Wasserstein Distance for the Fusion of Multisensor Multitarget Particle Filter Clouds

Daniel Danu, Thia Kirubarajan
Department of Electrical and Computer Engineering,
McMaster University, Canada.
danudg@univmail.cis.mcmaster.ca,
kiruba@mcmaster.ca

Abstract — In a multisensor multitarget tracking application, the evaluation of the cost of assigning particle filter clouds of different sensors as being estimates of the same target is an essential part in the particle cloud association. This paper treats the problem of evaluating the cost of particle filter clouds association based on the Wasserstein distance of different orders, analyzing the implications of clouds cardinality (for weighted particles), and of various resampling methods (for unweighted particles). As the Wasserstein distance at cloud level needs to have defined internally a metric at the particle level, the implications of using therein the Euclidean (for position components only) or Mahalanobis (including higher order components) distances are investigated. The cross-covariance of particle filter clouds is also estimated using the same Wasserstein distance and its introduction in the metric therein is explored. As a conclusion of various simulations, the design of the Wasserstein distance that is found to fit best the purpose of cloud-to-cloud association is presented.

Keywords: Tracking, particle filters, data association, data fusion, random probability measure.

1 Introduction

In the context of particle filters usage in multisensor target tracking, the fusion of clouds obtained on one sensor with the clouds obtained on other sensor is believed to be a step towards obtaining improved target states estimates [4]. At the first stage of the fusion process, the association of each sensor estimated targets is performed. One issue in estimating the distance between two particle filter clouds is that usually two particle clouds pertaining to the same true target have different support on the two sensors on which they are estimated. The Wasserstein distance has the advantage of not requiring the two probability distributions of which closeness is measured to have the same support, as it provides a “transportation plane” between the two clouds through the “transportation matrix” it estimates as well. Non-Gaussian particle clouds contain more information than the mean and covariance, therefore using more of this information in the association process is desirable. In [4] we proposed the Wasserstein distance for usage in computing the particle clouds association cost. In this paper properties of the Wasserstein distance as a metric for evaluating the closeness of particle clouds are explored. To our knowledge, this is the first analysis (beside our brief application in [4]) of the Wasserstein distance w.r.t particle filter clouds. The case when the particle clouds are, or can be approximated, by Gaussian distributions is also treated. As stated in [2], the particle clouds represent random probability measures of the target posterior deterministic probability density The weak convergence of the particle clouds as random probability measures towards the posterior density of the target was shown and analyzed in [2], [5]. The convergence of the Wasserstein distance estimate on particle filter clouds (discrete random measures) towards the distance between the underlying deterministic (unknown) posterior densities of the targets is demonstrated in this paper for the internal Euclidean distance used between particles. The effect of particle cloud cardinality on the distance, as well as of resampling methods is explored. The effect of using at particle level the Euclidean or Mahalanobis distance in the estimation of the Wasserstein distance is investigated. In order to apply the distance between two clouds, the particles need to be previously labeled or at least partitioned in clouds. There are several labeled particle filters proposed in the literature ([7], [8]). The transportation matrix can be used also for other purposes, as shown in Section 6.

2 Particle filter clouds as probability measures

In the context of this paper a particle cloud is a cluster of particles of the same label, representing the posterior probability distribution of a target state estimated by a particle filter [2], [5]. Each particle in a cloud is defined by its state $\xi$, weight $w_i$, and label $l_i$. The state indicates the particle location in the estimating space (which could consist of position, velocity in one, two, three dimensions, etc), the weight indicates the estimated probability of the target existence at the location given by the state in the state space and the label indicates the identity of the target estimated in a multitarget scenario. We use the notations $\Xi_s^{N,l} = \{\xi_s^j, w_s^j, I_s^j\}_{j=1}^N$ for a particle cloud of cardinality $N$ and label $l_i$ ($i = 1, \ldots, L$) on sensor $s$ ($s = 1, \ldots, S$), $\Xi_s^{N} = \{\xi_s^j,w_s^j\}_{j=1}^N$ when the label is not required, and $\Xi^{N} = \{\xi^j,w^j\}_{j=1}^N$ when the sensor information is not needed.
In the space defined by the particle state \( \xi \in \mathbb{R}^{n_x} \), where \( n_x \) is the size of the target state \( x \), the mass distribution of a labeled particle cloud is

\[
\Xi^N = \sum_{i=1}^{N} w_i \delta_{\xi_i}(\xi)
\]  

(1)

Equation (1) is a Monte Carlo approximation of the posterior probability distribution of the target state [2]. Given the randomness of the particle sampling (and resampling) processes, the particle cloud is a random measure which approximates this posterior [2]. If the same observation process \( Y_{0:t} \) (with different sample is used on different nodes (e.g. identical sensors, with same measurement noise on all sensors), then the \( \sigma \)-algebra generated by the observation process on all nodes is the same. If identical dynamic models are used on all sensors for a given target, the posterior density of the state conditioned on the \( \sigma \)-algebra of measurements is identical on all sensors [2]:

\[
\pi^y_{1:0} = ... = \pi^y_{S:0} \triangleq P(x_t \in A|\sigma(Y_{0:t}))
\]  

(2)

with \( A \in B(\mathbb{R}^{n_x}) \) for sensors \( s = 1 \ldots S \) and with the posterior densities \( \pi^y_{1:0} \) random probability measures. The posterior densities of the state conditioned on the observation samples available at each sensor are different at each estimating node:

\[
\pi^y_{1:0} \triangleq P(x_t \in A|y_{1:0}) \neq ... \neq \pi^y_{S:0} \triangleq P(x_t \in A|y_{S:0})
\]  

(3)

and they are deterministic probability measures (probability density functions). The cloud-to-cloud association process therefore is an association performed between probability measures. It could be performed on either one of the measures, random or deterministic, in the former case associating identical measures on each sensor for a target, while in the latter one associating for a given target the closest probability measures on each sensor. The choice of associating the deterministic probability measures in (3) is followed in this paper. Given the deterministic density of the target in (3) at one sensor, the particle filter cloud is a random probability measure that approximates it. In [2] it was shown that as a function of \( N \), \( \Xi^N \) is a sequence of random measures that converge weakly to the deterministic one:

\[
\lim_{N \to \infty} \Xi^N \to \pi
\]  

(4)

Therefore the problem of associating particle clouds is one of associating random probability measures and the association cost needs to be one that penalizes the dissimilarity between these measures.

## 2.1 Distance between probability measures

There are several distances between probability measures described in the literature, however most of them require a common support for the two measures. Some that do not are listed below.

(i) Levy-Prohorov distance between probabilities \( P_1, P_2 \):

\[
d^{LP}(P_1, P_2) = \inf \left\{ \epsilon | \forall B \in \mathcal{F}, P_1(B) \leq P_1(B^\epsilon) + \epsilon, P_2(B) \leq P_1(B^\epsilon) + \epsilon \right\}
\]

where \( B^\epsilon \) is the \( \epsilon \)-neighbourhood of the set \( B \) and \( \mathcal{F} \) is the \( \sigma \)-algebra on which \( P_1, P_2 \) are defined. This distance is used mostly for theoretical expositions and it is difficult to compute.

(ii) Haussdorf distance can be used between different sets, as is the case of particle clouds:

\[
d^H(P_1, P_2) = \max \left\{ \inf \{ \alpha > 0 | A^\alpha \supset B \}, \inf \{ \alpha > 0 | B^\alpha \supset A \} \right\}
\]

In the Haussdorf distance the boundary particles have the most impact while their internal distribution and weight is not taken into account, making it poorly suited for characterizing the similarity of discrete distributions.

(iii) Wasserstein distance, described next.

## 3 Wasserstein distance

The Wasserstein distance was introduced by the Russian statistician L. N. Vasershtein [12] in the paper [13] as:

\[
\ell_p(P_1, P_2) = \inf_{X \sim P_1, Y \sim P_2} \mathbb{E}[d(X, Y)^p]
\]

The German spelling was maintained in the literature, where it is also found under the names Kantorovich, Mallows, Earth Mover’s Distance, Kantorovich-Rubinstein, depending on its order \( p \) or domain of applications. (Vasershtein introduced the distance using \( p = 1 \), therefore the powers above are not explicitly visible).

The Wasserstein distance was introduced in the tracking literature through [3], for performance assessment, as the multitarget miss distance. In its statistical definition, it defines the dissimilarity between two probability densities \( f, g \), using the distance \( d \) (e.g. Euclidean, Mahalanobis) and the joint distribution \( h(\cdot, \cdot) \), whose marginals are the two densities \( \int h(\xi, \zeta) d\xi = g(\zeta), \int h(\xi, \zeta) d\zeta = g(\xi) \) [3]:
\[ d_p^W (f, g) = \inf_{h} \sqrt{\int \int d (\xi, \zeta)^p h (\xi, \zeta) d\xi d\zeta} \]  (5)

Equation (5) satisfies the three properties of a metric: distance is zero for (and only for) identical measures, symmetry, and triangular inequality.

In a mathematical formulation the Wasserstein distance defines the weak topology on the set of probability measures on a given space, and for dense probability measures on a given space, it defines a separable and complete metric [10].

The Wasserstein distance between the distributions \( P_1, P_2 \) of two random variables (r.v.) \( x_1, x_2 \in \mathbb{R}^n \) of means \( m_1, m_2 \) is always greater than the Euclidean distance between their means, as we have [11], corrected here for a typo in [11], as we believe the \( d_\mathcal{W}^2 (Q_1, Q_2) \) should enter the equation at the power of two:

\[ d_\mathcal{W}^2 (P_1, P_2) = \sqrt{\| m_1 - m_2 \|^2 + d_\mathcal{W}^2 (Q_1, Q_2)^2} \]  (6)

where \( Q_1, Q_2 \) are the distributions of r.v. \( x_1 - m_1 \) and \( x_2 - m_2 \) respectively. Therefore Wasserstein distance between r.v. that have the same mean can be non-zero. This could be useful in distinguishing between probability distributions that pertain to targets compared to probability distributions that pertain to clutter or other false alarms, if some information on the latter is known.

### 3.1 Wasserstein distance for particle clouds

For particle clouds \( \Xi_1, \Xi_2 \) the Wasserstein distance was introduced in [4] as:

\[ d_\mathcal{W}^2 (\Xi_1, \Xi_2) = \inf_{w} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w_{ij} d (\xi_i, \xi_j)^p \]  (7)

and having \( h \) defined as

\[ h(\xi, \zeta) = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w_{ij} \delta_{\xi_i} (\xi) \delta_{\zeta_j} (\zeta) \]  (8)

where \( w_{ij} \) are elements (taking values in the interval \( [0, 1] \) ) of the transportation matrix \( W \) between the two clouds and \( d \) is the distance established at particle level (e.g. Euclidian, Mahalanobis). One advantage of using the Wasserstein distance as a measure of closeness between particle clouds is that it does not require for the two probability densities to have the same support. The “transportation plane” defined by the matrix \( W \) between the two clouds through the matrix \( W \) makes this possible.

The transportation matrix \( W \) is similar to the Earth’s Movers Distance (EMD) defining the minimal work necessary to move the particles of one cloud into the other allowing transportation of fractional particles. Therefore, through the matrix \( W \), “fractional transports” are allowed, i.e. a particle from the source cloud can be transported (contribute) to several particles of the destination cloud, and several particles from the source cloud can contribute with different weights in building the weight of a particle in the destination cloud. For weighted particles, their weights are accounted for by the transportation matrix \( W \). This transportation matrix between the two clouds can be estimated through linear programming methods.

Comparing the definition of the Wasserstein distance in (2) with its application to particle clouds in (7) it can be seen that the joint probability density \( h (\xi, \zeta) \) is replaced by the transportation matrix \( W \).

\[ h \leftrightarrow W, \quad h (\xi_i, \zeta_j) \leftrightarrow w_{ij} \]  (9)

For example, given a particle cloud \( \Xi^N \) of \( N \) particles and a particle cloud \( \Xi^1 \) of just one particle of weight 1, the \( d_\mathcal{W}^2 \) between them is given by

\[ d_\mathcal{W}^2 (\Xi^N, \Xi^1) = \sqrt{\sum_{i=1}^{N} w_i \| \xi_i - x_0 \|^2} \]  (10)

The minimum (and unique) \( d_\mathcal{W}^2 \) distance between such clouds (with one degenerated - contracted - into a single particle) is obtained if \( \Xi^1 = \{ x_0 \} \) is positioned on the center of mass of \( \Xi^N \). The square value of this distance is equal to the variance of cloud \( \Xi^N \):

\[ \min_{x_0} d_\mathcal{W}^2 (\Xi^N, \Xi^1)^2 = \sigma^2 (\Xi^N) \]  (11)

\[ w_{ij} \sim p (\xi_i, \zeta_j | T) \]  (12)

### 3.2 Association cost of labeled particle clouds

For two sensors, \( s = 1, 2 \) we denote by \( \text{cost}(\Xi^{M_{1,s}}, \Xi^{M_{2,s}}) \) the cost of the hypothesis that clouds \( \Xi^{M_{1,s}}, \Xi^{M_{2,s}} \) are random probability measures of the posterior of the same target on both sensors. Labels \( l_1 \) and \( l_2 \) are the cloud labels on sensors 1 and 2, respectively and \( M_1, M_2 \) are their cardinalities. The cost based on the Wasserstein distance defined above for the clouds is
\[ \text{cost} \left( \Xi_{1}^{M,ilk}, \Xi_{2}^{M,ilk} \right) = \min_{\Xi} \frac{1}{N} \sum_{i=1}^{M} \sum_{j=1}^{M} w_{ij} d \left( \xi_{ij}, \xi_{ij}^{*} \right) \] (13)

subject to the constraints

\[ \sum_{i=1}^{M} w_{ij} = w_{ij}^{*}, \quad j = 1..M_{2}, \quad \sum_{j=1}^{M} w_{ij} = w_{i}, \quad i = 1..M_{1} \] (14)

Constraints above assure each particle in one cloud is fully assigned (using their full weights) to particles in the other cloud. [4].

For different \( p \) values, \( p \geq 1 \), \( d_{p}^{w} \) define equivalent topologies for probability distributions on the Euclidean space [10]. This means that for the same inner distance used between particles, \( d \left( \xi_{i}, \zeta_{i} \right) \), the ordering of \( d_{p}^{w} \) distances (association costs) between particle clouds is independent of \( p \). Therefore the associations resulting from using different orders in \( d_{p}^{w} \) are identical.

### 3.3 Convergence of the Wasserstein distance on particle clouds

The convergence of the particle filter as a sequence of random probability measures toward the deterministic posterior probability distribution of the target state was studied in [2], [5] w.r.t. number of particles and the resampling methods used. In this section the convergence of the Wasserstein distance estimated on particle clouds towards the Wasserstein distance between the deterministic (continuous) measurement-conditioned posterior densities is explored. The theoretical proof of convergence is given and the convergence is demonstrated empirically in the Section 5 as a function of

- number of weighted particles
- resampling method and number of particles
- internal distance used (Euclidean, Mahalanobis).

#### 3.3.1 Convergence of Euclidean Wasserstein distance on particle clouds

In [2], [5] it was shown that the mean square error of the particle filter estimator converges with a rate proportional to \( 1/N \) and that it is independent of the state dimension. Here the convergence of the Wasserstein distance between two particle filter clouds is explored.

Three properties of the Euclidean Wasserstein distance are listed below [6]:

P.1. For a sequence of (random) probability functions (measures) \( f_{n} \) that converges weakly to \( f \), the distance between \( f_{n} \) and \( f \) converges to zero:

\[ d_{2}^{w} \left( f_{n}, f \right) \to 0 \]

P.2. For two sequences of random measures \( f_{n}, g_{n} \), which converge weakly to two deterministic finite measures, \( f_{n} \to f \) and \( g_{n} \to g \), we have:

\[ d_{2}^{w} \left( f, g \right) \leq \lim \inf_{n \to \infty} d_{2}^{w} \left( f_{n}, g_{n} \right) \]

P.3. Triangle inequality:

\[ d_{2}^{w} \left( f, g \right) \leq d_{2}^{w} \left( f, h \right) + d_{2}^{w} \left( h, g \right) \]

The sequences of two particle filter clouds \( \Xi_{1}^{n}, \Xi_{2}^{n} \) of cardinalities \( n \), which are random probability measures, converge in mean square error to the posterior distribution of the estimated target [5]. Given the two independent sequences of measurements of the same target on the two sensors, the deterministic posteriors of the target on the two sensors are different. If we denote them by \( \pi_{1}, \pi_{2} \), based on P.1. we have

\[ \lim_{n \to \infty} d_{2}^{w} \left( \Xi_{1}, \pi_{1} \right) \to 0, \quad \lim_{n \to \infty} d_{2}^{w} \left( \Xi_{2}, \pi_{2} \right) \to 0 \] (17)

Using twice the triangle inequality (16) we get

\[ d_{2}^{w} \left( \Xi_{1}^{n}, \Xi_{2}^{n} \right) \leq d_{2}^{w} \left( \Xi_{1}^{n}, \pi_{2} \right) + d_{2}^{w} \left( \Xi_{2}^{n}, \pi_{2} \right) \]

\[ \leq d_{2}^{w} \left( \Xi_{1}^{n}, \pi_{1} \right) + d_{2}^{w} \left( \Xi_{2}^{n}, \pi_{2} \right) + d_{2}^{w} \left( \pi_{1}, \pi_{2} \right) \] (18)

therefore

\[ d_{2}^{w} \left( \Xi_{1}^{n}, \Xi_{2}^{n} \right) \leq \varepsilon(n) + \varepsilon(n) + d_{2}^{w} \left( \pi_{2}, \pi_{1} \right) \] (19)

with \( \lim_{n \to \infty} \varepsilon(n) = 0, \lim_{n \to \infty} \varepsilon(n) = 0 \). From (15) we have for the particle clouds \( \Xi_{1}^{n}, \Xi_{2}^{n} \)

\[ \lim \inf_{n \to \infty} d_{2}^{w} \left( \Xi_{1}^{n}, \Xi_{2}^{n} \right) \geq d_{2}^{w} \left( \pi_{1}, \pi_{2} \right) \] (20)

and combining (19) and (20) at the limit it results

\[ \lim_{n \to \infty} d_{2}^{w} \left( \Xi_{1}^{n}, \Xi_{2}^{n} \right) \to d_{2}^{w} \left( \pi_{2}, \pi_{1} \right) \] (21)

Therefore, for particle filter clouds converging to the posterior distribution of the targets, the Wasserstein distance
of order two converges to the one (of same order two) between the conditional posteriors of the target distribution.

### 4 Computation of the Wasserstein distance for particle clouds

The calculation of cloud-to-cloud association cost using Wasserstein distance as defined in (6)-(7) was carried out in [4] by linear programming (interior-point) methods. It becomes very intensive when the cardinality of the two clouds is high. Therefore, the implications of using in its computation a restricted number of highest weighted particles, or resampled clouds of lesser cardinalities are explored in following subsections. Also the resolution obtained in the Wasserstein distance by using the Euclidean or Mahalanobis distance at particle level is experimented.

#### 4.1 Gaussian particle clouds

For the particular case of particle clouds of Gaussian probability distributions, the Wasserstein distance of order 2 has a closed form expression [11]. Given two probability distributions $P_1, P_2$ of random variables in $\mathbb{R}^n$ with means $m_1, m_2 \in \mathbb{R}^n$ and covariances $P_1$ and $P_2 \in \mathbb{R}^{n \times n}$, the closed form expression is:

$$d_W^2(P_1, P_2) = \sqrt{\|m_1 - m_2\|^2 + tr(P_1 + trP_2 - 2tr(\sqrt{P_1 P_2})))^{1/2}}$$

(22)

Based on (6) and the derivation for distance of mean zero distributions in [11], we have for the special cases:

**n = 2:**

$$d_W^2(P_1, P_2) = \sqrt{\|m_1 - m_2\|^2 + tr(P_1 + trP_2 - 2[tr(P_1 P_2) + 2\sqrt{\det(P_1 P_2)})])^{1/2}}$$

(23)

and **n = 1:**

$$d_W^2(P_1, P_2) = \sqrt{\|m_1 - m_2\|^2 + tr(P_1 + trP_2 - 2\sqrt{\det(P_1 P_2)})}$$

(24)

In the remainder of Section 4, is presented the program of empirical studies through simulation to characterize the performance of the Wasserstein distance under various conditions. The results of these simulations are presented in Section 5.

#### 4.2 Restricted number of particles

The convergence in (21) is empirically explored, using Monte Carlo simulations for computation on a restricted number $N_0$ of weighted (of highest $N_0$ weights) particles in the cloud of cardinality $N \geq N_0$. Once the $N_0$ particles are selected in both clouds, their weights are rescaled to restore their total probability mass to unity.

#### 4.3 Resampled clouds

The result in (21) is empirically explored, using Monte Carlo simulations for different values of $N_r$ resampled (unweighted) particles using the multinomial and the stratified resampling methods, briefly described next.

##### 4.3.1 Multinomial resampling

A description of multinomial resampling can be found in [2], [14].

##### 4.3.2 Stratified resampling

Stratified resampling is a combination of deterministic and random methods and its description can be found in [14].

#### 4.4 Mahalanobis Wasserstein distance for particle clouds

The Mahalanobis Wasserstein distance allows the usage of all state components combined into the computation of the distance. It implies the computation of the sample covariance of each cloud:

$$\tilde{P} = \sum_{n=1}^{N} (\xi^n - \tilde{m})(\xi^n - \tilde{m})', \quad \tilde{m} = \sum_{n=1}^{N} \xi^n w^n$$

(25)

and based on it

$$d_{\text{Mah}}^W(\Xi^n, \Xi^\circ)$$

$$= \min_{\Pi} \left[ \sum_{j=1}^{N_r} \sum_{l=1}^{N_r} w_{jl} (\xi_{j,l} - \xi_{l,j})' \tilde{P}^{-1} (\xi_{j,l} - \xi_{l,j}) \right]$$

(26)

#### 4.5 Sample cross-covariance estimation using the Wasserstein distance

As the Wasserstein distance implies the computation of a joint distribution between clouds, this could be used in the estimation of the sample cross-covariance as below:

$$\tilde{P}_{12} = \sum_{j=1}^{N_1} \sum_{l=1}^{N_2} w_{1j} w_{2l} (\xi^n_{j,l} - \tilde{m}_1)(\xi^n_{j,l} - \tilde{m}_2)'$$

(27)
Simulations

For a challenging association, two trajectories of close targets T1, T2, that cross each other, and both having a high non-linear motion component, introduced by elevated process noise, were considered for simulation (Figure 1).

The first target starts at \( x_{01} = (-63.1, 40) \) m with a velocity \( v_{01} = (3.8, -1.8) \) m/s, and the second starts at \( x_{02} = (-35.1, 75) \) m with a velocity of \( v_{02} = (2.1, -3.1) \) m/s. Targets motion equations are simulated using

\[
x_k = F x_{k-1} + G v_k
\]

with

\[
F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} \frac{1}{2} T^2 & 0 \\ 1 & 0 \\ 0 & \frac{1}{2} T^2 \\ 0 & 0 & 1 \end{bmatrix}
\]

(28)

and with process noise standard deviation \( \sigma_v = [1.0, 0.1]^T \).

As shown in Figure 2, the crossings in x and y components happen slightly shifted in time (at around 5 sec. difference, therefore targets are distinct at all times). Two sensors (estimation nodes) provide particle clouds as estimates of targets states, each running the track-labeled probability hypothesis density filter (TLPHD), following the model in [8], here slightly modified to account for 2D motion (as in [4]) and multiple measurements falling within the region of an estimated target. Measurements on each of the two sensors (estimation nodes) are simulated at 1 sec. intervals following the equation

\[
y_k = H x_k + w_k
\]

(29)

with \( H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T \) and equal measurement noise standard deviation \( \sigma_w = [2.5, 2.5]^T \).

False measurements are Poisson distributed for each sensor with a rate of four per scan. Estimated tracks (including false ones) at each sensor for a sample run are shown in Figure 3.

For comparison, the association costs of clouds were computed using (i) \( d^E \) - Euclidean distance between clouds means (denoted \( d^E \) here), (ii) \( d^W_1 \) - Wasserstein distance of order 1, (iii) \( d^W_2 \) - Wasserstein distance of order 2, (iv) \( d^{\text{Mah}} \) - Wasserstein distance using Mahalanobis distance at particle level, and the corresponding ones on resampled particles: (v) \( d^E_r \), (vi) \( d^W_1 \), (vii) \( d^W_2 \), (viii) \( d^{\text{Mah}} \). Some sample variation of the distances above based on the type
(weighted or resampled) and number of particles used are presented in Figures 4-5 below.

Figure 4 Sample variation of distances vs. number of particles N at time k = 14 for clouds of matching targets.

Figure 5. Sample variation of distances vs number of particles N at time k = 16 for clouds of matching targets.

Figure 4 shows the Wasserstein distance between sample clouds (estimates on different sensors of the same target) on which all distances converge, either computed on resampled or unresampled particles, however the ones using resampled particles are closer to the limit for lower number of particles (Nr ~ 100). Figure 5 shows that in sample cases where the distance is computed based on the Mahalanobis distance at resampled particle level, one needs $N_r \geq 300$ in order to get a value close to convergence. In all cases, the distance computed on resampled particles and based on the Mahalanobis distance on resampled particles shows a steadiness for $N_r \geq 100$. In all cases the Wasserstein distances were computed using the linear program defined as in [4], of which solutions were obtained with the LIPSOL package [15]. One interesting fact is that times obtained for solving the linear program using resampled particles were significantly smaller than the ones for weighted particles (comparing the same cardinalities), especially for higher numbers of particles. Sample times are presented in Table 1 and Table 2 below.

Table 1. Times obtained for different Wasserstein distances on weighted particles.

<table>
<thead>
<tr>
<th>N0</th>
<th>$d_1^w$</th>
<th>$d_2^w$</th>
<th>$d_{Mah}^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.703</td>
<td>0.734</td>
<td>0.812</td>
</tr>
<tr>
<td>100</td>
<td>1.766</td>
<td>1.625</td>
<td>1.719</td>
</tr>
<tr>
<td>200</td>
<td>8.922</td>
<td>7.578</td>
<td>9.11</td>
</tr>
<tr>
<td>300</td>
<td>23.719</td>
<td>25.062</td>
<td>26.547</td>
</tr>
<tr>
<td>400</td>
<td>58.109</td>
<td>47.734</td>
<td>62</td>
</tr>
<tr>
<td>500</td>
<td>117.984</td>
<td>89.547</td>
<td>109.015</td>
</tr>
<tr>
<td>600</td>
<td>187.562</td>
<td>192.297</td>
<td>250.75</td>
</tr>
</tbody>
</table>

Table 2. Times obtained for different Wasserstein distances on resampled particles.

<table>
<thead>
<tr>
<th>N0</th>
<th>$d_1^w$</th>
<th>$d_2^w$</th>
<th>$d_{Mah}^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.719</td>
<td>0.672</td>
<td>0.719</td>
</tr>
<tr>
<td>100</td>
<td>1.438</td>
<td>1.484</td>
<td>1.547</td>
</tr>
<tr>
<td>200</td>
<td>6.219</td>
<td>5.969</td>
<td>6.156</td>
</tr>
<tr>
<td>300</td>
<td>17.875</td>
<td>18.562</td>
<td>21.485</td>
</tr>
<tr>
<td>400</td>
<td>41.171</td>
<td>44.203</td>
<td>41.687</td>
</tr>
<tr>
<td>500</td>
<td>65.859</td>
<td>77.407</td>
<td>78.468</td>
</tr>
<tr>
<td>600</td>
<td>112.922</td>
<td>138.125</td>
<td>123.719</td>
</tr>
</tbody>
</table>

6 Conclusions

The Wasserstein distance usage between particle filter clouds for target tracking was explored. The effects of particle cloud cardinality on the distance, as well as of resampling were investigated. Also the effect of using at particle level the Euclidean or Mahalanobis distance in the estimation of the Wasserstein distance was investigated. One finding from the simulations is that computing the distance on resampled particles is faster than using the same number of unresampled particles. Another finding is that the Wasserstein distance using the Mahalanobis distance is more stable (converges faster) than using the Euclidean one. Also the Mahalanobis distance enables the usage of all particle state components (e.g. usage of position, velocity, etc), a feature that the Euclidean distance does not allow. In the simulations studied, for the Wasserstein distance based on the Mahalanobis distance, using $N_0 \geq 100$ particles provided a stable association cost. Therefore, for an efficient implementation of the cloud-to-cloud association using the Wasserstein distance, the one based on the Mahalanobis distance at particle level, with a number of particles higher than 100 is proposed. The proposed
distance will be tested on more simulations with enhanced particle clouds skewness and kurtosis, to emphasize the extra information the Wasserstein distance brings compared to the Euclidian or Mahalanobis ones between the sample mean states (i.e. the extra information is brought by the second term under the square root of equation (6)).

References


