2019 IEEE 58th Conference on Decision and Control (CDC)
Palais des Congrès et des Expositions Nice Acropolis
Nice, France, December 11-13, 2019

Network Weight Estimation for Binary-Valued Observation Models

Yu Xing, Xingkang He, Haitao Fang, Karl Henrik Johansson

Abstract— **This paper studies the estimation of network weights for a class of systems with binary-valued observations. In these systems only quantized observations are available for the network estimation. Furthermore, system states are coupled with observations, and the quantization parts are unknown inherent components, which hinder the design of inputs and quantizers. In order to deal with the temporal dependency of observations and achieve the recursive estimation of network weights, a deterministic objective function is constructed based on the likelihood function by extending the dimension of observations and applying ergodic properties of Markov chains. By imposing an independent Gaussian assumption on disturbances, we show that the function is strictly concave and has a unique maximum identical to the true parameter vector, so in this way the estimation problem is transformed to an optimization problem. A recursive algorithm based on stochastic approximation techniques is proposed to solve this problem, and the strong consistency of the algorithm is established. Our recursive algorithm can be applied to online tasks like real-time decision-making and surveillance for networked systems. This work also provides a new scheme for the identification of systems with quantized observations.**

I. INTRODUCTION

The estimation problem of networks for dynamical systems is fundamental in diverse domains such as bioinformatics, communication, as well as social networks. For example, the knowledge of gene regulatory networks can deepen our understanding of diseases and development [1]. Besides, relationship networks among individuals contain information of group structures, which is crucial for the prediction of group behavior [2]. There are various formulations for the network estimation, e.g., topological inference [3], latent node identification [4], etc. This paper focuses on the first one, and we define networks as weighted graphs.

The estimation of network weights has attracted multidisciplinary attention for the last decades. [3] reviews methods of recovering complex networks from nonlinear dynamics. Also for nonlinear systems, [5] utilizes input design and a passivity approach to solve the estimation problem. Network estimation for consensus dynamics is considered in [6], in which the estimation problem is converted to a convex optimization one. Plenty of network estimation methods for opinion dynamics, such as DeGroot and Friedkin-Johnsen models, have also been investigated, such as compressed sensing [2], vector autoregressive processes [7], and least square algorithms [8]. Additionally, vast papers study the topic of graph signal processing, e.g., [9], [10], which focuses on processing signals from graphs and learning network topologies.

Most existing works concentrate on systems with continuous observations. In practical scenarios, however, agents often present discrete outputs rather than continuous ones [11], [12]. For instance, binary-valued signals may be the only information transmitted and observed in communication networks because of limited storage and bandwidth resources. Therefore, the study of network estimation for systems with quantized observations is necessary. To tackle this challenge, we resort to identification methods for quantized output systems.

The parameter estimation of quantized systems has developed rapidly in recent years. Based on full-rank periodic inputs, [13] introduces the optimal quasi-convex combination estimator. [14] replaces the assumption of full-rank periodic inputs by the one based on general quantized inputs. Under conditions of sufficiently rich inputs and prior knowledge of parameters, [15], [16] study a recursive projection algorithm for finite impulse response (FIR) systems. Besides, input conditions can be relaxed by designing adaptive quantizers [17], [18]. The expectation maximization (EM) algorithms are utilized to solve maximum likelihood estimation (MLE) problems for FIR systems in [19] and for ARX systems in [20], but they are batch algorithms. Finally, [21] investigates recursive identification of systems with binary outputs and ARMA noises by using stochastic approximation (SA) algorithms.

In this paper, we study the estimation of network weights for a class of binary-valued observation systems, which may not allow the design of inputs and quantizers. In these systems, agents present binary-valued outputs, which can be interpreted as true/false or active/inactive signals, and update their states based on these binary outputs. This update rule makes system states coupled with observations that cannot be modeled as selected or i.i.d. inputs as in [14], [15], [21]. Additionally, the quantization parts of the systems are unknown inherent components and cannot be designed like in [17], [18].

Our contributions are summarized as follows. We formulate a dynamical model over networks with binary-valued observations. The stability of outputs and the identifiability of the model are investigated in detail. To estimate network

This work is supported by National Key R&D Program of China (2016YFB0901900), the National Natural Science Foundation of China (61573345), Knut & Alice Wallenberg Foundation, and Swedish Research Council.

Yu Xing and Haitao Fang are with Key Lab of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, P. R. China yxing@amss.ac.cn; htfang@iss.ac.cn

Xingkang He and Karl Henrik Johansson are with Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden xingkang@kth.se; kallej@kth.se

978-1-7281-1398-2/19/$31.00 ©2019 IEEE

weights for this model, a recursive algorithm based on SA techniques is proposed.

More precisely, in order to deal with the temporal dependency of observations and achieve the recursive estimation of network weights, a deterministic objective function is constructed based on the likelihood function, by extending the dimension of observations and applying ergodic properties of Markov chains. By imposing an independent Gaussian assumption on disturbances, we show that the function is strictly concave and has a unique maximum identical to the true parameter vector. In this way, the estimation problem is transformed to an optimization problem. A recursive algorithm based on stochastic approximation techniques is proposed to solve this problem, and the strong consistency of the algorithm is established.

Unlike batch algorithms solving MLE problems in [19], [20], [23], our recursive algorithm can be applied to online tasks like real-time decision-making and surveillance for networked systems. This work also provides a new scheme for the identification of systems with quantized observations.

The remainder of this paper is organized as follows. Section II introduces some notations. We formulate the estimation problem in Section III, and study the model and its identifiability in Section IV. The estimation algorithm and numerical simulations are given in Section V. Section VI concludes the paper.

II. NOTATIONS

We represent column vectors by boldfaced lower-case or Greek letters, and their entries by lower-case letters with subscripts, e.g., a_i is the i-th entry of $\boldsymbol{a}$. By $\|\cdot\|$ we denote the Euclidean norm of a vector. Matrices and random vectors are written as upper-case letters such as A and X, but we will not emphasize the meaning unless this causes ambiguity. For a matrix A, its entries, rows, and transpose are denoted by a_{ij}, A_i, and A^T, respectively. The expectation of a random variable X is denoted by $E[X]$. For a sequence of random vectors, say $\{X_k, k \geq 0\}$, $X_{k,i}$ is used to represent the i-th entry of X_k.

Denote $|\boldsymbol{a}| = (|a_1|, \ldots, |a_n|)^T$ and $|A| = (|a_{ij}|)$, where $|x|$ is the absolute value of a real number x. The n-length all-zeros and all-ones vectors are written as $\mathbf{0}_n$ and $\mathbf{1}_n$, or simply $\mathbf{0}$ and $\mathbf{1}$. The symbol $\boldsymbol{e}_i$ denotes a unit vector with i-th entry being 1. A matrix A is called stochastic if $A\mathbf{1} = \mathbf{1}$, and called absolutely stochastic if $|A|\mathbf{1} = \mathbf{1}$. Let $\mathbb{R}^n$ be the n-dimensional Euclidean space. For a vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_{2m})^T \in \mathbb{R}^{2m}$, we define two projections from $\mathbb{R}^{2m}$ to $\mathbb{R}^m$: $h_F(\boldsymbol{x}) = (x_1, \ldots, x_m)^T$ and $h_L(\boldsymbol{x}) = (x_{m+1}, \ldots, x_{2m})^T$. That is to say, h_F collects the first m entries of $\boldsymbol{x}$, and h_L collects the last m entries of $\boldsymbol{x}$. Let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. We use $\mathcal{S}^m := \bigtimes_{i=1}^m \mathcal{U}_i$, $\mathcal{U}_i = \{0, 1\}$, to represent the Descartes product of m identical binary sets $\{0, 1\}$.

For a homogeneous and finite-state Markov chain $\{X_k\}$ taking values in a state space Ω, the transition probability from x to y is $P(x, y) = P\{X_1 = y|X_0 = x\}$, and the k-step transition probability from x to y is $P^k(x, y) = P\{X_k = y|X_0 = x\}$, $\forall x, y \in \Omega$. We say that y is reachable from x, if there exists $k \geq 1$ such that $P^k(x, y) > 0$. The Markov chain is said to be irreducible, if y is reachable from x for all $x, y \in \Omega$. The greatest common divisor of set $\{k \geq 1 : P^k(x, x) > 0\}$ is called the period of x, denoted by $d(x)$. The Markov chain is aperiodic if $d(x) = 1$ for all $x \in \Omega$. We call a probability distribution π on Ω as a stationary distribution, if $\forall y \in \Omega, \pi(y) = \sum_{x\in\Omega} \pi(x)P(x, y)$. The probability density function of the standard normal is represented by $\phi(x)$, and the cumulative density function by $\Phi(x)$, $x \in \mathbb{R}$.

III. PROBLEM FORMULATION

A. Motivation

In this paper, we consider the network estimation problem for a class of networked dynamics in which agents exchange binary-valued outputs and update according to these observations. The binary outputs can be viewed as two alternative actions of a person [24], good/failure conditions for a physical infrastructure or an economic unit [25], etc.

The dynamics have two significant characteristics: the outputs of an agent's neighbors generate a network cost for this agent, and the agent then changes its output by comparing this cost with a threshold of its own. For instance, in a collective behavior model called the threshold model [24], if for an individual, the number of people taking certain action exceeds a particular portion, generating a social cost, then this person will follow the others' actions. Other examples are cascade dynamics over networks [25], the model of binary decision with externalities in economics [26], and so on. The action generating process, specifically, updating outputs according to the average of neighbors' actions and one's threshold, can also be viewed as an optimal choice of a quadratic network utility function [27].

Since this kind of dynamics with binary-valued output and threshold rule are able to capture many realistic scenarios, a significant problem is whether we can estimate the underlying network weights from the limited binary-valued observations. This problem is crucial because one can further detect community or look for key nodes in the network, which deepens the understanding of the group and contributes to decision-making issues related to the networked dynamics.

B. Problem Formulation

Mathematically, we formulate the dynamics with binary-valued observations as follows, supposing that the network size $n \geq 2$,

$$\begin{aligned} Y_k &= AS_{k-1} + D_k, \\ S_k &= \mathcal{Q}(Y_k, \boldsymbol{c}), \end{aligned} \tag{1}$$

where $k \geq 1$, $Y_k = (Y_{k,1}, \ldots, Y_{k,n})^T$, $D_k = (D_{k,1}, \ldots, D_{k,n})^T$, $S_k = (S_{k,1}, \ldots, S_{k,n})^T$ are the state vector, the disturbance, and the observation vector at time t respectively. A is the weight matrix of the network, and $\boldsymbol{c} = (c_1, \ldots, c_n)^T$ is the unknown quantized threshold vector. $\mathcal{Q}(Y_k, \boldsymbol{c}) = (\mathbb{I}_{[Y_{k,1}>c_1]}, \ldots, \mathbb{I}_{[Y_{k,n}>c_n]})^T$ is the quantizer. Here $\mathbb{I}_{[\text{inequality}]}$

is an indicator function equal to 1 if the inequality in the bracket holds and equal to 0 otherwise.

Our main goal is to estimate the network weight matrix A and the quantization threshold vector $\boldsymbol{c}$. Denote these parameters by $\theta := \text{vec}\{(A\ \boldsymbol{c})\}$, where $(A\ \boldsymbol{c})$ is a matrix of dimension $n \times (n+1)$, and the vec$\{\cdot\}$ operator generates a vector from a matrix by stacking the transpose of its rows on one another. We propose a recursive algorithm based on SA techniques, and study the strong consistency.

For the weight matrix A, the ij-th entry represents the influence weight of j to i. To cover more situations, we do not assume that the row sums of A are 1, but if the assumption that A is stochastic or absolutely stochastic is made, then conditions on the disturbance can be relaxed, which is shown in Section IV-B. Negative weights are also permitted, which represent antagonistic relationships. We assume that $|A|$ has no row with zero sum, i.e., $|A_i|\mathbf{1} > 0$ for all i. This means that every agent has certain connections with others.

The disturbance D_t can be interpreted as an unknown external disturbance to the agents. In the following we follow an MLE approach, so we give the following standard normal assumption for the disturbance. The normal distribution assumption is not unusual for quantized systems, since it facilitates the computation of the MLE [19], [20], [23]. Furthermore, in Section V-A we show that, by imposing the Gaussian assumption, the estimation problem can be transformed to an optimization problem that can be solved via standard stochastic approximation algorithms. Additionally, if further assumption on the network weight matrix is made, then the variance of the Gaussian disturbances can be unknown, which is discussed in Section IV-B.

***Assumption** 1:* $\{D_{k,i}\}_{1\le i\le n, k\ge 1}$ are independent and identically distributed (i.i.d.) standard normal random variables, and independent of S_0.

IV. THE MODEL AND THE IDENTIFIABILITY

A. Stability of the Observation Sequence

This section investigates the stability of the observation sequence and the identifiability of the corresponding model set.

As in (1), the observation sequence $\{S_k, k \ge 0\}$ is a Markov chain with finite states. The existence of stationary distributions is a significant aspect of stochastic stability of Markov chains [28], and we have a straightforward result as follows.

***Theorem** 1:* (Stability) Suppose that Assumption 1 holds. Then the Markov chain $\{S_k\}$ defined by (1) is irreducible and aperiodic, and hence converges in distribution to a unique stationary distribution positive on $\mathcal{S}^n$ from any initial condition.

Proof: See the Appendix. ■

This theorem shows that the observation sequence can exhibit sufficient diversity for our estimation as long as the disturbance can surpass the influence of others to an agent, and make this agent display a different choice.

Define $\tilde{S}_k := (S_k^T\ S_{k-1}^T)^T, k \ge 1$. This chain is critical for our estimation. Note that $\{\tilde{S}_k, k \ge 1\}$ taking values in $\mathcal{S}^{2n}$ is also a Markov chain. For $k \ge 1$ and $\boldsymbol{s}_{k-1}, \boldsymbol{s}_k, \boldsymbol{s}_{k+1} \in \mathcal{S}^n$, there holds that

$$\begin{aligned} &P\left\{\tilde{S}_{k+1} = \begin{pmatrix} \boldsymbol{s}_{k+1} \\ \boldsymbol{s}_k \end{pmatrix} \Big| \tilde{S}_k = \begin{pmatrix} \boldsymbol{s}_k \\ \boldsymbol{s}_{k-1} \end{pmatrix}\right\} \\ &= P\{S_{k+1} = \boldsymbol{s}_{k+1} | S_k = \boldsymbol{s}_k\}. \end{aligned} \tag{2}$$

So $\{\tilde{S}_k\}$ is aperiodic. For states $(\boldsymbol{s}^T\ \boldsymbol{u}^T)^T, (\boldsymbol{x}^T\ \boldsymbol{y}^T)^T \in \mathcal{S}^{2n}$, since $\{S_k\}$ is irreducible, there exists $k \ge 1$ such that $P^k(\boldsymbol{x}, \boldsymbol{u}) > 0$. Moreover, from the proof of Theorem 1, $P(\boldsymbol{u}, \boldsymbol{s}) > 0$ holds. Hence it follows from (2) that

$$P\left\{\tilde{S}_{k+1} = \begin{pmatrix} \boldsymbol{s} \\ \boldsymbol{u} \end{pmatrix} \Big| \tilde{S}_0 = \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix}\right\} > 0,$$

which implies that $\{\tilde{S}_k\}$ is also irreducible, and further we have the following result:

***Theorem** 2 (Ergodicity):* Suppose that Assumption 1 holds. The Markov chain $\{\tilde{S}_k\}$ is irreducible and aperiodic, and converges in distribution to a unique stationary distribution positive on $\mathcal{S}^{2n}$, from any initial condition.

The next lemma illustrates the relation between $\{S_k\}$ and the stationary distribution of $\{\tilde{S}_k\}$. It says that for the stationary distribution of $\tilde{S}_k$, the conditional probability, of its first n entries taking the value of $\boldsymbol{s}_F$ given its last n entries equal to $\boldsymbol{s}_L$, is the same as the transition probability of $\{S_k\}$ from $\boldsymbol{s}_L$ to $\boldsymbol{s}_F$, which intuitively accords with the definition of $\{\tilde{S}_k\}$.

***Lemma** 1:* Suppose that Assumption 1 holds, and $\tilde{S}$ is subject to the stationary distribution of $\{\tilde{S}_k\}$. Then

$$P\{h_F(\tilde{S}) = \boldsymbol{s}_F | h_L(\tilde{S}) = \boldsymbol{s}_L\} = P(\boldsymbol{s}_L, \boldsymbol{s}_F),$$

for all $\boldsymbol{s}_F, \boldsymbol{s}_L \in \mathcal{S}^n$, where h_F and h_L are defined in Section II and $P(\cdot,\cdot)$ is the transition matrix of $\{S_k\}$.

Proof: The proof is omitted because of space limitation. ■

B. IDENTIFIABILITY

One of the central concerns in system identification is whether parameters of different values can determine an identical model [29]. For model (1), when we fix the distribution of disturbances in advance, the answer is negative by considering the following result.

***Theorem** 3:* (*Identifiability*) Suppose that Assumption 1 holds. Then distinct parameter vector θ corresponds to a distinct Markov chain $\{S_k\}$ defined by (1). That is to say, for two parameter vectors θ and $\hat{\theta}$ such that $a_{ij} \ne \hat{a}_{ij}$ or $c_i \ne \hat{c}_i$ for some $i, j \in \{1, 2, \ldots, n\}$, Markov chains $\{S_k\}$ and $\{\hat{S}_k\}$, defined by θ and $\hat{\theta}$ respectively, have different transition probability matrices.

Proof: See the Appendix. ■

If the noise assumption is relaxed to i.i.d. normal random variables with zero mean and unknown variance $\sigma > 0$, then the noise distribution function is $F(x) = \Phi(\frac{x}{\sigma})$, where $\Phi(\cdot)$ is the cumulative density function (c.d.f.) of the standard normal random variable. It follows from the proof of Theorem 3 that $c_i/\sigma = \hat{c}_i/\hat{\sigma}$, $a_{ij}/\sigma = \hat{a}_{ij}/\hat{\sigma}$, for all $1 \le i, j \le n$. This implies that the model (1) is unique up to constant multiples of the parameters.

In the literature, the influence weight matrix is often assumed to be row stochastic ($A_i\mathbf{1}=1, \forall 1\le i\le n$, and $a_{ij}\ge 0$, $\forall 1\le i,j\le n$) or absolutely row stochastic ($|A_i|\mathbf{1}=1$, $\forall 1\le i\le n$). Under this assumption, the unknown variance can also be estimated, because the model under these assumptions is equivalent to the original one under Assumption 1. To see this, denoting $B=\mathrm{diag}(a^1,\ldots,a^n)$ as the diagonal matrix with diagonal entries $a^1,\ldots,a^n$ with $a^i=|A_i|\mathbf{1}$, (1) can be written as

$$\begin{aligned}\tilde{Y}_k&=\tilde{A}S_{k-1}+\tilde{D}_k,\\ S_k&=\tilde{\mathcal{Q}}(\tilde{Y}_k),\end{aligned}\tag{3}$$

where $\tilde{Y}_k=B^{-1}Y_k$, $\tilde{A}=B^{-1}A$, $\tilde{D}_k=B^{-1}D_k$, and $\tilde{\mathcal{Q}}(\tilde{Y}_k)=(\mathbb{I}_{[\tilde{Y}_{k,1}>\tilde{c}_1]},\ldots,\mathbb{I}_{[\tilde{Y}_{k,n}>\tilde{c}_n]})^T$. Here $\tilde{c}_i=(a^i)^{-1}c_i$, and B^{-1} exists since $|A_i|\mathbf{1}>0$. So $\tilde{A}$ is absolutely row stochastic in (2), and $\tilde{D}_{t,i}$, $1\le i\le n$, become heterogeneous Gaussian noises with different variances. Under this condition, the identifiability still holds.

V. The Identification Algorithm

A. The Objective Function and Its Concavity

Recall that $\theta=\mathrm{vec}\{(A\ \ \boldsymbol{c})\}$ is the parameter vector to be estimated, and further denote $\theta^{(i)}=(A_i\ \ c_i)^T$. To avoid ambiguity, $\theta^*:=\mathrm{vec}\{(A^*\ \ \boldsymbol{c}^*)\}=(((\theta^*)^{(1)})^T,\ldots,(\theta^*)^{(n)})^T)^T$ is used to represent the true parameters. Given observation data $\{\boldsymbol{s}^k,0\le k\le T\}$, where $\boldsymbol{s}^k$ is the observation vector at time k, the log likelihood function is

$$\begin{aligned}&l(T;\theta)\\ &=\log P\{S_k=\boldsymbol{s}^k,0\le k\le T\}\\ &=\log\prod_{1\le k\le T}P\{S_k=\boldsymbol{s}^k|S_{k-1}=\boldsymbol{s}^{k-1}\}P\{S_0=\boldsymbol{s}^0\}\\ &=\log P\{S_0=\boldsymbol{s}^0\}+\sum_{1\le k\le T}\log P\{S_k=\boldsymbol{s}^k|S_{k-1}=\boldsymbol{s}^{k-1}\}\\ &=\log P\{S_0=\boldsymbol{s}^0\}+\sum_{1\le k\le T}\sum_{1\le i\le n}\log g_i(\tilde{\boldsymbol{s}}^k|\theta^{(i)}),\end{aligned}\tag{4}$$

where

$$g_i(\tilde{\boldsymbol{x}}|\theta^{(i)}):=(1-\Phi(c_i-A_ih_L(\tilde{\boldsymbol{x}})))^{\tilde{x}_i}\Phi(c_i-A_ih_L(\tilde{\boldsymbol{x}}))^{1-\tilde{x}_i},\tag{5}$$

$\tilde{\boldsymbol{x}}\in\mathcal{S}^{2n}$, and $(\tilde{\boldsymbol{s}}^t)^T:=((\boldsymbol{s}^t)^T(\boldsymbol{s}^{t-1})^T)$.

For fixed θ, $g_i(\tilde{\boldsymbol{x}}|\theta^{(i)})$ and $\nabla_{\theta^{(i)}}g_i(\tilde{\boldsymbol{x}}|\theta^{(i)})$ are bounded since $\tilde{\boldsymbol{x}}$ takes values in $\mathcal{S}^{2n}$. Thus, from the ergodicity of Markov chains (Theorem 17.1.7 in [28]), the following equations hold for the chain $\{\tilde{S}_k\}$ and fixed θ almost surely(a.s.):

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum\nolimits_{1\le k\le T}\sum\nolimits_{1\le i\le n}\log g_i(\tilde{S}_k|\theta^{(i)})\\ &=E\Big[\sum\nolimits_{1\le i\le n}\log g_i(\tilde{S}|\theta^{(i)})\Big],\end{aligned}$$

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum\nolimits_{1\le k\le T}\sum\nolimits_{1\le i\le n}\nabla_{\theta^{(i)}}\log g_i(\tilde{S}_k|\theta^{(i)})\\ &=E\Big[\sum\nolimits_{1\le i\le n}\nabla_{\theta^{(i)}}\log g_i(\tilde{S}|\theta^{(i)})\Big],\end{aligned}$$

where $\tilde{S}$ is subject to the stationary distribution of $\{\tilde{S}_k\}$.

Therefore, the function of θ

$$E\Big[\sum\nolimits_{1\le i\le n}\log g_i(\tilde{S}|\theta^{(i)})\Big]\tag{6}$$

will be used as an objective function to fulfill the estimation of θ^*. It has a good property:

Theorem 4: (*Strictly concavity of* (6)) Under Assumption 1, the function (6) is strictly concave with respect to θ over $\mathbb{R}^{n(n+1)}$, and the true parameter vector θ^* is the unique maximum point of (6).

Proof: Because of the space limitation, we only provide a sketch of the proof. For the first step, one need to show that the derivative can be passed under the expectation for both (6) and its gradient. Next, we have to show that the Hessian of (6) is negative definite over $\mathbb{R}^{n(n+1)}$. Finally, that θ^* is a maximum point of (6) needs to be proved. ■

Remark 1: This theorem is the key result of our paper. It shows that under the independent Gaussian assumption the network weight estimation algorithm can be transformed to an optimization problem, which can be addressed by using standard methods such as stochastic approximation algorithms.

Therefore, our estimation task turns to seeking the unique maximum point of this strictly concave function. However, $\tilde{S}$ cannot be directly obtained, so the observations $\{\tilde{S}_k\}$ are used to replace it. An SA algorithm is introduced in the next section, and it is verified that the true network can be indeed estimated by using the observation sequence.

B. Network Estimation Algorithm

We use the SA algorithm to deal with the estimation problem. For $1\le i\le n$ and $k\ge 1$, denote

$$K_i(\theta^{(i)},\tilde{S}_{k+1}):=\nabla_{\theta^{(i)}}\log g_i(\tilde{S}_{k+1}|\theta^{(i)}),$$

$$K(\theta,\tilde{S}_{k+1}):=(K_1(\theta^{(1)},\tilde{S}_{k+1}),\ldots,K_n(\theta^{(n)},\tilde{S}_{k+1}))^T,$$

where $\theta^T=((\theta^{(1)})^T,\ldots,(\theta^{(n)})^T)$, and g_i is defined in (5).

The estimation algorithm is as follows:

$$\theta_{k+1}=\theta_k+a_kK(\theta_k,\tilde{S}_{k+1}),\tag{7}$$

where $\theta_k^T=((\theta_k^{(1)})^T,\ldots,(\theta_k^{(n)})^T)$ is the estimate of θ^* at time k, which is the root of $E\{K(\theta,\tilde{S})\}$ as in Theorem 4, and a_k is the step size.

Remark 2: In this algorithm, we assume that θ_k is bounded. If this assumption does not hold, one can apply stochastic approximation algorithms with expanding truncations [30], in which the estimate θ_k is also bounded because of truncations. It is also verified that the number of truncations is finite a.s.

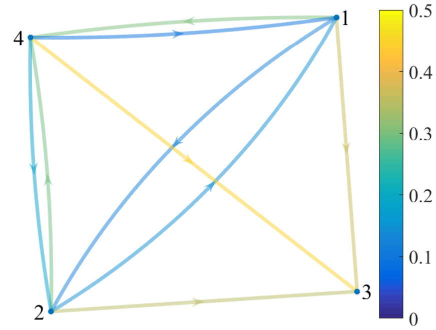

Fig. 1. The true network.

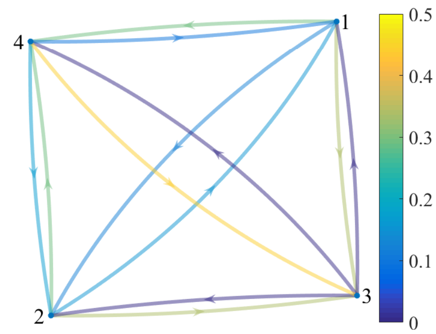

Fig. 2. The estimated network.

C. Strong Consistency of Estimation

In this section we show the strong consistency of the proposed algorithm, i.e., the estimate sequence converges to the true parameter vector with probability one. First, we introduce the following step-size condition, which is standard for SA algorithms.

***Assumption** 2:* Let a_k be the step size in (7), satisfying $a_k > 0$, $\sum_{k=1}^{\infty} a_k = \infty$, and $\sum_{k=1}^{\infty} a_k^2 < \infty$.

Under Assumptions 1 and 2, we have the following strong consistency result, indicating that the algorithm (7) converges to the true parameter vector θ^*.

***Theorem** 5:* (*Strong consistency*) Suppose that Assumptions 1 and 2 hold. Then the estimates θ_k of the algorithm (7) converges to θ^* a.s. from any fixed initial value, i.e.,

$$P\left\{\lim_{k\to\infty} \theta_k = \theta^*\right\} = 1,$$

where θ^* is the true parameter vector.

Proof: Because of the limitation of space, we omit the proof here. The theorem is verified mainly by validating the conditions of Theorem 2.5.1 in [30]. ■

D. NUMERICAL SIMULATIONS

We use an influence weight matrix with four individuals from an empirical study [31] to illustrate the consistency of the above algorithm. The weight matrix $\tilde{A}$ is given by

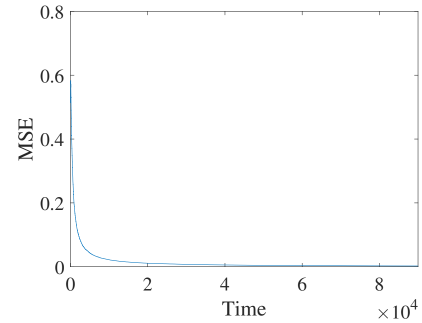

Fig. 3. The MSE of the proposed algorithm.

$$\tilde{A} = \begin{bmatrix} 0.220 & 0.120 & 0.360 & 0.300 \\ 0.147 & 0.215 & 0.344 & 0.294 \\ 0 & 0 & 1 & 0 \\ 0.090 & 0.178 & 0.446 & 0.286 \end{bmatrix}.$$

The noise is set to be independent white Gaussian with zero mean and variance 4, and $\tilde{c}$ is randomly selected as $\tilde{\boldsymbol{c}} = (0.13 \ 0.28 \ 0.08 \ 0.24)^T$. Therefore, as previous discussions, the parameters are identical to that $\boldsymbol{c} = (0.065 \ 0.14 \ 0.04 \ 0.12)^T$ and $A = \tilde{A}/2$ in our model.

We set the step size $a_k = 10/(k+200)$, and run the algorithm for 100 trials. In Figs. 1 and 2, the true network and the estimated network are presented with self-loop omitted. The two figures illustrate that the estimated network is close to the true one. It should also be noted that the estimates of a_{31}, a_{32}, and a_{34} are close to 0, indicating the network has no corresponding edges. Fig. 3 shows the mean square error (MSE), which at time k is defined as $\frac{1}{N}\sum_{i=1}^{N} \|\theta_k(i) - \theta^*\|^2$ with the number of trials $N = 100$, where $\theta_k(i)$ is the estimate of the i-th trial at time k.

VI. CONCLUSION

In this paper we study the estimation of network weights for a class of binary observation systems. These systems are distinctly different from models studied in the literature of quantized identification, because there is no room for the design of inputs and quantizers. We propose a recursive algorithm based on stochastic approximation techniques, and prove its consistency. Future work includes investigation of the convergence rate and asymptotical efficiency, generalization of the model and noise conditions, for example, discrete disturbances, and applications of the algorithm in practice.

APPENDIX

***Proof of Theorem 1**:*

Under Assumption 1, the probability transition matrix can be obtained via the following way:

$$\begin{aligned}
& P\{S_1 = \boldsymbol{s} | S_0 = \boldsymbol{u}\} \\
=\ & P\{A_i S_0 + D_{1,i} > c_i, \forall i \text{ s.t } s_i = 1, A_j S_0 + D_{1,j} \le c_j, \\
& \quad \forall j \text{ s.t. } s_j = 0 | S_0 = \boldsymbol{u}\} \\
=\ & P\{A_i \boldsymbol{u} + D_{1,i} > c_i, \forall i \text{ s.t } s_i = 1, A_j \boldsymbol{u} + D_{1,j} \le c_j, \\
& \quad \forall j \text{ s.t. } s_j = 0 | S_0 = \boldsymbol{u}\}
\end{aligned}$$

$$= P\{A_i\boldsymbol{u} + D_{1,i} > c_i, \forall i \text{ s.t } s_i = 1, A_j\boldsymbol{u} + D_{1,j} \le c_j, \\ \forall j \text{ s.t. } s_j = 0\}$$

$$= \prod_{1\le i\le n} (1 - \Phi(c_i - A_i\boldsymbol{u}))^{s_i}\Phi(c_i - A_i\boldsymbol{u})^{1-s_i} > 0, \quad (8)$$

for all $\boldsymbol{s}, \boldsymbol{u} \in \mathcal{S}^n$, $1 \le i \le n$. Therefore, the transition matrix of $\{S_k\}$ is irreducible and aperiodic, and the conclusion holds by Corollary 1.17 and Theorem 4.9 in [32]. □

Proof of Theorem 3:

From (8) in the proof of Theorem 1, we have the following

$$P\{S_1 = \boldsymbol{e}_i | S_0 = \boldsymbol{e}_j\} = (1 - \Phi(c_i - a_{ij}))\prod_{l\ne i}\Phi(c_l - a_{lj}),$$

$$P\{S_1 = \boldsymbol{0} | S_0 = \boldsymbol{e}_j\} = \prod_{1\le l\le n}\Phi(c_l - a_{lj}),$$

$$P\{S_1 = \boldsymbol{e}_i | S_0 = \boldsymbol{e}_j + \boldsymbol{e}_k\} = (1 - \Phi(c_i - a_{ij} - a_{ik}))\cdot \\ \prod_{l\ne i}\Phi(c_l - a_{lj} - a_{lk})$$

$$P\{S_1 = \boldsymbol{0} | S_0 = \boldsymbol{e}_j + \boldsymbol{e}_k\} = \prod_{1\le l\le n}\Phi(c_l - a_{lj} - a_{lk}),$$

where $1 \le i, j, k \le n$ and $k \ne j$, and the same for $\{\hat{S}_t\}$. Here Φ is the c.d.f. of standard normal distribution.

Suppose that $\{S_t\}$ and $\{\hat{S}_t\}$ have the same probability transition matrices. From Assumption 1 and the above equations, it follows that

$$\Phi(c_i - a_{ij}) = \Phi(\hat{c}_i - \hat{a}_{ij}) \\ \Phi(c_i - a_{ij} - a_{ik}) = \Phi(\hat{c}_i - \hat{a}_{ij} - \hat{a}_{ik}),$$

where $1 \le i, j, k \le n$ and $k \ne j$. Hence by the strictly increasing property of Φ,

$$c_i - a_{ij} = \hat{c}_i - \hat{a}_{ij} \\ c_i - a_{ij} - a_{ik} = \hat{c}_i - \hat{a}_{ij} - \hat{a}_{ik},$$

where $1 \le i, j, k \le n$ and $k \ne j$. Therefore, if we set $j = k + 1$ when $k < n$, and $j = 1$ when $k = n$, then we have for all $i, k \in \{1, \ldots, n\}$, $a_{ik} = \hat{a}_{ik}$. Consequently $c_i = \hat{c}_i$ holds for all $i \in \{1, \ldots, n\}$. □

References

[1] P. Dhaeseleer, S. Liang, and R. Somogyi, "Genetic network inference: from co-expression clustering to reverse engineering," *Bioinformatics*, vol. 16, no. 8, pp. 707–726, 2000.

[2] C. Ravazzi, R. Tempo, and F. Dabbene, "Learning influence structure in sparse social networks," vol. 5, no. 4, pp. 1976–1986, 2017.

[3] M. Timme and J. Casadiego, "Revealing networks from dynamics: an introduction," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 34, p. 343001, 2014.

[4] E. Nozari, Y. Zhao, and J. Cortés, "Network identification with latent nodes via autoregressive models," vol. 5, no. 2, pp. 722–736, 2018.

[5] M. Sharf and D. Zelazo, "Network identification: A passivity and network optimization approach," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 2107–2113, IEEE, 2018.

[6] S. Segarra, M. T. Schaub, and A. Jadbabaie, "Network inference from consensus dynamics," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 3212–3217, IEEE, 2017.

[7] C. Ravazzi, S. Hojjatinia, C. M. Lagoa, and F. Dabbene, "Randomized opinion dynamics over networks: influence estimation from partial observations," in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 2452–2457, IEEE, 2018.

[8] Y. Dong, W. Zhao, and Y. Xing, "The identification of social networks by the least-square algorithm," in *2018 37th Chinese Control Conference (CCC)*, pp. 1931–1936, IEEE, 2018.

[9] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[10] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[11] P. Frasca, R. Carli, F. Fagnani, and S. Zampieri, "Average consensus on networks with quantized communication," *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, vol. 19, no. 16, pp. 1787–1816, 2009.

[12] F. Ceragioli and P. Frasca, "Consensus and disagreement: The role of quantized behaviors in opinion dynamics," *SIAM Journal on Control and Optimization*, vol. 56, no. 2, pp. 1058–1080, 2018.

[13] L. Y. Wang, G. G. Yin, J.-F. Zhang, and Y. Zhao, *System Identification with Quantized Observations*. Springer, 2010.

[14] J. Guo, L. Y. Wang, G. Yin, Y. Zhao, and J.-F. Zhang, "Asymptotically efficient identification of FIR systems with quantized observations and general quantized inputs," *Automatica*, vol. 57, pp. 113–122, 2015.

[15] J. Guo and Y. Zhao, "Recursive projection algorithm on FIR system identification with binary-valued observations," *Automatica*, vol. 49, no. 11, pp. 3396–3401, 2013.

[16] T. Wang, M. Hu, and Y. Zhao, "Convergence properties of recursive projection algorithm for system identification with binary-valued observations," in *2018 Chinese Automation Congress (CAC)*, pp. 2961–2966, IEEE, 2018.

[17] D. Marelli, K. You, and M. Fu, "Identification of ARMA models using intermittent and quantized output observations," *Automatica*, vol. 49, no. 2, pp. 360–369, 2013.

[18] W. Zhao, H. Chen, R. Tempo, and F. Dabbene, "Recursive nonparametric identification of nonlinear systems with adaptive binary sensors," vol. 62, no. 8, pp. 3959–3971, 2017.

[19] B. I. Godoy, G. C. Goodwin, J. C. Agüero, D. Marelli, and T. Wigren, "On identification of FIR systems having quantized output data," *Automatica*, vol. 47, no. 9, pp. 1905–1915, 2011.

[20] J. C. Agüero, K. González, and R. Carvajal, "EM-based identification of ARX systems having quantized output data," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 8367–8372, 2017.

[21] Q. Song, "Recursive identification of systems with binary-valued outputs and with ARMA noises," *Automatica*, vol. 93, pp. 106–113, 2018.

[22] W. Pan, W. Dong, M. Cebrian, T. Kim, J. H. Fowler, and A. S. Pentland, "Modeling dynamical influence in human interaction: Using data to make better inferences about influence within social systems," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 77–86, 2012.

[23] B. I. Godoy, J. C. Agüero, R. Carvajal, G. C. Goodwin, and J. I. Yuz, "Identification of sparse FIR systems using a general quantisation scheme," *International Journal of Control*, vol. 87, no. 4, pp. 874–886, 2014.

[24] M. Granovetter, "Threshold models of collective behavior," *American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.

[25] D. J. Watts, "A simple model of global cascades on random networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, 2002.

[26] T. C. Schelling, "Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities," *Journal of Conflict Resolution*, vol. 17, no. 3, pp. 381–428, 1973.

[27] L. Blume and S. Durlauf, "Equilibrium concepts for social interaction models," *International Game Theory Review*, vol. 5, no. 03, pp. 193–209, 2003.

[28] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Springer Science & Business Media, 2012.

[29] L. Ljung, *System Identification: Theory for the User*. Prentice-hall, 1987.

[30] H.-F. Chen, *Stochastic Approximation and Its Applications*. Kluwer, Boston, MA, 2002.

[31] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.

[32] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, vol. 107. American Mathematical Soc., 2017.