The Chow-Liu Tree and the Optimal Tree in Multimarginal Schrödinger Bridge

Georgiy Bondar *gbondar@ucsc.edu*
Department of Applied Mathematics, University of California Santa Cruz, USA

Abhishek Halder *ahalder@{iastate,ucsc}.edu*
Department of Aerospace Engineering, Iowa State University, USA
Department of Applied Mathematics, University of California Santa Cruz, USA

There is an intriguing similarity between the well-known Chow-Liu tree derived in Chow & Liu (1968), and the minimum spanning tree (MST) solving the optimal multimarginal Schrödinger bridge problem (MSBP) derived in Bondar & Halder (2025). For details, see Bondar & Halder (2025, Sec. III-B, Remark 1).

In both cases, the solution is an algorithm where a complete (undirected) weighted graph is constructed on say $s \geq 3$ vertices. The vertices are given probability measures, or in the discrete case, probability vectors $\boldsymbol{\mu}_\sigma$ where the vertex index $\sigma \in \{1, 2, \dots, s\}$. The vertices have no self-loops.

For the Chow-Liu tree, the weights are the *mutual information* I_{σ_1,σ_2} for all vertex indices $\sigma_1 \neq \sigma_2$. The Chow-Liu tree in Chow & Liu (1968) is the *maximum weight spanning tree* for this complete weighted graph.

For Bondar & Halder (2025), the weights are

$$g_{\sigma_1\sigma_2} = \texttt{SB}_\eta\left(\mu_{\sigma_1}, \mu_{\sigma_2}\right) + H(\mu_{\sigma_1}) + H(\mu_{\sigma_2}),$$

where $\texttt{SB}_\eta\left(\mu_{\sigma_1}, \mu_{\sigma_2}\right)$ denotes the *optimal value of the bimarginal SBP* between $\mu_{\sigma_1}, \mu_{\sigma_2}$, i.e., *the entropy-regularized optimal transport cost with regularization parameter* $\eta > 0$. The $H(\boldsymbol{p}) := \langle \boldsymbol{p}, \log \boldsymbol{p}\rangle$ denotes the Shannon entropy for the prpobability vector $\boldsymbol{p}$. In Bondar & Halder (2025), the optimal MSBP graph structure is shown to be the *minimum weight spanning tree* for this complete weighted graph.

Despite their structural similarities, the corresponding objectives and their MSTs seem to have no simple relations even though one can be written in terms of the other. Let us elaborate.

Recall that the mutual information

$$\begin{aligned} I_{\sigma_1\sigma_2} &= \mathrm{D}_{\mathrm{KL}}\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2} \parallel \mu_{\sigma_1} \otimes \mu_{\sigma_2}\right) && (1)\\ &= H(\mu_{\sigma_1}) + H(\mu_{\sigma_2}) - H\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2}\right). && (2) \end{aligned}$$

The first equality is in terms of the KL divergence D_{KL} while the latter is in terms of entropy H (Cover & Thomas, 1991, eq. (2.41)).

letting $K_{\sigma_1\sigma_2} := \exp(-C_{\sigma_1\sigma_2}/\eta)$, the pairwise cost in Bondar & Halder (2025) can be re-written as

$$\begin{aligned} g_{\sigma_1\sigma_2} &= \texttt{SB}_\eta\left(\mu_{\sigma_1}, \mu_{\sigma_2}\right) + H(\mu_{\sigma_1}) + H(\mu_{\sigma_2})\\ &= \mathrm{D}_{\mathrm{KL}}\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2} \parallel K_{\sigma_1\sigma_2}\left(\mu_{\sigma_1} \otimes \mu_{\sigma_2}\right)\right) + H(\mu_{\sigma_1}) + H(\mu_{\sigma_2})\\ &= \mathrm{D}_{\mathrm{KL}}\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2} \parallel K_{\sigma_1\sigma_2}\left(\mu_{\sigma_1} \otimes \mu_{\sigma_2}\right)\right) + I_{\sigma_1\sigma_2} + H\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2}\right) && (3)\\ &= \mathrm{D}_{\mathrm{KL}}\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2} \parallel K_{\sigma_1\sigma_2}\left(\mu_{\sigma_1} \otimes \mu_{\sigma_2}\right)\right) + \mathrm{D}_{\mathrm{KL}}\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2} \parallel \mu_{\sigma_1} \otimes \mu_{\sigma_2}\right) + H\left(M^{\mathrm{opt}}_{\sigma_1\sigma_2}\right), && (4) \end{aligned}$$

wherein (3) follows from (2), while (4) follows from (1). From the above equalities, it is unclear how the MST with edge-cost $I_{\sigma_1\sigma_2}$ can be related to that with edge-cost $g_{\sigma_1\sigma_2}$.

A different way to relate $I_{\sigma_1\sigma_2}$ and $g_{\sigma_1\sigma_2}$ is via Peyré et al. (2019, Remark 4.2), which (up to scaling) unwinds the term $K_{\sigma_1\sigma_2} := \exp(-C_{\sigma_1\sigma_2}/\eta)$. This gives

$$g_{\sigma_1\sigma_2} = \langle C_{\sigma_1\sigma_2}, M^{\text{opt}}_{\sigma_1\sigma_2} \rangle + \eta I_{\sigma_1\sigma_2} + H(\mu_{\sigma_1}) + H(\mu_{\sigma_2}). \tag{5}$$

Again it is unclear from here, how the corresponding MSTs could be related.

Being unable to mathematically relate them, a numerical experiment in Bondar & Halder (2025, Sec. IV-A) computed both the Chow-Liu tree and the optimal tree for the MSBP, for the same problem data. This experiment found these two trees to be different. Could it be the case that one tree is a transformation of the other? The answer is unknown to the authors.

References

Georgiy A. Bondar and Abhishek Halder. Optimal multimarginal schrödinger bridge: Minimum spanning tree over measure-valued vertices. *IEEE Control Systems Letters*, pp. 1–1, 2025. doi: 10.1109/LCSYS.2025.3634944.

C Chow and Cong Liu. Approximating discrete probability distributions with dependence trees. *IEEE transactions on Information Theory*, 14(3):462–467, 1968.

Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 1991.

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.