18th International Conference on Information Fusion
Washington, DC - July 6-9, 2015

# A Method for Evaluating Performance of Joint Tracking and Classification

**Le Zhang** **Jian Lan** **X. Rong Li**

Center for Information Engineering Science Research (CIESR)
School of Electronics and Information Engineering
Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, P. R. China
lezhang1991@126.com, lanjian@mail.xjtu.edu.cn

Department of Electrical Engineering
University of New Orleans
New Orleans, LA70148, U.S.A
xli@uno.edu

***Abstract*—Joint tracking and classification (JTC) is rapidly gaining momentum recently. Algorithms have been proposed for this problem. However, performance of tracking and classification has been evaluated separately without considering their interdependence. In this paper, we propose a joint measure, named joint probability score (JPS), to account for tracking error, misclassification and their interdependence. The basic idea of JPS is to measure the difference between the cumulative distribution functions (CDFs) of the ideal JTC and the one to be evaluated. Moreover, performance of tracking and classification can be unified by applying CDF. The proposed method is unit free and positive definite. Also, it has close connections with stochastic dominance and the so-called continuous ranked probability score. Two examples illustrating our JPS are presented. The results demonstrate that JPS reflects well the joint performance of tracking and classification.**

## I. Introduction

Target tracking and target classification are practical problems in many military and civilian fields [1]. Usually they are solved separately. In some cases, these two problems are closely related. For example, different classes may have different kinematic models. Then the classification result may affect tracking and vice versa. Recently, joint tracking and classification (JTC) is gaining momentum. Some methods have been proposed, in which the problem is solved jointly [2] [3] [5] [9] [11]. These methods consider the interdependence between tracking and classification: good tracking may be beneficial to classification and vice versa.

Although performance evaluation is important, JTC related evaluation methods are rare. No matter how valid an algorithm is in theory, its performance must be evaluated for a number of purposes [8]. To our knowledge, existing evaluation methods can be classified into two classes.

The first one is to evaluate tracking and classification performance separately, as used in most JTC publications. For example, root mean-square error (RMSE) and/or average Euclidean error (AEE) are used to evaluate the performance of tracking, and probability of correct classification (PC) is applied to evaluate classification results [2] [3] [10] [11]. Such evaluation is unsatisfactory for JTC, because it cannot reflect the interdependence between tracking and classification. The problem is a joint one and we attempt to solve it jointly. Why do we evaluate the results separately? Take the example given in [15] for illustration here. We would like to recognize and track several possibly crossing ground targets simultaneously, say, a tank and a truck. The goal is to destroy the tank, but not the truck. Thus recognition and tracking in this problem are joint and so an accurate track with a recognition error could be a catastrophe, while a poor track with a correct recognition would lead to a miss. In the case that both estimation performance and classification performance are good, the joint performance may still not be good as the separate metrics measure the average performance of estimation and classification separately without accounting for the correspondence between them. Although an algorithm can possibly classify the target as a tank with a large PC and also provide a tracking result with a small RMSE, the latter does not necessarily correspond to the former, e.g., the obtained track does not satisfy the dynamic properties of the tank. However, this correspondence is needed for us to achieve the final goal. Thus, this algorithm may not have good joint performance.

The second one is based on the idea of mock data (see [4] [16] [9]). The basic idea is to compare some distance between real data and mock data generated by estimation and the data model determined by the classification results. The performance of JTC is evaluated in the data space. This method has the significant advantage that knowledge of the ground truth (e.g., the true target state and class) is not needed.

In this paper, we propose a measure that is unit free and unifies the performance evaluation of tracking and classification. A direct thought is to define a measure in the probability space. Take estimation performance evaluation as an example. A "natural" approach is to measure some "distance" $d(p_x, p_{\hat{x}})$ between the true probability density function (PDF) $p_x$ of the estimand $x$ and the PDF $p_{\hat{x}}$ of the estimator $\hat{x}$. Unfortunately, it is fundamentally flawed, as pointed out in [14]. The coupling between $x$ and $\hat{x}$ should also be considered. Thus using the PDF of estimation error $\tilde{x}$ seems a compelling way. However, using this PDF also has serious drawbacks. For example, in the ideal case, $p_{\tilde{x}}$ is a delta function $\delta(\tilde{x})$, which is hard for us to compare it with the estimation error PDF $\hat{p}_{\tilde{x}}$. To solve this, we propose to use the cumulative distribution function

(CDF) of estimation error norm. Moreover, by using CDF, it is easy to unify tracking and classification performance. The ideal/perfect case has no estimation error and misclassification. The CDF of the ideal case is thus a step function (see Figure 1(b)).

Then we give a measure to compare the CDF of estimation error norm and classification in the ideal case with that obtained by the JTC algorithm. The measure is required to be positive definite. This ensures that the best score can only correspond to ideal JTC. Otherwise, the measure can be cheated by an algorithm, to be shown in Section III. Absolute and relative versions of the measure are developed. They have several nice properties including being unit free, positive definite, as emphasized above, and they are joint measures accounting for the interdependence between tracking and classification. They also have close connections with stochastic dominance and the so-called continuous ranked probability score, which are well-known and widely used in many fields.

This paper is organized as follows. Section II is the main part of the paper, where we formulate the JTC evaluation problem and propose our method, including the absolute and relative versions. Their computation is also given in this section. Section III discusses serval properties of the proposed measure. Section IV presents illustrative examples, and Section V concludes this paper.

## II. Problem Formulation and Main Ideas

### A. Problem Formulation

Let

$$H_t = \{\text{true class is } t\}, D_i = \{\text{JTC decides on class } i\}. \quad (1)$$

What we can use in evaluation is state estimation error and corresponding classification results, which may take three different forms. The best case is that we know the joint probability density-mass function (PDMF) $p(\tilde{x}, D)$ or joint cumulative distribution function (JCDF) $P(\tilde{x}, D)$ of JTC where $D$ is the discrete random variable for decision with the range $\{D_1, \ldots, D_M\}$. However, it is usually hard to get in practice. We will focus on the following two cases.

(a) We can use the state estimation error $\tilde{x}^{(n)}$ and the corresponding classification result $D_i^{(n)}, i = 1, \ldots, M, n = 1, \ldots, N$, where $N$ is the samples size used in evaluation by, e.g., a Monte Carlo simulation, and $M$ is the number of candidate classes.

(b) We have state estimation error $\tilde{x}^{(n)}$ and classification results as soft decisions, e.g., each candidate class is assigned a probability $\mu_{i|t}^{(n)}$ under $H_t$ and

$$\sum_{i=1}^{M} \mu_{i|t}^{(n)} = 1 \quad (2)$$

Note that classification in (a) is actually a hard decision.

In the following, we will use the norm of estimation error $\|\tilde{x}\|$, denoted as $e$. For notational simplicity, we define

$$e_{i|t}^{(n)} = n^{\text{th}} \text{ sample of } e \text{ with } D_i \text{ under } H_t \quad (3)$$

### B. General Considerations

As mentioned above, we want to evaluate JTC in the probability space. Most existing methods in the probability space rely on PDFs. However, using a PDF in our problem has serious drawbacks. In the ideal case, the PDF of $e$ is a delta function $\delta(e)$, which is a generalized function that is 0 everywhere on the real line except at 0 with an integral of 1 over the entire line, as shown in Figure 1(a). It is hard to compare a PDF $p(e)$ with $\delta(e)$. A plausible alternative is to integrate the difference between these two PDFs. However, this method actually compares the difference only at point 0. Thus it is incomprehensive.

Using CDF can overcome these drawbacks. The CDF of the ideal case is a step function. We can compare the CDF $\hat{P}$ of the actual estimation error with it. What is more, every point of $\hat{P}$ is compared with the CDF of the ideal case. Moreover, the estimation error is often expressed in terms of samples, probably due to the widespread use of the Monte Carlo method. Thus it seems more reasonable to evaluate the results in terms of CDF [6].

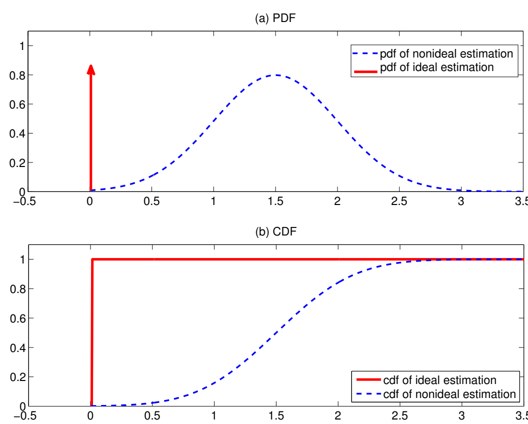


Fig. 1. PDFs and CDFs of the ideal estimation (delta function) and non-ideal estimation

Table 1 Probabilities of tracking and classification

| | $e \le \eta$ | $e > \eta$ | |
|---|---|---|---|
| $D_1$ | $p_{11}$ | $p_{12}$ | $p_{11} + p_{12}$ |
| ... | ... | ... | ... |
| $D_t$ | $p_{t1}$ | $p_{t2}$ | $p_{t1} + p_{t2}$ |
| ... | ... | ... | ... |
| $D_M$ | $p_{M1}$ | $p_{M2}$ | $p_{M1} + p_{M2}$ |
| | $\Sigma_{i=1}^{M} p_{i1}$ | $\Sigma_{i=1}^{M} p_{i2}$ | 1 |

### C. Joint Probability Score

Actually, the CDF $P\{e \le \eta\}$ is the probability of the event that the estimation error norm $e$ is not larger than a threshold $\eta$. For the JCDF of JTC, $P\{e \le \eta, D_i\}$ is the probability of the joint event $e \le \eta$ and the classification $D_i$. As shown in Table 1, $p_{t1} = P\{e \le \eta, D_t | H_t\}$ is the probability that

$e \leq \eta$ and the classification result is $D_t$, where $t$ is the true class. $p_{i2}, i \neq t$ is the probability that $e > \eta$ and the target is misclassified as the $i^{\text{th}}$ class (i.e.,$D_i$) under $H_t$.

Clearly, given $\eta$ the larger $P\{e \leq \eta, D_t|H_t\}$ is, the better the JTC performance is (high correct classification rate and small estimation error). (By the way, it can be proven that $1 - P\{e \leq \eta, D_t|H_t\} = \sum_{i \neq t} P\{(e > \eta) \cup (D_i)|H_t\}$ which is the overall probability of large estimation error (larger than $\eta$) or incorrect classifications ($i \neq t$).)

In view of the above, we can generally evaluate the JTC performance in terms of the following joint probability score (JPS):

$$\text{JPS} = \sum_{t=1}^{M} w_t \int_0^c (1 - P\{e \leq \eta, D_t|H_t\})^r d\eta \qquad (4)$$

where $w_t = P\{H_t\}$ is the prior probability that the true class is $t$. $r$ can be 1, 2, etc. By using $(1 - P\{e \leq \eta, D_t|H_t\})^r$, we measure JTC performance for any threshold $\eta$. By integrating $\eta$ from 0 to $c$, most significant points of the CDF are considered. $c$ is the upper limit of the integral. We will discuss its selection later.

Actually, $1 - P\{e \leq \eta, D_t|H_t\}$ is the difference between the CDF of the evaluated algorithm and that of the ideal one with $e = 0$ (no estimation error) and no misclassification.

When $r = 1$, it is the absolute difference in probability between the ideal JTC and the actual JTC. If $r = 2$, $(1-P\{e \leq \eta, D_t|H_t\})^2$ is the well known Brier score, which is one of the oldest probability verification metrics [13]. The Brier score can be seen as the mean-square error (MSE) of probability estimation.

### *D. Selection of Upper Limit for Integration*

A direct choice of $c$ is $\infty$, which seems reasonable as all probable values of $e$ are considered. However, for some cases, it may not be acceptable. In Figure 2, the JCDF of an estimation error norm $e$ with correct classification (e.g., $P\{e \leq \eta, D_t|H_t\}$) is shown. This JCDF is under the condition that not all classifications are correct. Note that JCDF approaches a steady value as $e$ increases.

Assume that for $\eta$ larger than the dash vertical line shown in Figure 2, the value of $(1 - P\{e \leq \eta, D_t|H_t\})^r$ remains to be $(1 - P\{D_t|H_t\})^r$, which is a nonzero constant. If the integral interval is infinite, the integral will also be infinite. To solve this problem, we suggest selecting a constant $c < \infty$. This is reasonable. First, if we use $\infty$ as the upper limit, once misclassification has a nonzero probability, no matter how small it is, the integral will be infinite. It is unacceptable if the evaluation of the JTC performance is totally dominated by classification performance. No matter how estimation error norm varies, the integral cannot reflect its influence. By choosing a finite $c$ which depends on a particular problem, we provide enough flexibility to balance tracking and classification. Second, in many practical problems, we may not be concerned with how large the estimation error is provided we know it is very large. For example, in road target tracking, position estimates that are too far away from the road may not be considered. Similarly, velocity estimates that are too large (say, 500km/h) for a car may be abandoned.

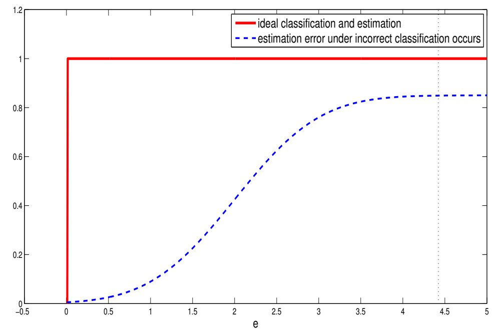


Fig. 2. JCDF of estimation error norm and correct classification under the condition that not all classifications are correct

For performance evaluation, a method depending on specifics of a particular problem has its appeal. The upper limit $c$ of integral should be chosen after sufficiently exploring the given problem and scenario.

### *E. Relative Version*

Equation (4) is an absolute measure in the sense that it is not with respect to any reference. Obviously, the measure is sensitive to the scenario used in the evaluation, including estimand's magnitude and data accuracy [8]. In this subsection, we propose a relative form of JPS (4) by choosing the worst case in theory as the reference.

Given $c$, the worst case is that the algorithm always misclassifies the target or the estimation error norms are all larger than $c$ (see Figure 3).

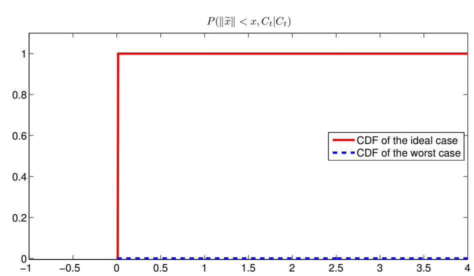


Fig. 3. The ideal and worst estimation and classification CDFs

In the worst case, the integrand of JPS (4) is always equal to 1 because the JTC always misclassifies the target or the estimation error is all larger than $c$. Thus JPS (4) equals $c$. The relative joint probability score (RJPS) can thus be defined as

$$\text{RJPS} = \frac{1}{c} \sum_{t=1}^{M} w_t \int_0^c (1 - P\{e \leq \eta, D_t|H_t\})^r d\eta \qquad (5)$$

In the ideal case mentioned above, RJPS $= 0$. In the worst case, it equals 1. For other cases, $(1-P\{e \le \eta, D_t|H_t\})^r \le 1$, and there is at least one point for which the strict inequality holds. So RJPS is less than 1. Thus, we have a measure with a range $[0, 1]$.

### F. Computation of JPS

In this subsection, JPS and RJPS are computed following the method and description given in [7]. In practice, we may get samples of estimation error and classification results (e.g., cases (a) and (b) in Section II.A) by Monte Carlo simulations. We use these samples to construct an empirical JCDF to approximate the true JCDF for comparison.

We need to rank the samples with classification $D_t$. For notational simplicity, from now on we assume the sample points of the estimation error norm $e$ defined in equation (3) are ordered:

$$e_{t|t}^{(j)} \le e_{t|t}^{(l)},\ j < l,\ j, l = 1, ..., N_t^c \tag{6}$$

where the superscript $(j)$ means the sample is generated in the $j^{\text{th}}$ Monte Carlo run, subscript $t|t$ means classification result is $D_t$ under $H_t$, and $N_t^c = \max\{n : e_{t|t}^{(n)} \le c\}$. $N_t^c$ is the number of samples which are smaller than $c$ with classification results $D_t$.

Note that for case (b) of Section II.A, an estimation error norm may correspond to more than one classification result. Each class is assigned a probability $\{\mu_{i|t}^{(n)}\}_{i=1}^M$, where $\mu_{i|t}^{(n)}$ is defined in equation (2). $\mu_{t|t}^{(n)}$ is the probability assigned to the true class $t$ under $H_t$ by the JTC algorithm.

For case (a), the results are hard decisions. The classification corresponding to an estimation error norm can be seen as one and only one $\mu_{i|t}^{(n)} = 1$, i.e., $\mu_{t|t}^{(n)}$ equals either 0 or 1.

Then the empirical JCDF with correct classification is

$$P\{e \le \eta, D_t|H_t\} = \frac{1}{N}\sum_{l=1}^{n} \mu_{t|t}^{(l)},\ \ e_{t|t}^{(n)} < \eta < e_{t|t}^{(n+1)} \tag{7}$$

Note that $\mu_{t|t}^{(l)}$ is the probability assigned to the true class on the $l^{\text{th}}$ run. See equation (2) for more details. It can be different for different runs.

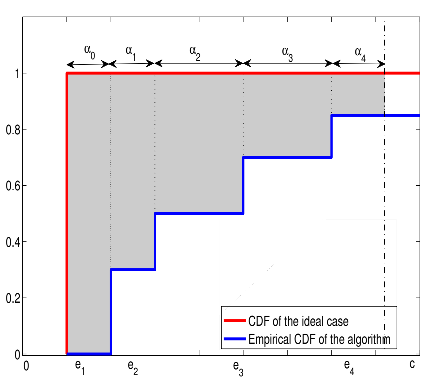


Fig. 4. Computation of JPS from the ideal CDF and empirical CDF

Using the empirical JCDF above, JPS (4) can be computed as follows:

$$\text{JPS} = \sum_{t=1}^{M} w_t \sum_{n=1}^{N_t^c} \alpha_{t|t}^{(n)} d_{t|t}^{(n)} \tag{8}$$

where $\alpha_{t|t}^{(n)}$ and $d_{t|t}^{(n)}$ are given in Table 2. $w_t$ is the prior probability that $H_t$ is the true class, defined right after equation (4). and $M$ is the candidate class number.

The relative JPS can then be obtained by:

$$\text{RJPS} = \frac{\text{JPS}}{c} \tag{9}$$

where $c$ is the maximum value of JPS, as analyzed above.

Table 2 Computation of JPS

| $n$ | 0 | $1, ..., N_t^c - 1$ | $N_t^c$ |
|---|---|---|---|
| $\alpha_{t\|t}^{(n)}$ | $e_{t\|t}^{(1)}$ | $e_{t\|t}^{(n+1)} - e_{t\|t}^{(n)}$ | $c - e_{t\|t}^{(N_t^c)}$ |
| $d_{t\|t}^{(n)}$ | 1 | $(1 - \sum_{l=1}^{n} \mu_{t\|t}^{(l)}/N)^r$ | $(1 - N_t^c/N)^r$ |

The computation can be summarized as follows:

1. Generate the true class $H_t$ according to its prior probability; get estimation and classification results.
2. Repeat step 1 for $N$ times.
3. Rank $e_{t|t}^{(n)}$ and arrange its corresponding classification result $\mu_{t|t}^{(n)}$ for $D_t$.
4. Select the upper limit $c$ and compute $\alpha_t^{(n)}$ and $d_t^{(n)}$ given in Table 2 for all $e_{t|t}^{(n)}$.
5. Get JPS and RJPS.

## III. Properties and Discussions

The proposed JPS has several attractive properties. In this section, we discuss these properties and show its connections with some well-known measures.

### A. Unit Free

The proposed JPS and RJPS are unit free as the integrand is probability. Using a unit free measure allows us to compare cases with different units. Moreover, as mentioned above, RJPS is a measure with a value between 0 and 1 and is thus not quite sensitive.

### B. Positive Definite

The proposed JPS and RJPS are positive definite. From equation (4), we have JPS = 0 if and only if the JCDF of the algorithm coincides with the JCDF of the ideal case. For other cases, JPS $> 0$.

This requirement seems simple; however, not all measures are positive definite. [6] gives an example to illustrate this, which is shown next. Consider estimating a Gaussian density $f_p$ with mean $\mu$ and variance $\sigma^2$. A possible measure using an observation is

$$S = (z - \mu)^2 + \sigma^2 \tag{10}$$

which seems to reward a smaller difference between the observation and the mean and a smaller variance of the estimation density. Unfortunately, an algorithm may cheat the measure, for example, by issuing $\sigma^2 = 0$. Thus, it is quite reasonable to require positive-definiteness.

### C. Joint Measure

The JPS measures the joint performance of tracking and estimation, accounting for the estimation error, misclassification, and their interdependence. We discuss the performance and the corresponding JPS values qualitatively in various cases.

As mentioned above, good JTC has a small estimation error and correct classification. In this case, $(1-P\{e \leq \eta, D_t|H_t\})^r$ is small for all $\eta$ and thus JPS is small.

On the other hand, terrible JTC always misclassifies the target or has a large estimation error norm. Because of incorrect classifications and large estimation errors, $(1-P\{e \leq \eta, D_t|H_t\})^r$ is large over a large interval of $\eta$, and so JPS is very large. Note that JPS measures the estimation and classification results jointly. The measure is sensitive to all distributional facets rather than just means and variance. Thus it is a comprehensive measure.

For other cases, for instance, the classification performance is poor, the estimation error with correct class is small and with incorrect classification is large. Thus $(1-P\{e \leq \eta, D_t|H_t\})^r$ is large, especially for small $\eta$.

An algorithm with good estimation has a small estimation error norm, that is, $P\{e \leq \eta\}$ is large. Similarly, an algorithm with good classification results has a large $P\{D_t|H_t\}$. Good JTC performance needs large $P\{e \leq \eta, D_t|H_t\}$. Note that $P\{e \leq \eta, D_t|H_t\} = P\{e \leq \eta|D_t, H_t\}P\{D_t|H_t\} \neq P\{e \leq \eta|H_t\}P\{D_t|H_t\}$. Even if an algorithm has good performance in estimation and classification separately, it still may not have good joint performance. Our JPS stresses the importance of their interdependence in the JTC evaluation.

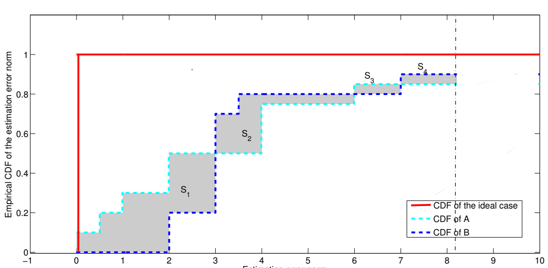

Fig. 5. Empirical CDFs and illustration of FSD and SSD

### D. Connections to Stochastic Dominance

For $r = 1$, JPS (4) can be interpreted as stochastic dominance, which is widely used in economics and investment theory. The definition of stochastic dominance given by [12] is that A dominates B in $\mathbf{U}_1$ if for all utility functions $U \in \mathbf{U}_1$ the inequality $E_A U(x) \geq E_B U(x)$ holds and for at least one utility $U_0 \in \mathbf{U}_1$ the strict inequality holds. $\mathbf{U}_1$ is the set of all utility functions $U$ that satisfy the first derivative $U' \geq 0$. This is known as the first degree stochastic dominance (FSD).

$A(x)$ and $B(x)$ are two CDFs of variable $x$. We say A dominates B by FSD for all $U \in \mathbf{U}_1$ if and only if $A(x) \geq B(x)$ for all values $x$, and there is at least one $x_0$ for which the strict inequality holds [12].

From Figure 5, the CDF of the ideal case dominates the JTC's CDF by FSD. The difference of these two CDFs is measured by the area between their CDFs.

A natural idea of ranking two algorithms is to compare their empirical CDFs directly. In some cases, it seems appealing. However, as shown in Figure 5, the CDFs of A and B may intersect, and thus the FSD cannot be applied. One may use the second degree stochastic dominance (SSD). In investment theory, users are assumed to be risk averters such that the utility function $U \in \mathbf{U}_2$, where $\mathbf{U}_2$ is the set of utilities that satisfy $U' \geq 0$ and the second derivative $U'' \geq 0$. Then, for A and B in Figure 5, we say A SSD B if and only if $S_1 + S_3 > S_2 + S_4$, where $S_i$ is the area enclosed between A and B from the lower intersection point of A and B to the higher point. It seems reasonable to do so as most investors dislike risk. For JTC performance evaluation, however, the assumption is not valid. So we recommend to use our method to get the JPS of each CDF first, and then algorithms can be ranked according to their JPS values.

### E. Connections to Continuous Ranked Probability Score

For $r = 2$, $c = \infty$, if classification results are all correct, (4) can be written as

$$\sum_{t=1}^{M} w_t \int_0^\infty (1 - P\{e \leq \eta, D_t|H_t\})^2 d\eta. \tag{11}$$

The above measure, known as the continuous ranked probability score (CRPS) [7], is widely used in meteorology. It can be seen as an extension of the (half) Brier score, $(p_t - p_e)^2$.

For deterministic estimation (i.e., $\|\tilde{x}\| = e$ with $Cov(\tilde{x}) = 0$), the CDF of the estimation is also a step function. $(1 - P\{e \leq \eta, D_t|H_t\})^2$, the integrand of formula (11), is 1 over the interval $[0, e]$ for $\eta$ and 0 over $(e, c]$, and then formula (11) turns out to be $e$, which is the estimation error norm. Thus, CRPS can be regarded as a weighted sum of norm. Moreover, our measure JPS with $r = 2$ can be treated as an extension of CRPS considering discrete and continuous variables jointly.

## IV. Illustrative Examples

In this section, two simple examples are provided to illustrate the proposed method. Both examples assume two candidate classes are characterized by different kinematic models:

(a) class $C_1$ with dynamic model CV [1]:

$$X_{k+1} = F_1 X_k + u_1 \tag{12}$$

where $u_1 \sim N(0, Q_1)$, and

$$F_1 = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

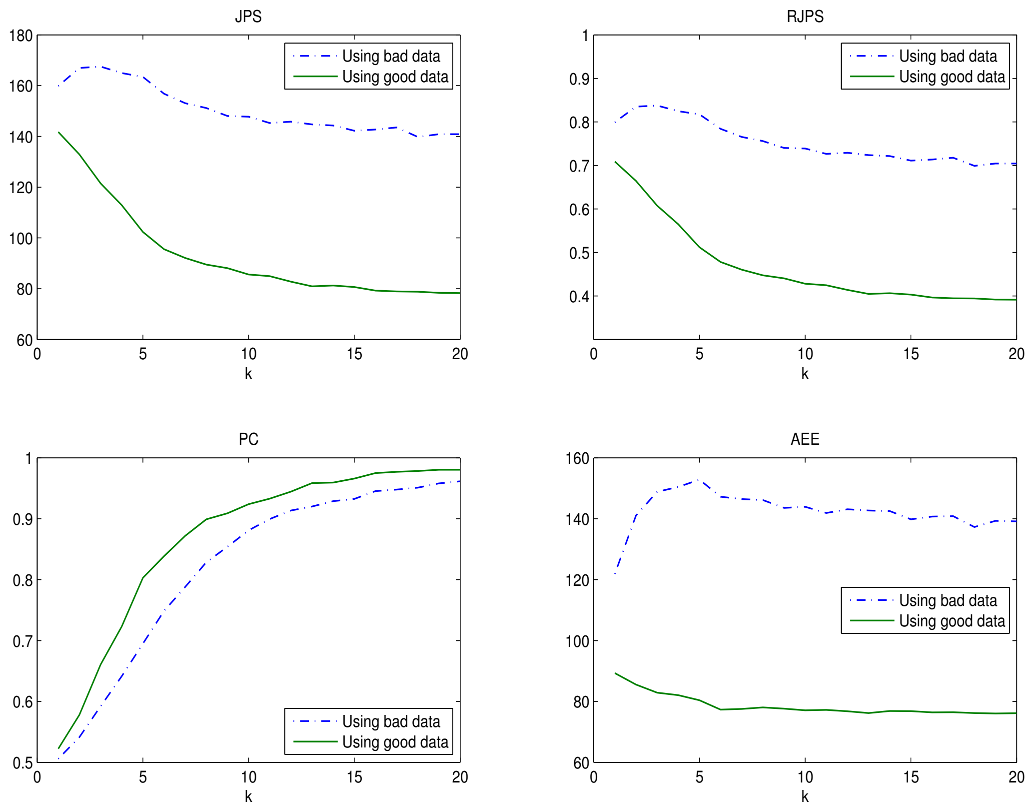


Fig. 6. Simulation results of the same algorithm using good data and bad data

$Q_1 = \text{diag}(160\text{m}^2, 160(\text{m/s})^2, 160\text{m}^2, 160(\text{m/s})^2)$.

(b) class $C_2$ with dynamic model CT [1]:

$$X_{k+1} = F_2 X_k + u_2 \tag{13}$$

$$F_2 = \begin{bmatrix} 1 & \sin(wT)/w & 0 & -(1-\cos(wT))/w \\ 0 & \cos(wT) & 0 & -\sin(wT) \\ 0 & (1-\cos(wT))/w & 1 & \sin(wT)/w \\ 0 & \sin(wT)/w & 0 & \cos(wT) \end{bmatrix}$$

where $u_2 \sim N(0, Q_2)$, $w = 0.45$rad/s, and

$$Q_2 = \text{diag}(250\text{m}^2, 250(\text{m/s})^2, 250\text{m}^2, 250(\text{m/s})^2).$$

The measurement equation is

$$Z_k^i = HX_k + v_i \tag{14}$$

with $v_i \sim N(0, R_i)$,

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$R_1 = \text{diag}(900\text{m}^2, 1000\text{m}^2)$, $R_2 = \text{diag}(8000\text{m}^2, 8000\text{m}^2)$.

The prior probability of two classes are $w_1 = 0.5$, $w_2 = 0.5$. True class and state are generated randomly according to this probability.

**Example 1** We compare the same algorithm using two sets of data with different quality. One (good data) is with a small covariance $R_1$, the other one (bad data) is with a large covariance $R_2$. Intuitively, the algorithm using the good data has better performance.

The curves of JPS, AEE [8], and correct classification rate are shown in Figure 6, where $\text{AEE}(\hat{x}) = \frac{1}{N}\sum_{i=1}^{N} e_i$, $N$ is the sample size and $e$ is the norm of estimation error. In the first subfigure of Figure 6, with the growth of data, the joint performance of JTC becomes better and better. Yet it is not clear for us to get this conclusion from the curves of AEE and the probability of correct classification rate (PC), as they are separate measures of two aspects.

**Example 2** Two commonly used methods, decision and estimation (D&E) and decision then estimation (DTE), are compared in this example. For D&E, tracking and classification are treated as two separate problems and solved without considering their interdependence even though both decision and estimation are optimal in their own senses [9]. For DTE, optimal decision is made first and then optimal estimation is done as if decision were always correct. The method partially considers the dependence of estimation on decision.

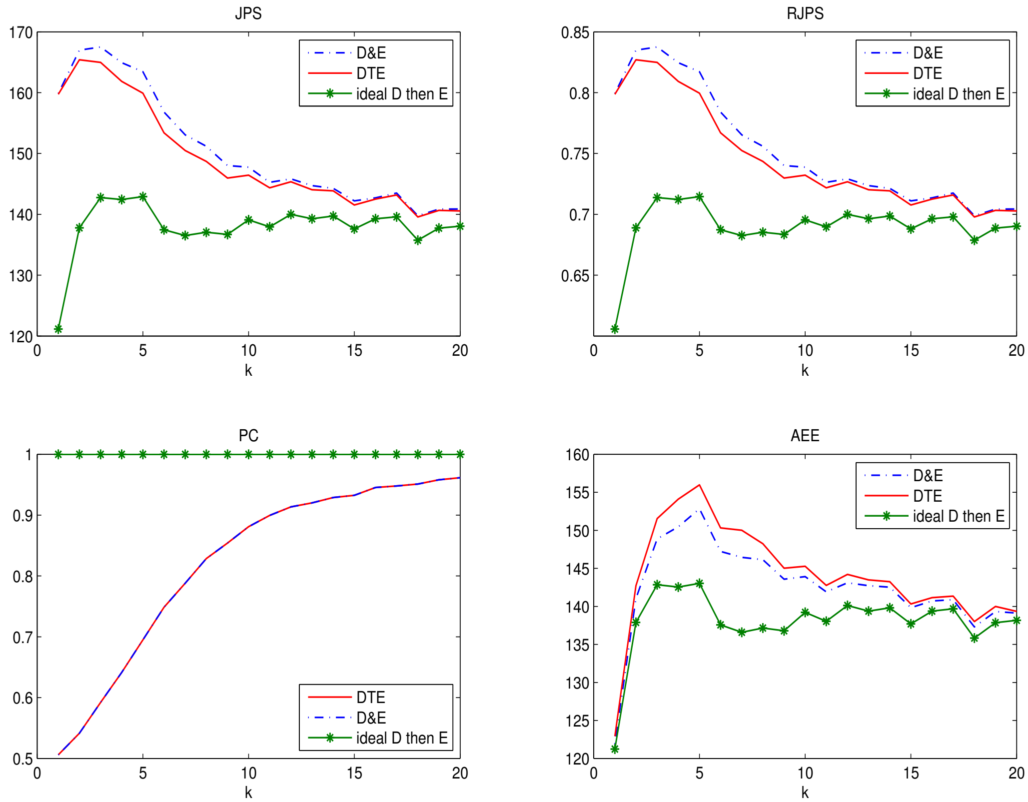


Fig. 7. Simulation results of decision and estimation, decision then estimation and ideal decision then estimation

The results are shown in Figure 7. For classification, DTE and D&E are the same. Note that the PC increases rapidly in the time interval 5s to 10s. The forth subfigure shows the AEE of D&E and DTE. D&E performs better than DTE in AEE as it is the optimal estimation. We can draw conclusions from the third and forth subfigures as follows. For classification, D&E is as good as DTE; for estimation, D&E is better than DTE. As D&E has the optimality in estimation and classification separately, we cannot say D&E performs better jointly than DTE.

Actually, DTE should have better joint tracking and classification performance than D&E, as DTE considers such dependence of estimation on classification while D&E does not consider any dependence. This result is obvious at the beginning of JTC because of the low PC. The first and second subfigures show that DTE is better than D&E. This illustrates that our measure indeed accounts for the interdependence between estimation and classification. It is misleading as mentioned above to use estimation and classification measures separately. Using a joint measure, such as our proposed JPS, is necessary for performance evaluation of JTC. The "ideal D then E" makes a perfect decision first and then does optimal estimation, and thus JPS reflects its performance of estimation only.

This example illustrates the statements in Section III.C and addresses the importance of using a joint measure to evaluate JTC.

## V. Conclusion

This paper deals with evaluation of JTC performance. The most widly used strategy by far is to evaluate tracking performance and classification performance separately. As the problem is a joint one, using separate metrics cannot reflect joint performance, especially the coupling between tracking and classification. We address the importance of using a joint measure in this problem. Based on this, we have proposed the method considering the performance of tracking and classification jointly, called JPS. Unlike common practice, we use CDF instead of PDF, as it is easier to compare and approximate CDFs and more comprehensive than using PDF. The JPS quantifies the difference in the CDFs of the ideal case and the algorithm outputs, which can reflect all distributional facets, and thus the measure is a comprehensive one. In JPS, the upper limit of the integral balances tracking and classification. It is better for a user to select the limit for each particular problem.

The relative JPS, is also provided, which has a value between 0 and 1 and is more insensitive. Computation of JPS is quite easy and the computational steps are given.

JPS is unit free and positive definite. Moreover, it can be interpreted by stochastic dominance for $r = 1$ and treated as an extension of the continuous ranked probability score for $r = 2$, which are well known and widely used in economic and meteorology. In addition to the JTC evaluation, the idea of JPS can be used in many areas as it unifies continuous and discrete variables. As demonstrated by the two examples presented, our proposed measures indeed account for estimation error, classification error, and their interdependence.

## REFERENCES


[1] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms, and Software*. New York: Wiley, 2001.

[2] J. Lan and X. R. Li, "Joint tracking and classification of extended object using random matrix," in *International Conference on Information Fusion*, (Istanbul, Turkey), July 2013, pp. 1550–1557.

[3] W. Cao, J. Lan, and X. R. Li, "Extended object tracking and classification based on recurisive joint decision and estimation," in *International Conference on Information Fusion*, (Istanbul, Turkey), July 2013, pp. 1670–1677.

[4] X. R. Li, M. Yang, and J. Ru, "Joint tracking and classification based on Bayes joint decision and estimation," in *Proceedings of International Conference on Information Fusion*, (Quebec City, Canada), July 2007, pp. 1421–1428.

[5] S. Challa and G. W. Pulford, "Joint target tracking and classification using radar and ESM sensors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 1039–1055, July 2001.

[6] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, March 2007.

[7] H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather and Forecasting*, vol. 15, pp. 559–570, May 2000.

[8] X. R. Li and Z. -L. Zhao, "Evaluation of estimation algorithms Part I: incomprehensive measures of performance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 4, pp. 1340–1358, October 2006.

[9] W. Cao, J. Lan, and X. R. Li, "Joint tracking and classification based on recursive joint decision and estimation using multi-sensor data," in *International Conference on Information Fusion*, (Salamanca, Spain), July 2014.

[10] W. Mei, G. -L. Shan, and X. R. Li, "An efficient Bayesian algorithm for joint target tracking and classification," in *International Conference on Signal Processing*, August 2004, pp. 2090–2098.

[11] L. -F. Sun, J. Lan, and X. R. Li, "Joint tracking and classification of extended object based on support functions," in *International Conference on Information Fusion*, (Salamanca, Spain), July 2014.

[12] H. Levy, *Stochastic Dominance: Investment Decision Making under Uncertainty*. New York: Springer, 2006.

[13] G. W. Brier, "Verification of forecasts expressed in terms of probability," *Monthly Weather Review*, vol. 78, no. 1, pp. 1–3, 1950.

[14] X. R. Li, Z. -L. Zhao, and Z. -S. Duan, "Error spectrum and desirability level for estimation performance evaluation," in *Workshop on Estimation, Tracking and Fusion: A Trbute to Fred Daum*, (Monterey, CA, USA), May 2007.

[15] X. R. Li, "Optimal Bayes joint decision and estimation," in *International Conference on Information Fusion*, (Quebec City, Canada), July 2007.

[16] Y. Liu and X. R. Li, "Recursive joint decision and estimation based on generalized Bayes risk," in *Proceedings of International Conference on Information Fusion*, (Chicago, USA), July 2011.