A Polynomial-Time Approximation Algorithm for Joint Probabilistic Data Association

Songhwai Oh and Shankar Sastry

***Abstract*— Joint probabilistic data association (JPDA) is a powerful tool for solving data association problems. However, the exact computation of association probabilities $\{\beta_{jk}\}$ in JPDA is NP-hard, where β_{jk} is the probability that j-th observation is from k-th track. Hence, we cannot expect to compute association probabilities in JPDA exactly in polynomial time unless $P = NP$. In this paper, we present a simple Markov chain Monte Carlo data association (MCMCDA) algorithm that finds an approximate solution to JPDA in polynomial time. For $\epsilon > 0$ and $0 < \eta < .5$, we prove that the algorithm finds good estimates of β_{jk} with probability at least $1 - \eta$ in time complexity $O(\epsilon^{-2} \log \eta^{-1} N(N \log N + \log(\epsilon^{-1})))$, where N is the number of observations.**

I. Introduction

The data association problem arises in many applications such as computer vision, surveillance, clustering, and mobile robots. In computer vision, the data association problem is known as the correspondence problem in which the objective is to determine which observation belongs to which feature [1], [2]. In target tracking, it is the problem of determining which observation is generated by which target or clutter.

Joint probabilistic data association (JPDA) is developed to solve the data association problem arises in multiple-target tracking [3]. JPDA is a suboptimal single-scan approximation to the optimal Bayesian filter, in which the associations between the "known" tracks and the *latest* observations are made sequentially. At each time step, instead of finding a single best association between latest observations and known tracks, JPDA enumerates all possible associations between observations and tracks and computes association probabilities $\{\beta_{jk}\}$, where β_{jk} is the probability that j-th observation is from k-th track. Given an association, the state of a target is estimated by a filtering algorithm and this conditional expectation of state is weighted by the association probability. Then the state of a target is estimated by summing over the weighted conditional expectations. It has proved very effective in a cluttered environment compared with the nearest neighbor approach which finds a single best association [3].

However, the exact calculation of association probabilities $\{\beta_{jk}\}$ in JPDA is NP-hard [4] since the related problem of finding the permanent of a 0-1 matrix is #P-complete [5]. To overcome the complexity of the problem, many approximation algorithms have been proposed. In [6], the "cheap" JPDA algorithm is developed and the association probabilities are estimated from a formula based on heuristics. In [7], a single-stage data association problem is considered and a leave-one-out heuristic is developed to avoid the enumeration of all possible associations. Sampling methods have been applied before. In [8], the Gibbs sampling method is applied to track a single target using measurements from a finite number of linear models, where the measurement to model association is unknown. In [9], a combination of Markov chain Monte Carlo (MCMC) and expectation-maximization (EM) is used to simultaneously track multiple vehicles using measurements from spatially separated sensors and learn the intrinsic parameters of the sensors. A combination of MCMC and EM is also used in [2] to solve the correspondence problem in computer vision. In [10], MCMC is applied to compute the association probabilities in JPDA and it is shown that MCMC outperforms Fitzgerald's cheap JPDA. Unfortunately, in all cases, the performance of an approximation algorithm for JPDA is measured in experiment only. The main contribution of this paper is the formal analysis of a simple sampling algorithm for JPDA, called the Markov chain Monte Carlo data association (MCMCDA) algorithm. We show that MCMCDA finds an approximate solution to JPDA in polynomial time. For $\epsilon > 0$ and $0 < \eta < .5$, we prove that the algorithm finds "good" estimates of β_{jk} with probability at least $1-\eta$ in time complexity $O(\epsilon^{-2} \log \eta^{-1} N(N \log N + \log(\epsilon^{-1})))$, where N is the number of observations (the precise definition of "good" estimates is given in Section V).

In [11], a general-purpose MCMCDA algorithm is developed to track an unknown number of targets. It has been shown that MCMCDA is computationally efficient compared to the multiple hypothesis tracker (MHT) [12] and outperforms MHT under extreme conditions, such as a large number of targets in a dense environment, low detection probabilities, and high false alarm rates [11]. The MCMCDA algorithm has been extended to sensor networks in a hierarchical manner to be scalable and it has been shown that MCMCDA is robust against sensor localization error, transmission failures and communication delays, *i.e.*, out-of-sequence measurements [13]. The MCMCDA algorithm presented in this paper can be considered as a special case of the algorithm presented in [11] for tracking a known number of targets.

The remainder of this paper is structured as follows. We summarize JPDA in Section II and describe the MCMC method in Section III. The MCMCDA algorithm is presented in Section IV and analysis about the algorithm is shown in Section V. We also present an experiment confirming our results in Section VI.

II. Joint Probabilistic Data Association

Joint probabilistic data association (JPDA) has been traditionally used with the Kalman filter, assuming linear dynamic and measurement models and a Gaussian noise model [3], and the Kalman filter is used in this paper for demonstration purpose. However, JPDA has been applied with a nonlinear filtering algorithm such as a particle filter [14]. We note that the proposed algorithm is applicable for both linear and nonlinear filters and our results can be easily generalized to the nonlinear case.

Let K be the number of targets moving around the surveillance region $\mathcal{R}$. The state dynamics of target k is modeled as

$$x_{t+1}^k = A_t^k x_t^k + G_t^k w_t^k, \tag{1}$$

for $t = 1, 2, \ldots$, where $x_t^k \in \mathbb{R}^{n_x}$ is the state of target k at time t, A_t^k and G_t^k are matrices with appropriate sizes, and w_t^k is a white Gaussian process with zero mean and covariance Q_t^k. The noisy observation of the state of a target is measured with a detection probability p_{d} which is less than unity. There are also false alarms and the number of false alarms has a Poisson distribution with a parameter $\lambda_{\mathrm{f}} V$ where V is the volume of $\mathcal{R}$ and λ_{f} is the false alarm rate per unit time, per unit volume. Let n_t be the number of observations at time t, including both noisy observations and false alarms. Let $y_t^j \in \mathbb{R}^{n_y}$ be the j-th observation at time t for $j = 1, \ldots, n_t$. Each target generates a unique observation at each sampling time if it is detected. The measurement model is

$$y_t^j = \begin{cases} C_t^j x_t^k + w_t'^j & \text{if } y_t^j \text{ is from } x_t^k \\ u_t^j & \text{otherwise,} \end{cases} \tag{2}$$

where $w_t'^j$ is a white Gaussian process with zero mean and covariance R_t^j, C_t^j is a matrix with an appropriate size, and $u_t^j \sim \mathrm{Unif}(\mathcal{R})$ are random processes for false alarms. Notice that, with probability $1 - p_{\mathrm{d}}$, the target is not detected and we call this a missing observation. Let $Y_t = \{y_t^j : 1 \le j \le n_t\}$ and $Y_{1:t} = \{Y_1, \ldots, Y_t\}$.

The authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720, {sho,sastry}@eecs.berkeley.edu.

This material is based upon work supported by the National Science Foundation under Grant No. EIA-0122599.

Now we summarize the joint probability data association (JPDA) filter [3]. Suppose that we have the following estimates from the previous filtering step $t-1$:

$$\begin{aligned}\hat{x}^k_{t-1|t-1} &:= \mathbb{E}\left[x^k_{t-1}|Y_{1:t-1}\right]\\ P^k_{t-1|t-1} &:= \mathbb{E}\left[s^k_{t-1|t-1}{s^k_{t-1|t-1}}^T|Y_{1:t-1}\right],\end{aligned}$$

where $s^k_{t-1|t-1} = x^k_{t-1} - \hat{x}^k_{t-1|t-1}$. For notational convenience, we assume that $A = A^k_t$, $G = G^k_t$ and $Q = Q^k_t$, for all k and t, and $C = C^j_t$ and $R = R^j_t$, for all j and t.

Step 1 (Prediction): For each k, compute

$$\begin{aligned}\hat{x}^k_{t|t-1} &:= \mathbb{E}\left[x^k_t|Y_{1:t-1}\right] = A\hat{x}^k_{t-1|t-1}\\ P^k_{t|t-1} &:= \mathbb{E}\left[s^k_{t|t-1}{s^k_{t|t-1}}^T|Y_{1:t-1}\right]\\ &= AP^k_{t-1|t-1}A^T + GQG^T,\end{aligned}$$

where $s^k_{t|t-1} = x^k_t - \hat{x}^k_{t|t-1}$.

Step 2 (Measurement Validation): The predicted observation for target k is $\hat{y}^k_{t|t-1} = C\hat{x}^k_{t|t-1}$. For each observation j, define the innovation $v^k_t(j) = y^j_t - \hat{y}^k_{t|t-1}$ and its covariance

$$\begin{aligned}B^k_t &= \mathbb{E}\left[v^k_t(j){v^k_t(j)}^T|Y_{1:t-1}\right]\\ &= CP^k_{t|t-1}C^T + R^T.\end{aligned}$$

Let $\Xi = [\xi_{jk}] \in \{0,1\}^{n_t\times K}$ be a validation matrix and $\xi_{jk} = 1$, *i.e.*, y^j_t is validated for target k, if and only if

$$v^k_t(j)^T(B^k_t)^{-1}v^k_t(j) < \delta, \tag{3}$$

where δ is an appropriate threshold. Without loss of generality, we assume that, for all j, $\sum_{k=1}^K \xi_{jk} \geq 1$, *i.e.*, all observations are validated with at least one target. If not, we can always resize the matrix Ξ and reduce n_t to make sure that each row vector has at least one non-zero element.

Step 3 (State Estimation): Let Ω be a set of all feasible joint association events at time t (for notational convenience, the subscript t is dropped). For each $\omega \in \Omega$, $\omega = \{(j,k)\}$, where (j,k) denotes an event that observation j is associated with target k. We represent a joint association event ω by a matrix $\hat{\Xi}(\omega) = [\hat{\xi}_{jk}(\omega)]$, where $\hat{\xi}_{jk}(\omega) = 1$ if the event (i,k) is true, otherwise, $\hat{\xi}_{jk}(\omega) = 0$. A joint association event is feasible when (i) it agrees with the validation matrix, *i.e.*, $\hat{\xi}_{jk}(\omega) \leq \xi_{jk}(\omega)$ for all j and k; (ii) an observation has at most one source, *i.e.*, $\sum_{k=1}^K \hat{\xi}_{jk}(\omega) \leq 1$ for all j; and (iii) a target has at most one observation, *i.e.*, $\sum_{j=1}^{n_t} \hat{\xi}_{jk}(\omega) \leq 1$ for all k. Notice that we use notations different from [3]. In particular, we did not introduce an additional column for "no target" so observation j is a false alarm if $\sum_{k=1}^K \hat{\xi}_{jk}(\omega) = 0$.

The state of a target can be estimated as

$$\begin{aligned}\mathbb{E}(x^k_t|Y_{1:t}) &= \sum_\omega \mathbb{E}(x^k_t|\omega, Y_{1:t})P(\omega|Y_{1:t}) \qquad (4)\\ &= \sum_{j=0}^{n_t} \mathbb{E}(x^k_t|\omega_{jk}, Y_{1:t})P(\omega_{jk}|Y_{1:t}),\end{aligned}$$

where ω_{jk} denotes the event $\{\omega \ni (j,k)\}$ and ω_{0k} denotes the event that no observation is associated with target k. Let $\beta_{jk} = P(\omega_{jk}|Y_{1:t})$. $\mathbb{E}(x^k_t|\omega_{jk}, Y_{1:t})$ can be computed easily by considering it as a single target estimation problem with a single observation. Hence, the computation of $\mathbb{E}(x^k_t|Y_{1:t})$ reduces to the computation of β_{jk}, where

$$\beta_{jk} = P(\omega_{jk}|Y_{1:t}) = \sum_{\omega:(j,k)\in\omega} P(\omega|Y_{1:t}). \tag{5}$$

The computation of β_{jk} requires a summation over the posteriors, hence the enumeration of all joint association events. JPDA is a method for computing expectations such as (4) using the association probabilities $\{\beta_{jk}\}$ in the presence of the identity uncertainty. As mentioned earlier, the exact calculation of $\{\beta_{jk}\}$ in JPDA is NP-hard [4] and it is the major drawback of JPDA.

Based on the parametric false alarm model, the posterior of ω can be computed as

$$\begin{aligned}P(\omega|Y_{1:t}) &= \frac{1}{Z}P(Y_t|\omega, Y_{1:t-1})P(\omega) \qquad (6)\\ &= \frac{1}{Z}\lambda_{\mathrm{f}}^{n_{\mathrm{f}}}p_{\mathrm{d}}^{n_{\mathrm{d}}}(1-p_{\mathrm{d}})^{K-n_{\mathrm{d}}}\prod_{j=1}^{n_t}\left[\mathcal{N}_{k_j}(y^j_t)\right]^{\tau_j},\end{aligned}$$

where Z is a normalizing constant; $n_{\mathrm{d}} = \#\{k : \sum_{j=1}^{n_t}\hat{\xi}_{jk}(\omega) = 1\}$ is the number of detections; $n_{\mathrm{f}} = n_t - n_{\mathrm{d}}$ is the number of false alarms; $\tau_j = \sum_{k=1}^K \hat{\xi}_{jk}(\omega)$; $k_j = k$ if and only if $\tau_j = 1$ and $\hat{\xi}_{jk}(\omega) = 1$ for given j; and $\mathcal{N}_{k_j}(y^j_t)$ is the Gaussian density function with mean $\hat{y}^{k_j}_{t|t-1}$ and covariance $B^{k_j}_t$.

Let $v^k_t = \sum_{j=1}^{n_t}\beta_{jk}v^k_t(j)$ be the combined innovation and $K^k_t = P^k_{t|t-1}C(B^k_t)^{-1}$ be the Kalman gain. Then the state of each target and its covariance are computed as follows (for more detail, see [3]).

$$\begin{aligned}\hat{x}^k_{t|t} &= \hat{x}^k_{t|t-1} + K^k_t v^k_t\\ P^k_{t|t} &= P^k_{t|t-1} - \left(\sum_{j=1}^{n_t}\beta_{jk}\right)K^k_t B^k_t {K^k_t}^T\\ &+ K^k_t\left(\sum_{j=1}^{n_t}\beta_{jk}v^k_t(j){v^k_t(j)}^T - v^k_t{v^k_t}^T\right){K^k_t}^T.\end{aligned}$$

III. Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) plays a significant role in many fields such as physics, statistics, economics, and engineering [15]. In some cases, MCMC is the only known general algorithm that finds a good approximate solution to a complex problem in polynomial time [16]. MCMC techniques have been applied to complex probability distribution integration problems, counting problems such as #P-complete problems, and combinatorial optimization problems [15], [16].

MCMC is a general method to generate samples from a distribution π by constructing a Markov chain $\mathcal{M}$ with states ω and stationary distribution $\pi(\omega)$. If we are at state $\omega \in \Omega$, we propose $\omega' \in \Omega$ following the proposal distribution $q(\omega,\omega')$. The move is accepted with an acceptance probability $A(\omega,\omega')$ where

$$A(\omega,\omega') = \min\left(1, \frac{\pi(\omega')q(\omega',\omega)}{\pi(\omega)q(\omega,\omega')}\right), \tag{7}$$

otherwise the sampler stays at ω, so that the detailed balance condition is satisfied, *i.e.*,

$$Q(\omega,\omega') = \pi(\omega)P(\omega,\omega') = \pi(\omega')P(\omega,\omega'), \tag{8}$$

for all $\omega,\omega' \in \Omega$, where $P(\omega,\omega') = q(\omega,\omega')A(\omega,\omega')$ is the transition probability from ω to ω' for $\omega' \neq \omega$. The described MCMC algorithm is known as the Metropolis-Hastings algorithm. If $\mathcal{M}$ is irreducible and aperiodic, then $\mathcal{M}$ converges to its stationary distribution by the ergodic theorem [17]. Hence, for a given bounded function f, the sample mean $\hat{f}$ of f over the sampled states converges to $\mathbb{E}_\pi f(\omega)$. Notice that (8) requires only the ability to compute the ratio $\pi(\omega')/\pi(\omega)$, avoiding the need to normalize π.

An ergodic chain $\mathcal{M}$ on state space Ω converges to its stationary distribution asymptotically. But a practical question is how fast $\mathcal{M}$ becomes close to stationarity. One way to measure the rate of convergence of $\mathcal{M}$ to stationarity is the "mixing time" of the Markov chain. Let P be the transition probabilities of $\mathcal{M}$ and let $P^t_x(\cdot)$ be the distribution of the state at time t given that $\mathcal{M}$ is started from the initial state $x \in \Omega$. If π is the stationary

distribution of $\mathcal{M}$, then the *total variation distance* at time t with initial state x is defined as

$$\Delta_x(t) = \|P_x^t - \pi\| = \max_{S \subset \Omega} |P_x^t(S) - \pi(S)| \tag{9}$$

The rate of convergence of $\mathcal{M}$ to stationarity can be measured by the *mixing time*:

$$\tau_x(\epsilon) = \min\{t : \Delta_x(s) \le \epsilon \text{ for all } s \ge t\}. \tag{10}$$

One approach to bound $\tau_x(\epsilon)$ of a Markov chain with a complex structure is the canonical path method [16]. In this paper, we consider a highly complex Markov chain, hence we use the canonical path method to bound $\tau_x(\epsilon)$ of the Markov chain simulated by the MCMCDA algorithm given in Section IV. For the remainder of this section, we describe the canonical path method.

For a finite, reversible and ergodic Markov chain $\mathcal{M}$ with state space Ω, consider an undirected graph $G = (V, E)$ where $V = \Omega$ and $E = \{(x, y) : Q(x, y) > 0\}$. Recall the definition of $Q(\cdot, \cdot)$ from (8). So an edge $(x, y) \in E$ indicates that the Markov chain $\mathcal{M}$ can make a transition from x to y or from y to x in a single step. For each ordered pair $(x, y) \in \Omega^2$, the canonical path γ_{xy} is a simple path from x to y in G. In terms of $\mathcal{M}$ the canonical path γ_{xy} is a sequence of legal transitions from x to y in $\mathcal{M}$. Let $\Gamma = \{\gamma_{xy} : x, y \in \Omega\}$ be the set of all canonical paths. Now the mixing time of the chain is related to the *maximum edge loading*:

$$\bar{\rho} = \bar{\rho}(\Gamma) = \max_e \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y)|\gamma_{xy}|. \tag{11}$$

If $\bar{\rho}$ is not so big, *i.e.*, no single edge is overloaded, then the Markov chain can move around fast and achieve the rapidly mixing property. The main result for the canonical path method is as follows [16], [18]:

Theorem 1: Let $\mathcal{M}$ be a finite, reversible, ergodic Markov chain with loop probabilities $P(x, x) \ge \frac{1}{2}$ for all states x. Let Γ be a set of canonical paths with maximum edge loading $\bar{\rho}$. Then the mixing time of $\mathcal{M}$ satisfies $\tau_x(\epsilon) \le \bar{\rho}(\log \pi(x)^{-1} + \log \epsilon^{-1})$, for any choice of initial state x.

IV. MCMC Data Association Algorithm

In this section, we describe the MCMC data association (MCMCDA) algorithm for approximating the association probabilities $\{\beta_{jk}\}$. When applied to a filtering problem, MCMCDA follows all the steps described in Section II except we use MCMCDA to estimate $\{\beta_{jk}\}$ in step 3. While the exact computation of $\{\beta_{jk}\}$ in JPDA is NP-hard, MCMCDA finds approximations to $\{\beta_{jk}\}$ in polynomial time with good fidelity as shown in Section V.

We fix the time t and the time index t is dropped from now. Let $N = n_t$ be the number of validated observations. We first reformulate our problem as a bipartite graph. Let $G = (U, V, E)$ be a bipartite graph, where $U = \{\hat{y}^k : 1 \le k \le K\}$ is a vertex set of predicted observations, $V = \{y^j : 1 \le j \le N\}$ is a vertex set of observations, and $E = \{(u, v) : u \in U, v \in V, (u - v)^T (B^{\kappa(u)})^{-1}(u - v) < \delta\}$ with $\kappa : U \to \{1, \ldots, K\}$ mapping predicted observation u to its target index. An edge $(u, v) \in E$ represents that observation v is validated for target u according to (3), hence, we are representing the validation matrix Ξ by E. Now a feasible joint event is a *matching* in G, *i.e.*, a subset $M \subset E$ such that no two edges in M share a vertex. The set of all feasible joint association events Ω can be represented as $\Omega \subset M_0(G) \cup \cdots \cup M_K(G)$, where $M_k(G)$ is a set of k-matchings in G. The posterior (6) of $\omega \in \Omega$ can be rewritten as

$$P(\omega|Y_{1:t}) = \frac{1}{Z} \lambda_{\mathrm{f}}^{N-|\omega|} p_{\mathrm{d}}^{|\omega|} (1 - p_{\mathrm{d}})^{K-|\omega|} \prod_{(u,v)\in\omega} \mathcal{N}_u(v), \tag{12}$$

where Z is a normalizing constant and $\mathcal{N}_u(v)$ is the Gaussian density function with mean u and covariance $B^{\kappa(u)}$.

The MCMC data association (MCMCDA) algorithm is an MCMC algorithm whose state space is the set of all feasible joint association events Ω and whose stationary distribution is the posterior (12). Each step of the MCMCDA algorithm is described in Algorithm 1, where we use the sampling method from [16]. A complete sampling strategy is described in Section V. In Algorithm 1, $A(\omega, \omega') = \min\left(1, \frac{\pi(\omega')}{\pi(\omega)}\right)$, where $\pi(\omega) = P(\omega|Y_{1:t})$ from (12). There are three MCMC moves and we name them for future reference: (i) an addition move proposes $\omega' = \omega + e$; (ii) a deletion move proposes $\omega' = \omega - e$; and (iii) a switch move proposes $\omega' = \omega + e - e'$.

Algorithm 1 MCMCDA (single step)

sample U from Unif$[0, 1]$
if $U < \frac{1}{2}$ **then**
  $\omega' = \omega$
else
  choose $e = (u, v) \in E$ uniformly at random
  if $e \in \omega$ **then**
    $\omega' = \omega - e$
  else if both u and v are unmatched in ω **then**
    $\omega' = \omega + e$
  else if exactly one of u and v is matched in ω and e' is the matching edge **then**
    $\omega' = \omega + e - e'$
  else
    $\omega' = \omega$
  end if
end if
$\omega = \omega'$ with probability $A(\omega, \omega')$

V. Analysis

Let $\mathcal{M}$ be the Markov chain simulated by Algorithm 1. Since the self-loop probability is nonzero, $\mathcal{M}$ is aperiodic. It can be easily seen that $\mathcal{M}$ is irreducible, *i.e.*, all states communicate, for example via the empty matching. In addition, the transitions described in Algorithm 1 satisfy the detailed balance condition (8) so $\mathcal{M}$ is reversible. Hence, by the ergodic theorem, the chain converges to its stationary distribution [17].

Let us first take a look at the complexity of the problem. As noted earlier, the state space of the Markov chain $\mathcal{M}$ is $\Omega \subset M_0(G) \cup \cdots \cup M_K(G)$, where $\mathcal{M}$ is the Markov chain simulated by Algorithm 1. For each k, $|M_k(G)| \le \binom{K}{k} \frac{N!}{(N-k)!}$ with equality if the subgraph of G with the k chosen vertices in U is a complete bipartite graph, *i.e.*, all observations are validated for all k chosen targets. Hence, we can bound the size of Ω as

$$\begin{aligned} |\Omega| &\le |M_0(G)| + \cdots + |M_K(G)| \\ &\le \sum_{k=0}^{K} \binom{K}{k} \frac{N!}{(N-k)!} =: \bar{\Omega}. \end{aligned} \tag{13}$$

Figure 1 shows this bound for $K = 5$ as a function of the number of observations. Certainly, the size of the state space grows exponentially as the number of targets or the number of observations increases, hence, the exact calculation of JPDA by enumeration is not feasible when the number of targets or the number of observations is large.

We first establish a few facts to prove the theorems below. In (12), the normalizing constant becomes

$$Z = \sum_{\omega\in\Omega} \left(\lambda_{\mathrm{f}}^{N-|\omega|} p_{\mathrm{d}}^{|\omega|} (1 - p_{\mathrm{d}})^{K-|\omega|} \prod_{(u,v)\in\omega} \mathcal{N}_u(v) \right). \tag{14}$$

We can bound each likelihood term as

$$\underline{L} \le \mathcal{N}_u(v) \le \bar{L},$$

for all $(u, v) \in E$, where

$$\begin{aligned} \bar{L} &= \max_{1\le k\le K} \left\{ \left((2\pi)^{n_y} |B^k| \right)^{-\frac{1}{2}} \right\} \\ \underline{L} &= \min_{1\le k\le K} \left\{ \left((2\pi)^{n_y} |B^k| e^{\delta} \right)^{-\frac{1}{2}} \right\}. \end{aligned}$$

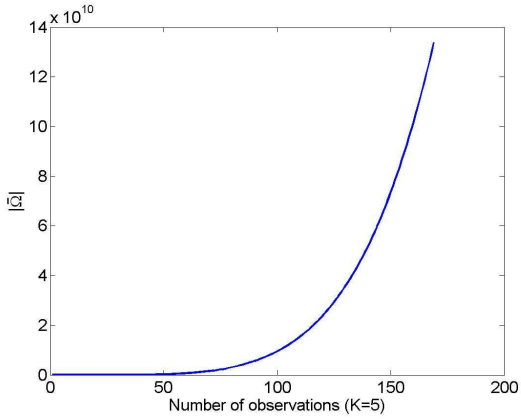

Fig. 1. $\bar{\Omega}$ as a function of the number of observations when $K = 5$

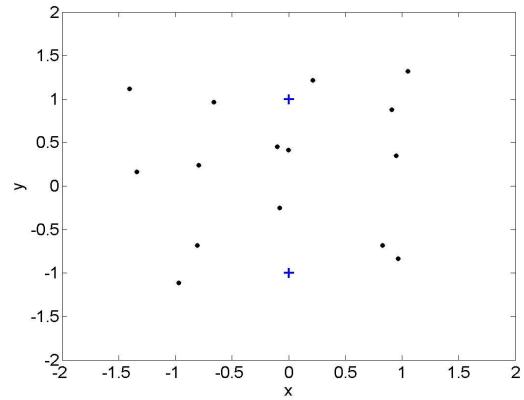

Fig. 2. Expected observations (crosses) and observations (dots)

The lower bound $\underline{L}$ is due to the measurement validation.

For Theorem 2 below, let $C = \frac{p_{\mathrm{d}}\bar{L}}{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}$, $D = \frac{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}{\underline{L}p_{\mathrm{d}}}$ and $R = \max\{1, C, D\}$. Also define $m_1 = \max\{1, \bar{L}\}$, $m_2 = \min\{1, \underline{L}\}$,

$$\begin{aligned} m_3(K,N) &= \max_{0\le k\le K}\{\lambda_{\mathrm{f}}^{N-k}p_{\mathrm{d}}^{k}(1-p_{\mathrm{d}})^{K-k}\}, \\ m_4(K,N) &= \min_{0\le k\le K}\{\lambda_{\mathrm{f}}^{N-k}p_{\mathrm{d}}^{k}(1-p_{\mathrm{d}})^{K-k}\}, \quad \text{and} \\ m_5(K,N) &= K\log\frac{m_1}{m_2} + \log\frac{m_3(K,N)}{m_4(K,N)} \\ &+ \sum_{k=1}^{K+1}\log k + \sum_{n=1}^{N}\log n. \end{aligned}$$

Remark 1: If $.5 < p_{\mathrm{d}} < 1$ and $\lambda_{\mathrm{f}} < 1 - p_{\mathrm{d}}$, then $m_3(K,N) = \lambda_{\mathrm{f}}^{N-K}p_{\mathrm{d}}^{K}$ and $m_4(K,N) = \lambda_{\mathrm{f}}^{N}(1-p_{\mathrm{d}})^{K}$. So $m_3(K,N)/m_4(K,N) = \left(\frac{p_{\mathrm{d}}}{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}\right)^{K}$ and K is the only remaining exponent.

Notice that the omitted proofs appear in Appendix.

Theorem 2: Suppose that $\lambda_{\mathrm{f}} > 0$ and $0 < p_{\mathrm{d}} < 1$. Then the mixing time of the Markov chain $\mathcal{M}$ is bounded by $\tau_x(\epsilon) \le 4R^4K^2N(m_5(K,N) + \log\epsilon^{-1})$ for all $x \in \Omega$.

Remark 2: Let $\bar{\tau}(\epsilon)$ be the upper bound found in Theorem 2. $\bar{\tau}(\epsilon)$ is polynomial in K and N. Under the assumptions in Remark 1, $\bar{\tau}(\epsilon) =$

$$O\left(K^2N\left(K\log\frac{Kp_{\mathrm{d}}}{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})} + N\log N + \log\epsilon^{-1}\right)\right).$$

If $m_3(K,N)/m_4(K,N)$ does not grow fast, *e.g.*, Remark 1, $\bar{\tau}(\epsilon) = O(K^2N(K\log K + N\log N + \log\epsilon^{-1}))$. If K is fixed, $\bar{\tau}(\epsilon) = O(N(N\log N + \log\epsilon^{-1}))$.

Let $p(\omega)$ be the distribution of the states of $\mathcal{M}$ after simulating Algorithm 1 for at least $\bar{\tau}(\epsilon)$ steps. Then the total variation distance satisfies $\|p-\pi\| \le \epsilon$. So we can sample from p to estimate $\{\beta_{jk}\}$. However, there is a small bias in our estimates since we are not sampling from π. The following theorem gives an upper bound on the number of samples needed for finding good estimates.

Theorem 3: Let $0 < \epsilon_1, \epsilon_2 \le 1$ and $0 < \eta < .5$. Suppose that $\|p - \pi\| \le \epsilon$ for $\epsilon \le \epsilon_1\epsilon_2/8$. Then, with a total of $504\epsilon_1^{-2}\epsilon_2^{-1}\lceil\log\eta^{-1}\rceil$ samples from p, we can find estimates $\hat{\beta}_{jk}$ for β_{jk} with probability at least $1-\eta$, such that, for $\beta_{jk} \ge \epsilon_2$, $\hat{\beta}_{jk}$ estimates β_{jk} within ratio $1+\epsilon_1$, *i.e.*, $(1-\epsilon_1)\beta_{jk} \le \hat{\beta}_{jk} \le (1+\epsilon_1)\beta_{jk}$, and, for $\beta_{jk} < \epsilon_2$, $\hat{\beta}_{jk} \le (1+\epsilon_1)\epsilon_2$.

Remark 3: Following Remark 2, for fixed K, $\bar{\tau}(\epsilon) = O(N(N\log N + \log\epsilon^{-1}))$. Combining this fact with Theorem 3, the time complexity of the overall procedure is $T = O(\epsilon_1^{-2}\epsilon_2^{-1}\log\eta^{-1}N(N\log N + \log(\epsilon_1^{-1}\epsilon_2^{-1})))$. Hence, with a total of T samples, Algorithm 1 finds estimates $\hat{\beta}_{jk}$ for β_{jk} with probability at least $1-\eta$, such that, for $\beta_{jk} \ge \epsilon_2$, $\hat{\beta}_{jk}$ estimates β_{jk} within ratio $1+\epsilon_1$, and, for $\beta_{jk} < \epsilon_2$, $|\hat{\beta}_{jk} - \beta_{jk}| \le (1+\epsilon_1)\epsilon_2$. We can simplify further by letting $\epsilon_0 = \epsilon_1\epsilon_2$. Then the time complexity is $O(\epsilon_0^{-2}\log\eta^{-1}N(N\log N + \log(\epsilon_0^{-1})))$.

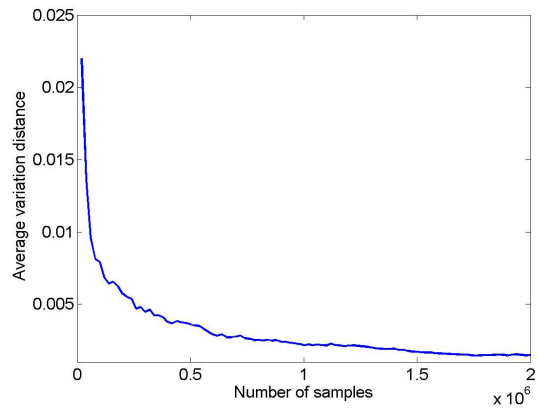

Fig. 3. Average variation Δ_β as a function of the number of samples

VI. Simulation Results

In this section, we show a simulation confirming our findings from last section. Since our goal is to estimate the association probabilities, we define the variation distance between two sets of association probabilities β_{jk} and $\hat{\beta}_{jk}$ by $\Delta_\beta = \max_{j,k}|\beta_{jk} - \hat{\beta}_{jk}|$. A simple case is chosen to demonstrate MCMCDA, in which two predicted observations are located at $[0,1]^T$ and $[0,-1]^T$ with $B^k = \mathrm{diag}(1,1)$ for $k \in \{1,2\}$. There are 15 observations as shown in Figure 2. Other parameters are: $\delta = 4$, $V = 16$, $\lambda_{\mathrm{f}} = .8125$, and $p_{\mathrm{d}} = .98$. In Figure 3, the average variation distance between two sets of association probabilities β_{jk} and $\hat{\beta}_{jk}$ from 10 independent runs is shown as a function of number of samples. β_{jk} are computed exactly by JPDA and $\hat{\beta}_{jk}$ are estimated by MCMCDA.

VII. Conclusions

JPDA is a powerful tool for solving data association problems but the exact computation of association probabilities in JPDA is NP-hard. Hence, for a large problem, we need to seek for an approximation algorithm. In this paper, we have presented an efficient approximation algorithm for JPDA based on Markov chain Monte Carlo data association (MCMCDA) and proved that the time complexity of the algorithm is polynomial in the size of the problem.

VIII. Acknowledgments

The first author would like to thank Prof. Stuart Russell for his comments and suggestions on the draft.

IX. Appendix

The proofs shown here parallel the proofs by Jerrum and Sinclair [16] in both structure and details; the main difference is the introduction of the non-uniform likelihood function into the posterior, allowing us to understand the relationship between the parameters in JPDA and the mixing time of the Markov chain.

A. Proof of Theorem 2

To prove Theorem 2, we need the following lemmas.

Lemma 1: Let $C = \frac{p_{\mathrm{d}}\bar{L}}{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}$ and $D = \frac{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}{\bar{L}p_{\mathrm{d}}}$. For any $\omega_0, \omega_1, \omega_2 \in \Omega$, if $\omega_1 = \omega_0 - e_0$, for some edge $e_0 \in \omega_0$, and $\omega_2 = \omega_1 - e_1$, for some edge $e_1 \in \omega_1$, then the following inequalities hold:

$$\begin{array}{rcl} \pi(\omega_0)/\pi(\omega_1) & \leq & C \\ \pi(\omega_0)/\pi(\omega_2) & \leq & C^2 \end{array} \quad \text{and} \quad \begin{array}{rcl} \pi(\omega_1)/\pi(\omega_0) & \leq & D \\ \pi(\omega_2)/\pi(\omega_0) & \leq & D^2. \end{array}$$

Proof: ω_0 and ω_1 are identical except that ω_1 is missing the edge e_0. So $|\omega_0| = |\omega_1| + 1$. If $e_0 = (u, v)$ and $k = |\omega_0|$,

$$\begin{aligned} \pi(\omega_0)/\pi(\omega_1) &= \frac{\lambda_{\mathrm{f}}^{N-k} p_{\mathrm{d}}^{k} (1-p_{\mathrm{d}})^{K-k}}{\lambda_{\mathrm{f}}^{N-(k-1)} p_{\mathrm{d}}^{k-1} (1-p_{\mathrm{d}})^{K-(k-1)}} \mathcal{N}_u(v) \\ &= \frac{p_{\mathrm{d}}}{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})} \mathcal{N}_u(v) \qquad \leq C. \end{aligned}$$

On the other hand,

$$\begin{aligned} \pi(\omega_1)/\pi(\omega_0) &= \frac{\lambda_{\mathrm{f}}^{N-(k-1)} p_{\mathrm{d}}^{k-1} (1-p_{\mathrm{d}})^{K-(k-1)}}{\lambda_{\mathrm{f}}^{N-k} p_{\mathrm{d}}^{k} (1-p_{\mathrm{d}})^{K-k}} \frac{1}{\mathcal{N}_u(v)} \\ &= \frac{\lambda_{\mathrm{f}}(1-p_{\mathrm{d}})}{p_{\mathrm{d}}} \frac{1}{\mathcal{N}_u(v)} \qquad \leq D. \end{aligned}$$

Since $\pi(\omega_0)/\pi(\omega_2) = \pi(\omega_0)/\pi(\omega_1) \times \pi(\omega_1)/\pi(\omega_2)$, by repeating the above argument twice, we get $\pi(\omega_0)/\pi(\omega_2) \leq C^2$. Similarly, we have $\pi(\omega_2)/\pi(\omega_0) \leq D^2$. ∎

Lemma 2: Let $R = \max\{1, C, D\}$, where C and D are defined in Lemma 1. Then the maximum edge loading of the Markov chain $\mathcal{M}$ is bounded as $\bar{\rho} \leq 4R^4K^2N$.

Proof: For each pair of matchings X, Y in G, we define the canonical path γ_{XY} as in [16]. Consider the symmetric difference $X \oplus Y$, where $X \oplus Y = (X - Y) \cup (Y - X)$. $X \oplus Y$ is a disjoint collection of paths in G including closed cycles, each of which has edges that belong to X and Y alternately. Suppose that we have fixed some arbitrary ordering on all simple paths in G, and designate a "start vertex" to each of the paths, which is arbitrary if the path is a closed cycle but must be an endpoint otherwise. This gives a unique ordering $P_1, P_2, \ldots, P_m$ on the paths appearing in $X \oplus Y$. The canonical path from X to Y involves "unwinding" each of the P_i in turn as follows. We need to consider two cases:

(i) *P_i is not a cycle*. Let P_i consist of the sequence $(v_0, v_1, \ldots, v_l)$ of vertices with the start vertex v_0. If $(v_0, v_1) \in Y$, perform a sequence of switching moves replacing (v_{2j+1}, v_{2j+2}) by (v_{2j}, v_{2j+1}) for $j = 0, 1, \ldots$, and finish with an addition move if l is odd. If $(v_0, v_1) \in X$, remove (v_0, v_1) and proceed as before for the reduced path $(v_1, \ldots, v_l)$.

(ii) *P_i is a cycle*. Let P_i consist of the sequence $(v_0, v_1, \ldots, v_{2l+1})$ of vertices, for $l \geq 1$, where v_0 is the start vertex, and $(v_{2j}, v_{2j+1}) \in X$ for $j = 0, \ldots, l$, with remaining edges belonging to Y. We first remove the edge (v_0, v_1). Now we are left with an open path O with endpoints v_0, v_1, with the start vertex v_k of O, for $k \in \{0, 1\}$. Then we unwind O as in (i) above but treating v_{1-k} as the start vertex to identify that it was a cycle.

Let t be an arbitrary edge in the Markov chain $\mathcal{M}$, *i.e.*, a transition from ω to $\omega' \neq \omega$. Let $cp(t) = \{(X, Y) : \gamma_{XY} \ni t\}$ be the set of canonical paths that use t. We define a function $\eta_t : cp(t) \to \Omega$ as in [16],

$$\eta_t(X, Y) = \begin{cases} X \oplus Y \oplus (\omega \cup \omega') - e_{XY_t}, & \\ \quad \text{if } t \text{ is a switch move and} & \\ \quad\quad \text{the current path is a cycle;} & \\ X \oplus Y \oplus (\omega \cup \omega'), & \text{otherwise,} \end{cases}$$

where e_{XY_t} is the edge in X adjacent to the start vertex that was removed first in (ii) above. $\eta_t(X, Y)$ is always a matching in G and η_t is injective as shown in [16]. Notice that the bipartite graph G considered here is a subset of the graphs considered in [16] so the arguments about η_t can be directly applied here.

Notice that

$$\begin{aligned} Q(t) &= Q(\omega, \omega') = \pi(\omega)P(\omega, \omega') \\ &= \frac{1}{2|E|} \min\{\pi(\omega), \pi(\omega')\}. \end{aligned} \tag{15}$$

Next, we bound $\pi(X)\pi(Y)$ and we need to consider four cases:

(i) *t is a deletion move*. We have $\omega' = \omega - e$ and $\eta_t(X, Y) = X \oplus Y \oplus (\omega \cup \omega')$. Since $\omega \cup \eta_t(X, Y)$ and $X \cup Y$ are identical when viewed as multisets,

$$\begin{aligned} \pi(X)\pi(Y) &= \pi(\omega)\pi(\eta_t(X, Y)) \\ &= \frac{2|E|Q(t)}{\min\{\pi(\omega), \pi(\omega')\}} \pi(\omega)\pi(\eta_t(X, Y)) \\ &= 2|E|Q(t) \max\left\{1, \frac{\pi(\omega)}{\pi(\omega')}\right\} \pi(\eta_t(X, Y)) \\ &\leq 2R|E|Q(t)\pi(\eta_t(X, Y)), \end{aligned}$$

where we used the identity (15) in the second equality and Lemma 1 for the last inequality.

(ii) *t is an addition move*. We have $\omega' = \omega + e$ and $\eta_t(X, Y) = X \oplus Y \oplus (\omega \cup \omega')$. Since $\omega \cup \eta_t(X, Y)$ and $X \cup Y$ are identical when viewed as multisets, using the arguments from (i),

$$\pi(X)\pi(Y) \leq 2R|E|Q(t)\pi(\eta_t(X, Y)).$$

(iii) *t is a switch move and the current path is a cycle*. Suppose $\omega' = \omega + e - e'$. Let $\omega_1 = \omega + e$. Then $\omega' = \omega_1 - e'$. Since $\frac{\pi(\omega)}{\pi(\omega')} = \frac{\pi(\omega_1)}{\pi(\omega')} \frac{\pi(\omega)}{\pi(\omega_1)}$, by Lemma 1, $\frac{\pi(\omega)}{\pi(\omega')} \leq CD \leq R^2$. Since $\eta_t(X, Y) = X \oplus Y \oplus (\omega \cup \omega') - e_{XY_t}$, the multisets $\omega \cup \eta_t(X, Y)$ differs from $X \cup Y$ only in that e and e_{XY_t} are missing from it. Hence, by Lemma 1,

$$\begin{aligned} \pi(X)\pi(Y) &\leq C^2\pi(\omega)\pi(\eta_t(X, Y)) \\ &= 2C^2|E|Q(t) \max\left\{1, \frac{\pi(\omega)}{\pi(\omega')}\right\} \pi(\eta_t(X, Y)) \\ &\leq 2R^4|E|Q(t)\pi(\eta_t(X, Y)). \end{aligned}$$

(iv) *t is a switch move and the current path is not a cycle*. This case is similar to (iii) but the multisets $\omega \cup \eta_t(X, Y)$ differs from $X \cup Y$ only in that e is missing from it. Hence, by Lemma 1,

$$\begin{aligned} \pi(X)\pi(Y) &\leq C\pi(\omega)\pi(\eta_t(X, Y)) \\ &= 2C|E|Q(t) \max\left\{1, \frac{\pi(\omega)}{\pi(\omega')}\right\} \pi(\eta_t(X, Y)) \\ &\leq 2R^3|E|Q(t)\pi(\eta_t(X, Y)). \end{aligned}$$

In summary, we have, in all cases,

$$\pi(X)\pi(Y) \leq 2R^4|E|Q(t)\pi(\eta_t(X, Y)).$$

Thus, for any transition t,

$$\begin{aligned} &\tfrac{1}{Q(t)} \textstyle\sum_{\gamma_{XY} \ni t} \pi(X)\pi(Y)|\gamma_{XY}| \\ &\leq 2R^4|E| \textstyle\sum_{\gamma_{XY} \ni t} \pi(\eta_t(X, Y))|\gamma_{XY}| \\ &\leq 4R^4K|E| \textstyle\sum_{\gamma_{XY} \ni t} \pi(\eta_t(X, Y)) \\ &\leq 4R^4K|E| \\ &\leq 4R^4K^2N \end{aligned}$$

where the second inequality follows from the fact that the length of any canonical path is bounded by $2K$, the third equality is due to the fact that η_t is injective and π is a probability distribution, and the last inequality follows from $|E| \leq KN$. Hence, $\bar{\rho} \leq 4R^4K^2N$. ∎

We now prove Theorem 2. $\mathcal{M}$ is a finite, reversible, ergodic Markov chain with loop probabilities $P(x,x) \geq \frac{1}{2}$ for all states x (see Section IV). Hence, by Theorem 1, we have

$$\tau_x(\epsilon) \leq \bar{\rho}(\log \pi(x)^{-1} + \log \epsilon^{-1}). \tag{16}$$

The upper bound for $\bar{\rho}$ is computed from Lemma 2. Now we just need to find the upper bound for $\pi(x)^{-1}$. From (14),

$$\begin{aligned} Z &\leq \sum_{\omega \in \Omega} m_1^K m_3(K,N) \\ &= m_1^K m_3(K,N)|\Omega| \\ &\leq m_1^K m_3(K,N) \sum_{k=0}^{K} \binom{K}{k} \frac{N!}{(N-k)!} \\ &\leq m_1^K m_3(K,N)(K+1)!N!, \end{aligned}$$

where the second inequality is by (13). Although this bound on Z is not tight, it will serve our purpose. For any $\omega \in \Omega$, $\pi(\omega) \geq \frac{1}{Z} m_2^K m_4(K,N)$ so

$$\begin{aligned} \frac{1}{\pi(\omega)} &\leq \frac{Z}{m_2^K m_4(K,N)} \\ &\leq \left(\frac{m_1}{m_2}\right)^K \frac{m_3(K,N)}{m_4(K,N)}(K+1)!N!. \end{aligned}$$

Hence,

$$\begin{aligned} \log \frac{1}{\pi(\omega)} &\leq \log\left(\left(\frac{m_1}{m_2}\right)^K \frac{m_3(K,N)}{m_4(K,N)}(K+1)!N!\right) \\ &= m_5(K,N). \end{aligned}$$

Putting all together, we have, for all initial state $x \in \Omega$, $\tau_x(\epsilon) \leq 4R^4K^2N(m_5(K,N) + \log \epsilon^{-1})$.

B. Proof of Theorem 3

Let $\beta_{\epsilon_2} = \{(j,k) : \beta_{jk} \geq \epsilon_2\}$. For now, assume $(j,k) \in \beta_{\epsilon_2}$, *i.e.*, $\beta_{jk} \geq \epsilon_2$. Let $X_{jk}(\omega) = \mathbb{I}((\hat{y}^k, y^j) \in \omega)$ where $\mathbb{I}$ is an indicator function. Notice that $\mathbb{E}_\pi(X_{jk}) = \pi(\omega_{jk}) = \beta_{jk}$, where $\omega_{jk} = \{\omega \in \Omega : (\hat{y}^k, y^j) \in \omega\}$. Since $\|p - \pi\| \leq \epsilon$,

$$\begin{aligned} |p(\omega_{jk}) - \pi(\omega_{jk})| &\leq \epsilon \leq \frac{\epsilon_1 \pi(\omega_{jk})}{8} \\ |\mathbb{V}\mathrm{ar}_p(X_{jk}) - \mathbb{V}\mathrm{ar}_\pi(X_{jk})| &\leq 3\epsilon \leq \frac{3\epsilon_1 \pi(\omega_{jk})}{8}. \end{aligned} \tag{17}$$

Let $\bar{\beta}_{jk} = \frac{1}{s}\sum_{i=1}^{s} X_{jk}(\omega_i)$ be the sample mean of s samples from p. Then $\mathbb{E}(\bar{\beta}_{jk}) = p(\omega_{jk})$ and $\mathbb{V}\mathrm{ar}(\bar{\beta}_{jk}) = \frac{1}{s}\mathbb{V}\mathrm{ar}_p(X_{jk})$. By Chebyshev's inequality,

$$P\left(|\bar{\beta}_{jk} - p(\omega_{jk})| > \frac{\epsilon_1}{3} p(\omega_{jk})\right) \leq \frac{9}{\epsilon_1^2 s} \frac{\mathbb{V}\mathrm{ar}_p(X_{jk})}{p(\omega_{jk})^2}. \tag{18}$$

Now if $|\bar{\beta}_{jk} - p(\omega_{jk})| \leq \frac{\epsilon_1}{3} p(\omega_{jk})$, from (17),

$$\begin{aligned} |\bar{\beta}_{jk} - \pi(\omega_{jk})| &\leq |\bar{\beta}_{jk} - p(\omega_{jk})| + |p(\omega_{jk} - \pi(\omega_{jk}| \\ &\leq \frac{\epsilon_1}{3} p(\omega_{jk}) + \frac{\epsilon_1}{8}\pi(\omega_{jk}) \\ &\leq \frac{\epsilon_1}{2}\pi(\omega_{jk}) \end{aligned} \tag{19}$$

and $\bar{\beta}_{jk}$ estimates $\pi(\omega_{jk})$ within ratio $1+\epsilon_1$. Since $\epsilon_1 < 1$ and $\mathbb{V}\mathrm{ar}_\pi(X_{jk}) \leq \pi(\omega_{jk})$,

$$\frac{\mathbb{V}\mathrm{ar}_p(X_{jk})}{p(\omega_{jk})^2} \leq \frac{\mathbb{V}\mathrm{ar}_\pi(X_{jk}) + \frac{3}{8}\pi(\omega_{jk})}{\left(\frac{7}{8}\pi(\omega_{jk})\right)^2} \leq \frac{2}{\pi(\omega_{jk})}. \tag{20}$$

Hence, by choosing $s = 72\epsilon_1^{-2}\epsilon_2^{-1}$ and using (18) and (20),

$$P\left(|\bar{\beta}_{jk} - p(\omega_{jk})| > \frac{\epsilon_1}{3} p(\omega_{jk})\right) \leq \frac{1}{4}, \tag{21}$$

that is, $\bar{\beta}_{jk}$ estimates $\pi(\omega_{jk})$ within ratio $1+\epsilon_1$ with probability at least $3/4$.

Now consider repeating the above experiment by an odd number t times, independently. Let $\hat{\beta}_{jk}$ be the median of the resulting t values of $\bar{\beta}_{jk}$. By (21), the probability that $\hat{\beta}_{jk}$ fails to approximate β_{jk} within ratio $1+\epsilon_1$ is at most

$$\begin{aligned} \sum_{i=(t+1)/2}^{t} \binom{t}{i}\left(\frac{1}{4}\right)^i\left(\frac{3}{4}\right)^{t-i} &\leq \left(\frac{1}{4}\right)^{t/2}\left(\frac{3}{4}\right)^{t/2} \sum_{i=(t+1)/2}^{t}\binom{t}{i} \\ &\leq \left(\frac{3}{16}\right)^{t/2} 2^t = \left(\frac{3}{4}\right)^{t/2}. \end{aligned}$$

Now let $t = 6\lceil \log \eta^{-1} \rceil + 1$, this probability is bounded above by η. Hence, with a total of st samples, $\hat{\beta}_{jk}$ estimates $\pi(\omega_{jk})$ within ratio $1+\epsilon_1$ with probability at least $1-\eta$ for $\beta_{jk} \geq \epsilon_2$. Notice that st is upper bounded by $504\epsilon_1^{-2}\epsilon_2^{-1}\lceil \log \eta^{-1} \rceil$.

Now consider β_{jk} that are smaller than ϵ_2. With probability at least $1-\eta$, for $(j,k) \in \beta_{\epsilon_2}$, $(1-\epsilon_1)\beta_{jk} \leq \hat{\beta}_{jk} \leq (1+\epsilon_1)\beta_{jk}$. So if $\hat{\beta}_{jk} \geq (1+\epsilon_1)\epsilon_2$, we must have $(j,k) \in \beta_{\epsilon_2}$. Hence, $\hat{\beta}_{jk} \leq (1+\epsilon_1)\epsilon_2$ for $\beta_{jk} < \epsilon_2$.

References

[1] I. Cox, "A review of statistical data association techniques for motion correspondence," *International Journal of Computer Vision*, vol. 10, no. 1, pp. 53–66, 1993.

[2] F. Dellaert, S. Seitz, C. Thorpe, and S. Thrun, "EM, MCMC, and chain flipping for structure from motion with unknown correspondence," *Machine Learning, special issue on Markov chain Monte Carlo methods*, vol. 50, pp. 45–71, 2003.

[3] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. San Diego, CA: Mathematics in Science and Engineering Series 179 Academic Press, 1988.

[4] J. Collins and J. Uhlmann, "Efficient gating in data association with multivariate distributed states," *IEEE Trans. Aerospace and Electronic Systems*, vol. 28, no. 3, 1992.

[5] L. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science*, vol. 8, pp. 189–201, 1979.

[6] R. Fitzgerald, "Development of practical PDA logic for multipltarget tracking by microprocessor," in *Multitarget-Multisensor Tracking: Advanced Applications*, Y. Bar-Shalom, Ed. Artech House: Norwood, MA, 1990.

[7] T. Huang and S. J. Russell, "Object identification in a Bayesian context," in *Proc. of the Fifteenth International Joint Conference on Artificial Intelligence*, 1997, pp. 1276–1283.

[8] N. Bergman and A. Doucet, "Markov chain Monte Carlo data association for target tracking," in *Proc. of IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2000.

[9] H. Pasula, S. J. Russell, M. Ostland, and Y. Ritov, "Tracking many objects with many sensors," in *Proc. IJCAI-99*, Stockholm, 1999.

[10] S. Cong, L. Hong, and D. Wicker, "Markov-chain Monte-Carlo approach for association probability evaluation," *IEE Proceedings of Control, Theory and Applications*, vol. 151, no. 2, pp. 185–193, March 2004.

[11] S. Oh, S. Russell, and S. Sastry, "Markov chain Monte Carlo data association for general multiple-target tracking problems," in *Proc. of the 43rd IEEE Conference on Decision and Control*, Paradise Island, Bahamas, Dec. 2004.

[12] D. Reid, "An algorithm for tracking multiple targets," *IEEE Transaction on Automatic Control*, vol. 24, no. 6, pp. 843–854, December 1979.

[13] S. Oh, L. Schenato, and S. Sastry, "A hierarchical multiple-target tracking algorithm for sensor networks," in *Proc. of the International Conference on Robotics and Automation*, Barcelona, Spain, April 2005.

[14] D. Schulz, W. Burgard, D. Fox, and A. Cremers, "Tracking multiple moving targets with a mobile robot using particle filters and statistical data association," in *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, 2001.

[15] I. Beichl and F. Sullivan, "The Metropolis algorithm," *Computing in Science and Engineering*, vol. 2, no. 1, pp. 65–69, 2000.

[16] M. Jerrum and A. Sinclair, "The Markov chain Monte Carlo method: An approach to approximate counting and integration," in *Approximations for NP-hard Problems*, D. Hochbaum, Ed. PWS Publishing, Boston, MA, 1996.

[17] G. Roberts, "Markov chain concepts related to sampling algorithms," in *Markov Chain Monte Carlo in Practice*, ser. Interdisciplinary Statistics Series, W. Gilks, S. Richardson, and D. Spiegelhalter, Eds. Chapman and Hall, 1996.

[18] P. Diaconis and D. Stroock, "Geometric bounds for eigenvalues of Markov chains," *Annals of Applied Probability*, vol. 1, pp. 36–61, 1991.