P}_{n|n-1}^{ST}\mathbf{C}_n^{\mathsf{T}} + \mathbf{R}_{n-1}^{ST},\\
\gamma_n &= \frac{\nu_0 + (n-1)d_y + \|\mathbf{Y}_n - \mathbf{Y}_{n|n-1}^{ST}\|_{[\mathbf{S}_n^{ST}]^{-1}}^2}{\nu_0 + nd_y},\\
\mathbf{K}_n^{ST} &= \mathbf{P}_{n|n-1}^{ST}\mathbf{C}_n^{\mathsf{T}}[\mathbf{S}_n^{ST}]^{-1},\\
\mathbf{X}_{n|n}^{ST} &= \mathbf{X}_{n|n-1}^{ST} + \mathbf{K}_n^{ST}(\mathbf{Y}_n - \mathbf{Y}_{n|n-1}^{ST}),\\
\mathbf{P}_{n|n}^{ST} &= \gamma_n(\mathbf{P}_{n|n-1}^{ST} - \mathbf{K}_n^{ST}\mathbf{S}_n^{ST}[\mathbf{K}_n^{ST}]^{\mathsf{T}}),\\
\mathbf{Q}_n^{ST} &= \gamma_n\mathbf{Q}_{n-1}^{ST},\\
\mathbf{R}_n^{ST} &= \gamma_n\mathbf{R}_{n-1}^{ST},
\end{aligned}
$$

where $\mathbf{R}_0^{ST} = \mathbf{R}^0$ and $\|\mathbf{x}\|_{\mathbf{A}}^2 = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}$ when $\mathbf{A}$ is a positive-definite matrix.

Remark 3. The difference between ST of Alg. 2 and ST2 proposed in [1, Sec. 3.1] under the name *exact filter* is that

$$
\mathbf{Q}_n^{ST2} = \mathbf{Q}_{n-1}^{ST2} = \cdots = \mathbf{Q}^0 \text{ and } \mathbf{R}_n^{ST2} = \mathbf{R}_{n-1}^{ST2} = \cdots = \mathbf{R}^0.
$$

This is because ST2 is based on a sequence of probability models $p_{n+1}(\mathbf{y}_{n+1},\mathbf{x}_{n+1},\mathbf{x}_n,s)$ of the form (2) such that $p_{n+1}(\mathbf{x}_n \mid s)$ matches $p_n(\mathbf{x}_n \mid \mathbf{y}_n,s)$, the posterior from the previous time-step. That is, future noise realisations are not included in these models.

III. PROPERTIES OF STUDENT'S t-FILTER

In this section we prove a number of properties of Student's t-filter of Alg. 2.

A. Basic Properties of Student's t-Filter

We begin by establishing some important connections between Student's t-filter, misspecified Kalman filter, and the optimal Kalman filter. For the analysis it is useful to define

$$
\xi_n := \prod_{i=1}^{n}\gamma_i, \quad \xi_0 := 1.
\tag{3}
$$

Proposition 4. *The misspecified Kalman filter and Student's t-filter of Algs. 1 and 2 admit the following relations:*

$$
\begin{aligned}
\mathbf{X}_{n|n}^{ST} &= \mathbf{X}_{n|n}^{MKF}, \quad \mathbf{P}_{n|n}^{ST} = \xi_n\mathbf{P}_{n|n}^{MKF}, \quad \mathbf{S}_n^{ST} = \xi_{n-1}\mathbf{S}_n^{MKF},\\
\mathbf{Q}_n^{ST} &= \xi_n\mathbf{Q}^0, \quad \mathbf{R}_n^{ST} = \xi_n\mathbf{R}^0.
\end{aligned}
$$

Proof. From the definition of ξ_n we immediately observe that $\mathbf{Q}_n^{ST} = \xi_n\mathbf{Q}^0$ and $\mathbf{R}_n^{ST} = \xi_n\mathbf{R}^0$. Since $\mathbf{S}_1^{ST} = \mathbf{S}_1^{MKF}$ and $\mathbf{K}_1^{ST} = \mathbf{K}_1^{MKF}$, it follows that $\mathbf{P}_{1|1}^{ST} = \gamma_1\mathbf{P}_{1|1}^{MKF} = \xi_1\mathbf{P}_{1|1}^{MKF}$. The induction assumption $\mathbf{P}_{n-1|n-1}^{ST} = \xi_{n-1}\mathbf{P}_{n-1|n-1}^{MKF}$ yields

$$
\mathbf{P}_{n|n-1}^{ST} = \xi_{n-1}(\mathbf{A}_n\mathbf{P}_{n-1|n-1}^{MKF}\mathbf{A}_n^{\mathsf{T}} + \mathbf{Q}^0) = \xi_{n-1}\mathbf{P}_{n|n-1}^{MKF}.
$$

Consequently,

$$
\mathbf{S}_n^{ST} = \xi_{n-1}(\mathbf{C}_n\mathbf{P}_{n|n-1}^{MKF}\mathbf{C}_n^{\mathsf{T}} + \mathbf{R}^0) = \xi_{n-1}\mathbf{S}_n^{MKF}
$$

and, due to cancellation of ξ_{n-1}, $\mathbf{K}_n^{ST} = \mathbf{K}_n^{MKF}$. Therefore

$$
\mathbf{P}_{n|n}^{ST} = \gamma_n\xi_{n-1}(\mathbf{P}_{n|n-1}^{MKF} - \mathbf{K}_n^{MKF}\mathbf{S}_n^{MKF}[\mathbf{K}_n^{MKF}]^{\mathsf{T}}) = \xi_n\mathbf{P}_{n|n}^{MKF}.
$$

This establishes that $\mathbf{P}_{n|n}^{ST} = \xi_n\mathbf{P}_{n|n}^{MKF}$. Because $\mathbf{K}_n^{ST} = \mathbf{K}_n^{MKF}$, we also have that $\mathbf{X}_{n|n}^{ST} = \mathbf{X}_{n|n}^{MKF}$. $\square$

Hence Student's t-filter can be interpreted as computing estimates ξ_n of the common scaling of the noise covariance matrices while producing the same state estimates as the MKF. More can be said when $\mathbf{Q}^0$, $\mathbf{R}^0$, and $\mathbf{P}_0^0$ are merely misscaled.

Proposition 5. *Suppose that the model (1) takes the form*

$$
\begin{aligned}
\mathbf{X}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_0,\lambda\mathbf{P}_0^0), &\text{(4a)}\\
\mathbf{X}_n \mid \mathbf{X}_{n-1} &\sim \mathcal{N}(\mathbf{A}_n\mathbf{X}_{n-1},\lambda\mathbf{Q}^0), &\text{(4b)}\\
\mathbf{Y}_n \mid \mathbf{X}_n &\sim \mathcal{N}(\mathbf{C}_n\mathbf{X}_n,\lambda\mathbf{R}^0) &\text{(4c)}
\end{aligned}
$$

for some $\lambda > 0$. Then, in addition to relations in Prop. 4, we have

$$
\mathbb{E}[\mathbf{X}_n \mid \mathbf{Y}_{1:n}] = \mathbf{X}_{n|n}^{KF} = \mathbf{X}_{n|n}^{MKF} = \mathbf{X}_{n|n}^{ST}.
$$

Proof. In a manner similar to the proof of Prop. 4, it can be proved that $\mathbf{P}_{n|n}^{KF} = \lambda\mathbf{P}_{n|n}^{MKF}$ and $\mathbf{S}_n^{KF} = \lambda\mathbf{S}_n^{MKF}$. From these equations it follows that $\mathbf{X}_{n|n}^{KF} = \mathbf{X}_{n|n}^{MKF}$.
$\square$

B. Asymptotic Properties Student's t-Filter

This section adduces some asymptotic properties of Student's t-filter. We prove that ξ_n tends to a value optimal in the sense of Kullback–Leibler divergence (Thm. 6) and that, under the model (4), ξ_n converges to λ (Cor. 7). The results are based on the recursion

$$
\begin{aligned}
\xi_n &= \frac{\nu_0 + (n-1)d_y + \|\mathbf{Y}_n - \mathbf{Y}_{n|n-1}^{ST}\|_{[\mathbf{S}_n^{ST}]^{-1}}^2}{\nu_0 + nd_y}\xi_{n-1}\\
&= \frac{\nu_0 + (n-1)d_y}{\nu_0 + nd_y}\xi_{n-1} + \frac{\|\mathbf{Y}_n - \mathbf{Y}_{n|n-1}^{ST}\|_{[\mathbf{S}_n^{MKF}]^{-1}}^2}{\nu_0 + nd_y}
\end{aligned}
$$

from which it follows by induction and Prop. 4 that

$$\xi_n = \frac{\nu_0 + \sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}}}{\nu_0 + nd_y}. \quad (5)$$

Under certain assumptions, ξ_n converges to a value that is Kullback–Leibler optimal in the class of probability models that (possibly incorrectly) assume the noise covariances are merely misscaled; this is the class defined in (6). Note that it is not necessary to assume that (4) is the true model.

Theorem 6. *Let the sequence $\mathbf{Y}_{1:n}$ be governed by the model (1) and denote its joint density by $p^n(\mathbf{y}_{1:n})$. Define the density*

$$q_\eta^n(\mathbf{y}_{1:n}) = \prod_{i=1}^{n} \mathcal{N}(\mathbf{y}_i; \mathbf{y}_{i|i-1}^{\text{MKF}}, \eta \mathbf{S}_i^{\text{MKF}}) \quad (6)$$

for $\eta > 0$ and the Kullback–Leibler divergence minimiser

$$\eta_n = \arg\min_{\eta > 0} D_{\text{KL}}(p^n \| q_\eta^n).$$

Assume that $\lim_{n\to\infty} \mathbb{E}[\xi_n]^2/n^2 = \lim_{n\to\infty} \mathbb{V}[\xi_n] = 0$. Then

$$\lim_{n\to\infty} \mathbb{E}[(\xi_n - \eta_n)^2] = 0,$$

where the expectation is taken with respect to p^n.

Proof. The Kullback–Leibler divergence is given by

$$D_{\text{KL}}(p^n \| q_\eta^n)$$
$$= \mathbb{E}[\log p^n(\mathbf{Y}_{1:n})] + \frac{nd_y}{2}\log(2\pi\eta) + \frac{1}{2}\sum_{i=1}^{n} \log |\mathbf{S}_i^{\text{MKF}}|$$
$$+ \frac{1}{2\eta}\sum_{i=1}^{n} \mathbb{E}\Big[\|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}}\Big].$$

Differentiation gives

$$\frac{dD_{\text{KL}}(p^n \| q_\eta^n)}{d\eta}$$
$$= \frac{nd_y}{2\eta} - \frac{1}{2\eta^2}\sum_{i=1}^{n} \mathbb{E}\Big[\|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}}\Big].$$

Therefore,

$$\eta_n = \frac{1}{nd_y}\sum_{i=1}^{n} \mathbb{E}\Big[\|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}}\Big],$$

and from (5) it follows that

$$\eta_n = \frac{\nu_0 + nd_y}{nd_y}\mathbb{E}[\xi_n] - \frac{\nu_0}{nd_y} = \left(1 + \frac{\nu_0}{nd_y}\right)\mathbb{E}[\xi_n] - \frac{\nu_0}{nd_y},$$

and

$$\mathbb{E}[(\xi_n - \eta_n)^2] = \mathbb{V}[\xi_n] + \left(\frac{\nu_0}{nd_y}\right)^2(\mathbb{E}[\xi_n] + 1)^2,$$

which vanishes as $n \to \infty$ by assumption. $\quad\square$

If the true model is (4), it is a straightforward consequence of Thm. 6 that ξ_n asymptotically attains the correct scaling since, in this case, the true model belongs to the class (6) of approximating models.
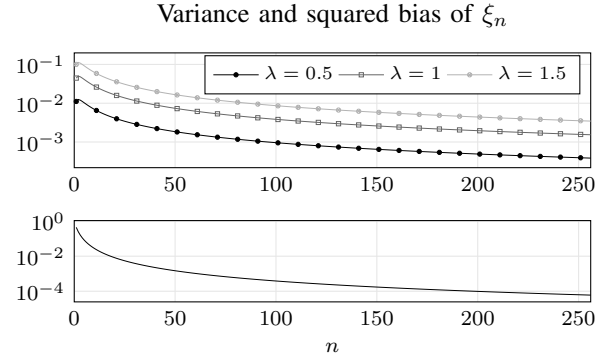
Variance and squared bias of ξ_n

Figure 1. Variance (above) and squared bias (below) of ξ_n for different λ. Note that the bias does not depend on λ. The value $\nu_0 = 10$ was used.

Corollary 7. *Under the model (4), $\eta_n = \lambda$ and*

$$\lim_{n\to\infty} \mathbb{E}[(\xi_n - \lambda)^2] = 0.$$

Proof. By assumptions of model (4), $p^n = q_\lambda^n$. Consequently, $D_{\text{KL}}(p^n \| q_\lambda^n) = 0$ and thus $\eta_n = \lambda$ is the global KL minimiser. Furthermore, from Prop. 4 it follows that

$$\|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}} = \lambda \|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{KF}}\|^2_{[\mathbf{S}_i^{\text{KF}}]^{-1}},$$

and $\|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{KF}}\|^2_{[\mathbf{S}_i^{\text{KF}}]^{-1}}$ are i.i.d. with a $\chi^2(d_y)$ distribution (see, e.g., [32]). From this it follows that

$$\xi_n = \frac{\nu_0}{\nu_0 + nd_y} + \frac{\lambda}{\nu_0 + nd_y}\sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{KF}}\|^2_{[\mathbf{S}_i^{\text{KF}}]^{-1}},$$

and

$$\mathbb{E}[\xi_n] = \frac{\nu_0 + \lambda nd_y}{\nu_0 + nd_y}, \quad \mathbb{V}[\xi_n] = \left(\frac{\lambda}{\nu_0 + nd_y}\right)^2 2nd_y. \quad (7)$$

The assumptions of Thm. 6 are thus satisfied and hence

$$\lim_{n\to\infty} \mathbb{E}[(\xi_n - \eta_n)^2] = \lim_{n\to\infty} \mathbb{E}[(\xi_n - \lambda)^2] = 0. \quad (8)$$

$\square$

Fig. 1 depicts variance and squared bias of ξ_n under the model (4).

C. Comparison to Maximal Likelihood

Here, ξ_n as an estimator of λ in (4) is compared to the maximum likelihood (ML) estimate. The derivative of the log-likelihood

$$\ell(\lambda) = \sum_{i=1}^{n} \log p(\mathbf{Y}_i \mid \mathbf{Y}_{1:i-1}),$$

is given by

$$\frac{d\ell}{d\lambda} = \frac{1}{2\lambda^2}\sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{ST}}\|^2_{[\mathbf{S}_i^{\text{MKF}}]^{-1}} - \frac{nd_y}{2\lambda}, \quad (9)$$

the ML estimate λ_n^{ML} is given by

$$\lambda_n^{\text{ML}} = \frac{\lambda}{nd_y}\sum_{i=1}^{n} \|\mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{ST}}\|^2_{[\mathbf{S}_i^{\text{KF}}]^{-1}}. \quad (10)$$
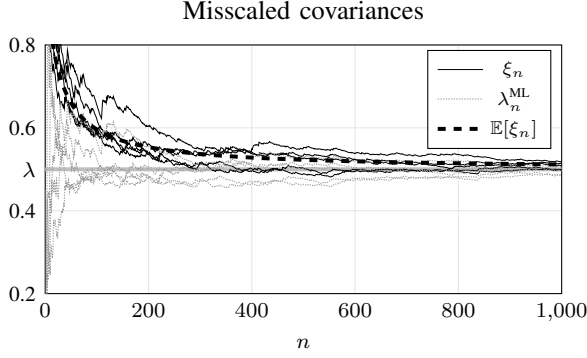
Misscaled covariances

Figure 2. Student's t-filter covariance scale estimates ξ_n and their theoretical mean $\mathbb{E}[\xi_n]$ and maximum likelihood estimates λ_n^{ML} given in (10) for five trajectories of the Wiener velocity model.

By Prop. 4 λ_n^{ML} is thus a scaled sum of i.i.d. $\chi^2(d_y)$ random variables. Therefore $\mathbb{E}[\lambda_n^{\text{ML}}] = \lambda$ and $\mathbb{V}[\lambda_n^{\text{ML}}] = 2\lambda/(nd_y)$. The relationship between ξ_n and λ_n^{ML} is elucidated in Thm. 8.

Theorem 8. *Under the model* (4), $\xi_n \to \lambda_n^{\text{ML}}$ *as* $\nu_0 \to 0$. *Furthermore*, $\Gamma^{-1}(s; \nu_0/2, \nu_0/2)$ *tends to Jeffrey's prior [33].*

Proof. From Prop. 4 and Eq. (5) it follows that

$$\lim_{\nu_0 \to 0} \xi_n = \frac{1}{nd_y} \sum_{i=1}^{n} \left\| \mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{ST}} \right\|_{[\mathbf{S}_i^{\text{MKF}}]^{-1}}^2$$
$$= \frac{\lambda}{nd_y} \left\| \mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}} \right\|_{[\lambda \mathbf{S}_i^{\text{MKF}}]^{-1}}^2$$
$$= \lambda_n^{\text{ML}}.$$

Furthermore, the log-likelihood derivative (9) is

$$\frac{d\ell}{d\lambda} = -\frac{nd_y}{2\lambda} + \frac{1}{2\lambda} \sum_{i=1}^{n} \left\| \mathbf{Y}_i - \mathbf{Y}_{i|i-1}^{\text{MKF}} \right\|_{[\mathbf{S}_i^{\text{KF}}]^{-1}}^2,$$

and hence the Fisher information is

$$\mathcal{I}(\lambda) = \mathbb{V}\left[\frac{d\ell}{d\lambda} \right] = \left(\frac{1}{2\lambda} \right)^2 \sum_{i=1}^{n} 2d_y.$$

Jeffrey's prior is then $p_{\text{J}}(\lambda) \propto \sqrt{\mathcal{I}(\lambda)} \propto \lambda^{-1}$ which is proportional to $\Gamma^{-1}(\lambda; \nu_0/2, \nu_0/2)$ as $\nu_0 \to 0$. $\square$

IV. A NUMERICAL EXAMPLE

For numerical examples of the Student's t-filter in Alg. 2 we use the Wiener velocity model with the model matrices

$$\mathbf{A}_n = \begin{bmatrix} \mathbf{I}_{2\times 2} & \Delta t \mathbf{I}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} \end{bmatrix} \quad \text{and} \quad \mathbf{C}_n = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

and true noise covariances

$$\mathbf{Q} = \lambda \begin{bmatrix} \frac{\Delta t^3}{3} \mathbf{I}_{2\times 2} & \frac{\Delta t^2}{2} \mathbf{I}_{2\times 2} \\ \frac{\Delta t^2}{2} \mathbf{I}_{2\times 2} & \Delta t \mathbf{I}_{2\times 2} \end{bmatrix} \text{ and } \mathbf{R} = \lambda \mathbf{I}_{2\times 2},$$

where $\lambda = 0.5$ and $\Delta t = 0.1$. The system was initialised with $\mathbf{P}_0 = \lambda \mathbf{I}_{4\times 4}$ and $\boldsymbol{\mu}_0 = \mathbf{0}_{4\times 1}$. To contrast ξ_n from λ_n^{ML}, similar when ν_0 is small by Thm. 8, we set $\nu_0 = 50$.
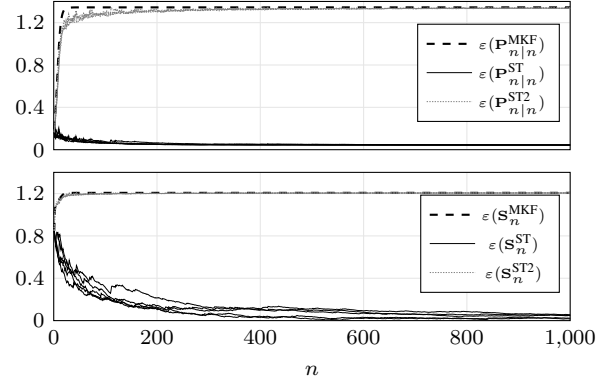
Structurally incorrect covariances

Figure 3. Frobenius errors (compared to the optimal Kalman filter) of the error covariance and measurement covariance produced by the ST, ST2, and MKF for five trajectories of the Wiener velocity model under structurally misspecified filter covariances (11). Note that the MKF covariances are data-independent.

First, we experiment with the scenario considered in Sec. III (i.e, covariance matrices the filter uses are misscaled versions of $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}_0$):

$$\lambda \mathbf{Q}^0 = \mathbf{Q}, \quad \lambda \mathbf{R}^0 = \mathbf{R}, \quad \text{and} \quad \lambda \mathbf{P}_0^0 = \mathbf{P}_0.$$

Fig. 2 depicts the scale estimate ξ_n produced by the Student's t-filter and the maximum likelihood estimate λ_n^{ML} of λ for five independent state trajectories. Also depicted is the mean of ξ_n, as given in (7). We see that the scale estimates converge to the the true value of λ as n increases.

In the second example we have a structural covariance misspecification:

$$\mathbf{Q}^0 = 0.1\,\mathbf{I}_{4\times 4}, \quad \mathbf{R}^0 = \mathbf{I}_{2\times 2}, \quad \text{and} \quad \mathbf{P}_0^0 = \mathbf{P}_0. \quad (11)$$

In this case it can only be expected that ξ_n converge to a value minimising some sort of joint distance between $\mathbf{Q}_n^{\text{ST}}$, $\mathbf{R}_n^{\text{ST}}$, and $\mathbf{P}_{n|n}^{\text{ST}}$ and $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{P}_{n|n}^{\text{KF}}$. Results appear in Fig. 3 for ST, ST2, and MKF in terms of the Frobenius errors

$$\varepsilon(\mathbf{P}_{n|n}) \coloneqq \left\| \mathbf{P}_{n|n} - \mathbf{P}_{n|n}^{\text{KF}} \right\|_{\text{Fro}} \text{ and } \varepsilon(\mathbf{S}_n) \coloneqq \left\| \mathbf{S}_n - \mathbf{S}_n^{\text{KF}} \right\|_{\text{Fro}},$$

latter of which is relevant in measurement gating [14, Sec. 2.3]. It is clear that Student's t-filter of Alg. 2 is superior in accuracy.

V. CONCLUSIONS

We analysed some properties of the Student's t-filter based on the joint model (2). Relations between the filter and misspecified Kalman filters were studied. Student's t-filter was shown to converge to a certain Kullback–Leibler optimal Kalman filter under some assumption as well as recover the correct scale when the noise covariance matrices are misscaled. The scale estimator was compared to the maximum likelihood estimator. Simulations involving the Wiener velocity model demonstrated the theoretical results.

An interesting future research direction would be to see if any analysis is possible for ST2 (recall Rmk. 3). However, as the formulation of this filter does not correspond to a joint probability model for the entire state and measurement sequence, different approaches and results ought to be sought.

References

[1] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy tailed process and measurement noise," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 5770–5774.

[2] F. Tronarp, R. Hostettler, and S. Särkkä, "Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise," in *International Conference on Information Fusion*, 2016, pp. 1859–1866.

[3] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "Robust Student's t based nonlinear filter and smoother," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2586–2596, 2016.

[4] Y. Huang, Y. Zhang, N. Li, S. M. Naqvi, and J. Chambers, "A robust Student's t based cubature filter," in *International Conference on Information Fusion*, 2016, pp. 9–16.

[5] M. Roth, "Kalman filtering for nonlinear systems and heavy-tailed noise," 2013, licenciate thesis, Linköpings universitet.

[6] ——, "Advanced Kalman filtering approaches to Bayesian state estimation," Ph.D. dissertation, Linköpings universitet, 2017.

[7] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust Student's *t*-based Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.

[8] J. Prüher, F. Tronarp, T. Karvonen, S. Särkkä, and O. Straka, "Student-*t* process quadratures for filtering of non-linear systems with heavy-tailed noise," in *International Conference on Information Fusion*, 2017.

[9] O. Straka and J. Duník, "Stochastic integration Student's-t filter," in *International Conference on Information Fusion*, 2017.

[10] P. Dong, Z. Jing, H. Leung, K. Shen, and J. Wang, "Student-*t* mixture labeled multi-Bernoulli filter for multi-target tracking with heavy-tailed noise," *Signal Processing*, vol. 152, pp. 331–339, 2018.

[11] F. Girón and J. Rojano, "Bayesian Kalman filtering with elliptically contoured errors," *Biometrika*, vol. 81, no. 2, pp. 390–395, 1994.

[12] P. J. Szablowski, "Elliptically contoured random variables and their application to the extension of the Kalman filter," *Computers & Mathematics with Applications*, vol. 19, no. 2, pp. 61–72, 1990.

[13] O. Barndorff-Nielsen, J. Kent, and M. Sørensen, "Normal variance-mean mixtures and *z* distributions," *International Statistical Review*, vol. 50, pp. 145–159, 1982.

[14] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, 1995.

[15] H. Kersting and P. Hennig, "Active uncertainty calibration in Bayesian ODE solvers," in *Uncertainty in Artificial Intelligence*, vol. 32, 2016, pp. 309–318.

[16] M. Schober, S. Särkkä, and P. Hennig, "A probabilistic model for the numerical solution of initial value problems," *Statistics and Computing*, 2018.

[17] F. Tronarp, H. Kersting, S. Särkkä, and P. Hennig, "Probabilistic solutions to ordinary differential equations as non-linear Bayesian filtering: A new perspective," 2018, arXiv preprint.

[18] R. K. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 15, no. 2, pp. 175–184, 1970.

[19] B. Friedland, "Estimating noise variances by using multiple observers," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-18, no. 4, pp. 442–448, 1982.

[20] J. Duník, M. Šimandl, and O. Straka, "Methods for estimating state and measurement noise covariance matrices: Aspects and comparison," in *IFAC Proceedings Volumes*, vol. 42, no. 10, 2009, pp. 372–377.

[21] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54, no. 3, pp. 596–600, 2009.

[22] T. Ardeshiri, E. Özkan, U. Orguner, and F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2450–2454, 2015.

[23] A. Assa and K. N. Plataniotis, "Adaptive Kalman filtering by covariance sampling," *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1288–1292, 2017.

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

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

[26] S. Sangsuk-Iam and T. E. Bullock, "Behavior of the discrete-time Kalman filter under incorrect noise covariances," in *IEEE Conference on Decision and Control*, 1987, pp. 1594–1600.

[27] ——, "Analysis of discrete-time Kalman filtering under incorrect noise covariances," *IEEE Transactions on Automatic Control*, vol. 35, no. 12, pp. 1304–1309, 1990.

[28] J. L. Willems and F. M. Callier, "Divergence of the stationary Kalman filter for correct and for incorrect noise variances," *IMA Journal of Mathematical Control and Information*, vol. 9, no. 1, pp. 47–54, 1992.

[29] S. Sangsuk-Iam, "Divergence of the discrete-time Kalman filter under incorrect noise covariances for linear periodic systems," in *American Control Conference*, 1994, pp. 1190–1194.

[30] S. S. Saab, "Discrete-time Kalman filter under incorrect noise covariances," in *American Control Conference*, 1995, pp. 1152–1156.

[31] S. Sangsuk-Iam and T. E. Bullock, "Analysis of continuous-time Kalman filtering under incorrect noise covariances," *Automatica*, vol. 24, no. 5, pp. 659–669, 1988.

[32] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, 2004.

[33] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 186, no. 1007, pp. 453–461, 1946.