Generalized optimal sub-pattern assignment metric
Abu Sajana Rahmathullah, Ángel F. García-Fernández, Lennart Svensson

Abstract—In this paper, we present the generalized optimal sub-pattern assignment (GOSPA) metric on the space of sets of targets. This metric is a generalized version of the unnormalized optimal sub-pattern assignment (OSPA) metric. The difference between unnormalized OSPA and GOSPA is that, in the proposed metric, we can choose a range of values for the cardinality mismatch penalty for a given cut-off distance $c$. We argue that in multiple target tracking, we should select the cardinality mismatch of GOSPA in a specific way, which is different from OSPA. In this case, the metric can be viewed as sum of target localization error and error due to missed and false targets. We also extend the GOSPA metric to the space of random finite sets, and show that both mean GOSPA and root mean squared GOSPA are metrics, which are useful for performance evaluation.

Index Terms—OSPA, multiple target tracking, metric, GOSPA, RFS

I. INTRODUCTION

In multiple target tracking (MTT) systems, targets can appear and disappear from a scene and MTT algorithms estimate this varying number of targets based on observations from sensors [16]. In this setting, the ground truth is a set that contains the true target states and the estimate is a set that contains the estimated target states. Then, it is possible that the algorithm has not estimated all the target states, leading to missed targets. It is also possible that the estimated set has false targets, i.e., estimated targets that are not close to any of the targets in the ground truth. It is highly desired to assess and compare the performance of MTT algorithms in a mathematically sound way, which boils down to computing the similarity between the ground truth and the estimated sets.

It is important that this similarity measure is a metric in the mathematical sense such that it satisfies non-negativity, definiteness, symmetry and triangle inequality [6, pp. 21]. These properties are relevant to be able to quantify the closeness between sets of objects. Specifically, the need for the triangle inequality is well motivated in [20, Sec. 1] and [17, Sec. 6.2.1]: if there is an estimate $A$ which is close to the ground truth and another estimate $B$ which is close to $A$, then the triangle inequality ensures that the estimate $B$ is also close to the ground truth, which agrees with the intuition.

The problem of designing suitable metrics between sets has been studied in the MTT literature. There are metrics such as the Hausdorff metric [12], the Wasserstein metric [12] (also referred to as optimal mass transfer metric in [24]) and the optimal sub-pattern assignment (OSPA) [24] metric to compute this distance between two sets of target states. It has been discussed with examples in [24] that the Hausdorff metric is insensitive to cardinality mismatches and to outliers. The Wasserstein metric on the other hand handles these issues better, but it lacks a consistent physical interpretation when the states have different cardinalities, besides other non-intuitive behaviors [24]. In [24], the OSPA metric was proposed to address these issues. This metric is also easily computed and has been used widely in the tracking community for performance evaluation and to obtain estimators [2], [9], [25]. The OSPA metric has also been adapted to handle sets of labeled targets [21].

Even though OSPA penalizes the cardinality mismatch, it fails to quantify it properly in certain scenarios. For instance, let us consider a scenario without any targets. Suppose one estimator indicates that there is one target (i.e., a false target) and another indicates that there are 1000 targets. In both cases, the OSPA metric has the same value. So according to OSPA, both estimators are equally accurate, which is counterintuitive. This issue is due to the normalization with the maximum cardinality in the OSPA definition. The OSPA metric without the normalization has been used in [25, Sec. IV] to obtain minimum mean OSPA estimate, such that the error adds over targets in the mean error. Even though [26] makes use of the unnormalized OSPA as a cost function, it has not been proved that it is a metric.

We argue that OSPA possesses a second weakness related to its expression for the cardinality mismatch. In MTT, detections, false and missed targets are more important concepts than cardinality mismatch [4, Chap. 13]. Therefore, it seems more reasonable to develop a metric that considers the errors due to these components. Instead, the OSPA metric considers localization and cardinality errors, which can lead to undesirable results. For example, consider a scenario with a target in the ground truth and two estimators - one estimator reported no target and the other estimator reported a false target (a target which is far from the real target). Both these estimators have identical OSPA metric, whereas the first estimate that only has a missed target is arguably better than the second one that has a false target in addition.

In this paper, we propose a metric, named generalized OSPA (GOSPA). In relation to unnormalized OSPA, GOSPA
includes an additional parameter that enables us to select the cardinality mismatch cost from a range of values that also includes the one used in OSPA. As desired, we show that for a specific selection of this parameter, the GOSPA metric is a sum of localization errors for the localized targets and a penalty for missed and false targets.

In practice, the ground truth is not known but is modeled as a random finite set (RFS) [16]. Algorithm evaluation is usually performed by averaging over the error of the estimates provided by the algorithm over the measurements. This implies that the estimates are also RFSs, so it is useful to have a metric that considers RFSs rather than finite sets. In the literature, there is no formal treatment of this problem to our knowledge. In this paper, we fill this gap by showing that the mean GOSPA and root mean squared GOSPA are metrics for RFSs of targets. This is an important result for performance evaluation and the design of estimators.

The outline of the rest of the paper is as follows. In Section II, we discuss the normalization issue in OSPA and illustrate with examples, how without normalization, OSPA works better in handling cardinality mismatch. In Section III, we proceed to generalize the unnormalized OSPA. We also discuss the roles of the different parameters in this section. In Section IV, we present and argue for the most appropriate choice of the GOSPA metric for MTT. In Section V, we discuss the roles of the different parameters in this section. In Section VI, we illustrate via simulations that the proposed choice of GOSPA provides expected results compared to the OSPA and unnormalized OSPA counterparts. Finally, conclusions are drawn in Section VII.

II. BACKGROUND

The main motivation for using a metric in estimation is to compare the estimates from different algorithms with the ground truth to assess their performance. To this purpose, the OSPA metric was introduced in [24] to compute the distance between finite sets of ground truth and estimates. Before we review the OSPA metric and discuss the normalization issues, let us first introduce the formal definition of a metric and some useful notations.

A. Metric

Definition 1. A metric $d_A(\cdot, \cdot)$ on a set $A$ is a function that satisfies the following properties for any $x, y, z \in A$ [22, Sec. 2.15]:

i. Non-negativity: $d_A(x, y) \geq 0$.
ii. Symmetry: $d_A(x, y) = d_A(y, x)$.
iii. Definiteness: $d_A(x, y) = 0 \iff x = y$.
iv. Triangle inequality: $d_A(x, y) \leq d_A(x, z) + d_A(z, y)$.

We emphasize here that the triangle inequality property, despite its abstractness, has a major practical importance in algorithm assessment [17, Sec. 6.2.1]. Suppose there are two estimates $y$ and $z$ for a ground truth $x$. Let us assume that according to $d_A$, the estimate $z$ is close to the ground truth $x$ and is also close to the other estimate $y$. Then, according to intuition, the second estimate $y$ should also be close to the ground truth $x$. This property is ensured by the triangle inequality property.

The triangle inequality also has practical implications to ensure the quality of approximate optimal estimators. Consider $x$ to be the ground truth and $z$ to be the optimal estimate, according to certain criterion. Let us assume that it is difficult to compute the optimal estimate $z$ and so we resort to an approximation $y$ of the optimal $z$. This happens often in practice. If the triangle inequality does not hold, that would mean that even if we have a good estimate $y$, close to the optimal $z$, which in turn is close to the ground truth $x$, it is possible that the distance from $y$ to the ground truth $x$ is high. This property is clearly not desirable.

B. OSPA metric

Let $d(\cdot, \cdot)$ denote a metric in $\mathbb{R}^N$ such that $d(x, y)$ is the distance between $x, y \in \mathbb{R}^N$ and let $d^{(c)}(x, y) = \min(d(x, y), c)$ be the cut-off metric associated with $d(x, y)$ [17, Sec. 6.2.1]. Let $\Pi_n$ be the set of all permutations in $\{1, \ldots, n\}$ for any $n \in \mathbb{N}$. Any element $\pi \in \Pi_n$ is a sequence $(\pi(1), \ldots, \pi(n))$. Let $X, Y$ be finite subsets of a bounded observation window $W \subset \mathbb{R}^N$, where $|A|$ denotes the cardinality of a set $A$. For $1 \leq p < \infty$ and $|X| \leq |Y|$, OSPA [24] is defined as

$$d^c_p(X, Y) \triangleq \left( \frac{1}{|Y|} \min_{\pi \in \Pi_Y} \sum_{i=1}^{|X|} d^{(c)}(x_i, y_{\pi(i)})^p + c^p(|Y| - |X|) \right)^{\frac{1}{p}}. \quad (1)$$

For $|X| > |Y|$, $d^c_p(X, Y) = d^c_p(Y, X)$. The $\infty$-OSPA is defined as

$$d^{(c)}(X, Y) \triangleq \begin{cases} \min_{\pi \in \Pi_Y} \max_{1 \leq i \leq |X|} d^{(c)}(x_i, y_{\pi(i)}) & |X| = |Y| \\ c & \text{otherwise} \end{cases}. \quad (2)$$

The metric properties of OSPA are proved in [24].

The OSPA metric in (1) is viewed as a sum of two components: the localization error contributed by the term $\sum_{i=1}^{|X|} d^{(c)}(x_i, y_{\pi(i)})^p$ and the cardinality error by the term $c^p(|Y| - |X|)$. We proceed to review the role of the different parameters in these error components. The parameter $p$ in OSPA determines the severity of penalizing the outliers in the localization component. That is, for a fixed value of $c$, the larger the value of $p$ is, the larger the pairs with large $d^{(c)}(x, y)$ get penalized. Also, the cardinality cost is penalized more, assuming that $c > 1$. In the limit when $p = \infty$, the maximum value of the cost is $c$ for
any cardinality mismatch. The parameter $c$ determines the maximum allowable localization error between a pair of points. It is also the cost for any cardinality mismatch.

Computing the OSPA metric is equivalent to solving a 2-D assignment problem. A close inspection of (3) shows that the OSPA metric involves finding the optimal choice of pair of points that minimizes the overall distance, given the distances for every pair of points. This problem is known as the 2-D assignment problem [3], [13]. Solutions to 2-D assignment problems have been used extensively by the tracking community for the data association problem that arises in the measurement update step with the radar point detection measurement model [4]. The same solution applies to computing the OSPA metric. Several methods exist: the modified auction algorithm [3], [18], Hungarian (also known as Munkres) [5], [15] and the Jonker-Volgenant-Castanon (JVC) [13] algorithms. All these algorithms have an asymptotic complexity of $O(|X|^2|Y|)$ with $|X| \leq |Y|$ [4].

C. Drawback with the normalization in OSPA

In this section, we illustrate that the normalization $\frac{1}{|Y|}$ in the OSPA metric can lead to counter-intuitive results. This effect can be illustrated with a simple example.

Example 1. Let us say the ground truth is $X = \emptyset$ and we have estimates $Y_j = \{y_1, \ldots, y_j\}$ from different algorithms indexed with $j \in \mathbb{N}$. Intuitively, the larger the value of $j$, the farther away we are from the ground truth and therefore, the distance metric should increase with $j$. However, the OSPA metric is $c$ for any $j \geq 1$. That is, according to the OSPA metric all these algorithms are equally accurate, which is not the desired evaluation in MTT.

This undesirable property of the OSPA metric is due to the normalization. However, if we remove this normalization from OSPA, the distance is $j^2 c$, which increases with $j$. This example is a clear motivation as to why the normalization should be removed from the OSPA metric to evaluate MTT algorithms. We refer to the OSPA metric without normalization as ‘unnormalized OSPA’ in the rest of this paper. The unnormalized OSPA was used in [25, Sec. IV] to obtain minimum mean OSPA estimate, such that “the measure adds over targets rather than averaging over targets”. However, as mentioned in the introduction, it has not been previously proved it is a metric, but we will present a proof of this as a particular case of the GOSPA metric introduced in the next section.

III. GENERALIZED OSPA METRIC

In this section, we present the generalized OSPA metric to measure the distance between finite sets of targets. To be precise, this metric is a generalization of the unnormalized OSPA metric. As we will see, the GOSPA metric contains an additional tuning parameter that considers a wider range of cardinality penalties than the OSPA metric.

Definition 3. Let $X = \{x_1, \ldots, x_{|X|}\}$ and $Y = \{y_1, \ldots, y_{|Y|}\}$ be finite subsets of $\mathbb{R}^N$ and let $0 < \alpha \leq 2$. For $1 \leq p < \infty$ and $|X| \leq |Y|$, the generalized OSPA function is defined as

$$d_p^{(c, \alpha)}(X, Y) \triangleq \left( \min_{\pi \in \Pi_{|Y|}} \sum_{i=1}^{|X|} \left( \min_{j \in |Y|} d_\pi(x_i, y_{\pi(i)})^p + \frac{c^p}{\alpha} (|Y| - |X|) \right) \right)^{\frac{1}{p}} \ .$$

(3)

If $|X| > |Y|$, $d_p^{(c, \alpha)}(X, Y) \triangleq d_p^{(c, \alpha)}(Y, X)$.

The corresponding GOSPA for $p = \infty$ is

$$d_\infty^{(c, \alpha)}(X, Y) \triangleq \min_{\pi \in \Pi_{|Y|}} \max_{1 \leq i \leq |X|} d_\pi(x_i, y_{\pi(i)}) \ .$$

(4)

Proposition 1. The GOSPA functions in (3) and (4) of Definition 3 are metrics.

Proof: It can be seen from the definition that the non-negativity, symmetry and definiteness holds. The triangle inequality is proved in Appendix A.

Remark. The region, where the GOSPA metric is defined, is over finite subsets of entire $\mathbb{R}^N$. On the contrary, in OSPA, we require the definition of a bounded window $W \subset \mathbb{R}^N$, see Definition 2. This bounded window $W$ is necessary in the triangle inequality proof of the OSPA metric in which dummy points chosen outside $W$ are used [24, App.]. In the GOSPA proof in Appendix A, we do not use the dummy points. The GOSPA proof includes the unnormalized OSPA and it can also be extended to OSPA so that the constraint $W$ is not necessary, see Appendix B.

A. Discussion

In GOSPA, similar to OSPA, it is possible to view the overall distance as the sum of localization and cardinality error components. In this subsection, we analyze this split up of GOSPA and discuss how the parameters $c$, $p$ and $\alpha$ influence these costs.

The role of the parameter $p$ in GOSPA is similar to that in OSPA [24]. A large $p$ penalizes outliers and cardinality mismatch severely. The role of the parameter $c$ is partly the same as in OSPA that it is the cut-off on the base distance. However, for the cardinality mismatch, the error contribution is $\frac{c^p}{\alpha}$ in GOSPA compared to $c^p$ in OSPA. The parameter $\alpha$ determines the cardinality error in relation to the maximum localization error.

The connection between the two parameters $c$ and $\alpha$ to these costs is easier to analyze when we set $p = 1$. In this case, the upper bound on $\alpha$, i.e., $\alpha = 2$ implies that the penalty for cardinality mismatch should at least be $\frac{c}{2}$ for every additional point in the larger set. In other words, the cost for cardinality mismatch should at least be half the maximum allowable localization error. If we make this cost smaller, we no longer have a metric as the triangle inequality
is not met. On the other hand, for a localization limit \( c \), the cardinality mismatch penalty can be set arbitrarily high by changing \( \alpha \).

We want to highlight two choices of \( \alpha \). When \( \alpha = 1 \), GOSPA reduces to unnormalized OSPA, thus proving that the unnormalized OSPA is a metric. As will be explained in the next section, when \( \alpha = 2 \), the GOSPA metric can be viewed as the sum of the localization errors of the localized targets plus the error due to missed and false targets. We will also argue that the penalty for cardinality mismatch (or missed/false target), \( \frac{c}{2} \), with this selection is more suitable for MTT than \( \alpha = 1 \) in which the cost for cardinality mismatch is the same as the cut-off \( c^p \).

IV. CHOICE OF \( \alpha \) IN MTT

In this section, we argue that the choice of \( \alpha = 2 \) in GOSPA is the most appropriate for analyzing the estimation error in multi object systems in terms of localized, missed and false targets. We also propose an alternative but equivalent description for the GOSPA metric for this choice of \( \alpha \). This new definition decomposes the distance metric as the sum of localization error due to localized targets and the error due to missed and false targets. For the sake of discussion, we set \( p = 1 \) in this section; the discussions are easily extended for any value of \( p \).

A. Motivating the choice of \( \alpha = 2 \)

In an MTT setting, suppose we have a ground truth \( X \) and estimate \( Y \). Consider two target states \( x \in X \) and \( y \in Y \), such that all the points in \( Y \) are far from \( x \) and all the points in \( X \) are far from \( y \). This implies that \( d(x, y) > c \), which in turn implies \( d(c)(x, y) = c \). In this case, we argue that the target \( x \) has been missed and that the estimator has instead presented a false target \( y \), and thus, both the points \( x \) and \( y \) are ‘unlocalized’. The value of \( c \) determines how far away they must be to be considered unlocalized. We refer to the point \( x \) in the ground truth as a ‘missed’ target and the point \( y \) in the estimate as a ‘false’ target for apparent reasons. On the other hand, we refer to the pair \((x, y)\) as ‘localized’ if \( d(c)(x, y) < c \) and \( x \) and \( y \) are matched according to the optimal permutation function.

According to OSPA and GOSPA, the cost contribution by a pair of missed target \( x \) and false target \( y \), associated to each other is \( d(c)(x, y) = c \). That is, the cost for a pair of unlocalized targets \( x \) and \( y \) is \( c \). This implies that the cost for a single unlocalized target should be \( \frac{c}{2} \). Intuitively this should be the value independently of whether or not an unlocalized target is associated to a target in the other set, because strictly speaking we should not associate target estimates and targets that are sufficiently far from each other. However, in OSPA, there is a disparity in the contribution by the unlocalized targets depending on whether they are part of the localization error component or the cardinality mismatch component in (1). The error contribution of an unlocalized target is \( \frac{c}{2} \) if it is accounted for in the localization component, as there is a cost \( c \) shared by this target and the associated unlocalized target in the other set. On the other hand, the error contribution is \( c \) if the unlocalized target is accounted for in the cardinality mismatch component. A similar disparity is also inherent to GOSPA when \( \alpha \neq 2 \).

We argue that \( \alpha = 2 \) is suitable in the GOSPA metric. With this choice, the contribution to the error by a pair of unlocalized targets \( x, y \) with \( d(c)(x, y) = c \), is \( c \). This term can be split up and viewed as \( x \) and \( y \) contributing \( \frac{c}{2} \) each. This is also in lines with the \( |Y| - |X| \) (assuming \( |Y| \geq |X| \)) false targets in \( Y \) that will contribute \( \frac{c}{2} \) each. That is, any unlocalized (missed or false) target always contributes \( \frac{c}{2} \) to the distance and we thus overcome the disparity inherent to other choices of \( \alpha \).

Given this insight, it is fair to look at the distance between \( X \) and \( Y \) as the sum of localization error due to the localized targets and error due to missed and false target. Consequently, in the rest of this section, we use the term ‘localization error’ to refer to the localization error due to the localized targets. This splitting of the error into localization error and error due to false and missed targets is illustrated in the example considered next.

Example 2. Consider the case where the ground truth is \( X = \{x_1, x_2\} \) and there are two estimates \( Y_a = \{y_1, y_2\} \) and \( Y_b = \{y_1\} \). The two estimates along with the ground truth are illustrated in Figures 1(a) and 1(b). Clearly, besides the localization error between \( x_1 \) and \( y_1 \), the estimate \( Y_a \) has a missed target \( x_2 \) and a false target \( y_2 \), whereas \( Y_b \) has just a missed target. So, according to intuition, \( Y_b \) in Figure 1(b) is a better estimate than \( Y_a \) in Figure 1(a). However, the OSPA metric \( d^{c}(X, Y) \) and the unnormalized OSPA metric \( d^{(1)}(X, Y) \) are \( \Delta + \frac{c}{2} \) and \( \Delta + c \) respectively for both estimates. On the contrary, the GOSPA metric shows a trend that agrees with the intuition, i.e., \( d^{(2)}(X, Y_a) = \Delta + c \) is larger than \( d^{(2)}(X, Y_b) = \Delta + \frac{c}{2} \). □

\begin{figure}
\centering
\includegraphics[width=\textwidth]{example1.png}
\caption{\(x\)'s are the points in set \( X \) and \( \bullet\)'s are the points in \( Y \). The permutation associations are shown using the dashed lines and the cut-off distances are shown on top of them. In this illustration, \( \Delta < c \).}
\end{figure}
B. Reformulating GOSPA for $\alpha = 2$

We propose an alternate expression for GOSPA when $\alpha = 2$ using assignment functions instead of permutation functions. Now, the breakdown of the metric as the sum of localization error and the error from false and missed target is explicit, which we consider insightful.

As mentioned before, the 2D assignment functions have already been used extensively in the tracking community for the data association problems [4, Sec. 6.5]. There are several ways of representing this assignment functions, the choice of which is often based on the application. We use assignment (also referred to as matching) set representation [23, Chap. 17]. An assignment set $\gamma$ between the pair of finite index sets $\{1,\ldots,|X|\}$ and $\{1,\ldots,|Y|\}$ is a set that has the following properties:

1) $\gamma \subseteq \{1,\ldots,|X|\} \times \{1,\ldots,|Y|\}$.
2) $(i,j), (i,j') \in \gamma \implies j = j'$.
3) $(i,j), (i',j) \in \gamma \implies i = i'$.

The second and third properties ensure that every $i$ and $j$ gets at most one assignment. Let $\Gamma^{(\{X\},\{Y\})}$ be the set of all such possible 2D assignment sets.

Lemma 2. We can rewrite the GOSPA metric for $\alpha = 2$ and $1 \leq p < \infty$ using the 2D assignment set as

$$d_p^{\{\alpha\}}(X,Y) \triangleq \min_{\gamma \in \Gamma^{(\{X\},\{Y\})}} \left( \sum_{(i,j) \in \gamma} d(x_i, y_j)^p + \frac{\varepsilon^p}{2} (|X| + |Y| - 2|\gamma|) \right)^{\frac{1}{p}}.$$  (5)

Note that the notion of cut-off metric $d^{\{\alpha\}}(\cdot,\cdot)$ is not needed in this representation. The above definition does not have an explicit cardinality mismatch term. The first term $\sum_{(i,j) \in \gamma} d(x_i, y_j)$ is the localization error and, the second term $\frac{\varepsilon^p}{2} (|X| + |Y| - 2|\gamma|)$ is the error due to the $|X| - |\gamma|$ missed targets and $|Y| - |\gamma|$ false targets, considering $X$ as the ground truth.

V. PERFORMANCE EVALUATION OF MTT ALGORITHMS

In the previous sections, we studied metrics between finite sets of targets. It was then implicitly assumed that the ground truth and the estimates are deterministic. However, MTT is often formulated as a Bayesian filtering problem where the ground truth is an RFS and the estimates are sets, which depend deterministically on the observed data [16]. For performance evaluation, in many cases, we average over several realizations of the data, so estimates are RFSs as well. Therefore, evaluating the performance of several algorithms is in fact a comparison between the RFS of the ground truth and the RFSs of the estimates. As in the case of deterministic sets [17, pp. 142], it is highly desirable to establish metrics for RFSs for performance evaluation.

In this section, we show that the average GOSPA, $E[d_p^{\{\alpha\}}(X,Y)]$ is a metric between RFSs. In fact, we show that a generalized version, $\sqrt[p]{E[d_p^{\{\alpha\}}(X,Y)^p]}$, is a metric. The case when $p = 2$ can be viewed as an extension of the root mean squared error (RMSE) metric between random vectors [19, Sec. 2.2] to RFSs.

Metrics have already been proposed and studied in the context of random vectors in $\mathbb{R}^N$. If we have a metric in $\mathbb{R}^N$, we have a metric on random vectors in $\mathbb{R}^N$ by taking the expected value [19, Sec. 2.2]. Then, a natural choice is to compute the average Euclidean distance, $E[\|x - y\|_2] \triangleq \int \int \|x - y\|_2 f(x,y) \, dx \, dy$, as a metric on random vectors, where $\|x - y\|_2$ is the Euclidean distance and $f(x,y)$ is the joint density between random vectors. However, a disadvantage is that it is more challenging to construct estimators based on it compared to the RMSE metric, which we discuss next.

The RMSE, $\sqrt{E[\|x - y\|_2^2]}$ is a commonly used metric between random vectors [19, Sec. 2.2]. To obtain the minimum RMSE estimate in practice, it is equivalent and common to minimize the mean squared error (MSE) $E[\|x - y\|_2^2]$, although the MSE and squared Euclidean distance $\|x - y\|_2^2$ are not metrics.

In the RFS case, there are estimators that are obtained by minimizing the mean squared OSPA [1], [7], [8], [11] for $p = 2$. This strategy is equivalent to minimizing the root mean squared OSPA metric. Note again, that squared OSPA and mean squared OSPA are not metrics, although their minimum coincides with the minimum of the root mean squared OSPA. The metric property of mean OSPA and root mean squared OSPA has not been established in the literature.

Similar to the Euclidean metric for vectors, one can use the GOSPA metric defined over finite sets to define metrics over RFSs. Following the approaches in the random vector case, root mean squared GOSPA and mean GOSPA seem like natural extensions to RFSs. Below, we establish that the more general case $\sqrt[p]{E[d_p^{\{\alpha\}}(X,Y)^p]}$ is a metric for GOSPA, which also includes the unnormalized OSPA metric.

Proposition 3. For $1 \leq p < \infty$, $c > 0$ and $0 < \alpha \leq 2$,

$$\sqrt[p]{E[d_p^{\{\alpha\}}(X,Y)^p]}$$
is a metric for any pair of RFSs $X$ and $Y$.

Proof: The proof of this proposition is provided in Appendix C. ■

Thus, the above proposition proves that mean GOSPA ($p = 1$) and root mean squared GOSPA with ($p = 2$) are also metrics. One can also easily extend the proof of the proposition to show that the mean OSPA and root mean squared OSPA are also metrics.

For the GOSPA analogue of RMSE, one can set $p = 2$ and use Euclidean distance for $d(\cdot,\cdot)$. Similar to the minimum
MSE estimators in random vectors, one can equivalently use the mean squared GOSPA $\mathbb{E} [d_2^{(c,2)}(X, Y)^2]$ for obtaining RFS estimators.

VI. ILLUSTRATIONS

In this section, we illustrate with examples how the mean GOSPA and root mean squared GOSPA metrics for $\alpha = 2$ present values that agree with the intuition compared to the OSPA and unnormalized OSPA metrics. We illustrate this for varying number of missed and false targets in the estimates.

In a Bayesian setting, the ground truth has a prior probability density function (PDF) and is observed through measurements. As previously mentioned, different estimators provide deterministic estimates based on these observed measurements. However, algorithm assessment is usually done by Monte Carlo simulation by averaging the errors with different measurements. This implies that ground truth and estimates are RFSs and we want to determine which estimate is the closest to the ground truth. Rather than providing a full simulation with prior and measurements, we assume that the ground truth and estimates are specific RFSs, which are easy to visualize and are useful to illustrate the major aspects of the proposed metrics.

We consider ground truth $X$ (See Figure 2) which is a multi-Bernoulli RFS composed of two independent Bernoulli RFSs, each with existence probability 1. The probability densities of the individual RFSs are Gaussian distributed with mean vectors $[-6, -6]'$ and $[0, 3]'$ and identity covariance matrices. Therefore, there are always two target present which are distributed independently with their corresponding PDFs.

We consider several estimate scenarios $Y$ for this ground truth with varying number of missed and false targets. The scenarios are illustrated in Figures 3(a) through 3(l). In all the cases, the estimate $Y$ is also a multi-Bernoulli RFS [17, Sec. 4.3.4]. In scenarios, where only one target in the ground truth is localized, the estimate $Y$ includes Bernoulli RFS with existence probability 1 and Gaussian density with mean $[-6.7, -5.1]'$ and identity covariance matrix. In scenarios where both targets in the ground truth are localized, the estimate $Y$ includes Bernoulli RFSs with existence probabilities 1 and Gaussian densities with mean vectors $[-6.7, -5.1]'$ and $[-1.8, 2.9]'$ and identity covariance matrices.

In scenarios, where the estimate reports false targets, the estimates have Bernoulli RFS with Gaussian density whose mean value is randomly generated away ($> c$) from the ground truth.

The results of these scenarios are tabulated in Table I. All the results are averaged over 1000 Monte-Carlo points obtained from the densities of the ground truth and estimates. The value of $c$ is set as 8. Euclidean metric is used as the base distance $d(\cdot, \cdot)$. The figures and the table have estimates with increasing number of missed targets when traversed across columns and increasing number of false targets when traversed across rows.

Let us first analyze the behavior of the different metrics for varying number of missed targets. Intuitively, as one traverses across rows, the distance between the RFSs should increase with increasing number of missed targets. This trend is observed with the GOSPA for $\alpha = 2$ and the OSPA metric for both $p = 1$ and $p = 2$. But, the unnormalized OSPA metric shows strange behaviors when there are false targets in the scenarios (the entries with red text in Table I). This is explained when we look at the unnormalized OSPA expression for these scenarios. If $n$ and $m$ are the number of false and missed targets, and $d_1 < c$ and $d_2 < c$ are the cut-off distances for the targets in the ground truth when localized, then the unnormalized OSPA when $n \geq 2$ is

$$d_p^{(c,1)} = \begin{cases} (nc^p)^{\frac{1}{p}} & m = 2 \\ (d_1^p + nc^p)^{\frac{1}{p}} & m = 1 \\ (d_1^p + d_2^p + nc^p)^{\frac{1}{p}} & m = 0 \end{cases}$$

For $n = 1$, $d_p^{(c,1)}$ takes the values $(2c^p)^{\frac{1}{p}}$, $(d_1^p + c^p)^{\frac{1}{p}}$ and $(d_1^p + d_2^p + c^p)^{\frac{1}{p}}$ respectively. Clearly, for $n \geq 2$, $d_p^{(c,1)}$ decreases with increasing number of missed targets $m$ for fixed $n$, which is not desirable.

Let us now analyze the behavior of the metrics for varying number of false targets. As the number of false targets increases, the metric should increase according to intuition. This intuitive trend is displayed by GOSPA for $\alpha = 2$ again. This trend is also observed in unnormalized OSPA, even though it is not a strict increase in the metric values (the entries with blue text in Table I). The OSPA metric on the other hand shows counter intuitive behavior as we discussed in Example 1 in Section II. When both the targets are missed, the OSPA metric is constant for varying number of false targets. Also, for the case with one missed target, OSPA and unnormalized OSPA have the same metric values when there are no false targets and when there is one false target. This trend is similar to the behavior we observed in Example 2.
Figure 3: The figures illustrate different scenarios showing the estimates. The red points represent samples from the Bernoulli RFS in $Y$ and the circles represent the data and the 2-sigma Gaussian ellipses. The dashed circles denote the $c$-radius from the center of these Gaussian components.
Table I: Table capturing the trends shown by mean metric and root mean squared metric for varying number of missed and false targets.

<table>
<thead>
<tr>
<th># false targets</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GOSPA</strong> for $p=2$</td>
<td>4.55</td>
<td>6.05</td>
<td>8</td>
<td>3.60</td>
<td>6.10</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>8.62</td>
<td>10.04</td>
<td>12</td>
<td>6.72</td>
<td>8.32</td>
<td>9.79</td>
</tr>
<tr>
<td>3</td>
<td>16.52</td>
<td>18.07</td>
<td>20</td>
<td>10.42</td>
<td>11.54</td>
<td>12.64</td>
</tr>
<tr>
<td>10</td>
<td>44.49</td>
<td>46.05</td>
<td>48</td>
<td>18.23</td>
<td>18.90</td>
<td>19.59</td>
</tr>
<tr>
<td><strong>OSPA</strong></td>
<td>2.27</td>
<td>5.02</td>
<td>8</td>
<td>2.55</td>
<td>5.88</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>4.20</td>
<td>5.02</td>
<td>8</td>
<td>5.07</td>
<td>5.88</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>5.70</td>
<td>6.51</td>
<td>8</td>
<td>6.39</td>
<td>7.02</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>7.04</td>
<td>7.45</td>
<td>8</td>
<td>7.37</td>
<td>7.65</td>
<td>8</td>
</tr>
<tr>
<td><strong>Unnormalized root OSPA</strong></td>
<td>4.55</td>
<td>10.04</td>
<td>16</td>
<td>3.60</td>
<td>8.32</td>
<td>11.31</td>
</tr>
<tr>
<td>1</td>
<td>12.62</td>
<td>10.04</td>
<td>16</td>
<td>8.79</td>
<td>8.32</td>
<td>11.31</td>
</tr>
<tr>
<td>10</td>
<td>84.49</td>
<td>82.05</td>
<td>80</td>
<td>25.54</td>
<td>25.40</td>
<td>25.29</td>
</tr>
</tbody>
</table>

VII. CONCLUSIONS

In this paper, we have presented the GOSPA metric, a generalized version of the unnormalized OSPA, which allows for a range of values for the cardinality mismatch penalty. We have also argued that the GOSPA for $\alpha = 2$, is the best selection for estimation in MTT to view the distance as sum of terms contributed to the localization of the localized targets and the penalty for the false and missed targets. In addition, we have extended GOSPA to compute metric in the space of RFSs, which is what we use for performance evaluation of algorithms via simulations. We have shown that both mean GOSPA and root mean squared GOSPA are metric between RFSs.

APPENDIX A

PROOF OF THE TRIANGLE INEQUALITY OF GOSPA

In the proof, an extension of Minkowski’s inequality [14, pp. 165] to sequences of different lengths by appending zeros to the shorter sequence is used. Let us say we have two sequences $(a_i)_{i=1}^m$ and $(b_i)_{i=1}^n$ such that $m \leq n$. We extend the sequence $(a_i)$ such that $a_i = 0$ for $i = m+1, \ldots, n$. Then, using Minkowski’s inequality on this extended sequence we get that

$$
\left( \sum_{i=1}^{m} |a_i| + |b_i|^p + \sum_{i=m+1}^{n} |b_i|^p \right)^{\frac{1}{p}} \leq \left( \sum_{i=1}^{m} |a_i|^p + \sum_{i=1}^{n} |b_i|^p \right)^{\frac{1}{p}}
$$

for $1 \leq p < \infty$. We use this result several times in our proof.

We would like to prove the triangle inequality:

$$
d_p^{(c,\alpha)}(X,Y) \leq d_p^{(c,\alpha)}(X,Z) + d_p^{(c,\alpha)}(Y,Z)
$$

for any three RFSs $X$, $Y$ and $Z$. The proof is dealt in three cases based on the values of $|X|$, $|Y|$ and $|Z|$. Without loss of generality, we assume $|Y| \geq |X|$ in all the three cases, since GOSPA is symmetric in $X$ and $Y$.

Case 1: $|X| \leq |Y| \leq |Z|$

For any $\pi \in \Pi_{|Y|}$,

$$
d_p^{(c,\alpha)}(X,Y) \leq \left( \sum_{i=1}^{|X|} d^{(c)}(x_i, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right)^p + \frac{c^p}{\alpha} \left( |Y| - |X| \right)^{\frac{1}{p}}.
$$

(9)

Using the triangle inequality on the cut-off metric $d^{(c)}(\cdot, \cdot)$, we get that for any $\pi \in \Pi_{|Y|}$ and for any $\sigma \in \Pi_{|Z|}$,

$$
d_p^{(c,\alpha)}(X,Y) \leq \left( \sum_{i=1}^{|X|} d^{(c)}(x_i, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right)^p + \frac{c^p}{\alpha} \left( |Y| - |X| \right)^{\frac{1}{p}} + \sum_{i=|X|+1}^{|Y|} d^{(c)}(z_{\sigma(i)}, y_{\pi(i)})^p + \frac{c^p}{\alpha} \left( |Z| - |Y| \right)^{\frac{1}{p}}
$$

(11)

$$
= \left( \sum_{i=1}^{|X|} d^{(c)}(x_i, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right)^p + \frac{c^p}{\alpha} \left( |Y| - |X| \right)^{\frac{1}{p}} + \sum_{i=|X|+1}^{|Y|} d^{(c)}(z_{\sigma(i)}, y_{\pi(i)})^p + \frac{c^p}{\alpha} \left( |Z| - |Y| \right)^{\frac{1}{p}}
$$

(12)

$$
\leq \left( \sum_{i=1}^{|X|} d^{(c)}(x_i, z_{\sigma(i)})^p + \frac{c^p}{\alpha} (|Z| - |X|) \right)^{\frac{1}{p}} + \left( \sum_{i=1}^{|Y|} d^{(c)}(z_{\sigma(i)}, y_{\pi(i)})^p + \frac{c^p}{\alpha} (|Z| - |Y|) \right)^{\frac{1}{p}}.
$$

(13)
The composition $\pi$ is used. Since $\pi$ is a bijection, we can invert $\pi$ to arrive at

$$
\begin{align*}
\frac{d_{p}^{(c, \alpha)}(X, Y)}{d_{p}(X, Y)} & \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} d_{p}^{(c)}(x_{i}, z_{\sigma(i)})^{p} + \frac{c_{p}^{p}}{\alpha} (|Z| - |X|) \right)^{\frac{1}{p}} + \left( \sum_{i=1}^{\lfloor |Y| \rfloor} d_{p}^{(c)}(z_{\pi^{-1}(\sigma(i))}, y_{i})^{p} + \frac{c_{p}^{p}}{\alpha} (|Z| - |Y|) \right)^{\frac{1}{p}}.
\end{align*}
$$

(14)

The composition $\pi^{-1} \circ \sigma$ will be a permutation on $\{1, \ldots, |Z|\}$. Let's denote this as $\tau$. So, for any $\tau, \sigma \in \Pi_{|Z|}$,

$$
\begin{align*}
\frac{d_{p}^{(c, \alpha)}(X, Y)}{d_{p}(X, Y)} & \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} d_{p}^{(c)}(x_{i}, z_{\sigma(i)})^{p} + \frac{c_{p}^{p}}{\alpha} (|Z| - |X|) \right)^{\frac{1}{p}} + \left( \sum_{i=1}^{\lfloor |Y| \rfloor} d_{p}^{(c)}(z_{\tau(i)}, y_{i})^{p} + \frac{c_{p}^{p}}{\alpha} (|Z| - |Y|) \right)^{\frac{1}{p}}.
\end{align*}
$$

(15)

which also holds for the $\sigma$ and $\tau$ that minimizes the first and the second term in the right hand side. This proves the triangle inequality for this case.

**Case 2: $|X| \leq |Z| \leq |Y|$**

As before, for any $\pi \in \Pi_{|Y|}$ and $\sigma \in \Pi_{|Z|}$,

$$
\begin{align*}
d_{p}^{(c, \alpha)}(X, Y) & \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} \left[ d^{(c)}(x_{i}, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right]^{p} \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |X|) \right)^{\frac{1}{p}} \\
& \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} \left[ d^{(c)}(x_{i}, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right]^{p} \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |X|) \right)^{\frac{1}{p}} \\
& \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} \left[ d^{(c)}(x_{i}, z_{\sigma(i)})^{p} + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)})^{p} \right]^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Z| - |X|) \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |Z|) \right)^{\frac{1}{p}} \\
& \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} \left[ d^{(c)}(x_{i}, z_{\sigma(i)})^{p} + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)})^{p} \right]^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Z| - |X|) \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |Z|) \right)^{\frac{1}{p}}.
\end{align*}
$$

(16)

**Case 3: $|Z| \leq |X| \leq |Y|$**

$$
\begin{align*}
d_{p}^{(c, \alpha)}(X, Y) & \leq \left( \sum_{i=1}^{\lfloor |X| \rfloor} d^{(c)}(x_{i}, y_{\pi(i)})^{p} + \frac{c_{p}^{p}}{\alpha} (|Y| - |X|) \right)^{\frac{1}{p}} \\
& \leq \left( \sum_{i=1}^{\lfloor |Z| \rfloor} d^{(c)}(x_{i}, y_{\pi(i)})^{p} + \frac{2c_{p}^{p}}{\alpha} (|X| - |Z|) \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |X|) \right)^{\frac{1}{p}}.
\end{align*}
$$

(19)

To get the above inequality, for $i = |X| + 1, \ldots, |Z|$, we have used the fact that $d^{(c)}(x_{i}, y_{\pi(i)})^{p} \leq c_{p}^{p} \leq 2c_{p}^{2}$ when $0 \leq \alpha \leq 2$,

$$
\begin{align*}
d_{p}^{(c, \alpha)}(X, Y) & \leq \left( \sum_{i=1}^{\lfloor |Z| \rfloor} \left[ d^{(c)}(x_{i}, z_{\sigma(i)}) + d^{(c)}(z_{\sigma(i)}, y_{\pi(i)}) \right]^{p} \right)^{\frac{1}{p}} + \frac{c_{p}^{p}}{\alpha} (|Y| - |Z|) \right)^{\frac{1}{p}}.
\end{align*}
$$

(20)

From here, the arguments are similar to the ones in the last two cases.

For the case when $p = \infty$ and $0 < \alpha \leq 2$, the triangle inequality follows immediately by using the result [14, pp. 203]

\[ \lim_{p \to \infty} \left( \sum_{i=1}^{m} |a_{i}|^{p} \right)^{\frac{1}{p}} = \max_{i=1, ..., m} |a_{i}| \text{ in the triangle inequality for } p < \infty. \]

**APPENDIX B**

**PROOF OF THE TRIANGLE INEQUALITY OF OSPA**

The proof is handled in three cases as in the GOSPA proof.

**Case 1: $|X| \leq |Y| \leq |Z|$**

For any $\pi \in \Pi_{|Y|}$,

$$
\begin{align*}
d^{(c)}(X, Y) & \leq \left[ \frac{1}{|Y|} \left( \sum_{i=1}^{\lfloor |X| \rfloor} d^{(c)}(x_{i}, y_{\pi(i)})^{p} + c_{p}^{p} (|Y| - |X|) \right) \right]^{\frac{1}{p}}.
\end{align*}
$$

(21)

We use the following result that if $0 \leq a \leq c_{p}^{p} |Y|$ and $|Y| \leq |Z|$, then $a \leq \frac{a}{|Y|} \leq \frac{a + c_{p}^{p} (|Z| - |Y|)}{|Z|}$ [24, App.] to change the normalization to $|Z|$:

$$
\begin{align*}
d^{(c)}(X, Y) & \leq \left[ \frac{1}{|Z|} \left( \sum_{i=1}^{\lfloor |X| \rfloor} d^{(c)}(x_{i}, y_{\pi(i)})^{p} + c_{p}^{p} (|Y| - |X|) \right) \right]^{\frac{1}{p}} + c_{p}^{p} (|Z| - |Y|) \right]^{\frac{1}{p}}.
\end{align*}
$$

(22)

From here, we can argue similar to the Case 1 and show that the triangle inequality holds.
where cases that have both infinite sums and integrals as it appears in triangle inequality is sketched below.

To prove the case, we use Minkowski’s inequality for infinite sums and integrals on the RHS:

\[
\int |\psi_n(x) + \phi_n(x)|^p \, dx \\
\leq \left( \int |\psi_n(x)|^p \, dx \right)^{\frac{1}{p}} + \left( \int |\phi_n(x)|^p \, dx \right)^{\frac{1}{p}}
\]

(27)

And then using Minkowski’s inequality for infinite sums on this, we get the RHS of (26).

Now, we use the above results for the triangle inequality of (25). Let us consider RFS X, Y and Z with joint distribution \( f(X, Y, Z) \):

\[
\sqrt[p]{\mathbb{E} \left[ d_p^{(c, \alpha)}(X, Y)^p \right]} \\
\leq \sqrt[p]{\mathbb{E} \left[ \left( d_p^{(c, \alpha)}(X, Z) + d_p^{(c, \alpha)}(Z, Y) \right)^p \right]} \\
= \left( \int \int \left( d_p^{(c, \alpha)}(X, Z) + d_p^{(c, \alpha)}(Z, Y) \right)^p \, f(X, Y, Z) \, \delta X \delta Y \delta Z \right)^{\frac{1}{p}}
\]

(28)

(29)

(30)

If we expand the set integrals in the above inequality, it is of the form

\[
\sqrt[p]{\mathbb{E} \left[ d_p^{(c, \alpha)}(X, Y)^p \right]} \\
\leq \left[ \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \int \int \\
\left( f_1\left( \{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}, \{z_1, \ldots, z_k\} \right) \\
+ f_2\left( \{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}, \{z_1, \ldots, z_k\} \right) \right)^p \\
\times d(x_1, \ldots, x_i) \, d(y_1, \ldots, y_j) \, d(z_1, \ldots, z_k) \right]^{\frac{1}{p}}
\]

(31)

where

\[
f_1(\{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}, \{z_1, \ldots, z_k\}) \\
d_p^{(c, \alpha)}(\{x_1, \ldots, x_i\}, \{z_1, \ldots, z_k\}) \\
\times \left( f(\{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}; \{z_1, \ldots, z_k\}) \right)^{\frac{1}{p}}
\]

(32)

and

\[
f_2(\{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}, \{z_1, \ldots, z_k\}) \\
d_p^{(c, \alpha)}(\{y_1, \ldots, y_j\}, \{z_1, \ldots, z_k\}) \\
\times \left( f(\{x_1, \ldots, x_i\}, \{y_1, \ldots, y_j\}; \{z_1, \ldots, z_k\}) \right)^{\frac{1}{p}}
\]

(33)
The multiple integrals and multiple sums in (31) can be considered as one major integral and sum. So, using Minkowski’s inequality in (26) for infinite sums and integrals on (31), we get

\[
\sqrt[p]{\mathbb{E}\left[\sum_i (X_i, Y_i)^p\right]} \\
\leq \left[ \int \int \int d_{p,\alpha}^{c}(X, Z)^p f(X, Y, Z)\delta X\delta Y\delta Z \right]^{\frac{1}{p}} \\
+ \left[ \int \int \int d_{p,\alpha}^{c}(Y, Z)^p f(X, Y, Z)\delta X\delta Y\delta Z \right]^{\frac{1}{p}} \\
= \left[ \int \int \int d_{p,\alpha}^{c}(X, Z)^p f(X, Z)\delta X\delta Z \right]^{\frac{1}{p}} \\
+ \left[ \int \int \int d_{p,\alpha}^{c}(Y, Z)^p f(Y, Z)\delta Y\delta Z \right]^{\frac{1}{p}},
\]

which finishes the proof of the triangle inequality. □

REFERENCES