BAYES OPTIMAL KNOWLEDGE EXPLOITATION FOR TARGET TRACKING WITH HARD CONSTRAINTS

***F. Papi*[1], *M. Podt*[1], *Y. Boers*[1], *G. Battistello*[2] *and M. Ulmke*[2]**

[1] *Thales Nederland BV, Sensors-TBU Radar Engineering*
Hengelo, The Netherlands, mail: francesco.papi@nl.thalesgroup.com
[2] *Fraunhofer FKIE, Sensor Data and Information Fusion Department*
Wachtberg, Germany, mail: giulia.battistello@fkie.fraunhofer.de

Keywords: Constrained Bayesian Filtering, Sequential Monte Carlo, Pseudo-Measurements, Rejection-Sampling, Kullback-Leibler Divergence.

Abstract

Nonlinear target tracking is a well known problem and its Bayes optimal solution, based on particle filtering techniques, is nowadays applied in high performance surveillance systems. Oftentimes, additional information about the environment and the target is available, and can be formalized in terms of constraints on target dynamics. Hence, a *Constrained* version of the Bayesian Filtering problem has to be solved to achieve optimal tracking performance.
In this paper we consider the *Constrained Filtering* problem for the case of perfectly known *hard* constraints. We clarify that in such a case the Particle Filter (PF) is still Bayes optimal if we can correctly model the constraints. We then show that from a Bayesian viewpoint, exploitation of the available knowledge in the prediction or in the update step are equivalent. Finally, we consider simple techniques to exploit constraints in the prediction and update steps of a PF, and use the Kullback-Leibler divergence to illustrate their equivalence through simulations.

1 Introduction

Many real life problems require the estimation of the state of a system that changes over time using a sequence of noisy measurements made on the system. Target tracking based on measurements collected by a radar or a similar sensor is an important application example.
Recently, accurate modeling has required the inclusion of nonlinearity and non-Gaussianity in the equations used for estimation purposes, thus making the Kalman Filter (KF) inapplicable in its basic form. Nonlinear extensions of the KF are the Extended Kalman Filter (EKF) [2], and the Unscented Kalman Filter (UKF) [10], both based on deterministic approximations. However, none of these methods is Bayes optimal.
Bayesian methods provide a rigorous framework based on the idea of constructing the a posteriori *probability density function* (PDF) of the state given all the available information, and then numerically approximating such a PDF. The *nonlinear filtering* problem is then recursively solved by a PF, which is a Monte Carlo based approximation of the Bayesian recursion, and nowadays represents the state of art in nonlinear target tracking. Such filters operate by propagating particles that are distributed according to the approximately true PDF of the state, and convergence to the true a posteriori distribution is guaranteed for a sufficiently large number of particles [6].
Oftentimes, additional information about the environment and the target is available, and can be formalized in terms of constraints on target dynamics. Many sources of external knowledge may be available like, for instance, the output of a classification algorithm, which provides valuable information on the model to be used for a target, and 2D/3D digital maps of the environment, which surely improve the modeling process (e.g., only track targets inside a road for ground vehicles [14]).
The idea of using state constraints to improve the tracking performance dates back to the 90s, when attempts to exploit hard linear equality constraints in Kalman Filtering led to the definition of the *Pseudo-Measurements* approach [12]. The key idea is to interpret the constraints as additional measurements, and it is proven to be optimal when using *hard* constraints.
Hard inequality state constraints are used in [4] to represent the known flight envelope (i.e., minimum and maximum velocities). The authors propose to obtain samples from a truncated distribution using a *Rejection-Sampling* approach. The proposed PF shows good performance, but might be practically unfeasible due to the computational load required.
A refined model with state dependent detection probability and clutter notch information is proposed in [11] for airborne GMTI based ground tracking. Equality and inequality constraints are used to model the known road network. Both Gaussian sum and particles based approximations are considered.
In [7] the use of PF for littoral tracking is proposed by formulating the problem as Joint Tracking and Classification (JTC), where a target class is assigned for each isolated land or water region. A similar approach is followed in [1], where the authors propose a modified version of the JTC-PF algorithm that uses class-dependent speed likelihoods.
In this paper we consider the Bayes formulation of the *Constrained Filtering* problem for the case of *hard* constraints. In particular, we define a constraint as being *hard* if it is perfectly known and the target does not violate it. We show that from a Bayesian viewpoint, exploitation of the available knowledge in

the prediction or in the update step are equivalent. This implies that for a sufficiently large number of particles, the *Pseudo-Measurements* approach can be seen as a practically feasible alternative to the *Rejection-Sampling* approach. We use simulations to additionally illustrate the equivalence of the two techniques by showing that the Kullback-Leibler divergence between the empirical distributions converges to zero.
The paper is organized as follows: in section 2 we review the Bayesian Filtering; section 3 is dedicated to the Constrained Filtering problem; in section 4 we review the two main techniques to exploit constraints in particle filtering, and report simulation results in section 5; section 6 collects our conclusions.

2 Bayesian Filtering Problem

In this section we briefly describe the *Bayesian Filtering Problem* and recall a convergence result [6] for its particle filtering based solution. Suppose the system is described by the following state and measurement equations:

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{w}_k \quad (1)$$

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k \quad (2)$$

where $\mathbf{x}_k \in \mathbb{R}^{n_x}$ is the system state, $\mathbf{z}_k \in \mathbb{R}^{n_z}$ the measurement vector, $\mathbf{w}_k \sim p_{\mathbf{w}_k}(\mathbf{w})$ the process noise, and $\mathbf{v}_k \sim p_{\mathbf{v}_k}(\mathbf{v})$ the measurement noise. Therefore, the Markov property holds for the system of eqs. (1)-(2), i.e.,

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \ldots, \mathbf{x}_0) = p(\mathbf{x}_k|\mathbf{x}_{k-1}) \quad (3)$$

where $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is known as the *transition Kernel*.
Let $\mathbf{Z}^k \triangleq \{ \mathbf{z}_0 \quad \mathbf{z}_1 \quad \ldots \quad \mathbf{z}_k \}$ be the sequence of measurements up to and including time k. Hence, the measurement $\mathbf{z}_k$ at time k is independent from past states, i.e.,

$$p(\mathbf{z}_k|\mathbf{z}_{k-1}, \ldots, \mathbf{z}_1, \mathbf{x}_k, \mathbf{x}_{k-1}, \ldots, \mathbf{x}_0) = p(\mathbf{z}_k|\mathbf{x}_k) \quad (4)$$

where $p(\mathbf{z}_k|\mathbf{x}_k)$ is known as the *likelihood function*.
Given a realization of $\mathbf{Z}^k$ associated with the system (1)-(2), the filtering problem aims at computing the conditional probability density $p(\mathbf{x}_k|\mathbf{Z}^k)$. Thus filtering consists of finding the a posteriori probability distribution of the system state conditioned on all past measurements. Let us assume that at time step $k-1$ the probability distribution $p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1})$ is available. Then Bayesian filtering is solved using a two step recursion:

- *Prediction Step*

$$p(\mathbf{x}_k|\mathbf{Z}^{k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1})\, d\mathbf{x}_{k-1} \quad (5)$$

  where $p(\mathbf{x}_k|\mathbf{Z}^{k-1})$ is the *predictive density* at time k.

- *Update Step*

$$p(\mathbf{x}_k|\mathbf{Z}^k) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}^{k-1})}{p(\mathbf{z}_k|\mathbf{Z}^{k-1})} \quad (6)$$

  where $p(\mathbf{z}_k|\mathbf{Z}^{k-1})$ is the *Bayes normalization constant*.

Various state estimators are obtained from the a posteriori PDF $p(\mathbf{x}_k|\mathbf{Z}^k)$, e.g., the *minimum variance* (MV) estimator, i.e.,

$$\hat{\mathbf{x}}_k^{MV} \triangleq \int_{\mathbb{R}^{n_x}} \mathbf{x}_k\, p(\mathbf{x}_k|\mathbf{Z}^k)\, d\mathbf{x}_k \quad (7)$$

or the *maximum a posteriori* (MAP) estimator, i.e.,

$$\hat{\mathbf{x}}_k^{MAP} \triangleq \underset{\mathbf{x}_k \in \mathbb{R}^{n_x}}{\arg\max}\ p(\mathbf{x}_k|\mathbf{Z}^k) \quad (8)$$

More in general, if $\boldsymbol{\phi}(\mathbf{x}_k) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_\phi}$ is a function of the state we want to estimate, most estimation algorithms compute an approximation of the conditional expectation:

$$\mathbf{E}\left(\boldsymbol{\phi}(\mathbf{x}_k)|\mathbf{Z}^k\right) = \int \boldsymbol{\phi}(\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}^k)\, d\mathbf{x}_k \quad (9)$$

The particle filter computes an approximation of (9) using the empirical filtering density [8]:

$$\hat{p}_N(\mathbf{x}_k|\mathbf{Z}^k) = \textstyle\sum_{i=1}^N w_k^i\, \delta_{\mathbf{x}_k^i}(\mathbf{x}_k), \quad \sum_{i=1}^N w_k^i = 1 \quad (10)$$

where N is the number of particles. Each particle $\mathbf{x}_k^i$ has an importance weight w_k^i associated to it, and $\delta_{\mathbf{x}_k^i}(\cdot)$ denotes the delta-Dirac mass located at $\mathbf{x}_k^i$.
Convergence results for the mean square error in approximating eq. (9) have been given in [6], i.e.,
Let the likelihood function $p(\mathbf{z}_k|\cdot)$ be bounded in the argument $\mathbf{x}_k \in \mathbb{R}^{n_x}$, i.e., $\|p\| < \infty$, and the system Kernel be weakly dependent on past state values, then for all $k \geq 0$ there exist a constant c such that for any function $\boldsymbol{\phi} \in B(\mathbb{R}^{n_x})$:

$$\mathbf{E}\left[\left((\hat{p}_N, \boldsymbol{\phi}) - (p, \boldsymbol{\phi})\right)^2\right] \leq c\, \frac{\|\boldsymbol{\phi}\|}{N} \quad (11)$$

where N is the number of particles, $B(\cdot)$ the set of Borel bounded functions in $\mathbb{R}^{n_x}$, and the notation $(p, \boldsymbol{\phi}) \triangleq \int p\phi$ is used. This roughly means that if the true optimal filter is quickly mixing, then uniform convergence in time of the particle filtering method is ensured. In practice, a sufficiently large number of particles is required. Theoretical analyses on the minimal number of particles are reported in [3].

3 Constrained Bayesian Filtering

Oftentimes additional information about the state is available. In fact, in the state space description of eq. (1), the state variables usually correspond to physical quantities of interest, and validity regions may be helpful for the filter design. For instance, speed constraints can be defined based on the solution of the classification problem and/or based on the type of terrain/sea traveled at the moment.
Here we specifically focus on *Constrained Bayesian Filtering* for the case of *hard* constraints,which are oftentimes encountered in practice. Examples are the tracking of ground vehicles moving on a road network, or the tracking of ships traveling on canals. In general, all the constraints arising from physical laws, e.g., the mass conservation for chemical reactions, are of

this type. We are interested in clarifying that the best way of using such information is through the Bayesian filtering recursion. Hence, let assume that external information is available in terms of nonlinear inequality constraints:

$$\mathbf{a}_k \leq \mathbf{C}_k(\mathbf{x}_k) \leq \mathbf{b}_k \tag{12}$$

where $\mathbf{C}_k : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_c}$, and the inequality sign holds for all elements. For convenience, let $\mathcal{C}_k$ be the set of all states satisfying the inequality constraint (12):

$$\mathcal{C}_k \triangleq \{\mathbf{x}_k : \mathbf{x}_k \in \mathbb{R}^{n_x}, \mathbf{a}_k \leq \mathbf{C}_k(\mathbf{x}_k) \leq \mathbf{b}_k\} \tag{13}$$

and $\mathcal{C}^k \triangleq \{\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_k\}$ be the sequence of $\mathcal{C}_k$ up to time k. From a Bayesian viewpoint, exploitation of external knowledge boils down to finding an approximation of

$$p(\mathbf{x}_k|\mathbf{Z}^k, \mathcal{C}_k) \propto \begin{cases} p(\mathbf{x}_k|\mathbf{Z}^k), & \text{if } \mathbf{x}_k \in \mathcal{C}_k \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where conditioning is performed also with respect to the sequence $\mathcal{C}^k$ of constrained state variables. Now, let us assume that we are able to define a two step prediction-update recursion like the one in eqs. (5) and (6) such that:

$$\begin{array}{ccc} p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1}, \mathcal{C}_{k-1}) & \xrightarrow{Prediction} & p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}_k) \\ p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}_k) & \xrightarrow{Update} & p(\mathbf{x}_k|\mathbf{Z}^k, \mathcal{C}_k) \end{array}$$

Then, for a sufficiently large number of particles, the empirical distribution $\hat{p}_N(\mathbf{x}_k|\mathbf{Z}^k, \mathcal{C}^k)$ converges to the true PDF in (14). Notice that we are implicitly assuming that eqs. (1)-(2) are a correct description of the continuous time evolution. If this is not the case, a large sampling time could mean that most or all particles will violate the constraint after prediction, this way generating a very depleted approximation of the posterior PDF. However, in such a situation one may choose a finer sampling time to enforce the constraints with improved performance.
In the following we define two Bayesian recursions for *Constrained Filtering*, in which knowledge is used in the prediction step in one case, and in the update step in the other case.

3.1 Using Knowledge in the Prediction Step

To exploit information in the prediction step, we define the following predictive PDF:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k) &= \\ &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{Z}^{k-1}, \mathcal{C}^k)\, d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k)\, p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1})\, d\mathbf{x}_{k-1} \end{aligned} \tag{15}$$

Proceeding, the Bayesian update step requires us to define the following a posteriori distribution:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}^k, \mathcal{C}^k) &= \frac{p(\mathbf{x}_k, \mathbf{z}_k, \mathbf{Z}^{k-1}, \mathcal{C}^k)}{p(\mathbf{z}_k, \mathbf{Z}^{k-1}, \mathcal{C}^k)} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{Z}^{k-1}, \mathcal{C}^k)\, p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k)\, p(\mathbf{Z}^{k-1}, \mathcal{C}^k)}{p(\mathbf{z}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k)\, p(\mathbf{Z}^{k-1}, \mathcal{C}^k)} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k)}{p(\mathbf{z}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k)} \end{aligned} \tag{16}$$

3.2 Using Knowledge in the Update Step

Here we define the predictive distribution without conditioning on the set $\mathcal{C}_k$, i.e.,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1}) &= \\ &= \int p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1})\, d\mathbf{x}_{k-1} \\ &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\, p(\mathbf{x}_{k-1}|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1})\, d\mathbf{x}_{k-1} \end{aligned} \tag{17}$$

And then update using both $\mathbf{z}_k$ and $\mathcal{C}_k$, i.e.,

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}^k, \mathcal{C}^k) &= \frac{p(\mathbf{x}_k, \mathbf{z}_k, \mathbf{Z}^{k-1}, \mathcal{C}_k, \mathcal{C}^{k-1})}{p(\mathbf{z}_k, \mathbf{Z}^{k-1}, \mathcal{C}_k, \mathcal{C}^{k-1})} \\ &= \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathcal{C}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1})}{p(\mathbf{z}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k)\, p(\mathcal{C}_k|\mathcal{C}^{k-1})} \end{aligned} \tag{18}$$

Let us now compare the a posteriori PDFs defined by eqs. (16) and (18). We conclude that the two Bayesian recursions are targeting the same a posteriori PDF if the following holds:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^k) &= \frac{p(\mathbf{x}_k, \mathbf{Z}^{k-1}, \mathcal{C}_k, \mathcal{C}^{k-1})}{p(\mathbf{Z}^{k-1}, \mathcal{C}_k, \mathcal{C}^{k-1})} \\ &= \frac{p(\mathcal{C}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{Z}^{k-1}, \mathcal{C}^{k-1})}{p(\mathcal{C}_k|\mathcal{C}^{k-1})} \end{aligned} \tag{19}$$

which is true thanks to Bayes theorem. In summary, exploitation of perfectly known *hard constraints* in the prediction or in the update step of the Bayesian recursion are equivalent. That is, for a large number of particles, a particle filtering approximation to the *constrained filtering recursion* will provide the same results independently from processing external knowledge in the prediction or in the update step of a PF.

4 Particle Filtering Methods

We now review two methods for PF based *Hard Constrained Filtering*: *(a)* the *Rejection-Sampling PF* which carries out the processing described in section 3.1, and *(b)* the *Pseudo-Measurements PF* which is a PF based approximation of the processing described in section 3.2.

4.1 *Rejection-Sampling (Prediction Step)*

A procedure to perform constrained sampling was introduced in [4]. Consider the conditional probability theorem:

$$p(x|A) = \frac{p(x, A)}{p(A)} \tag{20}$$

where A represents constraints on x such that $A = \{x :\ x \leq a\}$. Then the following are true:

$$p(x|A) = p(x|x \leq a) = \frac{p(x, x \leq a)}{p(x \leq a)} \tag{21}$$

$$p(x|A) = \begin{cases} \dfrac{p(x)}{p(x \leq a)}, & \text{if } x \leq a \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

In other words, the constrained PDF is the original $p(x)$ restricted to A and normalized. This is obtained using the PF described in Algorithm 1, where N_{eff} is the effective number of particles and $\beta \in (0, 1]$ is a suitable parameter, respectively, as defined in [9]. The solution is extremely simple, but is computationally expensive. In particular, the time required to perform rejection-sampling is not known a priori.

Algorithm 1: Rejection-Sampling Particle Filter

Input: $\left\{\mathbf{x}_{k-1}^i, w_{k-1}^i\right\}_{i=1}^N$ and the new measurement $\mathbf{z}_k$

Output: $\left\{\mathbf{x}_k^i, w_k^i\right\}_{i=1}^N$

while $i = 1, 2, \ldots, N$ ***(Prediction Step)*** **do**
  while $x_k^i \notin \mathcal{C}_k$ ***(Rejection-Sampling)*** **do**
    Generate a New Particle: $\mathbf{x}_k^i \sim p_k(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)$
  end
end

while $i = 1, 2, \ldots, N$ ***(Update Step)*** **do**
  Compute Weights: $\tilde{w}_k^i = w_{k-1}^i \, p(\mathbf{z}_k|\mathbf{x}_k^i)$;
end

Normalization Step: $w_k^i = \tilde{w}_k^i / \sum_{i=1}^N \tilde{w}_k^i \quad \forall \, i$;

Effective Sample Size: $N_{eff} = 1/\sum_{i=1}^N (w_k^i)^2$;

if $N_{eff} \leq \beta N$ ***(Resampling Step)*** **then**
  New Particles $\left\{\tilde{\mathbf{x}}_k^i, 1/N\right\}_{i=1}^N$ s.t. $P(\tilde{\mathbf{x}}_k^i = \mathbf{x}_k^i) = w_k^i$
end

4.2 *Pseudo-Measurements (Update Step)*

First introduced in [12], the *Pseudo-Measurements* approach interprets the constraints on the state variables as additional measurement equations. The main step is the definition of a constraint based likelihood function:

$$p(\mathcal{C}_k|\mathbf{x}_k^i) = \begin{cases} 1, & \text{if } \mathbf{a}_k \leq \mathbf{C}_k(\mathbf{x}_k^i) \leq \mathbf{b}_k \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

Thus, an additional likelihood function is used to evaluate the unnormalized weights in the PF algorithm, i.e.,

$$\tilde{w}_k^i = w_{k-1}^i \, \frac{p(\mathbf{z}_k|\mathbf{x}_k^i) \, p(\mathcal{C}_k|\mathbf{x}_k^i) \, p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{q_k(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)} \tag{24}$$

where the $\tilde{}$ sign is used to remember that normalization has to be performed. Furthermore, if we use the transition Kernel as proposal distribution, we have the common simplification:

$$\tilde{w}_k^i = w_{k-1}^i \, p(\mathbf{z}_k|\mathbf{x}_k^i) \, p(\mathcal{C}_k|\mathbf{x}_k^i) \tag{25}$$

which leads us to the definition of the PF described in Algorithm 2. The approach generally requires a large number of particles since many of them are discarded at each step. However, experience tells the computational load is strongly reduced compared to the *Rejection-Sampling PF*.

4.3 Discussion about the two filters

Let us now show that the presented algorithms provide the same results from a practical viewpoint. Let $\mathbf{X}^k \triangleq \left\{ \mathbf{x}_0 \quad \mathbf{x}_1 \quad \ldots \quad \mathbf{x}_k \right\}$ be the sequence of system states. Hence, the general expression for the weights of particles is:

$$w_k = \frac{p(\mathbf{Z}^k|\mathbf{X}^k) \, p(\mathbf{X}^k)}{p(\mathbf{Z}^k)} \tag{26}$$

which for sequential *constrained filtering* becomes:

$$\tilde{w}_k = w_{k-1} \, \frac{p(\mathbf{z}_k|\mathbf{x}_k) \, p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k)}{q(\mathbf{x}_k|\mathbf{X}^{k-1}, \mathbf{Z}^k)} \tag{27}$$

where we assume that we are able to evaluate the *constrained transition Kernel* $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}^k)$.
In *Rejection-Sampling* the following equality holds true:

$$q(\mathbf{x}_k|\mathbf{X}^{k-1}, \mathbf{Z}^k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k) \tag{28}$$

which yields for the evaluation of the weights:

$$\tilde{w}_k^{RS} = w_{k-1}^{RS} \, p(\mathbf{z}_k|\mathbf{x}_k) \tag{29}$$

In the *Pseudo-Measurements* approach we choose:

$$q(\mathbf{x}_k|\mathbf{X}^{k-1}, \mathbf{Z}^k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{30}$$

and use the *hard constrained likelihood*:

$$p(\mathcal{C}_k|\mathbf{x}_k) = \frac{p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k)}{p(\mathbf{x}_k|\mathbf{x}_{k-1})} = \begin{cases} 1 & \mathbf{x}_k \in \mathcal{C}_k \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

which yields for the evaluation of the weights:

$$\tilde{w}_k^{PS} = w_{k-1}^{PS} \, p(\mathbf{z}_k|\mathbf{x}_k) \, p(\mathcal{C}_k|\mathbf{x}_k) \tag{32}$$

which for the subset of particles verifying the constraints coincides with eq. (29). Hence, in the case of *hard* constraints and for a sufficiently large number of particles, the two methods provide the same results, thus allowing us to interpret the *Pseudo-Measurements PF* as a practically feasible alternative to the *Rejection-Sampling PF*.
A different reasoning can be followed in order to verify the correctness of the *Pseudo-Measurements PF*. The constraints in eq. (12) affect the target dynamics from a physical viewpoint.

Algorithm 2: Pseudo-Measurements Particle Filter

Input: $\left\{\mathbf{x}_{k-1}^i, w_{k-1}^i\right\}_{i=1}^N$ and the new measurement $\mathbf{z}_k$

Output: $\left\{\mathbf{x}_k^i, w_k^i\right\}_{i=1}^N$

while $i = 1, 2, \ldots, N$ ***(Prediction Step)*** **do**
  Generate a New Particle: $\mathbf{x}_k^i \sim p_k(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)$
end

while $i = 1, 2, \ldots, N$ ***(Update Step)*** **do**
  Compute Weights: $\tilde{w}_k^i = w_{k-1}^i \, p(\mathbf{z}_k|\mathbf{x}_k^i) \, p(\mathcal{C}_k|\mathbf{x}_k)$;
end

Normalization Step: $w_k^i = \tilde{w}_k^i / \sum_{i=1}^N \tilde{w}_k^i \quad \forall \, i$;

Effective Sample Size: $N_{eff} = 1/\sum_{i=1}^N (w_k^i)^2$;

if $N_{eff} \leq \beta N$ ***(Resampling Step)*** **then**
  New Particles $\left\{\tilde{\mathbf{x}}_k^i, 1/N\right\}_{i=1}^N$ s.t. $P(\tilde{\mathbf{x}}_k^i = \mathbf{x}_k^i) = w_k^i$
end

Hence, exploitation of such knowledge in the prediction step as described in section 3.1 is Bayes optimal if we can evaluate the *constrained Kernel* $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k)$. In addition, the convergence results of PFs are not affected by the chosen importance function $q(\mathbf{x}_k|\mathbf{X}^{k-1}, \mathbf{Z}^k)$ as long as the weights are re-scaled using eq. (27). Hence, we can use the *unconstrained transition Kernel* as importance function and evaluate the weights as:

$$\tilde{w}_k = w_{k-1} \frac{p(\mathbf{z}_k|\mathbf{x}_k)\, p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathcal{C}_k)}{p(\mathbf{x}_k|\mathbf{x}_{k-1})} \tag{33}$$

which exactly coincides with the equation used by Algorithm 2, thus proving again the correctness of the *Pseudo-Measurements PF* in the case of *hard constraints*.

5 Simulations

We are interested in verifying the practical equivalence between the *Pseudo-Measurements* and the *Rejection-Sampling* approaches. Despite the fact that particle filters are oftentimes used to approximate the MV estimate of eq. (7) or the MAP estimate of eq. (8), the use of point estimators in comparing particle filters might lead to completely wrong results. As described in [5], evaluation of the Kullback-Leibler Divergence (KLD) between the empirical distributions from the filters is suggested in this case. In fact, convergence towards zero of the KLD for an increasing number of particles will verify in practice the equivalence of the two techniques.
Let us briefly recall the KLD. Let a and b be two continuous densities on $\mathbb{R}^d$. The Kullback-Leibler Divergence $D_{KL}(a, b)$ between a and b is given by:

$$D_{KL}(a, b) \triangleq \mathbf{E}_p \left[\log \frac{a}{b}\right] = \int a(\mathbf{x}) \, \log \frac{a(\mathbf{x})}{b(\mathbf{x})} \, d\mathbf{x} \tag{34}$$

when support$(b) \subseteq$ support(a), otherwise $D_{KL}(a, b) = +\infty$. The KLD is always positive and equal to zero if and only if the two densities coincide. Hence, the KLD is appropriate to evaluate the closeness of a density to another. Let $\{\mathbf{x}^1, \mathbf{x}^2, \dots \mathbf{x}^n\}$ and $\{\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^m\}$ be i.i.d. samples drawn from a and b, respectively. An asymptotically unbiased and mean square consistent estimator for the KLD was introduced in [13]:

$$\hat{D}_{KL}(a, b) = \frac{d}{n} \sum_{i=1}^{n} \log \frac{\nu_{k_i}(i)}{\rho_{l_i}(i)} + \\ + \frac{1}{n} \sum_{i=1}^{n} (\psi(l_i) - \psi(k_i)) + \log \frac{m}{n-1} \tag{35}$$

where $\nu_{k_i}(i)$ is the Euclidian distance between $\mathbf{x}^i$ and its k_i-nearest neighbor in $\{\mathbf{y}^j\}$, $\rho_{l_i}(i)$ is the Euclidian distance between $\mathbf{x}^i$ and its l_i-nearest neighbor in $\{\mathbf{x}^j\}_{j \neq i}$, and ψ is the Digamma function. In [5] the estimator of eq. (35) is proven to be effective in high dimensional problems and when discrimination of close distributions is of interest.
We consider a simple 2D tracking problem as the one depicted in fig. 1, where a ship is traveling inside a known shipping lane. The chosen state vector is $\mathbf{x}_k = \begin{bmatrix} x_k & y_k & \dot{x}_k & \dot{y}_k \end{bmatrix}$,

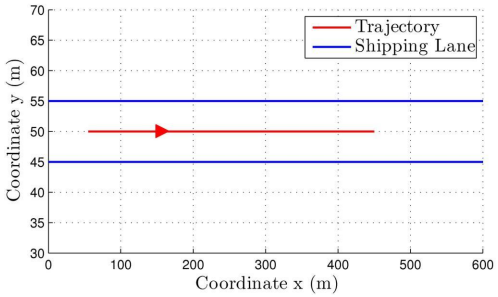

Figure 1: Scenario Used in Simulation

where (x_k, y_k) and $(\dot{x}_k, \dot{y}_k)$ are position and velocity vectors, respectively. A straight line Nearly Constant Velocity (NCV) model is used for the dynamics:

$$\mathbf{x}_k = \mathbf{A}\, \mathbf{x}_{k-1} + \mathbf{w}_k \tag{36}$$

where $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q})$ is zero-mean Gaussian process noise, and $\mathbf{A}$ and $\mathbf{Q}$ are known matrices [2].
A radar positioned at the Cartesian origin collects measurements of range, azimuth, and range rate. Hence, the nonlinear measurement function $\mathbf{h}_k(\cdot)$ in (2) takes the form:

$$\mathbf{h}_k(\mathbf{x}_k) \triangleq \begin{bmatrix} \sqrt{(x_k)^2 + (y_k)^2} \\ \text{atan2}(y_k, x_k) \\ \dfrac{x_k\, \dot{x}_k + y_k\, \dot{y}_k}{\sqrt{(x_k)^2 + (y_k)^2}} \end{bmatrix} \tag{37}$$

and $\mathbf{v}_k$ is zero-mean Gaussian noise. For the standard deviations we choose $\sigma_r = 25m$, $\sigma_\theta = 1\,\text{deg}$, and $\sigma_{\dot{r}} = 1m/\,\text{sec}$ in terms of range, azimuth, and range rate, respectively. The radar sampling time is $T_s = 1\,\text{sec}$.
Knowledge on the shipping lane is available as:

$$\mathcal{C}_k \triangleq \{\mathbf{x}_k : \mathbf{x}_k \in \mathbb{R}^{n_x}, 45 \leq y_k \leq 55\} \tag{38}$$

and implies the following knowledge-based likelihood:

$$p(\mathcal{C}_k|\mathbf{x}_k^i) = \begin{cases} 1, & \text{if } 45m \leq y_k \leq 55m \\ 0, & \text{otherwise} \end{cases} \tag{39}$$

In fig. 2 we report the results obtained over 100 Monte Carlo trials for an increasing number of particles. We report the Mean Value and the Variance for the estimated Kullback-Leibler Divergence for the two constrained PFs of Algorithms 1 and 2. Notice that since the two methods coincide (i.e., proven in section 3) the true Kullback-Leibler Divergence is 0 for an infinite number of particles. From fig. 2 it is possible to verify the convergence towards 0 for both the Mean and the Variance of the estimated KLD. This is important since it verifies the equivalence of the two methods from a practical viewpoint. However, a sufficiently large number of particles is required to achieve satisfactory low values of the KLD.

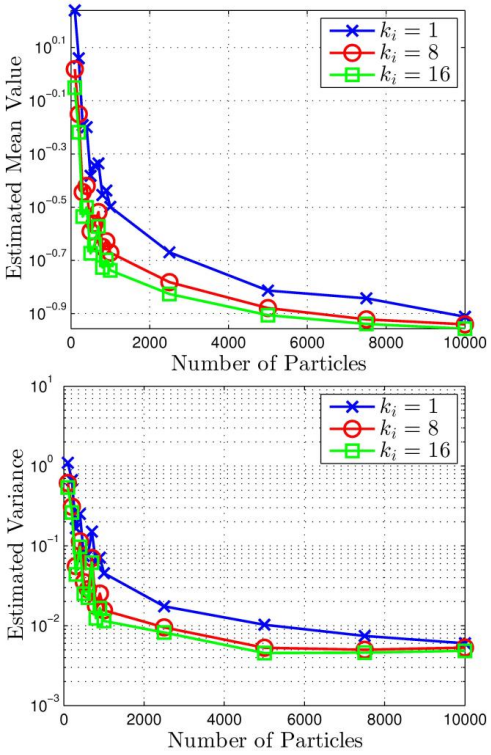

Figure 2: Mean Value and Variance of the estimated Kullback-Leibler Divergence over 100 Monte Carlo runs

6 Conclusion

This paper addresses the *Constrained Bayesian Filtering* problem for the case of perfectly known *hard* constraints. Oftentimes additional information about the state is available and can be formalized in terms of constraints on the state variables. Thus, Bayes optimal exploitation of external knowledge is achievable if a Particle Filter is used to solve the tracking. Furthermore, we showed that from a Bayesian viewpoint, exploitation of external knowledge in the prediction or in the update step of the filtering recursion are equivalent.

Two particle filtering based algorithms for *Constrained Filtering* are described and tested through simulations. By using an estimate of the Kullback-Leibler divergence, we numerically verified that as the number of particles increases, the two algorithms estimate the same a posteriori distribution, thus allowing us to interpret the *Pseudo-Measurements PF* as a practically feasible alternative to the *Rejection-Sampling PF*.

For future work we plan to apply *Constrained Filtering* to improve the tracking performance, and to extend the tracker functionalities in order to detect constraints violation.

7 Acknowledgment

The research leading to these results has received funding from the EUs Seventh Framework Programme under grant agreement n^o 238710. The research has been carried out in the MC IMPULSE project: https://mcimpulse.isy.liu.se

References

[1] D. Angelova and L. Mihaylova. Sequential monte carlo algorithms for joint target tracking and classification using kinematic radar information. *Proc. of the 7th Int. Conf. on Information Fusion*, Stockholm, Sweden, July 2004.

[2] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley and Sons, Inc., 2001.

[3] M. Bocquel, A. Lepoutre, O. Rabaste, and F. Le Gland. Optimisation of a particle filter for track-before-detect. *Proc. of the GRETSI 2011*, Bordeaux, France 2011.

[4] S. Challa and N. Bergman. Target tracking incorporating flight envelope information. *Proc. of the 3th Int. Conf. on Information Fusion, FUSION 2000*, 2:THC2/22–THC2/27, Paris, France, July 2000.

[5] R. Chou, M. Geist, M. Podt, and Y. Boers. Performance evaluation for particle filters. *Proc. of the 14th Int. Conf. on Information Fusion*, Chicago, IL, USA 2011.

[6] A. Doucet and D. Crisan. A survey of convergence results on particle filtering methods. *IEEE Transactions on Singal Processing*, 50(3):736–746, 2002.

[7] N. Gordon, S. Maskell, T. Kirubarajan, and M. Malick. Littoral tracking using particle filter. *Proc. of the 5th Int. Conf. on Information Fusion*, Annapolis, USA, July 2002.

[8] N. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F on Radar and Signal Processing*, 140(2):107113, 1993.

[9] F. Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Transaction on Aerospace and Electornic Systems*, 25(7):53–81, 2010.

[10] S. Julier and J. Uhlmann. Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401–422, 2004.

[11] W. Koch and M. Ulmke. Road-map assisted ground moving target tracking. *IEEE Transaction on Aerospace and Electornic Systems*, 42(4):1264–1274, 2006.

[12] M. Tahk and J. Speyer. Target tracking problems subject to kinematic constraints. *IEEE Transaction on Automatic Control*, 35(3):324–326, 1990.

[13] Q. Wang, S.R. Kulkarni, and S. Verd. Divergence estimation for multidimensional densities via k-nearest-neighbor distances. *IEEE Transactions on on Information Theory*, 55(5), 2009.

[14] C. Yang, M. Bakich, and E. Blasch. Nonlinear constrained tracking of targets on roads. *Proc. of the 8th Int. Conf. on Information Fusion*, pages 235–242, Philadelphia, USA, July 2005.