

# Incremental inference of collective graphical models

Rahul Singh, Isabel Haasler, Qinsheng Zhang, Johan Karlsson, and Yongxin Chen

**Abstract**—We consider incremental inference problems from aggregate data for collective dynamics. In particular, we address the problem of estimating the aggregate marginals of a Markov chain from noisy aggregate observations in an incremental (online) fashion. We propose a sliding window Sinkhorn belief propagation (SW-SBP) algorithm that utilizes a sliding window filter of the most recent noisy aggregate observations along with encoded information from discarded observations. Our algorithm is built upon the recently proposed multi-marginal optimal transport based SBP algorithm that leverages standard belief propagation and Sinkhorn algorithm to solve inference problems from aggregate data. We demonstrate the performance of our algorithm on applications such as inferring population flow from aggregate observations.

**Index Terms**—Markov processes, filtering, stochastic systems

## I. INTRODUCTION

THE problem of inference from aggregate data is widely studied in fields including machine learning, ecology, and social sciences [1], [2]. Similar problems also occur in the area of estimation and control, for instance in ensemble filtering [3]–[6]. In these applications, one aims to infer information about a group of agents in the case where only aggregate observations in the form of counts or contingency tables are provided [1]. Information about individuals may not be available due to, e.g., economical or privacy reasons [1]. For example, in bird migration analysis, individual trajectories are not readily accessible, but the number of birds in different areas can typically be counted from pictures. Another very present application is the spread of infectious diseases, where infections are usually under-reported, but inference of the true transmission dynamics is of utmost importance [7].

In control applications, individual dynamics are often estimated using for instance the Kalman filter [8]. In statistics, many methods are based on probabilistic graphical models (PGMs), e.g., the belief propagation algorithm [9]. However, neither approach is tractable when it comes to estimating the behavior of a large group of individuals simultaneously, let alone the fact that quite often only aggregate information is available. A number of methods have been developed to address this problem, for instance, the PGM framework has been extended to collective graphical models (CGMs), which is a formalism for inference and learning with aggregate data [1]. A CGM is a graphical model that describes the

This work was supported by the Swedish Research Council (VR), grant 2014-5870, SJTU-KTH cooperation grant and the NSF under grant 1901599 and 1942523.

R. Singh, Q. Zhang and Y. Chen are with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [{qzhang419,rasingh,yongchen}@gatech.edu](mailto:{qzhang419,rasingh,yongchen}@gatech.edu)

I. Haasler and J. Karlsson are with the Division of Optimization and Systems Theory, Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden. [haasler@kth.se](mailto:haasler@kth.se), [johan.karlsson@math.kth.se](mailto:johan.karlsson@math.kth.se)

relationship between the aggregated counts of individuals. Several algorithms for aggregate marginal inference within the CGM framework have been proposed including Sinkhorn belief propagation (SBP) [10], approximate MAP inference [11] and non-linear belief propagation [2]. Whereas the latter two suffer from instability and lack convergence guarantees, the SBP algorithm is guaranteed to converge in case the underlying graphical model is acyclic.

The SBP algorithm is based on multi-marginal optimal transport theory, which studies the problem of finding the most efficient transport plan between several distributions [12], [13]. It has been shown that the aggregate inference problem is equivalent to the entropic regularized formulation of a multi-marginal optimal transport problem [10]. With this equivalence, the celebrated Sinkhorn algorithm [13], [14] has been utilized to solve the aggregate inference problem.

Most estimation tasks in the control community involve dynamic systems and thus evolution over time. This dynamic nature adds a temporal component to CGM. As time evolves, the number of aggregate noisy observations and the size of the underlying graphical model increase constantly. Clearly, naive CGM inference algorithms such as SBP are not suitable for real-time operations as the computational complexity increases linearly with the length of the graph. Thus, to achieve real-time performance, we consider the problem of estimating the aggregate marginals of a Markov chain from noisy, aggregate observations in an incremental (online) fashion. In this problem,  $M$  individuals behave independently according to the same underlying Markov chain for multiple time steps, and, at each time step, a noisy aggregate observation is made. As a new aggregate observation comes in at time step  $t$ , the goal is to estimate the aggregate marginal at time  $t$ . In case of an individual's model ( $M = 1$ ), this coincides with the traditional PGM, and naive incremental inference [15] can be used by incorporating a sliding window filter to consider only a fixed number of most recent observations.

For collective dynamics, following the naive incremental approach, we build on the SBP algorithm and propose the sliding window Sinkhorn belief propagation (SW-SBP) algorithm for the incremental (online) aggregate inference problem. The SW-SBP algorithm employs a sliding window filter of length  $K$ , i.e., at each time step the  $K$  most recent observations are used in order to estimate the current hidden aggregate distribution. In order to capture previous information, we propose to add one node at the beginning of the window, for which we consider two different settings. In one setting, the marginal distribution of this node is specified. In the other one, the node introduces some prior potential. We evaluate the performance of our algorithm on a variety of scenarios, including population flow analysis, validating the working of our algorithm.

The contribution of this work is twofold. On one hand, this is

the first study of incremental inference in the CGM framework. Our method makes it possible to estimate the group behavior of a large collection of agents using aggregate observations in an online manner. On the other hand, this extends many incremental inference methods [15]–[17] to the CGM setting. Indeed, when specialized to individual dynamics, our method reduces to a standard incremental inference algorithm.

## II. BACKGROUND

### A. Probabilistic Graphical Models and Belief Propagation

Consider a distribution of  $J$  random variables with the same finite image space  $\mathcal{X}$ , where  $|\mathcal{X}| = d$ , and with some dependencies between them. A probabilistic graphical model (PGM) is a compact and intuitive representation of such a distribution, which describes the dependencies by a graph [18]. More precisely, for a graph  $G = (V, E)$ , the random variables are represented by the set of vertices  $V$ , and the dependencies by the set of edges  $E$ . The joint probability of the distribution of random variables can then be represented as

$$p(\mathbf{x}) := p(x_1, x_2, \dots, x_J) = \frac{1}{Z} \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j), \quad (1)$$

where  $\mathbf{x} = \{x_1, \dots, x_J\} \in \mathcal{X}^J$ ,  $\psi_{ij} \in \mathbb{R}$  are edge potentials, and  $Z \in \mathbb{R}$  is a normalization constant. The edge potential  $\psi_{ij}$  characterizes the strength of the dependency between the random variables at nodes  $i$  and  $j$ . Sometimes one also defines node potentials  $\phi_i(x_i)$ , for  $i \in V$ . However, these can be absorbed in the edge potentials  $\psi_{ij}$ , and in this work we choose the compact notation (1).

Often, one is interested in finding the distribution of one of the random variables, which is called the Bayesian marginal inference problem. An efficient method for this problem is the belief propagation algorithm, which updates the marginal distribution on the vertices, by sending messages (also called beliefs) between them [9]. Let  $N(i)$  denote the set of neighboring nodes of  $i$ . Then the message from variable node  $i$  to variable node  $j$  is

$$m_{i \rightarrow j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(x_i). \quad (2)$$

This message can be understood as the belief of node  $i$  about node  $j$ . The messages in (2) are updated iteratively over the graph. When the algorithm converges, the node and edge marginals are given by

$$b_i(x_i) \propto \prod_{k \in N(i)} m_{k \rightarrow i}(x_i) \quad (3a)$$

$$b_{ij}(x_i, x_j) \propto \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(x_i) \prod_{\ell \in N(j) \setminus i} m_{\ell \rightarrow j}(x_j). \quad (3b)$$

For an acyclic graph the belief propagation algorithm converges globally [19] and the estimated marginal distributions in (3) recover the true marginals exactly. Although convergence is not guaranteed for general graphs with cycles, in practice the belief propagation algorithm often performs well [20].

### B. Collective Graphical Models

Collective graphical models (CGMs) describe the distribution of the aggregate data from several populations, which are each sampled independently from a discrete graphical model [1]. Consider a PGM as in Section II-A with underlying graph  $G = (V, E)$ . Let  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}$  be  $M$  samples of the PGM according to its joint probability distribution (1). In particular, each sample is a tuple  $\mathbf{x}^{(m)} = (x_1^{(m)}, \dots, x_J^{(m)})$ , where  $x_i^{(m)} \in \mathcal{X}$ , for each  $i = 1, \dots, J$ .

Let  $X_i^{(m)}$  be the state of the  $m^{th}$  individual at node  $i$ , and let  $\mathbb{I}[\cdot]$  denote the indicator function. Then the aggregate node distribution  $\mathbf{n}_i \in \mathbb{R}^d$  and aggregate edge distribution  $\mathbf{n}_{ij} \in \mathbb{R}^{d \times d}$  are element-wise given by

$$n_i(x_i) = \sum_{m=1}^M \mathbb{I}[X_i^{(m)} = x_i], \quad \text{for } i \in V, \quad (4a)$$

$$n_{ij}(x_i, x_j) = \sum_{m=1}^M \mathbb{I}[X_i^{(m)} = x_i, X_j^{(m)} = x_j], \quad \text{for } (i, j) \in E. \quad (4b)$$

The collection of all the aggregate node and edge distributions is denoted as  $\mathbf{n}$ , i.e.,  $\mathbf{n} = \{\mathbf{n}_i, \mathbf{n}_{ij} | i \in V, (i, j) \in E\}$ . By construction all entries of  $\mathbf{n}$  are integers and they satisfy

$$\begin{aligned} \sum_{x_i} n_i(x_i) &= M, & \text{for } i \in V, \\ n_i(x_i) &= \sum_{x_j} n_{ij}(x_i, x_j), & \text{for } (i, j) \in E. \end{aligned} \quad (5)$$

From the underlying PGM  $p(\mathbf{x})$ , one can calculate the probability distribution of  $\mathbf{n}$ , and this is known as the CGM. Similar to PGM, an important problem in CGM is to infer the marginal distributions given some measurements. Multiple algorithms for aggregate marginal inference within the CGM framework have been proposed including the recent Sinkhorn belief propagation (SBP) [10], approximate MAP inference [11] and non-linear belief propagation (NLBP) [2]. The approximate MAP inference and NLBP often suffer from instability and lack of convergence, while SBP exhibits convergence guarantees for acyclic graphs.

### C. Sinkhorn Belief Propagation

The Sinkhorn belief propagation algorithm [10] is based on belief propagation and utilizes the celebrated Sinkhorn algorithm for multi-marginal optimal transport [13], [21], [22], in order to solve the aggregate inference problem efficiently.

Let  $G = (V, E)$  be a CGM, as in Section II-B, with joint aggregate distribution  $\mathbf{n}$ . Moreover, let  $\Gamma \subset \{1, \dots, J\}$  be a set of indices that represents all the nodes that are observed. That is, it holds  $\mathbf{n}_i = \mathbf{y}_i$  for a given set of observations  $\mathbf{y}_i$ , for  $i \in \Gamma$  (see [10] for more discussion on this observation model). In its variational form, the marginal inference problem for CGM with this observation model reads

$$\begin{aligned} \min_{\mathbf{n}} \quad & \text{KL}(\mathbf{n} \parallel \prod_{(i,j) \in E} \psi_{ij}(x_i, x_j)) \\ \text{s. t.} \quad & \mathbf{n}_i = \mathbf{y}_i, \quad \forall i \in \Gamma. \end{aligned} \quad (6)$$

When the underlying graph is a tree, the objective function of problem (6) is the same as the Bethe free energy [23]

$$F_{\text{Bethe}}(\mathbf{n}) = \sum_{i,j} \sum_{x_i, x_j} n_{ij}(x_i, x_j) \ln \frac{n_{ij}(x_i, x_j)}{\psi_{ij}(x_i, x_j)} - \sum_{i=1} (d_i - 1) \sum_{x_i} n_i(x_i) \ln n_i(x_i), \quad (7)$$

together with the consistency constraints (5). Here,  $d_i$  denotes the degree of node  $i \in V$ . Then the aggregate inference problem, with aggregate observations  $\mathbf{y}_i$  for  $i \in \Gamma$ , reads

$$\min_{\mathbf{n}_{ij}, \mathbf{n}_i} F_{\text{Bethe}}(\mathbf{n}) \quad (8a)$$

$$\text{s.t. } n_i(x_i) = y_i(x_i), \forall i \in \Gamma \quad (8b)$$

$$\sum_{x_j} n_{ij}(x_i, x_j) = n_i(x_i), \forall (i, j) \in E \quad (8c)$$

$$\sum_{x_i} n_i(x_i) = 1, \forall i \in V. \quad (8d)$$

Note that here (8b) corresponds to the aggregate observation constraints and (8c)-(8d) to the normalized consistency constraints (5). The solution to this problem is characterized by the following result.

*Theorem 1 ([10, Theorem 1]):* The solution to the aggregate inference problem (8) is characterized by

$$n_i(x_i) \propto \prod_{k \in N(i)} m_{k \rightarrow i}(x_i), \forall i \notin \Gamma \quad (9)$$

where  $m_{i \rightarrow j}(x_j)$  are fixed points of

$$m_{i \rightarrow j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(x_i); \\ \forall i \notin \Gamma, \forall j \in N(i), \quad (10a)$$

$$m_{i \rightarrow j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \frac{y_i(x_i)}{m_{j \rightarrow i}(x_i)}; \\ \forall i \in \Gamma, \forall j \in N(i). \quad (10b)$$

Note that the inference problem (6) can be viewed as an entropy regularized multi-marginal optimal transport problem [10], [13], [21]. As a result, one can utilize the efficient Sinkhorn algorithm to solve the inference problem (6). This can be accelerated further by merging it with belief propagation. To summarize, by combining Sinkhorn for multi-marginal optimal transport problem and Theorem 1, we establish Algorithm 1. We refer the reader to [10] for details of the derivation. Note that in contrast to other algorithms for the aggregate inference problem, which rely on an explicit observation model [2], SBP is guaranteed to converge when the underlying graph is a tree due to its foundations: Sinkhorn and belief propagation both converge [13], [19], [24], [25].

The expressions in (10) can be interpreted as messages between nodes, in analogy to the standard belief propagation method presented in Section II-A. In fact, the messages in (10) resemble the ones in (2). The messages (10a) can be understood as a scaling step, which guarantees that the constraints (8b) remain satisfied.

---

### Algorithm 1 Sinkhorn Belief Propagation (SBP)

---

```

Initialize the messages  $m_{i \rightarrow j}(x_j)$ 
Update  $m_{i \rightarrow j}(x_j)$  using (10)
while not converged do
  for  $i \in \Gamma$  do
    i) Update  $m_{i \rightarrow j}(x_j)$  using (10b)
    ii) Update all the messages on the path from  $i$  to  $i_{\text{next}}$ 
        according to (10a) (fix an order in  $\Gamma$ ,  $i_{\text{next}}$  is the next
        element to  $i$  in this order)
  end for
end while

```

---

### III. MAIN RESULTS

We consider *incremental* marginal inference problems from aggregate data within the CGM framework. In particular, we address the problem of estimating the aggregate marginals of a hidden Markov chain from noisy aggregate observations in an incremental (online) fashion. We follow the model described in Section II-C as the generative model of aggregate data. A hidden Markov model (HMM) is a Markov chain where the state is not directly observable. The joint distribution of an HMM factorizes as

$$p(\mathbf{x}, \mathbf{o}) = p(x_1) \prod_{t=1} p(x_{t+1} | x_t) p(o_t | x_t), \quad (11)$$

where  $\mathbf{x}, \mathbf{o}$  denotes state variable and observation respectively,  $p(x_1)$  is the initial distribution of the starting state,  $p(x_{t+1} | x_t)$  are the transition probabilities between hidden (unobserved) variables, and  $p(o_t | x_t)$  are the observation probabilities for time steps  $t = 1, 2, \dots$ . Note that the model given by (11) is a directed graphical model but it can be equivalently converted to the undirected graphical model (1) by viewing the transition and observation probabilities as edge potentials  $\psi_{ij}$  in (1).

In this problem,  $M$  individuals behave independently according to the same underlying hidden Markov chain for multiple time steps, and, at each time step, a noisy aggregate observation  $\mathbf{y}_t$  is made in terms of the number of individuals in each observation node  $o_t$ . As a new observation is made at time step  $t$ , our goal is to estimate the corresponding aggregate marginal at time  $t$ . The number of aggregate noisy observations and the size of the underlying graphical model increases with time. To perform inference from this full data is not suitable for real-time operations as the computational complexity of inference increases greatly with the size of the graph. Therefore, incremental inference is of great importance in that it can efficiently estimate the required aggregate marginals from new observations without needing access to the previous observations.

Building on the SBP algorithm, we propose a sliding window Sinkhorn belief propagation (SW-SBP) algorithm for the incremental (online) aggregate inference problem. Borrowing the idea from traditional incremental HMM inference [15], the SW-SBP algorithm employs a sliding window filter of length  $K$  so that at time step  $t$  ( $t > K$ ), only the  $K$  most recent observations are used in order to estimate the conditional probabilities. In addition to using only  $K$  previous observations,

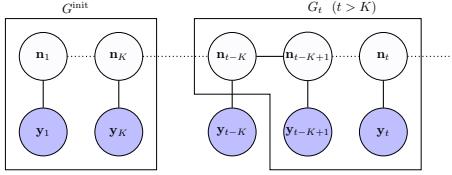


Fig. 1: Incremental HMM in SW-SBP.

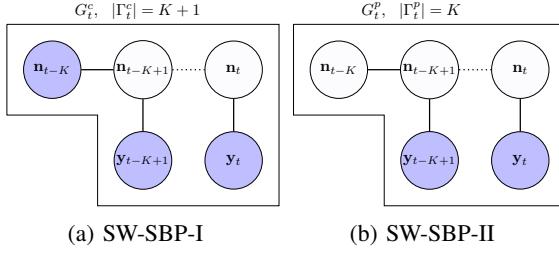


Fig. 2: Incremental HMM graph for the two methods.

we propose to add a node<sup>1</sup> to capture most of the information lost by discarding the previous observations. Denote this subgraph of the HMM model by  $G_t = (V_t, E_t)$  as shown in Figure 1. Here, the incremental time starts at  $t = K + 1$ . With this setting, we always have  $2K + 1$  nodes (in  $G_t$ ) at any time step  $t$  from which the aggregate marginal is estimated. The Bethe free energy (7) for the subgraph/sliding window  $G_t$  with the underlying HMM model (11) is

$$\begin{aligned} & \sum_{i=t-K+1}^t \left( \sum_{x_{i-1}, x_i} n_{i-1, i}(x_{i-1}, x_i) \ln \frac{n_{i-1, i}(x_{i-1}, x_i)}{p(x_i | x_{i-1})} + \right. \\ & \sum_{x_i, o_i} n_{i, i}(o_i, x_i) \ln \frac{n_{i, i}(o_i, x_i)}{p(o_i | x_i)} - 2 \sum_{x_i} n_i(x_i) \ln n_i(x_i) \Big) + \quad (12) \\ & \sum_{x_t} n_t(x_t) \ln n_t(x_t) \end{aligned}$$

We propose two methods for encoding information from the previous discarded observations in the node  $n_{t-K}$ . In the first method (SW-SBP-I) the node is constrained to be equal to the estimate of this node from the previous time step. In the second method (SW-SBP-II) previous information is imposed by updating the potential  $\phi_{t-K}$  on this node to the message  $m_{t-K \rightarrow t-K+1}$  obtained at the previous time step.

Figure 2 shows the incremental HMM graphs corresponding to the two proposed schemes. The graph corresponding to SW-SBP-I at time step  $t$  is  $G_t^c = (V_t^c, E_t^c)$  as shown in Figure 2a. Here, the node  $n_{t-K}$  is a fixed marginal node, which equals the estimated marginal on  $G_{t-1}^c$ . The index set of constraints  $\Gamma_t^c$  is described by the set of shaded nodes in Figure 2a. At time step  $t$ , SW-SBP-I solves the inference problem (8) with  $\Gamma = \Gamma_t^c$  and Bethe free energy given as in (12). The steps of SW-SBP-I are listed in Algorithm 2, where  $G^{\text{init}}$  is the initial graph with the first  $K$  number of observations (2 $K$  nodes) as in Figure 1.

For the SW-SBP-II case, the incremental graph at time  $t$  is represented by  $G_t^p = (V_t^p, E_t^p)$  as in Figure 2b. Here, the index set of constraints  $\Gamma_t^p$  does not contain  $t - K$ . Instead, the

<sup>1</sup>In case no extra node is added, the resulting algorithm is called naive SW-SBP.

---

## Algorithm 2 SW-SBP-I (constrained marginal)

---

```

Run SBP on initial graph  $G^{\text{init}}$ 
for  $t = K + 1, K + 2, \dots$ , do
  Constrain marginal  $n_{t-K}$  to be equal to its estimate
  obtained from  $G_{t-1}^c$ 
  Run SBP on  $G_t^c$ 
end for

```

---

potential  $\phi_{t-K}$  (which will be absorbed into  $p(x_{t-K+1} | x_{t-K})$  and is thus not explicit in (12)) is updated to be the message from  $n_{t-K}$  to  $n_{t-K+1}$  on the graph  $G_{t-1}^p$ . Algorithm 3 lists the steps of SW-SBP-II.

---

## Algorithm 3 SW-SBP-II (potential update)

---

```

Run SBP on initial graph  $G^{\text{init}}$ 
for  $t = K + 1, K + 2, \dots$ , do
  Update node potential  $\phi_{t-K}$  by message  $m_{t-K \rightarrow t-K+1}$ 
  obtained from  $G_{t-1}^p$ 
  Run SBP on  $G_t^p$ 
end for

```

---

*Remark 1:* The idea to replace the previous observations with a single node potential as in SW-SBP-II has been widely adopted in standard filtering settings. For instance, a Kalman filter [8] utilizes this idea to efficiently estimate the current state for linear dynamic systems.

*Remark 2:* When applied to a single HMM trajectory, that is, when all the measurements  $y_i$  are Dirac distributions, the SBP algorithm reduces [10] to the standard inference algorithm: Forward-backward algorithm [18] in HMM literature. This is due to the fact that when  $y_i$  is Dirac, the belief  $m_{i \rightarrow j}$  it outputs is independent of what it receives by (10b). Consequently, SW-SBP-II reduces to a standard incremental inference algorithm in HMM which also employs a potential node to capture discarded information when all the observations  $y_i$  are Diracs.

## IV. EVALUATION

We evaluate the performance of SW-SBP on multiple experiments including a toy example and a mobility pattern estimation problem. We compare the performance of different SW-SBP algorithms in terms of the estimation error with the baseline algorithm, which is SBP (Algorithm 1) on the whole graph without discarding previous measurements.

### A. Performance Comparison

To evaluate the performance of the proposed algorithms, we simulate a time-varying HMM with randomly chosen observations and evaluate the performance for different filter lengths  $K$  and different values of state dimension  $d$ . The observation is of the same dimension as the state. The transition and observation matrices are both generated randomly using  $500\mathcal{I} + 10\exp(E)$  where  $\mathcal{I}$  denotes the identity matrix and  $E$  is a random matrix generated from a standard Gaussian distribution. Figure 3 compares the performance of proposed methods for varying sliding window length. The results are averaged over 10 different trials. It is evident that the time

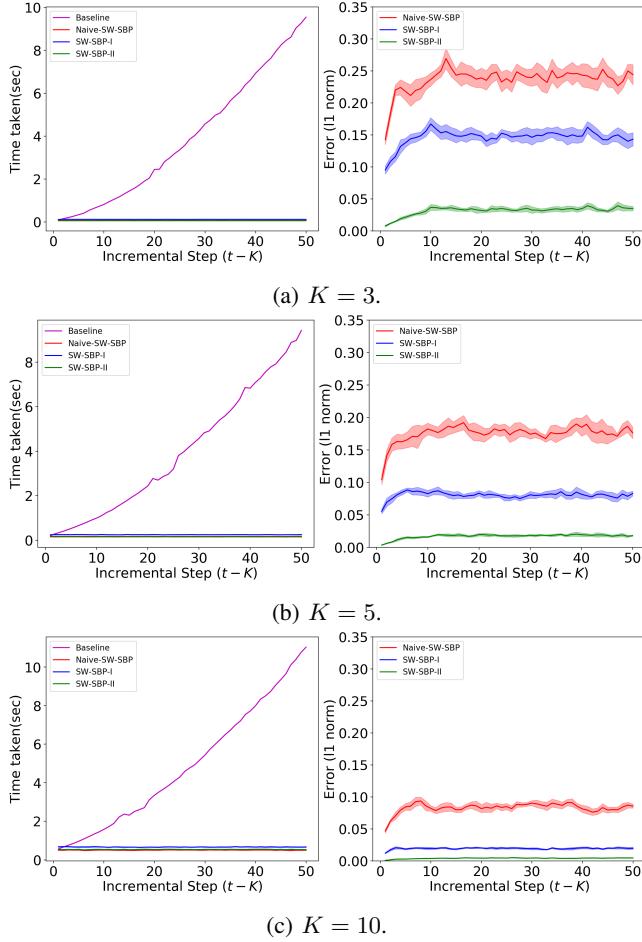


Fig. 3: Performance with  $d = 50$  for different window lengths  $K$ . The right column shows the  $\ell_1$ -norm error with respect to the baseline marginals and the left column shows the time consumption for the algorithms to converge. The solid lines denote the average over 10 trials and the shaded area represents the corresponding standard deviation.

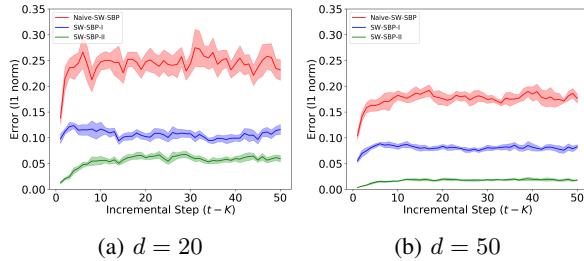


Fig. 4: Performance for different values of  $d$  with fixed window length of  $K = 5$ .

consumption of the baseline method increases rapidly with the size of the graph. Moreover, the convergence behaviors of Naive-SW-SBP, SW-SBP-I, and SW-SBP-II are similar. In terms of estimation errors, SW-SBP-I and SW-SBP-II outperform Naive-SW-SBP by a significant margin. Moreover, SW-SBP-II shows better error performance as compared to SW-SBP-I and as the length of the filter increases, the errors decrease. Figure 4 depicts the error performance for different values of  $d$  with

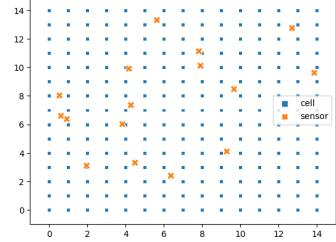


Fig. 5: Location of 16 sensors in a  $15 \times 15$  grid.

fixed window length of  $K = 5$ . It is observed from the figure that SW-SBP-I and SW-SBP-II perform similar for small values of  $d$ , outperforming Naive-SW-SBP algorithm, and SW-SBP-II performs better for large values of  $d$ . Note that the idea to update the potential in SW-SBP-II is inspired by the Kalman filter (see Remark 1). In standard inference of HMMs, which is a special case of our framework where  $y_i$  is a Dirac distribution for all  $i$ , the strategy in SW-SBP-II gives precise solutions, that is, SW-SBP-II is a precise incremental implementation of SBP. However, in cases where the measurements  $y_i$  are general distributions, SW-SBP-II only approximates SBP. Currently there is no rigorous justification why SW-SBP-II outperforms SW-SBP-I in our general setting.

### B. Bird Migration

Next, we consider the problem of estimating a mobility pattern of birds in a geographical area from noisy aggregate counts. Following the environment considered in [2], we simulate  $M$  individuals moving in an  $L \times L$  grid, aiming to reach the top-right of the grid following a Markov chain. The transition probabilities of this Markov chain follow a log-linear distribution that accounts for four factors: the distance between two positions, the angle between movement direction and the top-right direction, the angle between wind direction and the direction to goal (top-right in the grid), and preference to stay in the original cell. The weight for those four factors in the log-linear model is  $(5, 3, 1.6, 1)$ . The observations are made as the counts of individuals connected to the randomly distributed sensors in the grid (see Figure 5). These sensors constitute a distributed sensor network in which each sensor can only detect the present of birds nearby while is not able to output their locations. We assume that each individual is connected to exactly one of the sensors and the probability of the connection decreases exponentially as the distance between agent and the sensor increases. At time  $t = 0$ , the population is concentrated at the two clusters: one at the left-bottom and one at the center-bottom.

We simulate the Markov chain for  $M = 10000$  individuals independently and perform incremental inference using SW-SBP algorithms. We compare the performance of the algorithms in terms of the  $\ell_1$ -norm of the difference between the estimated marginals and the baseline marginals. Figure 7 shows the errors for the algorithm for 30 time steps on a  $15 \times 15$  grid with 16 randomly distributed sensors. It can be observed from the figure that SW-SBP-II performs significantly better than SW-SBP-I and Naive-SW-SBP. Moreover, the errors for all the three

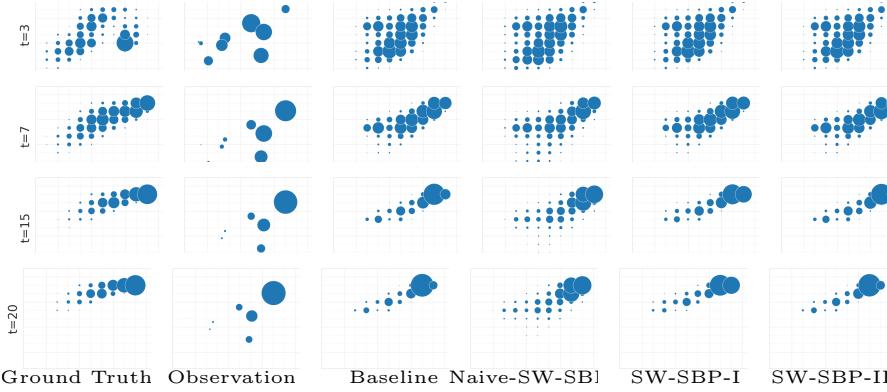


Fig. 6: Simulation of movement of  $M = 10000$  agents over  $15 \times 15$  grid with  $K = 3$ . The first column depicts the real movement of agents at different time steps, second column represents the aggregate sensor observations, third column depicts estimated aggregated positions using the baseline full graph. Fourth, fifth, and sixth columns represent the estimated positions obtained by different variants of SW-SBP. The size of the circles is proportional to the number of agents.

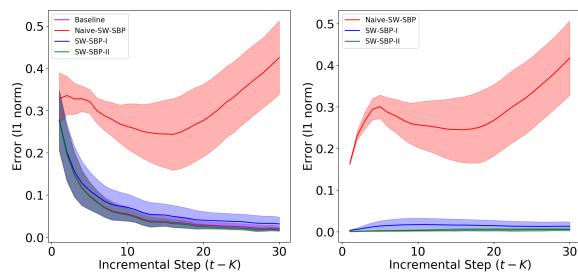


Fig. 7: Error comparison for bird migration experiment. The left and right figure show the error with respect to the ground truth and with respect to the baseline estimate, respectively.

methods decrease as the length of the sliding window filter increases. Figure 6 shows the true and estimated movements of the the population using different methods. As can be seen, SW-SBP-II infers the population movements closest to the baseline.

## V. CONCLUSION

In this paper, we proposed an algorithm for incremental inference from aggregate data in a Markov chain by employing a sliding window filter. The performance of proposed methods were demonstrated on multiple scenarios including estimation of bird mobility patterns. The present work focus on discrete setting and its extension to continuous state system is a future direction.

## REFERENCES

- [1] D. R. Sheldon and T. G. Dietterich, “Collective graphical models,” in *Advances in Neural Information Processing Systems*, 2011, pp. 1161–1169.
- [2] T. Sun, D. Sheldon, and A. Kumar, “Message passing for collective graphical models,” in *International Conference on Machine Learning*, 2015, pp. 853–861.
- [3] G. Evensen, “The ensemble kalman filter: Theoretical formulation and practical implementation,” *Ocean dynamics*, vol. 53, no. 4, pp. 343–367, 2003.
- [4] R. J. Lorentzen and G. Nævdal, “An iterative ensemble Kalman filter,” *IEEE Transactions on Automatic Control*, vol. 56, no. 8, pp. 1990–1995, 2011.
- [5] Y. Chen and J. Karlsson, “State tracking of linear ensembles via optimal mass transport,” *IEEE Control Systems Letters*, vol. 2, no. 2, pp. 260–265, 2018.
- [6] I. Haasler, A. Ringh, Y. Chen, and J. Karlsson, “Estimating ensemble flows on a hidden Markov chain,” in *58th IEEE Conference on Decision and Control*, 2019.
- [7] W. Yang, M. Lipsitch, and J. Shaman, “Inference of seasonal and pandemic influenza transmission dynamics,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 9, pp. 2723–2728, 2015.
- [8] R. E. Kalman, “A new approach to linear filtering and prediction problems,” 1960.
- [9] J. Pearl, “Probabilistic reasoning in intelligent systems: Networks of plausible inference,” *Morgan Kaufmann Publishers Inc.*, 1988.
- [10] R. Singh, I. Haasler, Q. Zhang, J. Karlsson, and Y. Chen, “Inference with aggregate data: An optimal transport approach,” in *Under Review*, 2020.
- [11] D. Sheldon, T. Sun, A. Kumar, and T. Dietterich, “Approximate inference in collective graphical models,” in *International Conference on Machine Learning*, 2013, pp. 1004–1012.
- [12] B. Pass, “Multi-marginal optimal transport: theory and applications,” *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 49, no. 6, pp. 1771–1790, 2015.
- [13] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, “Iterative Bregman projections for regularized transportation problems,” *SIAM Journal on Scientific Computing*, vol. 37, no. 2, pp. A1111–A1138, 2015.
- [14] M. Cuturi, “Sinkhorn distances: Lightspeed computation of optimal transport,” in *Advances in neural information processing systems*, 2013, pp. 2292–2300.
- [15] K. Murphy, “Dynamic bayesian networks: Representation, inference and learning,” *PhD thesis, University of California*, 2002.
- [16] G. Stamatescu, L. B. White, and R. Bruce-Doust, “Track extraction with hidden reciprocal chains,” *IEEE Transactions on Automatic Control*, vol. 63, no. 4, pp. 1097–1104, 2017.
- [17] L. B. White and H. X. Vu, “Maximum likelihood sequence estimation for hidden reciprocal processes,” *IEEE Transactions on Automatic Control*, vol. 58, no. 10, pp. 2670–2674, 2013.
- [18] M. J. Wainwright and M. I. Jordan, “Graphical models, exponential families, and variational inference,” *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [19] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Generalized belief propagation,” in *Advances in neural information processing systems*, 2001, pp. 689–695.
- [20] K. P. Murphy, Y. Weiss, and M. I. Jordan, “Loopy belief propagation for approximate inference: An empirical study,” in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.
- [21] L. Nenna, “Numerical methods for multi-marginal optimal transportation,” Ph.D. dissertation, 2016.
- [22] B. Pass, “On the local structure of optimal measures in the multi-marginal optimal transportation problem,” *Calculus of Variations and Partial Differential Equations*, vol. 43, no. 3–4, pp. 529–536, 2012.
- [23] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Constructing free-energy approximations and generalized belief propagation algorithms,” *IEEE Transactions on information theory*, vol. 51, no. 7, pp. 2282–2312, 2005.
- [24] J. Karlsson and A. Ringh, “Generalized Sinkhorn iterations for regularizing inverse problems using optimal mass transport,” *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1935–1962, 2017.
- [25] F. Elvander, I. Haasler, A. Jakobsson, and J. Karlsson, “Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion,” *Signal Processing*, 2020.