2016 European Control Conference (ECC)
June 29 - July 1, 2016. Aalborg, Denmark

State Estimation for Linear Systems with Additive Cauchy Noises: Optimal and Suboptimal Approaches

Robert Fonod[1], Moshe Idan[1] and Jason L. Speyer[2]

Abstract— **Only few estimation methods can converge in the presence of impulsive measurement and/or process noises without the use of augmented heuristic schemes. To understand the performance of these schemes, the optimal Idan/Speyer Cauchy Estimator (ISCE) is compared with the performance of the particle filter (PF) and Gaussian sum filter (GSF) as the convergence time of these estimators is allowed to increase. That is, the number of particles at each step for the PF and the number of Gaussian components at each step for the GSF are increased and their performance relative to the ISCE is numerically studied for scalar and two-state dynamic systems.**

I. Introduction

Impulsive measurement and process noises in stochastic state estimators have typically been handled by heuristic schemes that augment the estimation process. Recently, an analytical recursive nonlinear estimation scheme for multivariate linear systems with additive Cauchy distributed measurement and process noises has been developed. Cauchy uncertainties are impulsive and this filter is optimal in that it generates, in closed-form, the character function (CF) of the unnormalized conditional probability function of the state given the measurement history [1], [2]. From this character function, the conditional mean and conditional error variance in the presence of Cauchy distributed noise is obtained. However, there are several general estimation algorithms that may also be able to deal with impulsive noises [3]–[7]. Although those approaches are suboptimal, they may offer reasonable approximations also for the heavy tailed, Cauchy distributed noise environment.

Two of the most popular approximations are the particle filter (PF) and the Gaussian sum filter (GSF) in that they are shown to converge to the correct conditional density of the state as the number of terms increase. Therefore, they are implemented with some degree of approximation, producing a tradeoff between numerical efficiency and the estimation performance in constructing the conditional probability density function of the state given the measurement history and the resulting conditional mean and variance. Our objective is to compare the performance of these approximate filters to that of the optimal Cauchy filter. Performance is measured by restricting the filter approximation to an average computation time interval at each measurement update.

This work was supported by the United States-Israel Binational Science Foundation, Grant 2012122.

[1]Robert Fonod and Moshe Idan are with Department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa, 3200003, Israel. Email: {robert.fonod; moshe.idan}@technion.ac.il

[2]Jason L. Speyer is with Department of Mechanical and Aerospace Engineering, University of California, Los Angeles (UCLA), Los Angeles, 90095, USA. Email: speyer@g.ucla.edu

Although the closed-form analytical solution of the Cauchy filter provides the exact minimum variance estimates of the systems states given a measurement sequence, their computational complexity and memory burden becomes very high requiring an approximation in their implementation. A sliding window of data is used to limit the computation of the Cauchy filter. It is shown numerically that there is little differences in the conditional mean and variance for a window of six, eight, or even ten. Clearly, larger window size implies higher computational burden and memory requirements. The performance of the PF and GSF are tested and compared with the Cauchy estimator with a window of six and then eight.

II. Problem Formulation

Consider a discrete-time, single-input-single-output, multivariate, and time-invariant linear system described by

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Phi}\boldsymbol{x}_k + \boldsymbol{\Gamma} w_k, \tag{1}$$

$$z_k = \boldsymbol{H}\boldsymbol{x}_k + v_k, \tag{2}$$

with state vector $\boldsymbol{x}_k \in \mathbb{R}^n$, scalar measurement z_k, and known matrices $\boldsymbol{\Phi} \in \mathbb{R}^{n\times n}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times 1}$, and $\boldsymbol{H} \in \mathbb{R}^{1\times n}$. The noise inputs w_k and v_k are independent Cauchy distributed random variables with zero median and scaling parameters $\beta > 0$ and $\gamma > 0$, respectively. Their probability density functions (PDFs) and their characteristic functions are denoted p and ϕ, respectively, and are assumed to be time independent and given by

$$p_W(w_k) = \tfrac{\beta/\pi}{w_k^2+\beta^2} \quad \Rightarrow \quad \phi_W(\bar{\nu}) = e^{-\beta|\bar{\nu}|}, \tag{3}$$

$$p_V(v_k) = \tfrac{\gamma/\pi}{v_k^2+\gamma^2} \quad \Rightarrow \quad \phi_V(\bar{\nu}) = e^{-\gamma|\bar{\nu}|}, \tag{4}$$

where $\bar{\nu}$ is a scalar spectral variable.

The initial conditions at $k = 1$ are also assumed to be independent Cauchy distributed random variables. Specifically, each i-th element x_{1i} of the initial state vector $\boldsymbol{x}_1$ has a Cauchy PDF with a given median $\bar{x}_{1i}$ and a scaling parameter $\alpha_i > 0$, $i = 1, \ldots, n$. The joint PDF of the initial conditions and its characteristic function are given by

$$\begin{aligned} p_{X_1}(\boldsymbol{x}_1) &= \prod_{i=1}^{n} \frac{\alpha_i/\pi}{(x_{1i}-\bar{x}_{1i})^2+\alpha_i^2} \;\Rightarrow \\ \phi_{X_1}(\boldsymbol{\nu}) &= \prod_{i=1}^{n} e^{-\alpha_i|\nu_i|+j\bar{x}_{1i}\nu_i}, \end{aligned} \tag{5}$$

where ν_i is an element of the spectral variable $\boldsymbol{\nu} \in \mathbb{R}^n$.

The measurement history used in the estimation problem formulation is defined as $\boldsymbol{z}_{1:k} = \{z_1, \ldots, z_k\}$. The objective is to compute the minimum variance estimate of $\boldsymbol{x}_k$ given the measurement history of $\boldsymbol{z}_{1:k}$.

III. Optimal Solution - Cauchy Filter

In this section, a brief overview of the minimum variance Cauchy estimators, also known as Idan/Speyer Cauchy Estimator (ISCE) [8], for systems described by (1) and (2) is given, summarizing the main results of [1], [2], [9].

A. Scalar ISCE - PDF Approach

The original work derived the ISCE for single-state system using the PDF approach [9]. It was shown that under mild conditions on the system parameters, the PDF of the state at time step k given $\boldsymbol{z}_{1:k}$ can be expressed as

$$p(x_k|\boldsymbol{z}_{1:k}) = \sum_{i=1}^{k+2} \frac{a^i_{k|k} x_k + b^i_{k|k}}{(x_k - \sigma_{k|k})^2 + (\omega^i_{k|k})^2}. \tag{6}$$

Initialization and update rules for the series coefficients $a^i_{k|k}$, $b^i_{k|k}$, $\sigma^i_{k|k}$ and $\omega^i_{k|k}$ are to be found in [9]. It is apparent that all of the series coefficients must be updated at each time step, and that each measurement update increases the number of terms in the series. To avoid such impractical growth, [9] describes a truncation procedure that limits the number of terms in the series to a specified buffer length.

The approach above provides a closed-form expression for $p(x_k|\boldsymbol{z}_{1:k})$, which can be examined for its shape, as will be shown in this paper. Unfortunately, this approach was found to be insufficient when addressing multivariate systems due to its derivation specifics (partial fraction expansion of single-variable rational functions). As an alternative, the successful approach to address multivariate system utilizes the CF of the PDF of interest. The ISCE for single-state systems was re-derived using this approach in [1].

B. Multivariate ISCE - CF Approach

In this approach, instead of propagating $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$, its un-normalized characteristic function given by

$$\bar{\phi}_{\boldsymbol{x}_k|\boldsymbol{z}_{1:k}}(\boldsymbol{\nu}) = \int p(\boldsymbol{x}_k, \boldsymbol{z}_{1:k}) e^{j\boldsymbol{\nu}^T \boldsymbol{x}_k} d\boldsymbol{x}_k, \tag{7}$$

is propagated, while the normalization factor can be obtained by $p(\boldsymbol{z}_{1:k}) = \bar{\phi}_{\boldsymbol{x}_k|\boldsymbol{z}_{1:k}}(\boldsymbol{0})$. In [2] it was shown that (7) can be expressed as a growing sum of terms in the form

$$\bar{\phi}_{\boldsymbol{x}_k|\boldsymbol{z}_{1:k}}(\boldsymbol{\nu}) = \sum_{i=1}^{n_t(k|k)} g_i^{k|k}\left(y_{gi}^{k|k}(\boldsymbol{\nu})\right) \exp\left(y_{ei}^{k|k}(\boldsymbol{\nu})\right), \tag{8}$$

i.e., a sum of exponential terms multiplied by a coefficient function $g(\cdot)$. The argument of this coefficient function, $y_{gi}^{k|k}(\boldsymbol{\nu})$, and of the exponents, $y_{ei}^{k|k}(\boldsymbol{\nu})$, are complicated functions of the spectral vector $\boldsymbol{\nu}$. The details of the various parameters and functions of the above expression can be found in [2]. Those functions and parameters can also be used to determine the state estimate and estimation error covariance matrix in a closed-form.

The main difficulty with the above results is that the number of terms $n_t(k|k)$ grows rapidly with time. A windowing technique that allows reducing the number of terms using an approximation was recently presented in [10] for the two-state case. This version of the ISCE will be used in comparing the performance of the various filters in this study.

IV. Suboptimal Solution - Particle Filter

Particle filter, also known as sequential Monte Carlo (MC) method, is a set of algorithms implementing recursive Bayesian estimation, based on point mass representation of probability densities. For good surveys, see [5], [6], [11].

A. Sequential Importance Sampling

Most PFs are based on an algorithm known as sequential importance sampling (SIS), which is a MC technique for solving the Bayesian inference problem [6]. The main idea is to represent the required posterior $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ using a set of random samples with associated weights. Let $\{\boldsymbol{x}^i_k, \mu^i_k\}_{i=1}^{n_p}$ be such an approximation, where $\{\boldsymbol{x}^i_k\}_{i=1}^{n_p}$ is a set of support points (particles) with associated weights $\{\mu^i_k\}_{i=1}^{n_p}$, and n_p is the total number of particles. Then, using the principle of importance sampling and Bayes's rule, the posterior at time k can be approximated as [6]

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) \approx \sum_{i=1}^{n_p} \mu^i_k \delta(\boldsymbol{x}_k - \boldsymbol{x}^i_k), \tag{9}$$

where $\delta(\cdot)$ is the delta function. The weights are updated as

$$\mu^i_k \propto \mu^i_{k-1} \frac{p(z_k|\boldsymbol{x}^i_k) p(\boldsymbol{x}^i_k|\boldsymbol{x}^i_{k-1})}{q(\boldsymbol{x}^i_k|\boldsymbol{x}^i_{k-1}, z_k)}. \tag{10}$$

The symbol $\propto$ in (10) stands for "proportional to" and $q(\boldsymbol{x}^i_k|\boldsymbol{x}^i_{k-1}, z_k)$ is the chosen importance density used for sampling. Weights are normalized such that $\sum_{i=1}^{n_p} \mu^i_k = 1$. Based on the strong law of large numbers, the approximated posterior (9) approaches the true $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ as $n_p \to \infty$.

The SIS algorithm thus consists of recursive propagation of the weights and particles as each measurement is received sequentially. The minimum variance state estimate $\hat{\boldsymbol{x}}_k$ and the minimum error covariance $\boldsymbol{P}_k$ can be approximated by

$$\hat{\boldsymbol{x}}_k \approx \sum\nolimits_{i=1}^{n_p} \mu^i_k \boldsymbol{x}^i_k, \tag{11a}$$

$$\boldsymbol{P}_k \approx \sum\nolimits_{i=1}^{n_p} \mu^i_k (\boldsymbol{x}^i_k - \hat{\boldsymbol{x}}_k)(\boldsymbol{x}^i_k - \hat{\boldsymbol{x}}_k)^T. \tag{11b}$$

B. Choice of the Importance Density

The optimal importance density $q(\boldsymbol{x}^i_k|\boldsymbol{x}^i_{k-1}, z_k)$ was shown to be $p(\boldsymbol{x}_k|\boldsymbol{x}^i_{k-1}, z_k)$ [5]. This importance density, however, is not always known and thus can be used only in special cases, e.g., scalar system with Gaussian measurement and process noises. Hence, the most widely used density is the a priori distribution [6]

$$q(\boldsymbol{x}_k|\boldsymbol{x}^i_{k-1}, z_k) = p(\boldsymbol{x}_k|\boldsymbol{x}^i_{k-1}). \tag{12}$$

In this case, (10) reduces to

$$\mu^i_k \propto \mu^i_{k-1} p(z_k|\boldsymbol{x}^i_k). \tag{13}$$

It can be noticed that this choice of important density is independent of the measurement z_k. Thus, the state space is explored without any knowledge of the observation and this choice can fail if a new measurements appear in the tail of the prior or if the likelihood is too peaked in comparison to the prior. This strategy promotes a well known problem of the SIS algorithm, known as the degeneracy problem [6].

C. Degeneracy Problem

The degeneracy problem is a common phenomenon when after a few iterations, all but one particle will have negligible weight. This degeneracy results in a very poor approximation of $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ and a breakdown of the algorithm. It should be noted that this phenomenon occurs even if the optimal importance density is used, but is more severe when using the prior density $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}^i)$.

A suitable measure of degeneracy of the algorithm is the effective sample size n_p^{eff} defined in [6] as

$$n_p^{eff} = \frac{n_p}{1 + \text{Var}(\mu_k^{*i})}, \tag{14}$$

where $\mu_k^{*i} = p(\boldsymbol{x}_k^i|\boldsymbol{z}_{1:k})/q(\boldsymbol{x}_k^i|\boldsymbol{x}_{k-1}^i, z_k)$ is referred to as the "true weight". This measure cannot be evaluated because usually $q(\boldsymbol{x}_k^i|\boldsymbol{x}_{k-1}^i, z_k)$ is unknown. Therefore it is normally estimated by $\hat{n}_p^{eff} = 1/\sum_{i=1}^{n_p}(\mu_k^i)^2$, where μ_k^i is the normalized weight obtained using (13). Here $1 \leq n_p^{eff} \leq n_p$, where the upper bound is attained when all particles have the same weight, and the lower bound when all probability mass is at one particle. Note that small n_p^{eff} indicates severe degeneracy. The most common solutions to tackle the degeneracy problem is the use of resampling.

D. Resampling

Resampling discards particles that have low importance weights, as they do not contribute to the approximation, and multiplies particles that have high weight [7]. This is done by mapping the weighted measure $\{\boldsymbol{x}_k^i, \mu_k^i\}_{i=1}^{n_p}$ to an unweighed measure $\{\tilde{\boldsymbol{x}}_k^i, n_p^{-1}\}_{i=1}^{n_p}$ that still provides an approximation of $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$. In other words, this process concentrates the particles in the more important regions of state space. To prevent degeneracy, when $\hat{n}_p^{eff}$ is below a fixed threshold n_p^t, an appropriate resampling procedure shall be utilized.

Several resampling schemes exist, e.g., systematic, multinomial, stratified, residual, and regularized resampling. The choice of the resampling scheme affects the computational load as well as the MC approximation error, see the discussion in [6], [12]. In the current study, only the systematic resampling strategy was considered, as it was shown empirically to outperform the other methods for the Cauchy case.

Systematic resampling is among the preferred techniques by many authors and PF users since it is simple to implement, takes $\mathcal{O}(n_p)$ time, and minimizes the MC variation [6]. Let be assumed that the weights μ_k^i are normalized before resampling, i.e., $\sum_{i=1}^{n_p}\mu_k^i = 1$. The basic idea is to generate a random number μ_k^1 from the uniform distribution on $(0, 1/n_p]$. The remaining weights $\mu_k^j, j > 1$ are obtained deterministically according to: $\mu_k^j = \mu_k^1 + n_p^{-1}(j-1)$, $j = 2, \ldots, n_p$. Then, μ_k^j are used to select particles from $\{\boldsymbol{x}_k^i, \mu_k^i\}_{i=1}^{n_p}$. In the j-th selection, the particle $\boldsymbol{x}_k^i$ is chosen when the following condition is satisfied

$$c_k^{i-1} < \mu_k^j \leq c_k^i, \tag{15}$$

where $c_k^i = \sum_{l=1}^{i} w_k^l$. The probability of selecting $\boldsymbol{x}_k^i$ is the same as that of μ_k^i being in the interval bounded by the cumulative sum of the normalized weights as shown in (15).

V. Suboptimal Solution - Gaussian Sum Filter

Here the PDFs of the noise sequences and the initial state are treated using a Gaussian sum approximation. The presented filtering scheme is a generalization of the well-known GSF algorithm of Sorenson and Alspach [3].

A. Gaussian Sum Approximation

Let $\mathcal{N}(\boldsymbol{x}; \bar{\boldsymbol{x}}, \boldsymbol{P})$ denote a Gaussian PDF with argument $\boldsymbol{x}$, mean $\bar{\boldsymbol{x}}$, and covariance matrix $\boldsymbol{P}$. Suppose that the stationary Cauchy PDFs given by (3), (4), and (5) can be approximated arbitrarily close by a finite sum of linear combinations of Gaussian PDFs as

$$p_W(w_k) \approx \sum_{i=1}^{n_w} \mu_w^i \mathcal{N}\left(w_k; \bar{w}^i, Q^i\right), \tag{16a}$$

$$p_V(v_k) \approx \sum_{i=1}^{n_v} \mu_v^i \mathcal{N}\left(v_k; \bar{v}^i, R^i\right), \tag{16b}$$

$$p_{X_1}(\boldsymbol{x}_1) \approx \sum_{i=1}^{n_{1|0}} \mu_{1|0}^i \mathcal{N}\left(\boldsymbol{x}_1; \bar{\boldsymbol{x}}_{1|0}^i, \boldsymbol{P}_{1|0}^i\right), \tag{16c}$$

where μ_w^i, μ_v^i, and $\mu_{1|0}^i$ are nonnegative weighting factors satisfying $\sum_{i=1}^{n_w}\mu_w^i = \sum_{i=1}^{n_v}\mu_v^i = \sum_{i=1}^{n_{1|0}}\mu_{1|0}^i = 1$. The Gaussian sum mixture is always a valid density function and converges uniformly to any density of practical concern [3].

B. Measurement Update

Assume that at step k the a priori density $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1})$ can be expressed as a finite sum of Gaussian densities

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k-1}) = \sum_{i=1}^{n_{k|k-1}} \mu_{k|k-1}^i \mathcal{N}\left(\boldsymbol{x}_k; \bar{\boldsymbol{x}}_{k|k-1}^i, \boldsymbol{P}_{k|k-1}^i\right). \tag{17}$$

Then, given the k-th measurement z_k, the posteriori density $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ is given by

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \sum_{i=1}^{n_{k|k-1}} \sum_{j=1}^{n_v} \tilde{\mu}_{k|k}^{ij} \mathcal{N}\left(\boldsymbol{x}_k; \bar{\boldsymbol{m}}_{k|k}^{ij}, \boldsymbol{M}_{k|k}^{ij}\right), \tag{18}$$

where $\bar{\boldsymbol{m}}_{k|k}^{ij}$ and $\boldsymbol{M}_{k|k}^{ij}$ can be computed using Kalman-like equations for all $i = 1, \ldots, n_{k|k-1}$ and $j = 1, \ldots, n_v$ as

$$\bar{z}_k^{ij} = \boldsymbol{H}\bar{\boldsymbol{x}}_{k|k-1}^i + \bar{v}^j, \tag{19a}$$

$$S_k^{ij} = \boldsymbol{H}\boldsymbol{P}_{k|k-1}^i\boldsymbol{H}^T + R^j, \tag{19b}$$

$$\boldsymbol{K}_k^{ij} = \boldsymbol{P}_{k|k-1}^i\boldsymbol{H}^T(S_k^{ij})^{-1}, \tag{19c}$$

$$\bar{\boldsymbol{m}}_{k|k}^{ij} = \bar{\boldsymbol{x}}_{k|k-1}^i + \boldsymbol{K}_k^{ij}(z_k - \bar{z}_k^{ij}), \tag{19d}$$

$$\boldsymbol{M}_{k|k}^{ij} = \boldsymbol{P}_{k|k-1}^i - \boldsymbol{K}_k^{ij}S_k^{ij}(\boldsymbol{K}_k^{ij})^T. \tag{19e}$$

The weight factors $\tilde{\mu}_{k|k}^{ij}$ are updated using the following rule

$$\tilde{\mu}_{k|k}^{ij} = \frac{\mu_{k|k-1}^i \mu_v^j \mathcal{N}(z_k; \bar{z}_k^{ij}, S_k^{ij})}{\sum_{l=1}^{n_{k|k-1}}\sum_{m=1}^{n_v}\mu_{k|k-1}^l \mu_v^m \mathcal{N}(z_k; \bar{z}_k^{lm}, S_k^{lm})}. \tag{20}$$

It is obvious that $\tilde{\mu}_{k|k}^{ij} \geq 0$ and that $\sum_{i=1}^{n_{k|k-1}}\sum_{j=1}^{n_v}\tilde{\mu}_{k|k}^{ij} = 1$. Thus, (18) is a proper PDF. For notation convenience the double summation in (18) is restated as

$$p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k}) = \sum_{i=1}^{n_{k|k}} \mu_{k|k}^i \mathcal{N}\left(\boldsymbol{x}_k; \bar{\boldsymbol{x}}_{k|k}^i, \boldsymbol{P}_{k|k}^i\right), \tag{21}$$

where $n_{k|k} = (n_{k|k-1})(n_v)$, and $\mu_{k|k}^i$, $\bar{\boldsymbol{x}}_{k|k}^i$ and $\boldsymbol{P}_{k|k}^i$ are formed in an obvious fashion from $\tilde{\mu}_{k|k}^{ij}$, $\bar{\boldsymbol{m}}_{k|k}^{ij}$ and $\boldsymbol{M}_{k|k}^{ij}$.

C. Time Propagation

Suppose that $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ is given by (21). Then the prediction density $p(\boldsymbol{x}_{k+1}|\boldsymbol{z}_{1:k})$ is

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{z}_{1:k})=\sum_{i=1}^{n_{k|k}}\sum_{j=1}^{n_w}\tilde{\mu}^{ij}_{k+1|k}\mathcal{N}\Big(\boldsymbol{x}_{k+1};\bar{\boldsymbol{m}}^{ij}_{k+1|k},\boldsymbol{M}^{ij}_{k+1|k}\Big) \quad (22)$$

with

$$\bar{\boldsymbol{m}}^{ij}_{k+1|k} = \boldsymbol{\Phi}\bar{\boldsymbol{x}}^{i}_{k|k} + \boldsymbol{\Gamma}\bar{w}^{j}, \quad (23a)$$

$$\boldsymbol{M}^{ij}_{k+1|k} = \boldsymbol{\Phi}\boldsymbol{P}^{i}_{k|k}\boldsymbol{\Phi}^T + \boldsymbol{\Gamma}Q^j\boldsymbol{\Gamma}^T, \quad (23b)$$

$$\tilde{\mu}^{ij}_{k+1|k} = \mu^{i}_{k|k}\mu^{j}_{w}. \quad (23c)$$

For convenience, one can rewrite (22) as

$$p(\boldsymbol{x}_{k+1}|\boldsymbol{z}_{1:k})=\sum_{i=1}^{n_{k+1|k}}\mu^{i}_{k+1|k}\mathcal{N}\Big(\boldsymbol{x}_{k+1};\bar{\boldsymbol{x}}^{i}_{k+1|k},\boldsymbol{P}^{i}_{k+1|k}\Big), \quad (24)$$

where $n_{k+1|k}=(n_{k|k})(n_w)$, and $\mu^{i}_{k+1|k}$, $\bar{\boldsymbol{x}}^{i}_{k+1|k}$ and $\boldsymbol{P}^{i}_{k+1|k}$ are again formed from $\tilde{\mu}^{ij}_{k+1|k}$, $\bar{\boldsymbol{m}}^{ij}_{k+1|k}$ and $\boldsymbol{M}^{ij}_{k+1|k}$. Clearly, the definition of $p_{X1}(\boldsymbol{x}_1)$, given by (16c), has the form of (24) as does the a priori PDF assumed in (17).

Having the posterior density $p(\boldsymbol{x}_k|\boldsymbol{z}_{1:k})$ in the form of (21), the conditional mean and the estimation error covariance can be approximated in the GSF sense as [3]

$$\hat{\boldsymbol{x}}_k \approx \sum\nolimits_{i=1}^{n_{k|k}} \mu^{i}_{k|k}\bar{\boldsymbol{x}}^{i}_{k|k}, \quad (25a)$$

$$\boldsymbol{P}_k \approx \sum\nolimits_{i=1}^{n_{k|k}} \mu^{i}_{k|k}\big(\boldsymbol{P}^{i}_{k|k}+(\bar{\boldsymbol{x}}^{i}_{k|k}-\hat{\boldsymbol{x}}_k)(\bar{\boldsymbol{x}}^{i}_{k|k}-\hat{\boldsymbol{x}}_k)^T\big). \quad (25b)$$

It can be noticed that both $\hat{\boldsymbol{x}}_k$ and $\boldsymbol{P}_k$ are functions of the measurement z_k and that it has no recursive structure. The major disadvantage of the GSF is that the number of terms, $n_{k|k} = n_{1|0}\cdot(n_w\cdot n_v)^k/n_w$ at step k, increases exponentially in time and thus does the computational complexity.

D. Gaussian Sum Re-approximation

To avoid exponential growth of terms $n_{k|k}$, a seemingly tempting method of keeping Gaussian components with largest weights was found to be inefficient. Even if a weight of a Gaussian component is relatively small at a certain point, it might become large at the next step. Ignoring such a component might have catastrophic effects.

In this paper, after each measurement update, we suggest to re-approximate the densities by a reduced and fixed number of Gaussian components. This reduction is motivated by the observation that a relatively small number of Gaussian densities can approximate a large class of distributions.

For ease of notation, assume that the measurement updated Gaussian sum distribution at the given step is $p_a(\boldsymbol{x})$ having originally n_a terms,

$$p_a(\boldsymbol{x}) = \sum_{i=1}^{n_a}\mu^{i}_{a}\mathcal{N}\left(\boldsymbol{x};\bar{\boldsymbol{x}}^{i}_{a},\boldsymbol{P}^{i}_{a}\right). \quad (26)$$

Our objective is to approximate $p_a(\boldsymbol{x})$ by another Gaussian sum distribution $p_b(\boldsymbol{x}_k)$ with n_b terms ($n_b \ll n_a$),

$$p_b(\boldsymbol{x}) = \sum_{i=1}^{n_b}\mu^{i}_{b}\mathcal{N}\left(\boldsymbol{x};\bar{\boldsymbol{x}}^{i}_{b},\boldsymbol{P}^{i}_{b}\right). \quad (27)$$

The task is to choose proper values of μ^{i}_{b}, $\boldsymbol{x}^{i}_{b}$ and $\boldsymbol{P}^{i}_{b}$ such that the the following cost function

$$J = \int_{-\infty}^{\infty}\Big(p_a(\boldsymbol{x}) - p_b(\boldsymbol{x})\Big)^2\,d\boldsymbol{x} \quad (28)$$

is minimized, and the first two moments of the new Gaussian mixture match exactly those of the original one, i.e.,

$$\sum_{i=1}^{n_a}\mu^{i}_{a}\bar{\boldsymbol{x}}^{i}_{a} = \sum_{i=1}^{n_b}\mu^{i}_{b}\bar{\boldsymbol{x}}^{i}_{b}, \quad (29a)$$

$$\sum_{i=1}^{n_a}\mu^{i}_{a}\left(\boldsymbol{P}^{i}_{a}+\bar{\boldsymbol{x}}^{i}_{a}(\bar{\boldsymbol{x}}^{i}_{a})^T\right)=\sum_{i=1}^{n_b}\mu^{i}_{b}\left(\boldsymbol{P}^{i}_{b}+\bar{\boldsymbol{x}}^{i}_{b}(\bar{\boldsymbol{x}}^{i}_{b})^T\right), \quad (29b)$$

$$\sum_{i=1}^{n_b}\mu^{i}_{b}=1,\ \ \mu^{i}_{b}\geq 0,\ \ \boldsymbol{P}^{i}_{b}=(\boldsymbol{P}^{i}_{b})^T>0,\ \ \forall i=1\ldots n_b. \quad (29c)$$

A small J indicates that p_b is a very good approximation of p_a. Solving this constrained minimization problem generally involves computationally costly numerical integration and nonlinear optimization with respect to many parameters.

Note that the above re-approximation may incur significant information loss since in general there is no guarantee that the re-approximated distribution also preserves the higher order moments of the original one. It was incorporated here to obtained a computationally acceptable GSF approximation.

VI. Numerical Study

In the next, the performance of the two suboptimal algorithms, discussed in the previous sections, is numerically compared and analyzed to the optimal scalar and two-state ISCE. In both cases, the same process noise, $\beta = 0.1$, and measurement noise, $\gamma = 0.2$, parameters as well as noise sequences were used. These sequences are depicted in Fig. 1.

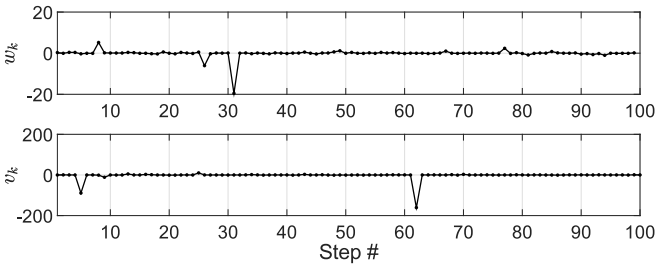

Fig. 1. Time evolution of the process and measurement noises

The implemented PF employs the systematic resampling procedure with the threshold parameter $n^t_p = 2/3n_p$. In the GSF case, the re-approximation technique presented in section V-D was implemented using standard constrained optimization tools of the Matlab environment.

A. Scalar Case

For the scalar case, the the following system parameters were chosen: $\Phi = 0.75$, $\Gamma = 1$, $H = 2$, and $\alpha = 0.5$. These parameters satisfy the condition $\gamma/|H| < \beta(1-\Phi) < \alpha$ of Assumption 4.1. in [9]. In the GSF case, the process noise, the measurement noise, and the initial state PDF were fitted

in the least-square sense with a weighted Gaussian sum of $n_w = 7$, $n_v = 7$, and $n_{1|0} = 9$ components, respectively.

Figures 2-3 are to visually compare the true posterior PDF $p(x_k|z_{1:k})$ of the ISCE at time step 8 with the approximations given by the PF and GSF with different number of particles (n_p) and Gaussian terms (n_b) kept at each step, respectively.

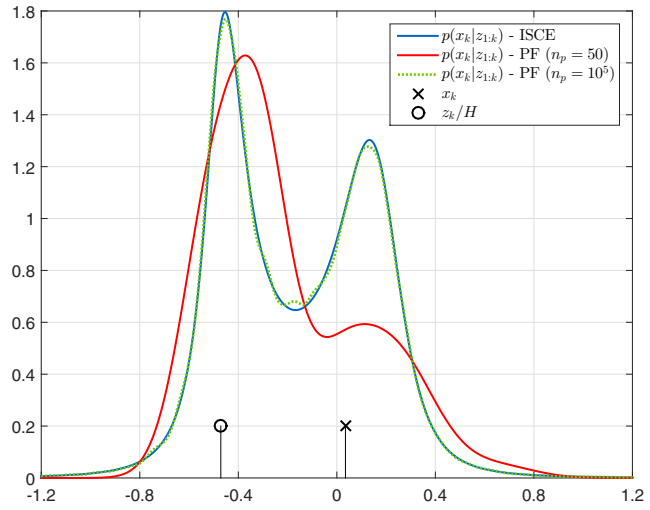

Fig. 2. Comparison of the posterior PDF with the PF.

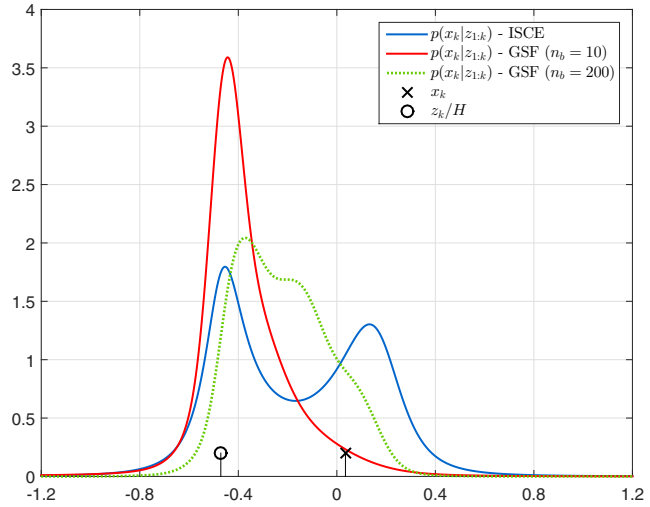

Fig. 3. Comparison of the posterior PDF with the GSF.

It can be seen from Fig. 2 that the PF with $10,000$ particles approximates reasonably well the true posteriori. However, the price to pay is the computational complexity. The average computation time of the PF with 50 particles is 4.5 times and of the PF with $10,000$ particles is $7,000$ times higher than the average computation time of the ISCE evaluated based on 100 steps, when carried out on the same computer.

On the other hand, the GSF's approximation is very poor even when 200 Gaussian components are kept at each time step, see Fig. 3. In this case, the computational burden with 10 terms kept is $1,000$ times and with 200 terms kept is $25,000$ times higher than the average computational burden of the ISCE evaluated based on 100 steps.

B. Two-state Case

In this case, the system parameters were chosen as follows:

$$\mathbf{\Phi} = \begin{bmatrix} 0.9 & 0.1 \\ -0.2 & 1.0 \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} 1.0 \\ 0.3 \end{bmatrix}, \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.10 \\ 0.05 \end{bmatrix},$$

and $\boldsymbol{H} = [\,1 \;\; 2\,]$. The system has stable eigenvalues at $0.95 \pm 0.1323j$. It is observable and complies with the necessary condition for the ISCE to exist [2], i.e., that $\boldsymbol{H\Gamma} \neq 0$. Here, the following values: $n_w = 3$, $n_v = 3$, and $n_{1|0} = 9$ were used to fit the respectively GSF's PDFs.

To allow fair comparison, the parameters of the PF and the GSF were selected such that their computational burden is similar to the ISCE implemented with the finite horizon approximation [10]. We have studied two cases. The first case considers a finite horizon (window) of 6 steps. To yield similar computation times, $n_p = 4,500$ particles were used for the PF and $n_b = 3$ Gaussian terms were kept for the GSF. The second case considers a window of 8 steps that takes longer to compute. Note that the performance of the ISCE for the 6 step and 8 step are indistinguishable. In this case $n_p = 60,000$ particles in PF were used and $n_b = 7$ Gaussian components were kept.

Figures 4-7 depict the performance of the ISCE, PF and GSF for various cases. In those plots we compare the estimation errors (solid lines) and the computed estimation error standard deviations (dotted lines) as obtained by the various filters. Figure 4 compares the performance of the ISCE and PF when the former uses a 6-step window while the latter uses $4,500$ particles. It can be seen that the PF can handle well the measurement outlier at step 5. However, after several steps, especially after encountering process noise outliers, the PF errors deviated significantly from the ISCE error. Moreover, those errors are not properly quantified by the estimated standard deviation obtained by the PF, as can be seen at larger time steps of above 30. This clearly demonstrates that $4,500$ particles are not enough to properly estimate the system states. The same conclusion can be drawn when examining the performance of the GSF depicted in Fig. 5. In this case, to have a comparable computing time, the GSF keeps only 3 Gaussian components. It clearly shows that the GSF performs very poorly, worse than the PF.

Figure 6 represents the case when far more particles are used. In this case, $60,000$ particles are used to match approximately the computational burden of the ISCE with an 8-step window. In this case the PF performance is comparable to that of the ISCE, except when a large process noise outlier is encountered at time step 62. In this case the PF overestimates the error covariance. Nonetheless, this has a nearly negligible effect on the overall estimation results. In contrast, the GSF with 7 terms, i.e., tuned to have a comparable computation time as the ISCE 8-window filter, is not capable to reproduce the performance of the Cauchy filter. This clearly demonstrates that the heavy-tail characteristics of the Cauchy noise environment cannot be captured by a finite number of Gaussian PDFs. Its performance deteriorates significantly when both process and measurement noise outliers occur.

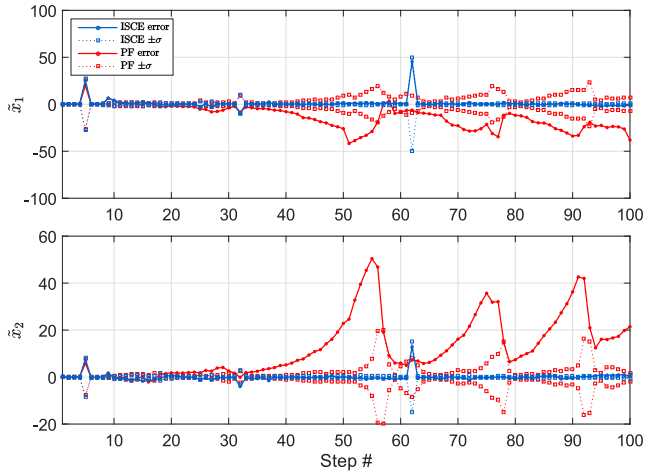

Fig. 4. Comparison of the ISCE using a window of 6 steps with the PF using 4,500 particles.

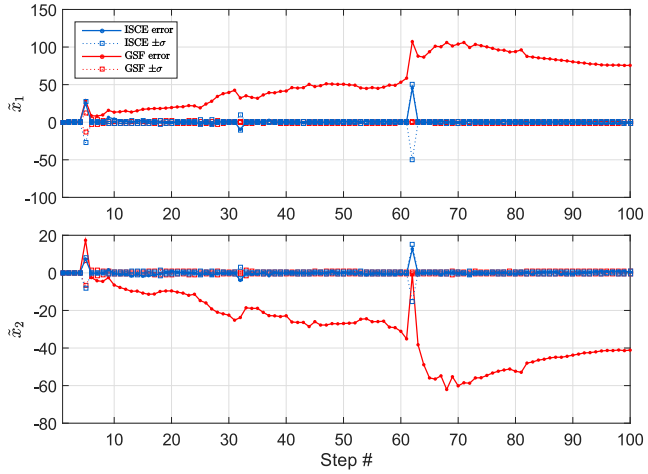

Fig. 5. Comparison of the ISCE using a window of 6 steps with the GSF keeping 3 terms at each step.

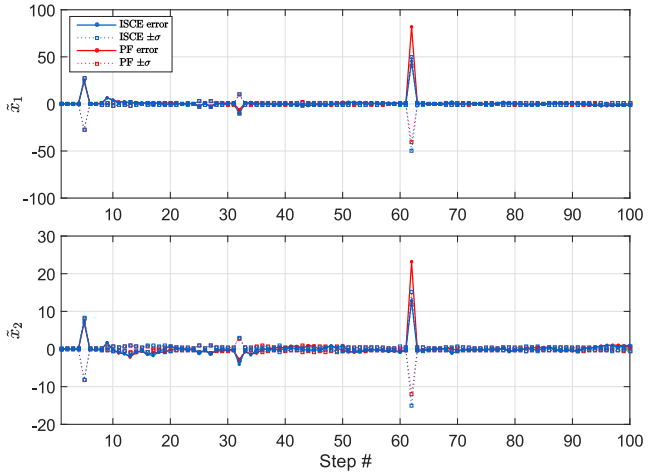

Fig. 6. Comparison of the ISCE using a window of 8 steps with the PF using 60,000 particles.

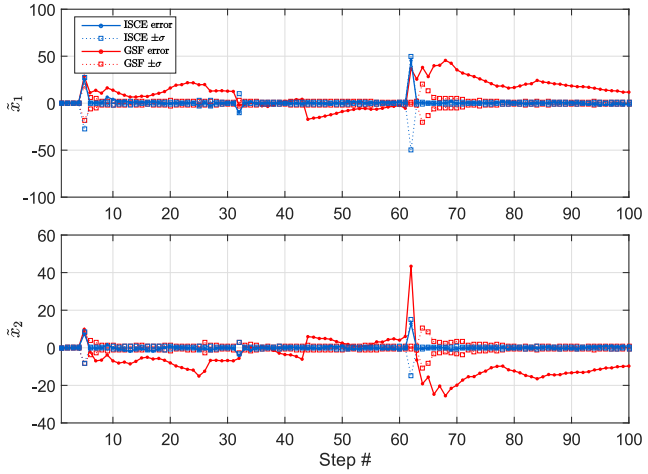

Fig. 7. Comparison of the ISCE using a window of 8 steps with the GSF keeping 7 terms at each step.

VII. Conclusion

In this paper, the performances of two suboptimal filtering algorithms have been numerically compared with the optimal Cauchy estimator. Simulation results for two-state case have revealed that both the PF and GSF perform quite poorly for a time interval dictated by the Cauchy filter for a window of six, but only the GSF perform quite poorly for a time interval a window of eight. Given the two-state ISCE's superior performance and its bounded computational cost, it represents a good candidate for real-time implementation of filtering problems in heavy-tailed Cauchy noise environment.

References

[1] M. Idan and J. L. Speyer, "State estimation for linear scalar dynamic systems with additive cauchy noises: Characteristic function approach," *SIAM Journal on Control and Optimization*, vol. 50, no. 4, pp. 1971–1994, 2012.

[2] ——, "Multivariate cauchy estimator with scalar measurement and process noises," *SIAM Journal on Control and Optimization*, vol. 52, no. 2, pp. 1108–1141, 2014.

[3] H. W. Sorenson and D. L. Alspach, "Recursive bayesian estimation using gaussian sums," *Automatica*, vol. 7, no. 4, pp. 465–479, 1971.

[4] D. L. Alspach and H. W. Sorenson, "Nonlinear bayesian estimation using gaussian sum approximations," *Automatic Control, IEEE Transactions on*, vol. 17, no. 4, pp. 439–448, 1972.

[5] A. Doucet, S. Godsill, and C. Andrieu, "On sequential monte carlo sampling methods for bayesian filtering," *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.

[6] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.

[7] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear /non-Gaussian Bayesian state estimation," in *IEE Proceedings F Radar and Signal Processing*, vol. 140, no. 2, 1993, pp. 107–113.

[8] J. R. Carpenter and A. K. Mashiku, "Cauchy drag estimation for low earth orbiters," in *AAS/AIAA Space Flight Mechanics Meeting*, Williamsburg, VA; United States, 2015.

[9] M. Idan and J. L. Speyer, "Cauchy estimation for linear scalar systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1329–1342, 2010.

[10] J. Fernandez, J. Speyer, and M. Idan, "Stochastic estimation for two-state linear dynamic systems with additive cauchy noises," *IEEE Transaction on Automatic Control*, vol. 60, no. 12, pp. 3367–3372, December 2015.

[11] A. Smith, A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo methods in practice*. Springer Verlag New York, 2013.

[12] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," *Signal Processing Magazine, IEEE*, vol. 32, no. 3, pp. 70–86, 2015.