A New Algorithm for Outlier Rejection in Particle Filters

Rohit Kumar, David Castañón, Erhan Ermis, Venkatesh Saligrama
Department of Electrical Engineering
Boston University,
Boston, MA 02115, USA
krohit@bu.edu, dac@bu.edu , ermis@bu.edu, srv@bu.edu

Abstract – *Filtering algorithms have found numerous application in various fields. One of the main factors that affect the performance of filtering algorithms is when the instrument recording the observations is faulty and yields observations which are outliers, that subsequently degrade the performance of the filter. A standard procedures to deal with this issue is to reject any measurement that is at least three standard deviations away from the predicted measurement. This method works very well for linear Gaussian estimation. For particle filter which does not require any Gaussian assumptions, the aforementioned noise rejection procedure yields poor performance. In this paper, we present a new outlier rejection procedure for particle filters that uses the theory from kernel density estimation and probability level sets. The proposed solution does not impose any constraint on the type of noise or the system transformation, and consequently the particle filter realizes its full potential. Simulation examples are presented in the end to show that our proposed algorithms works better than conventional outlier rejection algorithm.*

Keywords: Tracking, Particle Filtering, Estimation, Noise Rejection.

1 INTRODUCTION

Filtering algorithms have found wide range of applications in various areas like control theory, target tracking, economics, computer vision etc [1, 2, 3, 4]. One of the most popular linear filtering algorithms, the Kalman Filter, was presented in 1960 [5]. Five decades after its development, the Kalman filter filtering and its linearized extension to nonlinear systems, the Extended Kalman Filter (EKF), are still widely used for many applications. In recent years, there have been significant developments in computationally feasible approaches for non-linear filters for problems that violate the linear Gaussian assumptions of the Kalman filter. In particular, recent research efforts and advances in computational power has led to improved non-linear recursive filtering algorithms, such as the different versions of unscented Kalman filters (UKF) [6, 7] and particle filters [8].

Particle filters are the most common nonlinear filter algorithms that do not impose restrictions on the dynamical system model or the uncertain distributions of measurement noise [9, 10]. Particle filters represent the conditional probability density of the state in terms of samples (particles) and their weights. The filters operate recursively using prediction and update steps. At any given time 't,' particles are predicted to the next time '$t+1$' by sampling from an importance distribution; once observations are collected, an update step modifies the weights of the predicted particles. An additional resampling step is often incorporated to account for the weights to concentrate on few samples. The weight update steps gives higher weights to particles that are close to measurements and lower weights to particles away from the measurements.

Outlier measurements created by sensor failures often generate measured values that lie significantly outside the statistical confidence region for the predicted measurement values. When a particle filter receives an outlier measurement, the weight update will shift most of the weight to a few particles that are far from the correct state. Resampling will concentrate particles there, and lead to significant loss of performance. As an example, consider the application of dynamic localization of a robot estimating its location and pose from measurements of its surroundings. Outlier measurements can lead the robot to inconsistent position/pose estimations, which will propagate the error dynamically in subsequent iterations.

There are two principal ways of mitigating the effects of outlier measurements: Rejection and Accommodation [11]. In the rejection process one tries to determine via statistical methods if the received measurement is an outlier or not. If the measurement is found to be an outlier, then the prediction step of the filtering is performed and the update step is skipped. In the case of accommodation procedure, the received measurement is probabilistically weighted as whether it is a true measurement or an outlier. These probabilities are used in the weight update. The problem with accommodation approaches is that they require a statistical model for outliers in order to perform Bayesian inference.

In contrast, rejection approaches do not require a two-sided statistical model, but simply require a statistical model for "normal" measurements. In this paper, we focus on rejection approaches which have more robust assumptions.

The goal of rejection approaches is to reject a measurement as outlier, with low probability of rejecting normal measurements. For Gaussian measurements, one often uses the three standard deviation rule for rejecting measurements as outliers, proposed by Wright in 1800's [11]. He stated that if an observation is three standard deviation away from the mean then it can be rejected as an outlier, and the probability that a normal measurement would take this value is approximately 0.003. This model has been used extensively in the linear filtering literature. However for particle filters in nonlinear problems that include non-symmetric distribution, the three standard deviation rule can lead to significant errors. A better approach would be to exploit the structure of the distribution of the predicted state, as represented by the particles, and compute a 99.7th percentile interval for determine whether measurements should be declared outliers. However, note that there may be some values that have higher probability of happening outside the interval than inside, which is not the case for Gaussian distribution [12] (See Fig. 1).

Several authors have proposed techniques for outlier rejection in particle filter. In [13], the likelihoods of each updated particle are summed, and if the sum is below a threshold, the measurement is rejected as an outlier. In [14], the authors use selective updating, where each measurement is considered to be noisy with probability α and true measurement with probability $(1-\alpha)$. Consequently, a fraction α of the particles are not updated with the current measurement, and the remaining fraction of $(1-\alpha)$ are updated with the current measurements.

In this paper we present new algorithms for outlier rejection that incorporates ideas from kernel density estimation [15] and probability level sets [12, 16, 17]. It uses the predicted state particles as data samples to estimate the complete predicted distribution using kernel density estimation techniques, and to build a level set of confidence for rejecting measurements as outliers outside this set using statistical acceptance tests. We investigate variations of these algorithms with different computation requirements, and compare the performance of our outlier rejection algorithms to previous approaches proposed in the literature.

The rest of this paper is organized as follows: Section 2 presents the outlier rejection problem statement, and background on particle filter estimation and previous techniques for outlier rejection. Section 4 discusses briefly the kernel density estimation approaches and section 5 presents our new algorithms. Section 6 presents simulations of our algorithms. Section 7 discusses our conclusions and prospects for future growth.

2 Problem Statement and Background

We assume that we have a measurement device that can generate outlier observations. We assume that an observation is generated according to a normal measurement model with probability λ and an outlier with probability $1-\lambda$ as described below. Assume that the system model is given as:

$$\begin{aligned} \mathbf{x}_t &= f_t(\mathbf{x}_{t-1}, \mathbf{w}_t) \\ \mathbf{y}_t &= \begin{cases} h_t(\mathbf{x}_t, \mathbf{v}_t) & \text{if} \quad \beta < \lambda; \\ \mathbf{z}_t & \text{if} \quad \textit{otherwise}; \end{cases} \end{aligned} \tag{1}$$

where β is a Bernoulli random variable with probability λ, f_t is the system transfer function with noise $\mathbf{w}_t$, h_t is the observation function, $\mathbf{v}_t$ is the observation noise and $\mathbf{z}_t$ is some unknown random noise. For this paper we assume $\mathbf{z}_t$ to be uniformly distributed over some volume V of observation values. Note in general that neither λ nor the distribution of $\mathbf{z}_t$ are assumed to be known to the estimator, and the event that $\mathbf{z}_t$ is an outlier is not observed. Our goal is to estimate $E(\mathbf{x}_t|\mathbf{Y}^{[0,t]})$ where $\mathbf{Y}^{[0,t]} = (\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t)$.

The minimum mean square error estimate of of $\mathbf{x}_t$ given the measurements $\mathbf{Y}^{[0,t]}$ is given by the $E[\mathbf{x}|\mathbf{Y}^{[0,t]}]$. To compute this, we usually need to compute the conditional probability density $p(\mathbf{x}|\mathbf{Y}^{[0,t]})$. This density is computed recursively by following prediction and update step:

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{Y}^{[0,t-1]}) &= \int p(\mathbf{x}_t|\mathbf{x}|t-1)p(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]})d\mathbf{x}_{t-1} \\ p(\mathbf{x}_t|\mathbf{Y}^{[0,t]}) &= \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}^{[0,t-1]})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}^{[0,t-1]})d\mathbf{x}_t} \end{aligned}$$

In many situations, it is not possible to evaluate these integrals in closed form, and representation of the resulting conditional densities is complex. The particle filter addresses both of these problems by representing these densities approximately with a weighted set of samples (particles).

In case of particle filtering, let $\{\mathbf{x}_{t-1}^j, w_{t-1}^j\}_{j=1}^N$ be the particles and the corresponding weights obtained as an approximation of $p(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]})$. Therefore,

$$p(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]}) \approx \sum_{j=1}^{N} w_{t-1}^j \delta(\mathbf{x}_{t-1} - \mathbf{x}_{t-1|t-1}^j) \tag{2}$$

where N is the number of particles, and w_{t-1}^j denotes the weight of the j^{th} particle. Sometimes, it is not easy to sample particles from $p(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]})$ directly. Therefore importance sampling is used, i.e. samples are drawn from a known distribution and the sampled particles are subsequently weighted to produce a unbiased sampling estimator. The weights in particle filter are given as:

$$w_{t-1}^j = \frac{p(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]})}{q(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]})} \tag{3}$$

where q is the importance sampling distribution. To implement the filter in the recursive form, the importance sampling distribution is assumed to have following factorization:

$$q(\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{Y}^{[0,t]}) = q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{Y}^{[0,t]})q(\mathbf{x}_{t-1}|\mathbf{Y}^{[0,t-1]}) \quad (4)$$

Therefore the weight update equation at time t is given as:

$$w_t^j = \frac{p(\mathbf{y}_t|\mathbf{x}_t^j)p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j)p(\mathbf{x}_{t-1}^j|\mathbf{Y}^{[0,t-1]})}{p(\mathbf{y}_t|\mathbf{Y}^{[0,t-1]})q(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j,\mathbf{y}_t)q(\mathbf{x}_{t-1}^j|\mathbf{Y}^{[0,t-1]})} \quad (5)$$

$$= \frac{p(\mathbf{y}_t|\mathbf{x}_t^j)p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j)}{p(\mathbf{y}_t|\mathbf{Y}^{[0,t-1]})q(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j,\mathbf{y}_t)} \times \frac{p(\mathbf{x}_{t-1}^j|\mathbf{Y}^{[0,t-1]})}{q(\mathbf{x}_{t-1}^j|\mathbf{Y}^{[0,t-1]})} \quad (6)$$

$$\propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^j)p(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j)}{p(\mathbf{y}_t|\mathbf{Y}^{[0,t-1]})q(\mathbf{x}_t^j|\mathbf{x}_{t-1}^j,\mathbf{y}_t)} \times w_{t-1}^j \quad (7)$$

These weights (w_t^j) are normalized at each step and then depending on the variance of the weights, there is a resampling step performed to prevent impoverishment of particle weights.

There has been some work in the literature to find the best way to sample the particles from the importance distribution. However very less attention is given to develop particle filters that are robust to outlier measurements. As the last equation illustrates, the likelihood term $p(\mathbf{y}_t|\mathbf{x}_t^j)$ increases the weights of particles in the direction of the measurement. If the measurement is an outlier, the weights will increase for particles far from the real state values.

3 Current Techniques for Outlier Rejection

There are various statistical tests of testing whether the given measurement is a true measurement or an outlier. In this section, we discuss several such procedures.

The most commonly used procedure for outlier rejection is **Gating**. Gating assumes that the predicted probability density of the measurement has a Gaussian distribution. The basic approach is to form a validation region such that a normal measurement according to (1) (not an outlier) will fall in this region with a specified level of confidence. Therefore a measurement that lies inside this region is considered as true measurement; measurements outside this region are declared outliers.

Under the Gaussian density assumption, one can compute this region using a simple linearization argument based on the extended Kalman filter equations. Specifically,

$$(\mathbf{y}_t - h_t(\bar{\mathbf{x}}_t))^T\mathbf{S}^{-1}(\mathbf{y}_t - h_t(\mathbf{x}_t)) < \gamma^2 \quad (8)$$

where

$$\begin{aligned}
\mathbf{S} &= \mathbf{H}_t\bar{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T + \mathbf{R} \\
\bar{\mathbf{P}}_{t|t-1} &= \mathbf{F}_t\bar{\mathbf{P}}_{t-1|t-1}\mathbf{F}_t^T + \mathbf{Q} \\
\bar{\mathbf{x}}_{t-1} &= \sum_j^N \omega_{t-1}^j\mathbf{x}_{t-1}^j \\
\bar{\mathbf{P}}_{t-1} &= \sum_j^N \omega_{t-1}^j(\mathbf{x}_{t-1}^j - \bar{\mathbf{x}}_{t-1})(\mathbf{x}_t^j - \bar{\mathbf{x}}_{t-1})^T
\end{aligned}$$

where $\mathbf{F}_{t-1} = \frac{\partial f}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{t-1|t-1}}$, $\mathbf{H}_t = \frac{\partial h}{\partial \mathbf{x}}\Big|_{\hat{\mathbf{x}}_{t|t-1}}$ and γ is the threshold. The above equations start with the updated mean and covariance of the state from the particles after the update at $t-1$, and use the EKF equations to compute a Gaussian approximation to the measurement density. Alternative approaches using particle values of the predicted observations to compute the mean and variance are also easy to implement. The value of γ is set to achieve a maximum probability that the measurement may have come from one variable.

A different approach to rejection was proposed in [13]. Their strategy is based on rejecting measurements that are not close to any of the particles. Their criteria for rejection is

$$\sum_{j=1}^N p(\mathbf{y}_t|\mathbf{x}_t^j) < \eta \quad (9)$$

to sample state particles that are close to the observation. This strategy can be directly extended to our case for rejecting an observation if the sum of the likelihoods over all the particles is below a threshold. One main difficulty with this approach is that the threshold parameter does not correspond intuitively to statistics directly computable from the modeled probability density.

Another approach, introduced in [14], uses a selective update with the given measurements, where only a fraction of the particles are updated. That is, $(1-\alpha)$ particles are updated with the given measurement and α are only predicted, where α represents the expected corruption ratio and is given as:

$$\alpha = \begin{cases} 1 - \frac{p(y_t|x_t^m)}{P_t} & \text{if } \frac{p(y_t|x_t^m)}{P_t} < 1 \\ 0 & \text{if } \frac{p(y_t|x_t^m)}{P_t} > 1 \end{cases}$$

where $m = argmax_i\ p(y_t|x_t^i)$ and P_t is predefined threshold. Note that, if one particle is close to the measurement, then $\alpha = 0$ and all of the particles are updated. Otherwise, each particle is determined to be updated according to an independent Bernoulli random variable per particle, with probability $1-\alpha$ of having value 1.

4 Kernel Density Estimators

Kernel density estimator are non parametric techniques for estimating probability densities from samples. The kernel density estimator for a density f of a univariate random variable is given as:

$$\hat{f}(x) = \frac{1}{g}\sum_{i=1}^n w_i K(\frac{x-x_i}{g}) \quad (10)$$

where $\{x_i\}_{i=1}^n$ are n observations, g is the smoothing factor and K is the kernel function used. The kernel function is required to be symmetric, zero mean ($\int uK(u)du = 0$), and should integrate to one ($\int K(u)du = 1$). Standard kernel functions used in literature are: uniform $\left(\frac{1}{2}I(|u| \le 1)\right)$, triangle $((1-|u|)I(|u| \le 1))$, Gaussian $\left(\frac{1}{\sqrt{2\pi}}exp(-\frac{1}{2}u^2)\right)$, Cosine $\left(\frac{\pi}{4}cos(\frac{\pi}{2}u)I(|u| \le 1)\right)$, Epanechnikov $\left(\frac{3}{4}(1-u^2)I(|u|\leq 1)\right)$, etc, where $I(\cdot)$ is the indicator function. The kernel estimator can be seen as a sum of bumps centered around each particle where choice of the kernel determines the shape of the bump. Since the kernel is non-negative everywhere the estimated density is guaranteed to be non-negative. Furthermore, the estimated density will inherit the continuity and differentiability property of the kernels. For this paper, we shall use the Gaussian kernel. Consequently the estimated density will be smooth.

The smoothing factor is an important parameter of choice for kernel density estimation techniques. One simple way of selecting this factor is to minimize the mean integrated square error (MISE) of the estimator:

$$\begin{aligned} MISE(\hat{f}) &= \int MSE(\hat{f}(x))dx \\ &\approx \frac{1}{ng} \parallel K \parallel_2^2 + \frac{g^4}{4}\mu_2^2(K)(\int f''(x))^2 dx \end{aligned}$$

where $\parallel K \parallel_2^2$ indicates the norm and $\mu_2(K)$ indicates the variance of the kernel function. Thus $g_{opt} = \text{argmin}_g MISE(\hat{f})$ is given as:

$$g_{opt} = \left[\frac{\parallel K \parallel_2^2}{\mu_2^2(K)(\int f''(x))^2 dx}\right]^{1/5} n^{-1/5} \tag{11}$$

The only unknown in the above function is the term $(\int f''(x))^2 dx$, which is estimated using Silverman's theory assuming f to be a known density. If f is assumed to be normal density then g_{opt} is given by Silverman's rule of thumb as given below:

$$g_{rot} = 1.06 min\{\sigma, \frac{R}{1.34}\} n^{-1/5} \tag{12}$$

where σ is the variance of the data, and R is the inter quartile range ($R = X_{[0.75n]} - X_{[0.25n]}$).

For multivariate data, the kernel density estimator is given as:

$$\hat{f} = \sum_{i=1}^{n} \frac{w_i}{det(\mathbf{G})} \mathbf{K}(\mathbf{G}^{-1}(X - X_i)) \tag{13}$$

where $\mathbf{G}$ is the $d \times d$ positive definite smoothing matrix, and $\mathbf{K}$ is the multivariate kernel function with same restriction as that for univariate data. The best smoothing matrix for the multivariate data is still an active area of research. Some of the solutions mentioned in the literature are: a) An equal bandwidth g in all direction, i.e.,$\mathbf{G} = g\mathbf{I}_d$, where $\mathbf{I}_d$ is a d-dimensional identity matrix. b) Different bandwidth in all directions, i.e., $\mathbf{G} = diag(g_1, \ldots, g_d)$. For the diagonal kernel, g is selected using the following approximation:

$$g_j = \left(\frac{4}{d+2}\right)^{\frac{1}{d+4}} n^{\frac{-1}{d+4}} \sigma_j \tag{14}$$

where σ_j is the variance of the data along j^{th} direction.

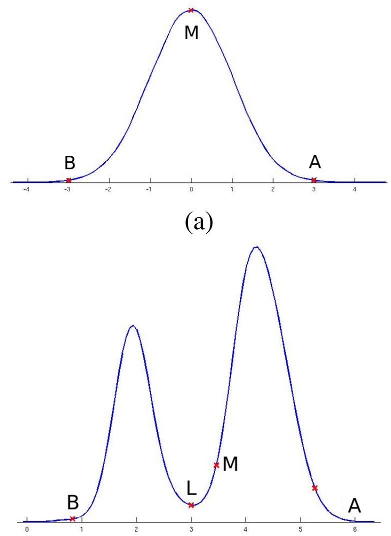

Figure 1: Figure shows an examples of the symmetric (Gaussian) and non-symmetric distribution. Points M indicates the mean of the distribution. Any measurement to the left of point B or the right of point A is rejected as noise (a) Points A and B indicate the points that is 3σ deviations away from mean (b) Point A is $99.95th$ percentile and point B is the $0.05th$ percentile. The likelihood of point L is lower than point A, although L will be accepted and A will be rejected.

5 New Outlier Rejection Algorithms

A common principle used in outlier rejection is to accept samples that have higher probability of occurrence and reject the ones with lower probability. For non symmetric distributions, there is no such center of distribution from which to construct symmetric intervals verifying the above principle as is the case in normal distribution (See Fig.(1)). One approach that avoids this inconvenience consists of choosing set with the highest density values, which corresponds to the level set of of the predicted measurement density function with probability equal to γ, and $1-\gamma$ being the given fixed error.

A main issue is how to estimate these level sets of the predicted measurement density function. To that effect, we will use non-parametric kernel density estimation techniques. As a result, our approach is well-suited for exploiting the advantages of particle filters in handling arbitrary non-linearities and non-Gaussian measurement noise.

Let $\{y_t^i\}_{i=1}^N$ be sample observation particles of the particle filter at time t, and is given as:

$$y_t^i = h_t(x_t^i) + v_t \tag{15}$$

where we have added the noise v_t to the predicted observations. We use these samples to interpolate the density using non-parametric density estimation technique as follows:

$$f_{y_p}(y_t) = \sum_{i=1}^{N} \frac{w_i}{g} K(\frac{y_t - y_t^i}{g}) \tag{16}$$

where g is the smoothing parameter, w_i is the weight associated with particle x_t^i, and K is the kernel used to estimation.

Given the predicted observation distribution, the goal is to test if the current observation is an outlier. We perform the hypothesis test:

$$\begin{aligned} H_0 &: \int_{\{x: f_{y_p}(x) < f_{y_p}(y_t)\}} f_{y_p}(x) dx \geq \gamma \\ H_1 &: else \end{aligned}$$

This integral can be readily calculated Gaussian distributions, but is much harder to compute for arbitrary Gaussian sum distributions as in (16). In our algorithm, we propose Monte Carlo techniques to evaluate this integral by sampling as:

$$\int_{x: f_{y_p}(x) < f_{y_p}(y_t)} f_{y_p}(x) dx \approx \frac{\sum_{i=1}^{T} I(f_{y_p}(y_i) < f_{y_p}(y_t))}{T} \quad (17)$$

where y_i are independent samples drawn from $f_{y_p}(y)$, $I(c)$ is an indicator function which is equal to one if the condition c is true else it is zero, T is the total number of samples. The test of (17) uses an empirical level of significance based on fractions of samples with higher probability density than the measurement being tested. If this fraction is above the threshold γ then hypothesis H_0 is accepted else H_1 is accepted. Fig. 2 shows an example of the proposed algorithm.

While the above solution is appealing and mathematically rigorous, computation of the required Monte Carlo integral may require lots of samples if the desired threshold is small. If we assume some additional structure, we may be able to reduce the required computations. We now assume that the observation noise in (1) is zero-mean additive Gaussian noise with known diagonal covariance with diagonal element σ^2.

The approach is based on using Gaussian kernels to fit the predicted average observations $h_t(x^i, t)$ without adding measurement noise. Then the density obtained by adding independent Gaussian noise can be obtained by

$$p_{y_p}(y_t) = \sum_{i=1}^{N} \frac{w_i}{g} K(\frac{y_t - h_t(x_t^i)}{g}) \quad (18)$$

this is nothing but a normal kernel with additive Gaussian noise. In this case we add the noise analytically rather than sampling as suggested in the earlier approach. Consequently the problem reduces to:

$$p_{y_p}(y_t) = \sum_{i=1}^{N} \frac{w_i}{g} K_1(\frac{y_t - h_t(x_t^i)}{g}) \quad (19)$$

where kernel K_1 is $\mathcal{N}(0, 1 + \sigma^2/g^2)$. The process of adding noise analytically reduces the computational complexity.

However, the main cause of the complexity in computing the level set is having to sample this density to empirically estimate this level set. We adopt an approximate approach, based on identifying the kernel that is closest to the observed value, and use the level set of this kernel density for the rejection value. Note that this level set will give a conservative level of significance. Mathematically, let

$$m = \text{argmin}_{i=1,\ldots,N} \parallel \{y_t - h_t(x_t^i)\} \parallel \quad (20)$$

Then a given measurement is accepted as true measurement is the condition in (21) is satisfied else it is rejected as noise.

$$(y_t - x_t^m)^T (g^2 + var(v_t))^{-1} (y_t - x_t^m) < \gamma^2 \quad (21)$$

where γ can be chosen for the specified level of significance under Gaussian assumptions.

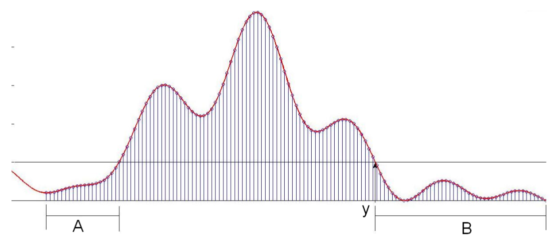

Figure 2: Figure shows a randomly sampled probability distribution function that is used to evaluate Eqn.(17). The observation y decides the probability level and then sampled to estimate the area under the level. Set A and Set B indicates the rejection region. In probability level set, the rejection region does not necessarily have to be contiguous.

6 Examples

In this section we present the simulation results to show that proposed particle filters work better than the conventional particle filter when outliers are present. The system used for simulations is summarized below.

$$\begin{aligned} x_t &= 0.5x_{t-1} + 25\frac{x_{t-1}}{1 + x_{t-1}^2} + 8\cos(1.2t) + w_t \\ y_t &= \begin{cases} \frac{x_t^2}{20} + v_t & \text{with probability } \lambda \\ z_t & \text{otherwise} \end{cases} \end{aligned} \quad (22)$$

where $\lambda = 0.7$, $v_t \sim \mathcal{N}(0, 2)$, $w_t \sim \mathcal{N}(0, 10)$ and z_t $U[-15, 15]$.

We ran experiments using the following algorithms: on a 3.6GHz linux machine with 3.5GB of memory:

DPF : Conventional particle filter with no outlier rejection

GPF : Particle filter with gating technique (3σ rule) from Section 3.

SLPF : Particle filter with summing likelihood outlier rejection in Eqn.(9) from Section 3. $\eta = 10^{-4}$ as used in [13].

SUPF : Particle filter with summing likelihood theory from Section 3. $P_t = 0.5$ is used.

Filter	MSE	Avg. Runtime in secs
DPF	8.6203	5.2915
GPF	7.4844	5.3753
SPF	7.9214	5.0900
SUPF	8.0917	5.0154
PF1	6.9063	13.4660
PF2	6.9397	13.7608
PF3	6.4829	5.1755

Table 1: RMS error in experiment for the different algorithms

PF1 : Proposed particle filter with full simulation with theory in Section 5. $\gamma = 0.1$ is used.

PF2 : Proposed particle filter with analytical noise incorporation with theory in Section 5. $\gamma = 0.1$ is used.

PF3 : Proposed particle filter with outlier rejection as in (21) with $\gamma = 3$. .

Each experiment was run with 100 particles for 50 time steps. The simulations are plotted in the Fig.(3), and the root mean square (RMS) and the average run time for the simulation are tabulated in Table 6. As one can see, PF1, PF2 and PF3 yield the best tracking results in RMS sense. However, both PF1 and PF2 take almost three times longer to to run (for $T = 10^4$). The reason for increase in run time is due to Eqn.(17) that requires evaluating T samples at each time step. Another side experiment was performed to run. In contrast, PF3 runs as fast as other algorithms, and has much better performance in the presence of outliers.

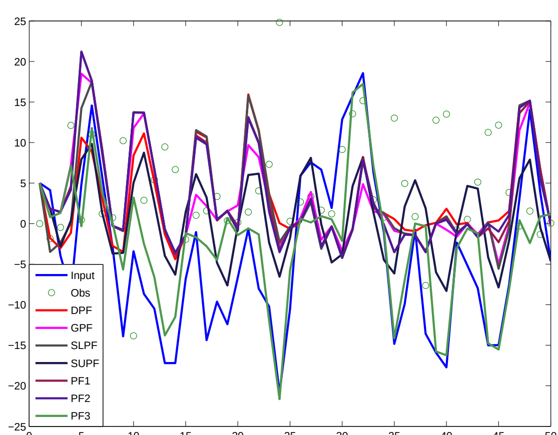

Figure 3: Graphs for the various filters to compare results.

7 Conclusion

In this paper, we presented new algorithms for incorporating outlier rejection in particle filters. These algorithms are based on statistical tests for acceptance of the normal hypothesis, and use empirical statistics to estimate the level of significance of these tests. We evaluated the algorithms on a nonlinear estimation problem, and compared their performance to other approaches proposed in the literature. our results establish that our algorithms have improved outlier rejection properties, albeit with some increased complexity in computation time.

We developed an approximation to our algorithms for the special case of additive Gaussian noise in nonlinear observations. The approximate statistical outlier rejection algorithm was as fast as the competing algorithms in the literature, and achieved performance comparable to the our slower algorithms.

In terms of future work, the above results need to be evaluated over a broader range of test problems, including multivariate problems where the computation of level sets are harder. In addition, because we are evaluating low probability events (outliers) using Monte Carlo sampling, one should investigate the use of importance sampling to reduce the number of samples involved in determining the boundary for outlier rejection.

8 Acknowledgment

We would like to thank Alex Ihler and Mike Mandel for their free kernel density estimation software. This work was sponsored by AFOSR.

References

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

[2] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[3] A. Gelb, *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.

[4] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond Kalman Filter:Particle Filter for Tracking Applications*. Artech House Publisher, 2004.

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

[6] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of The IEEE*, vol. 92, no. 3, pp. 401–422, March 2004.

[7] ——, "A general method for approximating nonlinear transformations for probability distributions," 1996.

[8] N. J. Gordon, S. D. J., and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian bayesian tracking," *Radar and Signal Processing, IEE Proceedings-F*, vol. 140, no. 2, pp. 107–113, 1993.

[9] 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, no. 2, pp. 174–188, Feburary 2002.

[10] 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.

[11] V. Barnett and T. Lewis, *Outlier in Statistical Data*. Wiley and Sons, 1996.

[12] J. N. Garcia, Z. Kutalik, K.-H. Cho, and O. Wolkenhauer, "Level sets and minimum volume sets of probability density functions," *International Journal of Approximate Reasoning*, vol. 34, no. 1, pp. 25–47, September 2003.

[13] X.-L. Hu, T. Schon, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1337–1348, April 2008.

[14] J.-S. Lee and W. K. Chung, "Robust particle filter localization by sampling from non-currupted window with incomplete map," *International Conference on Intelligent Robots and Systems, Nice, France*, pp. 1133–1139, September 2008.

[15] W. Hardle, M. Muller, S. Sperlich, and A. Werwatz, *Nonparametric and Semiparametric Models*. Springer Series in Statistics, 2004.

[16] C. Scott and E. Kolaczyk, "Nonparametric assessment of contamination in multivariate data using minimunm volume sets and fdr," *Journal of Computation and Graphical Statistics*, 2010.

[17] E. B. Ermis and V. Saligrama, "Distributed detection in sensor networks with limited range multi-modal sensors," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 843–858, February 2010.