Optimal Recursive Expert-Enabled Inference in Regulatory Networks

Amirhossein Ravari, Seyede Fatemeh Ghoreishi and Mahdi Imani

***Abstract*— Accurate inference of biological systems, such as gene regulatory networks and microbial communities, is a key to a deep understanding of their underlying mechanisms. Despite several advances in the inference of regulatory networks in recent years, the existing techniques cannot incorporate expert knowledge into the inference process. Expert knowledge contains valuable biological information and is often reflected in available biological data, such as interventions made by biologists for treating diseases. Given the complexity of regulatory networks and the limitation of biological data, ignoring expert knowledge can lead to inaccuracy in the inference process. This paper models the regulatory networks using Boolean network with perturbation. We develop an expert-enabled inference method for inferring the unknown parameters of the network model using expert-acquired data. Given the availability of information about data-acquiring objectives and expert confidence, the proposed method optimally quantifies the expert knowledge along with the temporal changes in the data for the inference process. The numerical experiments investigate the performance of the proposed method using the well-known p53-MDM2 gene regulatory network.**

***Index Terms*— Gene Regulatory Networks, Inference, Inverse Reinforcement Learning, Boolean networks.**

I. Introduction

Regulatory networks in systems biology are comprised of a large number of interacting components, such as bacteria, microbes, genes, and small molecules in gene regulatory networks and microbial communities [1], [2]. The interactions between elements of these regulatory networks control the ecosystem functioning and various cellular processes, such as stress response, DNA repair, and other mechanisms involved in complex diseases such as cancer [3]–[5]. Experts play critical roles in most practical analyses and decision-making of these biological systems. Experts' decisions are often reflected in available biological data, such as data collected during biological interventions for disease treatment and experiment perturbation for hypothesis testing. The expert decisions reflect the valuable expert perception of the complex biological systems, which, if quantified, can significantly help a deep understanding of their mechanism.

Several techniques have been developed in recent years for the inference of regulatory networks [6]–[8]. These methods aim to infer the unknown parameters of a regulatory network model using the available biological data. These include methods developed based on maximum likelihood, maximum aposteriori, or similar measures [9]–[11]. The complexity of the regulatory networks and the limitation in biological data (due to the huge cost associated with data acquiring) often prevent accurate inference using the existing techniques. Meanwhile, biological data are often collected so that their temporal changes might not elicit the interactions between components.

This paper investigates the incorporation of expert knowledge embedded in expert-acquired data to overcome data limitations and lack of excitation impacting the inference of regulatory networks. We model the regulatory networks with the Boolean network with perturbation [2], [12]. This model is shown to be effective in capturing the complex dynamics of regulatory networks, such as yeast and mammalian cell cycle networks, and gut microbial communities [13], [14]. Unlike the conventional inference techniques that aim to learn the unknown parameters of regulatory networks by investigating the temporal changes of data, we develop an expert-enabled inference method that incorporates the expert knowledge, which appears in terms of decisions/actions in available biological data, in the inference process. The exact optimal expert-enabled inference is achieved in domains with known expert confidence and data-acquiring objectives.

In particular, we study the inference in regulatory networks with missing interactions between the components, where the possible network models can be represented by a finite set. Without loss of generality, the maximum aposteriori (MAP) criterion is considered for the inference process. We derive the exact recursive and optimal expert-enabled MAP inference method, which consists of offline dynamic programming methods that run in parallel, along with online quantification of the expert knowledge and temporal changes in data upon observing a new measurement. We investigate the impact of expert confidence on the performance of the proposed method analytically and empirically. The performance of the proposed method in terms of accuracy is demonstrated through numerical experiments using the p53-MDM2 negative feedback loop network.

II. Mathematical Preliminaries

A. Regulatory Network Model

This paper employs the Boolean network with perturbation (BNp) model [2], [12], [15] for capturing the dynamics of regulatory networks. This model properly captures the stochasticity in regulatory networks, coming from intrinsic uncertainty or unmodeled parts of systems. Consider a system consisting of d components. The *state process* can be expressed as $\{\mathbf{x}_k; k = 0, 1, \ldots\}$, where $\mathbf{x}_k \in \{0, 1\}^d$ representing the

A. Ravari and M. Imani are with the Department of Electrical and Computer Engineering, and S.F. Ghoreishi is with the Department of Civil and Environmental Engineering and Khoury College of Computer Sciences at Northeastern University. Emails: ravari.a@northeastern.edu, f.ghoreishi@northeastern.edu, m.imani@northeastern.edu

activation/inactivation state of the components at time k. The component's state is updated at each discrete time through the following Boolean signal model:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}) \oplus \mathbf{u}_{k-1} \oplus \mathbf{n}_k\,, \tag{1}$$

for $k = 1, 2, \ldots$, where $\mathbf{u}_{k-1} \in \mathcal{U} \subset \{0,1\}^d$ is an external action (e.g., intervention, perturbation, etc.) at time step $k-1$, $\mathbf{n}_k \in \{0,1\}^d$ is a noise process at time k, "$\oplus$" denotes component-wise binary addition (the exclusive-or logic operation), and $\mathbf{f}$ is the network function. The noise process is modeled through d independent Bernoulli processes with parameter p (i.e., $0 \leq p \leq 0.5$) as: $\mathbf{n}_k(i) \sim \text{Bernoulli}(p)$, for $i = 1, ..., d$. The small values of p represent less noisy Boolean network models.

The network function $\mathbf{f}(.)$ can be expressed through a Boolean logic or pathway diagram models [11], [12], [14]. Without loss of generality, we consider the pathway diagram representation. According to this model, $\mathbf{f} = (f_1, \ldots, f_d)$ is a component form of Boolean function with the ith element as:

$$f_i(\mathbf{x}) = \begin{cases} 1, & \sum_{j=1}^{d} c_{ij}\mathbf{x}(j) + b_i > 0\,, \\ 0, & \sum_{j=1}^{d} c_{ij}\mathbf{x}(j) + b_i \leq 0\,, \end{cases} \tag{2}$$

where c_{ij} takes in $\{-1, 0, +1\}$ and represents the type of interaction from component j to component i, and b_i is a bias unit for component i, which takes a value in $\{-1/2, +1/2\}$. The $c_{ij} = +1$ represents a positive interaction, $c_{ij} = -1$ denotes a negative interaction, and $c_{ij} = 0$ corresponds to no interaction case. Meanwhile, the bias units are tie-breaking parameters that control the activation and inactivation of components in case of an equal number of activation and inactivation inputs.

It should be noted that the models denoted in (1) and (2) are well-known models for regulatory networks and are used here for the sake of illustration. Without loss of generality, the rest of the paper holds for a general class of Boolean network model of form $\mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{n}_k)$.

B. Optimal Policy in Regulatory Networks

The regulatory network model in (1) can be represented by a Markov decision process (MDP) with 5-tuple $\langle \mathcal{X}, \mathcal{U}, T, R, \gamma \rangle$, where $\mathcal{X} = \{0,1\}^d$ is the *state space*, $\mathcal{U}$ is the *action space*, $T : \mathcal{X} \times \mathcal{U} \times \mathcal{X}$ is the *state transition probability function* such that $p(\mathbf{x}' \mid \mathbf{x}, \mathbf{u})$ represents the probability of moving to state $\mathbf{x}'$ after taking action $\mathbf{u}$ in state $\mathbf{x}$, $R : \mathcal{X} \times \mathcal{U} \times \mathcal{X} \to \mathbb{R}$ is a bounded *reward function* such that $R(\mathbf{x}, \mathbf{u}, \mathbf{x}')$ encodes the reward earned when action $\mathbf{u}$ in taken in state $\mathbf{x}$ and system moves to state $\mathbf{x}'$, and $0 < \gamma < 1$ is a *discount factor*.

More formally, a deterministic stationary policy $\pi : \mathcal{X} \to \mathcal{U}$ for an MDP is a mapping from states to actions. The expected discounted reward function at state $\mathbf{x} \in \mathcal{X}$ after taking action $\mathbf{u} \in \mathcal{U}$ and the following policy π afterward is defined as:

$$Q^{\pi}(\mathbf{x}, \mathbf{u}) = \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t R(\mathbf{x}_t, \mathbf{u}_t, \mathbf{x}_{t+1}) \mid \mathbf{x}_0 = \mathbf{x}, \mathbf{u}_0 = \mathbf{u}, \mathbf{u}_{1:\infty} \sim \pi \Big]. \tag{3}$$

The optimal Q-function can be expressed as $Q^*(\mathbf{x}, \mathbf{u}) := Q^{\pi^*}(\mathbf{x}, \mathbf{u})$, which indicates the expected discounted reward after executing action $\mathbf{u}$ in state $\mathbf{x}$ and following optimal policy π^* afterward. An optimal (stationary) policy π^* attains the maximum expected return for all states as: $\pi^*(\mathbf{x}) = \text{argmax}_{\mathbf{u} \in \mathcal{U}} Q^*(\mathbf{x}, \mathbf{u})$, for $\mathbf{x} \in \mathcal{X}$.

III. Proposed Framework

A. Problem Formulation

Let $D_k = \{(\tilde{\mathbf{u}}_0, \tilde{\mathbf{x}}_1), (\tilde{\mathbf{u}}_1, \tilde{\mathbf{x}}_2), ..., (\tilde{\mathbf{u}}_{k-1}, \tilde{\mathbf{x}}_k)\}$ be the expert-acquired data up to time step k, where $\tilde{\mathbf{u}}_{r-1}$ is the taken action by an expert at time step $r-1$, which has led the system state to move from $\tilde{\mathbf{x}}_{r-1}$ to state $\tilde{\mathbf{x}}_r$. Let $R(.,.,.)$ represent the expert reward function, which denotes the reward associated with the main biological objectives. For instance, if the actions are interventions for treating diseases, the immediate reward measures the improvement achieved at any given step during the intervention process. Assuming $Q^*(.,.)$ to be the optimal Q-value for the expert reward function, we model the expert decisions/actions according to the following softmax policy [16]:

$$p(\mathbf{u} \mid \mathbf{x}) \propto \exp\left(\eta\, Q^*(\mathbf{x}, \mathbf{u})\right), \tag{4}$$

for $\mathbf{x} \in \mathcal{X}$ and $\mathbf{u} \in \mathcal{U}$; where $\eta > 0$ represents the confidence of expert. This softmax policy denotes that the expert might not take the optimal action (i.e., $\text{argmax}_{\mathbf{u} \in \mathcal{U}} Q^*(\mathbf{x}, \mathbf{u})$), as its information about the system dynamics or reward might be limited. Large values of η model more confident experts, whereas smaller values model more imperfect experts. The expert policy in (4) is widely used in the inverse reinforcement learning context for modeling expert/human behavior [17]–[19]. Another well-known policy to represent expert behavior is the ϵ-greedy policy [17]. According to this model, the expert policy at state $\mathbf{x}$ can be expressed as:

$$p(\mathbf{u} \mid \mathbf{x}) = \begin{cases} q + \frac{1-q}{|\mathcal{U}|} & \text{If } \mathbf{u} = \underset{\mathbf{u}' \in \mathcal{U}}{\text{argmax}} Q^*(\mathbf{x}, \mathbf{u}') \\ \frac{1-q}{|\mathcal{U}|} & \text{If } \mathbf{u} \neq \underset{\mathbf{u}' \in \mathcal{U}}{\text{argmax}} Q^*(\mathbf{x}, \mathbf{u}') \end{cases}, \text{ for } \mathbf{u} \in \mathcal{U}, \tag{5}$$

where q represents the expert confidence, taking its value in $0 \leq q \leq 1$; $q = 1$ and $q = 0$ represent an optimal and random expert, respectively. The expert model in (4) benefits from differentiability, whereas the expert model in (5) allows easier quantification of the expert confidence through parameter q. Without loss of generality, the rest of the paper considers the expert model in (4).

The expert-enabled inference consists of estimating the unknown parameters of the network model according to the available expert-acquired data. The unknown parameters in regulatory networks are often missing interactions between different components. The discrete nature of these parameters (i.e., c_{ij} or b_i parameters in model (2)) leads to a finite set of possible models for regulatory networks. For instance, given that n_1 regulatory interactions (i.e., c_{ij}) and n_2 bias units (i.e., b_i) are unknown, there are $M = 3^{n_1} \times 2^{n_2}$ different possible models. These models are represented by $\Theta = \{\boldsymbol{\theta}^1, ..., \boldsymbol{\theta}^M\}$, which correspond to M different regulatory network models.

Without loss of generality, we consider the maximum aposteriori (MAP) criterion for the inference process. The expert-enabled MAP inference given the expert-acquired data up to time step k (i.e., D_k) can be formulated as:

$$\begin{aligned}\hat{\boldsymbol{\theta}}_k^{\mathrm{EE-MAP}} &= \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \log P(\boldsymbol{\theta} \mid D_k) \\ &= \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \left[\log P(\boldsymbol{\theta}) + \log P(D_k \mid \boldsymbol{\theta})\right],\end{aligned} \tag{6}$$

where $P(.)$ is a probability mass function, and $P(\boldsymbol{\theta})$ and $P(D_k \mid \boldsymbol{\theta})$ denote the prior probability and the expert-enabled likelihood value for model $\boldsymbol{\theta}$, respectively. The prior probability for model $\boldsymbol{\theta}$ takes in $0 \leq P(\boldsymbol{\theta}) \leq 1$, with $\sum_{\boldsymbol{\theta}\in\Theta} P(\boldsymbol{\theta}) = 1$. If no prior information about the models is available, the non-informative (i.e., uniform) prior distribution can be considered, i.e., $P(\boldsymbol{\theta}) = 1/M$, for $\boldsymbol{\theta} \in \Theta$.

The logarithm of expert-enabled likelihood function, i.e., $\log P(D_k \mid \boldsymbol{\theta})$, can be expressed as:

$$\begin{aligned}&L_k^{\mathrm{EE}}(\boldsymbol{\theta}) := \log P(D_k \mid \boldsymbol{\theta}) \\ &= \log \prod_{r=1}^{k} P(\tilde{\mathbf{x}}_r, \tilde{\mathbf{u}}_{r-1} \mid \tilde{\mathbf{x}}_{1:r-1}, \tilde{\mathbf{u}}_{0:r-2}, \boldsymbol{\theta}) \\ &= \underbrace{\sum_{r=1}^{k} \log P(\tilde{\mathbf{x}}_r \mid \tilde{\mathbf{x}}_{r-1}, \tilde{\mathbf{u}}_{r-1}, \boldsymbol{\theta})}_{\text{State-Transition Term}} + \underbrace{\sum_{r=1}^{k} \log P(\tilde{\mathbf{u}}_{r-1} \mid \tilde{\mathbf{x}}_{r-1}, \boldsymbol{\theta})}_{\text{Expert-Knowledge Term}}.\end{aligned} \tag{7}$$

The last line in (7) is obtained using the Markovian property of the state transition in the Boolean network model in (1) and the softmax policy representing the expert behavior in (4). The existing inference techniques only consider the state-transition term in (7) for the inference process and ignore the expert-knowledge term. We will analyze the expert-knowledge term's impact on the inference performance in the next section and empirically demonstrate the high performance of the proposed expert-enabled inference method in the numerical experiments.

B. Recursive Formulation of Optimal Expert-Enabled Inference

This section describes the recursive formulation of the optimal expert-enabled MAP inference in (6). We start with a recursive representation of the expert-enabled log-likelihood function in (7) as:

$$\begin{aligned}&L_k^{\mathrm{EE}}(\boldsymbol{\theta}) = L_{k-1}^{\mathrm{EE}}(\boldsymbol{\theta}) \\ &+ \underbrace{\log P(\tilde{\mathbf{x}}_k \mid \tilde{\mathbf{x}}_{k-1}, \tilde{\mathbf{u}}_{k-1}, \boldsymbol{\theta})}_{l_k^D(\boldsymbol{\theta}):=\text{State-Transition Increment}} + \underbrace{\log P(\tilde{\mathbf{u}}_{k-1} \mid \tilde{\mathbf{x}}_{k-1}, \boldsymbol{\theta})}_{l_k^E(\boldsymbol{\theta}):=\text{Expert-Knowledge Increment}},\end{aligned} \tag{8}$$

where $L_{k-1}^{\mathrm{EE}}(\boldsymbol{\theta})$ is the previous expert-enabled log-likelihood function given D_{k-1}, and the state-transition and the expert-knowledge increments represent the addition to the likelihood after observing the last expert-acquired data, i.e., $(\tilde{\mathbf{u}}_{k-1}, \tilde{\mathbf{x}}_k)$.

Therefore, the optimal recursive formulation of the expert-enabled MAP inference in (6) can be expressed through:

$$\hat{\boldsymbol{\theta}}_k^{\mathrm{EE-MAP}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \Big[\log P(\boldsymbol{\theta}) + L_{k-1}^{\mathrm{EE}}(\boldsymbol{\theta}) + l_k^D(\boldsymbol{\theta}) + l_k^E(\boldsymbol{\theta})\Big]. \tag{9}$$

C. Exact Computation of State-Transition and Expert-Knowledge Increments

This section describes the computation of state-transition and expert-knowledge increments for the exact solution of the recursive expert-enabled MAP inference in (9).

1) Computation of State-Transition Increment: The evaluation of this term requires computing the probability of moving from state $\tilde{\mathbf{x}}_{k-1}$ to state $\tilde{\mathbf{x}}_k$ given the expert action $\tilde{\mathbf{u}}_{k-1}$ for all models. For model $\boldsymbol{\theta} \in \Theta$, this can be expressed as:

$$\begin{aligned}&l_k^D(\boldsymbol{\theta}) = \log P(\tilde{\mathbf{x}}_k \mid \tilde{\mathbf{x}}_{k-1}, \tilde{\mathbf{u}}_{k-1}, \boldsymbol{\theta}) \\ &= \log\left[p^{||\mathbf{f}^{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_{k-1})\oplus\tilde{\mathbf{u}}_{k-1}\oplus\tilde{\mathbf{x}}_k||_1}(1-p)^{d-||\mathbf{f}^{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_{k-1})\oplus\tilde{\mathbf{u}}_{k-1}\oplus\tilde{\mathbf{x}}_k||_1}\right] \\ &= ||\mathbf{f}^{\boldsymbol{\theta}}(\tilde{\mathbf{x}}_{k-1}) \oplus \tilde{\mathbf{u}}_{k-1} \oplus \tilde{\mathbf{x}}_k||_1 \log\frac{p}{1-p} + d\log(1-p),\end{aligned} \tag{10}$$

where $||.||_1$ is the absolute L-1 norm of a vector, and p is the parameter of the Bernoulli noise process in (1).

2) Computation of Expert-Knowledge Increment: Let $Q_{\boldsymbol{\theta}}^*(.,.)$ be the optimal model-specific Q-function for the network model parameterized by $\boldsymbol{\theta}$. Given that the softmax policy in (4) represents the expert policy, the expert-knowledge increment in (8) can be expressed as:

$$\begin{aligned}l_k^E(\boldsymbol{\theta}) &= \log P(\tilde{\mathbf{u}}_{k-1} \mid \tilde{\mathbf{x}}_{k-1}, \boldsymbol{\theta}) \\ &= \log\left[\frac{\exp\left(\eta\, Q_{\boldsymbol{\theta}}^*(\tilde{\mathbf{x}}_{k-1}, \tilde{\mathbf{u}}_{k-1})\right)}{\sum_{\mathbf{u}\in\mathcal{U}} \exp\left(\eta\, Q_{\boldsymbol{\theta}}^*(\tilde{\mathbf{x}}_{k-1}, \mathbf{u})\right)}\right], \text{ for } \boldsymbol{\theta} \in \Theta.\end{aligned} \tag{11}$$

Computation of the optimal model-specific Q-function for models $\boldsymbol{\theta} \in \Theta$ can be achieved by running parallel dynamic programming methods, such as value iteration and policy iteration [20]. Let $\{\mathbf{x}^1, \ldots, \mathbf{x}^{2^d}\}$ be an arbitrary enumeration of the possible Boolean state vectors. The expert reward function $R(.,.,.)$ can be represented in a matrix-form as:

$$(\mathbf{R}(\mathbf{u}))_{ij} = R\left(\mathbf{x}^j, \mathbf{u}, \mathbf{x}^i\right), \text{ for } i, j = 1, .., 2^d, \mathbf{u} \in \mathcal{U}. \tag{12}$$

We also define the *controlled transition matrix* of size $2^d \times 2^d$ associated with a Boolean network parameterized by $\boldsymbol{\theta}$ as:

$$\begin{aligned}(M_{\boldsymbol{\theta}}(\mathbf{u}))_{ij} &= P\left(\mathbf{x}_k = \mathbf{x}^i \mid \mathbf{x}_{k-1} = \mathbf{x}^j, \mathbf{u}_{k-1} = \mathbf{u}, \boldsymbol{\theta}\right), \\ &= p^{||\mathbf{f}^{\boldsymbol{\theta}}(\mathbf{x}^j)\oplus\mathbf{u}\oplus\mathbf{x}^i||_1}(1-p)^{d-||\mathbf{f}^{\boldsymbol{\theta}}(\mathbf{x}^j)\oplus\mathbf{u}\oplus\mathbf{x}^i||_1},\end{aligned} \tag{13}$$

for $i, j = 1, \ldots, 2^d$.

The optimal state value function for any model $\boldsymbol{\theta} \in \Theta$ can be expressed using the Bellman equation as [21], [22]:

$$\mathbf{V}_{\boldsymbol{\theta}}^*(j) = \max_{\mathbf{u}\in\mathcal{U}} \left[\sum_{i=1}^{2^d} (M_{\boldsymbol{\theta}}(\mathbf{u}))_{ij} \left[R(\mathbf{x}^j, \mathbf{u}, \mathbf{x}^i) + \gamma \mathbf{V}_{\boldsymbol{\theta}}^*(i)\right]\right], \tag{14}$$

for $j = 1, .., 2^d$. It is also shown in [21] that $\mathbf{V}_{\boldsymbol{\theta}}^*$ is the unique fixed point solution of the Bellman operator defined as:

$$T[\mathbf{V}_{\boldsymbol{\theta}}](j) = \max_{\mathbf{u}\in\mathcal{U}} \left[\sum_{i=1}^{2^d} (M_{\boldsymbol{\theta}}(\mathbf{u}))_{ij} \left[R(\mathbf{x}^j, \mathbf{u}, \mathbf{x}^i) + \gamma \mathbf{V}_{\boldsymbol{\theta}}(i)\right]\right], \tag{15}$$

for $j = 1, ..., 2^d$. The matrix form formulation of the Bellman operator in (14) can be expressed as:

$$T[\mathbf{V}_{\boldsymbol{\theta}}] = \max_{\mathbf{u}\in\mathcal{U}} \left[(\mathbf{R}(\mathbf{u}) \odot M_{\boldsymbol{\theta}}(\mathbf{u}))^T \mathbf{1}_{2^d\times 1} + \gamma M_{\boldsymbol{\theta}}^T(\mathbf{u})\, \mathbf{V}_{\boldsymbol{\theta}} \right], \tag{16}$$

where "max" is applied row-wise, $\mathbf{1}_{2^d\times 1}$ is a vector of size 2^d with all elements 1, and "$\odot$" denotes the component-wise multiplication of two matrices. It is proven in [21] that for any MDP (i.e., any model $\boldsymbol{\theta} \in \Theta$), the Bellman operator is a γ-contraction mapping, meaning that starting from any arbitrary $\mathbf{V}_{\boldsymbol{\theta}}^0$ and iteratively applying $\mathbf{V}_{\boldsymbol{\theta}}^{t+1} = T[\mathbf{V}_{\boldsymbol{\theta}}^t]$ for $t = 0, 1, ...$ leads to a fixed solution of the Bellman equation. Any fixed-point solution of Belman equation is an optimal solution for the MDP. Thus, for any $0 < \gamma < 1$ and any $\boldsymbol{\theta} \in \Theta$, the value iteration method will converge to the optimal state value function (see [21], for more information).

Let $\mathbf{V}_{\boldsymbol{\theta}}^0 = [0, \cdots, 0]^T$ be the initial value vector for model parameterized by $\boldsymbol{\theta}$. Sequentially performing the Bellman operator in (16) leads to a fixed point solution for the optimal model-specific state value function. Thus, one needs to repeat $\mathbf{V}_{\boldsymbol{\theta}}^{t+1} = T[\mathbf{V}_{\boldsymbol{\theta}}^t]$ till the time that the maximum difference between elements of value vectors in two consecutive iterations falls below a small prespecified threshold, i.e., $\max_{i\in\{1,..,2^d\}} |\mathbf{V}_{\boldsymbol{\theta}}^{t-1}(i) - \mathbf{V}_{\boldsymbol{\theta}}^t(i)| < \epsilon$.

Let $\mathbf{V}_{\boldsymbol{\theta}}^*$ be the fixed-point solution of (16) for model $\boldsymbol{\theta}$. The optimal model-specific Q-function for model $\boldsymbol{\theta}$ can be obtained as:

$$\begin{bmatrix} Q_{\boldsymbol{\theta}}^*(\mathbf{x}^1, \mathbf{u}) \\ \vdots \\ Q_{\boldsymbol{\theta}}^*(\mathbf{x}^{2^d}, \mathbf{u}) \end{bmatrix} = (\mathbf{R}(\mathbf{u}) \odot M_{\boldsymbol{\theta}}(\mathbf{u}))^T \mathbf{1}_{2^d\times 1} + \gamma M_{\boldsymbol{\theta}}^T(\mathbf{u}) \mathbf{V}_{\boldsymbol{\theta}}^*, \tag{17}$$

for $\mathbf{u} \in \mathcal{U}$. Replacing (17) into (11) leads to the exact computation of the expert-knowledge increment.

3) Analysis of Complexity and Impact of Objective and Expert's Confidence: The optimal recursive expert-enabled MAP inference can be achieved through offline and online steps. The offline process consists of running M parallel value iteration methods, each tuned to one specific network model. The computational complexity of the offline process is of order $O(2^{2d} \times |\mathcal{U}| \times M \times T)$, where T is the stopping iteration of the value iteration method. In the online process, the computational complexity of the state transition and expert knowledge increments are of order $O(d\times M)$ and $O(|\mathcal{U}|\times M)$, respectively. Therefore, aside from the offline computations, which can be done before starting the online inference process, the complexity of the online process grows linearly with the number of components in the network, possible network models, and the size of the action space.

The impact of the objective on the performance of the proposed expert-enabled inference method is described through a simple example. Consider a system with two possible network models; $\Theta = \{\boldsymbol{\theta}^1, \boldsymbol{\theta}^2\}$. The data-acquiring objective and the network models impact the optimal model-specific Q-functions in (17). Consider a simple case where two models have the same optimal model-specific Q-functions at states reflected in expert-acquired data (i.e., $Q_{\boldsymbol{\theta}^1}^*(\tilde{\mathbf{x}}_r, \mathbf{u}) = Q_{\boldsymbol{\theta}^2}^*(\tilde{\mathbf{x}}_r, \mathbf{u})$ for $r = 1, ..., k$ and $\mathbf{u} \in \mathcal{U}$); this represents a scenario where the expert would take the same actions if its perception about the model would be either $\boldsymbol{\theta}^1$ or $\boldsymbol{\theta}^2$. The expert-knowledge terms for these two models will be the same, and the expert-enabled inference becomes the same as the conventional inference technique. If given the objective, the model-specific policies are different at states reflected in expert-acquired data (i.e., $\pi_{\boldsymbol{\theta}^1}^*(\tilde{\mathbf{x}}_r) \neq \pi_{\boldsymbol{\theta}^2}^*(\tilde{\mathbf{x}}_r)$, for any $r \in \{1, ..., k\}$), then the expert actions would vary depending on its perception about the system models. Thus, the model that is more likely to create the actions reflected in expert-acquired data takes the largest expert-knowledge term. For large values of η (i.e., more confident experts), the expert-knowledge term becomes more distinguishable for the true model among others. By contrast, for small values of η, the expert-knowledge term becomes similar for various models and less impactful during the inference process. For a specific case of $\eta = 0$, representing the least confident expert (i.e., random action selection), the expert-knowledge term is equal for all models; thus, the expert-enabled MAP inference becomes the same as regular MAP inference.

IV. Numerical Experiments

In this section, the performance of the proposed framework is assessed using the p53-MDM2 Boolean network [23]. All results are averaged over 100 independent runs with random initial states. The parameter settings throughout the numerical experiments are: the process noise $p = 0.05$, the discount factor $\gamma = 0.95$, and the value iteration threshold $\epsilon = 0.01$. Finally, the initial probability for all models is assumed to be equal, i.e., $P(\boldsymbol{\theta}^i) = 1/M$, for $i = 1, ..., M$. The results are represented in terms of the maximum posterior probability and the average error of the inferred parameter, and comparisons are made with the optimal MAP inference technique.

The p53-MDM2 negative feedback loop gene regulatory network plays a critical role in various types of cancers, including the ovarian, esophageal, larynx, and lung [23], [24]. This network includes four genes and the stress input "dna_dsb", which shows the presence of DNA double-strand breaks. The state of four genes can be expressed through vector $\mathbf{x} = (\text{ATM}, \text{p53}, \text{Wip1}, \text{MDM2})$. The pathway diagram for the network is shown in Fig. 1; the solid arrows demonstrate positive interactions (i.e., $c_{ij} = +1$), and the blunt arrows demonstrate negative interactions (i.e., $c_{ij} = -1$). This Boolean model in (2) in stress response can be represented by:

$$\begin{bmatrix} c_{11}=0 & c_{12}=0 & c_{13}=-1 & c_{14}=0 \\ c_{21}=+1 & c_{22}=0 & c_{23}=-1 & c_{24}=-1 \\ c_{31}=0 & c_{32}=+1 & c_{33}=0 & c_{34}=0 \\ c_{41}=-1 & c_{42}=+1 & c_{43}=+1 & c_{44}=0 \end{bmatrix}, \begin{bmatrix} b_1=+1/2 \\ b_2=-1/2 \\ b_3=-1/2 \\ b_4=-1/2 \end{bmatrix}. \tag{18}$$

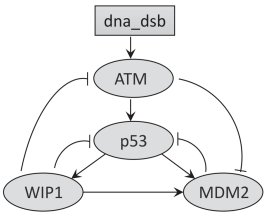

Fig. 1: The pathway diagram for the p53-MDM2 Boolean network.

The network in healthy conditions spends most of its time at state "0000", meaning that all 4 genes stay inactivated (see [25], for more information). This is, however, not the case for the network in cancerous conditions. The stress often leads to unnecessary activation of various genes and, consequently, uncontrolled cell proliferation. Therefore, for our experiments, we assume experts/biologists aim to keep the network at state "0000" through intervention. The intervention is achieved by experts through selecting one of the following control inputs at a time: $\mathcal{U} = \{\mathbf{u}^1 = [0,0,0,0], \mathbf{u}^2 = [1,0,0,0], \mathbf{u}^3 = [0,1,0,0], \mathbf{u}^4 = [0,0,1,0], \mathbf{u}^5 = [0,0,0,1]\}$. Note that $\mathbf{u}^1$ to $\mathbf{u}^5$ flip the value of none or a single gene at a time. This biological intervention can be expressed through the following expert reward function:

$$R(\mathbf{x}, \mathbf{u}, \mathbf{x}') = \frac{1}{4}\Big[- ||\mathbf{x}'||_1 - 0.5||\mathbf{u}||_1 \Big],$$

where the cost of any gene activation is $-1/4$, and the control input altering the gene state has the reward of $-1/8$. The negative reward value for the control represents the expert's desire to take minimum controls due to their potential side effects.

We assume the expert-acquired data are collected using the true network model in (18). We consider the following five interacting parameters to be unknown: c_{21}, c_{32}, c_{42}, c_{13} and c_{43}. Since each interaction can take in $\{-1, 0, +1\}$, there will be $M = 3^5 = 243$ possible network models.

In the first set of experiments, we consider data are acquired by confident experts with $\eta = 10$. The average number of value iteration steps for all models is 61. The average maximum posterior probability and the average connectivity error with respect to the number of expert-acquired data are shown in Fig. 2 and Fig. 3, respectively. The solid black lines correspond to the proposed expert-enabled MAP inference method, and the dashed red lines are associated with the regular MAP inference method. The shaded areas represent 68% confidence intervals for both methods. From Fig. 2, one can see that the proposed method achieves a much higher maximum posterior probability than regular MAP inference. This comes from incorporating the expert knowledge, which has helped to better distinguish between different possible network models. A similar trend can be seen in terms of the RMSE; a much smaller RMSE is obtained by the proposed method compared to the regular MAP inference method, which again demonstrates the importance of incorporating expert knowledge in the inference process.

Assigning an exact value to represent the expert confidence using parameter η in Boltzmann policy is challenging; since, according to (4), setting this value depends on the reward function and discount factor γ. This part of the experiment examines the impact of expert confidence on the performance of the proposed expert-enabled inference method. The average maximum posterior probability and error with respect to the number of data and expert confidence rate are shown in Fig. 4. Smaller average error and higher average maximum posterior probability are obtained by the proposed method under data acquired by more confident experts. The results of the proposed method become close to the conventional

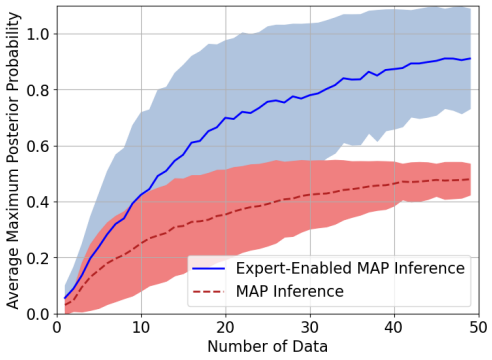

Fig. 2: The average maximum posterior probability for the p53-MDM2 network with 5 unknown regulatory interactions.

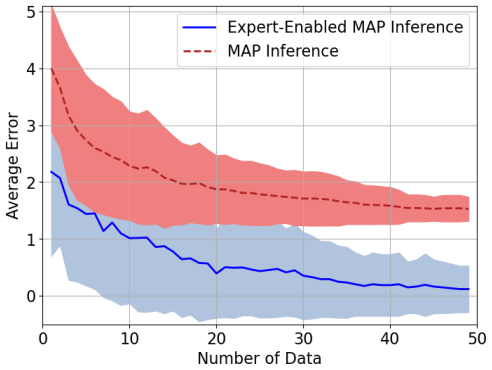

Fig. 3: The average error for the p53-MDM2 network with 5 unknown regulatory interactions.

MAP inference method as η approaches zero (i.e., the expert becomes random). The subplots in Fig. 4 represent the average results with respect to the expert confidence rate obtained according to 50 expert-acquired data. One can see a sudden change in the results as the value of η increases; for η values smaller than 10, a high average error is obtained, whereas for values of η larger than 10, the performance becomes better but stays almost similar across large η values. This demonstrates that having general knowledge about expert confidence (i.e., low or high confident expert) is often sufficient for setting the expert confidence rate.

V. Conclusion

This paper introduced an optimal expert-enabled maximum aposteriori (MAP) inference method for regulatory networks. Boolean network with perturbation is used for modeling regulatory networks. Unlike the existing inference techniques that only rely on temporal changes in data, the proposed method enables the optimal incorporation of expert knowledge reflected in available data for the inference process. The proposed method consists of offline and online steps: in the offline step, parallel dynamic programming methods tuned to possible network models are employed; in the online step, the outcomes of the offline step combined with the softmax policy (representing imperfect experts) are integrated for recursive and efficient expert-enabled inference. Through numerical experiments, we showed that the proposed framework increases the accuracy of the inference process more visibly in domains with more confident experts.

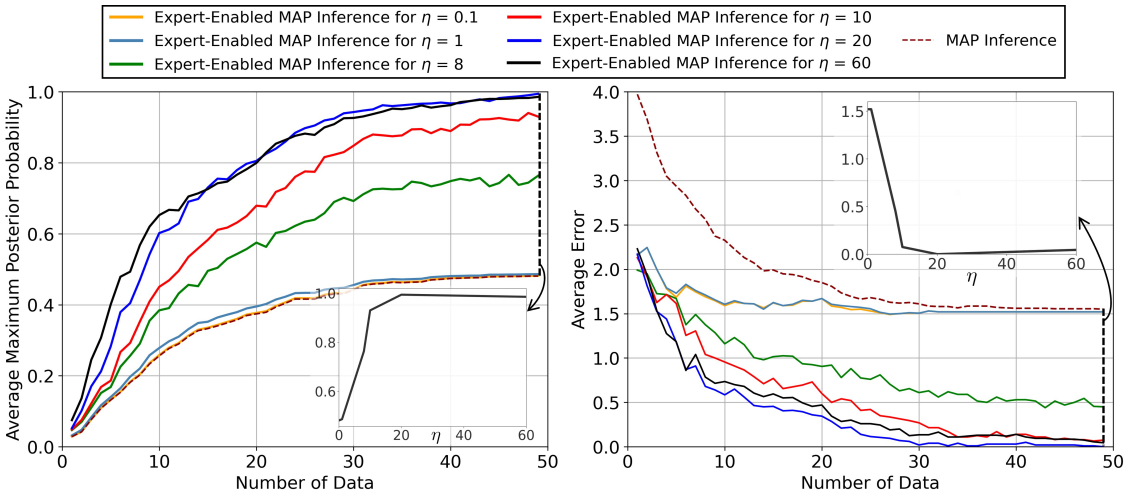

Fig. 4: The average error and maximum posterior probability with respect to the number of data and expert confidence rate using the p53-MDM2 network with 5 unknown regulatory interactions.

The future work examines the expert-enabled inference in domains with partially observable states, domains with unknown expert confidence or reward functions, as well as domains with complex state, action, and parameter spaces. We will also study the convergence of the expert-enabled inference methods in terms of the objective, network structure, and unknown parameters and compare them with conventional inference techniques.

Acknowledgment

This work has been supported in part by the National Institutes of Health award 1R21EB032480-01, National Science Foundation award IIS-2202395, ARMY Research Office award W911NF2110299, and Oracle Cloud credits and related resources provided by the Oracle for Research program.

References

[1] T. Timmermann, B. González, and G. A. Ruz, "Reconstruction of a gene regulatory network of the induced systemic resistance defense response in Arabidopsis using Boolean networks," *BMC bioinformatics*, vol. 21, no. 1, pp. 1–16, 2020.

[2] L. E. Chai, S. K. Loh, S. T. Low, M. S. Mohamad, S. Deris, and Z. Zakaria, "A review on the computational approaches for gene regulatory network construction," *Computers in biology and medicine*, vol. 48, pp. 55–65, 2014.

[3] V. Tiwari and D. M. Wilson III, "DNA damage and associated DNA repair defects in disease and premature aging," *The American Journal of Human Genetics*, vol. 105, no. 2, pp. 237–257, 2019.

[4] J. J. Chan and Y. Tay, "Noncoding RNA: RNA regulatory networks in cancer," *International journal of molecular sciences*, vol. 19, no. 5, p. 1310, 2018.

[5] J. Ule and B. J. Blencowe, "Alternative splicing regulatory networks: functions, mechanisms, and evolution," *Molecular cell*, vol. 76, no. 2, pp. 329–345, 2019.

[6] T. Akutsu, *Algorithms for analysis, inference, and control of Boolean networks*. World Scientific, 2018.

[7] I. Shmulevich, E. R. Dougherty, and W. Zhang, "From Boolean to probabilistic Boolean networks as models of genetic regulatory networks," *Proceedings of the IEEE*, vol. 90, no. 11, pp. 1778–1792, 2002.

[8] P. Trairatphisan, A. Mizera, J. Pang, A. A. Tantar, J. Schneider, and T. Sauter, "Recent development and biomedical applications of probabilistic Boolean networks," *Cell communication and signaling*, vol. 11, no. 1, pp. 1–25, 2013.

[9] I. Apostolopoulou and D. Marculescu, "Tractable learning and inference for large-scale probabilistic Boolean networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2720–2734, 2019.

[10] Y. Tan, F. L. Neto, and U. Braga-Neto, "PALLAS: Penalized maximum likelihood and particle swarms for inference of gene regulatory networks from time series data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.

[11] M. Imani and U. M. Braga-Neto, "Maximum-likelihood adaptive filter for partially observed Boolean dynamical systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 359–371, 2017.

[12] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Gene perturbation and intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 18, no. 10, pp. 1319–1331, 2002.

[13] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *Proceedings of the National Academy of Sciences*, vol. 100, no. 25, pp. 14796–14799, 2003.

[14] S. N. Steinway, M. B. Biggs, T. P. Loughran Jr, J. A. Papin, and R. Albert, "Inference of network dynamics and metabolic interactions in the gut microbiome," *PLoS computational biology*, vol. 11, no. 6, p. e1004338, 2015.

[15] Y. Xiao and E. R. Dougherty, "The impact of function perturbations in Boolean networks," *Bioinformatics*, vol. 23, no. 10, pp. 1265–1273, 2007.

[16] R. S. Sutton, A. G. Barto, *et al.*, *Introduction to reinforcement learning*, vol. 2. MIT press Cambridge, 1998.

[17] S. Arora and P. Doshi, "A survey of inverse reinforcement learning: Challenges, methods and progress," *Artificial Intelligence*, vol. 297, p. 103500, 2021.

[18] M. Imani and S. F. Ghoreishi, "Scalable inverse reinforcement learning through multifidelity Bayesian optimization," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[19] M. Imani and U. Braga-Neto, "Control of gene regulatory networks using Bayesian inverse reinforcement learning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 4, pp. 1250–1261, 2019.

[20] P. R. Kumar and P. Varaiya, *Stochastic systems: Estimation, identification, and adaptive control*, vol. 75. SIAM, 2015.

[21] D. P. Bertsekas, "Dynamic programming and optimal control," *Athena Scientific, Belmont, MA*, 2001.

[22] M. Imani and U. Braga-Neto, "Optimal control of gene regulatory networks with unknown cost function," in *Proceedings of the 2018 American Control Conference (ACC)*, IEEE, 2018.

[23] E. Batchelor, A. Loewer, and G. Lahav, "The ups and downs of p53: understanding protein dynamics in single cells," *Nature Reviews Cancer*, vol. 9, no. 5, p. 371, 2009.

[24] M. Fischer, "Conservation and divergence of the p53 gene regulatory network between mice and humans," *Oncogene*, vol. 38, no. 21, pp. 4095–4109, 2019.

[25] M. Imani and U. M. Braga-Neto, "Maximum-likelihood adaptive filter for partially observed Boolean dynamical systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 359–371, 2016.