Joint Tracking and Classification
Based on Bayes Joint Decision and Estimation

X. Rong Li, Ming Yang
Department of Electrical Engineering
University of New Orleans, New Orleans, LA 70148
Phone: 504-280-7416, Email: {xli, myang2}@uno.edu

Jifeng Ru
Arcon Corporation
Waltham, MA 02451
Email: jifeng@arcon.com

Abstract—Many problems involve both decision and estimation where the performance of decision and estimation affects each other. They are usually solved by a two-stage strategy: decision-then-estimation or estimation-then-decision, which suffers from several serious drawbacks. A more integrated solution is preferred. Such an approach was proposed in [14]. It is based on a new Bayes risk as a generalization of those for decision and estimation, respectively. It is Bayes optimal and can be applied to a wide spectrum of joint decision and estimation (JDE) problems. In this paper, we apply that approach to the important problem of joint tracking and classification of targets, which has received a great deal of attention in recent years. A simple yet representative example is given and the performance of the JDE solution is compared with the traditional methods. Issues with design of parameters needed for the new approach are addressed.

Keywords: Target tracking, target classification, estimation, decision, Bayes approach.

I. INTRODUCTION

Although some problems involve decision and/or estimation either alone or separately, good solutions to many other problems require joint decision and estimation (JDE). This is often the case for problems involving inter-dependent discrete- and continuous-valued uncertainties and thus decision and estimation affect each other.

The prevailing approach to JDE tries to make the best decision first disregarding estimation and then solve the estimation problem as if the decision were surely correct. This “decision-then-estimation” strategy has several serious flaws. It does not account for the possible decision errors in the subsequent estimation; on the other hand, decision is made regardless of the result of estimation. Another strategy is “estimation-then-decision.” It would not work well if estimation is significantly dependent on decision or estimation is not secondary. In the general case, a joint approach would be more promising than separate decision and estimation as well as decision then estimation or estimation then decision.

While JDE is an old problem and has been studied before, it has received increasing attention in recent years and efforts have been made to overcome the aforementioned drawbacks (see, e.g., [2]), especially for target inference (see, e.g., [1], [3], [4], [6]–[13], [15]–[20]).

In the companion paper [14], an integrated approach to JDE based on a novel Bayes risk as a generalization of those for decision and estimation, respectively, was proposed. In the optimization theoretic parlance, this approach has the potential of arriving at a globally optimal solution, which is inherently superior in performance to the conventional two-stage optimization strategy or separate decision and estimation, especially for problems where decision and estimation are highly correlated. The structure of the optimal solution in this framework and a converging JDE algorithm were presented in [14]. The power of the proposed framework is illustrated by applications to several difficult JDE problems in target inference, including track fusion and joint detection and tracking.

In this paper, we apply the JDE approach of [14] to joint tracking and classification (JTC) of targets. We consider a simple yet representative JTC example, where three types of data are available: the first type is useful for both tracking and classification; the second is particularly good for classification but not directly useful for tracking, the third is particularly good for tracking but not directly useful for classification. We present optimal decision, optimal estimation, decision-then-estimation, estimation-then-decision, and our proposed JDE solution in the Bayesian setting for this example. In this way, we demonstrate how the proposed JDE solution work and contrast its performance with the existing methods.

While the JDE approach of [14] is general, it relies on several design parameters. Only simple guidelines have been presented in [14]. In this paper, a case study is conducted concerning these design parameters that make a trade-off between decision and estimation performance.

Also, we propose a general method of evaluating performance of joint decision and estimation in a comprehensive way. To our knowledge, it is the only method, be it comprehensive or not, available that evaluates joint decision and estimation performance.

II. JOINT DECISION AND ESTIMATION

Numerous practical problems involve joint decision and estimation, where decision and estimation are inter-dependent. Consider a JDE problem in which \( x \) is the estimate (i.e., the quantity to be estimated) and the decision involves \( M \) candidates: \( D_1, \ldots, D_M \). The complete Bayes solution to the problem of estimating \( x \) using data \( z \) is its posterior density \( f(x|z) \). By the same token, a complete Bayesian solution

---

1Research supported in part by ARO grant W911NF-04-1-0274, NASA/LEQSF grant (2001-4)-01, and Navy through Planning Systems Contract # N00014-05-C-0382.
to the problem of deciding on candidates $D_1, \ldots, D_M$ is the set of posterior probabilities \{ $P\{D_1|z\}, \ldots, P\{D_M|z\}$ \}. We call this set of probabilities soft decisions, although this is unorthodox in decision theory. In the same spirit, when $x$ and $D$ are not independent, we may be interested in inferring them jointly in terms of the set of functions \{ $P([D, x])|z\}, \ldots, P([D, x]|z)$ \}, where $P([D, x]|z)$ is the posterior mixed “probability-density.” Here the hybrid quantity $(D, x) = (x(D_i), D_i(x))$ signifies the inter-dependence of $x$ and $D$. In some cases, a decision-then-estimation or an estimation-then-decision method is justifiable. In the general case, however, an integrated approach would be more promising than separate decision and estimation as well as the traditional two-stage strategies. Such an approach was presented in [14]. For convenience of the reader, its main results are presented below.

A. An Integrated Approach to JDE

The basic idea is to minimize the following Bayes risk for joint decision and estimation

$$\bar{R} = \sum_{i=1}^{M} \sum_{j=1}^{N} (\alpha_{ij}c_{ij} + \beta_{ij}E[C(x, \hat{x})|D_i, H_j])P\{D_i, H_j\}$$

(1)

where $D_i$ stands for the $i$th decision, which is equivalent to the event \{ $z \in D_i$ \}, where $D_i$ is the decision region for $D_i$ in the data space; $c_{ij}$ is the cost of decision $D_i$ when hypothesis $H_j$ is true; $C(x, \hat{x})$ is the estimation cost function; $E[C(x, \hat{x})|D_i, H_j]$ is the expected cost conditioned on the case that the decision is made on $D_i$ but $H_j$ is true; $\alpha_{ij}$ and $\beta_{ij}$ are relative weights of decision and estimation costs, which can be chosen to fit different requirements given by practical problems.

This new Bayes risk generalizes the traditional ones in that (a) it is a risk for joint decision and estimation, (b) the decision set and the hypothesis set need not be one-to-one correspondent, (c) it can handle problems with incommensurable subproblems of estimation, and (d) the coefficients $(\alpha_{ij}, \beta_{ij})$ provide additional flexibilities. For example, if $M = N$ and $D_i = \{ \text{deciding on } H_i \}$ for all $i$, then: (a) $\bar{R}$ with $(\alpha_{ij}, \beta_{ij}) = (1, 0)$ and $(0, 1)$ reduces to the traditional Bayes risk for decision and estimation, respectively; (b) $\bar{R}$ is simply the sum of the traditional Bayes risks for decision and estimation respectively if $(\alpha_{ij}, \beta_{ij}) = (1, 1)$; (c) $\bar{R}$ is a weighted sum of the product of decision and estimation costs if $(\alpha_{ij}, \beta_{ij}) = (0, c_{ij})$.

For any given $E[C(x, \hat{x})|D_i, H_j]$, to minimize $\bar{R}$, the optimal decision $D$ is

$$D = D_i, \quad \text{if } C_i(z) \leq C_k(z), \forall k$$

(2)

where the posterior cost is given by

$$C_i(z) = \sum_{j=1}^{N} (\alpha_{ij}c_{ij} + \beta_{ij}E[C(x, \hat{x})|D_i, H_j])P\{H_j|z\}$$

and given a set of decision regions $\{D_1, \ldots, D_M\}$ as a partition of the data space, the optimal estimator for (1) with $C(x, \hat{x}) = \hat{x}'\hat{x}$ is the following generalized conditional mean

$$\hat{x} = \sum_i 1(z; D_i)\hat{x}_i$$

(3)

where, for $z \in D_i$,

$$\hat{x}_i = E[x|z] = \sum_j E[x|z, H_j]P\{H_j|z\}$$

$$P\{H_j|z\} = \frac{\beta_{ij}P\{H_j|z\}}{\sum_k \beta_{ik}P\{H_k|z\}}$$

$$1(z; D_i) = \begin{cases} 1 & z \in D_i \\ 0 & \text{else} \end{cases}$$

and these quantities are undefined if $z \notin D_i$.

The optimal joint decision-estimate $(D, \hat{x})$ is the combination of the above optimal decision and optimal estimate. It can be obtained by the following iterative algorithm, which always converges.

1. Start from an arbitrary initial decision set $\{D_i\}_{r=0}$ as the initial partition $\{D_i\}$, where $r$ denotes the iteration index.
2. Estimation update: based on $\{D_i\}_r$, calculate $\hat{x}_i$ and $P\{D_i, H_j|z\}$ to obtain the optimal estimate $\hat{x}$ with the corresponding conditional mean square error.
3. Decision update: calculate $C_i(z)$ for each $D_i$, then $\bar{R}_r$ is obtained.
4. Compare $\bar{R}_r$ and $\bar{R}_{r-1}$. If the relative difference is small enough, stop the iteration, so $D_i$ with the smallest posterior cost at iteration $r$ is taken as the optimal decision and the corresponding $\hat{x}$ is taken as the optimal estimate; otherwise change decision regions to $\{D_i\}_r$ and repeat steps 2, 3, 4.

Simplification of this algorithm was discussed in [14]. Note that (3) reduces to the conditional mean if $\beta_{ij}$ is constant over $j$. Assume there is no decision error, namely, $P\{H_j|D_i, z\} = \delta_{i-j}$ if $z \in D_i$ and undefined otherwise, where $\delta_{i-j}$ is the Kronecker delta: $\delta_{i-j} = 1$ if $i = j$ and 0 otherwise. Then, the optimal estimator (3) reduces to

$$\hat{x} = \sum_i 1(z; D_i)E[x|z, H_i]$$

(4)

which is the estimator in the traditional “decision-then-estimation” strategy. It is the conditional mean based on a single “correct” decision, whereas (3) is a weighted sum of the conditional means to account for possible decision errors.

B. Performance Evaluation

While many practical problems involve JDE, their solutions are evaluated so far only in terms of decision performance and estimation performance, separately. We are not aware of any measure, comprehensive or not, for evaluating joint decision and estimation performance. In this section, we propose a systematic method and measure for evaluating a JDE algorithm comprehensively based on statistical distance between the original data and the mock data generated by the JDE algorithm.
A large class of JDE problems can be formulated as follows. The ground truth is that the observation data $z$ has a distribution $F(z|x)$; that is, $z \sim F(z|x)$, where $x$ is to be estimated and $s$ is unknown with discrete (or finitely many possible) values. Probably more often, the exact form of $F(z|x)$ is not known; rather, it is known that $z = h(s, x, v)$, where $v \sim F(v|x)$ with known $F(v|x)$. A JDE problem consists of two subproblems: decide on the $s$ value and estimate $x$. With the assumption that $s \in \{1, 2, \ldots, N\}$, it can be formulated as estimating $x$ and testing the hypotheses:

$$H_1 \text{ vs. } H_2 \text{ vs. } \cdots \text{ vs. } H_N$$

where $H_i$: $z \sim F_i(z|x)$ and $F_i(z|x) = F_i(z|x)|_{s=i}$. Let $\xi = (s, x)$. Then, the solution to a JDE problem is $\xi = (d, \hat{x})$, where $\hat{x}$ is the estimate of $x$ and $d \in \{1, 2, \ldots, N\}$ is the decision.

The basic idea of our method for evaluating JDE performance is to measure some distance between the true distribution $F(z|\xi)$ and the distribution $F_d(z|\hat{x})$ corresponding to the JDE result if they are available, such as in a simulation-based evaluation study, or some statistical distance between the original data $z$ and the mock data $\hat{z}$ generated by the JDE algorithm if the distributions are not available. Here, the mock data $\hat{z}$ is generated randomly by the JDE algorithm with the result $(d, \hat{x})$ via either the distribution $F_d(z|\hat{x})$ or the data model that converts a JDE result to the data, such as $\hat{z} = h(d, \hat{x}, v)$ with $v \sim F(v|d, \hat{x})$.

More specifically, let

$$\rho(z, \hat{z}) = \int \int \Delta[f(z|\xi), f_d(z|\hat{x})]dF(z|\xi)$$

where $\Delta[f(z|\xi), f_d(z|\hat{x})]$ is the “distance” between $f(z|\xi)$ and $f_d(z|\hat{x})$.

If both $s$ and $x$ are random, generate $\xi_i \sim f(\xi)$, $i = 1, \ldots, N_i$, where $f(\xi)$ is the prior distribution of $\xi$ for performance evaluation determined by the evaluator, which could differ from the one used in the JDE algorithm.

- For each pair $\xi_i = (s_i, x_i)$, generate $z_{ij} \sim f(z|\xi_i)$, $j = 1, \ldots, N_j$, each may be a vector consisting of multiple data points.
- For each $z_{ij}$, obtain $\hat{z}_{ij} = (d_{ij}, \hat{x}_{ij}) = g(z_{ij})$ by the JDE algorithm to be evaluated.

Let $\rho(z_i, \hat{z}_{ij}) = \Delta[f(z_i|\xi_i), f_d(z_i|\hat{x}_{ij})]$ be the “distance” between $f(z_i|\xi_i)$ and $f_d(z_i|\hat{x}_{ij})$. Compute the final performance metric

$$\rho(z, \hat{z}) = \int \int \Delta[f(z|\xi), f_d(z|\hat{x})]dF(z|\xi) \approx \frac{1}{N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \rho(z_i, \hat{z}_{ij})$$

We may use the total variation distance

$$\Delta_{tv}(F, F_d) = \frac{1}{2} \int |f(x) - f_d(x)|dx = \frac{1}{2} \sum |p_t(z_i) - p_d(z_i)|$$

where $p_t$ and $p_d$ are probability mass functions (pmfs). $\rho(z, \hat{z})$ will have a simple form when pmfs are involved:

$$\rho_{tv}(z, \hat{z}) \approx \frac{1}{2N_i N_j} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} |p_t(z_{ik}) - \hat{p}_d(\hat{z}_{ijk})|$$

### III. Target Tracking and Classification

Target tracking and classification (recognition) are coupled in many cases, but are often handled independently. For instance, while many tracking algorithms are mainly based on kinematic (e.g., radar, sonar) measurements and associated models, target classification is usually handled using identity or attribute data from, e.g., electronic support measure (ESM), high resolution radar, and acoustic, passive infrared, and seismic sensing modalities. The history in recent years has witnessed a great deal of effort in joint tracking and classification. In this section, we apply the integrated approach of [14] to joint tracking and classification via a simple yet representative example.

#### A. Problem Statement

Consider three types of data $z_1, z_2, z_3$ having the same size from three different sensors that are perfectly synchronized. For simplicity, our presentation below is for the case in which each data point is a scalar, but the approach works for the general vector case without difficulty. The first type is obtained from infrared imagers (or other energy-selective sensing devices) modeled as

$$z_{1j} = \theta x + v_j, \quad j = 1, \ldots, n$$

where $v_j \sim N(0, \sigma^2_v)$ is i.i.d. Gaussian noise; $x$ denotes the target state, which has a normal prior $N(\bar{x}, \sigma^2_x)$ and is independent of $v$; and the modulation term $\theta$ has two possible values,

$$H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1$$

which correspond to two possible classes of objects, e.g., humans and moving vehicles, respectively, since different classes of objects have different infrared features, assumed to be reflected in amplitude modulation.

The second type of data is obtained from ESM sensors (or some devices based on image, acoustic, or seismic features). The readings are in identity type and used to indicate different target classes with certain probabilities:

$$
\begin{align*}
P\{z_{2j} = \theta_0 | \theta = \theta_0\} &= 1 - p_0 \\
P\{z_{2j} = \theta_1 | \theta = \theta_0\} &= p_0 \\
P\{z_{2j} = \theta_0 | \theta = \theta_1\} &= 1 - p_1 \\
P\{z_{2j} = \theta_1 | \theta = \theta_1\} &= p_1
\end{align*}
$$

Assume $z_{21}, \ldots, z_{2n}$ are independent.

The third type of data is obtained from kinematic sensing devices, say, radar, modeled as

$$z_{3j} = x + w_j, \quad j = 1, \ldots, n$$

where $w_j \sim N(0, \sigma^2_w)$ is independent Gaussian noise.
The objective is not only to track the moving target \( x \) (with continuous uncertainty) but also to determine the target type \( \theta \). Note that type 2 and type 3 data do not provide direct information for tracking and classification, respectively. As a result, it is difficult to use type 2 data for tracking or use type 3 data for classification without joint decision and estimation.

### B. Conditional Independence

It is assumed that measurement errors from different sensors are independent and thus the three types of data are conditionally independent; that is,

\[
f(z_1, z_2, z_3|H_i, x) = f(z_1|H_i, x)f(z_2|z_1, H_i, x)f(z_3|z_1, H_i, x) = f(z_1|H_i, x)f(z_2|H_i)f(z_3|x)
\]

It turns out that type 1 and type 2 data are independent conditioned on the hypothesis (without \( x \)):

\[
f(z_1, z_2|H_i) = f(z_2|H_i)f(z_1|H_i) \tag{9}
\]

which follows from

\[
f(z_1, z_2|H_i) = \int f(z_1, z_2|H_i, x)f(x|H_i)dx = \int f(z_1|H_i, x)f(z_2|H_i)f(x)dx = f(z_2|H_i)f(z_1|H_i)
\]

Similar conditional independence holds between type 3 and type 2 data and between type 2 data and type 1 and type 3 data

\[
f(z_2, z_3|H_i) = f(z_2|H_i)f(z_3|H_i) \tag{10}
\]

\[
f(z_1, z_2, z_3|H_i) = f(z_2|H_i)f(z_1, z_3|H_i) \tag{11}
\]

but type 1 and type 3 data are not independent conditioned on the hypothesis (without \( x \))

\[
f(z_1, z_3|H_i) = \int f(z_1|H_i, x)f(z_3|x)f(z_3)f(x)dx = \int \mathcal{N}(z_1; \theta_i'x_1, \sigma^2_{w,1}I_n)\mathcal{N}(z_3; x_1, \sigma^2_{w,1}I_n)\mathcal{N}(x; \bar{x}, \sigma^2_x)dx
\]

where \( \mathcal{N}(y; \bar{y}, C_y) \) stands for the pdf of a Gaussian variable \( y \) with mean \( \bar{y} \) and covariance \( C_y \), and \( 1 = 1_n = [1, \ldots, 1]' \).

### C. Likelihood Functions

Under each hypothesis, type 1 and type 3 data, \( z = [z_1', z_3']' \), follow a linear model \( z = Hx + v \), where

\[
H_0: \quad H = H_0 = [\theta_0'1', 1']'
\]

\[
H_1: \quad H = H_1 = [\theta_1'1', 1']'
\]

and \( v = [v', w']' \) with \( R = \text{cov}(v) = \text{diag}(\sigma^2_{w,1}I_n, \sigma^2_{w,1}I_n) \).

Under each hypothesis, \( x \) and \( z \) are jointly Gaussian because they are two weighted sums of jointly Gaussian random variables \( x \) and \( v \) (since they are independent Gaussian). As such, \( z \) is Gaussian. Under \( H_i \),

\[
E[(z_1', z_3'|H_i)] = [\theta_i', 1']'\bar{x} = [\bar{z}_i', \bar{z}_3']'
\]

\[
\text{cov}([z_1', z_3'|H_i]) = [\theta_i'\sigma^2_{w,1}I_n + \sigma^2_{w,1}I_n] + R
\]

\[
= \begin{bmatrix}
\theta_i'\sigma^2_{11'} + \sigma^2_{11'} & \theta_i'\sigma^2_{11'} \\
\theta_i'\sigma^2_{11'} & \sigma^2_{11'} + \sigma^2_{w,1}I_n
\end{bmatrix} = C_{i3}
\]

where \( \bar{z}_3 = \bar{x}1, C_3 = \sigma^2_{11'} + \sigma^2_{w,1}I_n \), and, for \( i = 0, 1 \),

\[
\bar{z}_i = \theta_i'\bar{x}1, \quad C_i = \theta_i'\sigma^2_{11'} + \sigma^2_{w,1}I_n
\]

It thus follows that

\[
f(z_1, z_3|H_i) = \mathcal{N}([z_1', z_3']'; [\bar{z}_i', \bar{z}_3'], C_{i3}) \tag{12}
\]

\[
f(z_1|H_i) = \mathcal{N}(z_1; \bar{z}_i, C_i)
\]

\[
f(z_3|H_i) = \mathcal{N}(z_3; \bar{z}_3, C_3)
\]

Note that clearly \( f(z_1, z_3|H_i) \neq f(z_1|H_i)f(z_3|H_i) \).

Also, we clearly have

\[
f(z_2|H_i) = f(z_2|z_1, \ldots, z_n|H_i) = \prod_{j=1}^{n} [p_{ij} \delta_{zz_j - \theta_i} + (1 - p_{ij}) \delta_{zz_j - \theta_0}]
\]

### D. Classification by Bayesian Decision

The optimal Bayes decision minimizes the so-called Bayes risk

\[
\hat{R}_D = \sum_{i,j} c_{ij}P(\{H_i\}'|H_j)P(H_j)
\]

which is a special case of (1). It decides on the hypothesis \( H_i \) with the smallest posterior cost, that is, \( C_i(z) \leq C_l(z), \forall l \), where \( C_k(z) = \sum_j c_{kj}P(H_j|z) \) and \( z = (z_1, z_2, z_3) \).

In our example, we choose \( c_{00} = c_{11} = 0, c_{01} = c_{10} = 1 \). With this choice, the Bayes risk \( \hat{R}_D \) becomes the probability of decision error

\[
P_e = \sum_{i \neq j} P(\{H_i\}'|H_j)P(H_j)
\]

and thus the optimal Bayes decision becomes the minimum decision-error decision. It is well known that this amounts to maximum a posteriori (MAP) decision, which decides on the hypothesis having the maximum posterior probability.

It follows from Bayes’ theorem that the posterior probabilities are

\[
P\{H_0|z\} = \frac{P(H_0)f(z|H_0)}{P(H_0)f(z|H_0) + P(H_1)f(z|H_1)}
\]

\[
P\{H_1|z\} = 1 - P\{H_0|z\}
\]

While type 3 data \( z_3 \) is potentially helpful for decision (through some kind of estimation), it is not clear how it should be used in a purely decision setting. For example, if conditioned on \( x \) type 3 data is independent of the other types of data, as is the case for our problem, the likelihood ratio of \( H_1 \) to \( H_0 \) conditioned on \( x \) remains unchanged with or without type 3 data. As a result, in an implementation of
the optimal Bayes decision, type 3 data is often ignored in practice (although it can actually be used to help decision) and only the first two types of data (z₁ and z₂) are used. Now let \( z = (z₁, z₂) \) be the data used for decision.

Since each element in \( z₂ \) is Bernoulli distributed, to make decision on \( \theta \) is actually to infer Bernoulli parameter \( p \). It is known that \( z₂ = (1/n) \sum_{i=1}^{n} z₂i \) is a sufficient statistic for \( p \) [5]. Note that \( n z₂ \) counts the number \( γ \) of \( z₂i \)'s that are equal to 1. Thus, \( n z₂ \) has a binomial \((n, p)\) distribution:

\[
P(\hat{z}_2 = γ/n|H₁) = \binom{n}{γ} p^γ(1-p)^{n-γ}
\]

Thus, it follows from the conditional independence of type 1 and type 2 data that

\[
f(z|H₁) = f(z₁|H₁)f(z₂|H₁) = f(z₁|H₁)P(\hat{z}_2 = γ/n|H₁)
\]

\[
= N(z₁; \bar{z}, C₁) \binom{n}{γ} p^γ(1-p)^{n-γ}
\]

By assuming equal priors of both hypotheses, \( P(H₀) = P(H₁) \), from (13) it follows that

\[
P(H₀|z) = \left[ 1 + \left( \frac{p₀}{p₁} \right) \frac{(1-p₁)}{1-p₀} \frac{|C₀|^{1/2}}{|C₁|^{1/2}} \exp \left\{ \frac{1}{2} S \right\} \right]^{-1}
\]

\[
P(H₁|z) = 1 - P(H₀|z)
\]

where

\[
S = (z₁ - \bar{z})^T C₀^{-1}(z₁ - \bar{z}) - (z₁ - \bar{z})^T C₁^{-1}(z₁ - \bar{z})
\]

As such, the decision rule is

\[
S + 2 \ln \left[ \left( \frac{p₁}{p₀} \right) \frac{(1-p₁)}{1-p₀} \frac{|C₀|^{1/2}}{|C₁|^{1/2}} \right] \geq 2 \ln \frac{c₁₀ - c₀₀}{c₀₁ - c₁₁}
\]

which, for our choice of \((c₀₀, c₀₁, c₁₀, c₁₁) = (0, 1, 1, 0)\), is the same as the MAP decision decides on \( H₁ \) if \( P(H₁|z) > P(H₀|z) \).

### E. Tracking by Bayesian Estimation

The optimal Bayes estimator is the conditional mean \( \hat{x} = E[x|z] \). By the total expectation theorem,

\[
\hat{x} = E[x|z] = \sum_{i=0,1} E[x|z, H_i] P(H_i|z) = \sum_{i=0,1} \hat{x}_i P(H_i|z)
\]

\[
\text{MSE}(\hat{x}|z) = \sum_{i=0,1} [P_i + (\hat{x} - \hat{x}_i)(\hat{x} - \hat{x}_i)] P(H_i|z)
\]

By a similar argument as for the Bayesian decision, in Bayesian estimation, only type 1 and type 3 data, \( z₁ \) and \( z₃ \), are used directly and thus in the practical implementation, \( z = [z₁', z₃'] \). As explained in Sec. III-C, \( z \) and \( x \) are jointly Gaussian under each hypothesis. It is well known that, for the jointly Gaussian case, the conditional mean and its MSE matrix are given by, under \( H₁ \),

\[
\hat{x}_i = E[x|z, H₁] = \bar{x} + C₝ᵢCᵢ⁻¹(z - \bar{z})
\]

\[
P_i = \text{MSE}(\hat{x}_i|z, H₁) = Cₓ - CₓzᵢCᵢ⁻¹Cᵢzᵢ
\]

where \( Cₓzᵢ = \sigmaₓ²[\theta₁', 1'] \)

\[
Cᵢ = \sigmaᵢ²[\theta₁', 1'] + \text{diag}(\sigmaᵢ²Iₙ, \sigmaᵢ²Iₙ) = Cᵢ₃
\]

\[
zᵢ = [\theta₁, 1']^T \bar{x} = [z₁', \bar{z}'] \quad Cₓᵢ = \sigmaₓᵢ²
\]

The posterior probabilities of hypotheses are

\[
P(H₀|z) = \frac{P(H₀)f(z|H₀)}{P(H₀)f(z|H₀) + P(H₁)f(z|H₁)}
\]

\[
P(H₁|z) = 1 - P(H₀|z)
\]

where \( f(z|H₁) \) was given by (12). For equal prior probabilities of hypotheses, we have

\[
P(H₀|z) = \frac{f(z|H₀)}{f(z|H₀) + f(z|H₁)}
\]

### F. Classification before Tracking (Decision then Estimation)

In this traditional approach to JDE, a decision (target classification) is first made concerning the hypotheses \( H₀ \) and \( H₁ \), as in Sec. III-D, and then the target state \( x \) is estimated, as in Sec. III-E. More specifically, the data space is partitioned as \{\( D₀, D₁ \}\} by the decision rule first and the target state estimator is given by (4):

\[
\hat{x} = \sum_i 1(z|D_i)E[x|z, H_i]
\]

In other words, if the decision is “\( H₁ \)”, which should be made following Sec. III-D, then the estimate is \( \hat{x} = E[x|z, H₁] \) and MSE(\( \hat{x}|z, H₁ \)), given in Sec. III-E.

### G. Tracking before Classification (Estimation then Decision)

In this approach to JDE, the target state \( x \) is estimated first, as in Sec. III-E, and then a decision (target classification) is made concerning the hypotheses \( H₀ \) and \( H₁ \), as in Sec. III-D. More specifically, let \( \hat{x} \) be the Bayes target state estimator of Sec. III-E. Then

\[
f(z|\hat{x}, H₁) = f(z₁|H₁, \hat{x})f(z₂|H₁, \hat{x})
\]

\[
= N(z₁; \theta₁ \hat{x}, 1) \exp \left\{ -\frac{1}{2} \hat{x}^T \left( \frac{1}{\sigmaₚ₀^2} \right) (z₁ - \theta₁ \hat{x})^T \left( \frac{1}{\sigmaₚ₀^2} \right) (z₁ - \theta₁ \hat{x}) \right\}
\]

The Bayes test then decides on \( H₁ \) if

\[
\frac{f(z|\hat{x}, H₁)}{f(z|\hat{x}, H₀)} > \lambda = \frac{c₁₀ - c₀₀}{c₀₁ - c₁₁}
\]

Our choice \((c₀₀, c₀₁, c₁₀, c₁₁) = (0, 1, 1, 0)\) and equal prior probabilities of hypotheses \( P(H₀) = P(H₁) \) lead to \( \lambda = 1 \). Then the test can be simplified as

\[
n(\theta₁ - \theta₀) \hat{x} z₁₁ - (\theta₁ + \theta₀) \hat{x} / 2 + \ln \left[ \frac{p₁}{p₀} \frac{(1-p₁)}{1-p₀} \frac{|C₀|^{1/2}}{|C₁|^{1/2}} \right] \geq 0
\]

where the sample mean \( \bar{z}₁ = \frac{1}{n} \sum_{j=1}^{n} z₁j \).
The first type of data can serve both tracking and classification purposes directly. The second type is useful for decision but has no direct impact on estimation; the third type is useful for estimation but has no direct impact on decision. As a result, in the usual implementations of the traditional Bayes decision, Bayes estimation, and the two-stage approaches, one type of data is not used directly for either classification (decision) or tracking (estimation). However, our proposed joint approach uses all data without difficulty. Its performance is generally superior since more information is used.

The joint solution can be achieved by iteration. Although the iteration may start from any decision/estimation results, we choose the one with smaller JDE cost from the conventional solutions. For the choice of JDE cost weights \( \{\alpha_{ij}, \beta_{ij}\} \) in (1), the values of \( \alpha_{ij} \) will modify the decision cost in detail, and generally speaking, we would like to choose \( \beta_{01}, \beta_{10} < \beta_{00}, \beta_{11} \) so that the estimation errors costs are taken into account (otherwise the generalized Bayes risk will be dominated by the cost associated with decision errors).

For simplicity, here our JDE algorithm starts from the Bayes decision results: If the decision is “\( H_i \)”, i.e., \( z \in D_i \), from (3), we have

\[
\tilde{x}_i = \tilde{x}_i = \tilde{E}[x|z, D_i] = \int E[x|z, D_i, H_j]P(H_j|z) = \sum_j \tilde{x}_j P(H_j|z)
\]

where

\[
P(H_j|z) = \frac{\beta_{ij}P(H_j|z)}{\beta_{0j}P(H_0|z) + \beta_{1j}P(H_1|z)}, z \in D_i
\]

Then

\[
\tilde{x}_i = \tilde{x}_0P(H_0|z) + \tilde{x}_1P(H_1|z)
\]

\[
= \frac{\tilde{x}_0\beta_{0i}P(H_0|z) + \tilde{x}_1\beta_{1i}P(H_1|z)}{\beta_{0i}P(H_0|z) + \beta_{1i}P(H_1|z)}
\]

To calculate the posterior JDE cost

\[
R(z) = \sum_i \sum_j (c_{ij} + \beta_{ij}E[\tilde{x}^2|D_i, H_j, z])P(D_i, H_j|z)
\]

the key is to obtain the part imported by estimation. It can be shown that

\[
mse(\tilde{x}|D_i, H_j, z) \triangleq E[\tilde{x}^2|D_i, H_j, z] = \text{mse}(\tilde{x}_{ij}|z, D_i, H_j) + (\tilde{x}_{ij} - \tilde{x})(\tilde{x}_{ij} - \tilde{x})
\]

where, for \( z \in D_i \),

\[
\tilde{x}_{ij} = E[x|z, D_i, H_j] = E[x|z, H_j]
\]

\[
mse(\tilde{x}_{ij}|z, D_i, H_j) = \text{mse}(\tilde{x}_{ij}|z, H_j)
\]

and \( \tilde{x} \) is obtained by (3). Since, under \( z \in D_i \),

\[
\tilde{x}_{ij} - \tilde{x} = \tilde{x}_j - \sum_i 1(z; D_i)\tilde{x}_i = \tilde{x}_j - \tilde{x}_i
\]

we have

\[
E_{ij} \triangleq \text{mse}(\tilde{x}|D_i, H_j)
\]

\[
= E[E(\tilde{x}_{ij}|z, D_i, H_j)|D_i, H_j] + E(E(\tilde{x}_j - \tilde{x})'(\tilde{x}_j - \tilde{x})|D_i, H_j)
\]

\[
= \text{mse}(\tilde{x}_{ij}|D_i, H_j) + E[(\tilde{x}_j - \tilde{x}_i)'(\tilde{x}_j - \tilde{x}_i)|D_i, H_j]
\]

\[
= \text{mse}(\tilde{x}_j|D_i, H_j) + E[(\tilde{x}_j - \tilde{x}_i)'(\tilde{x}_j - \tilde{x}_i)|D_i, H_j]
\]

For each decision region,

\[
mse(\tilde{x}_{ij}|z, D_i, H_j) = E[E(\tilde{x}_{ij}|z, D_i, H_j)|D_i, H_j]
\]

\[
= \text{mse}(\tilde{x}_j|z, H_j)
\]

Since it does not depend on observations, we have

\[
mse(\tilde{x}_{ij}|D_i, H_j) = E[mse(\tilde{x}_{ij}|z, D_i, H_j)]
\]

\[
= \text{mse}(\tilde{x}_j|z, H_j)
\]

Note that under \( z \in D_i \)

\[
\tilde{x}_j - \tilde{x}_i = \tilde{x}_j - \sum_k \tilde{x}_k \beta_{ik}P(H_k|z)
\]

\[
= \frac{\tilde{x}_j - \tilde{x}_i}{\sum_k \beta_{ik}P(H_k|z)}
\]

Therefore

\[
\tilde{E}_{ij} \triangleq \tilde{E}.\tilde{E}_{ij} = \int_{z \in D_i} \tilde{E}_{ij}(\tilde{x}_j - \tilde{x}_i)'(\tilde{x}_j - \tilde{x}_i)|D_i, H_j)
\]

\[
= \int_{z \in D_i} (\tilde{x}_j - \tilde{x}_i)'(\tilde{x}_j - \tilde{x}_i)|D_i, H_j)
\]

\[
= E \left[ \frac{\tilde{y}_i\tilde{y}_i}{(\sum_k \beta_{ik}P(H_k|z))|D_i, H_j)} \right]
\]

where \( \tilde{y}_i = \sum_k (\tilde{x}_j - \tilde{x}_i)\beta_{ik}P(H_k|z) \). It can be obtained by the Monte-Carlo method numerically

\[
\tilde{E}_{ij} \approx \frac{1}{L_i} \sum_{k=1}^{L_i} \left[ \tilde{x}_j(\tilde{z}^{(i)}_k) - \tilde{x}_i(\tilde{z}^{(i)}_k) \right]' \left[ \tilde{x}_j(\tilde{z}^{(i)}_k) - \tilde{x}_i(\tilde{z}^{(i)}_k) \right]
\]

where \( L_i \) and \( L_i \) are the measurement counts in the Monte-Carlo simulation: Generate data points \( z_1, \ldots, z_{L_i} \) with the distribution \( F(z|H_j) \). Use the decision part to collect all the points \( z_1^{(i)}, \ldots, z_{L_i}^{(i)} \) in \( D_i \), where \( \sum_i L_i = L \).

Let \( c'_{ij} = \alpha_{ij}c_{ij} + \beta_{ij}E_{ij} \). If \( c'_{0j} > c'_{00} \) and \( c'_{01} > c'_{11} \), for the next iteration, decide on \( H_1 \) if

\[
S + 2 \ln \left[ \frac{(p_1)}{(p_0)} \frac{1}{(1 - p_0)} \frac{(1 - p_1)}{p_0} \right] \gamma^{-n-\gamma} \frac{C_{01}^{1/2}}{C_{11}^{1/2}} > 2 \ln \frac{c'_{10} - c'_{00}}{c'_{01} - c'_{11}}
\]

otherwise decide on \( H_1 \) if

\[
(c'_{01} - c'_{11}) \gamma^{-n-\gamma} \frac{C_{01}^{1/2}}{C_{11}^{1/2}} > 2 \ln \frac{c'_{00} - c'_{11}}{c'_{10} - c'_{00}}
\]

The most straightforward stopping criterion is to check the difference of \( \tilde{R} \) in two adjacent iterations, which is not easy to calculate. Alternatively, we may stop the iteration if \( \max_{i,j} |E_{ij} - E_{ij}^{(k+1)}| \) is below a threshold and there is no change in the decision (i.e., \( z \in (D_i^{(k)} \cap D_i^{(k+1)}) \), as explained in [14].
I. Performance Evaluation

In this example,
\[ f(z_1|\theta, x) = \mathcal{N}(z_1; \theta, x, \sigma^2) I \]
\[ f(z_2|\theta, x) = \prod_{j=1}^{n} \left[ p_i \delta_{z_{2j}-\theta_1} + (1-p_i) \delta_{z_{2j}-\theta_0} \right] \]
\[ f(z_3|\theta, x) = \mathcal{N}(z_3; x, \sigma^2) I \]
\[ f(z_1, z_2, z_3|\theta, x) = f(z_1|\theta, x) f(z_2|\theta, x) f(z_3|\theta, x) \]

Let \( \hat{f}(z|\theta, \hat{x}) = f(z|\theta, x)|_{(\theta, x) = (\hat{\theta}, \hat{x})} \)

Considering pmf, we have
\[ g(z|\theta, x) = f_1(z_1|\theta, x) p_2(z_2|\theta, x) f_3(z_3|\theta, x) \]
\[ = \mathcal{N}(z_1; \theta_1, x_1, \sigma^2_1) \mathcal{N}(z_3; x_1, \sigma^2_3) \prod_{j=1}^{n} \left[ p_i \delta_{z_{2j}-\theta_1} + (1-p_i) \delta_{z_{2j}-\theta_0} \right] \]

and
\[ \hat{g}(z|\theta, \hat{x}) = \mathcal{N}(z_1; \hat{\theta}_1, \hat{x}_1, \sigma^2_1) \mathcal{N}(z_3; x_1, \sigma^2_3) \prod_{j=1}^{n} \left[ \hat{p}_i \delta_{z_{2j}-\theta_1} + (1-\hat{p}_i) \delta_{z_{2j}-\theta_0} \right] \]

IV. Simulation Results

In our simulation example, the following parameter values were used
\[ \theta_0 = 1, \theta_1 = 2, p_0 = 0, p_1 = 0.65, \]
\[ \hat{x} = 1, \sigma^2_x = 0.5^2, \sigma^2_w = 1, \sigma^2_w = 0.5 \]

With data length \( n = 10 \), the simulation results are based on \( M = 500 \) Monte-Carlo runs.

A. Scenario 1: Data generated from \( H_0 \)

In this case, the simulated data is generated from \( H_0 \) (humans). The inference results are listed in Table I.

To obtain the above results, the weights of JDE cost are chosen as \( \alpha_{ij} = \alpha = 1 \), \( i, j = 0, 1 \). \( \beta_{00} = \beta_{11} \). \( \beta_{01} = \beta_{10} \). We add the constraint \( \sum_{ij} \beta_{ij} = B \) for the comparison purpose. Notice that the maximum possible mse (corresponding to the case where all classification results are incorrect) is around 0.35, to balance the decision and estimation impact, we chose \( B = 3 \).

The root-mean-square errors (RMSE) of tracking the target state and the probability of correct classification \( P_C \) are two conventional metrics for the tracking and classification performance, respectively, but not jointly. We are not aware of any measure in the literature for the overall joint performance.

The JDE performance metric \( \rho \) proposed in Sec. II-B is used to measure the joint performance. For \( N_j = 200, N_j = 5 \), the results are shown in the lower part of Table I.

Consider tracking errors. In the ideal case (always identify the target correctly) the root-mean square error is \( \text{RMSE}_{\text{ideal}} = 0.1767 \); for tracking without classification or tracking before classification, \( \text{RMSE}_E = \text{RMSE}_{E-D} = 0.2532 \), which equals \( \text{RMSE}_{\text{JDE}} \) with \( \beta_{ij}/\beta_{ii} = 1 \), as they should be; for classification before tracking, \( \text{RMSE}_{D-E} = 0.2403 \). This indicates that decision (classification) helps estimation (tracking) noticeably.

Consider classification performance now. For classification without tracking or classification before tracking, \( (P_C)_{D} = (P_C)_{D-E} = 0.798 \), which equals \( (P_C)_{\text{JDE}} \) with \( \beta_{ij}/\beta_{ii} = 1 \), as they should be; for tracking before classification, \( P_C = 0.894 \). This indicates that estimation (tracking) also helps decision (classification) significantly.

These results for tracking and classification verify that in the \( H_0 \) case for this example, decision and estimation will help each other. However, it is hard to compare the overall performance of the different strategies since none of the strategies is always better than the others in terms of both decision and estimation results. The JDE performance index \( \rho \) Sec. II-B is extremely useful in such a case where having an overall ranking is desirable.

We calculated the \( \rho \) values for the two conventional strategies: For classification before tracking, \( \rho_{E-D} = 0.9976 \times 10^{-2} \); for tracking before classification, \( \rho_{E-D} = 1.0105 \times 10^{-2} \). Note that \( \rho_{E-D} \) is not equal to \( \rho_{\text{JDE}} \) with \( \beta_{ij}/\beta_{ii} = 1 \), although \( \text{RMSE}_{E} = \text{RMSE}_{E-D} = \text{RMSE}_{\text{JDE}}(\beta_{ij}/\beta_{ii} = 1) \) and \( (P_C)_{D} = (P_C)_{D-E} = (P_C)_{\text{JDE}}(\beta_{ij}/\beta_{ii} = 1) \). As a distance measure, a smaller \( \rho \) value indicates better performance. Therefore we conclude that decision then estimation is indeed better than estimation then decision in terms of JDE performance.

By checking the \( \rho \) values of the proposed JDE solution listed in the table, we can see clearly that the JDE solution with \( \beta_{ij}/\beta_{ii} = 0 \) has the smallest \( \rho \) value, meaning that overall it outperforms the two existing strategies, although its decision performance is worse than that of the tracking before classification strategy. This weight choice agrees with our intuition that \( \beta_{01}, \beta_{10} \) should be smaller than \( \beta_{00}, \beta_{11} \).

B. Scenario 2: Data generated from \( H_1 \)

In this case, the simulated data was generated from \( H_1 \) (moving vehicles). The results are listed in Table II.

The weights of the JDE cost are the same as those in Scenario 1. For the tracking errors, \( \text{RMSE}_{\text{ideal}} = 0.1280; \text{RMSE}_{E} = \text{RMSE}_{E-D} = 0.1470; \text{and RMSE}_{D-E} = 0.1493. \) This time, decision (classification) does not help estimation (tracking). For the classification results, for classification without tracking or classification before tracking, \( P_C = 0.984 \); for tracking before classification, \( P_C = 0.882 \). Estimation (tracking) does not help decision (classification), either. This indicates that in this case, the conventional strategies cannot utilize the coupling between decision and estimation well to improve the inference results.

By checking the upper part of the table, we found that the JDE solution with \( \beta_{ij}/\beta_{ii} = 0 \) outperforms the decision-then-estimation and estimation-then-decision in decision and estimation performance concurrently. Then we checked the comprehensive performance index \( \rho \): For classification before tracking, \( \rho_{D-E} = 0.8122 \times 10^{-2} \); for tracking before
classification. \( \rho_{\text{E-D}} = 0.8740 \times 10^{-2} \). By comparing the corresponding \( \rho \) values of the JDE solution listed in the table, we can also draw the same conclusion that the JDE solution with \( \beta_{ij} / \beta_{ii} = 0 \) is better than the two existing strategies in terms of the overall performance.

### V. Conclusions and Discussion

In this paper, we applied the new proposed JDE framework to a joint target tracking and classification problem. The performance of the JDE solution was compared with two other existing strategies using both conventional methods and a joint performance index proposed by us. With an appropriate weight choice of the JDE cost, the JDE solution outperforms the other two strategies by overall evaluation. Note that the weight design of the JDE costs is problem dependent and subject to change based on the user’s preference. For instance, if we want to pay more attention to the estimation part, we should change based on the user’s preference. For instance, if we want to pay more attention to the estimation part, we should change based on the user’s preference. For simplicity, the idea was illustrated by batch processing. The solution to the dynamic case is under investigation.

### REFERENCES


