A Joint Multitarget Estimator for the Joint Target Detection and Tracking Filter

Erkan Baser*, Mike McDonald†, Thia Kirubarajan*, and Murat Efe‡

***Abstract*—This paper proposes a joint multitarget (JoM) estimator for the joint target detection and tracking (JoTT) filter. An efficient choice to the unknown JoM estimation constant (i.e., hypervolume around target state estimate) is proposed as a Pareto-optimal solution to a multi-objective nonlinear convex optimization problem. The multi-objective function is formulated as two convex objective functions in conflict. The first objective function is the information theoretic part of the problem and aims for entropy maximization, while the second one arises from the constraint in the definition of the JoM estimator and aims to improve the accuracy of the JoM estimates. The Pareto-optimal solution is obtained using the weighted sum method, where objective weights are determined as linear predictions from autoregressive models. In contrast to the marginal multitarget (MaM) estimator, the "target-present" decision from the JoM estimator depends on the spatial information as well as the cardinality information in the finite-set statistics (FISST) density. The simulation results demonstrate that the JoM estimator achieves better track management performance in terms of track confirmation latency and track maintenance than the MaM estimator for different values of detection probability. However, the proposed JoM estimator suffers from track termination latency more than the MaM estimator since the localization performance of the JoTT filter does deteriorate gradually after target termination.**

***Index Terms*—Target tracking, JoM estimator, Bernoulli RFS, JoTT filter, track management.**

I. Introduction

Target tracking is the process of estimating the state of a dynamic object by filtering noisy measurements in the presence of false alarms and missed detections. The whole process can be divided into track confirmation, track maintenance, and track termination functions. Hence, it is necessary to verify the existence of the target from the received measurements. A number of statistical algorithms have been proposed for the detection and tracking of single (or multiple) target(s) [1]. A recent innovation in the area of target detection and tracking is in the application of the Random Finite Sets (RFS) using the finite-set statistics (FISST) [2], [3].

The RFS formalism of the Bayesian multitarget filter provides a formal mechanism for propagating and updating FISST densities. Using the Almost Parallel Worlds Principle (APWOP) along with the relationship between the FISST probability and the measure theoretic probability, some statistical concepts and techniques in filtering theory and information theory can be established for the RFS formalism [2], [3]. However, the conventional single target state estimators (e.g., the maximum a posteriori (MAP) estimator and the expected a posteriori (EAP) estimator) are undefined for RFS based multitarget filters [2], [4]. Hence, two Bayesian optimal estimators were proposed to obtain the multitarget states from FISST densities. The first multitarget state estimator is called the marginal multitarget (MaM) estimator. This estimator only considers the cardinality information (i.e., the number of elements of a given RFS) in FISST densities. The second multitarget state estimator is called the joint multitarget (JoM) estimator. This estimator, as its name suggests, considers both the cardinality and spatial information related to multitarget states in FISST densities. These two estimators are Bayesian optimal, i.e., they minimize their Bayes risk functions. Recently, the minimum mean optimal sub-pattern assignment (MMOSPA) estimator in [5] was generalized for the probability hypothesis density (PHD) filter [6]. Thus, a theoretical basis also has been established for the commonly used k-means clustering method.

The multi-Bernoulli assumption on the RFS of targets represents each target independently by a parameter pair $\{q, f\}$ [8]. That is, for each target an independent Bernoulli RFS provides a unified statistical representation of target existence via the probability q and target states via the spatial probability density $f(x)$. Using the multi-Bernoulli RFS representation, tractable approximations of the multitarget Bayes filter, generally known as the multi-target multi-Bernoulli (MeMBer) filters, were developed [2], [9], [10]. In addition, the Bernoulli RFS formalism was used in the development of an exact solution to the single-target tracking problem. First, the integrated probabilistic data association (IPDA) filter [11] was formulated as an RFS based Bayes filter [12]. Then, this RFS formulation was extended by making use of a target birth model, state-dependent detection probability and arbitrary false alarm process in its framework. Thus, the joint target detection and tracking (JoTT) filter (also known as the Bernoulli filter) was developed with the objective of estimating the target existence probability along with its state(s) [2], [13]. For more detailed information regarding the theory, implementation and applications of Bernoulli filters, interested readers are referred to [8].

The performance of tracking algorithms and state estimators can be evaluated by metrics defined in terms of cardinality, time, and accuracy [14], [15]. The performance metrics should be determined according to which attributes of the tracking algorithm or the state estimator are selected to be monitored. For example, the mean OSPA (MOSPA) metric is appropriate

* The corresponding author. Address: Dept. of Electrical and Computer Engineering, McMaster University, 1280 Main St. West, Hamilton, ON L8S 4L8, Canada. E-mail: basere@mcmaster.ca
† Address: Radar Sensing and Exploitation Section, Defence Research and Development Canada, 3701 Carling Avenue, Ottawa, ON, K1A 0Z4, Canada.
‡ Address: Dept. of Electrical and Electronics Engineering, Ankara University, 50. Yil Kampusu, L Blok, Golbasi, Ankara, 06830, Turkey.

DRDC-RDDC-2015-P021

to reduce jitters and track coalescence [5], [7]. In addition, they should be consistent with the criteria that the tracking algorithm or state estimator is developed to optimize [5], [16]. Based on these facts, it is important to point out that the estimated states from the JoTT filter using the JoM estimator is identical to that using the MaM estimator if "target-present" decision is confirmed by these two estimators. Therefore, the performance metric(s) should be selected so as to monitor the cardinality and time attributes of these two estimators regarding track confirmation, track maintenance quality after the target birth, and track termination. There are numerous metrics defined in terms of cardinality and time. Nevertheless, the OSPA metric is defined as a rigorous and robust performance measure for the (multi)target Bayes filters [17], [18].

Even though the MaM estimator is used in MeMBer type filters, the exact use of the JoM estimator with these filters has not been studied so far. In this paper, we propose a JoM estimator to obtain the estimate of the target RFS from the JoTT filter. The proper choice to the unknwon JoM estimation constant (i.e., hypervolume around target state estimate) is obtained as a Pareto-optimal solution to a multi-objective nonlinear convex optimization problem. The multi-objective function is formulated as two convex objective functions in conflict. The first objective function is the information theoretic part of the problem and aims for entropy maximization, while the second one arises from the constraint in the definition of the JoM estimator and aims to improve the accuracy of the JoM estimates. The Pareto-optimal solution is obtained using the weighted sum method [19]–[22]. This method aggregates two or more objective functions into a single objective function using weights selected according to their relative importance. Then, the resulting single-objective optimization problem can be solved using any standard optimization technique [19], [20].

This paper is organized as follows: Section II provides the necessary background on information theory and multitarget state estimation. In Section III, the Bayesian optimal multitarget estimators (i.e., MaM and JoM estimators) are presented along with their evaluations for estimation of multitarget states. The proper choice to the JoM estimation constant is formulated in Section IV. For its Pareto-optimal solution, linear predictions of objective weights are proposed in Section V. The implementation of the JoM estimator for the JoTT filter under Gaussian assumptions is presented in Section VI. Simulation results are shown in Section VII. Finally, conclusions and future research directions are given in Section VIII.

II. Background

A. Concepts in Information Theory

In the following, we introduce some of the basic concepts of information theory. For the sake of completeness and clarity, we also summarize how each concept is utilized later.

Entropy: A random variable is statistically characterized by its probability density function (pdf). In traditional statistics, variance of a random variable is used to measure its uncertainty. However, in the information theoretic sense, entropy is a measure of the amount of uncertainty in a random variable [23]. For a discrete random variable x characterized by the probability mass function (pmf) $p(x)$ over its sample space $\mathcal{X}$, the entropy is computed as

$$H(p) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x)), \tag{1}$$

where $-\log(p(x))$ is called the self-information obtained by the observation of x. For the continuity of entropy, $0\log(0) = 0$, and thus zero probability does not change the uncertainty in x.

Entropy is a nonnegative measure, i.e., $H(x) \geq 0$ with the properties that $H(x)$ is maximized if $p(x)$ is uniform, and $H(x) = 0$ if there is no uncertainty in x, i.e., $p(x) = 0$ or 1 [23]. Hence, larger entropy means that less information is available for the realization of a random variable through its pmf [24].

Differential Entropy: For continuous random variables, the information theoretic uncertainty analogous to the entropy is called the differential entropy, and is defined as

$$H(f) = -\int_s f(x)\log(f(x))\,dx, \tag{2}$$

where S is the support set of the continuous pdf $f(x)$. Unlike the entropy, the differential entropy has values in the range $[-\infty, \infty]$. Therefore, its standalone value cannot be interpreted as the amount of uncertainty on a continuous time random variable. Besides, it makes sense within the definition of the following concepts.

Entropy and differential entropy will be utilized to analyze uncertainties related to the cardinality and spatial information in a FISST density, respectively. Thus, we can evaluate how appropriate the MaM and JoM estimators are for estimation of the multitarget states.

Asymptotic Equipartition Property: In information theory, the weak law of large numbers corresponds to asymptotic equipartition property (AEP) [23]. That is, given that $\tilde{x}_1, ..., \tilde{x}_n$ are independent and identically distributed (i.i.d.) random samples from $f(x)$, then the normalized self-information of this sequence weakly converges to the (differential) entropy of $f(x)$ with a small positive tolerance, i.e., $\tau > 0$ if n is large enough to satisfy [23]

$$\Pr\left(\left|-\frac{1}{n}\log f(\tilde{x}_1, ..., \tilde{x}_n) \to H(f)\right| < \tau\right) > 1-\delta, \tag{3}$$

where $\delta \to 0$ as $n \to \infty$ (proof is given by Chebyshev's inequality). The collection of these sequences forms typical set A_τ^n. Most of the total probability is contained in this set, i.e., $\Pr(A_\tau^n) > 1-\tau$ and is almost uniformly distributed [23] as

$$2^{-n(H(f)+\tau)} \leq \Pr(\tilde{x}_1, ..., \tilde{x}_n) \leq 2^{-n(H(f)-\tau)}. \tag{4}$$

Hence, if any statistical conclusion is drawn for a typical set, it would be true in general with high probability [23]. In addition, the volume of typical set is almost given by [23], [25]

$$Vol(A_\tau^n) \approx 2^{nH(f)}. \tag{5}$$

Then, the larger\smaller the differential entropy is, the more $f(x)$ disperses\concentrates over its support set S. Note that the typical set has the smallest volume, compared to all

possible sets that contain most of the total probability [23], [24].

Typical set of a standard Gaussian density will lead us to define another important set, where the sequences of mostly likely state estimates exist. Thus, our aim would be the entropy maximization by defining a uniform density over this set.

Quantization: The relationship between the entropy and the differential entropy is established by quantization. To see this, assume that the range of a continuous random variable x is divided into bins of Δ where $f(x)$ is continuous. Then, the entropy of the quantized random variable is given by

$$\begin{aligned} H(p) &= -\sum_{-\infty}^{\infty} p_i \log(p_i), \\ &= -\sum_{-\infty}^{\infty} f(x_i)\,\Delta \log(f(x_i)\,\Delta), \\ &= -\sum_{-\infty}^{\infty} f(x_i)\,\Delta \log(f(x_i)) - \log(\Delta), \end{aligned} \tag{6}$$

where the first term approaches $-\int_S f(x)\log(f(x))$ as $\Delta \to 0$. Thus, for n bit quantization of a continuous random variable, i.e., $\Delta = 2^{-n}$, the entropy increases with n as

$$H(p) = H(f) + n.$$

This means that in order to represent an n-bit quantized information from $x \sim f(x)$ the average number of bits required is $H(f) + n$ [23].

This concept will be utilized to analyze the entropy of a FISST density when the corresponding RFS is quantized. This analysis demonstrates an important fact about the selection of the JoM estimation constant.

Kullback-Leibler Divergence (Relative Entropy): Kullback-Leibler (KL) divergence is a statistical measure of the difference of a model or a theory based pdf $f(x)$ from a true or a reference pdf $f_t(x)$ on the same support set. If $f_t(x)$ is absolutely continuous with respect to $f(x)$ or $+\infty$ otherwise, KL divergence of $f(x)$ from $f_t(x)$ is defined as

$$\begin{aligned} K(f_t \| f) &= \int f_t(x) \log\left(\frac{f_t(x)}{f(x)}\right) dx, \\ &= \int f_t(x)\log(f_t(x))dx - \int f_t(x)\log(f(x))dx, \\ &= H(f_t \| f) - H(f_t), \end{aligned} \tag{7}$$

where the first term measures the uncertainty introduced by using a model or theory based $f(x)$ instead of the true or reference $f_t(x)$ while the second term is the differential entropy of $f_t(x)$. Hence, the more $f(x)$ resembles $f_t(x)$, the less is the information lost due to using $f(x)$. That is, $K(f \| f_t) \geq 0$ gets smaller values with equality if and only if $f(x) = f_t(x)$.

KL divergence is an important concept used in the development of other consistent concepts in information theory. For example, mutual information is a special case of KL divergence [23], and entropy maximization is in general formulated as the minimization of KL divergence instead of Shannon's entropy given by (1) and (2) [24], [26].

With the help of other relevant concepts KL divergence will be utilized to define the information theoretic part of the multi-objective optimization problem.

B. Multitarget State Estimation

In the following, we exemplify the problems of the MAP and EAP estimators when they are generalized for estimation of multitarget states. Then, we define the global MAP estimators, i.e., the GMAP-I and GMAP-II estimators, which were introduced in [27] and also known as the MaM and JoM estimators in [2], [4], respectively.

Consider the scenario in [2], [3], where a Bernoulli target moves in the one dimensional interval $[0, 2]$ with units given in meters. In addition, suppose that the target existence probability is set to 0.5 and if the Bernoulli target does exist, its spatial probability density is uniform over $[0, 2]$. That is, suppose that the FISST density in units of $\mathrm{m}^{-|X|}$ is

$$f(X) = \begin{cases} 0.5, \text{ if } X = \emptyset \\ 0.25\,\mathrm{m}^{-1}, \text{ if } \begin{cases} X = \{x\} \\ 0 \leq x \leq 2 \end{cases} \\ 0, \text{ otherwise} \end{cases}$$

First, we try to obtain the MAP estimate using $X^{MAP} = \arg\sup_X f(X)$. However, the MAP estimator is undefined since $f(\emptyset) = 0.5$ cannot be compared with $f(\{x\}) = 0.25\,\mathrm{m}^{-1}$. This problem would be eliminated by converting $f(X)$ into a unitless quantity by multiplying it with $\mathrm{m}^{|X|}$. Thus, we obtain the MAP estimate as $X^{MAP} = \emptyset$. However, this conversion results in a paradox. That is, if the Bernoulli target moved in the same interval with units given in kilometer instead of meter, this would result in $f(\{x\}) = 250\,\mathrm{m}^{-1}$. Thus, we would obtain the MAP estimate as $X^{MAP} = \{x\}$ after the conversion. That is, the change in unit of measurements from m to km also changes the MAP estimate [2], [3].

Now, using the set integral we try to obtain the EAP estimate from

$$\begin{aligned} X^{EAP} &= \int X f(X)\,\delta X, \\ &= \emptyset f(\emptyset) + \int_0^2 x f(\{x\})\,dx, \\ &= 0.5\,(\emptyset + 1\,\mathrm{m}). \end{aligned}$$

As indicated in [2], [3], the EAP estimator faces additional problems arising from ill-defined arithmetic operations on sets. Therefore, like the MAP estimator, the EAP estimator is undefined when generalized for estimation of multitarget states.

The GMAP-I and GMAP-II are Bayesian estimators, which are defined according to the minimization of the following cost functions [27]

$$C_0(X, Y) = \begin{cases} 0, \text{ if } |X| = |Y| \\ 1, \text{ if } |X| \neq |Y| \end{cases} \tag{8}$$

and

$$C(X, Y) = C_0(X, Y) + C_1(X, Y), \tag{9}$$

respectively. The second cost function in (9) takes into account the spatial information in a FISST density, i.e.,

$$C_1(X,Y)=\begin{cases} 0, \text{ if } \begin{cases} s=r, \\ (\ell_1,...,\ell_s)=(\varphi_{\beta_1},...,\varphi_{\beta_r}), \\ (x_1,...,x_s)=(y_{\beta_1},...,y_{\beta_r})\in K \end{cases} \\ 1, \text{ otherwise} \end{cases}$$

where the hybrid RFSs are defined as $X=\{\xi_1,...,\xi_s\}$ and $Y=\{\zeta_1,...,\zeta_r\}$ with their identities $(\ell_1,...,\ell_s)$ and $(\varphi_1,...,\varphi_r)$, i.e., $\xi_i=(x_i,\ell_i)$ for $i=1,...,s$ and $\zeta_i=(y_i,\varphi_i)$ for $i=1,...,r$. The RFSs consisting of $\forall x,y\in\mathbb{R}^n$ are surrounded by a closed ball K in $(\mathbb{R}^n)^r$ and are associated through a one-to-one function given by $\beta:(\ell_1,...,\ell_s)\rightarrow(\varphi_1,...,\varphi_r)$. Thus, the cost function in (8) just weights the cardinality discrepancy, whereas the cost function in (9) weights both the cardinality and spatial discrepancies. These properties of the GMAP-I and GMAP-II estimators will help us in evaluating the corresponding MaM and JoM estimators for estimation of multitarget states.

III. Multitarget Bayes Estimators

For RFSs with different cardinalities, their FISST densities have incommensurable scales (i.e., different physical dimensions). Furthermore, addition and subtraction operations on RFSs are not defined properly. Therefore, the multitarget analogues of the MAP and EAP estimators are undefined [2]–[4], [27]. Nevertheless, two MAP like multitarget estimators were proposed for FISST densities. In the following, we show how multitarget states are obtained using these Bayes estimators. In addition, we evaluate how appropriate they are for this purpose based on the results obtained from the analysis of uncertainties related to the cardinality and spatial information in a FISST density.

Marginal Multitarget (MaM) Estimator: The MaM estimate of an RFS is computed in a two-step procedure: first, the MAP estimate of the cardinality is determined:

$$\hat{n}^{MAP}\triangleq\arg\sup_n p_{|X|}(n), \tag{10}$$

where $|X|$ denotes the cardinality variable for the RFS X and is characterized by its probability mass function. That is, the cardinality distribution of the RFS X, given that $Z^{(k)}$ is the RFS of measurements at time k, is

$$p_{|X|}(n)\triangleq\frac{1}{n!}\int f_{k|k}\left(\{x_1,...,x_n\}\middle|Z^{(k)}\right)dx_1...dx_n. \tag{11}$$

Then, the MAP estimate of the multitarget states is determined from the corresponding FISST posterior density for the given cardinality estimate $n=\hat{n}^{MAP}$ as

$$\hat{X}^{MaM}=\arg\sup_{x_1,...,x_{\hat{n}^{MAP}}} f_{k|k}\left(\{x_1,...,x_{\hat{n}^{MAP}}\}\middle|Z^{(k)}\right). \tag{12}$$

The MaM estimator is Bayesian optimal [2], [4], [27]. However, it does not utilize all the information contained in the multitarget posterior density. Hence, it would be statistically unreliable when the target number is related to the spatial information in the FISST posterior density [2], [4]. That is, using the relationship between the FISST probability and measure theoretic probability, the differential entropy of an RFS X is given by [28], [29]

$$\begin{aligned} H(f_X)&=-\int f(X)\log\left(v^{|X|}f(X)\right)\delta X, \\ &=-\sum_{n=0}^{\infty}\frac{1}{n!}\int f(\{x_1,...,x_n\})\times \\ &\qquad\log\left(v^n f(\{x_1,...,x_n\})\right)dx_1...dx_n, \end{aligned} \tag{13}$$

where $v^{-|X|}$ is the unit of the FISST density $f(X)$. Note that the dependence of the FISST posterior density on the RFS $Z^{(k)}$ is dropped here for conciseness.

Substituting $f(\{x_1,...,x_n\})=n!p_{|X|}(n)f(x_1,...,x_n)$ into (13) yields

$$\begin{aligned} H(f_X)=&-\sum_{n=0}^{\infty}p_{|X|}(n)\int f(x_1,...,x_n)\times \\ &\log\left(n!v^n p_{|X|}(n)f(x_1,...,x_n)\right)dx_1...dx_n, \end{aligned} \tag{14}$$

and, after some algebraic manipulations, the differential entropy may be rewritten as the sum of the three terms, i.e.,

$$\begin{aligned} &H(f_X)= \\ &-\sum_{n=0}^{\infty}p_{|X|}(n)\log\left(p_{|X|}(n)\right)\int f(x_1,...,x_n)\,dx_1...dx_n+ \\ &-\sum_{n=0}^{\infty}p_{|X|}(n)\int f(x_1,...,x_n)\log\left(v^n f(x_1,...,x_n)\right)dx_1...dx_n+ \\ &-\sum_{n=0}^{\infty}p_{|X|}(n)\log(n!)\int f(x_1,...,x_n)\,dx_1...dx_n, \end{aligned} \tag{15}$$

where the first term is the entropy of the cardinality distribution:

$$\begin{aligned} H(p)&=\sum_{n=0}^{\infty}p_{|X|}(n)\log\left(p_{|X|}(n)\right)\int f(x_1,...,x_n)\,dx_1...dx_n, \\ &=-\sum_{n=0}^{\infty}p_{|X|}(n)\log\left(p_{|X|}(n)\right), \end{aligned}$$

and the second term is the average differential entropy of the joint pdf of $x_1,...,x_n$ over $p_{|X|}(n)$:

$$E\left[H(f_{X,n})\right]=\sum_{n=0}^{\infty}p_{|X|}(n)H(f_{x,n}).$$

The probability assigned to the FISST density with cardinality n, i.e., $f_{X,n}=f(\{x_1,...,x_n\})$, is uniformly distributed among joint pdfs $f_{x,n}=f(x_1,...,x_n)$ of $n!$ possible vectors for all permutations of $\{x_1,...,x_n\}$, i.e., $f_{x,n}$ are symmetric joint pdfs of $(x_{\sigma 1},...,x_{\sigma n})$, where σ indicates the permutation on the numbers $\{1,...,n\}$ [2], [29]. Hence, the third term indicates the information uncertainty due to change in the representation from RFSs, i.e., $\{x_1,...,x_n\}$, to vectors of indistinguishable points, i.e., $(x_1,...,x_n)$ [28], [29]:

$$E\left[\log(n!)\right]=\sum_{n=0}^{\infty}p_{|X|}(n)\log(n!),$$

The MaM estimator's cost function only penalizes the cardinality discrepancy between the true RFS and its estimate [27]. Therefore, the MaM estimator determines multitarget states without considering the uncertainty represented by the second and the third terms in the FISST densities.

Joint Multitarget (JoM) Estimator: In contrast to the MaM estimator, the JoM estimator determines the target number and multitarget states simultaneously from the FISST posterior density [2] as

$$\hat{X}^{JoM} = \arg\sup_{X} f_{k|k}\left(X \left| Z^{(k)}\right.\right) \frac{\varepsilon^{|X|}}{|X|!}, \tag{16}$$

where the parameter ε denotes a small constant (hereinafter called as the JoM estimation constant) and satisfies that $f\left(\{x_1, ..., x_n\}\right)\varepsilon^n \leq 1$ for all integers $n \geq 0$. However, there is a trade-off in the selection of ε. That is, smaller values of ε yield better accuracy in multitarget state estimates, but with slower convergence to the true multitarget states [2], [4]. In Appendix A, information theoretic analysis demonstrates that the uncertainty in multitarget state estimates cannot be improved by selecting too small values for ε.

Alternatively, the JoM estimator can be performed in a two-step procedure [2]. First, for integer values $n \geq 0$ the MAP estimates of the RFSs are computed from the corresponding posterior FISST densities:

$$\hat{X}^n = \arg\sup_{x_1,...,x_n} f\left(\{x_1, ..., x_n\} \left| Z^{(k)}\right.\right). \tag{17}$$

Then, using $\hat{X}^n$ for each n, the JoM estimate is determined as $\hat{X}^{JoM} = \hat{X}^{\hat{n}}$, where $\hat{n}$ denotes the solution to the following maximization problem:

$$\hat{n} = \arg\sup_{n} f\left(\{\hat{x}_1, ..., \hat{x}_n\} \left| Z^{(k)}\right.\right) \frac{\varepsilon^n}{n!}. \tag{18}$$

Like the MaM estimator, the JoM estimator is Bayesian optimal [2], [4], [27]. However, it is naturally more appropriate for the estimation of multitarget states since its cost function penalizes both discrepancies in cardinality and multitarget states [27]. In addition, it is known that the JoM estimator is statistically convergent [2], [4].

IV. Optimization of the JoM Estimation Constant

The differential entropy of a pdf is roughly represented by a uniform density over its typical set [23], [25]. However, typical sets do not include the sequences of all the most (least) probable state estimates [23], [25]. For example, Fig. 1 shows the cross-section of the typical set of a standard Gaussian density around a hypersphere centered at the origin of $\mathbb{R}^{n_x}$ [25], [30]. It can be seen that the typical set is represented by a thin shell bounded by two convex sets (see Appendix B). Instead, for log-concave pdfs (e.g., a Gaussian pdf) superlevel sets can be defined so as to include the sequences of most likely state estimates [30], [31]:

$$S_\lambda = \left\{x \in \mathbb{R}^{n_x} \middle| f\left(\tilde{x}_1, ..., \tilde{x}_n\right) \geq \lambda\right\}, \tag{19}$$

where $\tilde{x}_1, ..., \tilde{x}_n$ are i.i.d. samples drawn from the log-concave pdf $f(x)$, and λ is the supremum value of the uniform probability on the typical set for a small positive constant

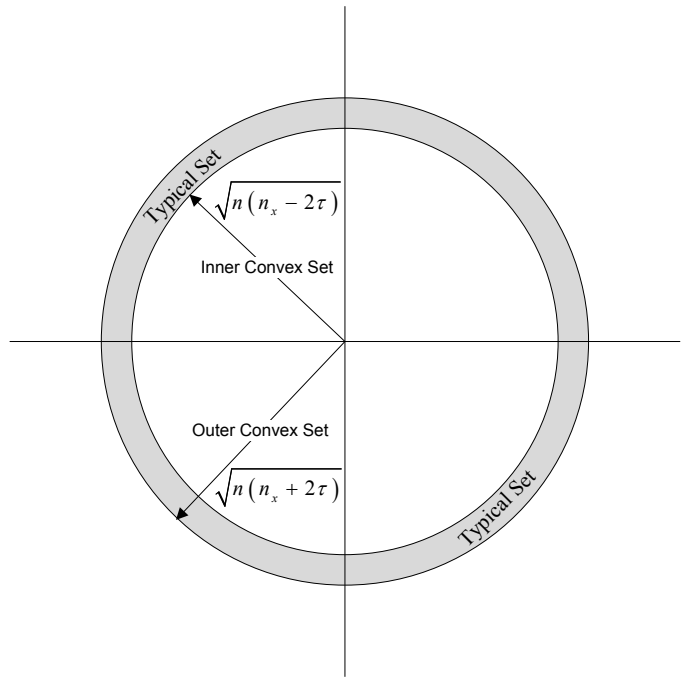

Fig. 1. The cross section of the typical set of the standard Gaussian density in $\mathbb{R}^{n_x}$.

τ, i.e., $\lambda = e^{-n(H(f)-\tau)}$ [23], where $H(f)$ is in nats. In particular, if x is Gaussian-distributed with mean μ and covariance matrix P in $\mathbb{R}^{n_x}$, i.e., $x \sim N(\mu, P)$, then substituting $H(f) = 0.5\log\left((2\pi e)^{n_x}|P|\right)$ [23] for λ yields

$$\lambda = \left((2\pi)^{n_x}|P|\right)^{-n/2} e^{-n\left(\frac{n_x}{2}-\tau\right)},$$

and the joint probability distribution of i.i.d. samples are given by

$$\begin{aligned} f\left(\tilde{x}_1, ..., \tilde{x}_n\right) &= \prod_{i=1}^{n} f\left(\tilde{x}_i\right), \\ &= f\left(\hat{x}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\tilde{x}_i-\mu\right)^T P^{-1}\left(\tilde{x}_i-\mu\right)}, \end{aligned}$$

where $f\left(\hat{x}\right) = \left((2\pi)^{n_x}|P|\right)^{-1/2}$.

Thus, the superlevel set given by (19) can be alternatively defined as

$$S_\lambda = \left\{\tilde{x} \in \mathbb{R}^{n_x} \middle| \frac{1}{n}\sum_{i=1}^{n}\left(\tilde{x}_i - \mu\right)^T P^{-1}\left(\tilde{x}_i - \mu\right) \leq n_x - 2\tau\right\}. \tag{20}$$

In general, this bounded and closed set includes the sequences of most likely random samples drawn from $f(x)$. However, our aim is to define a confined set that exclusively consists of good state estimates from the JoM estimator. To this end, the superlevel set in (20), when evaluated at $n = 1$, gives the least upper bound for this special subset as

$$S_\lambda^{(1)} = \left\{x \in \mathbb{R}^{n_x} \middle| \left(x - \mu\right)^T P^{-1}\left(x - \mu\right) \leq n_x - 2\tau\right\}, \tag{21}$$

where $0 < 2\tau < n_x$. This means that $S_\lambda^{(1)}$ is a hyperellipsoid (i.e., a convex set) with the centroid at μ in the region

surrounded by the inflection points of the Gaussian density $f(x)$.

The entropy maximization helps ignore spurious details like tail probabilities and side-lobes for which samples from these parts can be hardly ever observed [32]. Over bounded and closed sets, the entropy maximization is achieved by uniform densities [23]. Then, the KL divergence of $f(x)$ from the uniform density defined on $S_{\lambda}^{(1)}$, i.e., $u(x)=\varepsilon_{\lambda}^{-1}$ is given by

$$\begin{aligned} K(u\,\|f) &= \int u(x)\log\left(\frac{u(x)}{f(x)}\right)dx, \\ &= H(u\,\|f)-\log(\varepsilon_{\lambda}), \end{aligned} \tag{22}$$

where $\log(\varepsilon_{\lambda})$ is the differential entropy of $u(x)=\varepsilon_{\lambda}^{-1}$, i.e., $H(u)=\log(\varepsilon_{\lambda})$, and

$$\begin{aligned} H(u\,\|f) &= -\log(f(\hat{x}))+\frac{1}{2\varepsilon_{\lambda}}\int_{\varepsilon_{\lambda}}(x-\mu)^{T}P^{-1}(x-\mu)dx, \\ &\leq -\log(f(\hat{x}))+\frac{1}{2}(n_x-2\tau), \end{aligned}$$

where the last inequality follows from (21). Thus, the KL divergence in (22) can be rewritten as

$$K(u\,\|f)\ \leq -\log(f(\hat{x})\,\varepsilon_{\lambda})+\frac{1}{2}(n_x-2\tau), \tag{23}$$

where the first term on the right hand side is the approximated KL divergence of $f(x)$ from $u(x)=\varepsilon_{\lambda}^{-1}$ when ε_{λ} takes so small values, i.e., $n_x-2\tau\rightarrow 0$. Note that the sum on the right hand side of (23) is always nonnegative since $K(u\,\|f)\ \geq 0$ on ε_{λ}.

The volume of the hyperellipsoid $S_{\lambda}^{(1)}$ can be expressed in terms of τ as follows [33]:

$$\varepsilon_{\lambda}=C(n_x)\ |P|^{1/2}\ r^{n_x/2}, \tag{24}$$

where $r=n_x-2\tau$ is the critical value for the total probability of $f(x)$ in the hyperellipsoid, and $C(n_x)$ is the volume of the hypersphere with the unit radius in $\mathbb{R}^{n_x}$.

After substituting for ε_{λ} into (23), the problem at hand (i.e., determining the optimum volume of the hyperellipsoid) can be formulated as a nonlinear convex optimization problem that determines the optimum value of τ for the least upper bound of the KL divergence. That is,

$$\begin{aligned} &\text{minimize} && f_{o,I}(\tau)=-\log(f(\hat{x})\,\varepsilon_{\lambda})+\frac{1}{2}(n_x-2\tau), \\ &\text{subject to} && g_1(\tau)=-\tau\leq 0, \\ & && g_2(\tau)=-(n_x-2\tau)+\gamma_{\min}\leq 0, \end{aligned} \tag{25}$$

where $\gamma_{\min}$ is a small constant determined according to the chi-square table, considering the degree of freedom (i.e., n_x) and the probability of the confidence level indicating the smallest hyperellipsoid, e.g., $\Pr((n_x-2\tau)\geq\gamma_{\min})\geq 95\%$.

The convex optimization problem in (25) is solely formulated in terms of information theoretic sense. In other words, the objective function $f_{o,I}(\tau)$ in (25) is minimized as $n_x-2\tau\rightarrow n_x$ (see Appendix C for proof). Thus, the computation of the least upper-bound on the KL divergence through the optimization problem in (25) corresponds to the minimization of information gain in magnitude measured by

$$K(u\,\|f)\leq H(f)-\log(\varepsilon_{\lambda}).$$

In the JoM estimator, the selected hyperellipsoid surrounding the estimated states of targets should satisfy

$$\begin{aligned} \int_{\varepsilon_{\lambda}^{n}} f(\{x_1,...,x_n\})\,dx_1...dx_n &\triangleq \int_{\varepsilon_{\lambda}^{n}} f(x_1,...,x_n)\,dx_1...dx_n, \\ &\cong f(\hat{x}_1,...,\hat{x}_n)\,\varepsilon_{\lambda}^{n}, \end{aligned} \tag{26}$$

where the first expression follows from $f(\{x_1,...,x_n\})\triangleq n!\,f(x_1,...,x_n)$ and implies that the volume of the hyperellipsoid ε_{λ} for each target should be so small that only one permutation of the RFS is possible in the product space $\varepsilon_{\lambda}^{n}$, i.e., $\{x_1,...,x_n\}=(x_1,...x_n)$ [27]. However, as indicated in [2], setting $\varepsilon_{\lambda}^{n}$ to extremely small values would be impractical without considering the information provided by $f(x)$. In other words, $u(x)$ would be more informative than $f(x)$ as $n_x-2\tau\rightarrow 0$. However, this contradicts the information theoretic part of the optimization problem in (25), which aims for entropy maximization by minimizing information gain obtained using $u(x)$ instead of $f(x)$.

In contrast to single-objective optimization, there is usually no unique solution that simultaneously achieves the optimization of more than one objective function. Instead, in multi-objective optimization problems, Pareto-optimal solutions can be computed according to the relative importance of individual objective functions [19], [20]. For a vector of conflicting objective functions given by $F(x)=[f_1(x),...,f_N(x)]$ a solution x^* is said to be Pareto optimal if there does not exist another solution that dominates it [19]. That is, given that T is the feasible design space, there is no another point, $x\in\mathrm{T}$ satisfying $F(x)\leq F(x^*)$ and $f_i(x)<f_i(x^*)$ for at least one objective function. There are multiple methods for multi-objective optimization problems. However, the conversion of the multi-objective problem into a single-objective problem is the standard way of solving [19], [20].

To determine the optimum value of τ, two objective functions $f_{o,I}(\tau)$ and $f_{o,J}(\tau)$, which quantify entropy maximization and the accuracy of the JoM estimator, respectively, are in conflict with one another. An optimization problem with a single convex objective function can be defined by aggregating them with appropriately selected weights. However, a consistent Pareto-optimal solution to this optimization problem requires the normalization of these conflicting objective functions in different magnitudes [20], [21]. To this end, their extreme values are calculated at the vertex points of the Pareto-optimal set [20]. Specifically, for the problem at hand, first set $\tau=0$ to obtain the minimum of $f_{o,I}(\tau)$, i.e., $F_{o,I}^{Min}$ while setting $f_{o,J}(\tau)$ to its maximum value, i.e., $F_{o,J}^{Max}$. Then, set $\tau=0.5(n_x-\gamma_{\min})$ to obtain $F_{o,I}^{Max}$ and $F_{o,J}^{Min}$ for $f_{o,I}(\tau)$ and $f_{o,J}(\tau)$, respectively. Finally, the following robust normalization is performed for these conflicting objective functions [20], [21]:

$$f_{o,\xi}^{Trans}(\tau)=\frac{f_{o,\xi}(\tau)-F_{o,\xi}^{Min}}{F_{o,\xi}^{Max}-F_{o,\xi}^{Min}},\forall\xi\in\{I,J\}.$$

Thus, an optimization problem with a single convex objective function can be obtained as follows:

$$\begin{aligned} \text{minimize} \quad & f_m(\tau) = w_I f_{o,I}^{Trans}(\tau) + w_J f_{o,J}^{Trans}(\tau), \\ \text{subject to} \quad & g_1(\tau) = -\tau \leq 0, \\ & g_2(\tau) = -(n_x - 2\tau) + \gamma_{\min} \leq 0, \end{aligned} \tag{27}$$

where $f_{o,J}^{Trans}$ is the normalization of the objective function defined as

$$f_{o,J}(\tau) = \begin{cases} (n_x - 2\tau)^2 & \text{if } (n_x - 2\tau) > \gamma_{\min} \\ 0 & \text{otherwise,} \end{cases}$$

considering the accuracy of the JoM estimator.

In this paper, the weights of the conflicting objectives are determined as linear predictions from autoregressive (AR) models. The next section presents details about this process. However, the weights can also be chosen depending on the application and preference of decision maker(s) [19], [20].

The nonlinear convex optimization problem in (27) can be solved using any standard nonlinear optimization technique [19]. In addition, the solution is strictly Pareto optimal for the positive weights of the convex objective functions [20], [21] . In this paper, the sequential quadratic programming (SQP) is employed to find a Pareto-optimal solution to (27). The SQP iteratively solves a quadratic approximation to the Lagrangian function, in the sense that the sequence of solutions approaches to optimal solution satisfying the necessary Karush-Kuhn-Tucker (KKT) conditions [34], [35]. Note that there are many other ways to solve the above multi-objective optimization problem. The contribution of this paper is not in optimization, but in multitarget detection and state estimation. Thus, we have used a standard optimization approach that guarantees a Pareto-optimal solution without exhaustive comparison with other approaches.

In order to illustrate the geometrical interpretation of the weighted sum method, let us examine the nonlinear convex optimization problem in (27) with the following parameters: $P = \text{diag}\left([50, 50, 10, 10]'\right)$, $n_x = 4$ and $\gamma_{\min} = 0.297$ with the confidence probability of 99.9%. Considering the inequality constraints in (27) the feasible design space of τ, i.e., $\mathrm{T} = \{\tau \,|g_i(\tau) \leq 0, i = 1, 2\}$ is obtained as $\mathrm{T} = [0, 1.8515]$ [20], [22]. Thus, the feasible criterion space of the vector of the normalized objective functions, i.e., $F = \left[f_{o,I}^{Trans}(\tau), f_{o,J}^{Trans}(\tau)\right]$ is defined as $\Omega = \{F \,|\tau \in \mathrm{T}\}$ [20], [22]. Fig. 2 shows the relationship between the Pareto front and the normalized objective functions in the feasible criterion space. The Pareto front is the set of the non-dominated points, i.e., Pareto-optimal points in the criterion space [20]. As can be seen in Fig. 2, the Pareto front is a convex curve. Thus, a Pareto-optimal point can always be obtained depending on the weights of the conflicting objective functions [22], [36]. This is because for a given set of weights, the weighted sum method approximates the Pareto front as a line [36]:

$$f_{o,I}^{Trans}(\tau) = -\frac{w_J}{w_I} f_{o,J}^{Trans}(\tau) + \frac{1}{w_I} f_m(\tau^*),$$

where τ^* denotes a Pareto-optimal solution. For example, the SQP finds the Pareto-optimal solution as $\tau^* = 1.1674$ if the conflicting objective functions are considered equally important, i.e., $w_I = w_J = 0.5$. Thus, the Pareto-optimal point in the feasible design space is computed as $F = [0.1747, 0.1687]$. As expected, the normalized objective functions in conflict are penalized almost equally. In Fig. 2, the line with the slope -1 is tangent to the Pareto front at $F = [0.1747, 0.1687]$ and locally approximates the convex Pareto front.

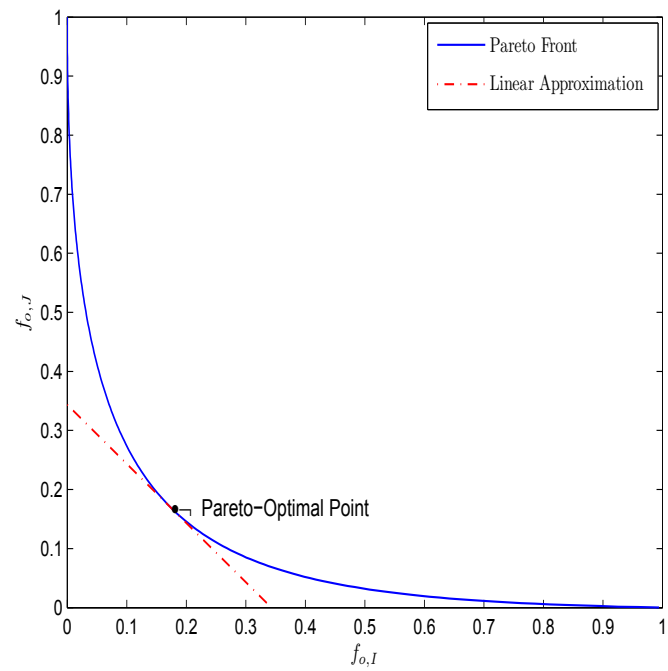

Fig. 2. Geometrical interpretation of the weighted sum method in the feasible criterion space.

V. Linear predictions of objective weights

AR models predict the current output of a stochastic process based on its previous outputs. The AR model of order N, denoted as $AR(N)$, is in general defined by [37]

$$x_k = c + \sum_{i=1}^{N} \alpha_i x_{k-i} + \vartheta_k,$$

where c denotes a constant for a non-zero mean value of x_k, $\{\alpha_i\}_{i=1}^{M}$ are predictor coefficients and ϑ_k is a white noise representing prediction error with zero mean and variance σ_ϑ^2. For linear predictions of the objective weights, we use the following $AR(1)$ model:

$$w_k = c + \alpha w_{k-1} + \vartheta_k, \tag{28}$$

where the predictor coefficient indicates linear relationship in this time series. For a wide sense stationary (WSS) process, the condition $|\alpha| < 1$ must be satisfied. In this case, the $AR(1)$ model is statistically characterized by [37]

$$\begin{aligned} E[w_k] &= \mu_w = \frac{c}{1-\alpha}, \\ \text{var}(w_k) &= \sigma_w^2 = \frac{\sigma_\vartheta^2}{1-\alpha^2}, \\ \text{cov}(w_k, w_{k-i}) &= \sigma_w^2 \alpha^i. \end{aligned}$$

Thus, the autocorrelation function between w_k and w_{k-i} decays to zero by α^i as $i \to \infty$. This means that the $AR(1)$ model is also stable, i.e., represents a predictable process.

The objective function $f_{o,J}(\tau)$ in (27) only considers the degree of freedom, i.e., n_x because of the definition of the hyperellipsoid in (21). Thus, substituting (24) into (26) for a Bernoulli target with parameter pair $\{q_k, f_k\}$ over the volume ε_λ results in

$$\int_{\varepsilon_\lambda} f_k(\{x\})\, dx \cong \; q_k f_k(\hat{x})\, \varepsilon_\lambda, \\ = q_k \frac{1}{2^{n_x/2}\Gamma\left(\frac{n_x}{2}\right)} (n_x - 2\tau)^{n_x/2},$$

where $f_k(x)$ is a Gaussian pdf and $\Gamma(\cdot)$ denotes the gamma function. Notice that the approximation is independent of P at time k, denoted as P_k. To consider the covariance of $f_k(x)$ implicitly in this approximation we determine the degree of correlation between $w_{J,k}$ and $w_{J,k-1}$ as

$$\beta_k = \frac{|P_{k-1}|^{1/2}}{|P_k|^{1/2}} \mathbf{1}_A(q_k),$$

where the first term is the ratio of infinitesimal volumes to locate a Bernoulli target with the same spatial probability at time k and $k-1$, respectively and $\mathbf{1}_A$ denotes an indicator function defined on the set $A = [q_{\min}, 1]$ [2]. The indicator function neglects changes in P_k before confirming a Bernoulli target with the threshold $q_{\min}$. Thus, we keep the weights at their initial states until a probable Bernoulli target is confirmed. In addition, for a stable process the correlation must decay to zero as time lag increases. For this purpose, we set $\alpha = \beta_k$ in (28) within its control limits as shown in Fig. 3.

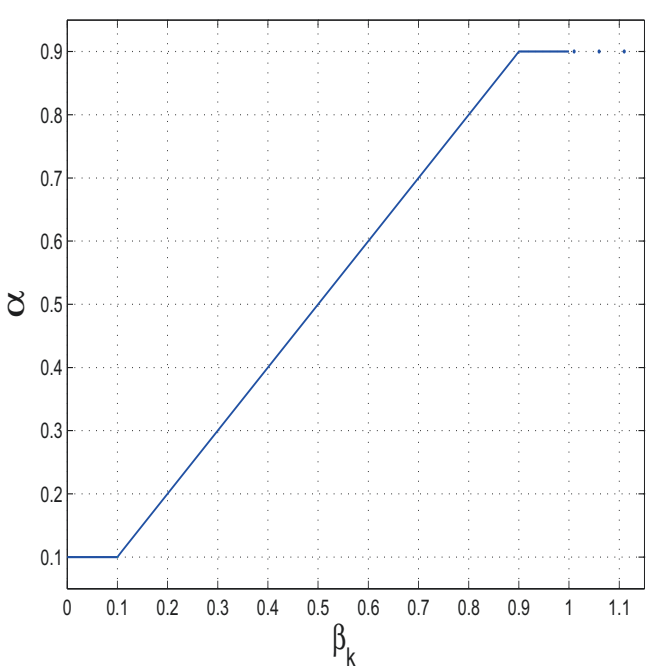

Fig. 3. Predictor coefficient of AR(1) model versus the degree of correlation between successive weights.

At this point, it is important to note that our $AR(1)$ model with the predictor coefficient evolving in time does not represent a WSS process. However, it would turn into a WSS process after the optimal JoTT filter converges to its steady-state with detections. Then, the predictor coefficient is set to $\alpha = 0.9$ according to Fig. 3 since successive changes in P_k would be small. Thus, the linear predictions monotonically approach to $\mu_{w,J} = 10c_J$, where $0.1 \le \mu_{w,J} \le 0.9$ in order to prevent that one objective completely dominates another in the multi-objective optimization. Since $f_k(x)$ is very peaky after the convergence, $f_{o,J}(\tau)$ becomes more important than $f_{o,I}(\tau)$ in (27). Hence, $\mu_{w,J}$ is set to its maximum value, i.e., $\mu_{w,J} = 0.9$ by $c_J = 0.09$.

Using $w_{I,k} + w_{J,k} = 1$, the the $AR(1)$ model for $w_{I,k}$ is defined by

$$w_{I,k} = 0.01 + \alpha w_{I,k-1} + \nu_k,$$

where ν_k is a white noise with zero mean and variance $\sigma^2_{w,I} = \sigma^2_{w,J}$ since $\nu_{I,k} = -\vartheta_{J,k}$. Similarly, after the convergence its linear predictions monotonically approach to $\mu_{w,I} = 0.1$.

On the other hand, the optimal JoTT filter gradually deteriorates after target death. Therefore, β_k takes values close to zero and with $\alpha = 0.1$ the linear predictions for $w_{J,k}$ and $w_{I,k}$ monotonically approach to their opposite means, i.e., $\mu_{w,J} = 0.1$ and $\mu_{w,I} = 0.9$, respectively. Consequently, $f_{o,I}(\tau)$ becomes more important than $f_{o,J}(\tau)$ in (27) as $f_k(x)$ disperses over ε_λ.

VI. Implementation of the JoM Estimator for the JoTT Filter

Suppose that at most one target is present. In this case, the RFS of a single target can be modeled as a Bernoulli RFS with the parameter pair (q_{k-1}, f_{k-1}). Thus, its FISST density is parameterized as

$$f_{k-1}(X) = \begin{cases} 1 - q_{k-1} & \text{if } X = \emptyset \\ q_{k-1} f_{k-1}(x) & \text{if } X = \{x\}, \end{cases} \tag{29}$$

where q_{k-1} is the existence probability of the target, and $f_{k-1}(x)$ is its spatial pdf if the target is present.

In the prediction step of the JoTT filter, the FISST density $f_{k-1}(X)$ propagated to time k is parameterized as follows [2], [13]:

$$q_{k|k-1} = p_B(1 - q_{k-1}) + q_{k-1} \int p_{S,k-1}(x) f_{k-1}(x)\, dx_{k-1}, \tag{30}$$

$$f_{k|k-1}(x) = \frac{1}{q_{k|k-1}} \left[(1 - q_{k-1}) p_B b_k(x) + q_{k-1} \langle f, p_S \psi \rangle\right], \tag{31}$$

where a newborn target is declared with probability p_B according to a birth density $b_k(x)$, i.e., the Bernoulli parameter pair (p_B, b_k), and

$$\langle f, p_S \psi \rangle = \int f_{k-1}(x) p_{S,k-1}(x) \psi_{k|k-1}(\cdot|x)\, dx_{k-1},$$

where $p_{S,k-1}(x)$ is the state-dependent target survival probability and if the target survives, its states evolve according to the Markov state transition density $\psi_{k|k-1}(\cdot|x)$.

Suppose that the single-sensor multitarget measurements at time k are modeled as

$$\mathrm{Z}_k = \Gamma_k\left(x\right)\mathrm{UC_k},$$

where C_k is the RFS of i.i.d. false alarms and $\Gamma_k\left(x\right)$ is the Bernoulli RFS of target-originated measurement with the parameter pair $(p_D(x), g_k(z|x))$, where $p_D(x)$ is the detection probability, and $g_k(z|x)$ is the measurement likelihood function.

In the original derivation of the JoTT filter, the false alarm process is modeled as an arbitrary RFS. If the Poisson false alarm RFS with mean rate λ_c and spatial pdf $c\left(z\right)$ is substituted for the arbitrary false alarm RFS, the original data update equations of the JoTT filter defined in [2], [13] have the form of

$$q_{k|k} = \frac{1 - f_{k|k-1}\left[p_D\right] + \sum\limits_{z\in Z_k}\frac{f_{k|k-1}\left[p_D g_k(z|\cdot)\right]}{\kappa(z)}}{q_{k|k-1}^{-1} - f_{k|k-1}\left[p_D\right] + \sum\limits_{z\in Z_k}\frac{f_{k|k-1}\left[p_D g_k(z|\cdot)\right]}{\kappa(z)}}, \quad (32)$$

$$f_{k|k}\left(x\right) = \frac{1 - p_D\left(x\right) + p_D\left(x\right)\sum\limits_{z\in Z_k}\frac{g_k(z|x)}{\kappa(z)}}{1 - f_{k|k-1}\left[p_D\right] + \sum\limits_{z\in Z_k}\frac{f_{k|k-1}\left[p_D g_k(z|\cdot)\right]}{\kappa(z)}} f_{k|k-1}\left(x\right). \quad (33)$$

where, in general, $f_{k|k-1}\left[x\right] = \int x\, f_{k|k-1}\left(x\right)dx$ and $\kappa\left(z\right) = \lambda_c\, c\left(z\right)$ is the intensity function of the Poisson false alarm RFS.

For the JoM estimator, the Bayesian risk function to be minimized is given by [27]

$$\int C\left(X, J\left(Z\right)\right) f\left(X\right)\delta X \approx 2 - p_{|X|}\left(|J|\right) - \frac{f\left(X\right)\varepsilon^{|J|}}{|J|!}, \quad (34)$$

where J denotes the JoM estimator, C is the cost function that penalizes both discrepancies in cardinality and multitarget states, and $p_{|X|}\left(|J|\right)$ is the cardinality distribution evaluated at the target number $|J|$.

Then, using the updated Bernoulli parameters from the JoTT filter, the JoM estimator confirms the presence of a single target if

$$2 - \left(1 - q_{k|k}\right) > 2 - q_{k|k} - q_{k|k}\, f_{k|k}\left(\hat{x}\right)\varepsilon, \quad (35)$$

where the left hand side is the Bayes risk function evaluated for the "no-target" case, i.e., $X = \emptyset$ and the right hand side is the Bayes risk function evaluated for the "target-present" case, i.e., $X = \{x\}$. Solving this inequality for $q_{k|k}$ yields the following test for "target-present" decision:

$$q_{k|k} > \frac{1}{2 + f_{k|k}\left(\hat{x}\right)\varepsilon}. \quad (36)$$

As in the original JoM estimator, first, the MAP estimate of $X = \{x\}$ is computed from the parameterized FISST density, i.e., $\left(q_{k|k}, f_{k|k}\right)$ where the spatial pdf $f_{k|k}$ has the Gaussian mixture form, i.e., $f_{k|k}\left(x\right) = \sum_{i=1}^{N_k} w_{k|k}^{(i)} f_{k|k}^{(i)}$, with the mixing weights satisfying that $\sum_{i=1}^{N_k} w_{k|k}^{(i)} = 1.0$. Before state estimation, pruning and merging of the Gaussian components are performed. Thus, the state estimation is obtained using the well-separated and significant Gaussian density components according to (17). For the selected Gaussian density component, its Pareto-optimal volume given by $T_{P,opt} = q_{k|k}\varepsilon_{P,opt}$ is computed. Then, the test for "target-present" decision in (36) is checked using $\varepsilon_{P,opt}$. That is, $f_{k|k}\left(\hat{x}\right)\varepsilon_{P,opt}$ is set to $\min\left(f_{k|k}\left(\hat{x}\right)\varepsilon_{P,opt}, 1/q_{k|k}\right)$. Consequently, if target is progressively better-localized, all of its probability mass would be almost located in $\varepsilon_{P,opt}$, i.e., $q_{k|k}\, f_{k|k}\left(\hat{x}\right)\varepsilon_{P,opt} \approx 1$ [2].

VII. Simulation Results

In this section, the proposed JoM estimator is compared with the MaM estimator. To do this, their track management performance using outputs of the JoTT filter is evaluated through the OSPA metric [17], [18]. The OSPA metric compares two finite sets X, and Y, considering the difference in their cardinalities (i.e., cardinality error) and the positional distance between their associated points (i.e., localization error) after an optimal assignment. The sensitivity of the OSPA metric to these two errors are controlled by the cut-off parameter c and the order parameter p. However, for a Bernoulli RFS the OSPA metric reduces to [38]

$$d_p^{(c)}\left(X, Y\right) = \begin{cases} 0 & \text{if } X = \emptyset,\ Y = \emptyset \\ c & \text{if } X = \emptyset,\ Y = \{y\} \\ c & \text{if } X = \{x\},\ Y = \emptyset \\ d^{(c)}\left(x, y\right) & \text{if } X = \{x\},\ Y = \{y\}, \end{cases}$$

where $d^{(c)}\left(x, y\right) = \min\left(c, d\left(x, y\right)\right)$ is the cut-off distance between the points in two non-empty Bernoulli RFSs. Thus, in this case, the OSPA metric is independent of the order parameter p. In addition, the major performance difference between the two estimators is expected to occur in the accuracy of their decisions on track confirmation, track maintenance, and track termination. Then, the cut-off parameter c must be set to a high value in order to make the OSPA metric sensitive to cardinality errors due to false and missing point estimates. In simulations, the OSPA metric is therefore computed with the parameters $p = 1$, and $c = 25$.

The target state vector comprises position and velocities in $\mathrm{x-y}$ directions, i.e., $x_k = [p_{x,k}, p_{y,k}, v_{x,k}, v_{y,k}]'$. If the target does survive with probability $p_S = 0.90$, its states evolve according to the coordinated turn model with the known turn rate Ω [33], [39], i.e., the state transition model is

$$x_k = F\left(\Omega\right)x_{k-1} + G\omega_{k-1},$$

where $\omega_{k-1} \sim N\left(0, Q_{k-1}\right)$ is the zero-mean Gaussian process noise with covariance matrix $Q_{k-1} = diag\left([0.1, 0.1]'\right)\mathrm{m/s^2}$, and the system matrices are

$$F\left(\Omega\right) = \begin{bmatrix} 1 & 0 & \frac{\sin(\Omega\mathrm{T})}{\Omega} & -\frac{1-\cos(\Omega\mathrm{T})}{\Omega} \\ 0 & 1 & \frac{1-\cos(\Omega\mathrm{T})}{\Omega} & \frac{\sin(\Omega\mathrm{T})}{\Omega} \\ 0 & 0 & \cos\left(\Omega\mathrm{T}\right) & -\sin\left(\Omega\mathrm{T}\right) \\ 0 & 0 & \sin\left(\Omega\mathrm{T}\right) & \cos\left(\Omega\mathrm{T}\right) \end{bmatrix},$$

$$G = \begin{bmatrix} \frac{T^2}{2} & 0 \\ 0 & \frac{T^2}{2} \\ T & 0 \\ 0 & T \end{bmatrix},$$

where T is the sampling interval and set at T $=$ 1s in simulations.

The single target tracking scenario runs for 40s. The target appears at time $k = 6$ and moves along a straight line with a constant speed of $|v| = 5\,\mathrm{m/s}$ in the x $-$ y directions until time $k = 20$. Then, it starts maneuvering at a constant turn rate of $|\Omega| = 2\,\mathrm{deg/s}$ and is terminated at time $k = 35$. The target birth is modeled as a Bernoulli RFS given by $\{q_b, f_b(x)\}$, where the birth existence probability is set at $q_b = 0.01$, and the spatial pdf is defined as $f_b(x) = N(\hat{x}_b, P_b)$ with mean $\hat{x}_b = [-70, 70, 0, 0]'$ and covariance matrix $P_b = diag\left([50, 50, 10, 10]'\right)$.

The target is detected by a sensor with state-independent detection probability p_D and the sensor has a linear Gaussian measurement model given by

$$z_k = Hx_k + \eta_k,$$

where $\eta_k \sim N(0, R_k)$ is the zero-mean Gaussian measurement noise with covariance matrix $R_k = diag\left([1, 1]'\right)\mathrm{m}$. With $I_{2\times2}$ and $0_{2\times2}$ denoting the $n \times n$ identity and zero matrices, respectively, the observation matrix is given by $H = [I_{2\times2}, 0_{2\times2}]$. In addition to noisy target-originated measurement, the received measurement set includes clutter points. In simulations, clutter is modeled as a Poisson RFS with the mean rate of $\lambda_c = 10$ per scan and uniform spatial distribution over the surveillance region $V = [-300\mathrm{m}, 300\mathrm{m}] \times [-300\mathrm{m}, 300\mathrm{m}]$, i.e., $c(z) = V^{-1}$. The performance of the two estimators is evaluated by running the same scenario for 500 Monte Carlo runs. In each trial, target-originated measurement, detected with p_D, and independent random clutters are generated. Fig. 4 shows the x and y components of the target trajectory, measurements and the position estimates obtained from the JoTT filter with $p_D = 0.80$ for one Monte Carlo trial.

In the JoTT filter, the Bernoulli RFS is represented as a Gaussian mixture. The maximum number of Gaussian components is set at $J_{\max} = 100$. They are pruned and merged at each time step with thresholds $T_{prune} = 10^{-3}$ and $T_{merge} = 4.0$, respectively according to the algorithm proposed in [40].

The track management performance of the proposed JoM estimator and the MaM estimator are shown in Fig. 5–7 for different values of the detection probability, ranging from high to moderately small values, i.e., $p_D = 0.95, 0.90, ..., 0.70$. The MaM estimator confirms "target-present" decision by comparing the existence probability $q_{k|k}$ with the hard threshold 0.5. However, the proposed JoM estimator confirms "target-present" decision by setting a lower margin than this hard threshold considering how well the JoTT filter localizes the target, i.e., the term $f_{k|k}(\hat{x})\,\varepsilon$ in (36). However, the maximum value of $f_{k|k}(\hat{x})\,\varepsilon$ is set by a confirmation threshold q_{min}. In simulations, q_{min} is set to 0.20. Thus, the track, for which $q_{k|k} > q_{min}$, is confirmed by the JoM estimator. In particular, the use of this threshold helps to prevent false point estimates before the target birth and after the target death.

In Fig. 5, it can be seen that the two estimators demonstrate almost the same track management performance in terms of track confirmation before the target birth at time $k = 6$. In addition, the initial track maintenance quality of the proposed JoM estimator with insignificant values of the lower margin is nearly the same as that of the MaM estimator. However, the JoTT filter localizes the target more accurately using target-originated measurements detected with high probability as time proceeds. Therefore, the lower margin than the hard threshold 0.5 becomes significant, so that the proposed JoM estimator does not prematurely declare track termination if

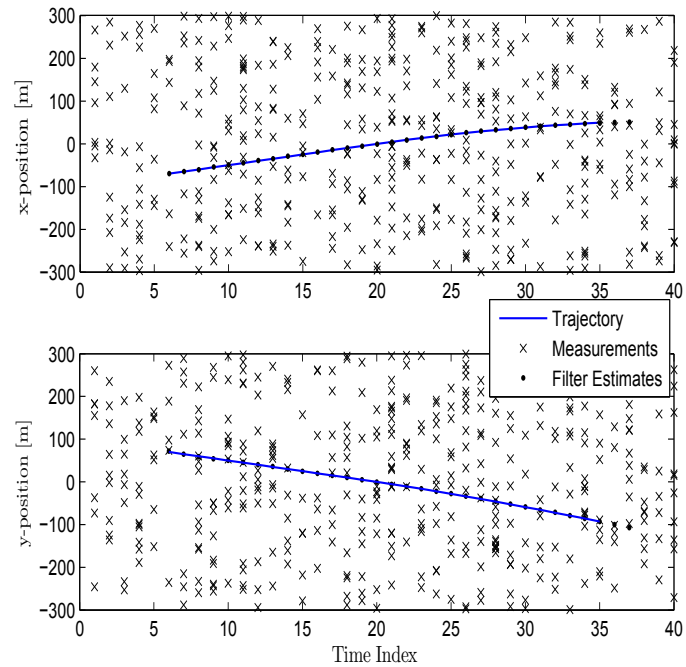

Fig. 4. x and y components of target trajectory, measurements and JoTT filter estimates.

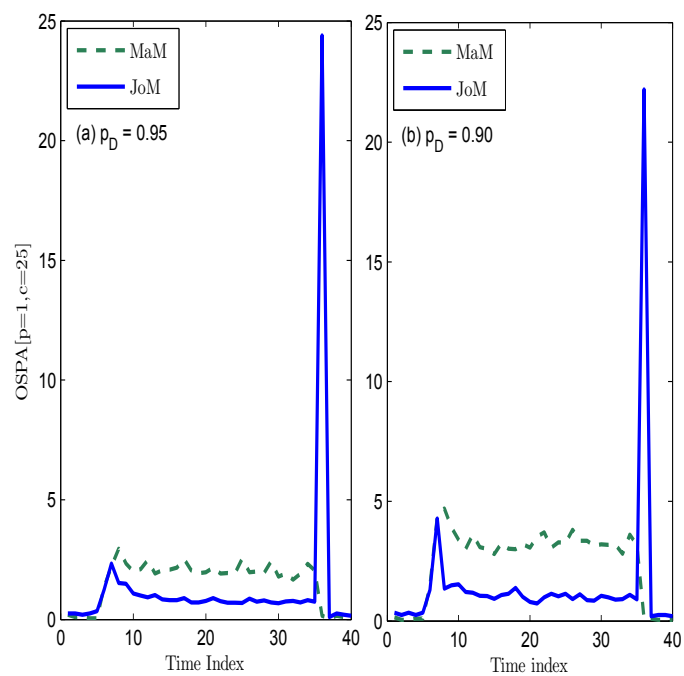

Fig. 5. 500 Monte Carlo run averages of the OSPA metric computed for the track management performance of the JoM and MaM estimators.

the target is miss-detected due to sensor imperfection. On the other hand, large values of the lower margin than the hard threshold 0.5 result in latency on track termination. That is, after the target is terminated at time $k = 35$, the localization performance of the JoTT filter does deteriorate gradually due to missed detections. Hence, the track termination decision is delayed in the proposed JoM estimator.

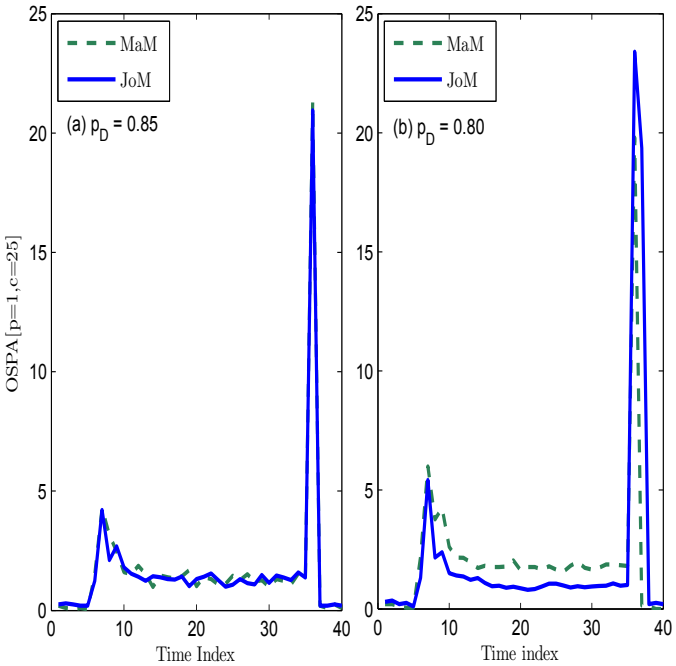

Fig. 6. 500 Monte Carlo run averages of the OSPA metric computed for the track management performance of the JoM and MaM estimators.

In Fig. 6(a), it can be seen that the track management performances of the two estimators are nearly the same during the tracking scenario. These results indicates that the decrease in the existence probability of target ($q_{k|k}$) cannot be compensated by the value of the lower margin computed in the proposed JoM estimator when the target is miss-detected. However, Fig. 6(b) shows that the track maintenance quality of the proposed JoM estimator is better than that of the MaM estimator after the target birth. That is, the value of the lower margin can compensate the decrease in $q_{k|k}$ due to target being miss-detected. Nevertheless, the proposed JoM estimator suffers from track termination latency more than the MaM estimator due to the statistics indicating a well-localized target obtained from the JoTT filter after time $k = 35$.

Finally, Fig. 7 shows the track management performances of the two estimators under moderately small detection probabilities. It can be seen that the initial track management performance of the proposed JoM estimator is better than that of the MaM estimator. More explicitly, the MaM estimator suffers much more from the track confirmation latency using the hard threshold 0.5 than the JoM estimator with insignificant values of the lower margin. In addition, the track maintenance quality of the proposed JoM estimator is better than that of the MaM estimator after a small period of time from the target birth. However, as in Fig. 6(b), the proposed JoM estimator confirms track termination with larger time delay after time $k = 35$, compared to the MaM estimator.

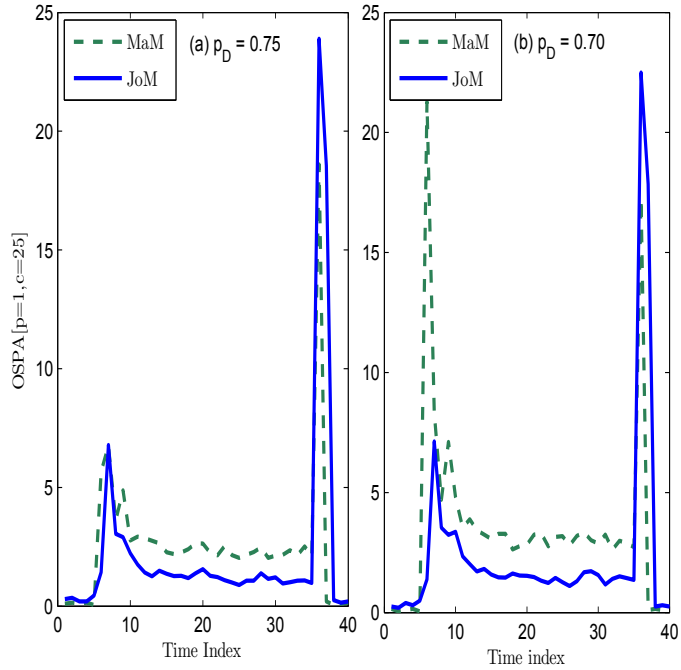

Fig. 7. 500 Monte Carlo run averages of the OSPA metric computed for the track management performance of the JoM and MaM estimators.

According to the $AR(1)$ models in Section V, time evolution of the weights in (27) for different values of detection probability is shown in Fig. 8 and Fig. 9. For considerably high detection probabilities, e.g., $p_D = 0.95$ and $p_D = 0.90$, the weights are adjusted as indicated in Section V, i.e., they monotonically approach to their means after the optimal JoTT filter converges to its steady-state with detections. However, if the detection probability is not so high or close to moderately small values, the weights are predicted based on the estimation error analysis in the optimal JoTT filter. Consequently, the linear predictions can be considered to be adaptive to the JoTT filter's performance.

VIII. Conclusions

In this paper, we have proposed an optimization algorithm to compute the optimal value of the unknown estimation constant in the JoM estimator. The optimization problem is defined in terms of two conflicting objective functions. The first objective function is defined in terms of the information theoretic sense and aims for entropy maximization by setting the estimation constant to its maximum permissible value. In contrast, the second one arises from the constraint in the definition of the JoM estimator and aims to improve the accuracy of the JoM estimates by setting the estimation constant to its minimum value determined by the probability of user's confidence level. We used a standard optimization approach that guarantees a Pareto-optimal solution.

The proposed JoM estimator is used in the JoTT filter and compared to the other MAP type multitarget estimator-called the MaM estimator. The simulation results demonstrate that the

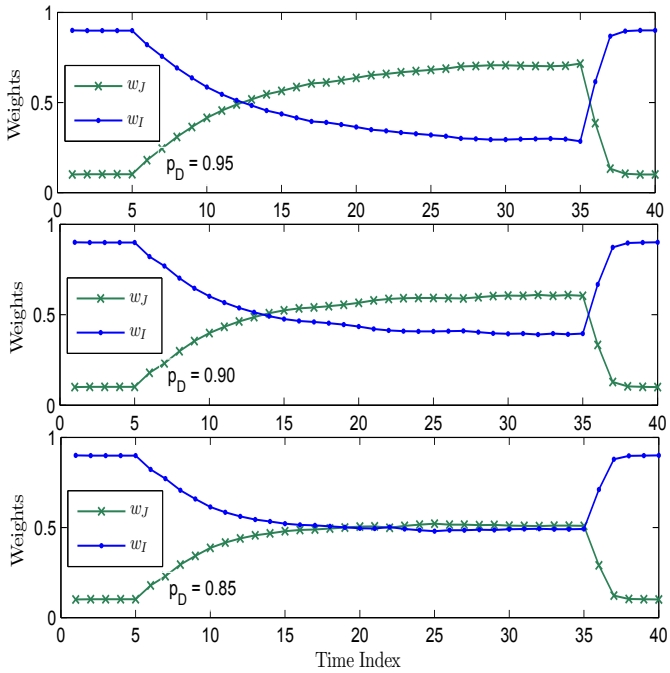

Fig. 8. 500 Monte Carlo run averages of the weights for high detection probabilities.

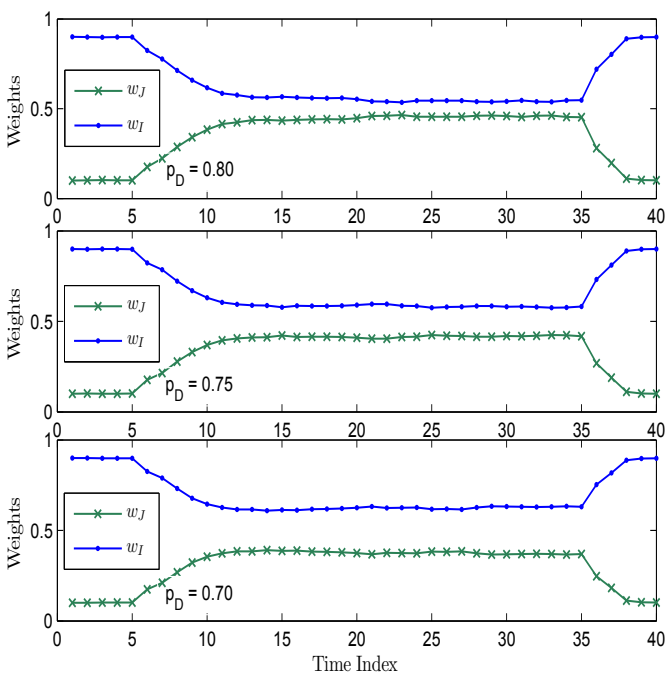

Fig. 9. 500 Monte Carlo run averages of the weights for moderately small detection probabilities.

track management performance of the proposed JoM estimator in terms of track confirmation latency, and track maintenance quality after target birth is better than that of the MaM estimator for different values of the detection probability, ranging from high to moderately small values. However, the proposed JoM estimator suffers from track termination latency more than the MaM estimator as the localization performance of the JoTT filter does deteriorate gradually after target termination.

Appendix A

To understand why selection of too small values for the JoM estimation constant does not ameliorate multitarget state estimates, quantize the FISST density $f(\{x_1, ..., x_n\})$ for all n into small and disjoint hyperspaces Δ^n with volume ε^n. Then using the relation $f(\{x_1, ..., x_n\}) \triangleq n! f(x_1, ..., x_n)$ the probability over a small hyperspace indexed by variable i, i.e., Δ_i^n [2] is computed as:

$$\begin{aligned} p_i(n) &= \frac{1}{n!} \int_{\Delta_i^n} f(\{x_1, ..., x_n\}) \, dx_1 ... dx_n, \\ &= \int_{\Delta_i^n} f(x_1, ..., x_n) \, dx_1 ... dx_n, \\ &\approx f(\hat{x}_{1_i}, ..., \hat{x}_{n_i}) \, \varepsilon^n. \end{aligned} \tag{A.1}$$

where $\hat{x}_{1_i}, ..., \hat{x}_{n_i}$ denotes the multitarget state estimates obtained from $f(\{x_1, ..., x_n\})$ in Δ_i^n. Note that if $f(\{x_1, ..., x_n\})$ is peaky over Δ_i^n, ε must be set to a small value to satisfy the following condition:

$$\sum_{n=0}^{\infty} \sum_{i:\Delta_i^n \in \mathcal{X}^n} p_i(n) \leq 1. \tag{A.2}$$

Thus, similar to the quantization of a continuous random variable, the entropy of the quantized RFS is defined as

$$H(X^{\Delta}) = -\sum_{n=0}^{\infty} \sum_{i:\Delta_i^n \in \mathcal{X}^n} p_i(n) \log(p_i(n)). \tag{A.3}$$

Upon substitution of $p_i(n) = f(\hat{x}_{1_i}, ..., \hat{x}_{n_i}) \, \varepsilon^n$ into the logarithmic function in (A.3), the entropy of the quantized RFS can be rewritten as

$$\begin{aligned} H(X^{\Delta}) &= -\sum_{n=0}^{\infty} \sum_{i:\Delta_i^n \in \mathcal{X}^n} p_i(n) \log(f(\hat{x}_{1_i}, ..., \hat{x}_{n_i}) \, \varepsilon^n), \\ &= -\sum_{n=0}^{\infty} \sum_{i:\Delta_i^n \in \mathcal{X}^n} p_i(n) \log(f(\hat{x}_{1_i}, ..., \hat{x}_{n_i})) - \\ &\quad \sum_{n=0}^{\infty} \sum_{i:\Delta_i^n \in \mathcal{X}^n} p_i(n) \log(\varepsilon^n), \end{aligned} \tag{A.4}$$

where the first term is the average self-information of the joint symmetric pdfs (i.e., $f(x_1, ..., x_n)$) over $\Delta^{(n)}$ and the second term is the average self-information of the uniform pdfs (i.e., $\mathcal{U}(x_1, ..., x_n) = \varepsilon^{-n}$) over $\Delta^{(n)}$.

For simplicity of analysis, assume that $f(\hat{x}_{1_i}, ..., \hat{x}_{n_i}) \approx 1.0$ over some hyperspaces Δ_i^n indexed by i^*. For the rest, $f(\hat{x}_{1_i}, ..., \hat{x}_{n_i}) \approx 0$ and thus from (A.1) the probability over those regions is $p_i(n) \approx 0$. In this case, using the convention $0 \log 0 = 0$ and $\log 1 = 0$, the first term in (A.4) is canceled and the entropy of the quantized RFS simplifies to

$$H(X^{\Delta}) \approx -\sum_{n=0}^{\infty} \sum_{i^*:\Delta_{i^*}^n \in \mathcal{X}^n} p_{i^*}(n) \log(\varepsilon^n), \tag{A.5}$$

Note that ε is small enough to satisfy the condition given by (A.2). Similar to typical sequences with equal probabilities in a typical set, most of the total probability is almost equally divided on some hyperspaces $\Delta_{i^*}^n$ indexed by i^*. Therefore,

selecting too small values for ε will not ameliorate the accuracy in multitarget state estimates. On the contrary, the entropy will get larger values due to the uncertainty regarding what multitarget state estimate is true.

Appendix B

For the standard Gaussian density $f(x)$ defined in $\mathbb{R}^{n_x}$, it follows from (3) that the amount of self-information associated with the outcome $(\tilde{x}_1, ..., \tilde{x}_n) \in A_n^\tau$ is

$$H(f) - \tau < -\frac{1}{n} \log f(\tilde{x}_1, ..., \tilde{x}_n) < H(f) + \tau, \quad \text{(B.1)}$$

where $\tilde{x}_1, ..., \tilde{x}_n$ are i.i.d. samples from $f(x)$, and $H(f) = 0.5 \log (2\pi e)^{n_x}$ [23].

Substituting for $-\log f(\tilde{x}_1, ..., \tilde{x}_n) = 0.5n \log (2\pi)^{n_x} + 0.5 \sum_{i=1}^n \tilde{x}_i^T \tilde{x}_i$ into (B.1) and making some algebraic manipulations yield

$$n(n_x - 2\tau) < \sum\nolimits_{i=1}^n \tilde{x}_i^T \tilde{x}_i < n(n_x + 2\tau), \quad \text{(B.2)}$$

where $\sum_{i=1}^n \tilde{x}_i^T \tilde{x}_i$ represents a thin shell around a hypersphere centered at the origin of $\mathbb{R}^{n_x}$ as claimed.

Appendix C

The nonlinear convex optimization problem in (25) is referred to as the primal problem [31]. The Lagrangian of the primal problem is written as

$$L(\tau, \lambda) = f_{o,I}(\tau) + \lambda_1 g_1(\tau) + \lambda_2 g(\tau), \quad \text{(C.1)}$$

where τ and $\lambda = (\lambda_1, \lambda_2)$ are called primal and dual variables, respectively.

According to the duality theorem, the dual problem has the same optimal solution with the primal problem if Slater's condition holds [35]. Associated with the primal problem, the dual function is defined as

$$\begin{aligned} g(\lambda) &= \min_\tau L(\tau, \lambda), \\ &= L(\tau^*, \lambda), \end{aligned} \quad \text{(C.2)}$$

where τ^* is the primal solution and the dual solutions to $g(\lambda)$, i.e., $\lambda^* = (\lambda_1^*, \lambda_2^*)$ are the Lagrange multipliers of the primal problem.

For any convex optimization problem with differentiable objective and constraint functions, the necessary and sufficient conditions to analyze the optimality of τ^*, and $\lambda^* = (\lambda_1^*, \lambda_2^*)$, are called the Karush-Kuhn-Tucker (KKT) conditions [31], [35]. That is, τ^*, and $\lambda^* = (\lambda_1^*, \lambda_2^*)$ must satisfy the following conditions

$$g_i(\tau^*) \leq 0, \quad \text{for } i = 1, 2 \quad \text{(C.3)}$$

$$\lambda_i^* \geq 0, \quad \text{for } i = 1, 2 \quad \text{(C.4)}$$

$$\lambda_i^* g_i(\tau^*) = 0, \quad \text{for } i = 1, 2 \quad \text{(C.5)}$$

and

$$\nabla_\tau L(\tau^*, \lambda^*) = \nabla_\tau f_{o,I}(\tau^*) + \sum\nolimits_{i=1}^2 \lambda_1^* \nabla_\tau \; g_i(\tau^*) = 0, \quad \text{(C.6)}$$

where (C.3) is called primal feasibility conditions of τ^*, (C.4) is called the dual feasibility conditions of $\lambda^* = (\lambda_1^*, \lambda_2^*)$, and (C.5) is called complementary slackness conditions. Thus, the last KKT condition verifies that τ^* is the global minimum point of $L(\tau, \lambda^*)$.

Based on the KKT conditions three possible cases are distinguished for optimality of τ^* , and $\lambda^* = (\lambda_1^*, \lambda_2^*)$:

1) The constraints are both inactive: this means that $\lambda_i^* = 0$, for $i = 1, 2$. Then, the optimal value of the primal variable is set to $\tau^* = 0$ to satisfy the last KKT condition as
$$\nabla_\tau L(\tau^*, \lambda^*) = \frac{n_x}{n_x - 2\tau^*} - 1 = 0. \quad \text{(C.7)}$$
2) The constraints are both active: this means that $\lambda_i^* > 0$ for $i = 1, 2$. Then, the complementary slackness conditions contradicts for optimality of τ. That is, (C.5) for $i = 1$ requires that $\tau^* = 0$, whereas (C.5) for $i = 2$ requires that $\tau^* = 0.5(n_x - \gamma_{\min})$ where $\gamma_{\min} \ll n_x$. Nevertheless, the optimal value of the primal variable becomes $\tau^* = 0$ if the probability of confidence is excessively set to $\gamma_{\min} = n_x$. Thus, the last KKT condition will have the form
$$\begin{aligned} \nabla_\tau L(\tau^*, \lambda^*) &= \frac{n_x}{n_x - 2\tau^*} - 1 - \lambda_1^* + 2\lambda_2^*, \\ &= -\lambda_1^* + 2\lambda_2^*, \end{aligned} \quad \text{(C.8)}$$
in which case, $\lambda_1^* = 2\lambda_2^*$. That is, the inequality constraint $g_2(\tau)$ turns into $g_2(\tau) : \tau \leq 0$. Then, the constraints $g_1(\tau)$ and $g_2(\tau)$ contradict each other unless they both turn into the equality constraint given by $\tau = 0$.
3) One active and one inactive constraint: this means that either $\lambda_1^* > 0$ and $\lambda_2^* = 0$ or $\lambda_1^* = 0$ and $\lambda_2^* > 0$. If $\lambda_1^* > 0$ and $\lambda_2^* = 0$, then the complementary slackness condition for $i = 1$ requires that $\tau^* = 0$ but the last KKT condition cannot be satisfied for $\tau^* = 0$. On the other hand, if $\lambda_1^* = 0$ and $\lambda_2^* > 0$, then the complementary slackness condition for $i = 2$ requires that $\tau^* = 0.5(n_x - \gamma_{\min})$ and again, the last KKT condition cannot be satisfied for $\lambda_2^* > 0$.

Consequently, the inequality constraints for the nonlinear convex problem are both inactive unless $n_x = \gamma_{\min}$. In addition, $\tau^* = 0$ is the optimal solution for the primal problem. That is, the convex objective function $f_{o,I}(\tau)$ given by (25) has a global minimum at $n_x - 2\tau^* = n_x$. Note that the inequality constraints $g_i(\tau)$, for $i = 1, 2$ are affine in addition to the convexity of $f_{o,I}(\tau)$, then Slater condition for the strong duality holds. Therefore, the strong duality indicates that the optimal solution to the primal problem $f_{o,I}(\tau)$ can be attained from the dual problem [31].

References

[1] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*, Storrs CT: YBS Publishing, 2011.

[2] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*, Norwood MA: Artech House, 2007.

[3] –, " ″Statistics 101″ for multisensor, multitarget data fusion", *IEEE Aerosp. and Electron. Sys. Magazine Part 2: Tutorials*, vol. 19, no. 1, pp. 53–64, Jan. 2004.

[4] –, "Why multi-Source, multitarget data fusion is tricky", *in Proc. of IRIS National Symposium on Sensor and Data Fusion*, May. 1999.

[5] D. F. Crouse, P. K. Willett, and Y. Bar-Shalom, "Developing a real-time track display that operators do not hate", *IEEE Trans. on Signal Processing*, vol. 59, no. 7, pp. 3441–3447, Jul. 2011.

[6] M. Baum, P. K. Willett, and U. D. Hanebeck, "MMOSPA-based track extraction in the PHD filter - a justification for k-means clustering", *in Proc. of the 53rd IEEE Conf. on Decision and Control*, CA, USA, Dec. 2014.

[7] M. Baum, P. K. Willett, and U. D. Hanebeck, "Polynomial-time algorithms for the exact MMOSPA estimate of a multi-object probability density represented by particles", *IEEE Trans. on Signal Processing*, vol. 63, no. 10, pp. 2476–2484, May. 2015.

[8] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, "A tutorial on Bernoulli filters: Theory, implementation and applications", *IEEE Trans. Signal Processing*, vol. 61, no. 13, pp. 3406–3430, Jul. 2013.

[9] B.-T. Vo, B.-N Vo, and A. Cantoni, "The cardinality balanced multi-target multi-Bernoulli filter and its implementations", *IEEE Trans. Signal Processing*, vol. 57, no. 2, pp. 409–423, Feb. 2009.

[10] E. Baser, T. Kirubarajan, and M. Efe, "Improved MeMBer filter with modeling of spurious targets", *in Proc. of 16th Int'l Conf. on Information Fusion*, Istanbul, Turkey, Jul. 2013.

[11] D. Musicki, R. Evans, and S. Stankovic, "Integrated probabilistic data association", *IEEE Trans. Automatic Control*, vol. 39, no .6, pp. 1237–1241, Jun. 1994.

[12] S. Challa, B.-N. Vo, and X. Wang, "Bayesian approaches to track existence - IPDA and random sets", *in Proc. 5th Int'l Conf. on Information Fusion*, vol.2, pp. 1228–1235, Annapolis, Maryland 2002.

[13] B.-T. Vo, "Random finite sets in multi-object filtering", PhD dissertation, School of Electrical, Electronic and Computer Engineering, University of Western Australia, Australia, Oct. 2008.

[14] A. A. Gorji, R. Tharmarasa, and T. Kirubarajan, "Performance measures for multiple target tracking problems", *in Proc. of 14th Int'l Conf. on Information Fusion*, Illinois, USA, Jul. 2011.

[15] R. L. Rothrock, and O. E. Drummond, "Performance metrics for multiple-sensor, multiple-target tracking", *in Proc. of SPIE Conf. on Signal and Data Processing of Small Targets*, Orlando, USA, Apr. 2000.

[16] X. R. Li, and Z. Zhao, "Evaluation of estimation algorithms part I: Incomprehensive measures of performance", *IEEE Aerosp. and Electron. Sys.*, vol. 42, no. 4, pp. 1340–1358, Jan. 2006.

[17] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters", *IEEE. Trans. Signal Processing*, vol. 59, no. 7, pp. 3447–3457, Aug. 2008.

[18] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo, "A metric for performance evaluation of tracking algorithms", *IEEE. Trans. Signal Processing*, vol. 59, no. 7, pp.3452–3457, Jul., 2011.

[19] P. Ngatchou, A. Zarei, and M. A. El-Sharkawi, "Pareto multi objective optimization", *in Proc. of the 13th Int'l Conf. on Intell. Syst. Application to Power Systems*, VA, USA, Nov. 2005.

[20] R. T. Marler, and J. S. Arora, "Survey of multi-objective optimization methods for engineering", *Structural and Multidisciplinary Optimization*, vol. 26, no. 6, pp.369–395, Springer, Mar. 2004.

[21] O. Grodzevich, and O. Romanko, "Discussions on normalization and other topics in multi-objective optimization", *in Proc. of Fields-MITACS Industrial Problem-Solving Workshop (FMIPW)*, Aug. 2006.

[22] R. T. Marler, and J. S. Arora, "The weighted sum method for multi-objective optimization", Structural and Multidisciplinary Optimization, vol. 41, no. 6, pp.853-862, Springer, Dec. 2009.

[23] T. M. Cover, and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York NY: J. Wiley and Sons, 2005.

[24] A. Lesne, "Shannon entropy: A rigorous mathematical notion at the crossroads between probability, information theory, dynamical systems and statistical physics", *Math. Struct. in Comp. Science*, 24 (4), in press, 2014.

[25] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge UK: Cambridge University Press, 2003.

[26] A. Hobson, and B.-K. Cheng, "A comparison of the Shannon and. Kullback information measures", *Journal of Statistical Physics*, Vol. 7, No. 4, 1973.

[27] I. Goodman, R. Mahler, and, H. Nguyen, *Mathematics of Data Fusion*, Norwell MA: Kluwer Academic Publisher, 1997.

[28] M. Rezaeian, and B.-N. Vo, "The entropy of random finite sets", *in Proc. of Int'l Symp. on Information Theory*, Seoul, Korea, 2009.

[29] D. J. Daley, and D. Vere-Jones, *An Introduction to the Theory of Point Process Volume II: General Theory and Structure*, 2nd ed. New York NY: Springer, 2008.

[30] S. Bobkov, and M. Madiman, "An equipartition property for high-dimensional log-concave distributions", *in Proc. of the 50th Annual Allerton Conf. on Commun., Control, and Comput.*, IL, USA, Oct. 2012.

[31] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge UK: Cambridge University Press, 2009.

[32] E. T. Jaynes, "On the rationale of maximum entropy methods", *Proc. IEEE*, vol. 70, pp. 939–952, Sept. 1982.

[33] S. S. Blackman, and R. F. Popoli, *Design and Analysis of Modern Tracking Systems*, Norwood MA: Artech House, 1999.

[34] P. T. Boggs, and J. W. Tolle, "Sequential quadratic programming", *Acta Numerica*, no. 4, pp.1–51, Jan. 1995.

[35] D. P. Bertsakes, *Nonlinear Programming*, 2nd ed. Belmont MA: Athena Scientific, 1999.

[36] X.-S. Yang, *Nature-Inspired Optimization Algorithms*, London UK: Elsevier, 2014.

[37] G. Kirchgssner, J. Wolters, and U. Hassler, *Introduction to Modern Time Series Analysis*, 2nd ed. New York NY: Springer, 2013.

[38] D. Clark, B.-T. Vo, and B.-N. Vo, "Forward-backward sequential Monte Carlo smoothing for joint target detection and tracking", *in Proc. of 16th Int'l Conf. on Information Fusion*, Seattle, USA, Jul. 2009.

[39] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, New York NY: J. Whiley and Sons, 2001.

[40] B.-N. Vo, and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter", *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

Erkan Baser received his B.Sc. degree in electronics engineering from Hacettepe University, Turkey in 2005 and M.Sc. degree in electrical and electronics engineering from Middle East Technical University, Turkey, in 2008.

From 2005 to 2009 he worked as a system test and algorithm development engineer for some defense companies in Turkey. Since 2013 he has been pursuing his PhD degree at McMaster University under the supervision of Dr. Kirubarajan. His research interests include point processes, filtering and estimation, sensor fusion, and multi-object tracking.

Michael McDonald received a B.Sc. (Hons) degree in applied geophysics in 1986 and an M.Sc. degree in electrical engineering in 1990, both from Queens University, Kingston, Canada. He received a Ph.D. in physics from the University of Western Ontario in London, Canada, in 1997.

He was employed at ComDev in Cambridge, Canada from 1989 through 1992 in their Space Science and Satellite Communications Departments and held a post-doctoral position in the Physics Department of SUNY at Stony Brook from 1996 through 1998 before commencing his current position as defence scientist in the radar systems section of Defence Research and Development Canada, Ottawa, Canada. His current research interests include the application of STAP processing and nonlinear filtering to the detection of small maritime and land targets as well as the development and implementation of passive radar systems.

Thia Kirubarajan (S95 M98 SM03) was born in Sri Lanka in 1969. He received the B.A. and M.A. degrees in electrical and information engineering from Cambridge University, England, in 1991 and 1993, and the M.S. and Ph.D. degrees in electrical engineering from the University of Connecticut, Storrs, in 1995 and 1998, respectively.

Currently, he is a professor in the Electrical and Computer Engineering Department at McMaster University, Hamilton, Ontario, Canada. He is also serving as an Adjunct Assistant Professor and the Associate Director of the Estimation and Signal Processing Research Laboratory at the University of Connecticut. His research interests are in estimation, target tracking, multisource information fusion, sensor resource management, signal detection, and fault diagnosis. His research activities at McMaster University and at the University of Connecticut are supported by the U.S. Missile Defense Agency, U.S. Office of Naval Research, NASA, Qualtech Systems, Inc., Raytheon Canada Ltd., and Defense Research Development Canada, Ottawa. In September 2001, he served in a DARPA expert panel on unattended surveillance, homeland defense, and counterterrorism. He has also served as a consultant in these areas to a number of companies, including Motorola Corporation, Northrop-Grumman Corporation, Pacific-Sierra Research Corporation, Lockhead Martin Corporation, Qualtech Systems, Inc., Orincon Corporation, and BAE systems. He has worked on the development of a number of engineering software programs, including BEARDAT for target localization from bearing and frequency measurements in clutter, and FUSEDAT for fusion of multisensor data for tracking. He has also worked with Qualtech Systems, Inc., to develop an advanced fault diagnosis engine.

Dr. Kirubarajan has published about 100 articles in the areas of his research interests, in addition to one book on estimation, tracking and navigation and two edited volumes. He is a recipient of Ontario Premiers Research Excellence Award (2002).

Murat Efe (S95 M99) received his B.Sc. and M.Sc. degrees in Electronics Engineering from Ankara University, Turkey in 1993 and 1994 respectively. He received his Ph.D. degree from the University of Sussex, UK in 1998. He worked as a postdoctoral research fellow at Ecole Polytechnique Federale de Lausanne in Switzerland following his Ph.D. He has been with the Electrical and Electronics Engineering Department at Ankara University since June 2000 where he both teaches and conducts research. Dr. Efe has been a consultant for several defense projects undertaken by different companies in Turkey. He was the General Co-Chair for the 16th International Conference on Information Fusion (Fusion13) that was held in Istanbul, Turkey between 9-12 July 2013. Currently he is a member of the Board of Directors of the International Society of Information Fusion and is serving as associate editor of the IEEE Transactions on Aerospace and Electronic Systems. His research interests include Kalman filtering, multi-target multi-sensor tracking, detection and estimation, cognitive radar, passive network sensing.