An Efficient Algorithm for Tracking Multiple Maneuvering Targets

Songhwai Oh and Shankar Sastry

Abstract—Tracking multiple maneuvering targets in a cluttered environment is a challenging problem. A combination of interacting multiple model (IMM) and joint probabilistic data association (JPDA) has been successfully applied to track multiple maneuvering targets. In IMM, the motion of a maneuvering target is approximated by a finite number of simple, distinct kinematic models. However, the exact computation of the combined approach has the time complexity which is exponential in the numbers of kinematic models and measurements. When applying JPDA and IMM, the numbers of targets and kinematic models are known, so we can design a tracking system suitable for the given numbers of targets and kinematic models. But the number of measurements is not known in advance, and it poses a serious problem in computing association probabilities in JPDA. Hence, for a large problem, we need to seek for an efficient approximation algorithm. In this paper, we present a randomized algorithm which finds approximations of association probabilities with good fidelity and prove that the time complexity of the algorithm is polynomial in the size of the problem.

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 [3].

Tracking multiple maneuvering targets in a cluttered environment is a challenging problem. A combination of interacting multiple model (IMM) [4] and joint probabilistic data association (JPDA) [3] has been successfully applied to track multiple maneuvering targets, e.g., IMM-JPDA [5]. In IMM, the motion of a maneuvering target is approximated by a finite number of simple, distinct kinematic models. For survey of IMM, see [6]. However, the exact computation has the time complexity which is exponential in the numbers of kinematic models and measurements. When applying JPDA and IMM, the numbers of targets and kinematic models are known, so we can design a tracking system suitable for the given numbers of targets and kinematic models, e.g., parallel computing. However, the number of measurements is not known in advance, and it poses a serious problem in computing association probabilities in JPDA.

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 \( \beta_{jk} \) 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].

The exact calculation of association probabilities \( \beta_{jk} \) in JPDA is NP-hard [7] since the related problem of finding the permanent of a matrix is #P-complete [8]. To overcome the complexity of the problem, many approximation algorithms have been proposed. Some heuristic approaches to approximate JPDA include a “cheap” JPDA algorithm [9], “suboptimal” JPDA [10] and “near-optimal” JPDA [11]. In [12], 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 [2], [13]. In [14], Markov chain Monte Carlo (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.

This paper presents a randomized algorithm, named Markov chain Monte Carlo data association (MCMCDA), for computing association probabilities required for IMM-JPDA and proves that the time complexity of the algorithm is polynomial in the number of kinematic models and the number of measurements. In [15], 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) [16] and outperforms MHT with heuristics, such as pruning, gating, clustering, \( N \)-scan-back logic and \( k \)-best hypotheses, under extreme conditions, such as a large number of targets in a dense environment, low detection probabilities, and high false alarm rates [15]. 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 [17]. In [18], it has been shown that a special case MCMCDA finds good estimates of association probabilities in polynomial time. This paper extends the results to the problem of tracking a known number of maneuvering targets.

The remainder of this paper is structured as follows. We describe a method to estimate states of multiple linear hybrid systems 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.
II. State Estimation of Multiple Linear Hybrid Systems

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

$$x_{t+1}^k = A_t^k(x_t^k)x_t^k + G_t^k(\nu_t^k)w_t^k(x_t^k), \quad \text{for } t = 1, 2, \ldots, $$

where $x_t^k \in \mathbb{R}^{n_x}$ is the state of target $k$ at time $t$, $\nu_t^k$ denotes the kinematic model representing the motion of target $k$ at time $t$; $A_t^k(\nu_t^k)$ and $G_t^k(\nu_t^k)$ are matrices with appropriate sizes; and $w_t^k(\nu_t^k)$ is a white Gaussian process with zero mean and covariance $Q_t^k(\nu_t^k)$. The evolution of $\nu_t^k$ is modeled by a finite state Markov chain taking values from $\{1, \ldots, M\}$ according to a transition probability matrix $P^m_k = [\pi_{jk}^\Omega]$. This linear hybrid system is also known as a jump linear system [19] and has been successfully applied to approximate a class of nonlinear systems [20].

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 (or outliers) and the number of false alarms has a Poisson distribution with a parameter $\lambda_f V$ where $V$ is the volume of $\mathcal{R}$ and $\lambda_f$ is the false alarm rate per unit volume. Let $n_i$ be the number of observations at time $t$, including both noisy observations and false alarms. Let $y_t^j \in \mathbb{R}^{n_x}$ 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

$$y_t^j = \begin{cases} C_t^j(\nu_t^k)x_t^k + w_t^j(\nu_t^k) & \text{if } y_t^j \text{ is from } x_t^k \\ u_t^j & \text{otherwise} \end{cases}$$

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

For notational convenience, we assume $A(\cdot) = A_t^k(\cdot)$, $G(\cdot) = G_t^k(\cdot)$, $Q(\cdot) = Q_t^k(\cdot)$, $C(\cdot) = C_t^j(\nu_t^k)$, $R(\cdot) = R_t^j(\cdot)$, and $P_m = P_m^k$, for all $k$, $t$, and $j$. Since we are operating at the filtering step, we further simplify our notations by dropping the subscript $t$. Let us denote the event \{$(\nu_t^k = i)$ by $\nu_t^k$ and let $\Omega$ be a set of all feasible joint association events. For each $\omega \in \Omega$, $\omega = \{(j, k)\}$, where $(j, k)$ denotes an event that $j$-th observation is associated with target $k$. Then the state of target $k$ can be estimated as

$$E(x_t^k|Y_{1:t}) = \sum_{i=1}^M E(x_t^k|\nu_t^k, Y_{1:t})P(\nu_t^k|Y_{1:t}), \quad \text{(3)}$$

where

$$E(x_t^k|\nu_t^k, Y_{1:t}) = \sum_{\omega} E(x_t^k|\omega, \nu_t^k, Y_{1:t})P(\omega|\nu_t^k, Y_{1:t}), \quad \text{(4)}$$

and

$$\omega = \{(j, k)\}$$

$$\omega_0$$

where $\omega_jk$ denotes the event \{$\omega \ni (j, k)$\} and $\omega_0$ denotes the event that no observation is associated with target $k$. Notice that other parameters, such as covariance matrices, can be computed similarly and interested readers are referred to [3], [5]. In (4), $E(x_t^k|\omega_jk, \nu_t^k, Y_{1:t})$ can be computed easily by considering it as a single target estimation problem where the current observation is the $j$-th observation. Hence, the computation of $E(x_t^k|\nu_t^k, Y_{1:t})$ reduces to the computation of association probability $\beta_{jk}(i)$, where

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

The computation of $\beta_{jk}(i)$ 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 $\beta_{jk}(i)$ in the presence of the identity uncertainty while IMM computes (3) using model posteriors $P(\mu_t^k|Y_{1:t})$. As mentioned earlier, the exact calculation of $\beta_{jk}(i)$ in JPDA is NP-hard [7] and it is the major drawback of JPDA. In next sections, we introduce a randomized algorithm which approximates $\beta_{jk}(i)$ and $P(\mu_t^k|Y_{1:t})$ without the enumerations of all joint association events and all possible combinations of kinematic models.

III. Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) plays a significant role in many fields such as physics, statistics, economics, and engineering [21]. In some cases, MCMC is the only known general algorithm that finds a good approximate solution to a complex problem in polynomial time [22]. MCMC techniques have been applied to complex probability distribution integration problems, counting problems such as #P-complete problems, and combinatorial optimization problems [21], [22].

MCMC is a general method to generate samples from a distribution $\pi$ by constructing a Markov chain $\mathcal{M}$ with states $\omega$ 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

$$A(\omega, \omega') = \min \left(1, \frac{\pi(\omega')q(\omega', \omega)}{\pi(\omega)q(\omega, \omega')} \right), \quad \text{(6)}$$

otherwise the sampler stays at $\omega$, so that the detailed balance condition is satisfied, i.e.,

$$Q(\omega, \omega') = \pi(\omega)P(\omega, \omega') = \pi(\omega')P(\omega, \omega'), \quad \text{(7)}$$

for all $\omega, \omega' \in \Omega$, where $P(\omega, \omega') = q(\omega, \omega')A(\omega, \omega')$ is the transition probability from $\omega$ to $\omega'$ 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 [23]. Hence, for a given bounded function $f$, the sample mean $f$ of $f$ over the sampled states converges to $E\pi f(\omega)$. Notice that (7) requires only the ability to compute the ratio $\pi(\omega')/\pi(\omega)$, avoiding the need to normalize $\pi$.

An ergodic chain $\mathcal{M}$ on state space $\Omega$ 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 $P_x^t(\cdot)$ be the distribution of the state at time $t$ given that $\mathcal{M}$ is started from the initial state $x \in \Omega$. If $\pi$ is the stationary distribution of $\mathcal{M}$, then the total variation distance at time $t$ with initial state $x$ is defined as

$$\Delta_x(t) = \|P_x^t - \pi\| = \max_{S \subset \Omega} |P_x^t(S) - \pi(S)| \quad \text{(8)}$$
The rate of convergence of $\mathcal{M}$ to stationarity can be measured by the mixing time:

$$\tau_x(\epsilon) = \min\{t : \Delta_x(s) \leq \epsilon \text{ for all } s \geq t\}.$$  (9)

One approach to bound $\tau_x(\epsilon)$ of a Markov chain with a complex structure is the canonical path method [22]. 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 a finite, reversible and ergodic Markov chain $\mathcal{M}$ with state space $\Omega$, consider an undirected graph $G = (V, E)$ where $V = \Omega$ and $E = \{(x, y) : Q(x, y) > 0\}$. For each ordered pair $(x, y) \in \Omega^2$, the canonical path $\gamma_{xy}$ is a simple path from $x$ to $y$ in $G$. In terms of $\mathcal{M}$ the canonical path $\gamma_{xy}$ 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:

$$\bar{\rho} = \rho(\Gamma) = \max_{x \in \Omega} \frac{1}{Q(e)} \sum_{\gamma_{xy} \ni e} \pi(x)\pi(y) |\gamma_{xy}|.$$  (10)

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 [22]:

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

IV. MCMC DATA ASSOCIATION ALGORITHM

In this section, we describe the MCMC data association (MCMCDA) algorithm for efficiently approximating the association probabilities $\beta_k(i)$ and model posteriors $P(\mu^k_t|Y_{1:t})$.

In JPDA, measurement validation is used to reduce the number of measurement considered in computation of association probabilities. The same measurement validation is used for MCMCDA but we later find that it is a critical step to approximate association probabilities in polynomial time. Let $\hat{y}^k$ be the predicted observation for target $k$ when $\mu^k_t$, i.e., $\hat{y}^k = C(\hat{\delta})\bar{E}(\hat{\gamma}_{ij}^k|\mu^k_t, Y_{1:t-1})$. Suppose there are $N$ observations and let $v_{j,k}^i = y^k - \hat{y}^k$ for $j = 1, \ldots, N$. The covariance of $v_{j,k}^i$ is $B_{k,i} = \mathbb{E}(v_{j,k}^i v_{j,k}^iT|\mu^k_t, Y_{1:t-1})$. For each target $k$, let $B_k = \max_{i,j} |B_{k,i}|$. The measurement $y^k$ is validated for target $k$, if and only if, for some $i$,

$$v_{j,k}^iT(B_k)^{-1}v_{j,k}^i < \delta,$$  (11)

where $\delta$ is an appropriate threshold. We assume that all measurements are validated with at least one target. If not, we can always consider the reduced problem, which consists of only validated measurements and targets with at least one validated measurement, and separately estimate states of targets with no validated measurement.

We encode the feasible joint association events in a graph. Let $G = (V_{\text{ex}}, E_{\text{ex}})$ be a graph with vertex set $V_{\text{ex}} = W \cup U \cup V$ and edge set $E_{\text{ex}} = F \cup E$, where $W = \{\mu^k_t : 1 \leq k \leq K, 1 \leq i \leq M\}$ is a set of kinematic models, $U = \{k : 1 \leq k \leq K\}$ is a set of target indices, $V = \{y^j : 1 \leq j \leq N\}$ is a set of observations, $F = \{(u, v) : \kappa(u) = u, w \in W, u \in U\}$ with $\kappa : W \rightarrow \{1, \ldots, K\}$ mapping kinematic model $w \in W$ to its target index, and $E = \{(u, v) : \exists i \text{ s.t. } (\hat{y}^u_i - v)^T(B^u_i)^{-1}(\hat{y}^u_i - v) < \delta, u \in U, v \in V\}$. An edge $(u, v) \in E$ represents that observation $v$ is validated for target $u$ according to (11). An edge $(u, v) \in F$ represents that the kinematic model of target $u$ is $w$. Note that all kinematic models $\{\mu^k_t : 1 \leq i \leq M\}$ of target $u$ are only connected to $u$. For future reference, let $G' = (U, V, E)$ and $G^m = (W, U, F)$; and let $i : W \rightarrow \{1, \ldots, M\}$ be a mapping from kinematic model $w \in W$ to its model index.

We now define a set of feasible events over $G$. Let $\Omega$ be a set of all feasible events defined over $G$ such that $\omega = (\omega^m, \omega^a) \in \Omega$ with $\omega^m \subset F$ and $\omega^a \subset E$ forms a subgraph in $G$. An event $\omega = (\omega^m, \omega^a) \in \Omega$ is feasible if $\omega^a$ is a matching in $G^m$ and $\omega^m$ is a $k$-matching in $G^m$. Based on the parametric false alarm model, the posterior of $\omega \in \Omega$ can be computed as

$$P(\omega|Y_{1:t}) = P(\omega^a|\omega^m)P(Y_{1:t}|\omega^m) = \frac{1}{Z} \frac{1}{N_{\omega}^m} P^{(\omega^m)}|\omega^a|P(Y_{1:t-1}|\omega^m) \times \prod_{(u, v) \in \omega^a} \mathcal{N}(u; \omega^a).$$  (12)

where $Z$ is a normalizing constant and $\mathcal{N}(u; \omega^m)$ is the Gaussian density function of variable $v$ with mean $\hat{y}^u_i$ and covariance $B^u_i$ for $(u, v) \in \omega^a$. Notice that $P(\mu^m_t|Y_{1:t-1})$ are computed from the interaction step of IMM [5].

The MCMC data association (MCMCDA) algorithm is an MCMC algorithm whose state space is the set of all feasible events $\Omega$ and whose stationary distribution is the posterior (12). Each step of the MCMCDA algorithm is described in Algorithm 1, where $A(\omega, \omega') = \min\left(1, \frac{\pi(\omega')}{\pi(\omega)}\right)$ and $\pi(\omega) = P(\omega|Y_{1:t})$ from (12). There are four MCMC moves and we name them for future reference: (i) a model switch move proposes $\omega' = \omega^a + e - e'$; (ii) an addition move proposes $\omega' = \omega^a + e$; (iii) a deletion move proposes $\omega'^a = \omega^a - e$; and (iv) a switch move proposes $\omega'^m = \omega^m + e - e'$.

Now suppose that Algorithm 1 is repeated for $S$ steps and let $w^m_\delta$ be the state of the chain at step $s$. Let $\hat{\rho}^k_t = \frac{\delta}{2} \sum_{s=1}^{S} \|((\mu^k_t, k) \in \omega^m_s)\|_{\beta_k(i)} = \frac{1}{2\rho_k^t} \sum_{s=1}^{S} \|((k, y') \in \omega^a_s, (\mu^k_t, k) \in \omega^m_s)\|_{\beta_k(i)}$ for $\rho_k^t > 0$ and $\beta_k(i) = 0$ for $\rho_k^t = 0$, where $\|\|$ is an indicator function. If the Markov chain is ergodic, by the ergodic theorem [23], $\beta_k(i) \rightarrow \beta_k(i)$ and $\hat{\rho}^k_t \rightarrow P(\mu^k_t|Y_{1:t})$ almost surely as $S \rightarrow \infty$. In the next section, we show that the Markov chain simulated by Algorithm 1 is ergodic and analyze the rate of this convergence.

If the chain converges to its stationary distribution fast, we can find good approximations with small $S$.

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 (7) so $\mathcal{M}$ is reversible. Hence, by the ergodic theorem, the chain converges to its stationary distribution [23].
Algorithm 1 (Multiple-model) MCMCDA (single step)

Sample $U$ from Unif[0,1]
if $U < \frac{5}{6}$ then
  $\omega' = \omega$
else
  $(\omega^m, \omega) = \omega$
  Choose $e = (u, v) \in E_{ex} \setminus \omega^m$ uniformly at random
  if $e \in F$ then
    $\omega' = \omega^m + e - e'$, where $e' = (w, v) \in \omega^m$
  else if $e \in \omega^a$ then
    $\omega'^a = \omega^a - e$
    else if both $u$ and $v$ are unmatched in $\omega^a$ then
      $\omega'^a = \omega^a + e$
    else if exactly one of $u$ and $v$ is matched in $\omega^a$ and
      $e'$ is the matching edge then
      $\omega'^a = \omega^a + e - e'$
    else
      $\omega'^a = \omega^a$
  end if
end if
end if
\[ \omega' = (\omega^m, \omega^a) \]
\[ \omega = \omega' \text{ with probability } A(\omega, \omega') \]

We first establish few facts. In (12), the normalizing constant is
\[ Z = \sum_{\omega \in \Omega} P(\omega^a|Y_{t-1})P(Y_t|\omega^a, \omega^m, Y_{t-1})P(\omega^m|Y_{t-1}). \]
We can bound each likelihood term as $L \leq N_u(v; \omega^m) \leq \bar{L}$, for all $(u, v) \in E_a$ and $\omega^m$, where
\[ \bar{L} = \max_{1 \leq k \leq K, 1 \leq i \leq M} \left\{ \left( (2\pi)^{n_a}|B^{k,i} \right)^{-\frac{1}{2}} \right\}. \]
\[ L = \min_{1 \leq k \leq K, 1 \leq i \leq M} \left\{ \left( (2\pi)^{n_a}|B^{k,i} \right)^{c(k,i)} \right\}. \]
Here, $B^{k,i}_a$ are positive definite matrices and $c(k, i) = \lambda_{\max}(B^{k,i})^{-1}/\lambda_{\min}(B^{k,i})^{-1}$, where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the maximum and minimum eigenvalues of matrix $A$, respectively. We assume that targets are maintained by covariance control and $c(k, i) \leq \bar{c}$ for some constant $\bar{c}$. Notice that the lower bound $\bar{L}$ is due to the measurement validation.

To prove Theorem 2, we need the following lemma. Note that the omitted proofs are given in appendix.

Lemma 1: For any $\omega_0, \omega_1, \omega_2 \in \Omega$, if $\omega_0^a = \omega_1^a = \omega_2^a$ and $\omega_1 = \omega_2 - e_0$, for some edge $e_0 \in \omega_1$, and $\omega_2 = \omega_1 - e_1$, for some edge $e_1 \in \omega_1$, then $\pi(\omega_0)/\pi(\omega_1) \leq C_1, \pi(\omega_0)/\pi(\omega_2) \leq C_2, \pi(\omega_1)/\pi(\omega_2) \leq D_1$, and $\pi(\omega_2)/\pi(\omega_0) \leq D_2$, where $C = \frac{p_d L}{\lambda(1-p_d)}$ and $D = \lambda(1-p_d)$.

Lemma 2: Suppose that $P(\mu^k_Y|Y_{t-1}) \geq \mu$ for all $1 \leq i \leq M$ and $1 \leq k \leq K$. For any $\omega_1, \omega_2 \in \Omega$, if $\omega_1^a = \omega_2^a$ and $\omega_1 = \omega_2 - e_1 + e_2$ for edges $e_1 \in \omega_1$ and $e_2 \in \omega_2$ with $e_1$ and $e_2$ sharing a common vertex, then $\frac{\pi(\omega_1)}{\pi(\omega_2)} \leq H$, where $H = \bar{L}/(\mu \bar{L})$.

Lemma 3: Suppose that $P(\mu^k_Y|Y_{t-1}) \geq \mu$ for all $1 \leq i \leq M$ and $1 \leq k \leq K$. Let $R = \max\{1, C, D, H\}$, where $C$ and $D$ are defined in Lemma 1 and $H$ is defined in Lemma 2. Then the maximum edge loading of the Markov chain $\mathcal{M}$ is bounded as $p \leq 6L^2K^2(N + M)$. 

Remark 1: In Lemma 3, we have assumed that $P(\mu^k_Y|Y_{t-1}) \geq \mu$ for all $i$ and $k$. If $P(\mu^k_Y|Y_{t-1}) < \mu$, then the contribution from model $i$ for target $k$ is small in (3) and its contribution can be safely ignored. Thus, one strategy is to ignore models whose priors are less than the threshold $\mu$. It can be seen as “model validation” similar to the measurement validation in IPDA. Hence, from now on, we assume that $P(\mu^k_Y|Y_{t-1}) \geq \mu$ for all $1 \leq i \leq M$ and $1 \leq k \leq K$.

For Theorem 2 below, define $m_1 = \max\{1, \bar{L}\}$, $m_2 = \min\{1, \bar{L}\}$,
\[ m_3(K, N) = \max_{0 \leq k \leq K} \{ \lambda_k^{1-k} \rho_k^2 (1 - p_d)K^{-k} \}, \]
\[ m_4(K, N) = \min_{0 \leq k \leq K} \{ \lambda_k^{1-k} \rho_k^2 (1 - p_d)K^{-k} \}, \]
\[ m_5(K, N) = K \log m_1 M \mu_2 + \log m_3(K, N) \mu_4(K, N) \]
\[ + \sum_{k=1}^{K+1} \log k + \sum_{n=1}^{N} \log n. \]

Theorem 2: Suppose that $\lambda > 0$ and $0 < p_d < 1$. Then the mixing time of the Markov chain $\mathcal{M}$ is bounded by $\tau(\epsilon) \leq 6L^2K^2(N + M)(m_5(K, N)\mu_2 \log \epsilon^{-1})$ for all $x \in \Omega$.

Remark 2: Let $\tau(\epsilon)$ be the upper bound found in Theorem 2. Since $m_5(K, N) \mu_2$ is polynomial in $K$ and $N$, $\tau(\epsilon)$ is polynomial in the number of kinematic models $K$ and the number of measurements $N$. If we assume that both $K$ and $M$ are fixed, $\tau(\epsilon) = \Omega(N \log N + \log \epsilon^{-1})$.

Let $p(\omega)$ be the distribution of the states of $\mathcal{M}$ after simulating Algorithm 1 at least $\tau(\epsilon)$ steps. Then the total variation distance satisfies $\|p - \pi\| \leq \epsilon$. Hence, for a given bounded function $f : \Omega \rightarrow \mathbb{R}$, we can estimate $\mathbb{E}_{\pi} f$ by the sample mean $f = \frac{1}{N} \sum_{s=1}^{N} f(\omega_s)$, where $\{\omega_s\}$ are sampled from $p$. However, there is a small bias in our estimates since we are not sampling from $\pi$. The following theorem from [18] gives an upper bound on the number of samples needed for finding good estimates.

Theorem 3: Let $0 < \epsilon_1, \epsilon_2 \leq 1$ and $0 < \eta < .5$. Suppose that $\|p - \pi\| \leq \epsilon \leq \epsilon_1 \epsilon_2 / 8$. If $f : \Omega \rightarrow [0, 1]$, then, with a total of $504 \epsilon_2^{-2} \epsilon_1^{-2} [\log \eta^{-1}]$ samples from $p$, we can find estimates $\hat{f}$ for $\mathbb{E}_{\pi} f$ with probability at least $1 - \eta$, such that, for $\mathbb{E}_{\pi} f \geq \epsilon_2$, $\hat{f}$ estimates $\mathbb{E}_{\pi} f$ within ratio $1 + \epsilon_1$, i.e., $(1 - \epsilon_1) \mathbb{E}_{\pi} f \leq \hat{f} \leq (1 + \epsilon_1) \mathbb{E}_{\pi} f$, and for $\mathbb{E}_{\pi} f < \epsilon_2$, $\hat{f} \leq (1 + \epsilon_1) \epsilon_2 / 2$.

Following Remark 2, for fixed $K$ and $M$, $\tau(\epsilon) = \Omega(N \log N + \log \epsilon^{-1})$. Combining this fact with Theorem 3, the time complexity of the overall procedure is $S = \Omega(\epsilon_1^{-1} \epsilon_2^{-2} \log \eta^{-1} N \log N + \log(\epsilon_2^{-1} \epsilon_1^{-1}))$. Hence, with a total of $S$ samples, Algorithm 1 finds estimates $\hat{f}$ for $\mathbb{E}_{\pi} f$ with probability at least $1 - \eta$, such that, for $\mathbb{E}_{\pi} f \geq \epsilon_2$, $\hat{f}$ estimates $\mathbb{E}_{\pi} f$ within ratio $1 + \epsilon_1$, and, for $\mathbb{E}_{\pi} f < \epsilon_2$, $\hat{f} \leq (1 + \epsilon_1) \epsilon_2 / 2$. We can simplify further by letting $\epsilon_0 = \epsilon_1 \epsilon_2 / 2$. Then the time complexity is $\Omega(\epsilon_2^{-2} \log \eta^{-1} N \log N + \log(\epsilon_2^{-1} \epsilon_1^{-1}))$.

Hence, we can estimate $P(\mu^k_i|Y_{t-1})$ by letting $f(\omega) = \|\mu^k_i - \omega\|$. The association probability $\beta_{jk}(i)$ can be estimated by $f_1 / f_2$ where $f_1(\omega) = 1(\{k, y\} \in \omega^a, (\mu^k_i, k) \in \omega^m)$ and $f_2(\omega) = \|\mu^k_i - \omega\|$. As long as $f_2 \geq \epsilon'$ for some $\epsilon' > 0$, we can find a good approximation of $\beta_{jk}(i)$.
APPENDIX

Proof of Lemma 1

\[ \pi(\omega_0) / \pi(\omega_1) = \frac{\lambda^N}{\lambda^k} \frac{p_d}{p_d} (1-p_d)^{k-1} N_0(v; \omega_0^p) \]

On the other hand,

\[ \pi(\omega_1) / \pi(\omega_0) = \frac{\lambda^N}{\lambda^k} \frac{p_d}{p_d} (1-p_d)^{k-1} N_0(v; \omega_1^p) \]

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

Proof of Lemma 2

Suppose that \( e_1 = (w_1, k) \) and \( e_2 = (w_2, k) \) and let \( i_1 = i(w_1) \) and \( i_2 = i(w_2) \). Since \( \omega_1^i = \omega_2^i \),

\[ \pi(\omega_1) / \pi(\omega_2) = \frac{P(\mu_i^k|Y_1 \cdot \cdot \cdot Y_{t-1})}{P(\mu_i^k|Y_1 \cdot \cdot \cdot Y_{t-1})} \frac{P(Y_1|\omega_1, Y_1 \cdot \cdot \cdot Y_{t-1})}{P(Y_1|\omega_2, Y_1 \cdot \cdot \cdot Y_{t-1})} \]

Now at most one observation is connected to the vertex \( k \). If there is an observation connected to \( k \), likelihoods \( P(\mu_i^k|Y_1 \cdot \cdot \cdot Y_{t-1}) \) and \( P(Y_1|\omega_2, Y_1 \cdot \cdot \cdot Y_{t-1}) \) only differ for this observation. Hence, the likelihood ratio in (14) is bounded above by \( \bar{L}_i / L \). Notice that \( \bar{L}_i / L \geq 1 \) since \( P(\mu_i^k|Y_1 \cdot \cdot \cdot Y_{t-1}) \geq \mu_i^k \).

Proof of Lemma 3

For \( X, Y \in \Omega \), the canonical path \( \gamma_{XY} \) is defined as follows. Consider the symmetric differences \( X^a \oplus Y^a \) and \( X^a \oplus Y^a \), where \( X \oplus Y = (X - Y) \cup (Y - X) \). \( X^a \oplus Y^a \) is a disjoint collection of paths in \( G_0^a \), each of which has edges that belong to \( X^a \) and \( Y^a \) alternately. \( X^a \oplus Y^a \) is a disjoint collection of paths in \( G_0^a \) including closed cycles, each of which has edges that belong to \( X^a \) and \( Y^a \) alternately. We first fix ordering on simple paths in \( G_0^a \) followed by simple paths in \( G_0 \). Then we can order each path in \( G_0^a \) by an “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 \( P_1, P_2, \ldots, P_n \) on the paths in \( X^{m \oplus Y^m} \) followed by \( P_{m+1}, P_{m+2}, \ldots, P_n \) on the paths appearing in \( X^a \oplus Y^a \). The canonical path from \( X \) to \( Y \) involves “unwinding” each of the \( P_i \) in turn as follows. We need to consider three cases:

(i) \( P_i \) is a path in \( X^a \oplus Y^a \). \( P_i \) contains three vertices since each target is connected to a single model. Let \( P_i \) consist of the sequence \((w_1, k, u_2)\), perform the switch move replacing \((w_1, k)\) with \((w_2, k)\).

(ii) \( P_i \) is a path in \( X^a \oplus Y^a \) and \( P_i \) is not a cycle. Let \( P_i \) consist of the sequence \((v_0, v_1, \ldots, v_l)\) of vertices with a start vertex \( v_0 \). If \( (v_0, v_1) \in Y \), perform a sequence of switching moves replacing \((v_{j+1}, v_{j+2})\) by \((v_{j+2}, v_{j+1})\) 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 \((v_1, \ldots, v_l)\).

(iii) \( P_i \) is a path in \( X^a \oplus Y^a \) and \( P_i \) is a cycle. Let \( P_i \) consist of the sequence \((v_0, v_1, \ldots, v_{l+1})\) of vertices, for \( l \geq 1 \), where \( v_0 \) is the start vertex, and \((v_{j+1}, v_{j+2})\) for \( j = 0, 1, \ldots, l \), with remaining edges belonging to \( Y \). We 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 \( v_0 \). For \( k \in \{0, 1\} \) then we unwound \( O \) as in (i) above but treating \( v_{1-k} \) as the start vertex to identify that it was a cycle.

Let \( q \) be an arbitrary edge in the Markov chain \( M \), i.e. a transition from \( \omega \) to \( \omega' \neq \omega \). Let \( cp(q) = \{(X, Y) : \gamma_{XY} \ni \omega \} \) be the set of canonical paths that use \( q \). We define a function \( \eta_{\omega} : cp(q) \rightarrow \Omega \) as follows:

\[ \eta_{\omega}(X, Y) = \left\{ \begin{array}{ll} X \oplus Y & \text{if } \omega \text{ is a model switch move;} \\
X \oplus Y \oplus (\omega \cup \omega') - e_{XY} & \text{if } q \text{ is a switch move and the current path is a cycle;} \\
X \oplus Y \oplus \omega & \text{otherwise}, \end{array} \right. \]

where \( e_{XY} \) is the edge in \( X \) adjacent to the start vertex that was removed first in (iii) above. Notice that \( \eta_{\omega}(X, Y) \in \Omega \) and an injective function. Since \( |F| = MK \),

\[ Q(q) = Q(\omega, \omega') = \pi(\omega)P(\omega, \omega') \]

Next, we bound \( \pi(X \cup \gamma(Y)) \) and we need to consider five cases:

(i) \( q \) is a model switch move. We have \( \omega' = \omega + e - e' \) and \( \omega \cup \eta_{\omega}(X, Y) \) and \( X \cup Y \) are identical when viewed as multisets.

(ii) \( q \) is a deletion move. We have \( \omega' = \omega - e \) and \( \eta_{\omega}(X, Y) = X \oplus Y \oplus (\omega \cup \omega') \). Since \( \omega \cup \eta_{\omega}(X, Y) \) and \( X \cup Y \) are identical when viewed as multisets,

(iii) \( q \) is an addition move. We have \( \omega' = \omega + e \) and \( \eta_{\omega}(X, Y) = X \oplus Y \oplus (\omega \cup \omega') \). Since \( \omega \cup \eta_{\omega}(X, Y) \) and \( X \cup Y \) are identical when viewed as multisets using the arguments from (i),

(iv) \( q \) is a switch move and the current path is a cycle. Suppose \( \omega' = \omega + e - e' \) and \( \omega' = \omega - e' \). Then \( \pi(\omega) \leq q(\omega, \omega') \) by Lemma 1, \( \pi(\omega) \leq |F| \leq R^2 \). Since \( \eta_{\omega}(X, Y) = X \oplus Y \oplus (\omega \cup \omega') - e_{XY} \), the multisets \( \omega \cup \eta_{\omega}(X, Y) \)
differs from $X \cup Y$ only in that $e$ and $e_{XY}$ are missing from it. Hence, by Lemma 1,

$$\pi(x)\pi(y) \leq C^2 \pi(\omega)\pi(\eta_t(X,Y))$$

where the second inequality follows from the fact that the length of $\eta$ is a probability distribution, and the last inequality follows from $E \leq KN$. Hence, $\bar{\rho} \leq 6R^4 K^2 (N + M)$.

Proof of Theorem 2

$M$ 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

$$\pi(x)\pi(y) \leq C^2 \pi(\omega)\pi(\eta_t(X,Y))$$

Thus, for any transition $q$,

$$\frac{1}{\sqrt{90}} \sum_{\pi(X,Y) \neq 0} \pi(x)\pi(y) |\gamma_{XY}|$$

$$\leq 2 \pi(E) + (M + 1)K \rho \max \left\{ 1, \frac{\rho(\omega)}{\phi(\omega)} \right\} \pi(\eta_t(X,Y))$$

In summary, we have

$$\pi(x)\pi(y) \leq 2R^4 (|E| + (M + 1)K) \rho \max \left\{ 1, \frac{\rho(\omega)}{\phi(\omega)} \right\} \pi(\eta_t(X,Y)).$$

Hence, $\log \frac{1}{\pi(\omega)} \leq \log \left( \frac{m_2}{m_2} \right) ^K \frac{m_3(K,N)}{m_4(K,N)} (K + 1)!N!$.

Putting all together, we have $\tau_e(\epsilon) \leq 6R^4 K^2 (N + M) (m_3(K,N) + \log e)$ for all initial state $x \in \Omega$.

REFERENCES


