BAYESIAN FILTERING FOR NONLINEAR STATE–SPACE MODELS IN SYMMETRIC α–STABLE MEASUREMENT NOISE

Jordi Vilà-Valls, Carles Fernández-Prades**, Pau Closas** and Juan A. Fernandez-Rubio**

*Universitat Politècnica de Catalunya (UPC) Jordi Girona 1-3, 08034, Barcelona, Spain.
** Centre Tecnològic de Telecom. de Catalunya (CTTC) Av. Friedrich Gauss 7, 08860, Castelldefels, Spain.
email: { jordi.vila-valls, juan.a.fernandez }@upc.edu, { carles.fernandez, pau.closas }@cttc.cat

ABSTRACT

Bayesian filtering appears in many signal processing problems, reason why it attracted the attention of many researchers to develop efficient algorithms, yet computationally affordable. In many real systems, it is appropriate to consider α–stable noise distributions to model possible outliers or impulsive behavior in the measurements. In this paper, we consider a nonlinear state–space model with Gaussian process noise and symmetric α–stable measurement noise. To obtain a robust estimation framework we consider that both process and measurement noise statistics are unknown. Using the product property of α–stable distributions we rewrite the measurement noise in a conditionally Gaussian form. Within this framework, we propose an hybrid sigma–point/Monte Carlo approach to solve the Bayesian filtering problem, what leads to a robust method against both outliers and a weak knowledge of the system.

1. INTRODUCTION

The problem under study concerns the derivation of efficient and robust methods to solve the recursive Bayesian filtering problem, which implies the on–line estimation of the time–varying unknown states of a system, using the incoming flow of information (observations) from the system, along with some prior statistical knowledge of the variation of such states. The solution to this problem does not exist in a general manner and we have to resort to suboptimal techniques. The Kalman filter (KF) provides the closed form solution to the optimal filtering problem in linear/Gaussian systems, assumptions that not always hold in real–life systems. The direct extension of the KF to nonlinear systems is the Extended Kalman filter (EKF), which uses a local linearization of the process and measurement nonlinear functions and applies the Kalman solution. This filter only gives acceptable performances in weak nonlinear systems and diverges easily.

A plethora of alternatives have been proposed in the last decade to solve the nonlinear estimation problem, among them, the family of Sigma–point (SP) filters [1, 2, 3] within the Gaussian framework, and the family of Sequential Monte Carlo (SMC) methods [4] for arbitrary noise distributions.

SMC methods provide a general framework to deal with nonlinear/non–Gaussian problems using a stochastic sampling approach to numerically approximate the integrals involved in the optimal solution. This broad suitability comes at expense of a high computational load, thus having limited practical use in time-constrained applications. SP methods use a deterministic sampling to approximate the integrals involved in the Bayesian filter solution. The key point is to assume that the measurement and process noise are independent random Gaussian processes. This leads to Gaussian state transition and measurement likelihood densities, which in turn reverts to a Gaussian posterior density [5]. Recently, new SP methods have been proposed [6, 3], numerically stable and easily extendable to high dimensional problems, mitigating the curse of dimensionality and divergence effects.

In many real–life systems, the additive Gaussian noise model does not hold and the noise statistics parameters are unknown. In these scenarios the methods based on the standard Kalman framework (KF, EKF and SP) give poor performances. α–stable distributions [7] have been shown to be appropriate in many signal processing applications [8], in particular, being useful to model the existence of outliers or impulsive behaviors on the measurement model [9].

When dealing with symmetric α–stable ($S\alpha S$) noise, the product property allows us to use a Scale Mixture of Normals (SMiN) representation of the α–stable probability density function (pdf) [10], which implies that a $S\alpha S$ distribution can be expressed in a conditionally Gaussian form. This idea has been applied to use Markov Chain Monte Carlo (MCMC) inference techniques [11] and to solve the Bayesian estimation problem for linear Time–Varying AutoRegressive (TVAR) models using a direct SMC solution [12].

In this contribution, we extend the results found in the literature to solve the robust Bayesian filtering problem for nonlinear multivariate state-space models where the measurement noise is α–stable distributed and the noise statistics parameters are unknown. The proposed method provides an alternative to the Kalman–Lévy filter [13] to deal with nonlinear systems and $S\alpha S$ noise, being robust against outliers while considering a weak knowledge about the system dynamics.

The paper is organized as follows: Section 2 introduces the α–stable distributions and the product property, being the key point to write the state–space model as conditionally Gaussian. Section 3 sets the state–space model and the estimation problem. In Sections 4 and 5 we derive the robust hybrid sigma–point/Monte Carlo method, and a target tracking application is used in Section 6 to evaluate the proposed algorithm.

2. α–STABLE DISTRIBUTIONS AND THE PRODUCT PROPERTY

In the literature, we find several equivalent definitions for α–stable distributions [7]. In all the cases, an α–stable distribution is characterized by four parameters: α refers to the index of stability, γ is the scale factor, and β and δ are the skewness and shift parameters. The characteristic function of an α–stable random variable is usually considered as a definition for stable distributions, because in a general case the closed form expression for the density function does not exist.

Definition 1. *An α-stable random variable (r.v.)* $X \sim$

This work has been partially supported by project 2009 SGR 1236 of the Catalan Administration (AGAUR), by the Spanish Science and Technology Commission: TEC 2010-19171 (MOSAIC), and TEC 2008-02685/TEC (NARRA) and by the European Commission in the framework of the COST Action IC0803 (RFCSET).

$S(\alpha, \gamma, \beta, \delta)$*, with parameters* $\alpha \in (0, 2]$*,* $\gamma \geq 0$*,* $\beta \in [-1, 1]$ *and* $\delta \in \mathbb{R}$*, has the following characteristic function*

$$\Phi_\alpha(u) = \begin{cases} exp\left(-\gamma^\alpha |u|^\alpha (1 - i\beta sign(u) \cdot \tan \frac{\pi\alpha}{2}) + i\delta u\right) \\ exp\left(-\gamma |u| (1 + i\beta \frac{2}{\pi} sign(u) \cdot \ln |u|) + i\delta u\right) \end{cases} \quad (1)$$

for $\alpha \neq 1$ *and* $\alpha = 1$*, respectively.*

Due to the non–existence of the variance of these distributions, all the techniques valid for the Gaussian case do not apply. An important aspect with α–stable distributions is that they generalize the central limit theorem (CLT), providing a solution for random variables with infinite variance. The index of stability or characteristic exponent α describes the tail of the distribution (smaller α, heavier tails), the skewness parameter β indicates if the distribution is right- or left-skewed, and the scale and shift parameters γ and δ are similar to the variance and the mean in a normal distribution, respectively, in the following sense:

$$\gamma Z + \delta \sim S(\alpha, \gamma, \beta, \delta), \quad (2)$$

where $Z \sim S(\alpha, 1, \beta, 0)$ is called standard α–stable random variable. We will use the shorthand notations $S(\alpha, 1, \beta, 0) = S(\alpha, \beta)$ and $S\alpha S = S(\alpha, \gamma, 0, \delta)$ for a symmetric α–stable distribution. A closed–form expression for the pdf of these distributions do not exist in general, but some exceptions are the Gaussian distribution $\mathcal{N}(\mu, \sigma^2) = S(2, \frac{\sigma}{\sqrt{2}}, 0, \mu)$, the Cauchy distribution $S(1, \gamma, 0, \delta)$, and the Lévy distribution $S(0.5, \gamma, 1, \delta)$ [7].

The *product property* of $S\alpha S$ distributions [7] states that every $S\alpha S$ r.v. can be expressed as the product of a $S\alpha S$ r.v. and a totally positive skewed α–stable r.v. A corollary of this property states that a $S\alpha S$ r.v. can always be represented as the product of a Gaussian r.v. and a positive totally skewed α–stable r.v., what is called a Scale Mixture of Normals (SMiN) representation [10].

Using this representation, every $S\alpha S$ r.v. can be expressed in a conditionally Gaussian form as follows:

$$\begin{aligned} z \sim S(\alpha, \gamma, 0, \delta), \;\; u \sim \mathcal{N}(0, 1) \text{ and } \lambda \sim S\left(\tfrac{\alpha}{2}, 1\right), \\ z = \delta + \gamma\sqrt{\lambda} u, \\ z|\lambda \sim \mathcal{N}(\delta, \gamma^2\lambda). \end{aligned} \quad (3)$$

This property is the key point to reformulate the nonlinear $S\alpha S$ measurement equation into a conditionally Gaussian form allowing the use of Gaussian techniques to solve the filtering problem.

3. STATE–SPACE MODEL

In this paper we are interested in nonlinear filtering problems where the process noise is Gaussian and the measurement noise is symmetric α–stable distributed, both being additive. The heavy–tailed measurement noise accounts for possible outliers or impulsive behavior in the observations, giving a more general and flexible framework than considering the standard Gaussian case [9].

The assumed discrete state-space model is expressed as

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{v}_k, \quad (4)$$
$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}^*_k, \quad (5)$$

where $k \in \mathbb{Z}$ refers to discret time instants, $\mathbf{x}_k \in \mathbb{R}^N$ and $\mathbf{y}_k \in \mathbb{R}^L$ are the states and the observations at time k, where the components $(y_{k,1}, ..., y_{k,L})$ are assumed to be independent. $\mathbf{f}$ and $\mathbf{h}$ are the process and measurement equations, known and nonlinear in a general case. $\mathbf{v} = \{\mathbf{v}_k, k \in \mathbb{Z}\}$ and $\mathbf{n}^* = \{\mathbf{n}^*_k, k \in \mathbb{Z}\}$ are the process (Gaussian) and observation ($S\alpha S$) noises, which are mutually independent with unknown statistics (*i.e.*, in real–life systems we do not have complete knowledge of the system dynamics)

$$\mathbf{v}_k \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{v,k}), \quad (6)$$
$$\mathbf{n}^*_k = \mathbf{D}_k \boldsymbol{\eta}_k, \quad (7)$$

where each element of $\boldsymbol{\eta}_k$ is standard $S\alpha S$ distributed:

$$\eta_{k,i} \sim S(\alpha_i, 0) \qquad \text{for } \; i = 1, \ldots, L,$$

and the scale factor diagonal matrix is defined as

$$\mathbf{D}_k = \text{diag}\left(\gamma_{n_k,1}, \ldots, \gamma_{n_k,L}\right).$$

Using the SMiN representation of an α–stable distribution, we can rewrite the measurement equation in a conditionally Gaussian form

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{n}_k, \quad (8)$$

where $\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{n,k})$ and

$$\boldsymbol{\Sigma}_{n,k} = \text{diag}\left(\gamma^2_{n_k,1}\lambda_{k,1}, \ldots, \gamma^2_{n_k,L}\lambda_{k,L}\right) \quad \lambda_{k,i} \sim S\left(\frac{\alpha_i}{2}, 1\right). \quad (9)$$

The robust Bayesian filtering problem within the conditionally Gaussian form concerns the recursive estimation of the states $\mathbf{x}_k$ and the unknown parameters of the system, namely the process noise covariance matrix $\boldsymbol{\Sigma}_{v,k}$ and for each element of $\mathbf{y}_k$ the triad $\boldsymbol{\phi}_{k,i} = (\gamma_{n_k,i}, \lambda_{k,i}, \alpha_i)^T$. We define the vector containing all the measurement noise parameters as $\boldsymbol{\phi}_k = \{\boldsymbol{\phi}_{k,i}, \text{ for } i \in [1, L]\}$ and we denote as $\boldsymbol{\theta}_k$ the overall parameter vector containing both process and measurement noise parameters.

4. HYBRID SQUARE–ROOT SIGMA–POINT/MONTE CARLO SOLUTION

The solution to the robust Bayesian filtering problem for the state–space model defined in (4) and (8) is given by the joint *a posteriori* distribution $p(\mathbf{x}_k, \boldsymbol{\theta}_k|\mathbf{y}_{1:k})$, which casts all the information about the states and the model contained in the observations. Its characterization allows us to obtain an optimal estimate respect any criterion, for example, the Mininum Mean Square Error (MMSE) or the Maximum a Posteriori (MAP) estimates. This pdf can be rewritten as

$$p(\mathbf{x}_k, \boldsymbol{\theta}_k|\mathbf{y}_{1:k}) = p(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{y}_{1:k}) p(\boldsymbol{\theta}_k|\mathbf{y}_{1:k}) \quad (10)$$

A direct application of a SMC method to obtain the joint estimation of the states and the parameters of the model is unviable because the dimensionality of the problem is too large and the method would collapse. To overcome this problem we profit of the underlying Gaussian structure. Under the knowledge of the noise statistics parameters $\boldsymbol{\Sigma}_{v,0:k}$ and $\boldsymbol{\phi}_{0:k}$, the state–space model is conditionally Gaussian and we can resort to standard Gaussian techniques to compute the *a posteriori* distribution $p(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{y}_{1:k})$. As the model is nonlinear we use a SP Kalman filtering method to find this pdf.

The SP Kalman filters are a family of powerful methods that use a deterministic sampling to approximate the integrals of the optimal Bayesian filter. The key idea is to compute the means and covariances used in the standard Kalman solution by propagating the sample set through the process and measurement nonlinear functions. A further refinement of SP schemes comes from the fact that, when we propagate the covariance matrix through a nonlinear function, the filter should preserve the properties of a covariance matrix, namely, its symmetry, and positive-definiteness. In practice, however, due to lack of arithmetic precision, numerical errors may lead to a loss of these properties. To circumvent this problem, a square-root filter is introduced to propagate the square-root of the covariance matrix instead of the covariance itself [6, 3].

We propose a square–root sigma–point Kalman filter using quadrature rules [6] to estimate the conditionally Gaussian filtering density $p(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{y}_{1:k})$. The method is sketched in Algorithm 1, where we use M sigma–points. We note that at time k the time update depends on $\boldsymbol{\Sigma}_{v,k}$ and the measurement update depends on $\boldsymbol{\phi}_k$. In the algorithm, $\mathbf{S} = \mathbf{Tria}(\boldsymbol{\Sigma})$ denotes a general triangularization algorithm (for instance, the QR decomposition), where $\mathbf{S}$ is a lower triangular matrix. The "/" operator refers to backward/forward substitution.

From the sigma–point solution we have that the *a posteriori* distribution and the likelihood density follow the normal distributions

$$p(\mathbf{x}_k|\boldsymbol{\theta}_{0:k}, \mathbf{y}_{1:k}) = \mathcal{N}\left(\hat{\mathbf{x}}_{k|k}, \boldsymbol{\Sigma}_{x,k|k}\right), \quad (11)$$

$$p(\mathbf{y}_k|\boldsymbol{\theta}_{0:k}, \mathbf{y}_{1:k-1}) = \mathcal{N}\left(\hat{\mathbf{y}}_{k|k-1}, \boldsymbol{\Sigma}_{y,k|k-1}\right), \quad (12)$$

where $\boldsymbol{\Sigma}_{y,i|j} = \mathbf{S}_{y,i|j}\mathbf{S}_{y,i|j}^T$. We note that the covariance matrix of the likelihood density is diagonal because the components of the observation, $y_{k,i}$, are independent, so we can write this pdf as a product of normal distributions (for simplicity, we drop the dependence of the likelihood on the process noise statistics parameters),

$$p(\mathbf{y}_k|\boldsymbol{\phi}_{0:k}, \mathbf{y}_{1:k-1}) = \prod_{i=1}^{L} p(y_{k,i}|\boldsymbol{\phi}_{0:k,i}, y_{1:k-1,i}), \quad (13)$$

where the likelihood density of $y_{k,i}$ is computed using the i^{th} element of the predicted measurement and the i^{th} element of the diagonal of the innovation covariance matrix

$$p(y_{k,i}|\boldsymbol{\phi}_{0:k,i}, y_{1:k-1,i}) = \mathcal{N}\left(\hat{\mathbf{y}}_{k|k-1,i}, \left[\boldsymbol{\Sigma}_{y,k|k-1}\right]_{ii}\right). \quad (14)$$

Although extended state–spaces can be defined by concatenating the original state vector $\mathbf{x}_k$ with a vector containing other unknown parameters of the state–space model (*i.e.*, parameters of $\mathbf{f}(\cdot)$ and $\mathbf{h}(\cdot)$), the elements of the process noise covariance matrix $\boldsymbol{\Sigma}_{v,k}$ cannot be jointly estimated along $\mathbf{x}_k$ by this approach, since sigma–point methods neglect the nonlinear dependency between $\mathbf{x}_k$ and the elements of $\boldsymbol{\Sigma}_{v,k}$. In other words, time update of Algorithm 1 is computed using only the first two moments of $\hat{\mathbf{x}}_{k|k-1}$, but it is easily shown that the third-order cross-conditional moment is non-zero [14] (*i.e.*, $\mathbf{x}$ and $\boldsymbol{\theta}$ are not jointly Gaussian), thus making $\boldsymbol{\Sigma}_{v,k}$ not identifiable.

Algorithm 1 Hybrid Square–root Sigma–point/Monte Carlo method for nonlinear Bayesian filtering in $S\alpha S$ noise

Require: $\mathbf{y}_{1:k}$, $\hat{\mathbf{x}}_0$, $\boldsymbol{\Sigma}_{x,0} = \mathbf{S}_{x,0|0}\mathbf{S}_{x,0|0}^T$, $\boldsymbol{\Sigma}_{v,1}$, $\boldsymbol{\Sigma}_{n,1}$.

Define M sigma–points and weights $\{\boldsymbol{\xi}_i, \omega_i\}_{i=1,\dots,M}$ by using quadrature or cubature rules [2, 3].
Set $\mathbf{W} = \mathrm{diag}(\sqrt{\omega_i})$
for $k = 1$ to ∞ **do**

Time update

Evaluate the sigma points:
$\mathbf{x}_{i,k-1|k-1} = \mathbf{S}_{x,k-1|k-1}\boldsymbol{\xi}_i + \hat{\mathbf{x}}_{k-1|k-1},\ i = 1, ..., M.$
Evaluate the propagated sigma points:
$\tilde{\mathbf{x}}_{i,k|k-1} = \mathbf{f}(\mathbf{x}_{i,k-1|k-1}).$
Estimate the predicted state:
$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=1}^{M} \omega_i \tilde{\mathbf{x}}_{i,k|k-1}.$
Estimate the square–root factor of the predicted error covariance:
$\mathbf{S}_{x,k|k-1} = \mathbf{Tria}\left(\left[\tilde{\mathcal{X}}_{k|k-1} \,\vdots\, \mathbf{S}_{\boldsymbol{\Sigma}_{v,k}}\right]\right)$, where:
$\mathbf{S}_{\boldsymbol{\Sigma}_{v,k}}$ is a square–root factor of $\boldsymbol{\Sigma}_{v,k}$ such that $\boldsymbol{\Sigma}_{v,k} = \mathbf{S}_{\boldsymbol{\Sigma}_{v,k}}\mathbf{S}_{\boldsymbol{\Sigma}_{v,k}}^T$, and
$\tilde{\mathcal{X}}_{k|k-1} = \left[\tilde{\mathbf{x}}_{1,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \ \cdots \ \tilde{\mathbf{x}}_{L,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right]\mathbf{W}.$

Measurement update

Evaluate the sigma points:
$\mathbf{x}_{i,k|k-1} = \mathbf{S}_{x,k|k-1}\boldsymbol{\xi}_i + \hat{\mathbf{x}}_{k|k-1},\ i = 1, ..., M.$
Evaluate the propagated sigma points:
$\tilde{\mathbf{y}}_{i,k|k-1} = \mathbf{h}(\mathbf{x}_{i,k|k-1}).$
Estimate the predicted measurement:
$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=1}^{M} \omega_i \tilde{\mathbf{y}}_{i,k|k-1}.$
Obtain $\mathbf{S}_{\Sigma_{n,k}}$ by Algorithm 3
Estimate the square–root of the innovation covariance matrix:
$\mathbf{S}_{y,k|k-1} = \mathbf{Tria}\left(\left[\mathcal{Y}_{k|k-1} \,\vdots\, \mathbf{S}_{\boldsymbol{\Sigma}_{n,k}}\right]\right)$, where:
$\mathbf{S}_{\boldsymbol{\Sigma}_{n,k}}$ denotes a square–root factor of $\boldsymbol{\Sigma}_{n,k}$ such that $\boldsymbol{\Sigma}_{n,k} = \mathbf{S}_{\boldsymbol{\Sigma}_{n,k}}\mathbf{S}_{\boldsymbol{\Sigma}_{n,k}}^T$, and
$\mathcal{Y}_{k|k-1} = \left[\tilde{\mathbf{y}}_{1,k|k-1} - \hat{\mathbf{y}}_{k|k-1} \ \cdots \ \tilde{\mathbf{y}}_{L,k|k-1} - \hat{\mathbf{y}}_{k|k-1}\right]\mathbf{W}.$
Estimate the cross–covariance matrix
$\boldsymbol{\Sigma}_{xy,k|k-1} = \mathcal{X}_{k|k-1}\mathcal{Y}_{k|k-1}^T$, where:
$\mathcal{X}_{k|k-1} = \left[\mathbf{x}_{1,k|k-1} - \hat{\mathbf{x}}_{k|k-1} \ \cdots \ \mathbf{x}_{L,k|k-1} - \hat{\mathbf{x}}_{k|k-1}\right]\mathbf{W}.$
Estimate the Kalman gain
$\mathbf{K}_k = \left(\boldsymbol{\Sigma}_{xy,k|k-1}/\mathbf{S}_{y,k|k-1}^T\right)/\mathbf{S}_{y,k|k-1}.$
Estimate the updated state
$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}\right).$
Estimate the square–root factor of the corresponding error covariance:
$\mathbf{S}_{x,k|k} = \mathbf{Tria}\left(\left[\mathcal{X}_{k|k-1} - \mathbf{K}_k\mathcal{Y}_{k|k-1} \,\vdots\, \mathbf{K}_k\mathbf{S}_{\boldsymbol{\Sigma}_{n,k}}\right]\right).$
Obtain $\mathbf{S}_{\Sigma_{v,k+1}}$ by Algorithm 2
end for

5. NOISE STATISTICS ESTIMATION

Standard estimation techniques such as Kalman-type filters require a complete knowledge of the process and measurement noise statistics. In real–life systems the full knowledge of the system dynamics is questionable and we have to derive estimation methods to circumvent this problem and to obtain a solution that is robust to such uncertainties.

5.1 Process noise statistics estimation

Noise covariance matrix estimation methods can be divided into several categories: correlation methods, covariance matching methods or subspace and minimax methods, to name a few [14]. In the model proposed in Section 3, the process noise is supposed to be an additive Gaussian zero–mean noise (see eq.(4)) with unknown covariance matrix. The measurement noise statistics parameters are estimated using a Monte Carlo method (Section 5.2) so the problem under concern in this section is the estimation of the process noise covariance matrix.

In this paper we use a covariance matching type method which is an extension to our nonlinear state–space model of the method first presented in [15]. An intuitive approximation of the noise term $\mathbf{v}_k$ is $\mathbf{q}_k = \hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}$. We want to formulate an estimate of $\boldsymbol{\Sigma}_v$ in terms of the sample covariance of $\hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}$. If we have at time K the set of i.i.d. samples $\{\mathbf{q}_i\}_{i=1:K}$ we can estimate the covariance of the sample set as

$$\hat{\mathbf{C}}_q = \frac{1}{K-1}\sum_{i=1}^{K}(\mathbf{q}_i - \hat{\mathbf{q}})(\mathbf{q}_i - \hat{\mathbf{q}})^T, \quad (15)$$

with $\hat{\mathbf{q}} = 1/K \sum_{i=1}^{K} \mathbf{q}_i$ the sample mean. The expected value of $\hat{\mathbf{C}}_q$ gives an indication of what is being estimated. In the linear case this expectation is given by [15]

$$\mathbb{E}[\hat{\mathbf{C}}_q]^L = \frac{1}{K}\sum_{j=1}^{K}\left(\mathbf{F}_j \mathbf{\Sigma}_{x,j-1|j-1}\mathbf{F}_j^T - \mathbf{\Sigma}_{x,j|j} + \mathbf{\Sigma}_v\right), \quad (16)$$

with $\mathbf{F}_j$ the state transition matrix at time j. The equivalent to our nonlinear problem can be written from the sigma-point solution (SP sample mean and covariance) as

$$\mathbb{E}[\hat{\mathbf{C}}_q]^{NL} = \frac{1}{K}\sum_{j=1}^{K}(\boldsymbol{\beta}_j + \mathbf{\Sigma}_v), \quad (17)$$

with $\boldsymbol{\beta}_j = \sum_{i=1}^{M} \omega_i \tilde{\mathbf{x}}_{i,j|j-1}\tilde{\mathbf{x}}_{i,j|j-1}^T - \hat{\mathbf{x}}_{j|j-1}\hat{\mathbf{x}}_{j|j-1}^T - \mathbf{\Sigma}_{x,j|j}$.

The unbiased estimator of $\mathbf{\Sigma}_v$ at time k within the sigma–point formulation is obtained from $\hat{\mathbf{C}}_q = \mathbb{E}[\hat{\mathbf{C}}_q]$ (using all the samples available) as

$$\hat{\mathbf{\Sigma}}_v = \frac{1}{k-1}\sum_{j=1}^{k}(\mathbf{q}_j - \hat{\mathbf{q}})(\mathbf{q}_j - \hat{\mathbf{q}})^T - \frac{1}{k}\sum_{j=1}^{k}\boldsymbol{\beta}_j. \quad (18)$$

If the process covariance matrix is time–varying we can use a finite length (L_q) sample set ($\{\mathbf{q}_i\}_{i=k-L_q+1:k}$) to estimate the covariance matrix, instead of the full sample set. The sequential estimation method for the process noise covariance matrix using the covariance matching type solution with L_q samples is sketched in Algorithm 2, where we note $(\mathbf{a_1} - \mathbf{a_2})^2 = (\mathbf{a_1} - \mathbf{a_2})(\mathbf{a_1} - \mathbf{a_2})^T$. As indicated in [15], for numerical reasons, the diagonal elements of the process covariance matrix are reset to the absolute value of their estimates.

Algorithm 2 Process noise covariance estimation method

1: Set $\mathbf{q}_k = \hat{\mathbf{x}}_{k|k} - \hat{\mathbf{x}}_{k|k-1}$.
2: Estimate the mean as $\hat{\mathbf{q}}_k = \hat{\mathbf{q}}_{k-1} + 1/L_q(\mathbf{q}_k - \mathbf{q}_{k-L_q})$.
3: Estimate the covariance matrix as

$$\begin{aligned}\hat{\mathbf{\Sigma}}_{v,k+1} = \quad & \hat{\mathbf{\Sigma}}_{v,k} + \tfrac{1}{L_q-1}\left\{(\mathbf{q}_k - \hat{\mathbf{q}}_k)^2 - (\mathbf{q}_{k-L_q} - \hat{\mathbf{q}}_k)^2\right.\\ & \left. + \tfrac{1}{L_q}(\mathbf{q}_k - \mathbf{q}_{k-L_q})^2 + \tfrac{L_q-1}{L_q}[\boldsymbol{\beta}_{k-L_q} - \boldsymbol{\beta}_k]\right\}\end{aligned}$$

5.2 Measurement noise statistics estimation

In this section we propose a method to obtain the unknown measurement noise parameters of a $S\alpha S$ distribution, which have to be estimated for a proper behavior of the algorithm. As the components $y_{k,i}$ are supposed to be independent, we can estimate the triad $\boldsymbol{\phi}_{k,i} = (\gamma_{n_k,i}, \lambda_{k,i}, \alpha_i)^T$ independently for each component. So we can use L parallel filters to estimate the triad instead of using a method to estimate the $3 \times L$ noise parameters at once, what might be computationally unaffordable.

We first define the evolution of the parameters to be estimated: the index of stability α_i and the scale parameter γ_i are static parameters, and $\lambda_{k,i}$ is a totally positive skewed α–stable random variable with $p(\lambda_{k,i}) \sim S(\alpha_i/2, 1)$. As we stated before, from the SP solution we are able to compute the likelihood $p(y_{k,i}|\boldsymbol{\phi}_{0:k}, y_{1:k-1,i})$. Using this density and the parameter evolution distributions we are able to construct a SMC solution to estimate the triad $\boldsymbol{\phi}_{k,i}$ for each element of the observation, $y_{k,i}$.

It is well known that estimating static parameters can induce a severe degeneracy of the particles involved in the SMC method. We use the solution given in [4, ch.10], where authors consider a refined artificial parameter evolution for the static parameter which avoids the underlying loss of information. Within this procedure the static parameters are supposed to be time-varying with transition densities

$$p(\alpha_{k,i}|\alpha_{k-1,i}) \sim \mathcal{N}\left(a\alpha_{k-1,i} + (1-a)\bar{\alpha}_{k-1,i}, h^2 V_{k-1,i}^{\alpha}\right) \quad (19)$$

$$p(\gamma_{n_k,i}|\gamma_{n_{k-1},i}) \sim \mathcal{N}\left(a\gamma_{n_{k-1},i} + (1-a)\bar{\gamma}_{k-1,i}, h^2 V_{k-1,i}^{\gamma}\right) \quad (20)$$

where the parameters are $a = (3b-1)/(2b)$, $h^2 = 1 - a^2$, and b is a discount factor typically around $0.95 - 0.99$. $\bar{\alpha}_{k-1,i}$, $V_{k-1,i}^{\alpha}$, $\bar{\gamma}_{k-1,i}$ and $V_{k-1,i}^{\gamma}$ are Monte Carlo sample means and variances of the Monte Carlo approximation to $p(\alpha_i|y_{1:k-1,i})$ and $p(\gamma_{n,i}|y_{1:k-1,i})$, respectively.

We choose as importance distribution the prior transition density $p(\boldsymbol{\phi}_{k,i}|\boldsymbol{\phi}_{k-1,i}) = p(\alpha_{k,i}|\alpha_{k-1,i})p(\gamma_{n_k,i}|\gamma_{n_{k-1},i})p(\lambda_{k,i})$, which implies that the importance weights are proportional to the likelihood function $p(y_{k,i}|\boldsymbol{\phi}_{0:k}, y_{1:k-1,i})$. The method is sketched in Algorithm 3, where N_p is the number of particles.

Algorithm 3 Measurement noise parameter estimation

1: **for** $i = 1$ to L **do**
2: Draw N_p samples from the importance distribution $\boldsymbol{\phi}_{k,i}^{(j)} \sim p(\boldsymbol{\phi}_{k,i}|\boldsymbol{\phi}_{k-1,i}^{(j)})$ for $j = 1, ..., N_p$
3: Compute and normalize the importance weights $w_k^{(j)} = \tilde{w}_{k-1}^{(j)} p(y_{k,i}|\boldsymbol{\phi}_{0:k}^{(j)}, y_{1:k-1,i})$ for $j = 1, ..., N_p$ $\tilde{w}_k^{(j)} = \frac{w_k^{(j)}}{\sum_{j=1}^{N_p} w_k^{(j)}}$
4: If necessary resample and set weigths to $1/N_p$
5: Estimate the noise parameter vector $\hat{\boldsymbol{\phi}}_{k,i} = \sum_{j=1}^{N_p} w_k^{(j)} \boldsymbol{\phi}_{k,i}^{(j)}$
6: **end for**

6. COMPUTER SIMULATIONS

In this section, in order to provide illustrative numerical results, we show how the proposed method performs in a radar target tracking example where the non–Gaussian measurement noise applies [8, 9]. We follow the setup used in [3] but assuming that the measurement noise is $S\alpha S$.

In this application, a target was moving in a 2-D plane and was tracked by a radar whose measurements were range and azimuth, $\mathbf{y}_k = [r_k, \psi_k]^T$. The states to be tracked were position, velocity and acceleration of the target. These were respectively gathered in vector $\mathbf{x}_k = [p_{x,k}, p_{y,k}, v_{x,k}, v_{y,k}, a_{x,k}, a_{y,k}]^T$. Both the trajectory and measurements were modeled as

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{I} & T \times \mathbf{I} & T^2/2 \times \mathbf{I} \\ \mathbf{0} & \mathbf{I} & T \times \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{x}_{k-1} + \mathbf{v}_k \quad (21)$$

$$\mathbf{y}_k = \begin{pmatrix} \sqrt{p_{x,k}^2 + p_{y,k}^2} \\ \arctan(p_{y,k}/p_{x,k}) \end{pmatrix} + \mathbf{n}_k \quad (22)$$

where T is the time–interval between measurements, set to 1 second. The Gaussian process noise was modeled as $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_v)$ and $\mathbf{\Sigma}_v = \mathrm{diag}(4, 4, 4, 4, 0.01, 0.01)$. Each component of the measurement noise was $S\alpha S$ distributed, $n_{k,i} \sim S(\alpha_i, \gamma_i, 0, 0)$, with $\alpha_1 = 1.4$, $\alpha_2 = 1.8$, $\gamma_1 = 7$ and $\gamma_2 = 0.0022$. These parameters imply that we might have strong outliers in the range measurements and weak outliers in the azimuth measurements. In a Gaussian scenario ($\alpha = 2$

and $\sigma^2 = 2\gamma^2$) the measurement noise covariance matrix is $\mathbf{R} = \text{diag}(98, 10^{-6})$.

The initial state estimate was drawn from $\mathcal{N}(\hat{\boldsymbol{x}}_0, \boldsymbol{\Sigma}_{x,0})$ for each Monte Carlo trial, with $\hat{\boldsymbol{x}}_0 = [2000, 2000, 20, 20, 0, 0]$ and $\boldsymbol{\Sigma}_{x,k} = \text{diag}(500, 500, 200, 200, 40, 40)$. We used 3 sigma-points/dimension so the method required $M = 3^6 = 729$ points. The unknown process covariance matrix was initialized to $100 \cdot \boldsymbol{\Sigma}_v$. We made 10000 independent Monte Carlo runs with 100 scans per run and we used the root-mean square error (RMSE) as the measure of performance. As the process noise covariance matrix is constant we use all the samples available ($L_q = k$). Concerning the SMC method initialization for measurement noise parameters estimation, the initial particles were drawn from $\mathcal{U}(1, 2)$ for α_i, from $\mathcal{U}(0, 5\gamma_i)$ for the estimation of $\gamma_{n_k,i}$ and from $S(\bar{\alpha}_{p,i}, 1, 1, 0)$ for the estimation of $\lambda_{k,i}$, with $\bar{\alpha}_{p,i}$ the mean of the initial particles for α_i. We used $N_p = 300$ particles.

In figure 1, we plot the RMSE of the estimation of the position obtained with the proposed method together with the Bayesian Cramér-Rao bound (BCRB) [16] associated with the Gaussian equivalent problem (without outliers). The BCRB is used as a benchmark to assess the ultimate performance achievable in the Gaussian case with known process and measurement noise statistics. As a reference we also plot the results obtained in the standard Gaussian case with known statistics using a Square-root Quadrature Kalman Filter (SQKF) and the α-stable case with known noise statistics.

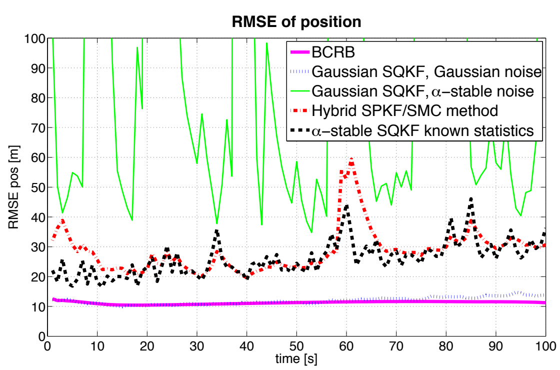

Figure 1: Postion RMSE for the hybrid SPKF/SMC method and the SQKF with known statistics for both Gaussian and α-stable cases. The Gaussian BCRB is plotted as a benchmark.

We can see that the performances obtained with the SQKF (dotted blue line) in the Gaussian case are optimal because they coincide with the BCRB (solid magenta line). When we introduced outliers (α-stable framework) the performances obtained with the SQKF still considering Gaussian noise were really poor (solid thin green line).

Considering a full knowledge of the statistics of the measurement noise (α-stable case with known statistics) we obtained reasonable good performances compared to the optimal case (dashed black line). The results obtained in this case are our reference for the hybrid SPKF/SMC ultimate achievable performance.

The results obtained with the proposed method (dot-dashed red line) are really encouraging. We can see that the proposed filter is able to deal with the impulsive behavior of the measurement noise and the unknown noise statistics in scenarios where other Gaussian filters fail, even with known statistic noise parameters, with a limited perfomance degradation.

7. CONCLUSIONS

This paper presented a solution to the robust Bayesian filtering problem for nonlinear state-space models with α–stable measurement noise, what has been proven to be a more appropriate representation of the measurement than the standard Gaussian case in many real–life systems. The method was validated by computer simulation in a target tracking application. We saw that the proposed method attains good performance results dealing correctly with outliers/impulsive behaviors in the measurement and unknown process and measurement noise statistics, while being computationally affordable when compared to standard methods.

REFERENCES

[1] R. V. der Merwe, Sigma-point Kalman filters for probabilistic inference in dynamic state-space models, Ph.D. thesis, OHSU, Portland, Oregon, USA (April 2004).

[2] I. Arasaratnam, S. Haykin, R. J. Elliot, Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature, Proc. of the IEEE 95 (5) (2007) 953–977.

[3] I. Arasaratnam, S. Haykin, Cubature Kalman filters, IEEE Trans. Automatic Control 54 (6) (2009) 1254–1269.

[4] A. Doucet, N. D. Freitas, N. Gordon (Eds.), Sequential Monte Carlo methods in practice, Springler, 2001.

[5] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, IEEE Trans. on Automatic Control 45 (5) (2000) 910–927.

[6] I. Arasaratnam, S. Haykin, Square-root quadrature Kalman filtering, IEEE Trans. Signal Processing 56 (6) (2008) 2589–2593.

[7] G. Samorodnitsky, M. S. Taqqu, Stable Non–Gaussian Random Processes: Stochastic models with infinite variance, Chapman and Hall, 1994.

[8] C. L. Nikias, M. Shao, Signal processing with alpha–stable distributions and applications, John Wiley and Sons, 1995.

[9] D. Middleton, Non–Gaussian noise models in signal processing for telecommunications: new methods and results for Class A and Class B noise models, IEEE Trans. on Information Theory 45 (4) (1999) 1129–1149.

[10] E. E. Kuruoglu, C. Molina, W. J. Fitzgerald, Approximation of α–stable probability densities using finite mixture of Gaussians, in: Proc. of the 9^{th} EUSIPCO, Typorama, Rhodes, 1998, pp. 989 – 992.

[11] S. J. Godsill, Inference in symmetric alpha-stable noise using MCMC and the slice sampler, in: Proc. IEEE ICASSP, Istanbul, Turkey, 2000, pp. 3806 – 3809.

[12] M. J. Lombardi, S. J. Godsill, On–line Bayesian estimation of signals in symmetric alpha-stable noise, IEEE Trans. on Signal Processing 54 (2) (2006) 775–779.

[13] D. Sornette, K. Ide, The Kalman-Lévy filter, Elsevier Physica D 151 (2001) 142–174.

[14] J. Duník, M. Simandl, O. Straka, Methods for estimating state and measurement noise covariance matrices: aspects and comparison, in: Proc. IFAC SYSID 2009, Saint-Malo, France, 2009, pp. 372 – 377.

[15] K. A. Myers, B. D. Tapley, Adaptive sequential estimation with unknown statistics, IEEE Trans. on Automatic Control 21 (8) (1976) 520–523.

[16] P. Tichavský, C. H. Muravchik, A. Nehorai, Posterior Cramér-Rao bounds for discrete-time nonlinear filtering, IEEE Trans. Signal Processing 46 (5) (1998) 1386–1396.