Overcomplete Radon Bases for Target Property Management in Sensor Networks

Xiaoye Jiang
Stanford University
Stanford, CA, 94305
xiaoyej@stanford.edu

Mo Li
Nanyang Tech University
Singapore, 639798
limo@ntu.edu.sg

Yuan Yao
Peking University
Beijing, China, 100871
yuany@math.pku.edu.cn

Leonidas J. Guibas
Stanford University
Stanford, CA, 94305
guibas@cs.stanford.edu

ABSTRACT

This paper presents a scalable algorithm for managing property information about moving objects tracked by a sensor network. Property information is obtained via distributed sensor observations, but will be corrupted when objects mix up with each other. The association between properties and objects then becomes ambiguous. We build a novel representation framework, exploiting an overcomplete Radon basis dictionary to model property uncertainty in such circumstances. By making use of the combinatorial structure of the basis design and sparse representations we can efficiently approximate the underlying probability distribution of the association between target properties and tracks, overcoming the exponential space that would otherwise be required. We conduct comparative simulations and the results validate the effectiveness of our approach.

Categories and Subject Descriptors

C.2 [**Computer Communication Networks**]: Distributed Systems; G.2.1 [**Discrete Mathematics**]: Combinatorics—*Permutations and Combinations*

General Terms

Theory, Algorithms

Keywords

Wireless Sensor Networks, Property Management, Homogeneous Spaces

IPSN'11, April 12–14, 2011, Chicago, Illinois.

1. INTRODUCTION

In this paper, we address the *property management problem*, in wireless sensor networks, which is related to the *identity management problem* [9]. Both problems arise in the context of accurately tracking and identifying multiple moving targets with distributed sensors in the field.

We first give a brief review of the *identity management problem*. Identity management for moving targets in the sensor networks was first introduced by Shin et al [24]. Suppose we have a sensor network which is tasked at tracking multiple, simultaneously moving targets in the monitored area. In such a setting, we assume that positions of the targets can be instantly tracked by the sensor nodes and we focus on distinguishing their identities during the tracking. When targets are well separated and good quality observations are obtained by the sensors, the problem factorizes nicely. Different sensor nodes can focus on different targets, forming collaboration groups to best determine target identities. The problem becomes more complicated, however, when two targets come close to each other, which leads to confusion as the signal signatures of two targets mix up. After the two targets separate again, their positions may become immediately distinguishable, but their identities can still be confused, and the sensors may no longer be able to tell who is who.

Such uncertainties about identities will be carried forward in time with each target, until good quality observations on their identities are obtained to allow disambiguation. How to achieve accurate and efficient disambiguation is subtle, e.g., when the identity of target A becomes clear due to a new observation from a sensor close to A, another target B which A has mixed up with earlier becomes unambiguous as well, see figure 1-(a). Thus when there are many moving targets with mixed trajectories, it becomes increasingly complicated for the sensor network to resolve such ambiguities globally. Such a problem is called the identity management problem and the major task in addressing the problem is to maintain a belief state for the correct association between target tracks and target identities with continuous input of target mixing events and updated identity observations from

sensors. The identity management problem poses a challenge for probabilistic inference as it needs to address the fundamental combinatorial challenge that there are a factorial number of possible associations to maintain between tracks and identities. There have been many works proposed to address the identity management problem.

Property management problem is another interesting problem related to the identity management problem. We note that in many cases, we do not need to distinguish the identity of every individual target or the sensors are not powerful enough to capture all target features for identification. Instead we may only get coarse property information on targets and such information will suffice for many applications. For example, when we track troops of different parties on the battlefield so as to infer which party the individuals are affiliated to, it would suffice that we correctly track their affiliations rather than their identities. Here their affiliations are regarded as properties associated with the targets. As a matter of fact, tracking such properties associated with targets is a more natural assumption for sensor networks because what sensors directly sense is property information rather than identity features. While there have been many efforts put in studying the identity management problem, however, few studies have been done for the property management problem in sensor networks.

In this paper, we conduct the first study on such a problem, where we focus on target properties rather than their identities. For simplicity of exposition, we restrict our attention to the basic case where the targets can be classified into two categories, red or blue. Similar to the identity management problem, confusion will arise when a red target and a blue target mix up with each other and then depart. Future property observations by sensors can help to resolve such ambiguities. The property management problem is to *maintain a belief state for the correct association between target tracks and target properties (red or blue) with target mixing events and updated property observations.*

The property management problem is closely related to the identity management problem — we can reveal target identities if we have enough target properties to differentiate the identities. Typically, $\mathcal{O}(\log n)$ target properties are adequate to completely identify an object out of n. Thus, property management provides us an alternative method to study the identity management problem, while the identity management solutions cannot be applied for property management. The property management problem has to be addressed separately, as multiple targets may share the same property and the permutation machinery assumption that each track corresponds to a unique identity no longer applies.

We summarize the contributions of our paper as follows:

1. To the best of our knowledge, this is the first work to address the property management problem with sensor networks.
2. We use novel overcomplete bases together with sparse approximation algorithms to represent uncertainties to achieve high accuracy as well as low computational overhead.

The paper is organized as follows: In Section 2, we introduce some related works. In Section 3, we use a Markov model to formulate the property management problem. A novel overcomplete Radon basis dictionary for representing uncertainties over homogeneous spaces is introduced in Section 4. In Section 5, scalable algorithms are provided based on the proposed framework. We conduct comparative simulations to validate this approach in Section 6. In Section 7, we conclude this paper.

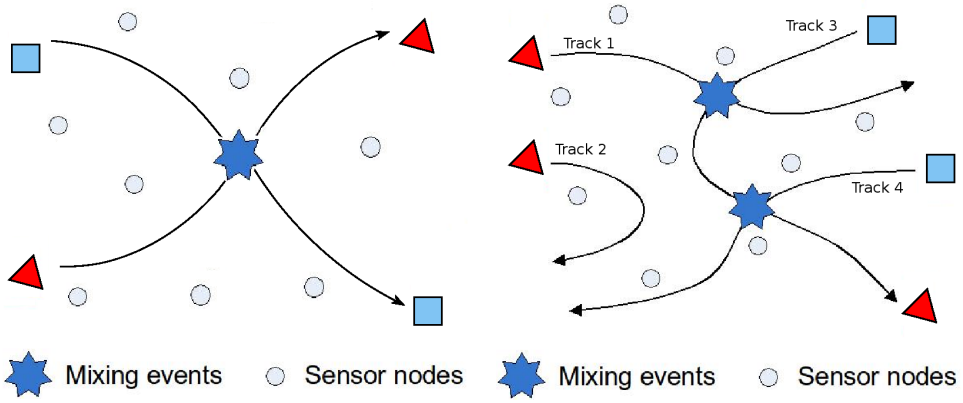
Table 1: Localized Mixing and Observation Data

Event #	Event Type
1	Tracks 1 and 3 mixed
2	Tracks 3 and 4 mixed
3	Observed target on track 3 is red

2. RELATED WORKS

The key computational challenge in the identity management problem is that the number of possible associations between tracks and target identities can be very large. To address such a problem, [24] uses the *belief matrix* to approximate the association probabilities, which collapses the factorial distribution to its first-order marginals (marginal probability that identity i is associated with track j). An alternative representation [22] is using an *information matrix* whose elements represent marginal log-likelihoods. Both methods provide efficient and scalable algorithms yet fail to characterize higher order marginals, such as the association probabilities between pairs of tracks and pairs of identities. The marginals of different orders are interconnected, thus the formulation becomes quickly unmanageable. Fortunately, there is an established mathematical theory that ideally suits to disentangling all the information: the *representation theory of permutation group* [6, 21, 23]. It turns out that one can define Fourier transforms for functions over all *permutations*, and low (high) order Fourier coefficients contain information about low (high) order marginals. [14] uses a general set of Fourier coefficients to represent uncertainty over permutations and demonstrates improvements against only using low order Fourier coefficients. Recently, [11] proposes an algorithm, called *Kronecker Conditioning*, which performs all probabilistic inference operations completely in the Fourier domain. Such a method can address any mixing or observation model and gains efficiency by truncating the Fourier expansions, allowing for a principled tradeoff between computational complexity and approximation accuracy. Though polynomial, the Fourier methods are still quite computational demanding when the number of targets is of even modest size. One way to mitigate the overhead is to factorize the problem into smaller clusters, so that highly certain individual or group associations can be pulled out of a global Fourier representation and represented compactly [12].

The identity management problem is not identical with the classical data association problem of maintaining correspondences between tracks and observations. In the identity management problem, the rate at which observations happen that are informative about target identities is not coupled to the rate of observations about target positions



(a) (b)

Figure 1: Targets move in a sensor network. (a) Identities of targets can get confused when they pass each other. (b) Four red or blue targets (denoted by red triangles and blue squares) moving around. We get confused about colors when tracks 1 crosses with track 3, then track 3 crosses with track 4. In the end, we observe target on track 3 is red.

and can be much lower. We note that a vast literatures already exist on the the data association problem, beginning with the multiple hypothesis testing approach (MHT) of Reid [20]. MHT is a deferred logic method in which past observations are exploited in forming new hypotheses when a new set of observations arises. Since the number of hypotheses can grow exponentially over time, various heuristics have been proposed to help cope with the complexity. For example, one can choose to maintain only the k best hypotheses for some parameter k [5], using Murty's algorithm [16]. But for the approximation to be effective, k still has to be exponential in the number of targets. A slightly more recent filtering approach is the joint probabilistic data association filter (JPDA) [3], which is a suboptimal single-stage approximation of the optimal Bayesian filter. JPDA makes associations sequentially and is unable to correct erroneous associations made in the past [19]. Even though it is more efficient than MHT, the calculation of the JPDA association probabilities is still an NP-hard problem [4]. Polynomial approximation algorithms to the JPDA association probabilities have recently been studied using Markov chain Monte Carlo (MCMC) methods [17, 18]. Generalized sensor models, e.g., binary sensors [2] which can only tell whether the target is moving toward the sensor or away from the sensor, as well as the related localizability problems [26] have also been considered in the literature. However, none of those proposed approaches can be used to address the property management problem. To the best of our knowledge, there is neither theoretical study on the problem itself nor algorithmic efforts in making distributed solutions in sensor networks.

Modern computer technologies have made it possible for us to deploy densely distributed sensor network systems. Such systems can hold up to hundreds of sensor nodes, which can perform lots of sensing and controlling tasks such as multi-target tracking [7], intrusion detection [1], ecosystem surveillance [15] and etc. The classical problem of reliably tracking also connects the vision community if one considers the camera network for object detection and recognition [25].

Given that the tracking literature is becoming mature, however, we note that the property management problem still needs to be addressed. In this problem, we don't assume a dense in time and accurate in space measurements of the target positions which are typically assumed in the tracking literature because those dense measurements are quite expensive to acquire. We also note that in a sensor network, often different types of sensors convey property information than positional information, and typically the former are more expensive. For example, we can imagine a network of simple proximity sensor that can be used to detect the presence of targets (and therefore provide information about locations), but which cannot differentiate the mobile targets from each other. These inexpensive sensors then can be augmented with a network of sparse but expensive camera sensors that can observe other properties of targets, such as colors, which help in differentiating or identifying the targets. The latter observations occur much less frequently, however. Thus, such a problem setup deserves the research attention from a theoretical perspective.

3. FORMULATION

We start with a simple tracking problem with four target tracks. As depicted in figure 1-(b), four targets, where two are red and the other two are blue, are moving within a field deployed with sensors. The sensors are capable of sensing target properties (red or blue). As we have mentioned earlier, the property management problem requires to maintain a belief state for the correct association between target tracks and target properties (red or blue), without distinguishing among red (blue) targets.

In this particular example, when the four targets are moving within the field, local sensors may report two types of events, namely, *mixing events* — two tracks get mixed when the targets get too close to each other; *observation events* — the target property on a particular track is clearly observed by a local sensor. Hence, a stream of localized data

is observed about the four tracks, which is recorded in table 1.

Assume initial colors of the targets are known. Then from table 1, we know track 2 never mixes with other tracks. Observing red target on track 3 will clarify all the ambiguities, e.g., targets on tracks $\{2, 3\}$ are red and targets on tracks $\{1, 4\}$ are blue. Such a simple example illustrates the combinatorial nature of the property management problem — reasoning on the mixing events allows us to determine which targets move along which tracks even though we only have partial observations on the tracks.

In the following, we introduce a Markov model to formulate the property management problem. The Markov model is constructed over homogeneous spaces which will be formally defined. Based on that, we provide probabilistic models for mixing and observation events.

3.1 The Markov Model

Consider we have n targets, k of them are red and $n-k$ of them are blue. We consider all possible $k-$subsets of the set $\{1, 2, \cdots, n\}$, so that each $k-$subset characterize a state indicating which k tracks have the red targets. We introduce a *Markov model* to model the uncertainty, which is represented by a probability distribution f over all $k-$subsets $x^{(t)}$ at time $t = 0, 1, 2, \cdots$. Such a distribution encodes the probability of an arbitrary $k-$subset of the tracks being red at time t. As figure 2 illustrates, we will update the distribution f over $x^{(t)}$ at each time step.

In figure 2, to model the conditional probability distribution $P(x^{(t)}|x^{(t-1)})$, we will work on a *mixing model* so that the mixing model reflects, for example, that the targets belonging to two tracks are swapped with some probability at a mixing event. To model the distribution $L(z^{(t)}|x^{(t)})$, we will work on an *observation model*, which captures the likelihood of observation $z^{(t)}$, given that targets on a $k-$subset of tracks $x^{(t)}$ are all red.

We focus on *filtering*, where one queries the Markov model for posterior at each time step, based on all past observations. Given distribution $f(x^{(t)}|z^{(0)}, \cdots, z^{(t)})$, we recursively compute $f(x^{(t+1)}|z^{(0)}, \cdots, z^{(t+1)})$ with two steps: a *rollup step* and a *conditioning step*. The rollup step multiplies the distribution in the mixing model and the distribution of the previous step:

$$f(x^{(t+1)}|z^{(0)}, \cdots, z^{(t)}) = \sum_{x^{(t)}} P(x^{(t+1)}|x^{(t)}) f(x^{(t)}|z^{(0)}, \cdots, z^{(t)}).$$

The conditioning step conditions the distribution on an observation $z^{(t+1)}$ using Bayes rule:

$$f(x^{(t+1)}|z^{(0)}, \cdots, z^{(t+1)}) \propto L(z^{(t+1)}|x^{(t+1)}) f(x^{(t+1)}|z^{(0)}, \cdots, z^{(t)}).$$

Since the space of all $k-$subsets of an $n-$set is of size $\binom{n}{k}$, a single iteration of the algorithm requires at least $\mathcal{O}(\binom{n}{k})$ operations, which is not polynomial with k. As will be detailed later, the approach that we advocate is to use a novel representation of distributions over all $k-$subsets, through which we can always maintain a compact representation for an arbitrary distribution. We will also present scalable algorithms such that updating distributions under the rollup and conditioning steps in such a representation framework requires polynomial complexity in both n and k.

The unique feature of this problem is that there is inherent dependencies among the property values, where the joint probability can not be factorized as products of marginal probabilities. Many typical compact representations, such as graphical models, cannot capture the inherent mutual exclusivity constraints associated with the problem where k targets are red reflect that the other $n-k$ targets are blue.

As final remarks of the Markov model, we assume throughout this paper, k is known as a priori. This is true in certain cases, e.g., tracking football players where for each player we only know the color information. It turns out that the general case where k is now known as a priori can be dealt with the same technical machinery proposed in this paper. We will come back to this point in the discussion section. We also focus the case where there are two possible colors associated with the targets in this paper. It turns out that for a more general case where there are multiple colors, the mathematics for characterizing the mutual exclusivity constraint will be much harder.

3.2 Permutation Groups and Homogeneous Spaces

In this section, we formally introduce *permutation groups* and *homogeneous spaces* which are used in modeling our problem. In mathematics, the set of all $k-$subsets of an $n-$set is known as a *homogeneous space*. Such a homogeneous space is associated with the *permutation group*, which acts on the homogeneous space in a transitive way.

Definition 1 A permutation on n elements is a one-to-one mapping of the set $\{1, \cdots, n\}$ onto itself and can be written as a tuple, $\sigma = [\sigma(1), \sigma(2), \cdots, \sigma(n)]$, where $\sigma(i)$ denotes where the i-th element is mapped with the permutation. The set of all permutations on n elements forms the *permutation group* S_n under the operation of function composition. We sometimes notate $\sigma = (i, j)$ which denote a swap of i with j.

Definition 2 The collection of all $k-$subsets of $\{1, 2, \cdots, n\}$ is a *homogeneous space*, denoted by X^k. The permutation group S_n acts on X^k in the following way: suppose $\sigma \in S_n$ and $x = \{x_1, x_2, \cdots, x_k\} \in X^k$, then $\sigma x = \sigma\{x_1, x_2, \cdots, x_k\} = \{\sigma(x_1), \sigma(x_2), \cdots, \sigma(x_k)\}$. It is easy to verify that S_n acts transitively on X^k(any $x, y \in X^k$ there exists a σ such that $\sigma x = y$).

A permutation acting on a homogeneous space models the process of how tracks followed by targets might be mixed or swapped, when the two targets approach or cross each other. It turns out that homogeneous spaces are natural for describing the Markov states of our system, and there is abundant mathematical structure that we can make use of to facilitate our formulation. Recently, [13] invented an approach based on the homogeneous space to study the clique detection problem in social networks. Although homogeneous spaces are well-studied objects in mathematics, to the best of our knowledge, they have not been used before to model

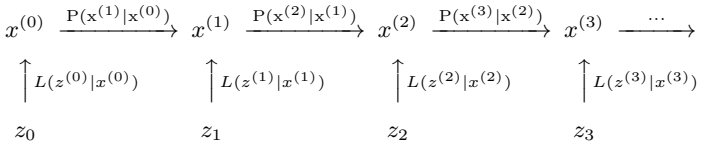


Figure 2: The Markov model for the property management problem.

such association problems.

Example 1 *In the previous example depicted in figure 1-(b), we consider the homogeneous space — all 2−subsets of $\{1,2,3,4\}$ (denoting the four tracks). At the beginning, targets on tracks $\{1,2\}$ are red. If tracks 1 and 3 swapped the targets, then targets on tracks $\{2,3\}$ will be red. Using mathematical terms, it can be stated in the following way: the permutation $(1,3)$ act on on $\{1,2\}$ (an element in X^2) will be $\{2,3\}$.*

3.3 Mixing Model

In this paper, we consider a particular class of probabilistic mixing models — that of random walks over the permutation group, which assumes that $x^{(t+1)}$ is generated from $x^{(t)}$ by drawing a random permutation $\sigma^{(t)}$ from some distribution $Q^{(t)}$ over the permutation group S_n. With such a probabilistic mixing model, we can write the rollup operation as a Markov transition matrix times the prior distribution. In our problem, $\sigma^{(t)} \in S_n$ represents a random permutation that might occur among tracks when they get too close to each other. As we have introduced in the previous section, a permutation $\sigma^{(t)}$ acts on a state $x^{(t)}$ in the homogeneous space X^k as $\sigma^{(t)}x^{(t)}$. Hence, the distribution over $x^{(t+1)}$ generated from $x^{(t)}$ by a random draw from the distribution Q over S_n is:

$$\begin{aligned} & f(x^{(t+1)}|z^{(0)},\cdots,z^{(t)}) = \sum_{x^{(t)}} P(x^{(t+1)}|x^{(t)}) f(x^{(t)}|z^{(0)},\cdots,z^{(t)}) \\ = & \sum_{(x^{(t)},\sigma):\ \sigma x^{(t)}=x^{(t+1)}} Q(\sigma) f(x^{(t)}|z^{(0)},\cdots,z^{(t)}) \\ = & \ T(x^{(t)},x^{(t+1)}) f(x^{(t)}|z^{(0)},\cdots,z^{(t)}), \end{aligned}$$

where $T(x,y) = \sum_{\sigma:\sigma x=y} Q(\sigma)$, meaning that all $Q(\sigma)$ such that $\sigma x = y$ will contribute to the $(x,y)-$entry of the transition matrix. In addition, we have the following theorem:

Theorem 1 *Let Q be a probability distribution on S_n, then Q induces a doubly stochastic Markov transition matrix for X^k with transitions: $T(x,y) = \sum_{\sigma:\sigma x=y} Q(\sigma)$.*

The above theorem gives an explicit formula for transition matrices of a distribution over the homogeneous space. As we will see later, transition matrices induced from the distribution Q also interact nicely with the homogeneous space, which can be utilized to simplify the computation of the rollup step.

In this paper, we consider the simplest probabilistic mixing model which assumes that with probability p, nothing happens to the two targets, and with probability $(1-p)$, the targets for tracks i and j are swapped (similar models are considered in [11]). With no priori knowledge about whether two targets will swap, we can simply take p to be $1/2$. If more information, e.g., moving speed, direction or other behavior of the targets, are available, then we can have a better estimate of how likely the two targets may swap. In any case, the probability distribution Q over S_n for this probabilistic mixing model is therefore:

$$Q(\sigma) = \begin{cases} p & \text{if } \sigma = id \\ 1-p & \text{if } \sigma = (i,j) \\ 0 & \text{otherwise} \end{cases}.$$

We note that there are special structures that we can explore in the transition matrix induced from the particular distribution Q over S_n, i.e., each row or column has either one nonzero entry (which must be 1) or two nonzero entries (which must be p and $1-p$). We will use such a fact to do the rollup operation in an efficient way.

Example 2 *We run the mixing update routines on the first two mixing events of the example in figure 1-(b). For each mixing event, we assume two tracks i and j swap targets with equal probability. Using the probabilistic mixing model we obtain distributions $f(x^{(t)})$ for $t = 0,1,2$ as shown in table 2. Here, $f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)})$ is abbreviated as $f(x^{(t)})$.*

3.4 Observation Model

In contrast to the rollup step, the conditioning step can potentially decrease uncertainty. We use Bayes rules to find the posterior distribution $P(x^{(t)}|z^{(t)})$ after observing some evidence $z^{(t)}$, which can be expressed as the following:

$$f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)}) = \frac{L(z^{(t)}|x^{(t)}) f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)})}{\sum_{x^{(t)}} L(z^{(t)}|x^{(t)}) f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)})}.$$

it requires two steps to compute the posterior — a *pointwise product* of prior $f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)})$ and likelihood $L(z^{(t)}|x^{(t)})$, followed by normalization step, which is computing $\sum_{x^{(t)}} L(z^{(t)}|x^{(t)}) f(x^{(t)}|z^{(0)},z^{(1)},\cdots,z^{(t)})$.

The simplest observation model assumes that we get observation z of the form: "see red on track i" (similar models are considered in [11, 14]). Now we assume all red (blue) targets have the same color histograms; sensors sense properties by a random draw from the color histogram of the target. If, for example, all red targets have 80% of red, 10% of blue and 10% of other colors (yellow, grey, etc.) while all blue targets have 70% of blue, 20% of red and 10% of other colors. Then the likelihood function for observation event z='see red on track i' given 'targets on $k-$subset tracks x are red' is:

- If $i \in x$, $L(z|x) = L(z|x \text{ are red}) = L(z|i \text{ is red}) = .8$.

Table 2: Updated Priors when Mixing Happens

	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{2,4\}$	$\{3,4\}$
$f(x^{(0)})$	1	0	0	0	0	0
$f(x^{(1)})$	1/2	0	0	1/2	0	0
$f(x^{(2)})$	1/2	0	0	1/4	1/4	0

• If $i \notin x$, $L(z|x) = L(z|x \text{ are red}) = L(z|x^c \text{ are blue}) = L(z|i \text{ is blue}) = .2$.
So we have

$$L(z|x) = \begin{cases} .8 & \text{if } i \in x \\ .2 & \text{if } i \notin x \end{cases}.$$

We conclude this section by pointing out that both the rollup step and conditioning step are of complexity at least $\mathcal{O}(\binom{n}{k})$, if we explicitly form the distribution on the homogeneous space. In the next section, we will explore the math structure of homogeneous spaces to overcome such a complexity issue.

4. EFFICIENT REPRESENTATION

In this section, we propose a novel representation framework to characterize distributions over homogeneous spaces.

4.1 Hierarchical Radon Basis

Let $L(X^k)$ be the set of all functions on the homogeneous space X^k. There is a special technique for decomposing $L(X^k)$. For each $0 \leq j \leq k$, define a matrix $\Phi^{j,k}$ as an $\binom{n}{k}$ by $\binom{n}{j}$ matrix where each row represents a $k-$subset and each column represents a $j-$subset. The entries in $\Phi^{j,k}$ are binary, indicating whether the $j-$subset is a subset of the $k-$subset. The matrix $\Phi^{j,k}$ can be interpreted as a mapping from functions on all $j-$subsets to functions on all $k-$subsets. The columns of the matrix $\Phi^{j,k}$ span an $\binom{n}{j}$ dimensional subspace of $L(X^k)$, which are called the *Radon bases* of order j. With a Radon basis we can represent a distribution over $k-$subsets by using coefficients on $j-$subsets, where $0 \leq j \leq k$.

In section 3.3, we noticed that a probability distribution Q over S_n can induce an action on a distribution over all $k-$subsets, i.e., Q induces a transition matrix for updating distributions over all $k-$subsets. However, there is nothing special about k when we define the transition matrix there. Thus we can generally let Q induce a transition matrix for updating distributions over all $j-$subsets, where $0 \leq j \leq k$. Here the transition matrix is $T^j(x,y) = \sum_{\sigma:\ \sigma x = y} Q(\sigma)$ where x, y are two $j-$subsets. The following proposition summarizes important properties regarding $\Phi^{j,k}(0 \leq j \leq k)$ and the transition matrices induced from Q.

Theorem 2 *Let Q be a distribution over S_n, we have the following three propositions:*

1. $R(\Phi^{0,k}) \subset R(\Phi^{1,k}) \subset \cdots \subset R(\Phi^{k,k})$ where $R(\cdot)$ denotes the range of a matrix.

2. Each subspace $R(\Phi^{j,k})$ is invariant under the action of the distribution Q.

3. The matrix $\Phi^{j,k}$ commutes with the action of the distribution Q, i.e., $\Phi^{j,k}T^j = T^k\Phi^{j,k}$ where T^j, T^k are transition matrices induced from the same distribution Q.

The first proposition states that the range of matrices $\Phi^{j,k}$ forms a hierarchical decomposition of $L(X^k)$. Bases that span $R(\Phi^{j,k})$ of small (large) j are efficient for approximating smooth (peaky) distributions over X^k respectively. We note that such a hierarchical basis design derives from the representation theory of permutation groups in such a way that each subspace is invariant under group actions [6]. The second proposition tells us that if $f \in L(X^k)$ lies in the range of $\Phi^{j,k}$ for some $j \leq k$, then after the rollup operation, the updated distribution still lies in the range of $\Phi^{j,k}$. The third proposition suggests us an efficient algorithm to update f when $f \in R(\Phi^{j,k})$. Suppose we have a distribution over $k-$subsets $f = \Phi^{j,k}c_j$ where c_j are coefficients on $j-$subsets, then because of this commutative property, the rollup operation for f can be performed simply by computing rollup operation for coefficients c_j:

$$T^k f = T^k\Phi^{j,k}c_j = \Phi^{j,k}T^j c_j = \Phi^{j,k}(T^j c_j).$$

One interesting fact about the Radon basis matrix $\Phi^{j,k}$ is that it has a pseudo-inverse $(\Phi^{j,k})^+$ which maps from functions on $k-$subsets to functions on $j-$subsets. $\Phi^{j,k}$ and $(\Phi^{j,k})^+$ satisfies that $(\Phi^{j,k})^+\Phi^{j,k}$ is identity and $\Phi^{j,k}(\Phi^{j,k})^+$ is an orthogonal projection. Given Radon bases consisting of delta functions on all $j-$subsets and $k-$subsets respectively, the (r,s) element of $(\Phi^{j,k})^+$ is $\frac{(-1)^{k-j}(k-j)}{(-1)^{|s-r|}|s-r|\binom{n-j}{|s-r|}}$, where $s-r$ means the set difference, i.e., $s-r = s \cap r^c$.

4.2 Overcomplete Basis Representation

Recent approaches [11, 14] in modeling the identity management problem keep compact representation of distributions over permutation groups by storing only low order Fourier coefficients. Clearly, similar ideas can be adopted here — we can use coefficients for low order Radon bases ($\Phi^{j,k}$ with small j) to represent a distribution over the homogeneous space X^k. Using a low order Radon basis, however, fails to characterize highly certain cases, e.g., a delta function on X^k which can be characterized by a single basis in $\Phi^{k,k}$, while a low order Radon basis is incapable of representing accurately such a peaky distribution; on the other hand, high order Radon bases are not efficient for representing smooth distributions over X^k, e.g., a constant function can be characterized by a single basis in $\Phi^{0,k}$, while one cannot have a compact representation for such a smooth function by using high order bases. Similar problems happen in the identity management problem, where a low order Fourier basis fails to represent peaky functions on permutation groups while high order Fourier bases are not efficient to represent smooth functions. In reaction to that, we propose to use an overcomplete Radon basis dictionary to represent distributions over homogeneous space X^k where we concate-

nate all $\Phi^{j,k}$'s, i.e.,

$$f = \Phi^{0,k}c_0 + \Phi^{1,k}c_1 + \cdots + \Phi^{k,k}c_k = \sum_{j=0}^{k} \Phi^{j,k}c_j,$$

where c_j are coefficients on $j-$subsets. By using a hierarchical overcomplete Radon basis, we will hopefully have sparse representation for any distribution over X^k.

Example 3 *In the example in figure 1-(b), the distribution f at $t = 2$ can be represented with coefficients $c_0 = 0$, $c_1 = [0, 1/4, 0, 0]^T$, and $c_2 = [1/4, 0, 0, 0, 0, 0]^T$ to f as the second column of $\Phi^{1,2}$ indicates three $2-$subsets which contain 2.*

5. ALGORITHM DESIGN

In this section, we design algorithms based on aforementioned theorems for updating probabilistic distributions over the homogeneous space X^k. We assume that, using overcomplete bases, sparse representations are available for the distributions over X^k which we work with. By making use of the combinatorial structure of the basis matrices and our sparse representation assumptions, we obtain efficient algorithms for updating distributions over X^k polynomial in n and k, which will be detailed in section 5.1 and section 5.2. When we keep updating the Radon basis coefficients, however, we may gradually lose sparsity. To resolve such an issue, we propose in section 5.3 sparse approximation algorithms to re-organize the coefficients and regain sparsity.

5.1 Algorithm for Rollup Step

Given a distribution f over homogeneous space X^k using overcomplete Radon bases, we have $f = \sum_{j=0}^{k} \Phi^{j,k}c_j$ where c_j are coefficients over $j-$subsets. If each $c_j (0 \le j \le k)$ is sparse, then *we can store f by only storing the nonzero values in each c_j and their corresponding column indices, which are $j-$subsets.* We can represent f in another way:

$$f = \sum_{\alpha} \Phi_\alpha c_\alpha,$$

where α is a subset of size at most k, and c_α is the coefficient for the basis column Φ_α.

We now describe the algorithm for updating Radon basis coefficients when mixing events happen. In particular, we consider probabilistic mixing models as described in section 3.3. When a mixing event happens, we need to perform an rollup operation: $f \leftarrow T^k f$ where T^k is a transition matrix for distribution f over all $k-$subsets, and T^k is induced from a probability Q on permutation group S_n. In section 4.1, we know that the basis matrices $\phi^{j,k}$ commute with the action of a distribution Q over S_n. Hence, we have

$$T^k f = T^k \sum_{j=0}^{k} \Phi^{j,k}c_j = \sum_{j=0}^{k} T^k \Phi^{j,k}c_j = \sum_{j=0}^{k} \Phi^{j,k}T^j c_j.$$

To update f, we only need to update coefficients c_j's as $c_j \leftarrow T^j c_j$, for $0 \le j \le k$.

When tracks t_i and t_j mix, according to the probabilistic mixing model in section 3.3, we have a distribution Q which takes nonzero values only on *id* and (t_i, t_j). Note that there is a special structure within the induced transition probability matrix — each column of the transition matrix T^j has either one nonzero entry (which is 1) or two nonzero entries (which are p and $1-p$). If we store f by using a collection of subsets α's with associated values c_α's, then we can efficiently get the result of the rollup operation. In cases where $t_i \in \alpha, t_j \in \alpha$, or $t_i \notin \alpha, t_j \notin \alpha$, $\Phi_\alpha c_\alpha$ will not be affected after updating, i.e., $T^k(\Phi_\alpha c_\alpha) = \Phi_\alpha c_\alpha$; in cases where $t_i \in \alpha, t_j \notin \alpha$, or $t_i \notin \alpha, t_j \in \alpha$, $\Phi_\alpha c_\alpha$ will be splitted into pc_α and $(1-p)c_\alpha$ on α and $\beta = \alpha\Delta\{t_i, t_j\}$, where Δ denotes the symmetric difference between two sets ($A\Delta B = (A \cup B) \cap (A \cap B)^c$), i.e., $T^k(\Phi_\alpha c_\alpha) = \Phi_\alpha(pc_\alpha) + \Phi_\beta(1-p)c_\alpha$.

Algorithm 1 Algorithm for Rollup Step

Input: A collection of subsets with associated values $I = \{(\alpha, c_\alpha) : c_\alpha \neq 0\}$
Output: A collection of subsets with associated values O
Procedure:
Initialize $O \leftarrow \{\}$
for each $(\alpha, c_\alpha) \in I$ **do**
  if $t_i \in \alpha, t_j \in \alpha$ or $t_i \notin \alpha, t_j \notin \alpha$ **then**
    $O \leftarrow O \cup \{(\alpha, c_\alpha)\}$
  else if $t_i \in \alpha, t_j \notin \alpha$ or $t_i \notin \alpha, t_j \in \alpha$ **then**
    $\beta \leftarrow \alpha\Delta\{t_i, t_j\}$
    Retrieve value v_α associated with α, if α exist in O; otherwise set $v_\alpha = 0$
    Retrieve value v_β associated with β, if β exist in O; otherwise set $v_\beta = 0$
    $O \leftarrow O \cup \{\alpha, v_\alpha + pc_\alpha\}$
    $O \leftarrow O \cup \{\beta, v_\beta + (1-p)c_\beta\}$
  end if
end for

We have the following theorem regarding algorithm 1.

Theorem 3 *Suppose $f = \sum_\alpha \Phi_\alpha c_\alpha$, N is the number of nonzero coefficients c_α's and k is the number of red targets. The rollup algorithm can generate output in $\mathcal{O}(kN \log N)$ computational time. The size of nonzero coefficients in the output is at most $2N$.*

Example 4 *In the example in figure 1-(b), at time $t = 1$ the distribution over $2-$subsets can be stored by*

$$I = \{(\{1, 2\}, 1/2), (\{2, 3\}, 1/2)\}.$$

After the mixing event happens between tracks 3 and 4 where we have a probability distribution Q over S_4:

$$Q(\sigma) = \begin{cases} 1/2 & \text{if } \sigma = id \\ 1/2 & \text{if } \sigma = (3,4) \\ 0 & \text{otherwise} \end{cases}.$$

The updated distribution can be stored as:

$$O = \{(\{1, 2\}, 1/2), (\{2, 3\}, 1/4), (\{2, 4\}, 1/4)\}.$$

5.2 Algorithm for Conditioning Step

Two computation phases are involved in the conditioning step. First, a pointwise product needs to be computed, and secondly, we need to compute the normalizing constant. Note that in the observation model in section 3.4, the likelihood function $L(z|x)$ are of the form:

$$L(z|x) = \begin{cases} a & \text{if } i \in x \\ b & \text{if } i \notin x \end{cases}.$$

Algorithm 2 Algorithm for Conditioning Step— Computing Posterior Distribution

Input: Two collection of subsets with associated values $I_1 = \{(\alpha, c_\alpha) : c_\alpha \neq 0\}, I_2 = \{(\beta, l_\beta) : l_\beta \neq 0\}$
Output: A collection of subsets with associated values O
Procedure:
for each $(\alpha, c_\alpha) \in I_1$ **do**
  for each $(\beta, l_\beta) \in I_2$ **do**
    if $|\alpha \cup \beta| \leq k$ **then**
      Retrieve value v_γ associated with γ, if γ exist in O; otherwise set $v_\gamma = 0$
      $\gamma \leftarrow \alpha \cup \beta$
      $v_\gamma \leftarrow v_\gamma + c_\alpha l_\beta$
      $O \leftarrow O \cup \{(\gamma, v_\gamma)\}$
    end if
  end for
end for
Compute normalizing constant $Z \leftarrow \sum_\gamma \binom{n-|\gamma|}{k-|\gamma|} c_\gamma$
Divide each c_γ in O by Z

Such a likelihood function $L(z|x)$ lies in the space of $R(\Phi^{1,k})$. If we define a function L_1 on $1-$subset as follows:

$$L_1(j) = \begin{cases} a - \frac{k-1}{k} b & \text{if } j = i \\ \frac{1}{k} b & \text{if } j \neq i \end{cases},$$

it is easy to verify that $L(z|x) = R^{1,k} L_1$. We can even express $L(z|x)$ in a more compact way: $L(z|x) = b\Phi_\emptyset + (a - b)\Phi_{\{i\}}$.

Given $f(x) = \sum_\alpha \Phi_\alpha c_\alpha$ and $L(z|x) = \sum_\beta \Phi_\beta c_\beta$ where β's are at most $1-$subsets, the pointwise product is

$$f(x) \cdot L(z|x) = (\sum_\alpha \Phi_\alpha c_\alpha) \cdot (\sum_\beta \Phi_\beta l_\beta) = \sum_{\alpha,\beta} c_\alpha l_\beta (\Phi_\alpha \cdot \Phi_\beta),$$

where the last equality is due to the distributive law for the pointwise product operation.

For the basis vector corresponding to α and the basis vector corresponding to β, the pointwise product between Φ_α and Φ_β can be estimated as:

$$\Phi_\alpha \cdot \Phi_\beta = \begin{cases} \Phi_{\alpha \cup \beta} & \text{if } |\alpha \cup \beta| \leq k \\ \mathbf{0} & \text{if } |\alpha \cup \beta| > k \end{cases}$$

The normalizing constant actually equals to the l_1 norm of $f(x) \cdot L(z|x)$. If we have $f(x) \cdot L(z|x) = \sum_\gamma \Phi_\gamma c_\gamma$, then $Z = \sum_\gamma |\Phi_\gamma|_1 c_\gamma = \sum_\gamma \binom{n-|\gamma|}{k-|\gamma|} c_\gamma$.

The algorithm for updating Radon basis coefficients when an observation event happens is summarized in algorithm 2, which can be used to deal with general likelihood functions. In the special case where the likelihood function can be compactly represented as a linear combination of $\Phi_\emptyset$ and $\Phi_{\{i\}}$, we have the following theorem:

Theorem 4 *Suppose $f = \sum_\alpha \Phi_\alpha c_\alpha$, N is the number of nonzero coefficients c_α's and k is the number of red targets. The conditioning algorithm can generate output in $\mathcal{O}(kN \log N)$ computational time. The size of nonzero coefficients in the output is at most $2N$.*

We remark that in the conditioning algorithm, the normalizing step is not essential because if we do not normalize the distribution, the result is still accurate up to a multiplication constant. We may even choose a special likelihood function $L(z|x) = b\Phi_\emptyset + (a - b)\Phi_{\{i\}}$ where $b = 1$ with the benefit that if $f = \sum_\alpha \Phi_\alpha c_\alpha$ is the prior distribution, then c_α will not change after the conditioning step when the reported target i does not belong to α.

Algorithm 3 Orthogonal Matching Pursuit Algorithm

Input: Basis dictionary Φ, distribution f, stopping criteria ϵ
Output: Residual r, coefficients x, indices Λ
Procedure:
while $\|r\| \geq \epsilon$ **do**
  Measure correlations $c \leftarrow \Phi^T r$
  $\Lambda \leftarrow \Lambda \cup \{\arg\max_j c(j)\}$
  $x \leftarrow \arg\min_{z:\text{supp}(z) \subset \Lambda} \|f - \Phi z\|_2$
  $r \leftarrow f - \Phi x$
end while

5.3 Sparse Approximation

We have developed algorithms for updating Radon basis coefficients at the rollup step and conditioning step. For the rollup step, coefficients only propagate within Radon bases of the same order while for the conditioning step coefficients may propagate to Radon bases of higher orders, i.e., the more observations we have, the more we will be certain about which $k-$subset has the red targets.

The one step rollup and conditioning algorithms are quite scalable. As we keep updating the Radon basis coefficients using the rollup and conditioning algorithms, however, we may need more and more coefficients to represent the distribution over all $k-$subsets (As revealed by theorem 3 and 4, we may need up to two times more coefficients to characterize the distribution). As a result, the number of coefficients used to represent the distribution may grow exponentially as we proceed with the rollup steps and conditioning steps.

To overcome the exponential growth of number of bases used in the representation, we develop an approximation algorithm to re-organize the Radon basis coefficients such that we can always keep compact representation of the distribution. This is possible since we used overcomplete Radon bases, which for any distribution, there are more than one way to characterize it. Thus we can search for a sparse approximation representation for the distribution if it is not represented in a compact way.

5.3.1 Orthogonal Matching Pursuit

We note that after a series of mixing events happen, distributions on the homogeneous space X^k become smoother and their energies gradually concentrate to subspaces spanned by low order Radon bases. More precisely, for any distribution $f \in L(X^k)$, if we consider the l_2 distance between f and the orthogonal projection of f to the subspace spanned by the columns in $\Phi^{j,k}$ $(0 \leq j \leq k)$, it is easy to prove that such a distance will decrease after each rollup step. For example, in the special case where $j = 0$, it reduces to the conclusion that the l_2 distance between f and the uniform distribution will decrease after each rollup step. In such sense, we should introduce lower order Radon bases to see if we can more efficiently represent the distribution f after a series of mixing events.

In [8], a greedy algorithm, *Orthogonal Matching Pursuit (OMP)* (see algorithm 3) is proposed to solve the sparse

approximation problem over redundant dictionaries, which works by greedily searching for bases most correlated with the residual and use them to fit the distribution. The Orthogonal Matching Pursuit algorithm also has better theoretical guarantees about quality of the approximation, given that the dictionary has smaller incoherent parameter [8]. In our case, we can use the entire overcomplete Radon basis dictionary as input basis dictionary to OMP, yet with a large incoherent parameter. However, the computational burden of searching a combinatorial size basis dictionary is unaffordable. We observe that in our probabilistic mixing model (which assumes that at each timestep only two targets may swap), targets will not get well mixed very quickly. For example, coefficients c_α on the set α will evenly spread its energy to other subsets of the same size only if there is a target i in α which well mixes with all targets in α^c, with the basis $\Phi_{\alpha-\{i\}}$ being efficient in representing the distribution after those mixing events.

As a result, we may adaptively downsample the whole overcomplete Radon basis dictionary according to current representation of f. Given $f = \sum_{\alpha\in I} \Phi_\alpha c_\alpha$, we downsample a subset of the basis dictionary, e.g., $\{\beta : \beta \subset \alpha, |\alpha - \beta| \leq s, \alpha \in I\}$ (i.e., the basis β which is a subset of some $\alpha \in I$ and β differs from α at most s elements). For example, if $s = 1$ and $\alpha = \{1, 2, \cdots, k\} \in I$, then all $(k-1)$-subset of α together with α itself are sampled as candidate bases for approximation. If we use N bases to represent f, then we will downsample at most kN bases. With fewer bases sampled, we also achieve a smaller incoherence parameter.

For the computational complexity of the Orthogonal Matching Pursuit algorithm in our case, we note that, given $f = \sum_\alpha \Phi_\alpha c_\alpha$, the inner product of a basis Φ_β in the basis dictionary Φ and f can be computed in polynomial time because $\langle \Phi_\beta, f\rangle = \sum_\alpha \langle \Phi_\beta, \Phi_\alpha\rangle c_\alpha$, where

$$\langle \Phi_\beta, \Phi_\alpha\rangle = \begin{cases} \binom{n-|\alpha\cup\beta|}{k-|\alpha\cup\beta|} & \text{if } |\alpha\cup\beta| \leq k \\ \mathbf{0} & \text{if } |\alpha\cup\beta| > k \end{cases}$$

Solving a least square problem can also be done in polynomial time because the least square solution is $(\Phi^T\Phi)^{-1}\Phi^T f$, where estimating $\Phi^T\Phi$ reduces to evaluating inner product of two bases in the basis dictionary and estimating $\Phi^T f$ reduces to evaluating the inner product between a basis in the basis dictionary and f. So we have the following theorem:

Theorem 5 *Suppose $f = \sum_\alpha \Phi_\alpha c_\alpha$, N is the number of nonzero coefficients c_α's and k is the number of red targets. If we downsample at most kN bases and use OMP algorithm to generate an m-term approximation solution, then the complexity of the OMP algorithm is $\mathcal{O}(k^2N(N+m)m + (k+m)m^3)$.*

Example 5 *In the example in figure 1-(b), we see that the distribution f can be represented as*

$$f = \Phi_{\{1,2\}} \cdot .5 + \Phi_{\{2,3\}} \cdot .25 + \Phi_{\{2,4\}} \cdot .25.$$

If we run OMP on f, the basis $\Phi^{\{2\}}$ and $\Phi^{\{1,2\}}$ can be identified which help us to re-organize coefficients for f as

$$f \approx \Phi_{\{2\}} \cdot .25 + \Phi_{\{1,2\}} \cdot .25.$$

In such a simple example, the approximation is highly accurate.

Figure 3: A view of the simulated data.

5.3.2 *Thresholding*

We may also choose to do thresholding on Radon basis coefficients to maintain sparse approximation to the distribution. Such a technique is especially useful after a sequence of observation events happen, since conditioning steps may result in exponential decay of Radon basis coefficients.

Given $f = \sum_\alpha \Phi_\alpha c_\alpha$, where c_α's are Radon basis coefficients. If we always insist on the positiveness of the Radon basis coefficients, then one can directly estimate the l_1 norm of $\Phi_\alpha c_\alpha$ as $\binom{n-|\alpha|}{k-|\alpha|}c_\alpha$. Based on l_1 norm contribution of $\Phi_\alpha c_\alpha$ to the distribution f, we can threshold off the insignificant bases to maintain a sparse approximation of f. Such an algorithm is quite scalable and in practice works well.

Finally we should note that OMP can help to represent the distribution on homogeneous spaces by lower order bases while thresholding do not have such a property.

5.4 Discussion

In this section, we give brief discussion on several issues related to the proposed algorithms.

• *Error Propogation:* Whenever we do sparse approximation to approximate f by f', we introduce errors. Mixing events always shrink the approximation error $\|f - f'\|$ which is due to the fact that mixing matrices have eigenvalues bounded by 1, while observation events do not. However, we note that n independent observations of all the targets will drive both f and f' to converge to delta distributions. In this sense, approximation errors can be under control.

• *Sparsity Propogation:* Theoretically one may need up to two times more coefficients to represent the updated distribution. In practice, if mixing and observation events happen locally, sparsity can always be kept to a relatively low level.

• *Timing for Sparsification:* In practice, we do sparsification after a sequence of mixing events happen or a sequence of observations happen such that the true distribution becomes more smooth or peaky while the coefficients used are too many.

• *Positiveness of Coefficients:* Clearly, rollup and conditioning algorithms will keep the positiveness of the coeffi-

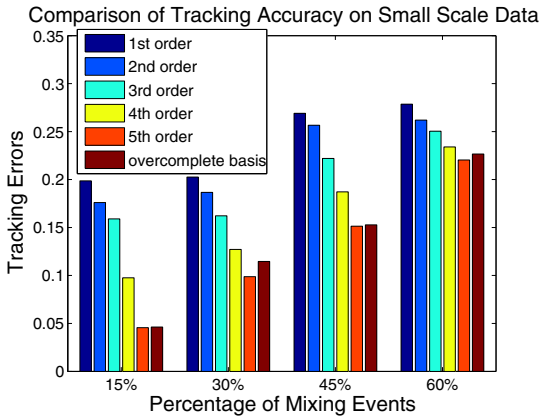


Figure 4: Tracking errors with different percentages of mixing events.

cients. However, OMP algorithm may result in approimating the distribution with negative coefficients. In practice, we preserve positiveness of the coefficients by projecting on the positive cone.

• *Prediction:* By using pseudo-inverse $(\Phi^{j,k})^+$ mentioned in section 4.1, we are able to compute the score for each target. We pick out k tracks which have the hightest scores and predict the targets on those k tracks are red. This is equivalent to looking at orthogonal projection of f to subspaces spanned by $\Phi^{1,k}$ and find out which $k-$subset has the largest weight.

• *Comparative Methods:* Similar as the approaches taken in [12, 14] in studying identity management problem, we can approximate the distribution over X^k by using low order Radon bases, i.e., $f \approx R^{j,k}c_j$ with small j's so that we only store c_j which is a vector of length $\binom{n}{j}$. Though polynomial in n, low order Radon bases are incapable of characterizing peaky distribuion over X^k. In section 6, we will compare our approach with using only low order Radon basis coefficients.

• *k is not known as a priori:* For a more general case where k - the number of red targets is not known as a priori, we can still use the Radon bases to address the problem. Essentially the rollup step and the conditioning step will be unaffected, because we can see that the algorithm for the rollup step is independent of k, while the algorithm for the conditioning step can also be adapted to the case that k is not known (we just pretend the 'if' statement is not there). For the sparse approximation step, we note that since the spaces spanned by Radon bases of different orders have a particular hierarchical structure, thus we can approximate high (low) order Radon bases coefficients with nearing low (high) order Radon bases coefficients, depending on the scenario is highly uncertain (certain). Such an operation will be independent of k.

• *Connections between Property Management and Identity Management:* As for the difference between property management and identity management, property management tries to infer less information and therefore can be easier and less costly to implement. At the same time, in many settings, information about properties can be sufficient for the network needs, e.g., to differentiate friend from enemy.

The approach for the property management problem provides an alternative method for the identity management problem. From a mathematical point of view, suppose we have n targets, then we can code them using $\mathcal{O}(\log n)$ bits such that each identity has a unique binary code of length $\mathcal{O}(\log n)$. For each bit, all the targets can be classified as either red or blue depending on whether the bit is 0 or 1. Then based on probabilistic beliefs on each bit, we can infer the target identity. One can generalize such an approach where properties of the targets act as features which they collaboratively determine the target identities.

On the other hand, it turns out that the Fourier basis coefficients used for the identity management problem [11, 12] can be collapsed to the Radon basis coefficients discussed in this paper. The intuition is that if we have a distribution over the permutation group (each permutation assigns target positions to the identities), and if the first k identities have color property red, then by summing up the probabilities over all permutations such that it maps a particular $k-$subset to the first k identities, i.e., we don't care the permutation within, essentially we are collpasing the fourier basis coefficients for permutation groups to the Radon basis coefficients for the homogeneous spaces. Thus, using Radon bases to address the property management problem can be viewed as an approach which collapse the probability distributions given by the Fourier coefficients for the identity management problem.

6. EVALUATION

In this section, we perform several experiments to illustrate the effectiveness and efficiency of the proposed approach. We use the Delta3D game engine to generate simulated crowds of up to 100 moving targets wearing either red or blue clothes and walking around in an outdoor market [10]; figure 3 depicts a snapshot view of the simulated crowd. Such a simulation approach allows us to obtain accurate ground truth for big crowds than was feasible in usual physical testbed. The data contains interesting movement patterns and we can extract mixing and observation events directly from the data. We log a mixing event whenever two targets get close to each other within some distance and an observation event whenever one target is separated from all the other targets for some distance. We can control the percentages of mixing events by adjusting the distance parameters as well. We measure tracking errors using the fraction of mislabeled target properties over the tracks.

We first run a small scale experiments where there are 10 targets, 5 red and 5 blue. The homogeneous space X^5 is of size $\binom{10}{5} = 252$. As illustrated in figure 4, four sets of experiments with different percentages of mixing events were performed, reflecting scenarios of highly certainty to highly confusion. For each set of experiments, we run the overcomplete basis algorithm and compare with using only low order Radon basis coefficients. We measure tracking errors using the fraction of mislabeled target properties over the tracks. When mixing events happen rarely, using high order Radon basis coefficients can greatly help to improve tracking accu-

Table 3: Large Scale Experiments

Methods	#Targets	Tracking Errors	Running Time
1st Order Radon Basis	$n = 20$	0.3167	0.02(s)
	$n = 60$	0.2804	0.17(s)
	$n = 100$	0.2836	0.46(s)
3rd Order Radon Basis	$n = 20$	0.2977	29.79(s)
	$n = 60$	0.2845	119.45(s)
	$n = 100$	0.2891	912.21(s)
Overcomplete Basis	$n = 20$	0.1727	29.69(s)
	$n = 60$	0.1751	292.90(s)
	$n = 100$	0.1823	1342.35(s)

Table 4: Identity Management Tracking Accuracy

Methods	Fourier Approach (order)	Radon Approach (#Properties)
$n = 20$	0.3842 (1)	0.4543 (3)
	0.3740 (2)	0.4189 (5)
	0.3603 (3)	0.3592 (7)
$n = 60$	0.3811 (1)	0.4428 (5)
	0.3701 (2)	0.3741 (7)
	0.3595 (3)	0.3624 (9)
$n = 100$	0.3814 (1)	0.4250 (7)
	0.3569 (2)	0.3809 (9)
	0.3499 (3)	0.3447 (11)

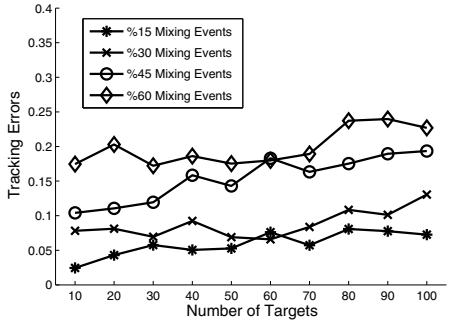


Figure 5: Tracking errors with large scale data.

racy; while if mixing events happen frequently, using high order Radon basis coefficients does not help much to improve tracking accuracy. This is reasonable since if mixing events happen rarely, distributions can be well characterized by a high order Radon basis while low order bases are not sufficient to characterize distributions; on the other hand, if mixing events happen frequently, distributions can be well characterized by low order Radon bases, so using a high order Radon basis would not provide additional benefits. Our overcomplete basis approach uses on average about 50 bases to characterize the distribution over X^5. The tracking accuracy is almost comparable to completely storing the distribution on X^5, which requires storing 252 coefficients.

From the small scale experiment, we can see that there is a fundamental trade off between the number of coefficients used and the tracking accuracy. With more bases used, we can track targets better, however, we can not use as many bases as we want if the problem size goes large because in the extreme case we would use exponentially many bases. Moreover, the Heisenberg uncertainty principle plays an important role in our experiments. The scenario where there are very few mixing events can be well characterized by using only the low order Radon bases; while the scenario where there are a lot of observation events can be well characterized by using only the high order Radon bases. That's why we see greater improvements of the tracking accuracy by using more high order Radon bases for the case where the targets have few mixings; while there are relatively smaller improvements of tracking accuracy by using more high order Radon bases if targets are well mixed. Thus, using overcomplete Radon basis is a good way to balance high tracking accuracy requirement and computational efficiency.

Our algorithm shows great benefits on tracking accuracy and computational time in larger scale experiments. When there are n ($n \geq 20$) targets and half of them are red, it would be impossible to store the entire distribution f. Thus we compare our approach with using only low order Radon basis coefficients which approximates f, i.e., we approximate f by $R^{j,k}c_j$ with small j. In our experiments, there are 1000 time steps and half of them are mixing events. We use up to 100, 300 and 500 bases in the overcomplete basis algorithm to characterize the distribution. From table 3, we see that our approach improves the tracking accuracy greatly compared with only using low order Radon basis coefficients and the running time of our approach is comparable to using Radon basis coefficients of order 3. We also run the large scale experiments to see how the tracking accuracy varies with changes of the percentages of mixing events. As shown in figure 5, the tracking errors increase when mixing events happen more frequently, which coincides with the intuition. The tracking errors also increase with the increment of the number of targets. The explanation for this is that we use linear growing number of bases to approximate a distribution of whose size is exponentially growing.

We finally set up an experiment which compares using the properties to infer the target identities with the existing Fourier approach for identity management problem. In this experiment, each target has many properties (features), which collaboratively determine its identity. It can be seen from table 4 that the more features are used, the better tracking accuracy can be achieved. The tracking accuracy of two approaches are comparable.

7. CONCLUSION

In this paper, we studied the *property management problem*. A novel method which uses an overcomplete Radon basis dictionary to represent uncertainties is proposed. We developed scalable algorithms to efficiently update the Radon basis coefficients, together with approximation algorithms which maintain a sparse approximation of the true distribution. Compared with other possible methods, the proposed approach achieves better performance in tracking accuracy with tolerable computation overhead.

8. ACKNOWLEDGEMENT

The authors would like to thank the shepherd, Bhaskar Krishnamachari, for his constructive feedback and valuable input. Thanks also to Kyle Heath, Qixing Huang, and anonymous reviewers for reading this paper and giving valuable comments. Xiaoye Jiang and Leonidas Guibas wish to acknowledge the support of ARO grants W911NF-10-1-0037 and W911NF-07-2-0027, as well as NSF grant CCF 1011228 and a gift from the Google Corporation. Mo Li acknowledges support from COE SUG/RSS 20Aug201013/14 in Nanyang Technological University of Singapore. Yuan Yao acknowledges supports from the National Basic Research Program of China (973 Program 2011CB809105), NSFC (61071157), Microsoft Research Asia, and a professorship in the Hundred Talents Program at Peking University.

9. REFERENCES

[1] A. Arora, P. Dutta, S. Bapat, V. Kulathumani, H. Zhang, V. Naik, V. Mittal, H. Cao, M. Gouda, Y. Choi, T. Herman, S. Kulkarni, U. Arumugam, M. Nesterenko, A. Vora, and M. Miyashita. A line in the sand: a wireless sensor network for target detection, classification, and tracking. *Computer Networks*, 46:605–634, 2004.

[2] J. Aslam, Z. Butler, F. Constantin, V. Crespi, G. Cybenko, and D. Rus. Tracking a moving object with a binary sensor network. In *Proceedings of ACM SenSys*, 2003.

[3] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.

[4] J. Collins and J. Uhlmann. Efficient gating in data association with multivariate distributed states. *IEEE Transaction on Aerospace and Electronic Systems*, 28:909–916, 1992.

[5] I. Cox and S. Hingorani. An efficient implementation of reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. In *Proceedings of ICPR*, 1994.

[6] P. Diaconis. *Group Representations in Probability and Statistics*. Institute of Mathematical Statistics, 1988.

[7] P. Dutta, J. Hui, J. Jeong, S. Kim, C. Sharp, J. Taneja, G. Tolle, K. Whitehouse, and D. Culler. Trio: enabling sustainable and scalable outdoor wireless sensor network deployments. In *Proceedings of IEEE/ACM IPSN*, 2006.

[8] A. Gilbert, S.Muthukrishnan, and M. J. Strauss. Approximation of functions over redundant dictionaries using coherence. In *Proceedings of ACM-SIAM SODA*, 2003.

[9] L. J. Guibas. The identity management problem – a short survey. In *Proceedings of International Conference on Information Fusion*, 2008.

[10] K. Heath and L. Guibas. Multi-person tracking from sparse 3d trajectories in a camera sensor network. In *Proceedings of IEEE ICDSC*, 2008.

[11] J. Huang, C. Guestrin, and L. J. Guibas. Fourier theoretic probabilistic inference over permutations. *Journal of Machine Learning Reserach*, 10:997–1070, 2009.

[12] J. Huang, C. Guestrin, X. Jiang, and L. J. Guibas. Exploiting probabilistic independence for permutations. In *Proceedings of AISTATS*, 2009.

[13] X. Jiang, Y. Yao, and L. Guibas. Stable identification of cliques with restricted sensing. In *Neural Information Processing Systems (NIPS) - Learning with Orderings workshop*, 2009.

[14] R. Kondor, A. Howard, and T. Jebara. Multi-object tracking with representations of the symmetric group. In *Proceedings of AISTATS*, 2007.

[15] L. Mo, Y. He, Y. Liu, J. Zhao, S.-J. Tang, X. Li, and G. Dai. Canopy closure estimates with greenorbs: sustainable sensing in the forest. In *Proceedings of the ACM Sensys*, 2009.

[16] K. Murty. An algorithm for ranking all the assignments in order of increasing cost. *Operations Research*, 16:682–687, 1968.

[17] S. Oh, S. Russell, and S. Sastry. Markov chain monte carlo data association for general multiple-target tracking problems. In *Proceedings of IEEE CDC*, 2004.

[18] S. Oh and S. Sastry. A polynomial-time approximation algorithm for joint probabilistic data association. In *Proceedings of the American Control Conference*, 2005.

[19] A. Poore. Multidimensional assignment and multitarget tracking. In *Proceedings of Partitioning Data Sets – DIMACS Workshop*, volume 19, pages 169–196, 1995.

[20] D. Reid. An algorithm for tracking multiple targets. *IEEE Transaction on Automatic Control*, 6:843–854, 1979.

[21] B. Sagan. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer-Verlage, 2001.

[22] B. Schumitsch, S. Thrun, G. Bradski, and K. Olukotun. The information-form data association filter. In *Proceedings of NIPS*, 2005.

[23] J.-P. Serre. *Linear Representation of Finite Groups*. Springer-Verlag, 1977.

[24] J. Shin, L. J. Guibas, and F. Zhao. A distributed algorithm for managing multi-target identities in wireless ad-hoc sensor networks. In *Proceedings of IEEE/ACM IPSN*, 2003.

[25] D. Xie, T. Yan, D. Ganesan, and A. Hanson. Design and implementation of a dual-camera wireless sensor network for object retrieval. In *Proceedings of IEEE/ACM IPSN*, 2008.

[26] Z. Yang, Y. Liu, and X. Li. Beyond trilateration: On the localizability of wireless ad-hoc networks. *IEEE/ACM Transactions on Networking*, 18(6):1806–1814, 2010.