Cross-Entropy Method for K-Best Dependent-Target Data Association Hypothesis Selection

Shozo Mori, and Chee Chong
BAE Systems, Advanced Information Technologies
Los Altos, CA, U.S.A.
{shozo.mori,chee.chong}@baesystems.com

Abstract ***- This paper is concerned with probabilistic evaluation of multiple-frame data association hypotheses in multiple-target tracking problems, in particular, when targets are not necessarily independent a priori. Multiple-target tracking problems with dependent targets naturally arise whenever targets interact with each other, as they move in congested traffic, or as they actively coordinate their movements in other situations. This paper develops a Bayesian data association hypothesis evaluation formula for dependent targets. Because the resulting formula does not have a multiplicative or log-linear form, the best hypothesis cannot be selected by integer linear programming or multi-dimensional assignment algorithms commonly used to solve data association problems in multiple target tracking. Instead, we propose to use Reuven Rubinstein's cross-entropy method as a possible solution. A K-best hypothesis selection extension will be discussed as an application of the generalized Murty's algorithm. This paper focuses on the theoretical aspects as the first step of a solution concept development.***

Keywords: Multiple target tracking, dependent targets, K-best data association hypothesis selection, Reuven Rubinstein's cross-entropy method, generalized Murty's algorithm.

1 Introduction

This paper is concerned with calculation of the probabilities of multiple-frame data association hypotheses when tracking multiple objects, called targets, i.e., in multiple-target tracking problems, as defined in [1-5], particularly when targets are not necessarily independent a priori. Multiple-target tracking problems with dependent targets naturally arise whenever targets interact with each other, as they move in congested traffic, or they are in coordinated movements.

As shown in [1-5], data association hypotheses can be represented as a consistent set of tracks, each of which is a set of measurements that are hypothesized to originate from the same target. As shown in [3] for the batch-processing form, and in [4] for the recursive form, the a posteriori probability of a data association hypothesis is obtained as the normalization of the product of the track likelihoods[1] of the tracks in the hypothesis. It was shown in [5] that a set of independence and Poisson assumptions is a sufficient condition for the Bayesian evaluation formula to be multiplicative. A necessary condition was later shown in [6].

Since the consistency among tracks in any given data association hypothesis can be expressed by a set of linear equalities or inequalities with a binary matrix and a binary vector, the multiplicative likelihood or the additive log-likelihood allows the association problem to be solved by general zero-one integer linear programming algorithms [3], or more specialized so-called multiple-dimensional assignment algorithms, notably using relaxation methods as shown in [7,8]. On the other hand, when targets are not independent from each other a priori, the evaluation formula may not have a multiplicative or additive form anymore, as shown later in this paper. Then, the problem cannot be solved by any algorithm that relies on the multiplicity or the additivity of the objective function.

Thus, *the objective of this paper is to develop a formula and an algorithm for evaluating a posteriori probability of data association hypotheses when targets are not necessarily independent a priori.* Specifically, we propose the use of *Reuven Rubinstein's cross-entropy method* [9] for selecting the maximum a posteriori (MAP) data association hypothesis. It is claimed [9] that the Reuven Rubinstein's cross entropy method provides a very effective stochastic optimization algorithm for a wide range of combinatorial optimization problems. Recently it was reported [10] that the cross-entropy method can be effectively used to solve multiple target tracking data association problems. Although [10] addresses the data association problems when the targets are independent a priori, the cross entropy method does not require the objective function to be linear, and this paper tries to exploit that fact.

Stochastic optimization algorithms, including the cross-entropy method and other well-known methods such as the *Markov Chain Monte Carlo* (*MCMC*) *method* [11,18], are attractive alternatives to numerical

---

[1] By the likelihood, we mean any quantity proportional to a certain probability, or un-normalized probability density or mass function, not a likelihood in a classical statistics sense.

optimization approaches such as integer linear programming [3] and relaxation-based linear-objective multiple-dimensional assignment algorithms [7,8], to solve data association problems, i.e., to obtain the MAP association hypothesis, when the effective target density rises above certain thresholds, as reported in [12]. Unlike the MCMC method, as described in [11,18], the cross-entropy method does not require the likelihood function to be multiplicative or the log-likelihood function to be additive.

Another important objective of solving the data association problem for multiple target tracking is to assess the uncertainty of the MAP data association hypothesis in addition to the problem of finding the MAP hypothesis itself. The *generalized Murty's algorithm* described in [13] provides a means for obtaining approximate uncertainty assessment by finding the *k-best multi-frame data association hypotheses*. Unlike the other k-best algorithms such as the relaxation-based method [14], the generalized Murty's algorithm does not require linear objective functions. Thus, the combination of the Reuven Rubinstein's cross-entropy method and the generalized Murty's algorithm can be used to select the MAP hypothesis and its approximate uncertainty assessment.

The next section, Section 2, defines a class of multiple target tracking problems where the targets are not necessarily independent a priori. Section 3 derives a formula for computing the probabilities of data association hypotheses, and defines a zero-one integer programming problem with an objective function that is not necessarily linear. Section 4 describes a method for selecting the single best MAP hypothesis, as well as a method for selecting the k-best solutions to the problem, thus providing an approximate method for the uncertainty assessment of the selected data association hypothesis. Some concluding remarks will be given in Section 5.

2 Models with Dependent Targets: Problem Statement

Let us consider a problem of *tracking* or dynamically estimating the states of an unknown number *n* of objects, called *targets*, as defined as a joint Markovian stochastic process[2] $(x_i(t))_{i=1}^n$ with continuous time parameter *t* as $t \in [t_0, \infty)$, on the target state space *E*, a Euclidian space[3]. It is assumed that, given number *n* of targets, the joint target state process is defined by a joint initial probability density, $\overline{f}_{nt_0}((x_i(t_0))_{i=1}^n)$, and a joint transition probability density, $f_{n\Delta t}((x_i(t+\Delta t))_{i=1}^n \mid (x_i(t))_{i=1}^n)$, for each $t \in [t_0, \infty)$ and $\Delta t > 0$.

We assume the *permutability* (or the symmetry or the exchangeability) of the targets, by requiring the initial condition and the transition probability to be permutable, i.e., for any given *n*, $\overline{f}_{nt_0}((x_{\alpha(i)}(t_0))_{i=1}^n) = \overline{f}_{nt_0}((x_i(t_0))_{i=1}^n)$ and $f_{n\Delta t}((x_{\alpha(i)}(t+\Delta t))_{i=1}^n \mid (x_{\alpha(i)}(t))_{i=1}^n) = f_{n\Delta t}((x_i(t+\Delta t))_{i=1}^n \mid (x_i(t))_{i=1}^n)$, for any permutation α on $\{1,...,n\}$. We assume the number *n* of targets is fixed[4] but random and generally unknown, with a given probability mass function $\overline{p}(n)$ and a finite mean $\overline{\nu} \overset{\text{def}}{=} \sum_{n=1}^{\infty} n\overline{p}(n) < \infty$. The randomness of the number (cardinality) *n* of targets, together with the permutability of the joint state process, defines the unique characteristics[5] of multiple target tracking problems, and distinguishes it from other estimation problems.

Consider a series, $y_1, y_2, \ldots$, of *measurement frames*, each of which, y_k, is a finite sequence of measurements as, $y_k = (y_{kj})_{j=1}^{m_k}$, taken at the same time $t_k \in [t_0, \infty)$ by the same sensor (or almost at the same time so that any two measurements in each frame do not originate from the same target), with length m_k, and $t_0 \le t_1 \le t_2 \le \cdots$. Each y_{kj} is the *j*-th *measurement* in the *k*-th frame, which lies in measurement space E_{Mk} that is a measurable set in a Euclidean space (See Footnote 3). As in many other dynamical state estimation problems, a key problem is to evaluate the a posteriori probability[6] $P(x_i(t_K) \mid (y_k)_{k=1}^K)$ for a given *K*. However, in many multiple target tracking applications, we are interested also in data association, i.e., determining the measurements that originate from the same targets. This paper is concerned with such situations.

Data association hypotheses are defined as follows: For each frame $y_k = (y_{kj})_{j=1}^{m_k}$, define a random integer-valued function a_k, called the *assignment function*, such that[7] $\mathrm{Dom}(a_k) \subseteq \{1,....,n\}$, $\mathrm{Im}(a_k) \subseteq \{1,....,m_k\}$, and a_k is an one-to-one function with probability one. When $a_k(i) = j$, the *j*-th measurement of the *k*-th frame is said to *originate from* the *i*-th target. When $i \notin \mathrm{Dom}(a_k)$, the

---

[2] $(\xi_i)_{i=1}^n$ is a shorthand of $(\xi_1, ..., \xi_n)$, a finite sequence of length *n*.

[3] The target state space, as well as the measurement spaces, can be a so-called hybrid space to accommodate discrete state components, or more generally, any locally compact Hausdorff space satisfying the second axiom of countability. But we maintain the Euclidean assumption for the sake of simplicity.

[4] Largely the fixed-number-of-target assumption is for the sake of simplicity but we should note that this assumption does not imply a fixed number of detected targets.

[5] The random cardinality and the permutability make the targets effectively modeled as a *finite point process* or *random finite set*, as discussed in [15].

[6] We use *P* as the generic symbol for conditional or unconditional probability density or mass functions.

[7] For any function *a*, by $\mathrm{Dom}(a)$ we mean the *domain* of the function *a*, and by $\mathrm{Im}(a)$ we mean the *image* (or the *range*) of *a*, i.e., $\mathrm{Im}(a) = \{a(i) \mid i \in \mathrm{Dom}(a)\}$.

i-th target is said to be *not detected* in frame y_k, and when $j \notin \mathrm{Im}(a_k)$, the *j*-th measurement is a *false alarm*.

For a given integer *K*, let

$$\lambda = \{\{(k,j) \mid k = 1,\ldots,K,\ j = a_k(i)\} \mid i \in \bigcup\nolimits_{k=1}^{K} \mathrm{Dom}(a_k)\} \tag{1}$$

Then we call any possible realization of λ a *data association hypothesis on cumulative frames* $(y_k)_{k=1}^{K}$. It follows from this definition that any data association hypothesis λ on $(y_k)_{k=1}^{K}$ is a consistent (non-overlapping) set of nonempty *tracks* on $(y_k)_{k=1}^{K}$, where a track on $(y_k)_{k=1}^{K}$ is defined as any subset τ of $\bigcup_{k=1}^{K}\{k\} \times \{1,\ldots\ldots,m_k\}$, such that, for any $k = 1,\ldots,K$, there is at most one $j \in \{1,\ldots,m_k\}$ for which we have $(k,j) \in \tau$ (definition of a frame). The consistency requirement for any hypothesis λ means that, for any distinct pair (τ_1,τ_2) of tracks in λ, we must have $\tau_1 \cap \tau_2 = \emptyset$. Let the set of all the tracks on $(y_k)_{k=1}^{K}$ be denoted by $\mathcal{T}_K$. Then we can define the set $\mathit{\Lambda}_K$ of all the data association hypotheses by

$$\mathit{\Lambda}_K = \left\{ \lambda \subseteq \mathcal{T}_K \,\middle|\, \begin{array}{l} \emptyset \notin \lambda \text{, and } \tau_1 \cap \tau_2 = \emptyset \\ \text{if } \tau_1 \in \lambda, \tau_2 \in \lambda \text{ and } \tau_1 \neq \tau_2 \end{array} \right\} \tag{2}$$

A track τ on $(y_k)_{k=1}^{K}$ defined above is completely equivalent to an integer-valued function defined on the set $\{1,\ldots,K\}$, where $j = \tau(k)$ for $k \in \{1,\ldots,K\}$ if the track includes the pair (k,j) of indices, $\tau(k) = 0$ otherwise. Any track τ on $(y_k)_{k=1}^{K}$ defined this way satisfies $\tau(k) \in \{0,1,\ldots,m_k\}$ for every $k \in \{1,\ldots,K\}$. We will use these two representations of a track interchangeably in the rest of this paper whenever there is no fear of confusion. We call the original representation the *set-theoretic representation*, while the latter the *mapping representation*. When the empty track $\emptyset$ is used as the function representation, we define as $\emptyset(k) = 0$ for any *k*. We will also use notation $y_{k0} = \theta$ for convenience, where θ is the symbol for “nothing”, signifying the mis-detection at frame *k*.

The probabilistic evaluation of a data association hypothesis $\lambda \in \mathit{\Lambda}_K$ *is to numerically evaluate the a posteriori probability* $P(\lambda \mid (y_k)_{k=1}^{K})$. The maximum a posteriori (MAP) data association hypothesis $\hat{\lambda}$ is the hypothesis that maximizes $P(\lambda \mid (y_k)_{k=1}^{K})$, and K-best hypothesis selection is to select the top K ranked hypotheses in terms of $P(\lambda \mid (y_k)_{k=1}^{K})$. The K-best hypothesis selection for sufficiently large K can be used to approximately calculate the posterior probability of the MAP hypothesis, which we define as our uncertainty assessment solution. Section 3 derives a probabilistic evaluation formula, and Section 4 discusses hypothesis selection problems, based on the evaluation formula developed in Section 3.

The set of additional assumptions described below defines a general class of target/sensor models used to determine the scope of the problems for this paper. First of all, we assume the conditional independence of the frames, in terms of measurements and the underlying random assignments as,

$$\begin{aligned} &P(((y_{kj})_{j=1}^{m_k})_{k=1}^{K}, (a_k)_{k=1}^{K} \mid ((x_i(t_k))_{i=1}^{n})_{k=1}^{K}) \\ &\qquad = \prod_{k=1}^{K} P((y_{kj})_{j=1}^{m_k}, a_k \mid (x_i(t_k))_{i=1}^{n}) \end{aligned} \tag{3}$$

for each *K*.

For each frame *k*, we assume the target-wise independent detection as

$$\begin{aligned} &P(\mathrm{Dom}(a_k) \mid (x_i(t_k))_{i=1}^{n}) = \\ &\left(\prod_{i \in \mathrm{Dom}(a_k)} p_{Dk}(x_i(t_k)) \right) \left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(a_k)}}^{n} (1 - p_{Dk}(x_i(t_k))) \right) \end{aligned} \tag{4}$$

with the probability $p_{Dk}(x)$ of a target at state $x \in E$ being detected in frame *k*.

Let us assume that the errors contained in measurements originating from targets are target-wise independent, and that the values of false alarms are independent from each other and from target detections, so that we have

$$\begin{aligned} &P((y_{kj})_{j=1}^{m_k} \mid m_k, a_k, (x_i(t_k))_{i=1}^{n}) = \\ &\left(\prod_{i \in \mathrm{Dom}(a_k)} p_{Mk}(y_{k a_k(i)} \mid x_i) \right) \left(\prod_{\substack{j=1 \\ j \notin \mathrm{Im}(a_k)}}^{m_k} p_{Fk}(y_{kj}) \right) \end{aligned} \tag{5}$$

with the target-state-to-measurement transition probability density[8] p_{Mk} that generally depends on (k,j), and the false alarm value probability density p_{Fk}. We assume[9]

---

[8] Typically each measurement y_{kj} is given in terms of a pair of mean and covariance for a given Euclidean space E_{Mk}. The lack of p_{Mk} the dependence on the measurement index *j* implicitly implies all the measurement shares the same measurement error statistics. We can consider the measurement error statistics that are dependent of the index *j* by considering the “reported” measurement error statistics as being included in the measurement value y_{kj}.

[9] This strict positivity assumption, as well as the Poisson assumption mentioned later, is not essential but made here for the sake of simplicity.

$p_{Fk}(\eta) > 0$ for all $\eta \in E_{Mk}$. Furthermore, we assume that the number[10] $m_{Fk} \overset{\text{def}}{=} m_k - \#(\mathrm{Dom}(a_k)) = m_k - \#(\mathrm{Im}(a_k))$ of false alarms is independent of anything else and is modeled by a Poisson random variable with mean $\nu_{Fk} > 0$.

Finally, we assume that each random assignment a_k is *totally random* in the sense that, given a hypothesized set $\mathrm{Dom}(a_k)$ of indices of detected targets and the number m_k of measurements in frame *k*, all the possible $m!/(m - \#(\mathrm{Dom}(a_k)))!$ realizations of a_k are equally likely, i.e.,

$$P(a_k \mid \mathrm{Dom}(a_k), m_k) \equiv \frac{(m_k - \#(\mathrm{Dom}(a_k)))!}{m_k\,!} \tag{6}$$

By putting everything together, including (4-6),we have

$$P((y_{kj})_{j=1}^{m_k} \mid (x_i(t_k))_{i=1}^{n}) = \frac{e^{-\nu_{Fk}}}{m_k\,!} \sum_{a_k \in \mathcal{A}(n,m_k)} \left(\prod_{\substack{j=1 \\ j \notin \mathrm{Im}(a_k)}}^{m_k} \gamma_{Fk}(y_{kj}) \right) \left(\prod_{i \in \mathrm{Dom}(a_k)} p_{Mk}(y_{ka_k(i)} \mid x_i(t_k)) p_{Dk}(x_i(t_k)) \right) \left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(a_k)}}^{n} (1 - p_{Dk}(x_i(t_k))) \right) \tag{7}$$ [11]

where $\gamma_{Fk}(y_{kj}) \overset{\text{def}}{=} \nu_{Fk} p_{Fk}(y_{kj})$ is the false alarm density at y_{kj}. This joint target-state-to-frame transition probability density (7), together with the conditional independence assumption (3), gives us a standard Bayesian filtering equation. However, because the problem for this paper is to numerically evaluate each data association hypothesis, which will be done in the next section, eqn. (7) is not useful for us.

3 Evaluation of Data Association Hypotheses When Targets Are Not Necessarily Independent A Priori

As defined in the previous section, our problem is, given cumulative frames $(y_k)_{k=1}^{K} = ((y_{kj})_{j=1}^{m_k})_{k=1}^{K}$, to evaluate data association hypotheses λ on $(y_k)_{k=1}^{K}$, i.e., to numerically calculate the a posteriori probability $P(\lambda \mid (y_k)_{k=1}^{K})$.

Since a data association hypothesis λ explicitly depends on the number m_k of measurements in every frame $k = 1,...,K$, $P(\lambda \mid (y_k)_{k=1}^{K})$ must be evaluated through $P(\lambda \mid (y_k)_{k=1}^{K}) = P((y_k)_{k=1}^{K})^{-1} P(\lambda, (y_k)_{k=1}^{K})$. The joint probability-mass-density function $P(\lambda, (y_k)_{k=1}^{K})$ can be expanded by random assignments $(a_k)_{k=1}^{K}$ and the joint target state $((x_i(t_k))_{k=1}^{K})_{i=1}^{n}$, using all the assumptions made in the previous section, as

$$\begin{aligned} &P(\lambda, (y_k)_{k=1}^{K}) \\ &= \sum_{n=0}^{\infty} \int_{E^{Kn}} \sum_{(a_k)_{k=1}^{K}} P(\lambda, (y_k)_{k=1}^{K}, (a_k)_{k=1}^{K}, ((x_i(t_k))_{k=1}^{K})_{i=1}^{n}) \\ &= \prod_{k=1}^{K} e^{-\nu_{Fk}} \prod_{j=1}^{m_k} \gamma_{Fk}(y_{kj})^{-1} \sum_{n=\#(\lambda)}^{\infty} \frac{n!\,\bar{p}(n)}{(n - \#(\lambda))!} Q_K(\lambda, n) \end{aligned} \tag{8}$$

where the second summation on the second line of (8) is meant to be over all the assignments $(a_k)_{k=1}^{K}$ that *supports* hypothesis λ through eqn. (1), and

$$Q_K(\lambda, n) \overset{\text{def}}{=} \int_{E^n} \bar{f}_{nt_0}((x_i(t_0))_{i=1}^{n}) \left(\prod_{k=1}^{K} \int_{E^n} \left(\prod_{i \in \mathrm{Dom}(\bar{a}_k)} q_{Mk}(y_{k\bar{a}_k(i)} \mid x_i(t_k)) \right) \left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_k)}}^{n} (1 - p_{Dk}(x_i(t_k))) \right) f_{n\ (t_k - t_{k-1})}((x_i(t_k))_{i=1}^{n} \mid (x_i(t_{k-1}))_{i=1}^{n}) \prod_{i=1}^{n} dx_i(t_k) \right) \left(\prod_{i=1}^{n} dx_i(t_0) \right) \tag{9}$$

with $q_{Mk}(y_{kj} \mid x_i) \overset{\text{def}}{=} p_{Mk}(y_{kj} \mid x_i) p_{Dk}(x_i) / \gamma_{Fk}(y_{kj})$, assuming[12] $t_0 < t_1 < \cdots < t_K$. $(\bar{a}_k)_{k=1}^{K}$ in (9) is any assignment function for which eqn. (1) holds with $(a_k)_{k=1}^{K} = (\bar{a}_k)_{k=1}^{K}$ for a given hypothesis λ and *n*.

One way for calculating the likelihood $Q_K(\lambda, n)$ is a recursive calculation that uses the sequence, $\lambda_{|1}, \lambda_{|2}, ..., \lambda_{|K-1}$, of *predecessors* or *restrictions* of hypothesis λ on cumulative frames, $y_1, (y_k)_{k=1}^{2}, ..., (y_k)_{k=1}^{K-1}$, respectively, as being defined as[13]

---

[10] By $\#(X)$ we mean the cardinality of (the number of elements in) any set *X*.

[11] $\mathcal{A}(n,m)$ is the set of all the one-to-one functions *a* defined on a subset $\mathrm{Dom}(a)$ of set $\{1,...,n\}$ taking values in set $\{1,...,m\}$.

[12] Any inequality "<" in $t_0 < t_1 < \cdots < t_K$ can be replaced by "$\leq$" by adjusting eqn. (10) in an obvious way.

[13] "\" is the set-theoretic subtraction operator, i.e., defined as $A \setminus B = \{a \in A \mid a \notin B\}$.

$\lambda_{|k} = \left\{ \tau \cap \left(\bigcup_{\kappa=1}^{k} \{\kappa\} \times \{1,\dots,m_\kappa\} \right) \middle| \tau \in \lambda \right\} \setminus \{\varnothing\}$, and the recursion defined by

$$Q_k(\lambda_{|k}, n) = M_k(\lambda_{|k}, n) Q_{k-1}(\lambda_{|k-1}, n) \tag{10}$$

$$\begin{aligned} &\bar{f}_{n\,t_k}((x_i(t_k))_{i=1}^n \mid \lambda_{|k-1}) = \\ &\int_{E^n} f_{n\,(t_k - t_{k-1})}((x_i(t_k))_{i=1}^n \mid (x_i(t_{k-1}))_{i=1}^n) \\ &\hat{f}_{n\,t_{k-1}}((x_i(t_{k-1}))_{i=1}^n \mid \lambda_{|k-1}) \prod_{i=1}^n dx_i(t_{k-1}) \end{aligned} \tag{11}$$

and

$$\begin{aligned} &\hat{f}_{n\,t_{k-1}}((x_i(t_{k-1}))_{i=1}^n \mid \lambda_{|k-1}) = \\ &M_{k-1}(\lambda_{|k-1}, n)^{-1} \left(\prod_{i \in \mathrm{Dom}(\bar{a}_{k-1})} q_{Mk-1}(y_{k-1\,\bar{a}_{k-1}(i)} \mid x_i(t_{k-1})) \right) \\ &\left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_{k-1})}}^n (1 - p_{Dk-1}(x_i(t_{k-1}))) \right) \bar{f}_{nt_{k-1}}((x_i(t_{k-1}))_{i=1}^n \mid \lambda_{|k-2}) \end{aligned} \tag{12}$$

with

$$\begin{aligned} M_k(\lambda_{|k}, n) = &\int_{E^n} \left(\prod_{i \in \mathrm{Dom}(\bar{a}_k)} q_{Mk}(y_{k\bar{a}_k(i)} \mid x_i(t_k)) \right) \\ &\left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_k)}}^n (1 - p_{Dk}(x_i(t_k))) \right) \bar{f}_{nt_k}((x_i(t_k))_{i=1}^n \mid \lambda_{|k-1}) \\ &\prod_{i=1}^n dx_i(t_k) \end{aligned} \tag{13}$$

using appropriate initial conditions.

Eqns. (10-13) define a process that is a direct extension of the independent-target recursive track likelihood calculation typically used by track-oriented multiple-hypothesis tracking algorithms.

On the other hand, the likelihood function $Q_K(\lambda, n)$ defined by (9) can be rewritten as

$$\begin{aligned} Q_K(\lambda, n) = &\int_{E^{nK}} \prod_{k=1}^K \left(\prod_{i \in \mathrm{Dom}(\bar{a}_k)} q_{Mk}(y_{k\bar{a}_k(i)} \mid x_i(t_k)) \right) \\ &\left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_k)}}^n (1 - p_{Dk}(x_i(t_k))) \right) P(((x_i(t_k))_{k=1}^K)_{i=1}^n \mid n) \prod_{i=1}^n dx_i(t_k) \end{aligned} \tag{14}$$

which provides us with an alternative way of calculating the likelihood $Q_K(\lambda, n)$.

For this alternative method, we assume that there is an effective way to calculate the a posteriori target state probability density

$$\begin{aligned} &P(((x_i(t_k))_{k=1}^K)_{i=1}^n \mid \lambda, n, (y_k)_{k=1}^K) = Q_K(\lambda, n)^{-1} \\ &\prod_{k=1}^K \left(\prod_{i \in \mathrm{Dom}(\bar{a}_k)} q_{Mk}(y_{k\bar{a}_k(i)} \mid x_i(t_k)) \right) \\ &\left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_k)}}^n (1 - p_{Dk}(x_i(t_k))) \right) P(((x_i(t_k))_{k=1}^K)_{i=1}^n \mid n) \end{aligned} \tag{15}$$

Then we also assume that the conditional probability density $P(((x_i(t_k))_{k=1}^K)_{i=1}^n \mid \lambda, n, (y_k)_{k=1}^K)$ can be well approximated by a Gaussian distribution. Because eqn. (15) holds for any value of $((x_i(t_k))_{k=1}^K)_{i=1}^n$, by choosing $((x_i(t_k))_{k=1}^K)_{i=1}^n$ to be equal to the mean vector $((\hat{x}_i(t_k))_{k=1}^K)_{i=1}^n$ of the Gaussian distribution, and assuming that the a priori probability distribution $P(((x_i(t_k))_{k=1}^K)_{i=1}^n \mid n)$ can be considered as a constant $\bar{p}_0$ around that mean vector, we have

$$\begin{aligned} Q_K(\lambda, n) = &\bar{p}_0 \det(2\pi \hat{V}_{nK}) \\ &\prod_{k=1}^K \left(\prod_{i \in \mathrm{Dom}(\bar{a}_k)} q_{Mk}(y_{k\bar{a}_k(i)} \mid \hat{x}_i(t_k)) \right) \\ &\left(\prod_{\substack{i=1 \\ i \notin \mathrm{Dom}(\bar{a}_k)}}^n (1 - p_{Dk}(\hat{x}_i(t_k))) \right) \end{aligned} \tag{16}$$

This technique for calculating the likelihood was originally introduced as an effective way for calculating normalizing constants of various Bayesian formulas arisen from Gaussian distributed estimation problems in [16], and was recently re-discovered as a batch-oriented track likelihood calculation in [17]. In eqn. (16), matrix $\hat{V}_{nK}$ is the $nKd \times nKd$ joint target state estimation error covariance matrix of the estimate $((\hat{x}_i(t_k))_{k=1}^K)_{i=1}^n$, where d is the dimension of the target state space E. Typically the estimate $((\hat{x}_i(t_k))_{k=1}^K)_{i=1}^n$ can be obtained from an iterative method to maximize the a posteriori joint target state probability density, and we assume it can be well approximated as a conditional mean vector.

Let the set of all the non-empty tracks be denoted by $\breve{\mathcal{T}}_K \overset{\text{def}}{=} \mathcal{T}_K \setminus \{\varnothing\}$ and an enumeration of $\breve{\mathcal{T}}_K$ be $\tau_1, \tau_2, \dots, \tau_{\mathrm{n}}$. Then the collection of all the subsets of $\breve{\mathcal{T}}_K$, i.e., the power set $2^{\breve{\mathcal{T}}_K}$, is one-to-one to the space $\{0,1\}^{\mathrm{n}}$ of n-dimensional binary vectors, through a function $\varphi : 2^{\breve{\mathcal{T}}_K} \to \{0,1\}^{\mathrm{n}}$, defined by $\tau_i \in \lambda$ if and only if $\xi_i = 1$ whenever $\varphi(\lambda) = (\xi_i)_{i=1}^{\mathrm{n}}$. Let an enumeration of all the measurements in $\bigcup_{k=1}^K \{k\} \times \{1,\dots,m_k\}$ be $(k_1, j_1), (k_2, j_2), \dots, (k_{\mathrm{m}}, j_{\mathrm{m}})$, and define a binary matrix

$A = ((A_{ii'})_{i=1}^{\mathrm{m}})_{i'=1}^{\mathrm{n}} \in \{0,1\}^{\mathrm{m\times n}}$ by $A_{ii'} = 1$ if and only if $(k_i, j_i) \in \tau_{i'}$. Then we have

$$\Lambda_K = \{\varphi^{-1}(\xi) \mid \xi \in \{0,1\}^{\mathrm{n}} \text{ and } A\xi \le b\} = \varphi^{-1}(D_K) \quad (17)$$

with $D_K = \{\xi \in \{0,1\}^{\mathrm{n}} \mid A\xi \le b\}$, where b is an m-dimensional vector whose elements are all 1's, and the comparator "$\le$" in (17) is the component-by-component vector-comparator.

Then the a posteriori probability $P(\lambda \mid (y_k)_{k=1}^K)$ can be translated from a functional on the space Λ_K of data association hypotheses to a functional $\phi : D_K \to [0,\infty)$ on the subset D_K of $\{0,1\}^{\mathrm{n}}$, through $\phi(\xi) = P(\varphi^{-1}(\xi) \mid (y_k)_{k=1}^K)$, $\xi \in D_k$, or, $\phi(\varphi(\lambda)) = P(\lambda \mid (y_k)_{k=1}^K)$, $\lambda \in \Lambda_K$. It follows from (8) that

$$\begin{aligned}\phi(\varphi(\lambda)) &= P(\lambda \mid (y_k)_{k=1}^K) \\ &= C \sum_{n=\#(\lambda)}^{\infty} \frac{n!\,\overline{p}(n)}{(n-\#(\lambda))!} Q_K(\lambda, n)\end{aligned} \quad (18)$$

with a normalizing constant *C*.

Suppose that, given the number *n* of targets, the targets are *independent, identically distributed* (i.i.d.), in the sense that

$$\overline{f}_{nt_0}((x_i(t_0))_{i=1}^n) = \prod_{i=1}^n \overline{f}_{1t_0}(x_i(t_0)) \quad (19)$$

and

$$\begin{aligned}& f_{n\Delta t}((x_i(t+\Delta t))_{i=1}^n \mid (x_i(t))_{i=1}^n) \\ &\quad = \prod_{i=1}^n f_{1\Delta t}(x_i(t+\Delta t) \mid x_i(t))\end{aligned} \quad (20)$$

Then, (9) can be decomposed as

$$Q_K(\lambda, n) = \overline{\nu}^{-n} \left(\ell_K(\emptyset)\right)^{n-\#(\lambda)} \prod_{\tau\in\lambda} \ell_K(\tau) \quad (21)$$

if $n \ge \#(\lambda)$, 0 otherwise, where $\ell_K(\tau)$ is the track likelihood of track τ defined by

$$\begin{aligned}\ell_K(\tau) &= \overline{\nu} \int_E \overline{f}_{1t_0}(x'_0) \\ &\left(\prod_{k=1}^K \int_E q_{Mk}(y_{k\tau(k)} \mid x'_k) f_{1(t_k - t_{k-1})}(x'_k \mid x'_{k-1}) dx'_k \right) dx'_0\end{aligned} \quad (22)$$

with $q_{Mk}(\theta \mid x) \overset{\text{def}}{=} 1 - p_{Dk}(x)$ for each $x \in E$.

By substituting (22) into (21) and then into (18), we have the hypothesis evaluation formula for the i.i.d. case, as

$$\begin{aligned}\phi(\varphi(\lambda)) &= P(\lambda \mid (y_k)_{k=1}^K) \\ &= C \left(\sum_{n=\#(\lambda)}^{\infty} \frac{n!}{\overline{\nu}^n} \frac{\hat{\nu}_K^{n-\#(\lambda)}}{(n-\#(\lambda))!} \overline{p}(n) \right) \left(\prod_{\tau\in\lambda} \ell_K(\tau) \right)\end{aligned} \quad (23)$$

with $\hat{\nu}_K \overset{\text{def}}{=} \ell_K(\emptyset)$, which was originally shown in [6].

Furthermore, if the a priori probability distribution is Poisson, i.e., if $\overline{p}(n) = e^{-\overline{\nu}}\overline{\nu}^n / n!$, $n = 0,1,2,\ldots$, then eqn. (23) becomes the familiar data association hypothesis evaluation formula

$$\phi(\varphi(\lambda)) = P(\lambda \mid (y_k)_{k=1}^K) = C' \prod_{\tau\in\lambda} \ell_K(\tau) \quad (24)$$

with $C' = C \exp\left(-(\overline{\nu} - \hat{\nu}_K)\right)$, which is commonly used for almost all the MHT algorithms including [3-5]. As shown in the above, this simple multiplicative evaluation function (24) is a direct consequence of the i.i.d. and Poisson assumptions. For the non-Poisson but i.i.d. cases, the formal multiplicity with respective to the track likelihoods is preserved, as seen in (23), but it contains a coefficient $\sum_{n=\#(\lambda)}^{\infty} \frac{n!}{\overline{\nu}^n} \frac{\hat{\nu}_K^{n-\#(\lambda)}}{(n-\#(\lambda))!} \overline{p}(n)$ that is *not* multiplicative with respect to the set of tracks.

Thus, this section has derived a class of zero-one integer programming problems with generally non-multiplicative or nonlinear (using the logarithm) objective functions, for multiple target tracking problems when the targets are not necessarily independent or the number of targets is not necessarily Poisson a priori.

4 K-Best Hypothesis Selection Using Cross-Entropy Method and Generalized Murty's Algorithm

Given a subset[14] D of $\{0,1\}^{\mathrm{n}}$, defined by a binary matrix $A \in \{0,1\}^{\mathrm{m\times n}}$ and a binary vector $b \in \{0,1\}^{\mathrm{m}}$ whose components are all ones, as $D = \{\xi \in \{0,1\}^{\mathrm{n}} \mid A\xi \le b\}$, consider the problem for maximizing the functional $\phi : D \to [0,\infty)$ which is not necessarily multiplicative or linear, as defined in the previous section. First, let us consider the problem of selecting the single best solution $\hat{\xi} \in D$ (such that $\phi(\hat{\xi}) = \max\{\phi(\xi) \mid \xi \in D\}$).

---

[14] In this section, we dropped the index *K* that represents the cumulative frames $(y_k)_{k=1}^K$. We will use indices *k* in ways different from those in previous sections.

As mentioned before, when the objective function is not multiplicative or linear, we cannot use the usual integer linear programming technique [3] or the relaxation method [7,8]. Even some stochastic methods such as the MCMC algorithms described in [11,12] require a multiplicative or linear objective function. This is why we propose the use of Reuven Rubinstein's cross-entropy method [9] to solve data association problems when the targets are not necessarily independent or even when they are independent but the number of targets is not necessarily Poisson. Recently the cross-entropy method was applied to solve the multiple target tracking data association problems in [10], which focuses on i.i.d.-Poisson problems.

A simple cross entropy algorithm for solving our optimization problem can be described as a recursive algorithm to update the pair $(c_k, \pi_k) \in [0,\infty) \times [0,1]^{\mathrm{n}}$ of *objective level threshold* c_k and *probability mass function* π_k, for each step $k = 0,1,2,\ldots$. A version of cross-entropy method recursion can de defined as follows:

1) Given $(c_k, \pi_k) = (c_k, (\pi_{ki})_{i=1}^{\mathrm{n}}) \in [0,\infty) \times [0,1]^{\mathrm{n}}$, generate N_S i.i.d. samples, $(X_{ks})_{s=1}^{N_S} = ((X_{ksi})_{i=1}^{\mathrm{n}})_{s=1}^{N_S}$, with probability

$$\mathrm{Prob}\{X_{ks} = \xi = (\xi_i)_{i=1}^{\mathrm{n}} \;;\; (\pi_{ki})_{i=1}^{\mathrm{n}}\} = \frac{\mathbb{I}\{A\xi \le b\} \prod_{i=1}^{\mathrm{n}} \pi_{ki}^{\xi_i} (1-\pi_{ki})^{(1-\xi_i)}}{\sum\limits_{\xi' = (\xi')_{i=1}^{\mathrm{n}} \in \{0,1\}^{\mathrm{n}}} \mathbb{I}\{A\xi' \le b\} \prod_{i=1}^{\mathrm{n}} \pi_{ki}^{\xi'_i} (1-\pi_{ki})^{(1-\xi'_i)}} \tag{25}$$

where $\mathbb{I}\{A\xi \le b\} = 1$ if $A\xi \le b$, $= 0$ otherwise, for every $\xi \in \{0,1\}^{\mathrm{n}}$.

2) For a given $\rho \in (0,1)$, let c_{k+1} be the $(1-\rho)$-quantile level, i.e.,

$$c_{k+1} = \max\left\{ c \in [0,\infty) \,\middle|\, (1/N_S)\sum_{s=1}^{N_S} \mathbb{I}\{\phi(X_{ks}) \ge c\} \ge \rho \right\} \tag{26}$$

where $\mathbb{I}\{\phi(\xi) \ge c\} = 1$ if $\phi(\xi) \ge c$, $= 0$ otherwise.

3) Update $\pi_k = (\pi_{ki})_{i=1}^{\mathrm{n}}$ to obtain $\pi_{k+1} = (\pi_{k+1\,i})_{i=1}^{\mathrm{n}}$, by

$$\begin{aligned} (\pi_{k+1\,i})_{i=1}^{\mathrm{n}} &= \left(\frac{\sum_{s=1}^{N_S} \mathbb{I}\{\phi(X_{ks}) \ge c_{k+1}\} X_{ksi}}{\sum_{s=1}^{N_S} \mathbb{I}\{\phi(X_{ks}) \ge c_{k+1}\}} \right)_{i=1}^{\mathrm{n}} \\ &\in \underset{\pi \in [0,1]^{\mathrm{n}}}{\mathrm{ArgMax}} \frac{1}{N_S} \sum_{s=1}^{N_S} \mathbb{I}\{\phi(X_{ks}) \ge c_{k+1}\} \ln(\mathrm{Prob}\{X_{ks} \;;\; \pi_k\}) \end{aligned} \tag{27}$$

The process can be initialized by $c_0 = 0$ and $\pi_0 = (\pi_{0i})_{i=1}^{\mathrm{n}}$ with $\pi_{0i} \equiv 1/2$, and $\rho = 0.01$ is recommended in [9]. Another sampling scheme, an alternative to (25), is proposed in [10]. We expect the probability distribution $\pi_k = (\pi_{ki})_{i=1}^{\mathrm{n}}$ to converge to the optimal solution.

The basic principle behind the cross entropy method is, as the iteration steps are accumulated, to generate better and better samples, i.e., what are called *elite samples* in [10]. Therefore, if we accumulate all the samples, $\{\{X_{ks}\}_{s=1}^{N_S}\}_{k=1}^{K}$, up to the last step *K*, it may be possible to accumulate all the "good" samples that, when we take the best K samples, may constitute an approximation of the best K solutions.

When the cross entropy optimization process does not accumulate enough number of top-ranked solutions, we may use the *generalized Murty's algorithm* [13] to explicitly assemble the K-best solutions. The generalized Murty's algorithm is a generalization of Murty's algorithm [19] that is a K-best ranking algorithm for bipartite assignment problems and that has been applied to the measurement-to-track assignment problems in [20]. The generalized Murty's algorithm [13], as well as the original Murty's algorithm [19], is what we may call a *meta algorithm* that can use any applicable optimization algorithm as a sub-algorithm.

Moreover, the Murty's or the generalized Murty's algorithm itself does not require the objective function to be multiplicative or linear. Therefore, we can use the generalized Murty's algorithm that utilizes the cross-entropy algorithm as a sub-algorithm. In order to apply the generalized Murty's algorithm, however, all the constraints must be made to be equality constraints, i.e., we must have $A\xi = b$, instead of $A\xi \le b$, as the joint constraint, by augmenting the solution space $\{0,1\}^{\mathrm{n}}$ for additional slack variables, explicitly for each false alarm hypothesis. We must modify eqns. (25-27) accordingly, but necessary modifications are rather obvious.

5 Conclusion

We have developed a data association hypothesis evaluation formula for multiple target tracking problems where the targets are not necessarily independent from each other a priori. We showed that the most noticeable departure from the i.i.d.-Poisson cases is the loss of the multiplicity of the likelihood or the additivity of the log-likelihood function. This loss prevents us from using commonly used optimal data association selection algorithms such as zero-one integer linear programming, relaxation-method-based multidimensional assignment algorithms, or even the MCMC method.

As a possible solution, we proposed to use the cross-entropy method (which was discovered only recently as a method for solving data association problems in [10]), since it does not require the objective function to be multiplicative or linear. We also proposed the use of the generalized Murty's algorithm for the K-best optimization extension of the cross entropy method, mainly because the generalized Murty's algorithm does not require multiplicative or linear objective functions either.

This paper presents only the concept of the proposed method. Although the solution sampling for the cross-entropy method is straightforward and can be done very fast, the numerical evaluation for each data association hypothesis may be extremely costly depending on the complexity of the problem. In fact, in some situations, even with linear objective functions, numerical calculation of track likelihood of each track requires significant resources. In order for the technical approach presented in this paper to be made practical, therefore, we need to develop effective numerical calculation of the hypothesis likelihoods, which is currently in development.

Acknowledgment

The authors wish to express their sincere gratitude to the late Dr. Jean-Pierre Le Cadre, for suggesting the cross entropy method for solving data association problems, in the discussions that followed the presentation of [18] at FUSION 2008, in Cologne, Germany. We deeply regret his untimely passing in July, 2009.

References

[1] Yaakov Bar-Shalom, and Xiao-Rong Li, *Multitarget-multisensor tracking: principles and techniques*, YBS Publishing, Storrs, CT, 1995.

[2] S. S. Blackman, and R. Popoli, *Design and analysis of modern tracking systems*, Artech House, Norwood, MA, 1999.

[3] Charles L. Morefield, *Application of 0-1 integer programming to multi-target tracking problems*, IEEE Transactions on Automatic Control, Vol. AC-23, pp. 302-312, June 1977.

[4] Donald B. Reid, *An algorithm for tracking multiple targets*, IEEE Transactions on Automatic Control, Vol. AC-24, pp. 843-854, Dec., 1979.

[5] Shozo Mori, Chee-Yee Chong, Edison Tse, and Richard P. Wishner, *Tracking and classifying multiple targets without a priori identification*, IEEE Transaction on Automatic Controls, Vol. AC-33, No. 2, pp. 401-409, 1986.

[6] Shozo Mori, and Chee-Yee Chong, *Evaluation of data association hypotheses: non-Poisson i.i.d. cases*, in Proc. 7th International Conference on Information Fusion, pp. 1133 – 1140, Stockholm, Sweden, July 2004.

[7] Krishna R. Pattipati, Somnath Deb, and Yaakov Bar-Shalom, *Passive multisensor data association using a new relaxation algorithm,* in Multitarget-Multisensor Tracking: Advanced Applications, ed. by Yaakov Bar-Shalom, Artech House, 1990.

[8] Aubrey B. Poore, and Nenad Rijavec, *A new class of methods for solving data association problems arising from multitarget tracking*, in Proc. 1991 American Automatic Control Conference, Vol. 3, pp. 2302-2304, Boston, MA, 1991.

[9] Reuven Y. Rubinstein, and Dirk P. Kroese, *The cross-entropy method – a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning*, Springer Science, 2004.

[10] Daniel Sigalov, and Nahum Shimkin, *Cross-entropy based data association for multi target tracking*, in Proc. 3rd International Conference on Performance Evaluation Methodologies and Tools, Athens, Greece, October 2008.

[11] Songhwai Oh, Stuart Russell, and Shankar Sastry, *Markov chain Monte Carlo data association for general multipel-target tracking problems*, in Proc. 43rd IEEE Conference on Decision and Control, pp. 735-742, Vol. 1, Atlantis, Paradise Island, the Bahamas, December 2004.

[12] Shozo Mori, and Chee-Yee Chong, *Evaluation of a posteriori probabilities of multi-frame data association hypotheses*, in Proc. SPIE Symposium on Signal and Data Processing of Small Targets, San Diego, August 2007.

[13] Evan Fortunato, William Kreamer, Shozo Mori, Chee-Yee Chong, Gregory Castanon, *Generalized Murty's algorithm with application to multiple hypothesis tracking*, in Proc. 10th International Conference on Information Fusion, Quebec, Quebec, Canada, July 2007.

[14] Robert L. Popp, Krishna R. Pattipati, and Yaakov Bar-Shalom, *M-best s-d assignment algorithm with application to multiple target tracking*, IEEE Transactions on Aerospace and Electronic Systems, Vol. 37, No. 1, pp. 22 – 39, January 2001.

[15] Shozo Mori, and Chee-Yee Chong, *Point process formalism of multitarget tracking*, in Proc. 5th International Conference on Information Fusion, Annapolis, MD, July 2002.

[16] Chee-Yee Chong, Shozo Mori, and Kuo-Chu Chang, *Distributed multitarget multisensor tracking*, Chapter 8 in *Multitarget-multisensorm tracking: advanced applications*, edited by Y. Bar-Shalom, Artech House, MA, 1990.

[17] Charles Gershfield, *Iterated linearized batch processing method for new object tracks*, BAE Systems, AIT Technical Note, BA-AIT-SSA-2009-1, August 2009.

[18] Shozo Mori, and Chee-Yee Chong, *Markov chain Monte Carlo method for evaluating multi-frame data association hypotheses*, in Proc. 11th International Conference on Information Fusion, Cologne, Germany, July 2008.

[19] Katta G. Murty, *An algorithm for ranking all the assignments in order of increasing cost*, Operations Research, Vol. 16, No. 3, pp. 682 – 687, May-June, 1968.

[20] Ingemar J. Cox, and Matt L. Miller, *On finding ranked assignments with application to Multitarget tracking and motion correspondence*, IEEE Transactions on Aerospace and Electronic Systems, Vol. 32, No. 1, pp. 486-489, 1995.