

# Student- $t$ process quadratures for filtering of non-linear systems with heavy-tailed noise

Jakub Prüher, Filip Tronarp, Toni Karvonen, Simo Särkkä, and Ondřej Straka

This article has been published in *Proceedings of 20th International Conference on Information Fusion*, held in Xian, China, on July 10–13, 2017. When citing this article, use the following reference:

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 *20th International Conference on Information Fusion*, 2017.

DOI: 10.23919/ICIF.2017.8009742

**Copyright.** ©2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# Student- $t$ Process Quadratures for Filtering of Non-Linear Systems with Heavy-Tailed Noise

Jakub Prüher\*, Filip Tronarp†, Toni Karvonen†, Simo Särkkä†, and Ondřej Straka\*

\*Department of Cybernetics, University of West Bohemia, Pilsen, Czech Republic

†Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland

**Abstract**—The aim of this article is to design a moment transformation for Student- $t$  distributed random variables, which is able to account for the error in the numerically computed mean. We employ Student- $t$  process quadrature, an instance of Bayesian quadrature, which allows us to treat the integral itself as a random variable whose variance provides information about the incurred integration error. Advantage of the Student- $t$  process quadrature over the traditional Gaussian process quadrature, is that the integral variance depends also on the function values, allowing for a more robust modelling of the integration error. The moment transform is applied in nonlinear sigma-point filtering and evaluated on two numerical examples, where it is shown to outperform the state-of-the-art moment transforms.

## I. INTRODUCTION

State estimation problems arise in many engineering fields and sciences, such as global positioning system [1], tracking [2], [3] and finance [4]. In this paper, we are interested in designing a robust filter applicable in cases where the process and measurement noises are heavy-tailed. Student- $t$  filter for linear systems was presented in [5], where it was found to increase robustness with respect to assumptions on the noise statistics in the sense of mean error. It was later extended to the non-linear and non-additive noise case in [6] by generalising the Unscented transform to integration with respect to a Student- $t$  distribution; the robustness with respect to noise assumptions was also replicated. The key component of any local filtering algorithm is a moment transform (MT), which is responsible for computing the moments of a random variable transformed through a non-linear function. When the dynamical system is linear, the transformed moments can be computed analytically.

In the case of non-linear systems, numerical approximations are required. Many such approximations have been proposed over the years, including the well-known Unscented transform [7] (an instance of a fully symmetric integration rule [8], [6]), and the Gauss–Hermite quadrature [9], [10].

All the above-mentioned MTs are based on numerical quadrature rules that are approximative and make errors that go unaccounted for. Recently, [11] proposed to leverage the Bayesian quadrature (BQ) approach and presented a Gaussian process quadrature (GPQ) MT which is able to account for the integration error in a principled manner. In particular, Gaussian filters with the GPQ MTs were shown to exhibit better filter self-assessment [11], [12] properties than their classical quadrature counterparts. It is thus reasonable to expect similar improvements in the setting of Student- $t$  filtering [13], [6].

The foundation of the GPQ is the Gaussian process (GP) regression model [14], which is used as a surrogate probabilistic model of the integrand. Given the evaluation points (sigma-points), the mean function of the conditioned GP approximates the integrand, while the GP predictive variance informs about the lack of knowledge of the function behaviour in places where it was not evaluated. As the numerical quadrature approximations seek to work with as few integrand evaluations as possible to constrain the computational load, it is of paramount importance that the surrogate model predictive variance is as accurate as possible.

An attractive alternative to the GP is the Student- $t$  process (TP) regression [15], [16], which has the advantage that the model predictive variance also depends on the observed function values unlike in the case of the GP model. It can thus provide more accurate and robust predictive variances, which directly translate into the resulting integral quadrature approximations. Motivated by findings in [15] where it was concluded that “TP has many if not all of the benefits of GPs, but with increased modelling flexibility at no extra cost.”, we aim to leverage the TP regression model for the design of a Student- $t$  process quadrature (TPQ) MT. As far as we are aware, the use of TP in a quadrature context has not been previously attempted, although it has been suggested in [17, Section 2.1].

It should be noted that the transformation of moments is a more general problem as it arises in other applications, such as sensor system design [18] and optimal control [19]; however, the focus of this article is in sigma-point filtering [20]. We combine the proposed TPQ moment transform with the recent approach by [6] to sigma-point filtering of non-linear systems with heavy-tailed noise.

The article is structured as follows, Section II describes the Student- $t$  process quadrature, which is applied in Section III for moment transformation design. Section IV applies the proposed TPQ moment transform in Student- $t$  sigma-point filtering, while Section V presents the numerical experiments. Finally, Section VI concludes the article.

## II. STUDENT- $t$ PROCESS QUADRATURE

The key problem in moment transformations pertains to the computation of the moments of a transformed random variable. Given a random variable  $\mathbf{x} \in \mathbb{R}^D$  with a density function  $p$

and a non-linear integrand  $\mathbf{g}: \mathbb{R}^D \rightarrow \mathbb{R}^E$ , the goal of an MT is to compute

$$\boldsymbol{\mu} = \mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = \int \mathbf{g}(\mathbf{x})p(\mathbf{x}) \, d\mathbf{x}, \quad (1)$$

$$\boldsymbol{\Pi} = \mathbb{C}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = \int (\mathbf{g}(\mathbf{x}) - \boldsymbol{\mu})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\mu})^\top p(\mathbf{x}) \, d\mathbf{x}, \quad (2)$$

$$\mathbf{C} = \mathbb{C}_{\mathbf{x}}[\mathbf{x}, \mathbf{g}(\mathbf{x})] = \int (\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{g}(\mathbf{x}) - \boldsymbol{\mu})^\top p(\mathbf{x}) \, d\mathbf{x}. \quad (3)$$

In general, the above integrals cannot be evaluated analytically. There are essentially two ways of approximating them. The first one is to linearize the integrand (in the context of state estimation, this yields the extended Kalman filter and its relatives) and the second to use numerical quadrature which leads to variants of sigma-point filters. This section develops a probabilistic numerical integration rule based on modelling the integrand as a Student- $t$  process. The input density is considered to be the multivariate Student- $t$ , given by

$$\begin{aligned} p(\mathbf{x}) &= \text{St}(\mathbf{x} \mid \mathbf{m}, \boldsymbol{\Sigma}, \nu) \\ &= Z \left[ 1 + \frac{1}{\nu} (\mathbf{x} - \mathbf{m})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{m}) \right]^{-\frac{\nu+D}{2}}, \end{aligned} \quad (4)$$

where  $Z = [\Gamma((\nu+D)/2)]/[\Gamma(\nu/2)|\nu\pi\boldsymbol{\Sigma}|^{1/2}]$  is the normalizing constant and the mean  $\mathbf{m}$ , the scaling matrix  $\boldsymbol{\Sigma}$  and the degrees of freedom  $\nu$  are parameters. Note that  $\boldsymbol{\Sigma} = \frac{\nu-2}{\nu}\mathbf{P}$  defines relationship between the scaling and the covariance matrix. Its application to the MT problem is given in Section III.

A quadrature (or a sigma-point) rule is an approximation to the expectation (integral) of the transformed random variable  $\mathbf{g}(\mathbf{x})$  of the form

$$\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = \int \mathbf{g}(\mathbf{x})p(\mathbf{x}) \, d\mathbf{x} \approx \sum_{i=1}^N w_i \mathbf{g}(\mathbf{x}_i) \quad (5)$$

where  $w_i \in \mathbb{R}$  are the non-zero *weights* and  $\mathbf{x}_i \in \mathbb{R}^D$  are the *sigma-points* (or *nodes*). The classical approach to selecting the sigma-points and the weights is to choose them so that the rule (5) is exact whenever the coordinates of  $\mathbf{g}$  are multivariate polynomials of low degree. The most popular sigma-point rules used in Gaussian assumed density filtering such as the Unscented transform [7], an instance of a fully symmetric integration rule [8], and iterated Gauss–Hermite quadrature [9] belong to this category.

Alternatively, one can embrace the philosophy of *probabilistic numerics* [21], [22], [23], [24] and view the process of numerical computation of an integral as a problem of statistical inference. In this setting, the integrand is modelled as a stochastic process conditioned on the evaluations  $\mathbf{g}(\mathbf{x}_i)$  at the sigma-points  $\mathbf{x}_i$ . Injecting additional uncertainty through a stochastic process into the problem might seem counter-productive at first. However, after examining (5), we come to realize that the quadrature rule sees the function only through a finite number of function values—how the integrand behaves at other points is unknown. A probabilistic model for the integrand allows us to acknowledge this uncertainty and induces

a posterior probability distribution over the integral (5) itself. The posterior integral mean estimates the value of the integral while the posterior variance is construed as a model of the integration error. Numerical approximations of this sort go by the name *Bayesian quadrature* (BQ) [25], [26], [17].

Gaussian processes [14] have been a popular modelling choice in the BQ as prior distributions on functions due to their favourable analytical properties. Namely, when a GP distributed function is mapped through an integral, which is a *linear* operator, the integration result is also Gaussian distributed. GPQ has been applied to filtering problems in [27], [28] (a more general presentation on the MT problem can be found in [11]).

In this article, we consider the TP regression model as an attractive alternative to the GP, which we believe has potential to bring about significant improvements for the reasons outlined below. Since Student- $t$  distribution is invariant under affine transformations, TPs retain the favourable analytical properties of GPs, while providing increased modelling flexibility [15], namely:

- the distribution of the integral itself is Student- $t$ ,
- as opposed to GP, the model predictive variance additionally depends on the function values, which allows for more precise uncertainty modelling,
- and finally, a TP contains a GP as a special case (for infinite degrees of freedom (DoF)).

#### A. Student- $t$ Process Regression Model

From the perspective of the BQ, the TP regression model is a tool for modelling uncertainty in the knowledge of the numerically integrated function. Consider a real-valued function  $g: \mathbb{R}^D \rightarrow \mathbb{R}$  which is assigned a TP prior, such that  $g(\mathbf{x}) \sim \mathcal{TP}(0, k(\mathbf{x}, \mathbf{x}'), \nu_g)$ . This implies that for any finite collection of points  $\mathbf{x}'_1, \dots, \mathbf{x}'_m$  the function values are jointly Student- $t$  distributed with the degrees of freedom (DoF)  $\nu_g > 2$ . That is,

$$[g(\mathbf{x}'_1) \ \cdots \ g(\mathbf{x}'_m)] \sim \text{St}\left(\mathbf{0}, \frac{\nu_g-2}{\nu_g}\mathbf{K}, \nu_g\right), \quad (6)$$

where the kernel (covariance) matrix  $\mathbf{K}$  is made up of pairwise kernel evaluations, so that  $[\mathbf{K}]_{ij} = k(\mathbf{x}'_i, \mathbf{x}'_j; \boldsymbol{\theta})$ , where  $\boldsymbol{\theta}$  are the kernel parameters. For brevity, dependence on  $\boldsymbol{\theta}$  will be made explicit only when absolutely necessary. The choice of the positive-definite kernel  $k$  is up to the user and usually reflects expected smoothness of the underlying function. Given  $N$  observations  $\mathbf{y} = [g(\mathbf{x}_1), \dots, g(\mathbf{x}_N)]^\top$  at the evaluation points  $\mathbf{x}_1, \dots, \mathbf{x}_N$ , conditioning on the data  $\mathcal{D} = \{(\mathbf{x}_i, g(\mathbf{x}_i))\}_{i=1}^N$  results in a TP posterior with mean and variance [15], [16]

$$\mathbb{E}_g[g(\mathbf{x}) \mid \mathcal{D}] = \mathbf{k}^\top(\mathbf{x})\mathbf{K}^{-1}\mathbf{y}, \quad (7)$$

$$\mathbb{V}_g[g(\mathbf{x}) \mid \mathcal{D}] = \gamma \left[ k(\mathbf{x}, \mathbf{x}) - \mathbf{k}^\top(\mathbf{x})\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}) \right], \quad (8)$$

where  $\gamma = (\nu_g - 2 + \mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y}) / (\nu_g - 2 + N)$  and  $[\mathbf{k}(\mathbf{x})]_i = k(\mathbf{x}, \mathbf{x}_i)$ .

The posterior mean is identical to that of the GP regression, but the posterior variance (8) has the additional data-dependent scaling coefficient  $\gamma$ . This dependency on the function evaluation means that the TP regression is often more informative about the true underlying function than GP regression. Furthermore, the DoF is an additional tunable parameter allowing for control of the heavy-tailed process behaviour. The lower the DoF, the heavier the tails and vice versa. For  $\nu_g \rightarrow \infty$  a GP is recovered, which means the GP regression can be interpreted as a special case of the TP regression. Worth noting is that for increasing DoF the scaling factor becomes less dependent on the function values, eventually degrading to the GP predictive variance. Intuitively, one would assume that the predictive variance of the function would be affected by its own observations. For the GP this is not the case, implying that it is possible to know the predictive variance before the function observations are even obtained, which can be advantageous in certain applications. We argue, however, that in BQ applications, where limited datasets are encountered and an accurate quantification of uncertainty is crucial, the TP should be preferred, because of its superior characterization of uncertainty. The difference in the predictive variance can be seen in Figure 1, where the GP and TP are compared using the same values of kernel parameters. The mean functions of both models, which approximate the true underlying function, are identical. The TP is able to inflate the predictive variance due to its heavy-tailed nature, resulting in more realistic uncertainties in the function behaviour given the available data.

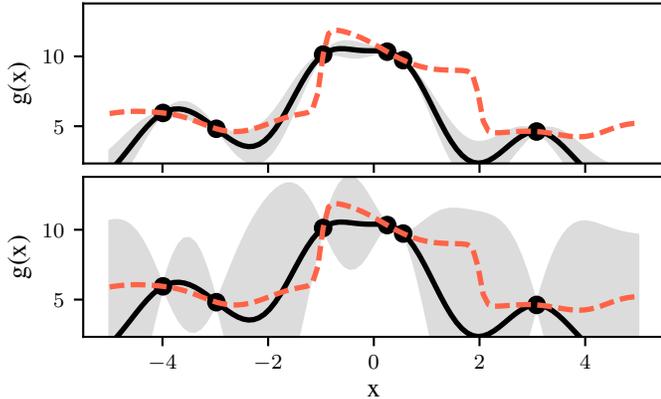


Figure 1: Comparison of predictive moments of the Gaussian process (top) and the Student- $t$  process (bottom) regression models using the same set of kernel parameters. The DoF of the TP model was set to  $\nu_g = 10$ . The true function (dashed red), the posterior mean (solid black) and predictive variance (gray band).

### B. Student- $t$ Process Quadrature Integral Moments

A TP posterior distribution over the integrand  $g(\mathbf{x})$  induces a posterior distribution over the integral  $\mathbb{E}_{\mathbf{x}}[g(\mathbf{x})] =$

$\int g(\mathbf{x})p(\mathbf{x}) d\mathbf{x}$ . The mean and variance of this distribution can be directly evaluated [26], [17]:

$$\mathbb{E}_g[\mathbb{E}_{\mathbf{x}}[g(\mathbf{x})] | \mathcal{D}] = \mathbb{E}_{\mathbf{x}}[\mathbb{E}_g[g(\mathbf{x}) | \mathcal{D}]] = \mathbf{q}^\top \mathbf{K}^{-1} \mathbf{y}, \quad (9)$$

$$\begin{aligned} \mathbb{V}_g[\mathbb{E}_{\mathbf{x}}[g(\mathbf{x})] | \mathcal{D}] &= \mathbb{E}_{\mathbf{x}, \mathbf{x}'}[\mathbb{C}_g[g(\mathbf{x}), g(\mathbf{x}') | \mathcal{D}]] \\ &= \gamma [\mathbb{E}_{\mathbf{x}, \mathbf{x}'}[k(\mathbf{x}, \mathbf{x}')] - \mathbf{q}^\top \mathbf{K}^{-1} \mathbf{q}], \end{aligned} \quad (10)$$

where  $\mathbf{q} = \mathbb{E}_{\mathbf{x}}[\mathbf{k}(\mathbf{x})]$ . In Equation (9), the mean of the integral is identical to integrating the TP posterior mean function. The only difference to the GPQ is in the integral posterior variance (10) having a data-dependent coefficient. Note that the integral posterior mean (9) is indeed a quadrature rule of the form (5) with the weights  $w_i = [\mathbf{q}^\top \mathbf{K}^{-1}]_i$ . Integrals of general vector functions  $\mathbf{g}: \mathbb{R}^D \rightarrow \mathbb{R}^E$  can be evaluated by applying the above equations to each function output independently with the same kernel and DoF used for each output.

### III. STUDENT- $t$ PROCESS QUADRATURE MOMENT TRANSFORM

In this section, we apply the results for the integral moments from the previous section for derivation of our proposed general purpose TPQ moment transform. The results from this section are aimed to be applied later in Section IV for approximation of the state and measurement moments in a Student- $t$  sigma-point filtering setting.

First, consider a general case of non-linear transformation of arbitrarily distributed random variable

$$\mathbf{y} = \mathbf{g}(\mathbf{x}), \quad \mathbf{x} \sim p(\mathbf{x}) \quad (11)$$

where  $\mathbf{g}: \mathbb{R}^D \rightarrow \mathbb{R}^E$  is such that the first two moments of  $\mathbf{y}$  exist. In the BQ, when the uncertainty in  $\mathbf{g}(\mathbf{x})$  is introduced through a posterior stochastic process regression model, the general BQ approximations to the moments in Equations (1) to (3) need to account for it, resulting in

$$\mathbb{E}[\mathbf{y}] = \mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] \approx \mathbb{E}_{\mathbf{g}, \mathbf{x}}[\mathbf{g}(\mathbf{x})], \quad (12)$$

$$\mathbb{C}[\mathbf{y}] = \mathbb{C}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] \approx \mathbb{C}_{\mathbf{g}, \mathbf{x}}[\mathbf{g}(\mathbf{x})], \quad (13)$$

$$\mathbb{C}[\mathbf{x}, \mathbf{y}] = \mathbb{C}[\mathbf{x}, \mathbf{g}(\mathbf{x})] \approx \mathbb{C}_{\mathbf{g}, \mathbf{x}}[\mathbf{x}, \mathbf{g}(\mathbf{x})]. \quad (14)$$

The BQ transformed mean can be written out, using the law of total expectation, as

$$\mathbb{E}_{\mathbf{g}, \mathbf{x}}[\mathbf{g}(\mathbf{x})] = \mathbb{E}_{\mathbf{g}}[\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})]] = \mathbb{E}_{\mathbf{x}}[\mathbb{E}_{\mathbf{g}}[\mathbf{g}(\mathbf{x})]], \quad (15)$$

which shows that the expectation of the integral is equivalent to integrating the model mean function (cf. Equation (9)). Applying the law of total covariance, we can decompose the transformed covariance in two ways

$$\mathbb{C}_{\mathbf{g}, \mathbf{x}}[\mathbf{g}(\mathbf{x})] = \mathbb{E}_{\mathbf{g}}[\mathbb{C}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})]] + \mathbb{C}_{\mathbf{g}}[\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})]] \quad (16)$$

$$= \mathbb{E}_{\mathbf{x}}[\mathbb{C}_{\mathbf{g}}[\mathbf{g}(\mathbf{x})]] + \mathbb{C}_{\mathbf{x}}[\mathbb{E}_{\mathbf{g}}[\mathbf{g}(\mathbf{x})]]. \quad (17)$$

The first decomposition reveals the fact that the general BQ MT incorporates integral variance as an additional term in transformed variance. By now it is evident that the MT based on BQ can be interpreted as a principled covariance inflation scheme. Note that for deterministic  $\mathbf{g}$  the covariance decomposition in Equation (16) would revert back to the

classical definition of the covariance. In the BQ, however, the integrand uncertainty is crucial for modelling the integration error. The second decomposition is computationally beneficial because the individual terms can be computed in closed form for suitable kernel-density pairs.

In order to derive the TPQ MT for  $\mathbf{x} \sim \text{St}(\mathbf{m}, \mathbf{P}, \nu)$ , we utilize the familiar stochastic decoupling substitution  $\mathbf{x} = \mathbf{m} + \mathbf{L}\boldsymbol{\xi}$ , which allows for casting the expectations in terms of a standard Student- $t$  random variable  $\boldsymbol{\xi} \sim \text{St}(\mathbf{0}, \mathbf{I}, \nu)$ , so that

$$\mathbb{E}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = \mathbb{E}_{\boldsymbol{\xi}}[\mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})] \approx \mathbb{E}_{\mathbf{g}, \boldsymbol{\xi}}[\mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})], \quad (18)$$

$$\mathbf{C}_{\mathbf{x}}[\mathbf{g}(\mathbf{x})] = \mathbf{C}_{\boldsymbol{\xi}}[\mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})] \approx \mathbf{C}_{\mathbf{g}, \boldsymbol{\xi}}[\mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})], \quad (19)$$

$$\mathbf{C}_{\mathbf{x}}[\mathbf{x}, \mathbf{g}(\mathbf{x})] = \mathbf{C}_{\boldsymbol{\xi}}[\boldsymbol{\xi}, \mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})] \approx \mathbf{C}_{\mathbf{g}, \boldsymbol{\xi}}[\boldsymbol{\xi}, \mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})], \quad (20)$$

where  $\mathbf{L}\mathbf{L}^\top = \mathbf{P}$ . For notational brevity we define  $\tilde{\mathbf{g}}(\boldsymbol{\xi}) \triangleq \mathbf{g}(\mathbf{m} + \mathbf{L}\boldsymbol{\xi})$ . Using Equations (9) and (15), the transformed mean becomes

$$\mathbb{E}_{\mathbf{g}, \boldsymbol{\xi}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})] = \mathbb{E}_{\boldsymbol{\xi}}[\mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]] = \mathbf{Y}^\top \mathbf{K}^{-1} \mathbb{E}_{\boldsymbol{\xi}}[\mathbf{k}(\boldsymbol{\xi})], \quad (21)$$

where  $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_E]$  contains observations of each output of  $\mathbf{g}$  in columns. For the transformed covariance we employ Equation (16) and obtain

$$\mathbf{C}_{\mathbf{g}, \boldsymbol{\xi}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})] = \mathbb{E}_{\boldsymbol{\xi}}[\mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})] \mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]^\top] - \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^\top + \mathbb{E}_{\boldsymbol{\xi}}[\mathbf{C}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]], \quad (22)$$

where, after plugging in from Equation (9), the first term becomes

$$\mathbb{E}_{\boldsymbol{\xi}}[\mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})] \mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]^\top] = \mathbf{Y}^\top \mathbf{K}^{-1} \mathbb{E}_{\boldsymbol{\xi}}[\mathbf{k}(\boldsymbol{\xi}) \mathbf{k}(\boldsymbol{\xi})^\top] \mathbf{K}^{-1} \mathbf{Y} \quad (23)$$

and the third term is

$$\mathbb{E}_{\boldsymbol{\xi}}[\mathbf{C}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]] = \text{diag}([s_{\text{TP}}^1 \ \dots \ s_{\text{TP}}^E]) \quad (24)$$

where

$$s_{\text{TP}}^e = \gamma_e [\mathbb{E}_{\boldsymbol{\xi}}[k(\boldsymbol{\xi}, \boldsymbol{\xi})] - \text{tr}(\mathbb{E}_{\boldsymbol{\xi}}[\mathbf{k}(\boldsymbol{\xi}) \mathbf{k}(\boldsymbol{\xi})^\top] \mathbf{K}^{-1})] \quad (25)$$

and  $\gamma_e = (\nu_g - 2 + \mathbf{y}_e^\top \mathbf{K}^{-1} \mathbf{y}_e) / (\nu_g - 2 + N)$ . Finally, for the input-output covariance we have

$$\mathbf{C}_{\mathbf{g}, \boldsymbol{\xi}}[\boldsymbol{\xi}, \tilde{\mathbf{g}}(\boldsymbol{\xi})] = \mathbf{L} \mathbb{E}_{\boldsymbol{\xi}}[\boldsymbol{\xi} \mathbb{E}_{\mathbf{g}}[\tilde{\mathbf{g}}(\boldsymbol{\xi})]] = \mathbf{L} \mathbb{E}_{\boldsymbol{\xi}}[\boldsymbol{\xi} \mathbf{k}(\boldsymbol{\xi})^\top] \mathbf{K}^{-1} \mathbf{Y}. \quad (26)$$

The following definition gathers our results so far.

**Definition 1** (General TPQ moment transform). The general Student- $t$  process quadrature approximation to the joint distribution of  $\mathbf{x} \sim \text{St}(\mathbf{m}, \frac{\nu-2}{\nu} \mathbf{P}, \nu)$  and a transformed random variable  $\mathbf{y} = \mathbf{g}(\mathbf{x})$  is given by

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \text{St} \left( \begin{bmatrix} \mathbf{m} \\ \hat{\boldsymbol{\mu}} \end{bmatrix}, \frac{\nu-2}{\nu} \begin{bmatrix} \mathbf{P} & \hat{\mathbf{C}} \\ \hat{\mathbf{C}}^\top & \hat{\boldsymbol{\Pi}} \end{bmatrix}, \nu \right) \quad (27)$$

where the transformed moments are

$$\hat{\boldsymbol{\mu}} = \mathbf{Y}^\top \mathbf{w}, \quad (28)$$

$$\hat{\boldsymbol{\Pi}} = \mathbf{Y}^\top \mathbf{W} \mathbf{Y} - \hat{\boldsymbol{\mu}} \hat{\boldsymbol{\mu}}^\top + \mathbf{S}, \quad (29)$$

$$\hat{\mathbf{C}} = \mathbf{L} \mathbf{W}_c \mathbf{Y}, \quad (30)$$

$$\mathbf{S} = \text{diag}([s_{\text{TP}}^1 \ \dots \ s_{\text{TP}}^E]), \quad (31)$$

$$s_{\text{TP}}^e = \frac{\nu_g - 2 + \mathbf{y}_e^\top \mathbf{K}^{-1} \mathbf{y}_e}{\nu_g - 2 + N} [\bar{k} - \text{tr}(\mathbf{Q} \mathbf{K}^{-1})] \quad (32)$$

and  $\mathbf{Y} = [\mathbf{y}_1 \ \dots \ \mathbf{y}_E]$ , where the  $e$ -th column  $\mathbf{y}_e = [g_e(\mathbf{x}_1) \ \dots \ g_e(\mathbf{x}_N)]^\top$  contains function values of the  $e$ -th output of  $\mathbf{g}(\mathbf{x})$ . The sigma-points are given by  $\mathbf{x}_i = \mathbf{m} + \mathbf{L}\boldsymbol{\xi}_i$ , where  $\mathbf{L}\mathbf{L}^\top = \mathbf{P}$ . The elements of the kernel matrix are  $[\mathbf{K}]_{ij} = k(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j; \boldsymbol{\theta})$  and the TPQ weights are  $\mathbf{w} = \mathbf{K}^{-1} \mathbf{q}$ ,  $\mathbf{W} = \mathbf{K}^{-1} \mathbf{Q} \mathbf{K}^{-1}$  and  $\mathbf{W}_c = \mathbf{R} \mathbf{K}^{-1}$ , where

$$[\mathbf{q}]_i = \mathbb{E}_{\boldsymbol{\xi}}[k(\boldsymbol{\xi}, \boldsymbol{\xi}_i; \boldsymbol{\theta})], \quad (33)$$

$$[\mathbf{Q}]_{ij} = \mathbb{E}_{\boldsymbol{\xi}}[k(\boldsymbol{\xi}, \boldsymbol{\xi}_i; \boldsymbol{\theta}) k(\boldsymbol{\xi}, \boldsymbol{\xi}_j; \boldsymbol{\theta})], \quad (34)$$

$$[\mathbf{R}]_{*j} = \mathbb{E}_{\boldsymbol{\xi}}[\mathbf{x} k(\boldsymbol{\xi}, \boldsymbol{\xi}_j; \boldsymbol{\theta})], \quad (35)$$

$$\bar{k} = \mathbb{E}_{\boldsymbol{\xi}}[k(\boldsymbol{\xi}, \boldsymbol{\xi}; \boldsymbol{\theta})]. \quad (36)$$

The notation  $[\mathbf{R}]_{*j}$  stands for the  $j$ th column of the matrix  $\mathbf{R}$ . The set of unit sigma-points  $\{\boldsymbol{\xi}_i : i = 1, \dots, N\}$  can be chosen *arbitrarily*.

The transform is general in a sense that it can, in principle, operate with any kernel. Since decoupling is used in the moment integrals, the kernel expectations in (33)–(36) do not depend on the parameters of the distribution of  $\mathbf{x}$  and the TPQ weights can be fully pre-computed. This fact significantly eases computational burden when our proposed TPQ moment transform is later applied in Student- $t$  sigma-point filtering.

From the above summary it is apparent that the TPQ MT requires evaluation of the kernel expectations in (33)–(36). The popular radial basis function (RBF) kernel, given by

$$k(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\theta}) = s^2 \exp \left( -\frac{1}{2} (\boldsymbol{\xi} - \boldsymbol{\xi}')^\top \boldsymbol{\Lambda}^{-1} (\boldsymbol{\xi} - \boldsymbol{\xi}') \right), \quad (37)$$

where  $s, \boldsymbol{\Lambda} = \text{diag}([\ell_1^2 \ \dots \ \ell_D^2])$  are kernel parameters collected into a vector  $\boldsymbol{\theta}$ , has been used in previous work [28]. This kernel admits a closed-form evaluation of the expectations in (33)–(36) for a Gaussian distributed  $\mathbf{x}$ . However, in our case, where  $\mathbf{x}$  is Student- $t$  distributed, we have been unable to find any kernel admitting closed-form solution. For this reason, we used the RBF kernel and resorted to the standard Monte Carlo numerical approximation. Fortunately, in filtering this is not really a problem, because the expectations, and consequently the TPQ weights, can be fully pre-computed offline.

#### IV. TPQ STUDENT- $t$ FILTER

The aim of this section is to present the TPQ Student- $t$  filter harnessing the proposed TPQ MT presented in Section III and enriching the work of [6] with the BQ philosophy. We first review the general Student- $t$  filtering framework for non-additive noise setting, and then later apply the TPQ MT for construction of the TPQ Student- $t$  filter.

Consider the following discrete-time state-space model

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1}), \quad (38)$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k), \quad (39)$$

where (38) describes the evolution of the system state  $\mathbf{x}_k$  in time and (39) describes the process by which measurements  $\mathbf{z}_k$  are generated. The function  $\mathbf{f}: \mathbb{R}^{d_x} \times \mathbb{R}^{d_q} \rightarrow \mathbb{R}^{d_x}$  is the system dynamics and  $\mathbf{h}: \mathbb{R}^{d_x} \times \mathbb{R}^{d_r} \rightarrow \mathbb{R}^{d_z}$  is the measurement function. The variables  $\mathbf{q}_k$  and  $\mathbf{r}_k$  represent the zero-mean process and measurement noises with known covariance matrices  $\mathbf{Q}_k$  and  $\mathbf{R}_k$ , respectively.

The source of novelty in Student- $t$  filter comes from the conditioning formula for Student- $t$  random variables [29], [5]. The measurement update equations can be derived by assuming that the state and the measurement are jointly Student- $t$  distributed, such that

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix} \sim \text{St} \left( \begin{bmatrix} \mathbf{m}_{k|k-1}^x \\ \mathbf{m}_{k|k-1}^z \end{bmatrix}, \frac{\nu-2}{\nu} \begin{bmatrix} \mathbf{P}_{k|k-1}^x & \mathbf{P}_{k|k-1}^{xz} \\ \mathbf{P}_{k|k-1}^{zx} & \mathbf{P}_{k|k-1}^z \end{bmatrix}, \nu \right). \quad (40)$$

Then, the conditioned state is distributed according to  $\mathbf{x}_k | \mathbf{z}_{1:k} \sim \text{St}(\mathbf{m}_{k|k}^x, \frac{\nu^*-2}{\nu^*} \mathbf{P}_{k|k}^x, \nu^*)$  with the statistics given by

$$\mathbf{m}_{k|k}^x = \mathbf{m}_{k|k-1}^x + \mathbf{P}_{k|k-1}^{xz} (\mathbf{P}_{k|k-1}^z)^{-1} (\mathbf{z}_k - \mathbf{m}_{k|k-1}^z), \quad (41)$$

$$\mathbf{P}_{k|k}^x = \frac{\nu_x - 2 + \beta}{\nu_x - 2 + d_z} \left( \mathbf{P}_{k|k-1}^x - \mathbf{P}_{k|k-1}^{xz} (\mathbf{P}_{k|k-1}^z)^{-1} \mathbf{P}_{k|k-1}^{zx} \right), \quad (42)$$

$$\beta = (\mathbf{z}_k - \mathbf{m}_{k|k-1}^z)^\top (\mathbf{P}_{k|k-1}^z)^{-1} (\mathbf{z}_k - \mathbf{m}_{k|k-1}^z), \quad (43)$$

$$\nu^* = \nu + d_z. \quad (44)$$

These equations constitute the Student- $t$  measurement update rule, where  $\mathbf{m}_{k|k}^x$  and  $\mathbf{P}_{k|k}^x$  are the filtered state mean and covariance, respectively. It is instructive to consider the case where  $\mathbf{z}_k$  is Gaussian, which entails  $\beta \sim \chi^2(d_z)$  and  $\mathbb{E}[\beta] = d_z$ . Thus the posterior covariance  $\mathbf{P}_{k|k}^x$  either increases or decreases depending on the outcome of  $\beta$  in relation to its expected value under a  $\chi^2$  assumption.

The means and covariances in (40), necessary for the update, can be approximated by *any* moment transform. The predicted state mean  $\mathbf{m}_{k|k-1}^x$  and covariance  $\mathbf{P}_{k|k-1}^x$  are computed by using the system dynamics  $\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1})$  in place of the general non-linear transformation from Equation (11) with the input variable distributed according to

$$\begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{q}_{k-1} \end{bmatrix} \sim \text{St} \left( \begin{bmatrix} \mathbf{m}_{k-1|k-1}^x \\ 0 \end{bmatrix}, \frac{\nu-2}{\nu} \begin{bmatrix} \mathbf{P}_{k-1|k-1}^x & 0 \\ 0 & \mathbf{Q} \end{bmatrix}, \nu \right). \quad (45)$$

The same applies for the measurement mean  $\mathbf{m}_{k|k-1}^z$ , covariance  $\mathbf{P}_{k|k-1}^z$  and cross-covariance  $\mathbf{P}_{k|k-1}^{xz}$ , which are computed by using the measurement function  $\mathbf{h}(\mathbf{x}_k, \mathbf{r}_k)$  in place of the non-linear transformation from Equation (11) with input variable distributed according to

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{r}_k \end{bmatrix} \sim \text{St} \left( \begin{bmatrix} \mathbf{m}_{k|k-1}^x \\ 0 \end{bmatrix}, \frac{\nu-2}{\nu} \begin{bmatrix} \mathbf{P}_{k|k-1}^x & 0 \\ 0 & \mathbf{R} \end{bmatrix}, \nu \right). \quad (46)$$

Once the approximations to the predictive state and measurement moments are available, a joint Student- $t$  approximation in Equation (40) can be formed with the desired degrees of freedom  $\nu$ . This subsequently enables the use of the update rule given by Equations (41) to (44).

Notice, the DoF update in Equation (44) entails that  $\nu^* \rightarrow \infty$  for  $k \rightarrow \infty$ , which means the filter will asymptotically behave as a Kalman filter. In order to continue to operate at desired DoF  $\nu$ , we opt for the moment matching perspective used in [5], [6] and assume that the conditioned state  $\mathbf{x}_k | \mathbf{z}_{1:k} \sim \text{St}(\mathbf{m}_{k|k}^x, \frac{\nu-2}{\nu} \mathbf{P}_{k|k}^x, \nu)$ .

The entire procedure of is outlined in Algorithm 1. Note that any sigma-points can be used for the TPQ MT in this algorithm.

**Algorithm 1** (One step of the TPQ Student- $t$  filter).

*Input:* filtered mean  $\mathbf{m}_{k-1|k-1}^x$ , filtered covariance  $\mathbf{P}_{k-1|k-1}^x$ , desired DoF  $\nu$

*Output:* filtered mean  $\mathbf{m}_{k|k}^x$  and covariance  $\mathbf{P}_{k|k}^x$

- 1) Use the TPQ MT with kernel parameters  $\theta_f$  to compute the predictive state moments  $\mathbf{m}_{k|k-1}^x$  and  $\mathbf{P}_{k|k-1}^x$  using (28) and (29) assuming the input is distributed according (45).
- 2) Use TPQ MT with kernel parameters  $\theta_h$  to compute the predictive measurement moments  $\mathbf{m}_{k|k-1}^z$ ,  $\mathbf{P}_{k|k-1}^z$  and  $\mathbf{P}_{k|k-1}^{xz}$  using (28) to (30) assuming the input is distributed according (46).
- 3) Use the Student- $t$  measurement update in (41) and (42) to compute the filtered mean  $\mathbf{m}_{k|k}^x$  and covariance  $\mathbf{P}_{k|k}^x$ .
- 4) Fix the DoF by putting  $\text{St}(\mathbf{x}_k | \mathbf{m}_{k|k}^x, \frac{\nu-2}{\nu} \mathbf{P}_{k|k}^x, \nu) \approx \text{St}(\mathbf{x}_k | \mathbf{m}_{k|k}^x, \frac{\nu^*-2}{\nu^*} \mathbf{P}_{k|k}^x, \nu^*)$ .

## V. EXPERIMENTAL RESULTS

In this section, we compare the performance of the proposed TPQ-based Student- $t$  filters and the Student- $t$  filter introduced in [6], which is based on classical quadrature. In all experiments, we are measuring the filter error by the root mean square error (RMSE)

$$\text{RMSE} = \left( \frac{1}{K} \sum_{k=1}^K \|\mathbf{x}_k - \mathbf{m}_{k|k}^x\|^2 \right)^{1/2}. \quad (47)$$

Since the BQ-based MTs, such as TPQ or GPQ, are primarily focused on incorporating additional uncertainty by inflating the estimated covariance, we used the inclination indicator (INC) [12] as a metric which takes into account the estimated state covariance. The indicator is given by

$$\text{INC} = \frac{10}{K} \sum_{k=1}^K \log_{10} \frac{(\mathbf{x}_k - \mathbf{m}_{k|k}^x)^\top (\mathbf{P}_{k|k}^x)^{-1} (\mathbf{x}_k - \mathbf{m}_{k|k}^x)}{(\mathbf{x}_k - \mathbf{m}_{k|k}^x)^\top \Sigma_k^{-1} (\mathbf{x}_k - \mathbf{m}_{k|k}^x)}, \quad (48)$$

where  $\Sigma_k$  is the sample mean-square-error (MSE) matrix, which can be computed from samples of the true system state trajectories. When the indicator is  $\text{INC} = 0$  the estimator is said to be *balanced*, which is to say that the estimated covariance is on average equal to the true state MSE matrix.

For  $\text{INC} > 0$  the estimator is said to be *optimistic* while for  $\text{INC} < 0$  it is considered *pessimistic*.

### A. Univariate Non-Stationary Growth Model

In the first numerical illustration, we consider the univariate non-stationary growth model (UNGM), which is often used for benchmarking purposes [30], [31]. The system is given by the following set of equations

$$x_k = 0.5x_{k-1} + \frac{25x_{k-1}}{1 + x_{k-1}^2} + 8 \cos(1.2k) + q_{k-1}, \quad (49)$$

$$z_k = 0.05x_k^2 + r_k. \quad (50)$$

The initial conditions were drawn from  $x_0 \sim \mathcal{N}(0, 1)$ . Outliers in the state noise  $q_k$  and measurement noise  $r_k$  were simulated with Gaussian mixtures, such that  $q_k \sim 0.8\mathcal{N}(0, \sigma_q^2) + 0.2\mathcal{N}(0, 10\sigma_q^2)$  and  $r_k \sim 0.8\mathcal{N}(0, \sigma_r^2) + 0.2\mathcal{N}(0, 100\sigma_r^2)$ , where  $\sigma_q^2 = 10$  and  $\sigma_r^2 = 0.01$ . We simulated 500 trajectories for 250 time steps, which were used for evaluation of the RMSE and INC performance metrics. All tested filters used a state-

	RMSE	STD	INC	STD
UKF	8.6924	0.1517	3.0012	0.1539
SF	17.4461	0.6236	51.8733	0.4417
TPQSF( $\nu_g = 3$ )	7.5683	0.1091	1.5837	0.1561
TPQSF( $\nu_g = 4$ )	6.8323	0.1044	2.3384	0.1713
TPQSF( $\nu_g = 10$ )	6.1423	0.0154	5.6910	0.0324
TPQSF( $\nu_g = 100$ )	7.4399	0.1550	12.5120	0.1676
TPQSF( $\nu_g = 500$ )	7.5709	0.1546	13.2104	0.1623
GPQSF	7.6766	0.1554	13.5926	0.1595

Table I: Performance of TPQSF compared in terms of average RMSE and INC. Standard deviations of the criteria were estimated by bootstrapping. For increasing DOF parameter  $\nu_g$  of the TP regression model the performance approaches that of the GPQSF.

space model with an initial condition distributed according to  $x_0 \sim \text{St}(0, \frac{\nu-2}{\nu}1, \nu)$  and the following noise statistics  $q_k \sim \text{St}(0, \frac{\nu-2}{\nu}\sigma_q^2, \nu)$ ,  $r_k \sim \text{St}(0, \frac{\nu-2}{\nu}\sigma_r^2, \nu)$ , where  $\nu = 4$ .

We compared the RMSE and INC of our proposed TPQSF with the SF [6], the UKF [7] and a Student- $t$  filter using the GPQ MT [28] (abbreviated GPQSF). The TPQSF and the GPQSF will be collectively referred to as the BQ filters. Student- $t$  filters used the same 3rd degree fully symmetric sigma-point set with  $\kappa = 0$  and the filter DoF fixed at  $\nu = 4$ . The kernel parameters for all BQ filters were set to  $\theta_f = [3 \ 1]$  and  $\theta_h = [3 \ 3]$ . Table I reports MC simulation averages of both metrics along with bootstrapped variances [32] of the averages (using 10,000 samples). It is evident that the TPQSFs can outperform all the classical filters (UKF, SF) as well as the GPQSF in terms of both metrics. The values of INC, being closer to zero, indicate increased estimate credibility. For increasing DoF of the Student- $t$  process model, we observe the performance of TPQSFs approaching that of GPQSF, which is an expected behaviour, since TPQSF with  $\nu_g = \infty$  is equivalent to GPQSF.

### B. Radar Tracking with Glint Noise

As a second illustration, we consider tracking of a moving object where the range and bearing measurements are corrupted with glint noise. We adopt the example from [33], where the tracking scenario is described by the following state-space model

$$\mathbf{x}_k = \begin{bmatrix} 1 & \tau & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} \tau^2/2 & 0 \\ \tau & 0 \\ 0 & \tau^2/2 \\ 0 & \tau \end{bmatrix} \mathbf{q}_{k-1} \quad (51)$$

$$\mathbf{z}_k = \begin{bmatrix} \sqrt{x_k^2 + y_k^2} \\ \text{atan2}(y_k, x_k) \end{bmatrix} + \mathbf{r}_k \quad (52)$$

with the system state being defined as  $\mathbf{x}_k = [x_k \ \dot{x}_k \ y_k \ \dot{y}_k]$ . The state components  $x_k$  and  $y_k$  are the Cartesian coordinates of the moving object and the pair  $\dot{x}_k, \dot{y}_k$  stands for the velocity in the respective directions. During simulations, the discretization interval was  $\tau = 0.5$  s, the initial state was drawn from  $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$  with  $\mathbf{m}_0 = [10\ 000 \text{ m} \ 300 \text{ ms}^{-1} \ 1000 \text{ m} \ -40 \text{ ms}^{-1}]$  and  $\mathbf{P}_0 = \text{diag}([10\ 000 \text{ m}^2 \ 100 \text{ m}^2 \text{s}^{-2} \ 10\ 000 \text{ m}^2 \ 100 \text{ m}^2 \text{s}^{-2}])$ . The state noise is Gaussian distributed, such that  $\mathbf{q}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$  with covariance  $\mathbf{Q} = \text{diag}([50 \text{ m}^2 \text{s}^{-4} \ 5 \text{ m}^2 \text{s}^{-4}])$ . The glint noise in the measurements is modelled by a Gaussian mixture

$$\mathbf{r}_k \sim (1 - \beta)\mathcal{N}(\mathbf{0}, \mathbf{R}_1) + \beta\mathcal{N}(\mathbf{0}, \mathbf{R}_2) \quad (53)$$

with  $\mathbf{R}_1 = \text{diag}([50 \text{ m}^2 \ 0.4 \text{ mrad}^2])$  and  $\mathbf{R}_2 = \text{diag}([5000 \text{ m}^2 \ 16 \text{ mrad}^2])$ , where  $\beta$  is the glint noise probability.

As in the UNGM experiment, we compared the performance of our proposed TPQSF with the SF, the standard UKF and the GPQSF. The UKF used  $\kappa = 0$ , following the usual heuristic recommendation. For the TPQSF we considered two settings of the TP model DoF parameter,  $\nu_g = 2.2$  and  $\nu_g = 4$ . All of the Student- $t$  filters assumed that the initial state, the state noise and the measurement noise were characterized by the Student- $t$  distribution, such that

$$\mathbf{x}_0 \sim \text{St}(\mathbf{m}_{0|0}^x, \frac{\nu_x-2}{\nu_x}\mathbf{P}_{0|0}^x, \nu_x), \quad (54)$$

$$\mathbf{q}_k \sim \text{St}(\mathbf{0}, \frac{\nu_q-2}{\nu_q}\mathbf{Q}, \nu_q), \quad (55)$$

$$\mathbf{r}_k \sim \text{St}(\mathbf{0}, \frac{\nu_r-2}{\nu_r}\mathbf{R}, \nu_r) \quad (56)$$

where the initial state estimate was  $\mathbf{m}_{0|0}^x = [10\ 175 \text{ m} \ 295 \text{ ms}^{-1} \ 980 \text{ m} \ -35 \text{ ms}^{-1}]$ , the initial covariance  $\mathbf{P}_{0|0}^x = \mathbf{P}_0$  and DoF parameters were  $\nu_x = 1000$ ,  $\nu_q = 1000$  and  $\nu_r = 4.0$ . The kernel parameters for the BQ filters were set to  $\theta_f = [1 \ 100 \ 100 \ 100 \ 100]$  for the dynamics model and  $\theta_f = [0.05 \ 10 \ 100 \ 10 \ 100]$  for the measurement model.

The filter performance was evaluated by simulating 1,000 trajectories, each 100 time steps long, and computing the Monte Carlo averages of the performance scores. Figure 2 shows box-plots of the time-averaged RMSE scores. The left pane shows that the UKF and SF have more extreme outliers than the proposed TPQSF, while in the right pane we see that

the classical SF is better in terms of median RMSE. It is worth noting that because TPQ-based filters have a tunable DoF parameter, they were able to achieve improved median RMSE over the GPQ-based filter. From Figure 3, showing

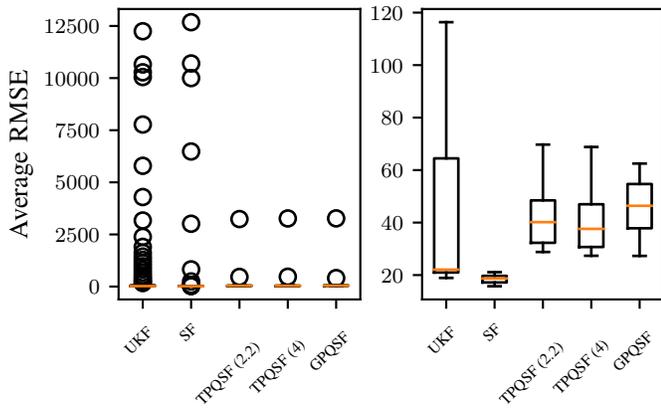


Figure 2: Overall filter RMSEs shown with outliers (left) and a detail without outliers (right). The proposed TPQ-based filters have less extreme outliers, whereas the median RMSE favours the classical quadrature-based SF.

the time-averaged INC score, we can deduce that the BQ filters provide more balanced estimates on average, whereas the classical filters are excessively optimistic in their estimates. This behaviour is in accordance with our expectations, because the BQ filters account for the additional functional uncertainty as described in Section III. Table II shows the mean of the

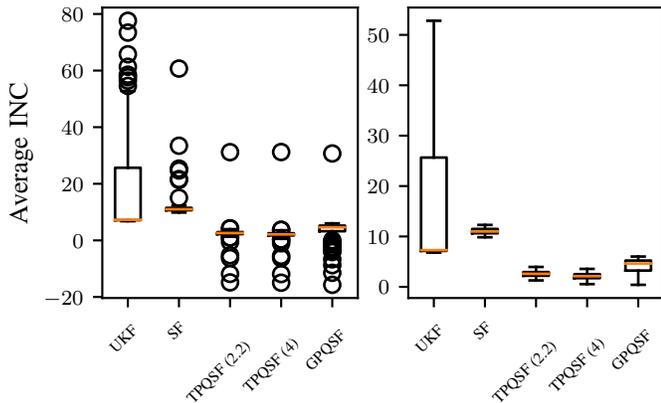


Figure 3: Overall filter INCs shown with (left) and without (right) outliers. The proposed TPQ-based filters display improved INC with most outliers in the pessimistic direction, whereas the UKF and SF are excessively optimistic.

overall average RMSE and INC along with their standard deviations, which were estimated by bootstrapping with 10,000 samples. Evidently, TPQSFs drastically improve the mean of the overall average RMSE and, as mentioned previously, provide much more balanced state estimates.

	RMSE	STD	INC	STD
UKF	803.99	231.62	18.37	1.80
SF	457.49	200.88	12.22	0.58
TPQSF( $\nu_g = 2.2$ )	77.29	32.17	2.39	0.38
TPQSF( $\nu_g = 4$ )	75.54	31.95	1.92	0.39
GPQSF	81.04	32.17	3.52	0.45

Table II: Overall RMSEs for the radar tracking example. The average RMSE favours the TPQ-based filters. Our proposed filters also give more balanced state estimates on average, as shown by the inclination indicator (INC) being closer to zero.

## VI. CONCLUSION

In this article, we proposed a Student- $t$  filter with a moment transform based on the Student- $t$  process quadrature. The MT is able to acknowledge the limited extent of knowledge of the integrated function when evaluated at finite number of evaluation points. The proposed TPQ moment transform was applied in sigma-point Student- $t$  filtering and its performance evaluated on two numerical examples. Overall, the results indicate that the proposed TPQSF is superior at self-assessing its own performance and consequently provides more balanced estimates.

## ACKNOWLEDGEMENT

Prüher and Straka were supported by the Czech Science Foundation, project no. GA 16-19999J and the Czech Ministry of Education, Youth and Sports, project no. LO1506. Tronarp and Särkkä were supported by Academy of Finland projects 266940, 273475 and 295080. Karvonen was supported by Aalto ELEC Doctoral School.

## REFERENCES

- [1] M. S. Grewal, L. R. Weill, and A. P. Andrews, *Global Positioning Systems, Inertial Navigation, and Integration*. Wiley, 2007.
- [2] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House, 1999.
- [3] M. Li, B. H. Kim, and A. I. Mourikis, “Real-time motion tracking on a cellphone using inertial sensing and a rolling-shutter camera,” in *2013 IEEE International Conference on Robotics and Automation*, May 2013, pp. 4712–4719.
- [4] R. Bhar, *Stochastic filtering with applications in finance*. World Scientific, 2010.
- [5] M. Roth, E. Özkan, and F. Gustafsson, “A Student’s  $t$  filter for heavy tailed process and measurement noise,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 5770–5774.
- [6] F. Tronarp, R. Hostettler, and S. Särkkä, “Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise,” in *2016 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1859–1866.
- [7] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” in *Proceedings of the IEEE*, vol. 92, no. 3, March 2004, pp. 401–422.

- [8] J. McNamee and F. Stenger, "Construction of fully symmetric numerical integration formulas," *Numerische Mathematik*, vol. 10, no. 4, pp. 327–344, 1967.
- [9] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, 2000.
- [10] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, 2006.
- [11] J. Prüher and O. Straka, "Gaussian process quadrature moment transform," 2017, preprint, arXiv:1701.01356.
- [12] X. R. Li and Z. Zhao, "Measuring Estimator's Credibility: Noncredibility Index," in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1–8.
- [13] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate filter for nonlinear systems with heavy tailed measurement noises," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4209–4213.
- [14] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [15] A. Shah, A. G. Wilson, and Z. Ghahramani, "Student- $t$  processes as alternatives to Gaussian processes," in *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 33, 2014, pp. 877–885.
- [16] A. Solin and S. Särkkä, "State space methods for efficient inference in Student- $t$  process regression," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2015)*, ser. JMLR Workshop and Conference Proceedings, vol. 38, 2015, pp. 885–893.
- [17] F.-X. Briol, C. J. Oates, M. Girolami, M. A. Osborne, and D. Sejdinovic, "Probabilistic integration: A role for statisticians in numerical analysis?" 2016, preprint, arXiv:1512.00933.
- [18] H. Zangl and G. Steiner, "Optimal Design of Multiparameter Multisensor Systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 7, pp. 1484–1491, July 2008.
- [19] I. M. Ross, R. J. Proulx, and M. Karpenko, "Unscented guidance," in *2015 American Control Conference*, July 2015, pp. 5605–5610.
- [20] S. Särkkä, *Bayesian Filtering and Smoothing*. New York: Cambridge University Press, 2013.
- [21] P. Diaconis, "Bayesian numerical analysis," in *Statistical decision theory and related topics IV*, S. S. Gupta and J. O. Berger, Eds. Springer-Verlag New York, 1988, vol. 1, pp. 163–175.
- [22] A. O'Hagan, "Some Bayesian numerical analysis," *Bayesian Statistics*, vol. 4, pp. 345–363, 1992.
- [23] P. Hennig, M. A. Osborne, and M. Girolami, "Probabilistic numerics and uncertainty in computations," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 471, no. 2179, 2015.
- [24] J. Cockayne, C. Oates, T. Sullivan, and M. Girolami, "Bayesian probabilistic numerical methods," 2017, preprint, arXiv:1702.03673.
- [25] A. O'Hagan, "Bayes–Hermite quadrature," *Journal of Statistical Planning and Inference*, vol. 29, no. 3, pp. 245–260, November 1991.
- [26] T. Minka, "Deriving quadrature rules from Gaussian processes," Statistics Department, Carnegie Mellon University, Tech. Rep., November 2000.
- [27] S. Särkkä, J. Hartikainen, L. Svensson, and F. Sandblom, "On the relation between Gaussian process quadratures and sigma-point methods," *Journal of Advances in Information Fusion*, vol. 11, no. 1, pp. 31–46, 2016.
- [28] J. Prüher and M. Šimandl, "Bayesian quadrature variance in sigma-point filtering," in *Informatics in Control, Automation and Robotics, 12th International Conference, ICINCO 2015 Colmar, Alsace, France, 21-23 July, 2015 Revised Selected Papers*, ser. Lecture Notes in Electrical Engineering, J. Filipe, O. Gusikhin, M. K., and J. Sasiadek, Eds. Springer International Publishing, 2016, vol. 370.
- [29] O. Barndorff-Nielsen, J. Kent, and M. Sørensen, "Normal variance-mean mixtures and  $z$  distributions," *International Statistical Review*, no. 50, pp. 145–159, 1982.
- [30] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113, 1993.
- [31] G. Kitagawa, "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.
- [32] L. Wasserman, *All of Nonparametric Statistics*. Springer, 2007.
- [33] I. Arasaratnam, S. Haykin, and R. Elliott, "Discrete-Time Nonlinear Filtering Algorithms Using Gauss–Hermite Quadrature," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.