

Network Learning with Directional Sign Patterns

Anqi Dong, Can Chen, Tryphon T. Georgiou, *Life Fellow, IEEE*

Abstract—Complex systems can be effectively modeled via graphs that encode networked interactions, where relations between entities or nodes are often quantified by signed edge weights, e.g., promotion/inhibition in gene regulatory networks, or encoding political or friendship differences in social networks. However, it is often the case that only an aggregate consequence of such edge weights that characterize relations may be directly observable, as in protein expression of in gene regulatory networks. Thus, learning edge weights poses a significant challenge that is further exacerbated for intricate and large-scale networks. In this article, we address a model problem to determine the strength of sign-indefinite relations that explain marginal distributions that constitute our data. To this end, we develop a paradigm akin to that of the Schrödinger bridge problem and an efficient Sinkhorn type algorithm (more properly, Schrödinger-Fortet-Sinkhorn algorithm) that allows fast convergence to parameters that minimize a relative entropy/likelihood criterion between the sought signed adjacency matrix and a prior. The formalism that we present represents a novel generalization of the earlier Schrödinger formalism in that marginal computations may incorporate weights that model directionality in underlying relations, and further, that it can be extended to high-order networks – the Schrödinger-Fortet-Sinkhorn algorithm that we derive is applicable all the same and allows geometric convergence to a sought sign-indefinite adjacency matrix or tensor, for high-order networks.

I. INTRODUCTION

Complex interactions between multiple subsystems can often be encoded in a networked structure that models respective relations [22]. These relations may be binary as in ordinary graphs with vertices and edges, trinary and so on, as in higher-order relations. The strength of respective interactions may be assigned as a weight in corresponding edges, or their higher-order analogues. Yet, contrary to typical accounts in the vast literature on the subject, interactions may be sign-indefinite, signifying e.g., promotion/inhibition or friend/unfriend in biological and social networks, respectively. Moreover, perceived aggregate effects of such relations may have an added directional bias. In many such examples, acquiring nodal information is relatively straightforward, while directly and precisely learning edge weights presents a significant challenge. Numerous computational methods have been proposed to learn and quantify edge

weights in networks, including cross-correlation [16], mutual information [27], system identification [10], and graph transformer neural networks [29]. However, most of these methods require adequate time-series data on node states, which might not be easily accessible in general.

In our recent work [14], we presented a framework for identifying sign-distinguishable edge weights by utilizing prior knowledge of the sign-indefinite structure, encoded in a sign-indefinite “prior” adjacency matrix, and nodal statistics that constitute data. The formalism involves minimizing a suitable relative entropy functional between sign-indefinite measures, which can be solved using a suitable generalization of the well-known Sinkhorn algorithm (more properly, Schrödinger-Fortet-Sinkhorn algorithm, since it was explicit in the work of Fortet [15] on the Schrödinger bridge problem). The generalization consists in modifying the iteration steps by scaling with values computed as the positive roots of a quadratic polynomial – the classical Sinkhorn “diagonal-scaling” iteration applies when the adjacency matrix is sign-definite (with entries ≥ 0) [6], [18]. The proposed method was applied to determine the promotion and inhibition interactions in gene regulatory networks [14].

Whereas determining the sign (promotion/inhibition) in multi-subsystem/multi-species relations may be tangible, determining the precise strength may be challenging as it is not typically accessible in complex and large-scale networks. For instance, in gene regulatory networks, protein expression reveals marginal information on nodes, and statistics on increased/decreased expression provide a sign for the promotion/inhibitive coupling between nodes. However, the strength of such pair-wise or, more generally, multi-gene interactions cannot be directly observed due to stupendously complicated effects across the network. Yet it is of at most importance in inferring potential fitness advantages on organisms by their combined effect [26]. Then, in ecological networks, quantifying interspecies interactions can also be challenging due to the intricate complexity and variability of the overall activity across the network, including environmental factors, habitat structures, and species behaviors [24]. Such scenarios motivate our problem, where partial information at the level of nodes can be assumed reliable, whereas the strength (but not the sign) of respective interactions is less so. Thereby, we consider directional marginal-wise sign templates that may encode the combined effect of interactions at the level of nodes, and assume nodal data as given, from which we seek to identify strength levels of pair-wise or multi-node interactions.

Evidently, in real-world systems that may be impacted by higher-order interactions, describing these solely using

Anqi Dong is with the Division of Decision and Control Systems and Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. anqidong.andy@gmail.com

Can Chen is with the School of Data Science and Society and the Department of Mathematics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. canc@unc.edu

Tryphon T. Georgiou are with the Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697, USA. tryphon@uci.edu

networks may lead to a loss of higher-order information. A higher-order network (or a hypergraph) is a generalization of a network, in which its hyperedges can connect any number of nodes [1]. Examples of higher-order networks encompass various domains, including email communication networks, metabolic networks, and protein-protein interaction networks [2], [3]. Computational methods and the influx of concepts such as entropy [4], controllability [5], and various similarity measures [25] are rapidly entering the toolbox for dealing with higher-order networks. The contribution in the present paper can be viewed from this angle, formulating a basic problem of inference of multi-node interactions based on observed partial information and priors.

Thus, in this article, we formulate the problem and propose a framework aimed at tackling the inverse problem of estimating edge weights in a network by minimizing the Kullback–Leibler divergence between a prior and a posterior concerning directional marginal-wise sign patterns. Similar to our previous work [14], this problem can be solved using a Sinkhorn-like algorithm. This algorithm amounts to implementing coordinate ascent to maximize a concave functional, leading to linear convergence rate. The algorithm represents a significant generalization of the standard solver of the entropic regularized optimal transport problem and the Schrödinger bridges problem, which has attracted increasing attention among the fields of control [7], [8], machine learning [8], and image processing [23]. Furthermore, it applies to higher-order networks, where we determine hyperedge weights based on directional marginal-wise sign-indefinite adjacency tensors (generalizations of adjacency matrices).

The organization of the remaining article is as follows. In Section II, we provide necessary background knowledge of signed networks and the Sinkhorn algorithm. Section III presents the generalized Sinkhorn algorithm along with established theoretical results for our proposed problem. Section IV discusses the generalization to general higher-order networks with the proposed method. Section V includes three numerical examples to validate the algorithm. Finally, we conclude with future directions in Section VI.

II. PRELIMINARIES

A. Signed Networks

Signed networks refer to network structures where the relationships between nodes are sign-indefinite ($\mathbf{A}_{ij} \gtrless 0$) [30]. These networks capture not only affinity between nodes but also the nature of respective interactions, whether these represent cooperation, friendship, antagonism, rivalry, or other types of expression. Signed networks are prevalent across diverse domains, spanning sociology [28], ecology [11], and signal processing [12]. Mathematically, a signed network can be represented by a sign-indefinite adjacency matrix, where its entries take values of -1 , 0 , and 1 , or by a suitably weighted counterpart.

Signed higher-order networks generalize signed networks by representing relationships among multiple entities or nodes with positive or negative signs. When all hyperedges link the exact same k number of nodes, the higher-order

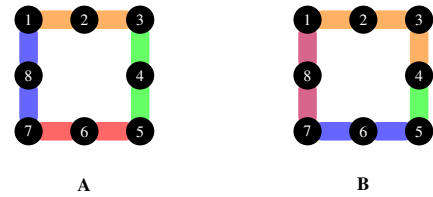


Fig. 1: Every hyperedge in the example is colored by one color. **(A)** 3-uniform higher-order network with hyperedges $e_1 = \{1, 2, 3\}$, $e_2 = \{3, 4, 5\}$, $e_3 = \{5, 6, 7\}$ and $e_4 = \{8, 7, 1\}$. **(B)** Non-uniform higher-order network with hyperedges $e_1 = \{1, 2, 3, 4\}$, $e_2 = \{4, 5\}$, $e_3 = \{5, 6, 7\}$ and $e_3 = \{1, 8, 7\}$.

network is termed k -uniform (Figure 1). Every k -uniform higher-order network with n nodes can be represented by a k th-order n -dimensional adjacency tensor (i.e., a multi-dimensional array generalized from matrices and vectors) [4], [5]. Thus, similar to signed networks, a signed k -uniform higher-order network can be characterized by a sign-indefinite adjacency tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times \dots \times n}$ defined as

$$\mathcal{A}_{i_1 i_2 \dots i_k} = \begin{cases} 1 & \text{if } e \text{ forms a positive interaction} \\ -1 & \text{if } e \text{ forms a negative interaction} \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where e denotes the hyperedge containing node i_1, i_2, \dots, i_k .

B. Schrödinger Bridges Problem

The Schrödinger Bridges Problem (SBP) originated as a problem in large deviations theory, as the problem to minimize the Kullback-Leibler divergence, a measure of dissimilarity between probability distributions [7], [9], between prior and posterior, so that the latter agrees with estimated statistics. Thus, if $\mathbf{Q} \in \mathbb{R}_{\geq 0}^{n \times n}$ with non-negative entries represents a “prior” joint probability between two random variables, and if the corresponding to positive probability vectors $\mathbf{p} \in \mathbb{R}_{> 0}^n$ and $\mathbf{q} \in \mathbb{R}_{> 0}^n$ are not consistent with \mathbf{Q} , then SBP seeks to adjust the values in \mathbf{Q} so as to derive a “posterior” $\mathbf{P} \in \mathbb{R}_{\geq 0}^{n \times n}$ that is consistent with the specified marginals. The posterior matrix $\mathbf{P} \in \mathcal{C}$ is defined as the set of admissible elements from a feasible set $\mathcal{C} = \mathcal{C}_0 \cap \mathcal{C}_1 \cap \mathcal{C}_2$, where $\mathcal{C}_0 := \{\mathbf{P} \mid \mathbf{P}_{ij} \geq 0\}$, $\mathcal{C}_1 := \{\mathbf{P} \mid \sum_{j=1}^n \mathbf{P} = \mathbf{p}\}$, and $\mathcal{C}_2 := \{\mathbf{P} \mid \sum_{i=1}^n \mathbf{P} = \mathbf{q}\}$.

The standard SBP seeks \mathbf{P} as the minimizer of the Kullback-Leibler divergence to the prior \mathbf{Q} , i.e., as the solution to

$$\mathbf{P}^* = \arg \min_{\mathbf{P} \in \mathcal{C}} \text{KL}(\mathbf{P}|\mathbf{Q}) = \arg \min_{\mathbf{P} \in \mathcal{C}} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{ij} \log \left(\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}} \right).$$

Throughout we follow the convention that $\log(0/0) = 0$. The objective function is strictly convex, and once the feasible set \mathcal{C} is non-empty, an optimizer \mathbf{P}^* always exists. Additionally, a closed-form solution can be derived from the first-order

Algorithm 1 Sinkhorn algorithm

- 1: Initialize $\boldsymbol{\nu} \in \mathbb{R}^n$, e.g., setting $\boldsymbol{\nu} = \mathbf{0}$.
 - 2: For $j = 1, 2, \dots, n$, determine

$$\mu_j = \log(p_j / \sum_{i=1}^n \mathbf{Q}_{ij} \exp(-\nu_i)).$$
 - 3: For $i = 1, 2, \dots, n$, determine

$$\nu_i = \log(q_i / \sum_{j=1}^n \mathbf{Q}_{ij} \exp(-\mu_j)).$$
 - 4: Repeat steps 2 and 3 until convergence.
-

optimality condition by employing the Lagrangian, i.e.,

$$\begin{aligned} \mathcal{L}(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\nu}) := & \sum_{ij | \mathbf{Q}_{ij} \neq 0} \mathbf{P}_{ij} \log\left(\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}}\right) + \sum_{j=1}^n \mu_j \left(\sum_{i=1}^n \mathbf{P}_{ij} - q_j\right) \\ & + \sum_{i=1}^n \nu_i \left(\sum_{j=1}^n \mathbf{P}_{ij} - p_i\right), \end{aligned}$$

where μ_j is the j th entry of $\boldsymbol{\mu}$ (similarly for q_j , ν_i , and p_i). Therefore, $\mathbf{P}_{ij}^* = \mathbf{Q}_{ij} \exp(-\nu_i - \mu_j)$ with $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are the dual variables for the marginal constraints

$$\sum_{j=1}^n \mathbf{P}_{ij}^* = \mathbf{p} \quad \text{and} \quad \sum_{i=1}^n \mathbf{P}_{ij}^* = \mathbf{q}. \quad (2)$$

The solution \mathbf{P}_{ij}^* has to satisfy the marginal constraints in \mathcal{C}_1 and \mathcal{C}_2 , and the optimizer can be thus obtained by the Sinkhorn scheme that iteratively updates the two dual variables (Algorithm 1). The linear convergence rate of the Sinkhorn algorithm can be established using the Hilbert metric, and interested readers are referred to [23] for further details.

The SBP can be generalized to the multidimensional setting, taking the form

$$\mathcal{P}^* = \arg \min_{\mathcal{P} \in \mathcal{C}} \sum_{i_1 i_2 \dots i_k} \mathcal{P}_{i_1 i_2 \dots i_k} \left(\log \left(\frac{\mathcal{P}_{i_1 i_2 \dots i_k}}{\mathcal{Q}_{i_1 i_2 \dots i_k}} \right) \right), \quad (3)$$

where $\mathcal{Q} \in \mathbb{R}_{\geq 0}^{n \times n \times \dots \times n}$ represents the given prior and $\mathcal{P} \in \mathbb{R}_{\geq 0}^{n \times n \times \dots \times n}$ denotes the posterior. The posterior \mathcal{P} satisfies the marginal constraint $\mathcal{P} \in \mathcal{C}_0 \cap \mathcal{C}_1 \cap \dots \cap \mathcal{C}_k$ where each \mathcal{C}_ℓ is defined as $\mathcal{C}_\ell = \{\mathcal{P} \mid \sum_{i_1 \dots i_k / i_\ell} \mathcal{P} = \mathbf{p}^{(\ell)}\}$ for every marginal vector $\mathbf{p}^{(\ell)} \in \mathbb{R}_{> 0}^n$. Similar to the standard Sinkhorn iteration in Algorithm 1, a closed-form solution of the optimizer can be derived, obtained by a gradient ascent method with, similarly, a linear convergence rate [21].

III. SINKHORN WITH DIRECTIONAL SIGN TEMPLATES

The underlying principle in our previous work [13], [14] assumes a predetermined prior, with a given sign template applied across all directions of the marginal constraints. However, such an assumption may not always be practical for real-world applications. In complex systems, precisely capturing relationships can be challenging due to uncertainties or the inability to directly measure certain edge signs. This scenario would lead to the creation of multiple distinct signed networks, each representing alternative interpretations of the system's dynamics. For instance, in a social network

where friendships and animosities exist, the absence of direct observations for several relationships may lead to the construction of multiple signed social networks with differing assumptions about uncertain relationships. Since sign patterns may be subjective and their precise determination may be impossible under certain circumstances, a framework that allows for directional sign templates as an initial guess is necessary and critical.

We introduce a non-trivial generalization of the Sinkhorn algorithm, namely Sinkhorn with directional sign templates. We first start with the classical two-dimensional case by considering a prior matrix $\mathbf{Q} \in \mathbb{R}_{\geq 0}^{n \times n}$ and two specified positive marginal vectors $\mathbf{p} \in \mathbb{R}_{> 0}^n$ and $\mathbf{q} \in \mathbb{R}_{> 0}^n$. In contrast to the previous work, the problem now admits two directional marginal-wise sign patterns. The constraints for the posterior matrix $\mathbf{P} \in \mathbb{R}_{\geq 0}^{n \times n}$ in (2) are thus generalized to

$$\mathcal{C}'_1 := \{\mathbf{P} \mid \sum_{j=1}^n [\mathbf{X} * \mathbf{P}]_{ij} = p_j\}, \quad \mathcal{C}'_2 := \{\mathbf{P} \mid \sum_{i=1}^n [\mathbf{Y} * \mathbf{P}]_{ij} = q_i\}, \quad (4)$$

where $\mathbf{X} \in \mathbb{R}^{n \times n}$ and $\mathbf{Y} \in \mathbb{R}^{n \times n}$ are two distinct sign-indefinite adjacency matrices of the network, and $*$ denotes the element-wise multiplication.

Problem 1: The SBP seeks the most likely posterior $\mathbf{P} \in \mathbb{R}_{\geq 0}^{n \times n}$ with respect to a given prior $\mathbf{Q} \in \mathbb{R}_{\geq 0}^{n \times n}$, achieved by minimizing the relative entropy

$$\mathbf{P}^* = \arg \min_{\mathbf{P}} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{ij} \left(\log \left(\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}} \right) - 1 \right) \quad (5)$$

with the convention $\log(0/0) = 0$. We intentionally add -1 at the end for simplicity in later computations. The posterior \mathbf{P} is also required to meet the marginal constraints

$$\sum_{i=1}^n \mathbf{X}_{ij} \mathbf{P}_{ij} = p_j \quad \text{and} \quad \sum_{j=1}^n \mathbf{Y}_{ij} \mathbf{P}_{ij} = q_i. \quad (6)$$

Our problem considers two distinct sign patterns, indicated by \mathbf{X} and \mathbf{Y} , for different marginals while maintaining a convex objective. First, we demonstrate that the optimizer has a closed-form solution under the feasibility assumption.

Assumption 1 (Existence): The solution to Problem 1 exists. In other words, the set of possible solutions is defined as $\{\mathbf{P} \mid \mathbf{P} \in \mathcal{C}'_1 \cap \mathcal{C}'_2\} \cap \{\mathbf{P} \mid KL(\mathbf{P} \mid \mathbf{Q}) < +\infty\}$ is non-empty.

Note that a simple feasibility test may not exist for the problem. It is well-known that certain moment problems lack a straightforward feasibility test, as discussed in [17].

Proposition 1 (Closed-form solution): The optimizer \mathbf{P} has a closed-form solution computed as

$$\mathbf{P}_{ij}^* = \begin{cases} \mathbf{Q}_{ij} \exp(-\mu_j \mathbf{X}_{ij} - \nu_i \mathbf{Y}_{ij}) & \text{for } \mathbf{X}_{ij} \neq 0, \mathbf{Y}_{ij} \neq 0 \\ 0 & \text{for } \mathbf{X}_{ij} = 0, \mathbf{Y}_{ij} = 0 \end{cases},$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\nu} \in \mathbb{R}^n$ are the Lagrangian multipliers of the constraints in (4).

Proof: We first rewrite the Lagrangian \mathcal{L} as

$$\begin{aligned} \mathcal{L}(\mathbf{P}, \boldsymbol{\mu}, \boldsymbol{\nu}) &:= \sum_{ij | \mathbf{Q}_{ij} \neq 0} \mathbf{P}_{ij} \left(\log\left(\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}}\right) - 1 \right) \\ &+ \sum_j \mu_j \left(\sum_{i=1}^n \mathbf{X}_{ij} \mathbf{P}_{ij} - p_j \right) + \sum_i \nu_i \left(\sum_{j=1}^n \mathbf{Y}_{ij} \mathbf{P}_{ij} - q_i \right). \end{aligned} \quad (7)$$

The first-order optimality condition $\partial\mathcal{L}/\partial\mathbf{P}_{ij} = 0$ yields $\log(\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}}) + \mu_j \mathbf{X}_{ij} + \nu_i \mathbf{Y}_{ij} = 0$, for i, j with $\mathbf{Q}_{ij} > 0$. Otherwise, $\mathbf{P}_{ij} = 0$. Therefore, the optimizer must have the following functional dependence on the Lagrange multipliers, i.e., $\mathbf{P}_{ij}^* = \mathbf{Q}_{ij} \exp(-\mu_j \mathbf{X}_{ij} - \nu_i \mathbf{Y}_{ij})$, if $\mathbf{Q}_{ij} \neq 0$, and $\mathbf{P}_{ij}^* = 0$ otherwise. ■

The optimal kernel can be rewritten in a more specific form

$$\mathbf{P}_{ij}^* = \begin{cases} \mathbf{Q}_{ij} \exp(-\mu_j - \nu_i), & \text{if } \mathbf{X}_{ij} > 0, \mathbf{Y}_{ij} > 0, \\ \mathbf{Q}_{ij} \exp(-\mu_j + \nu_i), & \text{if } \mathbf{X}_{ij} > 0, \mathbf{Y}_{ij} < 0, \\ \mathbf{Q}_{ij} \exp(+\mu_j - \nu_i), & \text{if } \mathbf{X}_{ij} < 0, \mathbf{Y}_{ij} > 0, \\ \mathbf{Q}_{ij} \exp(+\mu_j + \nu_i), & \text{if } \mathbf{X}_{ij} < 0, \mathbf{Y}_{ij} < 0, \\ 0, & \text{if } \mathbf{X}_{ij} = 0, \mathbf{Y}_{ij} = 0. \end{cases} \quad (8)$$

To write (8) in a more compact form, we define indicator matrices \mathbf{Q}^{++} , \mathbf{Q}^{+-} , \mathbf{Q}^{-+} , and \mathbf{Q}^{--} as

$$\begin{aligned} \mathbf{Q}^{++} &:= \{\mathbf{Q}_{ij} \mid \mathbf{X}_{ij} > 0, \mathbf{Y}_{ij} > 0\}, \\ \mathbf{Q}^{+-} &:= \{\mathbf{Q}_{ij} \mid \mathbf{X}_{ij} > 0, \mathbf{Y}_{ij} < 0\}, \\ \mathbf{Q}^{-+} &:= \{\mathbf{Q}_{ij} \mid \mathbf{X}_{ij} < 0, \mathbf{Y}_{ij} > 0\}, \\ \mathbf{Q}^{--} &:= \{\mathbf{Q}_{ij} \mid \mathbf{X}_{ij} < 0, \mathbf{Y}_{ij} < 0\}. \end{aligned}$$

It is straightforward to see that $\mathbf{Q} = \mathbf{Q}^{++} + \mathbf{Q}^{+-} + \mathbf{Q}^{-+} + \mathbf{Q}^{--}$. We also define variables $\alpha_j = \exp(\mu_j)$ and $\beta_i = \exp(\nu_i)$. The optimizer (8) thus can be rewritten as

$$\mathbf{P}_{ij}^* = \mathbf{Q}_{ij}^{++} \alpha_j \beta_i + \mathbf{Q}_{ij}^{+-} \alpha_j \beta_i^{-1} + \mathbf{Q}_{ij}^{-+} \alpha_j^{-1} \beta_i + \mathbf{Q}_{ij}^{--} \alpha_j^{-1} \beta_i^{-1},$$

which satisfies the constraints in (4), i.e.,

$$\begin{aligned} \sum_{i=1}^n (\mathbf{Q}_{ij}^{++} \beta_i + \mathbf{Q}_{ij}^{+-} \beta_i^{-1}) \alpha_j + (\mathbf{Q}_{ij}^{-+} \beta_i + \mathbf{Q}_{ij}^{--} \beta_i^{-1}) \alpha_j^{-1} &= p_j, \\ \sum_{j=1}^n (\mathbf{Q}_{ij}^{++} \alpha_j + \mathbf{Q}_{ij}^{-+} \alpha_j^{-1}) \beta_i + (\mathbf{Q}_{ij}^{+-} \alpha_j + \mathbf{Q}_{ij}^{--} \alpha_j^{-1}) \beta_i^{-1} &= q_i. \end{aligned}$$

Additionally, to simplify the constraints, we define

$$\begin{aligned} a_j^{\{\mathbf{Q}^{++}, \mathbf{Q}^{+-}\}}(\boldsymbol{\beta}) &:= \sum_{i=1}^n (\mathbf{Q}_{ij}^{++} \beta_i + \mathbf{Q}_{ij}^{+-} \beta_i^{-1}), \\ b_j^{\{\mathbf{Q}^{-+}, \mathbf{Q}^{--}\}}(\boldsymbol{\beta}) &:= \sum_{i=1}^n (\mathbf{Q}_{ij}^{-+} \beta_i + \mathbf{Q}_{ij}^{--} \beta_i^{-1}), \\ a_i^{\{\mathbf{Q}^{++}, \mathbf{Q}^{-+}\}}(\boldsymbol{\alpha}) &:= \sum_{j=1}^n (\mathbf{Q}_{ij}^{++} \alpha_j + \mathbf{Q}_{ij}^{-+} \alpha_j^{-1}), \\ b_i^{\{\mathbf{Q}^{+-}, \mathbf{Q}^{--}\}}(\boldsymbol{\alpha}) &:= \sum_{j=1}^n (\mathbf{Q}_{ij}^{+-} \alpha_j + \mathbf{Q}_{ij}^{--} \alpha_j^{-1}). \end{aligned}$$

We can then rewrite the constraints as

$$p_j = a_j^{\{\mathbf{Q}^{++}, \mathbf{Q}^{+-}\}}(\boldsymbol{\beta}) \alpha_j - b_j^{\{\mathbf{Q}^{-+}, \mathbf{Q}^{--}\}}(\boldsymbol{\beta}) \alpha_j^{-1}, \quad (9a)$$

$$q_i = a_i^{\{\mathbf{Q}^{++}, \mathbf{Q}^{-+}\}}(\boldsymbol{\alpha}) \beta_i - b_i^{\{\mathbf{Q}^{+-}, \mathbf{Q}^{--}\}}(\boldsymbol{\alpha}) \beta_i^{-1}. \quad (9b)$$

Algorithm 2 Generalized Sinkhorn algorithm

1: Initialize $\boldsymbol{\beta} \in \mathbb{R}^n$, e.g., setting $\boldsymbol{\nu} = \mathbf{1}$.

2: Compute \mathbf{Q}^{++} , \mathbf{Q}^{+-} , \mathbf{Q}^{-+} , and \mathbf{Q}^{--} .

3: For $j = 1, 2, \dots, n$, determine

$$\alpha_j = g\left(a_j^{\{\mathbf{Q}^{++}, \mathbf{Q}^{+-}\}}(\boldsymbol{\beta}), b_j^{\{\mathbf{Q}^{-+}, \mathbf{Q}^{--}\}}(\boldsymbol{\beta}), p_j\right). \quad (11a)$$

4: For $i = 1, 2, \dots, n$, determine

$$\beta_i = g\left(a_i^{\{\mathbf{Q}^{++}, \mathbf{Q}^{-+}\}}(\boldsymbol{\alpha}), b_i^{\{\mathbf{Q}^{+-}, \mathbf{Q}^{--}\}}(\boldsymbol{\alpha}), q_i\right). \quad (11b)$$

5: Repeat steps 2 and 3 until convergence.

6: Compute the optimizer \mathbf{P}^* .

We observe that α_j can be explicitly computed from the marginal constraint p_j , while keeping the vector $\boldsymbol{\beta}$ fixed. Similarly, β_i can be computed from q_i . To see this, we can consider the following function $f(x) = ax + bx^{-1}$, where a and b are both positive. Thus, for a given value c , a solution to $f(x) = c$ can be readily obtained as the positive root of a quadratic equation, computed as

$$x = \left(\frac{-c + \sqrt{c^2 + 4ab}}{2b} \right) =: g(a, b, c). \quad (10a)$$

For our purposes, the case where $b = 0$ is also of interest, i.e.,

$$x = \left(\frac{a}{c} \right) =: g(a, 0, c). \quad (10b)$$

The proposed generalized Sinkhorn algorithm for two directional sign templates is summarized in Algorithm 2.

Proposition 2 (Convergence): Algorithm 2 converges to optimal with linear convergence rate under Assumption 1.

Proof: The convergence of the algorithm can be understood through the strong duality of the problem. We first substitute the optimizer \mathbf{P}_{ij}^* into the Lagrangian (7), obtaining

$$h(\boldsymbol{\mu}, \boldsymbol{\nu}) = \sum_{i=1}^n \sum_{j=1}^n -\mathbf{P}_{ij}^* - \mu_j p_j - \nu_i q_i. \quad (12)$$

The dual problem is therefore a maximization problem with a strictly concave objective defined as

$$\arg \max_{\boldsymbol{\mu}, \boldsymbol{\nu}} h(\boldsymbol{\mu}, \boldsymbol{\nu}), \quad (13)$$

where the Slater's condition is satisfied, ensuring that strong duality holds.

Given that the objective is strictly convex, we can obtain the optimizer \mathbf{P}^* using a coordinate ascent strategy, while ensuring strong duality holds. Specifically, the coordinate-wise optimization of the dual (13), i.e., $\mu_j = \arg \max_{\mu_j} h(\boldsymbol{\mu}, \boldsymbol{\nu})$ and $\nu_i = \arg \max_{\nu_i} h(\boldsymbol{\mu}, \boldsymbol{\nu})$, gives that

$$\frac{\partial h(\mu_j, \nu_i)}{\partial \mu_j} = \sum_{i=1}^n \mathbf{P}_{ij}^* - p_j = 0, \quad \frac{\partial h(\mu_j, \nu_i)}{\partial \nu_i} = \sum_{j=1}^n \mathbf{P}_{ij}^* - q_i = 0$$

at each step, leading to an update for μ_j and ν_i to satisfy (9) and (11). Therefore, the generalized Sinkhorn iteration in

Algorithm 2 inherits the linear convergence rate of coordinate ascent [21, Theorem 2.1], as in the standard Sinkhorn. ■

Last but not least, the problem and obtained results (including Algorithm 2 and Proposition 2) can be extended to the multi-marginal setting for higher-order networks, with a predetermined k th-order n -dimensional prior $\mathcal{Q} \in \mathbb{R}_{>0}^{n \times n \times \dots \times n}$ and k positive marginals $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}, \dots, \mathbf{p}^{(k)} \in \mathbb{R}_{>0}^n$. Moreover, every marginal constraint set for the posterior $\mathcal{P} \in \mathbb{R}_{\geq 0}^{n \times n \times \dots \times n}$ is then defined as

$$\mathcal{C}_\ell'' := \{ \mathcal{P} \mid \sum_{i_1 i_2 \dots i_k / i_\ell} \mathcal{X}_{i_1 i_2 \dots i_k}^{(\ell)} \mathcal{P}_{i_1 i_2 \dots i_k} = p_{i_\ell}^{(\ell)} \},$$

where $\mathcal{X}^{(\ell)} \in \mathbb{R}^{n \times n \times \dots \times n}$ are directional marginal-wise sign-indefinite adjacency tensors. Thus, all the admissible solution \mathcal{P} has to be in the set $\mathcal{C}'' = \mathcal{C}_1'' \cap \mathcal{C}_2'' \cap \dots \cap \mathcal{C}_k''$.

Problem 2: The SBP seeks the most likely posterior $\mathcal{P} \in \mathbb{R}_{\geq 0}^{n \times n \times \dots \times n}$ with respect to a given prior $\mathcal{Q} \in \mathbb{R}_{\geq 0}^{n \times n \times \dots \times n}$, achieved by minimizing the relative entropy (3) over admissible solutions $\mathcal{P} \in \mathcal{C}''$.

The problem is well-defined when it is feasible, and its solution can be obtained similarly to the two-dimensional case (Algorithm 2). Concerning convergence, the argument is analogous to our discussion in Proposition 2, as it relies on the gradient ascent nature of the problem [21].

IV. GENERAL HIGH-ORDER NETWORKS

Every k -uniform higher-order network, where hyperedges have the same cardinality k , is known to possess a k th-order tensor representation (1). However, representing general (or non-uniform) higher-order networks is not straightforward due to the potential variance in hyperedge cardinality. Nevertheless, a tensor representation can be achieved by converting the original non-uniform higher-order network to a uniform one through the addition of ‘virtual’ nodes. These virtual nodes carry no nodal information but merely serve as placeholders to ensure every hyperedge has the same cardinality. Consequently, the converted uniform higher-order network features k^{\max} cardinality for every hyperedge, with k^{\max} representing the largest hyperedge cardinality in the original non-uniform higher-order network. A simple example of such conversion is illustrated in Figure 2. Ultimately, the corresponding marginal distribution maintains the same number of nodes, with nodal information on every virtual node set to zero. Hence, this problem can be adapted to Problem 2.

V. NUMERICAL EXAMPLES

We conducted three numerical examples to illustrate our framework and the efficiency of the generalized Sinkhorn algorithm. The code for all experiments can be found at the following link: <https://github.com/dytroshut/marginal-wise-CDC>.

A. Ecological Network

In ecological networks, interactions between species are often understood qualitatively, meaning that the signs of

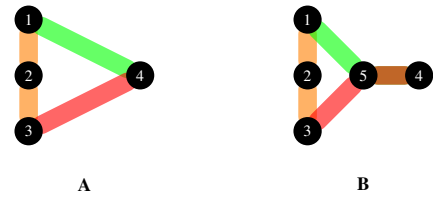


Fig. 2: Uniformity conversion. (A) A non-uniform higher-order network with hyperedges $e_1 = \{1, 2, 3\}$, $e_2 = \{1, 4\}$, and $e_3 = \{3, 4\}$ is converted to a 3-uniform higher-order network by adding virtual node 5. (B) The resulting 3-uniform higher-order network with hyperedges $e_1 = \{1, 2, 3\}$, $e_2 = \{1, 4, 5\}$, and $e_3 = \{3, 4, 5\}$.

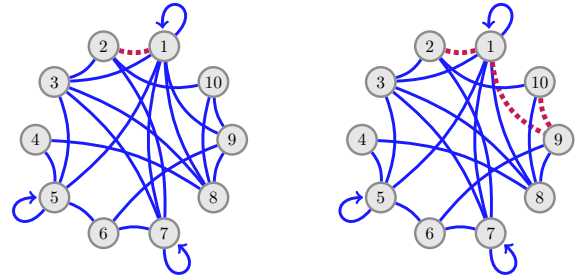


Fig. 3: The topology of a 10-node network with sign templates \mathbf{X} (left) and \mathbf{Y} (right).

interactions (such as inhibition or promotion) can be determined with reasonable confidence, but quantifying the actual magnitudes may be challenging [19], [20]. Under certain instances, even determining the signs of interactions can be obscure. Thus, it is necessary to consider multiple potential sign patterns when learning the magnitudes of these interactions. In this example, we considered an ecological network where ten species are represented by nodes, and their species-wise interactions are represented by edges. The abundance of each species is normalized and represented as marginal distributions given by

$$\mathbf{p} = [0.1, 0.05, 0.05, 0.15, 0.2, 0.05, 0.03, 0.07, 0.25, 0.05],$$

$$\mathbf{q} = [0.05, 0.1, 0.05, 0.2, 0.07, 0.15, 0.05, 0.25, 0.03, 0.05].$$

The two marginal distributions originated from two potential sign templates \mathbf{X} and \mathbf{Y} , with negative signs located at $\mathbf{X}_{1,2}, \mathbf{Y}_{1,2}, \mathbf{Y}_{1,9}, \mathbf{Y}_{9,10}$ with their permutations.

Specifically, the coefficient between Species 1 and 2 is known to be negative, while the interactions between Species 1 and 9, as well as between Species 9 and 10, are uncertain. The topologies of both of the 10-node signed ecological networks are depicted in Figure 3. The prior is set as $\mathbf{Q} = |\mathbf{X}| = |\mathbf{Y}|$ for simplicity, without any initial guess of the edge weights. The posterior (14) obtained using Algorithm 2 represents the maximum likelihood estimation based on the observed information from the ecological networks.

B. Convergence

To demonstrate the convergence rate of the proposed framework, we presented a multi-marginal case study in-

$$\mathbf{P} = \begin{bmatrix} 0.0108 & 0.0990 & 0.0817 & 0 & 0.0412 & 0 & 0.0167 & 0 & 0.0194 & 0.0292 \\ 0.1497 & 0 & 0.1195 & 0 & 0 & 0 & 0.0291 & 0 & 0 & 0.0511 \\ 0.0286 & 0.0002 & 0 & 0 & 0.0023 & 0 & 0.0009 & 0.0180 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0168 & 0 & 0 & 0.1332 & 0 & 0 \\ 0.0010 & 0 & 0.0063 & 0.1439 & 0.0038 & 0.0450 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0059 & 0 & 0.0024 & 0 & 0.0417 & 0 \\ 0.0006 & 0.0002 & 0.0035 & 0 & 0 & 0.0249 & 0.0009 & 0 & 0 & 0 \\ 0 & 0 & 0.0025 & 0.0561 & 0 & 0 & 0 & 0 & 0.0104 & 0.0011 \\ 0.0850 & 0 & 0 & 0 & 0 & 0.0801 & 0 & 0.0536 & 0 & 0.0314 \\ 0.0015 & 0.0006 & 0 & 0 & 0 & 0 & 0 & 0.0453 & 0.0027 & 0 \end{bmatrix} \quad (14)$$

volving tensors (Problem 2), where both the prior and posterior are third-order tensors associated with a higher-order network. The marginal violations are defined as $\log \left(\left\| \sum_{i_1 i_2 i_3 / i_\ell} [\mathcal{X}^{(\ell)} * \mathcal{P}]_{i_1 i_2 i_3} - \mathbf{p}^{(\ell)} \right\| \right)$ for $\ell = 1, 2, 3$. Both the convergence and its linear convergence rate in terms of marginal violations can be found in our arXiv version.

VI. CONCLUSION

In this article, we introduced a generalized Sinkhorn algorithm with directional sign indicators, where each sign pattern serves as an initial guess along with the corresponding marginal, to quantitatively learn edge weights in a (higher-order) network. We also investigated the convergence and convergence rate of the proposed method. The current framework has demonstrated efficiency in the maximum likelihood estimation of (high-order) networks. Furthermore, given that a prior and a posterior are high-dimensional arrays, the problem may encounter the ‘‘curse of dimensionality,’’ as both memory and computation complexity increase exponentially with the size of the problem. To mitigate this challenge, it is worthwhile to apply tensor decomposition techniques.

REFERENCES

- [1] Claude Berge. *Hypergraphs: combinatorics of finite sets*, volume 45. Elsevier, 1984.
- [2] Can Chen, Chen Liao, and Yang-Yu Liu. Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning. *Nature Communications*, 14(1):2375, 2023.
- [3] Can Chen and Yang-Yu Liu. A survey on hyperlink prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [4] Can Chen and Indika Rajapakse. Tensor entropy for uniform hypergraphs. *IEEE Transactions on Network Science and Engineering*, 7(4):2889–2900, 2020.
- [5] Can Chen, Amit Surana, Anthony M Bloch, and Indika Rajapakse. Controllability of hypergraphs. *IEEE Transactions on Network Science and Engineering*, 8(2):1646–1657, 2021.
- [6] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal steering of a linear stochastic system to a final probability distribution, Part I. *IEEE Transactions on Automatic Control*, 61(5):1158–1169, 2015.
- [7] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169:671–691, 2016.
- [8] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4:89–113, 2021.
- [9] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. *SIAM Review*, 63(2):249–313, 2021.
- [10] Alessandro Chiuso and Gianluigi Pillonetto. System identification: A machine learning perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:281–304, 2019.
- [11] Mark RT Dale and Marie-Josée Fortin. *Quantitative analysis of ecological networks*. Cambridge University Press, 2021.
- [12] Thomas Dittrich and Gerald Matz. Signal processing on signed graphs: Fundamentals and potentials. *IEEE Signal Processing Magazine*, 37(6):86–98, 2020.
- [13] Anqi Dong, Tryphon T. Georgiou, and Allen Tannenbaum. Data assimilation for sign-indefinite priors: A generalization of Sinkhorn’s algorithm. *arXiv preprint: 2308.11791*, 2023.
- [14] Anqi Dong, Tryphon T. Georgiou, and Allen Tannenbaum. Promotion/inhibition effects in networks: A model with negative probabilities. *arXiv preprint: 2307.07738*, 2023.
- [15] Montacer Essid and Michele Pavon. Traversing the schrödinger bridge strait: Robert fortet’s marvelous proof redux. *Journal of Optimization Theory and Applications*, 181(1):23–60, 2019.
- [16] Matteo Garofalo, Thierry Nieuws, Paolo Massobrio, and Sergio Martinoia. Evaluation of the performance of information theory-based methods and cross-correlation to estimate the functional connectivity in cortical networks. *PloS one*, 4(8):e6482, 2009.
- [17] Tryphon T Georgiou. Relative entropy and the multivariable multidimensional moment problem. *IEEE Transactions on Information Theory*, 52(3):1052–1066, 2006.
- [18] Tryphon T. Georgiou and Michele Pavon. Positive contraction mappings for classical and quantum Schrödinger systems. *Journal of Mathematical Physics*, 56(3), 2015.
- [19] Clark Jeffries. Qualitative stability and digraphs in model ecosystems. *Ecology*, 55(6):1415–1419, 1974.
- [20] DO Logofet and NB Ulianov. Sign stability in model ecosystems: a complete class of sign-stable patterns. *Ecological modelling*, 16(2-4):173–189, 1982.
- [21] Zhi-Quan Luo and Paul Tseng. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.
- [22] Mehran Mesbahi. *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010.
- [23] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.
- [24] Daniel B Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, 2011.
- [25] Amit Surana, Can Chen, and Indika Rajapakse. Hypergraph similarity measures. *IEEE Transactions on Network Science and Engineering*, 10(2):658–674, 2022.
- [26] Emmanuel Tannenbaum and Eugene I Shakhnovich. Solution of the quasispecies model for an arbitrary gene network. *Physical Review E*, 70(2):021903, 2004.
- [27] Alejandro F Villaverde, John Ross, Federico Morán, and Julio R Banga. Mider: network inference with mutual information distance and entropy reduction. *PloS one*, 9(5):e96732, 2014.
- [28] Weiwei Yuan, Kangya He, Donghai Guan, and Guangjie Han. Edge-dual graph preserving sign prediction for signed social networks. *IEEE Access*, 5:19383–19392, 2017.
- [29] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph transformer networks. *Advances in neural information processing systems*, 32, 2019.
- [30] Thomas Zaslavsky. Signed graphs. *Discrete Applied Mathematics*, 4(1):47–74, 1982.