## Acknowledgments

We thank James Thornton for the DSB codebase[5] and useful discussions. AD acknowledges support from the Lagrange Mathematics and Computing Research Center. AD and MN would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme *The mathematical and statistical foundation of future data-driven engineering* when work on this paper was undertaken. MN acknowledges funding from the grant SCAI (ANR-19-CHIA-0002).

## References

Acciaio, B., Backhoff-Veraguas, J., and Carmona, R. Extended mean field control problems: stochastic maximum principle and transport perspective. *SIAM journal on Control and Optimization*, 57 (6):3666–3693, 2019.

Agueh, M. and Carlier, G. Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

Álvarez-Esteban, P. C., Del Barrio, E., Cuesta-Albertos, J., and Matrán, C. A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications*, 441(2): 744–762, 2016.

Bayraktar, E., Cox, A. M., and Stoev, Y. Martingale optimal transport with stopping. *SIAM Journal on Control and Optimization*, 56(1):417–433, 2018.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

Bunne, C., Papaxanthos, L., Krause, A., and Cuturi, M. Proximal optimal transport modeling of population dynamics. In *International Conference on Artificial Intelligence and Statistics*, pp. 6511–6528. PMLR, 2022.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pp. 213–229. Springer, 2020.

Carlier, G. On the linear convergence of the multimarginal Sinkhorn algorithm. *SIAM Journal on Optimization*, 32(2):786–794, 2022.

Cattiaux, P., Conforti, G., Gentil, I., and Léonard, C. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021.

Chen, T., Liu, G.-H., and Theodorou, E. A. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. *International Conference on Learning Representations*, 2022.

Chen, T., Liu, G.-H., Tao, M., and Theodorou, E. A. Deep momentum multi-marginal Schrödinger bridge. 2023. doi: 10.48550/ARXIV.2303.01751.

Chen, Y., Georgiou, T., Pavon, M., and Tannenbaum, A. Robust transport over networks. *IEEE Transactions on Automatic Control*, 62(9):4675–4682, 2016.

Chen, Y., Conforti, G., and Georgiou, T. T. Measure-valued spline curves: An optimal transport viewpoint. *SIAM Journal on Mathematical Analysis*, 50(6):5947–5968, 2018.

Chen, Y., Georgiou, T. T., and Pavon, M. Optimal transport in systems and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 4, 2021.

Chizat, L. Doubly regularized entropic Wasserstein barycenters. *arXiv preprint arXiv:2303.11844*, 2023.

Claici, S., Chien, E., and Solomon, J. Stochastic Wasserstein barycenters. In *International Conference on Machine Learning*, pp. 999–1008. PMLR, 2018.

[5] https://github.com/JTT94/diffusion_schrodinger_bridge

Cohen, S., Arbel, M., and Deisenroth, M. P. Estimating barycenters of measures in high dimensions. *arXiv preprint arXiv:2007.07105*, 2020.

Cotar, C., Friesecke, G., and Klüppelberg, C. Density functional theory and optimal transportation with Coulomb cost. *Communications on Pure and Applied Mathematics*, 66(4):548–599, 2013.

Csiszár, I. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, pp. 146–158, 1975.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26, 2013.

Cuturi, M. and Doucet, A. Fast computation of Wasserstein barycenters. In *International Conference on Machine Learning*, pp. 685–693. PMLR, 2014.

Dai Pra, P. A stochastic control approach to reciprocal diffusion processes. *Applied Mathematics and Optimization*, 23(1):313–329, 1991.

De Bortoli, V., Thornton, J., Heng, J., and Doucet, A. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34: 17695–17709, 2021.

Dupuis, P. and Ellis, R. S. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011.

Eisenberger, M., Toker, A., Leal-Taixé, L., and Cremers, D. Deep shells: Unsupervised shape correspondence with optimal transport. *Advances in Neural Information Processing Systems*, 33: 10491–10502, 2020.

Fan, J., Taghvaei, A., and Chen, Y. Scalable computations of Wasserstein barycenter via input convex neural networks. *arXiv preprint arXiv:2007.04462*, 2020.

Feydy, J., Charlier, B., Vialard, F.-X., and Peyré, G. Optimal transport for diffeomorphic registration. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, pp. 291–299. Springer, 2017.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. POT: Python optimal transport. *The Journal of Machine Learning Research*, 22(1):3571–3578, 2021.

Fortet, R. Résolution d'un système d'équations de M. Schrödinger. *Journal de Mathématiques Pures et Appliquées*, 19(1-4):83–105, 1940.

Haasler, I., Ringh, A., Chen, Y., and Karlsson, J. Multimarginal Optimal Transport with a tree-structured cost and the Schrödinger bridge problem. *SIAM Journal on Control and Optimization*, 59(4):2428–2453, 2021.

Haussmann, U. G. and Pardoux, E. Time reversal of diffusions. *The Annals of Probability*, pp. 1188–1205, 1986.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Hyvärinen, A. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Karcher, H. Riemannian center of mass and so called Karcher mean. *arXiv preprint arXiv:1407.2087*, 2014.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Knight, P. A. The Sinkhorn–Knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275, 2008.

Kober, H. A theorem on Banach spaces. *Compositio Mathematica*, 7:135–140, 1940.

Koller, D. and Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

Korotin, A., Li, L., Solomon, J., and Burnaev, E. Continuous Wasserstein-2 barycenter estimation without minimax optimization. *arXiv preprint arXiv:2102.01752*, 2021.

Korotin, A., Egiazarian, V., Li, L., and Burnaev, E. Wasserstein iterative networks for barycenter estimation. *arXiv preprint arXiv:2201.12245*, 2022.

Kullback, S. Probability densities with given marginals. *The Annals of Mathematical Statistics*, 39 (4):1236–1243, 1968.

LeCun, Y. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

Léonard, C. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete & Continuous Dynamical Systems-A*, 34(4):1533–1574, 2014.

Li, L., Genevay, A., Yurochkin, M., and Solomon, J. M. Continuous regularized Wasserstein barycenters. *Advances in Neural Information Processing Systems*, 33:17755–17765, 2020.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *International Conference on Learning Representations*, 2023.

Luo, Z.-Q. and Tseng, P. On the convergence of the coordinate descent method for convex differentiable minimization. *Journal of Optimization Theory and Applications*, 72(1):7–35, 1992.

Makkuva, A., Taghvaei, A., Oh, S., and Lee, J. Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning*, pp. 6672–6681. PMLR, 2020.

Marino, S. D. and Gerolin, A. An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. *Journal of Scientific Computing*, 85(2):27, 2020.

Mi, L., Yu, T., Bento, J., Zhang, W., Li, B., and Wang, Y. Variational Wasserstein barycenters for geometric clustering. *arXiv preprint arXiv:2002.10543*, 2020.

Minsker, S., Srivastava, S., Lin, L., and Dunson, D. Scalable and robust Bayesian inference via the median posterior. In *International Conference on Machine Learning*, pp. 1656–1664. PMLR, 2014.

Nichol, A. Q. and Dhariwal, P. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171, 2021.

Nutz, M. Introduction to entropic optimal transport, 2021.

Pele, O. and Werman, M. Fast and robust Earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision*, pp. 460–467. IEEE, 2009.

Peluchetti, S. Diffusion bridge mixture transports, Schrödinger bridge problems and generative modeling. *arXiv preprint arXiv:2304.00917*, 2023.

Peyré, G., Cuturi, M., et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

Rabin, J., Peyré, G., Delon, J., and Bernot, M. Wasserstein barycenter and its application to texture mixing. In *Scale Space and Variational Methods in Computer Vision: Third International Conference, SSVM 2011, Ein-Gedi, Israel, May 29–June 2, 2011, Revised Selected Papers 3*, pp. 435–446. Springer, 2012.

Ruschendorf, L. Convergence of the Iterative Proportional Fitting procedure. *The Annals of Statistics*, pp. 1160–1174, 1995.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Schmitz, M. A., Heitz, M., Bonneel, N., Ngole, F., Coeurjolly, D., Cuturi, M., Peyré, G., and Starck, J.-L. Wasserstein dictionary learning: Optimal transport-based unsupervised nonlinear dictionary learning. *SIAM Journal on Imaging Sciences*, 11(1):643–678, 2018.

Schrödinger, E. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2(4):269–310, 1932.

Shi, Y., De Bortoli, V., Campbell, A., and Doucet, A. Diffusion Schrödinger bridge matching. *arXiv preprint arXiv:2303.16852*, 2023.

Sinkhorn, R. and Knopp, P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

Solomon, J., Rustamov, R., Guibas, L., and Butscher, A. Wasserstein propagation for semi-supervised learning. In *International Conference on Machine Learning*, pp. 306–314. PMLR, 2014.

Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., and Guibas, L. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4):1–11, 2015.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.

Srivastava, S., Li, C., and Dunson, D. B. Scalable Bayes via barycenter in Wasserstein space. *The Journal of Machine Learning Research*, 19(1):312–346, 2018.

Staib, M., Claici, S., Solomon, J. M., and Jegelka, S. Parallel streaming Wasserstein barycenters. *Advances in Neural Information Processing Systems*, 30, 2017.

Stroock, D. W. and Varadhan, S. S. *Multidimensional diffusion processes*, volume 233. Springer Science & Business Media, 1997.

Su, Z., Wang, Y., Shi, R., Zeng, W., Sun, J., Luo, F., and Gu, X. Optimal mass transport for shape matching and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (11):2246–2259, 2015.

Valiente, G. *Algorithms on Trees and Graphs*, volume 112. Springer, 2002.

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. Solving Schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134, 2021.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

## Appendix organization

First, additional notation is introduced in Appendix A. Then, we briefly recall some notions on undirected and directed trees in Appendix B. Similarly, martingale problems are introduced in Appendix C. The proofs of the main manuscript and additional theoretical results on Tree Schrödinger Bridges are given in Appendix D. Additional details on our consideration of the tree-based static SB problem are described in Appendix E. Details on the implementation of TreeDSB are given in Appendix F and the experiments are investigated in Appendix G.

## A Additional notation

For any finite set $\mathsf{E}$, we equivalently refer to the cardinal of $\mathsf{E}$ as $\mathrm{card}(\mathsf{E})$ or $|\mathsf{E}|$. Let $(\mathsf{X}, \mathcal{X})$ be a measurable space. For any $x \in (\mathbb{R}^d)^{\ell+1}$ and any $m \in \{0, \dots, \ell\}$, let $x_{-m} = (x_0, \dots, x_{m-1}, x_{m+1}, \dots, x_\ell)$. For any family of measures $\{\nu_j\}_{j\in\{0,\dots,\ell\}}$ defined on $(\mathsf{X}, \mathcal{X})$ and any $i \in \{0, \dots, \ell\}$, let $\nu_{-i} = \bigotimes_{j\in\{0,\dots,\ell\}\setminus\{i\}} \nu_j$. Let $I = \{i_1, \dots, i_q\} \subset \{1, \dots, \ell\}$ and $\mu \in \mathscr{P}^{(\ell)}$ such that $\mu \ll \mathrm{Leb}$. We define $I^c = \{1, \dots, \ell\}\backslash I$ and denote it by $\{i_1^c, \dots, i_{\bar{q}}^c\}$ where $\bar{q} = \ell - q$. We denote the marginal of $\mu$ along $I$ by $\mu_I$, *i.e.*, $\mu_I \in \mathscr{P}^{(q)}$ and we have for any $\mathsf{A} \in \mathcal{B}((\mathbb{R}^d)^q)$, $\mu_I(\mathsf{A}) = \int_{\mathsf{X}} \mu(x) \prod_{j=1}^q \delta_{x_{i_j}}(\mathsf{A}_j)\mathrm{d}x$. In addition, note that $\mu_I \ll \mathrm{Leb}$. We denote the conditional distribution of $\mu$ given $I$ by $\mu_{|I}(\cdot|\cdot)$, *i.e.*, $\mu_{|I}(\cdot|\cdot) \in \mathscr{P}^{(\bar{q})} \times (\mathbb{R}^d)^q$ and we have for any $y \in (\mathbb{R}^d)^q$ and any $\mathsf{A} \in \mathcal{B}((\mathbb{R}^d)^{\bar{q}})$, $\mu_{|I}(\mathsf{A}|y) = \int_{\mathsf{X}} \mu(x)/\mu_I(y) \prod_{j=1}^q \delta(x_{i_j} - y_j) \prod_{j'=1}^{\bar{q}} \delta_{x_{i^c_{j'}}}(\mathsf{A}_{j'})\mathrm{d}x$. Remark that for any $y \in (\mathbb{R}^d)^q$, $\mu_{|I}(\cdot|y) \ll \mathrm{Leb}$. For any subset $\mathsf{J} \subset \mathsf{I}^c$ with $\mathrm{card}(\mathsf{J}) = q_{\mathsf{J}}$, we also define $\mu_{\mathsf{J}|\mathsf{I}}(\cdot|\cdot) \in \mathscr{P}^{(q_{\mathsf{J}})} \times (\mathbb{R}^d)^q$ such that for any $y \in (\mathbb{R}^d)^q$, $\mu_{\mathsf{J}|\mathsf{I}}(\cdot|y) = \{\mu_{|\mathsf{I}}(\cdot|y)\}_{\mathsf{J}}$. For a collection of functions $\{f_i\}_{i\in\mathsf{I}}$, with $\mathsf{I} \subset \{1, \dots, n\}$ and $n \in \mathbb{N}$ such that $f_i : \mathbb{R}^d \to \mathbb{R}$, we define $\oplus_{i\in\mathsf{I}} f_i : (\mathbb{R}^d)^n \to \mathbb{R}$ such that for any $x = (x_1, \dots, x_n) \in (\mathbb{R}^d)^n$, $\oplus_{i\in\mathsf{I}} f(x) = \sum_{i\in\mathsf{I}} f_i(x_i)$.

## B Introduction to trees

**Undirected tree.** An undirected graph $\mathsf{T} = (\mathsf{V}, \mathsf{E})$, with vertices $\mathsf{V}$ and edges $\mathsf{E}$, is said to be an *undirected tree* if it is *acyclic* and *connected* (Valiente, 2002, Definition 1.19.). In particular, we have $\mathrm{card}(\mathsf{E}) = \mathrm{card}(\mathsf{V}) - 1$. The undirected edge between two nodes $v_1$ and $v_2$ is similarly denoted by $\{v_1, v_2\}$ or $\{v_2, v_1\}$. We say that $\mathsf{T}' = (\mathsf{V}', \mathsf{E}')$ is a *sub-tree* of $\mathsf{T}$ if $\mathsf{T}'$ is an undirected tree with vertices $\mathsf{V}' \subset \mathsf{V}$ and edges $\mathsf{E}' \subset \mathsf{E}$. For any vertex $v \in \mathsf{V}$, we define the set of its *neighbours* $\mathsf{N}_v$ as the set of vertices $v' \in \mathsf{V}$ such that $\{v, v'\} \in \mathsf{E}$. The integer $\mathrm{card}(\mathsf{N}_v)$ is referred to as the degree of $v$. The vertices with degree 1 are called *leaves*, and we denote the set of leaves by $\mathsf{V}_\mathsf{L} \subset \mathsf{V}$. The (unique) *path* in $\mathsf{T}$ between two vertices $v$ and $v'$ is the sequence of two-by-two distinct edges $\{\{v_i, v_{i+1}\}\}_{i=1}^n$ (with $n \geq 1$) such that $v_k = v_{k+1}$ for any $k \in \{1, \dots, n\}$ such that $k = 0 \bmod(2)$, $v_1 = v$ and $v_{n+1} = v'$. This path can be seen as a linear sub-tree of $\mathsf{T}$, and we define $n$ as the *length* of this path. We say that $\mathsf{T}$ is *weighted* if there exists a map $w : \mathsf{E} \mapsto \mathbb{R}_+$; in this case, $w(\{v_1, v_2\})$, or equivalently $w(\{v_2, v_1\})$ (also denoted by $w_{v_1,v_2}$ or $w_{v_2,v_1}$ ) is called the weight of the edge $\{v_1, v_2\}$. The tree $\mathsf{T}$ is said to be *rooted* in $r \in \mathsf{V}$ if $r$ defines a partial ordering $\leq_{\mathsf{T},r} \subset \mathsf{V} \times \mathsf{V}$ such that for any $v_1, v_2 \in \mathsf{V}$, $v_1 \leq_{\mathsf{T},r} v_2$ if the node $v_1$ lies on the unique path between $r$ and $v_2$.

**Directed tree.** Consider a directed graph $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$ rooted in $r \in \mathsf{V}$. Any directed edge $e \in \mathsf{E}_r$ from $v_1 \in \mathsf{V}$ to $v_2 \in \mathsf{V}$ is denoted by $(v_1, v_2)$. $\mathsf{T}_r$ is a said to be a *directed tree* rooted in $r$ if (i) the underlying undirected graph $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ is an undirected tree rooted in $r$ and (ii) any $(v_1, v_2) \in \mathsf{E}_r$ is directed according to the partial ordering $\leq_{\mathsf{T},r}$, *i.e.*, $\{v_1, v_2\} \in \mathsf{E}$ and $v_1 \leq_{\mathsf{T},r} v_2$. For any vertices $(v, v') \in \mathsf{V} \times \mathsf{V}$ such that $v \leq_{\mathsf{T},r} v'$, the (unique) *path* in $\mathsf{T}_r$ from $v$ to $v'$, denoted by $\mathrm{path}_{\mathsf{T}_r}(v, v')$, is defined as the directed version of the path in $\mathsf{T}$ between $v$ and $v'$ (viewed as a sub-tree of $\mathsf{T}$), which is rooted in $v$. We say that $\mathsf{T}_r$ is *weighted*, if $\mathsf{T}$ is weighted and the edges of $\mathsf{T}_r$ have the same weights as the corresponding undirected edges of $\mathsf{T}$. For any $(v_1, v_2) \in \mathsf{E}_r$, we denote this weight by $w_{v_1,v_2}$. We say that $\mathsf{T}_r$ is the (unique) *directed version* of $\mathsf{T}$ rooted in $r$. It is endowed with a canonical vertex numbering $\zeta : \mathsf{V} \to \{0, \dots, \mathrm{card}(\mathsf{V}) - 1\}$, corresponding to a depth-first traversal of its nodes, starting from the root $r$ (Valiente, 2002, Definition 3.1.). This numbering is consistent with the partial ordering on $\mathsf{T}$, *i.e.*, if $v_1 \leq_{\mathsf{T},r} v_2$, $\zeta(v_1) \leq \zeta(v_2)$, and satisfies $\zeta(r) = 0$. In the rest of the paper, we will write in an equivalent manner $v$ or $\zeta(v)$.

For any vertices $(v_1, v_2) \in \mathsf{E} \times \mathsf{E}$ such that $v_1 \leq_{\mathsf{T},r} v_2$, $\mathrm{path}_{\mathsf{T}_r}(v_1, v_2)$ corresponds to the ordered set of edges in $\mathsf{E}_r$ which define the ordered path between two vertices $v_1$ and $v_2$. For any vertex $v \in \mathsf{V}$, we define:

(a) the set of its *children* $\mathsf{C}_v$ as the set of vertices $v' \in \mathsf{V}$ such that $(v, v') \in \mathsf{E}_r$. In particular, for any $v \in \mathsf{V}_L$, the set of leaves, one has $\mathsf{C}_v = \emptyset$.

(b) its *parent* as the unique vertex $p(v)$ such that $(p(v), v) \in \mathsf{E}_r$, if $v \neq r$ (the parent of the root is not defined).

Note that $\mathsf{N}_r = \mathsf{C}_r$ and, for any vertex $v \in \mathsf{V}\backslash\{r\}$, $\mathsf{N}_v = \{p(v)\} \cup \mathsf{C}_v$.

**Definition 8** (Tree-structured directed Probabilistic Graphical Model (PGM))**.** *Consider a directed tree $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$. The directed PGM induced by $\mathsf{T}_r$ (Koller & Friedman, 2009, Definition 3.4.), denoted by $\mathscr{P}_{\mathsf{T}_r}$, is the family of distributions $\pi \in \mathscr{P}^{(|\mathsf{V}|)}$ which have a Markovian factorization along $\mathsf{T}_r$, i.e.,*

$$\mathscr{P}_{\mathsf{T}_r} = \{\pi \in \mathscr{P}^{(|\mathsf{V}|)} : \pi = \pi_r \bigotimes\nolimits_{(v,v') \in \mathsf{E}_r} \pi_{v'|v}\} \ .$$

**Lemma 9.** *Consider an undirected tree $\mathsf{T} = (\mathsf{V}, \mathsf{E})$. Let $(r, r') \in \mathsf{V} \times \mathsf{V}$. Let $\mathsf{T}'$ be a sub-tree of $\mathsf{T}$ with vertices $\mathsf{V}'$ such that $r' \in \mathsf{V}'$. Denote by $\mathsf{T}'_{r'}$ the directed version of $\mathsf{T}'$ rooted in $r'$. Then, for any $\pi \in \mathscr{P}_{\mathsf{T}_r}$, we have $\pi_{\mathsf{V}'} \in \mathscr{P}_{\mathsf{T}'_{r'}}$.*

*Proof.* Let $(r, r') \in \mathsf{V} \times \mathsf{V}$. We denote by $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$, respectively $\mathsf{T}_{r'} = (\mathsf{V}, \mathsf{E}_{r'})$, the directed version of $\mathsf{T}$ rooted in $r$, respectively $r'$. We define the paths $\mathsf{P}_{r,r'} = \mathrm{path}_{\mathsf{T}_r}(r, r') \subset \mathsf{E}_r$ and $\mathsf{P}_{r',r} = \mathrm{path}_{\mathsf{T}_{r'}}(r', r) \subset \mathsf{E}_{r'}$. It is easy to see that

(a) $\mathsf{E}_r \backslash \mathsf{P}_{r,r'} = \mathsf{E}_{r'} \backslash \mathsf{P}_{r',r}$,

(b) $\mathsf{P}_{r,r'} = \{(v_2, v_1) : (v_1, v_2) \in \mathsf{P}_{r',r}\}$,

(c) $\mathsf{P}_{r',r} = \{(v_2, v_1) : (v_1, v_2) \in \mathsf{P}_{r,r'}\}$.

Let $\pi \in \mathscr{P}_{\mathsf{T}_r}$. First note that for any $(v_1, v_2) \in \mathsf{E}_r$, we have by Bayes decomposition $\pi_{v_1}\pi_{v_2|v_1} = \pi_{v_2}\pi_{v_1|v_2} = \pi_{v_1,v_2}$. Then it comes

$$\begin{aligned}
\pi &= \pi_r \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_r} \pi_{v_2|v_1} \\
&= \pi_r \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{P}_{r,r'}} \pi_{v_2|v_1} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_r \backslash \mathsf{P}_{r,r'}} \pi_{v_2|v_1} \\
&= \pi_r \bigotimes\nolimits_{(v_2,v_1) \in \mathsf{P}_{r',r}} \pi_{v_2|v_1} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_{r'} \backslash \mathsf{P}_{r',r}} \pi_{v_2|v_1} \\
&= \pi_{r'} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{P}_{r',r}} \pi_{v_2|v_1} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_{r'} \backslash \mathsf{P}_{r',r}} \pi_{v_2|v_1} \\
&= \pi_{r'} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_{r'}} \pi_{v_2|v_1} \ ,
\end{aligned}$$

and therefore, we have $\pi \in \mathscr{P}_{\mathsf{T}_{r'}}$.

Let $\mathsf{T}'$ be a sub-tree of $\mathsf{T}$ with vertices $\mathsf{V}'$ such that $r' \in \mathsf{V}'$. First note that $\mathsf{E}'_{r'} \subset \mathsf{E}_{r'}$. Using the previous computation, we have for any $\mathsf{A} \in \mathcal{B}((\mathbb{R}^d)^{|\mathsf{V}'|})$,

$$\begin{aligned}
\pi_{\mathsf{V}'}(\mathsf{A}) &= \int_{(\mathbb{R}^d)^{|\mathsf{V}|}} \pi_{r'}(x_{r'}) \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_{r'}} \pi_{v_2|v_1}(x_{v_2}|x_{v_1}) \prod\nolimits_{v' \in \mathsf{V}'} \delta_{x_{v'}}(\mathsf{A}_{v'}) \mathrm{d}x \\
&= \int_{(\mathbb{R}^d)^{|\mathsf{V}|-|\mathsf{V}'|}} \{\pi_{r'}(\mathsf{A}_{r'}) \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}'_{r'}} \pi_{v_2|v_1}(\mathsf{A}_{v_2}|x_{v_1})\} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}_{r'} \backslash \mathsf{E}'_{r'}} \pi_{v_2|v_1}(x_{v_2}|x_{v_1}) \mathrm{d}x_{\mathsf{V} \backslash \mathsf{V}'} \\
&= \{\pi_{r'} \bigotimes\nolimits_{(v_1,v_2) \in \mathsf{E}'_{r'}} \pi_{v_2|v_1}\}(\mathsf{A}) \ ,
\end{aligned}$$

which proves that $\pi_{\mathsf{V}'} \in \mathscr{P}_{\mathsf{T}'_{r'}}$. □

**Discretized undirected tree.** Let $N \geq 1$. Consider an undirected tree $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with weights $w$. We say that $\mathsf{T}^{(N)} = (\mathsf{V}^{(N)}, \mathsf{E}^{(N)})$ is a $N$-discretized version of $\mathsf{T}$ if it is an undirected tree with weights $w^{(N)}$ such that

(a) $\mathsf{V}^{(N)} = \mathsf{V} \bigsqcup \cup_{\substack{e \in \mathsf{E}, \\ k \in \{1,\ldots,N-1\}}} \{v_e^k\}$,

(b) $\mathsf{E}^{(N)} = \cup_{e \in \mathsf{E}} \cup_{k=0,\dots,N-1} \left\{\{v_e^k, v_e^{k+1}\}\right\}$ with the convention that the vertices $v_e^N$ and $v_e^N$ are defined such that $\{v_e^0, v_e^N\} = e$,

(c) $\sum_{e \in \mathrm{path}_{\mathsf{T}}(v,v')} 1/w_e^{(N)} = 1/w_{v,v'}$, if $\{v, v'\} \in \mathsf{E}$.

Remark that the leaves of $\mathsf{T}^{(N)}$ are exactly the original leaves of $\mathsf{T}$ and that $\mathsf{T}^{(1)} = \mathsf{T}$. The non-uniqueness of $\mathsf{T}^{(N)}$ comes from the freedom of choice on the weights of its edges.

**Discretized directed tree.** Let $N \geq 1$. Consider a directed tree $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$ rooted in $r \in \mathsf{V}$ with weights $w$. We say that $\mathsf{T}_r^{(N)} = (\mathsf{V}^{(N)}, \mathsf{E}_r^{(N)})$ is a $N$-discretized version of $\mathsf{T}_r$ if it is the directed version of $\mathsf{T}^{(N)}$ rooted in $r$, where $\mathsf{T}^{(N)}$ is a $N$-discretized version of the underlying undirected tree of $\mathsf{T}_r$.

# C Background on martingale problems

In this section, we introduce the background on Stochastic Differential Equations (SDEs) and weak solutions of SDEs following the framework of (Stroock & Varadhan, 1997, Section 10.1, page 249). We recall that $\mathrm{C}_0^\infty(\mathbb{R}^d)$ is the space of infinitely differentiable real-valued functions which vanish at infinity. In addition, we have that $\mathcal{S}_+^d$ is the space of $d \times d$, symmetric, non-negative matrices.

**Definition 10.** *Let $T > 0$ or $T = +\infty$, $\sigma : [0, T) \times \mathbb{R}^d \to \mathcal{S}_+^d$ and $b : [0, T) \times \mathbb{R}^d \to \mathbb{R}^d$, locally bounded measurable functions. We define the* infinitesimal generator, *$\mathcal{A}$, given for any $f \in \mathrm{C}_0^\infty(\mathbb{R}^d)$, $t \in [0, T)$ and $x \in \mathbb{R}^d$ by*

$$\mathcal{A}_t(f)(x) = \langle b_t(x), \nabla f(x)\rangle + \tfrac{1}{2}\langle \sigma_t(x)\sigma_t(x)^\top, \nabla^2 f(x)\rangle. \tag{4}$$

*We say that a probability measure $\mathbb{P}$* satisfies the martingale problem for *$\mathcal{A}$ if for any $t \in [0, T)$ and $f \in \mathrm{C}_0^\infty(\mathbb{R}^d)$, we have that $(f(\mathbf{X}_t) - \int_0^t \mathcal{A}_s(f)(\mathbf{X}_s)\mathrm{d}s)_{s \in [0,t]}$ is a $\mathbb{P}$-martingale.*

In the main document, see Section 2, we say that "a path measure $\mathbb{P}$ is associated with $\mathrm{d}\mathbf{X}_t = b(t, \mathbf{X}_t)\mathrm{d}t + \sigma(t, \mathbf{X}_t)\mathrm{d}\mathbf{B}_t$ with $(\mathbf{B}_t)_{t \geq 0}$ a $d$-dimensional Brownian motion" if $\mathbb{P}$ solves the martingale problem associated with $\mathcal{A}$ given by (4). Unless specified, we always assume that such a path measure exists and is unique. Below, we recall the following theorem, see (Stroock & Varadhan, 1997, Theorem 10.2.2), which gives sufficient conditions for the existence and uniqueness of solutions to the martingale problem.

**Theorem 11.** *Assume that for any $x \in \mathbb{R}^d$ we have*

$$\inf\{\langle \theta, \sigma\sigma^\top(s, x)\theta\rangle \, : \, \theta \in \mathbb{R}^d, \ \|\theta\| = 1, \ s \in [0, T]\} > 0,$$
$$\lim_{y \to x} \sup\{\|\sigma(s, x) - \sigma(s, y)\| \, : \, s \in [0, T]\} = 0.$$

*In addition, assume that there exists $C > 0$ such that for any $x \in \mathbb{R}^d$*

$$\sup\{\|\sigma\sigma^\top(t, x)\| \, : \, s \in [0, T]\} + \sup\{\langle x, b(t, x)\rangle \, : \, s \in [0, T]\} \leq C(1 + \|x\|^2).$$

*Then, there exists a unique solution to the martingale problem with initialization $x_0 \in \mathbb{R}^d$.*

# D Theoretical results on Tree Schrödinger Bridges

We respectively provide in Appendix D.1, Appendix D.2 and Appendix D.3 the proofs of the results of the main manuscript presented in Section 3, Section 4 and Section 5. Finally, we make a detailed comparison between our setting and the framework of Haasler et al. (2021) in Appendix D.4. In the rest of this section, we consider an undirected tree $\mathsf{T} = (\mathsf{V}, \mathsf{E})$, where $|\mathsf{V}| = \ell + 1$, and some subset $\mathsf{S} \subset \mathsf{V}$ which we denote by $\mathsf{S} = \{i_0, \dots, i_{K-1}\}$. We define $\mathsf{S}^c = \mathsf{V} \backslash \mathsf{S}$.

## D.1 Proofs of Section 3

Proposition 1 is straightforward to obtain by combining the definition of the Brownian motion with the definition of $\pi^0$ given in (2). The following lemma details the recursion relation between the (mIPF) iterates, which is key to prove Proposition 2.

**Lemma 12.** *Let $(\pi^n)_{n\in\mathbb{N}}$ be the sequence given by (mIPF). Let $n \in \mathbb{N}$, $k_n = (n-1) \bmod (K)$, $k_n + 1 = n \bmod (K)$. Denote by $\mathsf{T}_{k_n}$, respectively $\mathsf{T}_{k_n+1}$ with edges $\mathsf{E}_{k_n+1}$, the directed version of $\mathsf{T}$ rooted in $i_{k_n}$, respectively in $i_{k_n+1}$. We have:*

*(i)* $\pi^n \in \mathscr{P}_{\mathsf{T}_{k_n}}$,

*(ii)* $\pi^{n+1} = \mu_{i_{k_n+1}} \bigotimes_{(v,v')\in\mathsf{E}_{k_n+1}} \pi^n_{v'|v}$. *In particular, for any* $(v,v') \in \mathsf{E}_{k_n+1}$, $\pi^{n+1}_{v'|v} = \pi^n_{v'|v}$.

*Proof.* We show the result (i) by recursion on $n \in \mathbb{N}$, and will deduce (ii) from the proof. Using (2), we first have $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$, where $r$ is chosen as $i_{K-1}$, see Section 3. Thus, we obtain the result (i) at step $n = 0$. Assume now that $\pi^n \in \mathscr{P}_{\mathsf{T}_{k_n}}$ for some $n \in \mathbb{N}$.

Consider the paths $\mathsf{P}_n = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$ and $\mathsf{P}_{n+1} = \mathrm{path}_{\mathsf{T}_{k_n+1}}(i_{k_n+1}, i_{k_n})$. Note that these two paths have the same length, denoted by $J$, and contain the same vertices, denoted by $\mathsf{V}_n$. Let $\pi \in \mathscr{P}^{(\ell+1)}$ such that $\mathrm{KL}(\pi|\pi^n) < +\infty$. We have the following decomposition

$$\mathrm{KL}(\pi|\pi^n) = \mathrm{KL}(\pi_{\mathsf{V}_n}|\pi^n_{\mathsf{V}_n}) + \int_{(\mathbb{R}^d)^{J+1}} \mathrm{KL}(\pi_{|\mathsf{V}_n}|\pi^n_{|\mathsf{V}_n})\mathrm{d}\pi_{\mathsf{V}_n}(x_{\mathsf{V}_n}) \,.$$

Hence, the $(n+1)$-th iterate of (mIPF) is given by $\pi^{n+1} = \pi^{n+1}_{\mathsf{V}_n} \otimes \pi^n_{|\mathsf{V}_n}$, with

$$\pi^{n+1}_{\mathsf{V}_n} = \mathrm{argmin}\{\mathrm{KL}(\pi|\pi^n_{\mathsf{V}_n}) \,:\, \pi \in \mathscr{P}^{(J+1)},\ \pi_{i_{k_n+1}} = \mu_{i_{k_n+1}}\} \,.$$

Since $\pi^n \in \mathscr{P}_{\mathsf{T}_{k_n}}$, we have (i) $\pi^n_{|\mathsf{V}_n} = \bigotimes_{(v,v')\in\mathsf{E}_{k_n}\setminus\mathsf{P}_n} \pi^n_{v'|v}$ and (ii) $\pi^n_{\mathsf{V}_n} \in \mathscr{P}_{\mathsf{P}_{n+1}}$ by Lemma 9, where $\mathsf{P}_{n+1}$ is viewed as a directed tree rooted in $i_{k_n+1}$. Defining $\mathsf{V}_{n+1} = \mathsf{V}_n\setminus\{i_{k_n+1}\}$, we thus have $\pi^n_{\mathsf{V}_n} = \pi^n_{i_{k_n+1}} \otimes \pi^n_{\mathsf{V}_{n+1}|i_{k_n+1}}$ where $\pi^n_{\mathsf{V}_{n+1}|i_{k_n+1}} = \bigotimes_{(v,v')\in\mathsf{P}_{n+1}} \pi^n_{v'|v}$.

Let $\pi \in \mathscr{P}^{(J+1)}$ such that $\pi_{i_{k_n+1}} = \mu_{i_{k_n+1}}$ and $\mathrm{KL}(\pi|\pi^n_{\mathsf{V}_n}) < +\infty$. Similarly to the previous computation, we have the following decomposition

$$\begin{aligned}\mathrm{KL}(\pi|\pi^n_{\mathsf{V}_n}) &= \mathrm{KL}(\pi_{i_{k_n+1}}|\pi^n_{i_{k_n+1}}) + \int_{\mathbb{R}^d} \mathrm{KL}(\pi_{|i_{k_n+1}}|\pi^n_{\mathsf{V}_{n+1}|i_{k_n+1}})\mathrm{d}\pi_{i_{k_n+1}}(x_{i_{k_n+1}})\\ &= \mathrm{KL}(\mu_{i_{k_n+1}}|\pi^n_{i_{k_n+1}}) + \int_{\mathbb{R}^d} \mathrm{KL}(\pi_{|i_{k_n+1}}|\pi^n_{\mathsf{V}_{n+1}|i_{k_n+1}})\mathrm{d}\mu_{i_{k_n+1}}(x_{i_{k_n+1}}) \,.\end{aligned}$$

Therefore, we obtain

$$\pi^{n+1}_{\mathsf{V}_n} = \mu_{i_{k_n+1}} \otimes \pi^n_{\mathsf{V}_{n+1}|i_{k+1}} = \mu_{i_{k+1}} \bigotimes_{(v,v')\in\mathsf{P}_{n+1}} \pi^n_{v'|v} \,.$$

Noting that $\mathsf{E}_{k_n}\setminus\mathsf{P}_n = \mathsf{E}_{k_n+1}\setminus\mathsf{P}_{n+1}$ and recalling that $\pi^{n+1} = \pi^{n+1}_{\mathsf{V}_n} \otimes \pi^n_{|\mathsf{V}_n}$, it finally comes

$$\pi^{n+1} = \mu_{i_{k_n+1}} \bigotimes_{(v,v')\in\mathsf{P}_{n+1}} \pi^n_{v'|v} \bigotimes_{(v,v')\in\mathsf{E}_{k_n+1}\setminus\mathsf{P}_{n+1}} \pi^n_{v'|v} = \mu_{i_{k_n+1}} \bigotimes_{(v,v')\in\mathsf{E}_{k_n+1}} \pi^n_{v'|v} \,. \tag{5}$$

Therefore, $\pi^{n+1} \in \mathscr{P}_{\mathsf{T}_{k_n+1}}$, which achieves the recursion for (i), and we obtain (ii) by (5). □

Hence, Lemma 12 shows that the (mIPF) iterates admit a Markovian factorization on $\mathsf{T}$, and can be defined recursively using the edges of $\mathsf{T}$. We now provide the proof of Proposition 2.

*Proof of Proposition 2.* We will prove this result by recursion on $n \in \mathbb{N}$. Observe that the initialisation is directly given by Proposition 1. Assume now that the result of Proposition 2 stands for some $n \in \mathbb{N}$. Let $k_n = (n-1) \bmod (K)$, $k_n + 1 = n \bmod (K)$. Denote by $\mathsf{T}_{k_n}$ with edges $\mathsf{E}_{k_n}$, respectively $\mathsf{T}_{k_n+1}$ with edges $\mathsf{E}_{k_n+1}$, the directed version of $\mathsf{T}$ rooted in $i_{k_n}$, respectively in $i_{k_n+1}$. For any vertex $v$ of $\mathsf{T}_{k_n+1}$, we define $p(v)$ as the (unique) parent of $v$ and $c(v)$ as the unique child of $v$ when it exists. Consider the $(n+1)$-th dynamic iterate defined by (a) and (b), *i.e.*, $(\mathbb{P}^{n+1}_{(v,v')})_{(v,v')\in\mathsf{E}_{k_n+1}}$. To prove that this iterate has the properties stated in Proposition 2, we proceed by recursion on the edges of $\mathsf{T}_{k_n+1}$, following the bread-first order in $\mathsf{T}_{k_n+1}$. In this order, the edge $(i_{k_n+1}, c(i_{k_n+1}))$ is the first to be considered. Remark that $c(i_{k_n+1})$ is well defined since $i_{k_n+1}$ is a leaf of $\mathsf{T}$.

Here, we denote $T_{c(i_{k_n+1}),i_{k_n+1}}$ by $T$. By construction, we have $\mathbb{P}^{n+1}_{(i_{k_n+1},c(i_{k_n+1}))} = \mu_{i_{k_n+1}} \otimes (\mathbb{P}^n_{(c(i_{k_n+1}),i_{k_n+1})})^R_{|0}$. By recursion assumption, $\mathbb{P}^n_{(c(i_{k_n+1}),i_{k_n+1})} \in \mathscr{P}(\mathrm{C}([0,T],\mathbb{R}^d))$ since

$(c(i_{k_n+1}), i_{k_n+1}) \in \mathsf{E}_{k_n}$. Then, $\mathbb{P}^{n+1}_{(i_{k_n+1}, c(i_{k_n+1}))}$ is a well defined path measure on $[0, T]$. By definition of the (mIPF) sequence, we have $\mu_{i_{k_n+1}} = \pi^{n+1}_{i_{k_n+1}}$. By recursion assumption, we also have that $\mathrm{Ext}(\mathbb{P}^n_{(c(i_{k_n+1}), i_{k_n+1})}) = \pi^n_{c(i_{k_n+1}), i_{k_n+1}}$. Hence, it comes that $(\mathbb{P}^n_{(c(i_{k_n+1}), i_{k_n+1})})^R_{T|0} = \pi^n_{c(i_{k_n+1})|i_{k_n+1}} = \pi^{n+1}_{c(i_{k_n+1})|i_{k_n+1}}$, where the last equality comes from Lemma 12. Finally, we obtain that $\mathrm{Ext}(\mathbb{P}^{n+1}_{(i_{k_n+1}, c(i_{k_n+1}))}) = \pi^{n+1}_{i_{k_n+1}, c(i_{k_n+1})}$, which proves the initialisation.

Assume now that $\mathbb{P}^{n+1}$ is well defined and has the right properties, up to some edge in $\mathsf{T}_{k_n+1}$. Consider the following edge, denoted by $(v, v') \in \mathsf{E}_{k_n+1}$, in the breadth-first order. By edge recursion, we have that $\mathrm{Ext}(\mathbb{P}^{n+1}_{(p(v),v)}) = \pi^{n+1}_{p(v),v}$, and thus $\mathbb{P}^{n+1}_{(p(v),v),T_{p(v),v}} = \pi^{n+1}_v$. Define the path $\mathsf{P}_n = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. Then, we face two cases.

(i) Either $(v, v') \in \mathsf{E}_{k_n} \backslash \mathsf{P}_n$. Then, we have by (a) that

$$\mathbb{P}^{n+1}_{(v,v')} = \mathbb{P}^{n+1}_{(p(v),v),T_{p(v),v}} \otimes \mathbb{P}^n_{(v,v')|0} = \pi^{n+1}_v \otimes \mathbb{P}^n_{(v,v')|0}$$

In particular, $\mathbb{P}^{n+1}_{(v,v')}$ is a well defined path measure on $[0, T_{v,v'}]$. Since $(v, v') \in \mathsf{E}_{k_n}$, $\mathrm{Ext}(\mathbb{P}^n_{(v,v')}) = \pi^n_{v,v'}$ by recursion assumption. In particular, $\mathbb{P}^n_{(v,v'),T_{v,v'}|0} = \pi^n_{v'|v} = \pi^{n+1}_{v'|v}$ where the last equality comes from Lemma 12. We thus have $\mathrm{Ext}(\mathbb{P}^{n+1}_{(v,v')}) = \pi^{n+1}_{v,v'}$.

(ii) Or $(v', v) \in \mathsf{P}_n$. Then, we have by (b) that

$$\mathbb{P}^{n+1}_{(v,v')} = \mathbb{P}^{n+1}_{(p(v),v),T_{v,v'}} \otimes (\mathbb{P}^n_{(v',v)})^R_{|0} = \pi^{n+1}_v \otimes (\mathbb{P}^n_{(v',v)})^R_{|0}$$

In particular, $\mathbb{P}^{n+1}_{(v,v')}$ is a well defined path measure on $[0, T_{v,v'}]$. Here, $(v', v) \in \mathsf{E}_{k_n}$ and thus, $\mathrm{Ext}(\mathbb{P}^n_{(v',v)}) = \pi^n_{v',v}$ by recursion assumption. In particular, $(\mathbb{P}^n_{(v',v)})^R_{T_{v',v}|0} = \pi^n_{v'|v} = \pi^{n+1}_{v'|v}$ where the last equality comes from Lemma 12. We thus have $\mathrm{Ext}(\mathbb{P}^{n+1}_{(v,v')}) = \pi^{n+1}_{v,v'}$.

This achieves the recursion. □

### D.2 Proofs of Section 4

**Remark on assumption A1.** Although **A**1 *is not needed to establish the result of Proposition 3, Corollary 4 and Proposition 5*, it is however crucial in the proof of convergence of (mIPF) stated in Proposition 6. Nevertheless, we choose to keep **A**1 as an assumption in the statement of every theoretical result presented in Section 4 for sake of clarity.

**Additional definitions.** We define the set $\mathscr{P}_\mathsf{S} = \cap_{i \in \mathsf{S}} \mathscr{P}_i$, where $\mathscr{P}_i = \{\pi \in \mathscr{P}^{(\ell+1)} : \pi_i = \mu_i\}$, *i.e.*, $\mathscr{P}_\mathsf{S}$ is the set of all probability measures $\pi \in \mathscr{P}^{(\ell+1)}$ which verify

$$\int_{(\mathbb{R}^d)^{\ell+1}} f_i(x_i) \mathrm{d}\pi(x_{0:\ell}) = \int_{\mathbb{R}^d} f_i(x_i) \mathrm{d}\mu_i(x_i) ,$$

for any family of bounded measurable functions $\{f_i\}_{i \in \mathsf{S}} \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})^K$. Since $\mathbb{R}^d$ is separable, there exists a dense family of functions $\{f_i^k\}_{k \in \mathbb{N}^*, i \in \mathsf{S}}$, with $f_i^k \in \mathrm{L}^\infty(\mu_i)$ for any $k \in \mathbb{N}^*$ and any $i \in \mathsf{S}$, such that $\pi \in \mathscr{P}_\mathsf{S}$ if and only if

$$\int_{(\mathbb{R}^d)^{\ell+1}} f_i^k(x_i) \mathrm{d}\pi(x_{0:\ell}) = \int_{\mathbb{R}^d} f_i^k(x_i) \mathrm{d}\mu_i(x_i)$$

or equivalently, upon centering $f_i^k$,

$$\int_{(\mathbb{R}^d)^{\ell+1}} f_i^k(x_i) \mathrm{d}\pi(x_{0:\ell}) = 0 .$$

In the rest of the section, we consider such family $\{f_i^k\}_{k \in \mathbb{N}^*, i \in \mathsf{S}}$.

For any $n \in \mathbb{N}^*$, we also define $\mathscr{P}^n_\mathsf{S} = \cap_{i \in \mathsf{S}} \mathscr{P}^n_i$, where $\mathscr{P}^n_i = \{\pi \in \mathscr{P}^{(\ell+1)} : \int_{(\mathbb{R}^d)^{\ell+1}} f_i^k(x_i) \mathrm{d}\pi(x_{0:\ell}) = 0,\ \forall k \in \{1, \ldots, n\}\}$. In particular, we have

$$\mathscr{P}_\mathsf{S} = \cap_{n \in \mathbb{N}^*} \mathscr{P}^n_\mathsf{S} . \tag{6}$$

Finally, (static-mSB) can be rewritten as

$$\pi^\star = \mathrm{argmin}\{\mathrm{KL}(\pi \mid \pi^0) : \pi \in \mathscr{P}_\mathsf{S}\} . \tag{7}$$

**Proof of Proposition 3 and Corollary 4.** In this part of the section, we present an extension of the theoretical results from Nutz (2021) to the multi-marginal setting. We first present two technical results, Lemma 13 and Lemma 14, which are respectively adapted from (Nutz, 2021, Lemma 2.10.) and (Nutz, 2021, Lemma 2.11.).

**Lemma 13.** *Let $\{\tilde{\mu}_j\}_{j\in\mathsf{S}^c}$ be a family of probability measures defined on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We define $\tilde{\pi}^0 = \bigotimes_{i\in\mathsf{S}} \mu_i \bigotimes_{j\in\mathsf{S}^c} \tilde{\mu}_j$. Let $\mathsf{A} \in \bigotimes_{m=0}^{\ell} \mathcal{B}(\mathbb{R}^d)$ such that $\tilde{\pi}^0(\mathsf{A}) = 1$. Then, for $\tilde{\pi}^0$-almost any $x^\star \in \mathsf{A}$, there exists a family of sets $\{\mathsf{X}_m^0\}_{m=0}^{\ell} \subset (\mathbb{R}^d)^{\ell+1}$ such that*

*(a) $\mu_i(\mathsf{X}_i^0) = 1$ for any $i \in \mathsf{S}$, and $\tilde{\mu}_j(\mathsf{X}_j^0) = 1$ for any $j \in \mathsf{S}^c$,*

*(b) $\mathsf{A}^0 = \mathsf{A} \cap (\prod_{m=0}^{\ell} \mathsf{X}_m^0)$ satisfies $x^\star \in \mathsf{A}^0$ and*

$$(x_0^\star, \ldots, x_{m-1}^\star, x_m, x_{m+1}^\star, \ldots, x_\ell^\star) \in \mathsf{A}^0, \forall x \in \mathsf{A}^0, \forall m \in \{0, \ldots, \ell\} \; .$$

*Proof.* Consider such set $\mathsf{A}$. We define for any $m \in \{0, \ldots, \ell\}$ the set

$$\mathsf{X}_m = \{u \in \mathbb{R}^d : \tilde{\pi}_{-m}^0(\mathsf{A}_m^u) = 1\} \; ,$$

where $\mathsf{A}_m^u = \{y \in (\mathbb{R}^d)^\ell : (y_0, \ldots, y_{m-1}, u, y_m, \ldots, y_{\ell-1}) \in \mathsf{A}\}$.

Take $i \in \mathsf{S}$. Assume that $\mu_i(\mathsf{X}_i) < 1$. We recall that $\tilde{\pi}^0 = \tilde{\pi}_{-i}^0 \otimes \mu_i$. Using Fubini's theorem and that $\int_{\mathsf{A}_i^{x_i}} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) < 1$ for any $x_i \notin \mathsf{X}_i$, we have

$$\begin{aligned}
1 = \tilde{\pi}^0(\mathsf{A}) &= \textstyle\int_{\mathsf{A}} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) \otimes \mathrm{d}\mu_i(x_i) \\
&= \textstyle\int_{\mathbb{R}^d} \{\int_{\mathsf{A}_i^{x_i}} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i})\} \mathrm{d}\mu_i(x_i) \\
&= \textstyle\int_{\mathsf{X}_i} \{\int_{\mathsf{A}_i^{x_i}} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i})\} \mathrm{d}\mu_i(x_i) + \int_{\mathsf{X}_i^c} \{\int_{\mathsf{A}_i^{x_i}} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i})\} \mathrm{d}\mu_i(x_i) \\
&< \mu_i(\mathsf{X}_i) + \mu_i(\mathsf{X}_i^c) = 1 \; ,
\end{aligned}$$

which is absurd. Therefore, we obtain $\mu_i(\mathsf{X}_i) = 1$, and similarly, we have $\tilde{\mu}_j(\mathsf{X}_j) = 1$ for any $j \in \mathsf{S}^c$. For any $y \in (\mathbb{R}^d)^\ell$, any $m \in \{0, \ldots, \ell\}$, we define the set

$$\bar{\mathsf{A}}_m^y = \{u \in \mathbb{R}^d : (y_0, \ldots, y_{m-1}, u, y_m, \ldots, y_{\ell-1}) \in \mathsf{A}\} \; .$$

Let $i \in \mathsf{S}$. We have by Fubini's theorem

$$\begin{aligned}
1 = \tilde{\pi}^0(\mathsf{A}) &= \textstyle\int_{\mathsf{A}} \mathrm{d}\mu_i(x_i) \otimes \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) \\
&= \textstyle\int_{(\mathbb{R}^d)^\ell} \{\int_{\bar{\mathsf{A}}_i^{x_{-i}}} \mathrm{d}\mu_i(x_i)\} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) \\
&= \textstyle\int_{\prod_{m=0, m\neq i}^{\ell} \mathsf{X}_i} \{\int_{\bar{\mathsf{A}}_i^{x_{-i}}} \mathrm{d}\mu_i(x_i)\} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) \; ,
\end{aligned}$$

where the last equality comes from the fact that $\mu_i(\mathsf{X}_i) = 1$ for any $i \in \mathsf{S}$, $\tilde{\mu}_j(\mathsf{X}_j) = 1$ for any $j \in \mathsf{S}^c$ and that $\tilde{\pi}^0 = \bigotimes_{i\in\mathsf{S}} \mu_i \bigotimes_{j\in\mathsf{S}^c} \tilde{\mu}_j$. Consequently, there exists a measurable set $\mathsf{A}_{-i} \subset \prod_{m=0, m\neq i}^{\ell} \mathsf{X}_i$ such that the following properties hold: (a) $\mu_i(\bar{\mathsf{A}}_i^y) = 1$ for any $y \in \mathsf{A}_{-i}$, (b) $\tilde{\pi}_{-i}^0(\mathsf{A}_{-i}) = 1$. Similarly, this result holds for any $j \in \mathsf{S}^c$, *i.e.*, there exists a measurable set $\mathsf{A}_{-j} \subset \prod_{m=0, m\neq j}^{\ell} \mathsf{X}_i$ such that the following properties hold: (a) $\tilde{\mu}_j(\bar{\mathsf{A}}_j^y) = 1$ for any $y \in \mathsf{A}_{-j}$, (b) $\tilde{\pi}_{-j}^0(\mathsf{A}_{-j}) = 1$. We consider such sets $\{\mathsf{A}_{-m}\}_{m=0}^{\ell}$ for the rest of the proof and finally define the set

$$\tilde{\mathsf{A}} = \cap_{m=0}^{\ell} \tilde{\mathsf{A}}_m \; ,$$

where $\tilde{\mathsf{A}}_m = \mathsf{A}_{-m} \times \{u \in \bar{\mathsf{A}}_m^y : y \in \mathsf{A}_{-m}\}$. By definition, we have $\tilde{\mathsf{A}} \subset \mathsf{A} \cap \prod_{m=0}^{\ell} \mathsf{X}_m$, using the fact that $\tilde{\mathsf{A}}_m \subset \mathsf{A}$ for any $m \in \{0, \ldots, \ell\}$. In addition, for any $i \in \mathsf{S}$, we get by Fubini's theorem

$$\tilde{\pi}^0(\tilde{\mathsf{A}}_i) = \textstyle\int_{\tilde{\mathsf{A}}_i} \mathrm{d}\mu_i(x_i) \otimes \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) = \int_{\mathsf{A}_{-i}} \{\int_{\bar{\mathsf{A}}_i^{x_{-i}}} \mathrm{d}\mu_i(x_i)\} \mathrm{d}\tilde{\pi}_{-i}^0(x_{-i}) = \tilde{\pi}_{-i}^0(\mathsf{A}_{-i}) = 1 \; ,$$

and similarly, we get $\tilde{\pi}^0(\tilde{\mathsf{A}}_j) = 1$ for any $j \in \mathsf{S}^c$. We can deduce that $\tilde{\pi}^0(\tilde{\mathsf{A}}) = 1$ since $\tilde{\pi}^0(\tilde{\mathsf{A}}^c) \le \sum_{m=0}^{\ell} \tilde{\pi}^0(\tilde{\mathsf{A}}_m^c) = 0$.

Let $x^\star \in \tilde{\mathsf{A}}$. In particular, $x^\star \in \mathsf{A}$. We define the set $\mathsf{A}^0 = \mathsf{A} \cap (\prod_{m=0}^{\ell} \mathsf{X}_m^0)$, where $\mathsf{X}_m^0 = \mathsf{X}_m \cap \bar{\mathsf{A}}_m^{x^\star_{-m}}$ for any $m \in \{0, \dots, \ell\}$. We now establish the result of Lemma 13.

We first prove (a). Let $i \in \mathsf{S}$. Since $x^\star \in \tilde{\mathsf{A}}$, we have $x^\star \in \tilde{\mathsf{A}}_i$ and therefore $x^\star_{-i} \in \mathsf{A}_{-i}$. By definition of $\mathsf{A}_{-i}$, we obtain that $\mu_i(\bar{\mathsf{A}}_i^{x^\star_{-i}}) = 1$ and thus,

$$\mu_i(\{\mathsf{X}_i^0\}^{\mathrm{c}}) \leq \mu_i(\mathsf{X}_i^{\mathrm{c}}) + \mu_i(\{\bar{\mathsf{A}}_i^{x^\star_{-i}}\}^{\mathrm{c}}) = 0,$$

which gives $\mu_i(\mathsf{X}_i^0) = 1$, and similarly, we have $\tilde{\mu}_j(\mathsf{X}_j^0) = 1$ for any $j \in \mathsf{S}^{\mathrm{c}}$.

We now prove (b). Let $m \in \{0, \dots, \ell\}$. Since $x^\star \in \tilde{\mathsf{A}} \subset \mathsf{A}$, we get $x^\star_m \in \bar{\mathsf{A}}_m^{x^\star_{-m}}$. Using that $\tilde{\mathsf{A}} \subset \mathsf{A} \cap_{m=0}^{\ell} \mathsf{X}_m$, we get $x^\star \in \mathsf{A}^0$. Let $x \in \mathsf{A}^0$. We denote $x^m = (x^\star_0, \dots, x^\star_{m-1}, x_m, x^\star_{m+1}, \dots, x^\star_\ell)$. We need to show that $x^m \in \mathsf{A}$ and $x^m \in \prod_{j=1}^{\ell} \mathsf{X}_j^0 = \prod_{j=1}^{\ell} (\mathsf{X}_j \cap \bar{\mathsf{A}}_j^{x^\star_{-m}})$. First, since $x_j^m = x_j$ or $x_j^\star$ for any $j \in \{0, \dots, \ell\}$, and $x \in \mathsf{A}^0$ and $x^\star \in \mathsf{A}^0$, we get that for any $j \in \{0, \dots, \ell\}$, $x_j^m \in \mathsf{X}_j$. Similarly, for any $j \in \{0, \dots, \ell - 1\}$, $x_j^m \in \bar{\mathsf{A}}_j^{x^\star_m}$. Therefore, we get that $x^m \in \prod_{j=1}^{\ell} (\mathsf{X}_j \cap \bar{\mathsf{A}}_j^{x^\star_{-m}})$. Since $x_m \in \mathsf{A}_m^{x^\star_{-m}}$ (because $x \in \prod_{j=1}^{\ell} (\mathsf{X}_j \cap \bar{\mathsf{A}}_j^{x^\star_{-m}})$), we get that $x \in \mathsf{A}$, which concludes the proof. $\square$

**Lemma 14.** *Let $\mathsf{A}^0 \subset (\mathbb{R}^d)^{\ell+1}$. For any $m \in \{0, \dots, \ell\}$, we denote $\mathsf{X}_m^0 = \mathrm{proj}_m(\mathsf{A}^0)$. We make the following assumptions.*

*(a) Assume there exists $x^\star \in \mathsf{A}^0$ such that for any $x \in \mathsf{A}^0$, for any $m \in \{0, \dots, \ell\}$, we have $(x^\star_0, \dots, x^\star_{m-1}, x_m, x^\star_{m+1}, \dots, x^\star_\ell) \in \mathsf{A}^0$.*

*(b) Assume there exists a family of functions $\{\varphi^n_{i_k}\}_{n \in \mathbb{N}^*, k \in \{0, \dots, K-1\}}$ with $\varphi^n_{i_k} : \mathsf{X}^0_{i_k} \to [-\infty, +\infty)$ such that for any $n \in \mathbb{N}^*$ and any $k \in \{0, \dots, K-2\}$, we have $\varphi^n_{i_k}(x^\star_{i_k}) = 0$.*

*(c) Denote $F^n(x) = \sum_{k=0}^{K-1} \varphi^n_{i_k}(x_{i_k})$ for any $x \in \mathsf{A}^0$. Assume that for any $x \in \mathsf{A}^0$, $F(x) = \lim_{n\to\infty} F^n(x)$ exists and is such that $F(x) \in [-\infty, +\infty)$ with $F(x^\star) \in \mathbb{R}$.*

*Then, for any $i \in \mathsf{S}$, for any $x_i \in \mathsf{X}_i^0$, $\varphi_i(x_i) = \lim_{n\to\infty} \varphi_i^n(x_i)$ exists and is such that $\varphi_i(x_i) \in [-\infty, +\infty)$.*

*Proof.* Consider $\mathsf{A}^0 \subset (\mathbb{R}^d)^{\ell+1}$ such that assumptions (a), (b) and (c) hold. Remark that we have $F^n(x^\star) = \varphi^n_{i_{K-1}}(x^\star_{i_{K-1}})$.

Let $x \in \mathsf{A}^0$. We denote $x^m = (x^\star_1, \dots, x^\star_{m-1}, x_m, x^\star_{m+1}, \dots, x^\star_l)$ for any $m \in \{0, \dots, \ell\}$. In particular, we have $x^m \in \mathsf{A}^0$ by assumption (a). Let us define

$$\begin{aligned} \varphi_{i_k}(x_{i_k}) &= F(x^{i_k}) - F(x^\star), \quad \forall k \in \{0, \dots, K-2\}\,, \\ \varphi_{i_{K-1}}(x_{i_{K-1}}) &= F(x^{i_{K-1}})\,. \end{aligned}$$

Using assumption (c), we have $\varphi_i(x_i) \in [-\infty, +\infty)$ for any $i \in \mathsf{S}$. Let $k \in \{0, \dots, K-2\}$. We have by definition of $F^n$,

$$\varphi^n_{i_k}(x_{i_k}) = F^n(x^{i_k}) - \textstyle\sum_{\substack{m=0 \\ m \neq k}}^{K-1} \varphi^n_{i_m}(x^\star_{i_m}) = F^n(x^{i_k}) - F^n(x^\star)\,,$$

where we used assumption (b) in the last equality. Since $x^{i_k} \in \mathsf{A}^0$ and $x^\star \in \mathsf{A}^0$, we have by assumption (c),

$$\lim_{n\to\infty} \varphi^n_{i_k}(x_{i_k}) = F(x^{i_k}) - F(x^\star) = \varphi_{i_k}(x_{i_k})\,.$$

Furthermore, by combining the definition of $F^n$ with assumption (b), we have

$$\lim_{n\to\infty} \varphi^n_{i_{K-1}}(x_{i_{K-1}}) = F(x^{i_{K-1}}) = \varphi_{i_{K-1}}(x_{i_{K-1}}),$$

which concludes the proof. $\square$

In what follows, before proving Proposition 3, we respectively show in Proposition 15 and Proposition 16 how **A**2 and **A**3 can be satisfied in the case where $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$, as in (2), that is

$$\pi^0 = \pi_r^0 \bigotimes_{(v,v')\in\mathsf{E}_r} \pi^0_{v'|v} .$$

**Proposition 15.** *Let $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$. Assume that $\pi_r^0 = \mu_r$ if $r \in \mathsf{S}$ or $\pi_r^0 = \mathrm{N}(m_r, \sigma_r \mathrm{Id})$, with $m_r \in \mathbb{R}^d$ and $\sigma_r > 0$ if $r \in \mathsf{S}^{\mathrm{c}}$. In addition, assume that for any $(v, v') \in \mathsf{E}_r$, $\pi^0_{v'|v}(\cdot|x_v) = \mathrm{N}(x_v, \sigma_{v,v'}\mathrm{Id})$ with $\sigma_{v,v'} > 0$. Finally, assume that for any $i \in \mathsf{S}$, $\int_{\mathbb{R}^d} \|x\|^2 \mathrm{d}\mu_i(x) < +\infty$ and $\mathrm{H}(\mu_i) < +\infty$. Then* **A**2 *is satisfied.*

*Proof.* Let $\pi = \otimes_{i\in\mathsf{S}}\mu_i \otimes_{i\in\mathsf{S}^{\mathrm{c}}} \nu_i$ with $\nu_i$ any Gaussian measure with positive definite covariance matrix. First, we have that

$$\mathrm{KL}(\pi \mid \pi^0) = \mathrm{KL}(\pi_r \mid \pi_r^0) + \textstyle\sum_{(v,v')\in\mathsf{E}_r} \int_{\mathbb{R}^d} \mathrm{KL}(\pi_{v'|v}|\pi^0_{v'|v})\mathrm{d}\pi_v \ .$$

For any $(v, v') \in \mathsf{E}_r$, there exists $C_{v,v'} \geq 0$ such that

$$\begin{aligned}\textstyle\int_{\mathbb{R}^d} \mathrm{KL}(\pi_{v'|v}|\pi^0_{v'|v})\mathrm{d}\pi_v &\leq C_{v,v'} - \mathrm{H}(\pi_{v'}) + \textstyle\int_{\mathbb{R}^d\times\mathbb{R}^d} \|x_v - x_{v'}\|^2/(2\sigma^2_{v,v'})\mathrm{d}\pi_v \otimes \pi_{v'}(x_v, x_{v'}) \\ &\leq C_{v,v'} - \mathrm{H}(\pi_{v'}) + (1/\sigma^2_{v,v'}) \textstyle\int_{\mathbb{R}^d} \|x_v\|^2 \mathrm{d}\pi_v(x_v) + (1/\sigma^2_{v,v'}) \int_{\mathbb{R}^d} \|x_{v'}\|^2 \mathrm{d}\pi_{v'}(x_{v'}) < +\infty.\end{aligned}$$

We conclude the proof upon remarking that $\mathrm{KL}(\pi_r \mid \pi_r^0) < +\infty$. □

**Proposition 16.** *Let $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$. Assume that $\pi_r^0 = \mu_r$ if $r \in \mathsf{S}$ or $\pi_r^0 = \mathrm{N}(m_r, \sigma_r \mathrm{Id})$, with $m_r \in \mathbb{R}^d$ and $\sigma_r > 0$ if $r \in \mathsf{S}^{\mathrm{c}}$. In addition, assume that for any $(v, v') \in \mathsf{E}_r$, $\pi^0_{v'|v}(\cdot|x_v) = \mathrm{N}(x_v, \sigma_{v,v'}\mathrm{Id})$ with $\sigma_{v,v'} > 0$. Finally, assume that for any $i \in \mathsf{S}$, $\mu_i$ admits a positive density w.r.t. the Lebesgue measure. Then* **A**3 *is satisfied.*

*Proof.* We have that $\pi^0$ admits a positive density w.r.t the Lebesgue measure. Letting $\tilde{\pi}^0 = \bigotimes_{i\in\mathsf{S}} \mu_i \bigotimes_{j\in\mathsf{S}^{\mathrm{c}}} \tilde{\mu}_j$ where $\tilde{\mu}_j$ which admits a positive density w.r.t. the Lebesgue measure for any $j \in \mathsf{S}^{\mathrm{c}}$, we get that $\tilde{\pi}^0$ admits a positive density w.r.t. the Lebesgue measure and therefore $\pi^0 \sim \tilde{\pi}^0$, which concludes the proof. □

Using the preliminary results presented above, we are now ready to prove Proposition 3.

*Proof of Proposition 3.* Assume **A**1 and **A**2. Since $\mathscr{P}_{\mathsf{S}}$ is convex and closed in total-variation norm, there exists a probability distribution $\pi^\star$ solution to (7), or equivalently to (static-mSB), by using **A**2 with (Csiszár, 1975, Theorem 2.1.). Moreover, this solution is unique by strict convexity of $\mathrm{KL}(\cdot \mid \pi^0)$.

We now turn to the proof of existence of potentials defining $(\mathrm{d}\pi^\star/\mathrm{d}\pi^0)$, by adapting the arguments of (Nutz, 2021, Section 2.3.). Define $\nu^n = \mathrm{argmin}\{\mathrm{KL}(\pi \mid \pi^0) : \pi \in \mathscr{P}^n_{\mathsf{S}}\}$ for any $n \in \mathbb{N}^*$. Since $\{\mathscr{P}^n_{\mathsf{S}}\}_{n\in\mathbb{N}^*} \subset \mathscr{P}^{(\ell+1)}$ is a decreasing sequence of sets that are convex and closed in total-variation norm such that (6) holds, we get from (Nutz, 2021, Proposition 1.17.) with **A**2 that

$$\lim_{n\to\infty} \|\nu^n - \pi^\star\|_{\mathrm{TV}} = 0 \ ,$$

or equivalently

$$\lim_{n\to\infty} \|(\mathrm{d}\nu^n/\mathrm{d}\pi^0) - (\mathrm{d}\pi^\star/\mathrm{d}\pi^0)\|_{\mathrm{L}^1(\pi^0)} = 0 \ . \tag{8}$$

Following (Nutz, 2021, Example 1.18), there exists a family of bounded measurable functions $\{\varphi_i^n\}_{n\in\mathbb{N}^*,i\in\mathsf{S}}$ with $\varphi_i^n : \mathbb{R}^d \to \mathbb{R}$ such that for any $n \in \mathbb{N}^*$

$$(\mathrm{d}\nu^n/\mathrm{d}\pi^0) = \exp[\textstyle\bigoplus_{i\in\mathsf{S}} \varphi_i^n] \ . \tag{9}$$

We consider such family $\{\varphi_i^n\}_{n\in\mathbb{N}^*,i\in\mathsf{S}}$ for the rest of the proof. By combining (8) and (9), we obtain, up to extraction,

$$(\mathrm{d}\pi^\star/\mathrm{d}\pi^0) = \lim_{n\to\infty} \exp[\textstyle\bigoplus_{i\in\mathsf{S}} \varphi_i^n] \quad \pi^0\text{-a.s.} \ . \tag{10}$$

We now define the following sets

$$\mathsf{A}^\star = \{x \in (\mathbb{R}^d)^{\ell+1} : \lim_{n\to\infty} \bigoplus_{i\in\mathsf{S}} \varphi_i^n(x_i) \in [-\infty, +\infty)\} \ ,$$
$$\mathsf{B}^\star = \{x \in (\mathbb{R}^d)^{\ell+1} : \lim_{n\to\infty} \bigoplus_{i\in\mathsf{S}} \varphi_i^n(x_i) > -\infty\} \subset \mathsf{A}^\star$$

Using (10), we have $\pi^0(\mathsf{A}^\star) = 1$. Using **A**3, it comes $\tilde{\pi}^0(\mathsf{A}^\star) = 1$. Moreover, we also get that $\pi^\star(\mathsf{B}^\star) = 1$ by (10). Thus, it comes $\pi^0(\mathsf{B}^\star) > 0$, and $\tilde{\pi}^0(\mathsf{B}^\star) > 0$ using **A**3.

We then apply Lemma 13 to $\tilde{\pi}^0$ and $\mathsf{A} = \mathsf{A}^\star$. Since $\tilde{\pi}^0(\mathsf{B}^\star) > 0$, it implies that there exists $x^\star \in \mathsf{B}^\star$ and a measurable set $\mathsf{A}^0 \subset \mathsf{B}^\star$ verifying the properties (a) and (b). Following (Nutz, 2021, Corollary 2.12), we may assume without loss of generality in the statement of Lemma 13 that the sets $\mathsf{X}_m^0$ are measurable with $\prod_{m=0}^{\ell} \mathsf{X}_m^0 \subset \mathsf{A}$. In this case, we obtain that $\mu_i(\mathrm{proj}_i(\mathsf{A}^0)) = 1$ for any $i \in \mathsf{S}$.

We now aim at applying Lemma 14 to the set $\mathsf{A}^0$. Remark that $\mathsf{A}^0$ directly satisfies assumption (a). For any $n \in \mathbb{N}^*$, consider the following transformation of the functions $\{\varphi_i^n\}_{i\in\mathsf{S}}$

$$\varphi_{i_k}^n \leftarrow \varphi_{i_k}^n - \varphi_{i_k}^n(x_{i_k}^\star), \quad \forall k \in \{0, \dots, K-2\} \ ,$$
$$\varphi_{i_{K-1}}^n \leftarrow \varphi_{i_{K-1}}^n + \textstyle\sum_{k=0}^{K-2} \varphi_{i_k}^n(x_{i_k}^\star) \ .$$

For any $i \in \mathsf{S}$, we restrict $\varphi_{i_k}^n$ to $\mathsf{X}_{i_k}^0$, so that the family $\{\varphi_i^n\}_{n\in\mathbb{N}^*, i\in\mathsf{S}}$ now verifies assumption (b). Finally, since $\mathsf{A}^0 \subset \mathsf{A}^\star$ and $x^\star \in \mathsf{B}^\star$, we directly obtain assumption (c).

Therefore, Lemma 14 may be applied. It provides us with the family of functions $\{\varphi_i\}_{i\in\mathsf{S}}$ defined by $\varphi_i : \mathsf{X}_i^0 \to [-\infty, +\infty)$ with $\varphi_i = \lim_{n\to\infty} \varphi_i^n$ $\mu_i$-a.s. for any $i \in \mathsf{S}$. Since $\mu_i(\mathrm{proj}_i(\mathsf{A}^0)) = 1$ for any $i \in \mathsf{S}$, we may extend the functions $\varphi_i$ to $\mathbb{R}^d$. In particular, we can find a family of functions $\{\psi_i^\star\}_{i\in\mathsf{S}}$ with $\psi_i^\star : \mathbb{R}^d \to [-\infty, +\infty)$ such that $\psi_i^\star = \varphi_i$ $\mu_i$-a.s. Note that these functions are measurable as limits of measurable functions.

Since $\pi^0 \sim \tilde{\pi}^0$ by **A**3, (10) turns into

$$(\mathrm{d}\pi^\star/\mathrm{d}\pi^0) = \exp[\textstyle\bigoplus_{i\in\mathsf{S}} \psi_i^\star] \quad \pi^0\text{-a.s.} \ . \tag{11}$$

Finally, we show that the functions $\psi_i^\star$ are $\mu_i$-a.s. finite. Let $i \in \mathsf{S}$. Let us define $\mathsf{A}_i = \{x_i \in \mathbb{R}^d : \psi_i^\star(x_i) = -\infty\}$. Using (11), we obtain $(\mathrm{d}\pi^\star/\mathrm{d}\pi^0)(\mathsf{A}_i \times (\mathbb{R}^d)^\ell) = 0$. Since $\pi_i^\star = \mu_i$, we have

$$\mu_i(\mathsf{A}_i) = \pi^\star(\mathsf{A}_i \times (\mathbb{R}^d)^\ell) = \textstyle\int_{\mathsf{A}_i \times (\mathbb{R}^d)^\ell} (\mathrm{d}\pi^\star/\mathrm{d}\pi^0)\mathrm{d}\pi^0 = 0 \ ,$$

which gives the result. □

We now turn to the proof of Corollary 4, which states that the iterates of (mIPF) can be expressed via potentials, in the same manner as the solution $\pi^\star$ to (static-mSB).

*Proof of Corollary 4.* Assume **A**1, **A**2 and **A**3. We prove the result of this corollary by recursion on $n \in \mathbb{N}^*$. First take $n = 1$. In this case, the first iteration of (mIPF) is a multi-marginal SB problem of the form (static-mSB) where $S = \{i_0\}$ with reference measure $\pi^0$. Therefore, using **A**2 and **A**3, we can apply Proposition 3 and obtain existence of $\psi_{i_0}^1 : \mathbb{R}^d \to \mathbb{R}$ such that

$$(\mathrm{d}\pi^1/\mathrm{d}\pi^0) = \exp[\psi_{i_0}^1] \quad \pi^0\text{-a.s} \ .$$

By taking $\psi_{i_k}^0 = 0$ for $k \in \{1, \dots, K-1\}$, we thus obtain the result at step $n = 1$.

Now assume that the result is verified for some $n \in \mathbb{N}^*$, with $k_n = (n-1) \bmod(K)$. We define $k_n + 1 = n \bmod(K)$ and $q_n \in \mathbb{N}$ as the quotient of the Euclidean division of $n$ by $K$. In this case, the $(n+1)$-th iteration of (mIPF) is a multi-marginal SB problem of the form (static-mSB) where $\mathsf{S} = \{i_{k_n+1}\}$ with reference measure $\pi^n$. Using (13), we have that **A**2 is satisfied for this new (static-mSB) problem. **A**1 and **A**3 are satisfied for this problem, given the form of $\pi^n$. Therefore, we can apply Proposition 3 and obtain existence of $\psi_{i_{k_n+1}}^{q_n+1} : \mathbb{R}^d \to \mathbb{R}$ such that

$$(\mathrm{d}\pi^{n+1}/\mathrm{d}\pi^n) = \exp[\psi_{i_{k_n+1}}^{q_n+1}] \quad \pi^n\text{-a.s} \ . \tag{12}$$

By assumption, we have that $\pi^n \ll \pi^0$. Hence, we obtain $\pi^{n+1} \ll \pi^0$ and thus,

$$(\mathrm{d}\pi^{n+1}/\mathrm{d}\pi^0) = (\mathrm{d}\pi^{n+1}/\mathrm{d}\pi^n)(\mathrm{d}\pi^n/\mathrm{d}\pi^0) \quad \pi^0\text{-a.s} \ .$$

By combining (12) with the result of the recursion at step $n$, we directly obtain the result at step $n+1$, which achieves the proof. □

**Proofs of Proposition 5 and Proposition 6.** In this part of the section, we establish the proofs of results related to the convergence of (mIPF), respectively Proposition 5 and Proposition 6, which can be seen as a natural extension of (Ruschendorf, 1995, Proposition 2.1.) and (Ruschendorf, 1995, Theorem 3.1.).

*Proof of Proposition 5.* Under **A**1 and **A**2, we obtain by Proposition 3 existence and uniqueness of a solution to (static-mSB), which we denote by $\pi^\star$. Since $\pi^\star \in \mathscr{P}_{\mathsf{S}}$, using recursively (Csiszár, 1975, Theorem 3.12.), the fact that $\{\pi_{i_k} = \mu_{i_k} \ : \ \pi \in \mathscr{P}^{(|\mathsf{V}|)}\}$ is convex for any $k \in \{0, \ldots, K-1\}$ and (mIPF), we obtain

$$\mathrm{KL}(\pi^\star \mid \pi^0) = \mathrm{KL}(\pi^\star \mid \pi^n) + \textstyle\sum_{i=1}^n \mathrm{KL}(\pi^i \mid \pi^{i-1}) \ . \tag{13}$$

Therefore, we have $\sum_{i=1}^\infty \mathrm{KL}(\pi^i \mid \pi^{i-1}) \leq \mathrm{KL}(\pi^\star \mid \pi^0) < \infty$ and thus,

$$\lim_{i \to +\infty} \mathrm{KL}(\pi^i \mid \pi^{i-1}) = 0 \ . \tag{14}$$

Let $n \in \mathbb{N}^*$ with $n > 2K$, $k \in \{0, \ldots, K-1\}$ and let $q_n \in \mathbb{N}$ be the quotient of the Euclidean division of $n-1$ by $K$. We define $n_k = q_n K + k + 1$ with $(n_k - 1) = k \bmod(K)$ if $n_k \leq n$. Otherwise, we set $n_k = (q_n - 1)K + k + 1$ with $(n_k - 1) = k \bmod(K)$. Note that we always have $|n - n_k| \leq 2K$. In particular, we have $\pi^{n_k}_{i_k} = \mu_{i_k}$ by definition of (mIPF). Therefore, we obtain

$$\begin{aligned}
\|\pi^n_{i_k} - \mu_{i_k}\|_{\mathrm{TV}} &\leq \|\pi^n - \pi^{n_k}\|_{\mathrm{TV}} \\
&\leq \|\pi^n - \pi^{n-1}\|_{\mathrm{TV}} + \ldots + \|\pi^{n_k+1} - \pi^{n_k}\|_{\mathrm{TV}} && \text{(triangle inequality)} \\
&\leq (2\mathrm{KL}(\pi^n \mid \pi^{n-1}))^{1/2} + \ldots + (2\mathrm{KL}(\pi^{n_k+1} \mid \pi^{n_k}))^{1/2} \ , && \text{(Pinsker's inequality)}
\end{aligned}$$

where each term goes to 0 as $n \to +\infty$ in the last inequality by (14), which achieves the proof. □

We now turn to the proof Proposition 6, which requires several preliminary technical results. For the rest of this section, we define, for any $n \in \mathbb{N}$, $q_n$ as the quotient of the Euclidean division of $n-1$ by $K$ (in particular, $q_0 = -1$).

**Schrödinger equations.** Under **A**1, **A**2 and **A**3, we know from Proposition 3 that the unique solution $\pi^\star$ to (static-mSB) can be $\pi^0$-a.s. written as $(\mathrm{d}\pi^\star / \mathrm{d}\pi^0) = \exp[\bigoplus_{i \in \mathsf{S}} \psi_i^\star]$, where $\{\psi_i^\star\}_{i \in \mathsf{S}}$ are measurable potentials, referred to as *Schrödinger potentials*. These functions are determined by the fixed-point *Schrödinger equations*

$$\psi_i^\star(x_i) = \log[r_i(x_i) / \textstyle\int_{(\mathbb{R}^d)^\ell} \exp[\sum_{j \in \mathsf{S} \setminus \{i\}} \psi_j^\star(x_j)] h(x_{0:\ell}) \mathrm{d}\nu_{-i}(x_{-i})] \quad \mu_i\text{-a.s.}, \quad \forall i \in \mathsf{S} \ ,$$

which are obtained by marginalising $\pi^\star$ along its constrained marginals. This family of potentials is not unique. Indeed, for any family of real numbers $\{\lambda_{i_k}\}_{k \in \{0, \ldots, K-2\}}$, we have

$$(\mathrm{d}\pi^\star / \mathrm{d}\pi^0) = \exp[\textstyle\bigoplus_{i \in \mathsf{S}} \tilde{\psi}_i] \ ,$$

where $\tilde{\psi}_{i_k} = \psi^\star_{i_k} + \tilde{\lambda}_{i_k}$ for any $k \in \{0, \ldots, K-1\}$ with $\tilde{\lambda}_{i_k} = \lambda_{i_k}$ if $k \in \{0, \ldots, K-2\}$ and $\tilde{\lambda}_{i_{K-1}} = -\sum_{i=0}^{K-2} \lambda_{i_k}$.

**Remark on the initialisation of (mIPF).** Consider a probability measure $\bar{\pi}^0 \in \mathscr{P}^{(\ell+1)}$ of the form

$$(\mathrm{d}\bar{\pi}^0 / \mathrm{d}\pi^0) = \exp[\textstyle\bigoplus_{i \in \mathsf{S}} \psi_i^0] \ , \tag{15}$$

where $\{\psi_i^0\}_{i \in \mathsf{S}}$ is a family of measurable potentials with $\psi_i^0 : \mathbb{R}^d \to \mathbb{R}$ such that $\left|\int_{\mathbb{R}^d} \psi_i^0 \mathrm{d}\mu_i\right| < \infty$ for any $i \in \mathsf{S}$. Then, for any $\pi \in \mathscr{P}_{\mathsf{S}}$, we have

$$\mathrm{KL}(\pi \mid \pi^0) = \mathrm{KL}(\pi \mid \bar{\pi}^0) + \textstyle\int_{(\mathbb{R}^d)^K} \bigoplus_{i \in \mathsf{S}} \psi_i^0 \mathrm{d}\pi = \mathrm{KL}(\pi \mid \bar{\pi}^0) + \sum_{i \in \mathsf{S}} \int_{\mathbb{R}^d} \psi_i^0 \mathrm{d}\mu_i \ .$$

Hence, (static-mSB) is equivalent to the multi-marginal SB problem

$$\mathrm{argmin}\{\mathrm{KL}(\pi | \bar{\pi}^0) \ : \ \pi \in \mathscr{P}^{(\ell+1)}, \ \pi_i = \mu_i \ , \forall i \in \mathsf{S}\} \ .$$

We refer to (Peyré et al., 2019, Proposition 4.2) for the EOT counterpart of this result. This means that the solutions of the multi-marginal Schrödinger Bridge problem are invariant by multiplication of the reference measure by potentials on the *fixed* marginals. Consequently, the initialisation of the (mIPF) sequence may be chosen as $\bar{\pi}^0$ instead of $\pi^0$.

**For sake of clarity, we now refer to the reference probability measure of (static-mSB) as $\bar{\pi}$ or $\pi^{-1}$ and to the initialisation of the (mIPF) iterates as $\pi^0$.**

**Solving (mIPF) with potentials.** To prove the convergence of the (mIPF) iterates to the solution $\pi^\star$ given by Proposition 3, we first rewrite these iterates with potentials, following the form of $\pi^\star$.

To do so, we recursively define the sequence of potentials $\{\psi_i^n\}_{n\in\mathbb{N},i\in\mathsf{S}}$ by

$$\begin{aligned}&\psi_{i_0}^0 = \ldots = \psi_{i_{K-2}}^0 = 0 \;, \\ &\psi_{i_{K-1}}^0(x_{i_{K-1}}) = \log(r_{i_{K-1}}(x_{i_{K-1}})/\textstyle\int_{(\mathbb{R}^d)^\ell} h(x_{0:\ell})\mathrm{d}\nu_{-i_{K-1}}(x_{-i_{K-1}})) \;,\end{aligned} \tag{16}$$

and for any $n \in \mathbb{N}^*$ and $k \in \{0,\ldots,K-1\}$

$$\begin{aligned}\psi_{i_k}^{q_n+1}(x_{i_k}) = \log[r_{i_k}(x_{i_k})/\textstyle\int_{(\mathbb{R}^d)^\ell} \exp[\bigoplus_{\ell=0}^k \psi_{i_\ell}^{q_n+1}(x_{i_\ell}) \bigoplus_{m=k+1}^{K-1} \psi_{i_m}^{q_n}(x_{i_m})] \\ \times h(x_{0:\ell})\mathrm{d}\nu_{-i_k}(x_{-i_k})] \;,\end{aligned} \tag{17}$$

recalling that $q_n$ is the quotient of the Euclidean division of $n-1$ by $K$.

We now define the sequence of probability measures $\{\pi^n\}_{n\in\mathbb{N}}$ by

$$\mathrm{d}\pi^n/\mathrm{d}\bar{\pi} = \exp[\textstyle\bigoplus_{\ell=0}^{k_n} \psi_{i_\ell}^{q_n+1} \bigoplus_{m=k_n+1}^{K-1} \psi_{i_m}^{q_n}],\; k_n = (n-1)\bmod(K),\; n = q_n K + k_n + 1. \tag{18}$$

In particular, we have $(\mathrm{d}\pi^0/\mathrm{d}\bar{\pi}) = \exp[\oplus_{\ell=0}^{K-1}\psi_{i_\ell}^0] = \exp[\psi_{i_{K-1}}^0]$, and thus $\int_{\mathbb{R}^d} \psi_{i_{K-1}}^0 \mathrm{d}\mu_{i_{K-1}} = \mathrm{KL}(\mu_{i_{K-1}} \mid \bar{\pi}_{i_{K-1}})$. Consequently, $\pi^0$ can be chosen as the initialisation of (mIPF), following the previous remark, if we assume that $\mathrm{KL}(\mu_{i_{K-1}} \mid \bar{\pi}_{i_{K-1}}) < \infty$. In (TreeSB) with $r = i_{K-1}$, the latter assumption is directly verified since we choose $\bar{\pi}_{i_{K-1}} = \mu_{i_{K-1}}$.

Let $n \in \mathbb{N}$, with $k_n = (n-1)\bmod(K)$, $k_n + 1 = n \bmod(K)$. Using (16) and (17), we get that $\pi_{i_{k_n}}^n = \mu_{i_{k_n}}$. Moreover, we have

$$\mathrm{d}\pi^n/\mathrm{d}\pi^{n-1} = \exp[\psi_{i_{k_n}}^{q_n+1} - \psi_{i_{k_n}}^{q_n}] \;, \tag{19}$$

with the convention that $\psi_{i_{K-1}}^{-1} = 0$. In particular, we obtain that $\pi_{|i_{k_n+1}}^{n+1} = \pi_{|i_{k_n+1}}^n$.

In conclusion, the sequence $\{\pi^n\}_{n\in\mathbb{N}}$ defined in (18) verifies $\pi^{n+1} = \mu_{i_{k_n+1}}\pi_{|i_{k_n+1}}^n$ for any $n \in \mathbb{N}$. By decomposition property of the Kullback-Leibler divergence, this sequence solves (mIPF) with initialisation $\pi^0$. *We consider such iterates in the following.*

Since $\pi_{i_{k_n}}^n = \mu_{i_{k_n}}$, we have that

$$\mathrm{KL}(\pi^n \mid \pi^{n-1}) = \textstyle\int_{\mathbb{R}^d} (\psi_{i_{k_n}}^{q_n+1} - \psi_{i_{k_n}}^{q_n})\mathrm{d}\mu_{i_{k_n}} \;. \tag{20}$$

Before proving a multi-marginal counterpart to (Ruschendorf, 1995, Lemma 4.1), we state and prove the following result.

**Proposition 17.** *Let $\pi_0, \pi_1$ two probability measures on $\mathbb{R}^d$ such that $\pi_0 \ll \pi_1$. Then, denoting $f = \mathrm{d}\pi_0/\mathrm{d}\pi_1$, the following assertions are equivalent:*

*(a)* $\mathrm{KL}(\pi_0 \mid \pi_1) < +\infty$

*(b)* $\int_{\mathbb{R}^d} |\log(f)(x)|\mathrm{d}\pi_0(x) < +\infty$

*(c)* $\int_{\mathbb{R}^d} \log(f)(x)\mathbb{1}_{f(x)>1}\mathrm{d}\pi_0(x) < +\infty$

*If one of these conditions is satisfied then $\int_{\mathbb{R}^d} |\log(f)(x)|\, \mathrm{d}\pi_0 \leq \mathrm{KL}(\pi_0 \mid \pi_1) + 2/\mathrm{e}$.*

*Proof.* First, note that

$$\textstyle\int_{\mathbb{R}^d} |\log(f)(x)|\mathbb{1}_{f<1}\mathrm{d}\pi_0(x) \leq \int_{\mathbb{R}^d} |\log(f)(x)f(x)|\mathbb{1}_{f<1}\mathrm{d}\pi_1(x) \leq 1/\mathrm{e} \;, \tag{21}$$

where we have used that for any $u \in [0,1]$, $|u\log(u)| \leq 1/\mathrm{e}$. We have that (b) implies (c). Using the previous result we have that (c) implies (b). Hence (c) and (b) are equivalent. In addition, it is clear that (b) implies (a). Finally (this is more of a convention), we have that $\mathrm{KL}(\pi_0 \mid \pi_1) = \int_{\mathbb{R}^d} \log(f)(x)\mathbb{1}_{f(x)>1}\mathrm{d}\pi_0(x) + \int_{\mathbb{R}^d} \log(f)(x)\mathbb{1}_{f(x)<1}\mathrm{d}\pi_0(x) < +\infty$. Using (21) this implies (c). Finally, we have

$$\begin{aligned}\textstyle\int_{\mathbb{R}^d} |\log(f)(x)|\,\mathrm{d}\pi_0(x) &= \textstyle\int_{\mathbb{R}^d} \log(f)(x)\mathrm{d}\pi_0(x) - 2\int_{\mathbb{R}^d} \log(f)(x)\mathbb{1}_{f(x)<1}\mathrm{d}\pi_0(x) \\ &\leq \mathrm{KL}(\pi_0 \mid \pi_1) + 2/\mathrm{e},\end{aligned}$$

which concludes the proof. □

We begin with the following lemma which controls the integral of the potentials uniformly w.r.t. $n \in \mathbb{N}$. It can be seen as the *multi-marginal* counterpart of (Ruschendorf, 1995, Lemma 4.1).

**Lemma 18.** *Assume* **A**4*. There exist* $\{c_i\}_{i\in\mathsf{S}} \in (0,+\infty)^K$ *such that for any function* $f : (\mathbb{R}^d)^{\ell+1} \to \mathbb{R}$ *of the form* $f = \bigoplus_{i\in\mathsf{S}} f_i$*, we have*

$$c_i \|f\|_{\mathrm{L}^1(\pi^\star)} \geq \|f_i\|_{\mathrm{L}^1(\mu_i)} , \quad \forall i \in \mathsf{S} . \tag{22}$$

*For any* $n \in \mathbb{N}^\star$*, we have*

*(a)* $\sum_{i\in\mathsf{S}} \int_{\mathbb{R}^d} \psi_i^n \mathrm{d}\mu_i \leq \mathrm{KL}(\pi^\star \mid \bar{\pi}) < \infty$,

*(b)* $\int_{(\mathbb{R}^d)^{\ell+1}} (\bigoplus_{i\in\mathsf{S}} \psi_i^\star - \bigoplus_{i\in\mathsf{S}} \psi_i^n) \mathrm{d}\pi^\star \leq \mathrm{KL}(\pi^\star \mid \bar{\pi}) < \infty$,

*(c)* $\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} |\psi_i^n| \mathrm{d}\mu_i < \infty, \ \forall i \in \mathsf{S}$.

*Proof.* First, we have that (22) is a direct consequence of (Kober, 1940, Theorem 1) and **A**4. Let us now prove (a). Using (20), we have

$$\begin{aligned}
\textstyle\sum_{m=0}^{Kn} \mathrm{KL}(\pi^m \mid \pi^{m-1}) &= \textstyle\sum_{\ell=0}^{n-1} \sum_{k=0}^{K-1} \mathrm{KL}(\pi^{\ell K+k+1} \mid \pi^{\ell K+k}) + \mathrm{KL}(\pi^0 \mid \pi^{-1}) \\
&= \textstyle\sum_{\ell=0}^{n-1} \sum_{i\in\mathsf{S}} \int_{\mathbb{R}^d} (\psi_i^{\ell+1} - \psi_i^\ell) \mathrm{d}\mu_i + \int_{\mathbb{R}^d} (\psi_{i_{K-1}}^0 - \psi_{i_{K-1}}^{-1}) \mathrm{d}\mu_{i_{K-1}} \\
&= \textstyle\sum_{i\in\mathsf{S}} \sum_{\ell=0}^{n-1} \int_{\mathbb{R}^d} (\psi_i^{\ell+1} - \psi_i^\ell) \mathrm{d}\mu_i + \int_{\mathbb{R}^d} (\psi_{i_{K-1}}^0 - \psi_{i_{K-1}}^{-1}) \mathrm{d}\mu_{i_{K-1}} \\
&= \textstyle\sum_{i\in\mathsf{S}} \int_{\mathbb{R}^d} (\psi_i^n - \psi_i^0) \mathrm{d}\mu_i + \int_{\mathbb{R}^d} (\psi_{i_{K-1}}^0 - \psi_{i_{K-1}}^{-1}) \mathrm{d}\mu_{i_{K-1}} \\
&= \textstyle\sum_{i\in\mathsf{S}} \int_{\mathbb{R}^d} \psi_i^n \mathrm{d}\mu_i \leq \mathrm{KL}(\pi^\star \mid \bar{\pi}). \ ,
\end{aligned}$$

where the last inequality follows the proof of Proposition 5.

Since the first term in the inequality of (b) is equal to $\mathrm{KL}(\pi^\star \mid \pi^{nK})$, we obtain (b) using that $\mathrm{KL}(\pi^\star \mid \pi^{nK}) \leq \mathrm{KL}(\pi^\star \mid \bar{\pi})$ following the proof of Proposition 5.

Let us now prove (c). Since $\mathrm{KL}(\pi^\star \mid \bar{\pi}) < \infty$, using Proposition 17, we have that $\bigoplus_{i\in\mathsf{S}} \psi_i^\star \in \mathrm{L}^1(\pi^\star)$. From (b) and Proposition 17, we also get that $\bigoplus_{i\in\mathsf{S}} (\psi_i^\star - \psi_i^n) \in \mathrm{L}^1(\pi^\star)$, and thus $\int_{(\mathbb{R}^d)^{\ell+1}} \left|\bigoplus_{i\in\mathsf{S}} (\psi_i^\star - \psi_i^n)\right| \mathrm{d}\pi^\star \leq C_0$ with $C_0 > 0$. Therefore, we have

$$\textstyle\int_{(\mathbb{R}^d)^{\ell+1}} \left|\bigoplus_{i\in\mathsf{S}} \psi_i^n\right| \mathrm{d}\pi^\star \leq \int_{(\mathbb{R}^d)^{\ell+1}} \left|\bigoplus_{i\in\mathsf{S}} \psi_i^\star\right| \mathrm{d}\pi^\star + \int_{(\mathbb{R}^d)^{\ell+1}} \left|\bigoplus_{i\in\mathsf{S}} (\psi_i^\star - \psi_i^n)\right| \mathrm{d}\pi^\star \leq 2C_0 .$$

Using (22), we conclude with **A**4 that for any $i \in \mathsf{S}$, we have

$$\textstyle\int_{\mathbb{R}^d} |\psi_i^n| \, \mathrm{d}\mu_i \leq 2c_i C_0 ,$$

which concludes the proof of (c). □

The next lemma gives an explicit expression for $\mathrm{KL}(\pi^n \mid \bar{\pi})$. It can be seen as the *multi-marginal* counterpart of (Ruschendorf, 1995, Lemma 4.2).

**Lemma 19.** *For any* $n \in \mathbb{N}$*, with* $k_n = (n-1) \bmod(K)$*, we have*

$$\begin{aligned}
\mathrm{KL}(\pi^n \mid \bar{\pi}) = \textstyle\int_{\mathbb{R}^d} \psi_{i_{k_n}}^{q_n+1} \mathrm{d}\mu_{i_{k_n}} + \sum_{\ell=0}^{k_n-1} \int_{\mathbb{R}^d} \psi_{i_\ell}^{q_n+1} \exp[\psi_{i_\ell}^{q_n+1} - \psi_{i_\ell}^{q_n+2}] \mathrm{d}\mu_{i_\ell} \\
+ \textstyle\sum_{m=k_n+1}^{K-1} \int_{\mathbb{R}^d} \psi_{i_m}^{q_n} \exp[\psi_{i_m}^{q_n} - \psi_{i_m}^{q_n+1}] \mathrm{d}\mu_{i_m} .
\end{aligned}$$

*Proof.* Let $n \in \mathbb{N}$, with $k_n = (n-1) \bmod(K)$. Using (18), we have

$$\mathrm{KL}(\pi^n \mid \bar{\pi}) = \textstyle\int_{\mathbb{R}^d} \psi_{i_{k_n}}^{q_n+1} \mathrm{d}\mu_{i_{k_n}} + \sum_{\ell=0}^{k_n-1} \int_{\mathbb{R}^d} \psi_{i_\ell}^{q_n+1} \mathrm{d}\pi_{i_\ell}^n + \sum_{m=k_n+1}^{K-1} \int_{\mathbb{R}^d} \psi_{i_m}^{q_n} \mathrm{d}\pi_{i_m}^n . \tag{23}$$

Consider $m \in \{k_n+1, \ldots, K-1\}$. Let $m_n$ be the closest integer to $n$ such that $m_n > n$ and $m = (m_n - 1) \bmod(K)$. By (19), we have

$$\mathrm{d}\pi^n = \exp[\textstyle\bigoplus_{j=k_n+1}^m \psi_{i_j}^{q_n} - \psi_{i_j}^{q_n+1}] \mathrm{d}\pi^{m_n}.$$

Using (19) recursively, we obtain

$$\mathrm{d}\pi_{i_m}^n = \exp[\psi_{i_m}^{q_n} - \psi_{i_m}^{q_n+1}] \mathrm{d}\pi_{i_m}^{m_n} , \tag{24}$$

where we recall that $\pi^{m_n}_{i_m} = \mu_{i_m}$.

Consider now $\ell \in \{0, \dots, k_n - 1\}$. Let $\ell_n$ be the closest integer to $n$ such that $\ell_n > n$ and $\ell = (\ell_n - 1) \bmod (K)$. By (19), we have

$$\mathrm{d}\pi^n = \exp[\textstyle\bigoplus_{j=k_n+1}^{K-1}\{\psi^{q_n}_{i_j} - \psi^{q_n+1}_{i_j}\} \bigoplus_{j'=0}^{\ell}\{\psi^{q_n+1}_{i_{j'}} - \psi^{q_n+2}_{i_{j'}}\}]\mathrm{d}\pi^{\ell_n},$$

and using (19) recursively, we obtain

$$\mathrm{d}\pi^n_{i_\ell} = \exp[\psi^{q_n+1}_{i_\ell} - \psi^{q_n+2}_{i_\ell}]\mathrm{d}\pi^{\ell_n}_{i_\ell}, \tag{25}$$

where we recall that $\pi^{\ell_n}_{i_\ell} = \mu_{i_\ell}$. We conclude the proof upon combining (23), (24) and (25). □

We are now ready to prove a *uniform integrability* result which is the multi-marginal counterpart of (Ruschendorf, 1995, Lemma 4.4). Before stating Lemma 21, we prove the following well-known lemma. We recall that a sequence $(\Psi_n)_{n\in\mathbb{N}}$ such that for any $n \in \mathbb{N}$, $\Psi_n \in \mathrm{L}^1(\mu)$, is *uniformly integrable* w.r.t. $\mu$ if (i) $\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} |\Psi_n| \,\mathrm{d}\mu < +\infty$ and (ii) for any $\varepsilon > 0$, there exists $K > 0$ such that for any $n \in \mathbb{N}$, $\int_{\overline{\mathrm{B}}(0,K)^c} |\Psi_n| \,\mathrm{d}\mu \leq \varepsilon$.

**Lemma 20.** *Let $f : \mathbb{R} \to \mathbb{R}$, convex and non-decreasing on $[A, +\infty)$ with $A > 0$ and $\lim_{x\to+\infty} f(x)/x = +\infty$. Assume that $\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} f(|\Psi_n|)\mathrm{d}\mu < +\infty$. Then, $(\Psi_n)_{n\in\mathbb{N}}$ is uniformly integrable w.r.t. $\mu$.*

*Proof.* Since $f$ is convex, using Jensen's inequality, we get that $\sup_{n\in\mathbb{N}} f(\int_{\mathbb{R}^d} |\Psi_n| \,\mathrm{d}\mu) < +\infty$ and since $\lim_{x\to+\infty} f(x)/x = +\infty$ we have $\sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} |\Psi_n| \,\mathrm{d}\mu < +\infty$. Let $\varepsilon > 0$, there exists $K > 0$ such that for any $x > K$, $x \leq \varepsilon f(x)/B$ with $B = \sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} f(|\Psi_n|)\mathrm{d}\mu < +\infty$. Therefore, we have for any $n \in \mathbb{N}$

$$\textstyle\int_{\overline{\mathrm{B}}(0,K)^c} |\Psi_n| \,\mathrm{d}\mu \leq (\varepsilon/B) \int_{\overline{\mathrm{B}}(0,K)^c} f(|\Psi_n|)\mathrm{d}\mu \leq \varepsilon,$$

which concludes the proof. □

**Lemma 21.** *Assume* **A**4 *and* **A**5*. Then, $\{\exp[\bigoplus_{i\in\mathsf{S}} \psi^n_i]\}_{n\in\mathbb{N}}$ is uniformly integrable w.r.t. $\bar{\pi}$.*

*Proof.* It is enough to show that the sequence $\{f(\exp[\bigoplus_{i\in\mathsf{S}} \psi^n_i)]\}_{n\in\mathbb{N}}$ is bounded in $\mathrm{L}^1(\bar{\pi})$, where $f : u \mapsto u\log(u)$ is continuous, convex and such that $\lim_{u\to\infty} f(u)/u = +\infty$, see Lemma 20. Let $n \in \mathbb{N}$. We have

$$\begin{aligned}
&\textstyle\int_{(\mathbb{R}^d)^{\ell+1}} f(\exp[\bigoplus_{i\in\mathsf{S}} \psi^n_i)]\mathrm{d}\bar{\pi} = \mathrm{KL}(\pi^{nK} \mid \bar{\pi}) \\
&\quad = \textstyle\int_{\mathbb{R}^d} \psi^n_{i_{K-1}} \mathrm{d}\mu_{i_{K-1}} + \sum_{k=0}^{K-2} \int_{\mathbb{R}^d} \psi^n_{i_k} \exp[\psi^n_{i_k} - \psi^{n+1}_{i_k}]\mathrm{d}\mu_{i_k} && \text{(Lemma 19)} \\
&\quad = \textstyle\sum_{k=0}^{K-1} \int_{\mathbb{R}^d} \psi^n_{i_k} \mathrm{d}\mu_{i_k} + \sum_{k=0}^{K-2} \int_{\mathbb{R}^d} \psi^n_{i_k} \{\exp[\psi^n_{i_k} - \psi^{n+1}_{i_k}] - 1\}\mathrm{d}\mu_{i_k} \\
&\quad \leq \textstyle\mathrm{KL}(\pi^\star \mid \bar{\pi}) + (\bar{c} + 1) \sum_{k=0}^{K-2} \int_{\mathbb{R}^d} \psi^n_{i_k} \mathrm{d}\mu_{i_k} && \text{(Lemma 18-(a), \textbf{A}5)} \\
&\quad \leq \textstyle\mathrm{KL}(\pi^\star \mid \bar{\pi}) + (\bar{c} + 1) \sum_{k=0}^{K-2} \sup_{n\in\mathbb{N}} \int_{\mathbb{R}^d} \left|\psi^n_{i_k}\right| \mathrm{d}\mu_{i_k} < \infty \,. && \text{(Lemma 18-(c))}
\end{aligned}$$

□

With the preliminary results stated above, we are now ready to prove Proposition 6.

*Proof of Proposition 6.* Using **A**4 and **A**5, we have, by Lemma 21, uniform integrability of $\{\exp[\bigoplus_{i\in\mathsf{S}} \psi^n_i]\}_{n\in\mathbb{N}}$ in $\mathrm{L}^1(\bar{\pi})$. Therefore, the sequence $\{\pi^{nK}\}_{n\in\mathbb{N}}$ is relatively compact with respect to the weak topology of $\sigma(\mathrm{L}^1(\bar{\pi}), \mathrm{L}^\infty(\bar{\pi}))$, denoted as the $\tau$-topology. We recall that $\lim_{n\to\infty} \mathrm{KL}(\pi^{nK+1} \mid \pi^{nK}) = 0$. This implies that $\{\pi^{nK+1}\}_{n\in\mathbb{N}}$ is also relatively $\tau$-compact. By trivial recursion, we obtain that the sequences $\{\pi^{nK+k}\}_{n\in\mathbb{N}}$, where $k \in \{2, \dots, K-1\}$ are also relatively $\tau$-compact. Therefore, $\{\pi^n\}_{n\in\mathbb{N}}$ is relatively $\tau$-compact and $\tau$-sequentially compact.

We consider an increasing function $\Phi : \mathbb{N} \to \mathbb{N}$ such that $\{\pi^m\}_{m\in\Phi(\mathbb{N})}$ is a $\tau$-convergent subsequence, and we denote by $\tilde{\pi}$ its limit for this topology. In particular, $\tilde{\pi} \in \mathscr{P}_{\mathsf{S}}$ by Proposition 5. We assume without loss of generality that $\Phi(\mathbb{N}) \subset K\mathbb{N}$.

Using the lower semi-continuity of the Kullback-Leibler divergence (Dupuis & Ellis, 2011, Lemma 1.4.3), we get

$$\mathrm{KL}(\tilde{\pi} \mid \bar{\pi}) \le \liminf \mathrm{KL}(\pi^m \mid \bar{\pi}) \le \limsup \mathrm{KL}(\pi^m \mid \bar{\pi}) \, .$$

Consider $k \in \{0, \dots, K-2\}$. By (19), we have

$$\frac{\mathrm{d}\mu_{i_k}}{\mathrm{d}\pi^{nK+k}_{i_k}} = \frac{\mathrm{d}\pi^{nK+k+1}_{i_k}}{\mathrm{d}\pi^{nK+k}_{i_k}} = \frac{\mathrm{d}\pi^{nK+k+1}}{\mathrm{d}\pi^{nK+k}} = \exp[\psi^{n+1}_{i_k} - \psi^n_{i_k}] \, ,$$

and thus,

$$\|\mu_{i_k} - \pi^{nK+k}_{i_k}\|_{\mathrm{TV}} = (1/2)\int_{\mathbb{R}^d} \left|\mathrm{d}\pi^{nK+k}_{i_k}/\mathrm{d}\mu_{i_k} - 1\right| \mathrm{d}\mu_{i_k} = (1/2)\int_{\mathbb{R}^d} \left|\exp[\psi^n_{i_k} - \psi^{n+1}_{i_k}] - 1\right| \mathrm{d}\mu_{i_k} \, .$$

With Proposition 5, we obtain that $\{\exp[\psi^n_{i_k} - \psi^{n+1}_{i_k}]\}_{n\in\mathbb{N}}$ converges to 1 in $\mathrm{L}^1(\mu_{i_k})$. In addition using the uniform integrability of $\{\psi^n_{i_k}\}_{n\in\mathbb{N}}$ and **A**5, we get

$$\limsup_{n\to+\infty} \int_{\mathbb{R}^d} \psi^n_{i_k} \exp[\psi^n_{i_k} - \psi^{n+1}_{i_k}]\mathrm{d}\mu_{i_k} = \limsup_{n\to+\infty} \int_{\mathbb{R}^d} \psi^n_{i_k}\mathrm{d}\mu_{i_k} \, .$$

We denote $m = K\ell$. Since $\mathrm{KL}(\pi^m \mid \bar{\pi}) = \int_{\mathbb{R}^d} \psi^\ell_{i_{K-1}}\mathrm{d}\mu_{i_{K-1}} + \sum_{k=0}^{K-2}\int_{\mathbb{R}^d} \psi^\ell_{i_k}\exp[\psi^\ell_{i_k} - \psi^{\ell+1}_{i_k}]\mathrm{d}\mu_{i_k}$ by Lemma 19, we finally have

$$\mathrm{KL}(\tilde{\pi} \mid \bar{\pi}) \le \limsup\{\textstyle\sum_{k=0}^{K-1}\int_{\mathbb{R}^d} \psi^\ell_{i_k}\mathrm{d}\mu_{i_k}\} \le \mathrm{KL}(\pi^\star \mid \bar{\pi})$$

where the last inequality comes from Lemma 18.

Since $\tilde{\pi}_i = \mu_i$ for any $i \in \mathsf{S}$, using Proposition 5, we have $\tilde{\pi} = \pi^\star$ by uniqueness of $\pi^\star$. Hence, $\pi^\star$ is the only limit point of $\{\pi^n\}_{n\in\mathbb{N}}$ in the $\tau$-topology. In particular, $\mathrm{KL}(\pi^n \mid \bar{\pi}) \to \mathrm{KL}(\pi^\star \mid \bar{\pi})$. Since $\mathscr{P}_{\mathsf{S}}$ is convex, this last result implies $\|\pi^\star - \pi^n\|_{\mathrm{TV}} \to 0$, see the proof of Theorem 2.1 in Csiszár (1975). □

We finish this section by highlighting that **A**5 is stronger than (Ruschendorf, 1995, B1). A natural extension of the latter assumption would consist of having a guarantee on the $(K-1)$ first potentials given by (17), as presented below.

**A6.** *There exist $0 < \underline{c} < \bar{c}$ such that for any $k \in \{0, \dots, K-2\}$, we have $\underline{c} \le \exp(-\psi^1_{i_k}) \le \bar{c}$.*

Under **A**6, (Ruschendorf, 1995, Lemma 4.3) can be adapted as written below.

**Lemma 22.** *Assume **A**6. Then, for any $n \in \mathbb{N}^*$*

*(a) for any $k \in \{0, \dots, K-2\}$, there exists $\alpha_{n,k} \in \mathbb{N}$ such that*

$$\underline{c} \cdot (\underline{c}/\bar{c})^{\alpha_{n,k}(K-2)} \le \exp[\psi^{n-1}_{i_k} - \psi^n_{i_k}] \le \bar{c} \cdot (\bar{c}/\underline{c})^{\alpha_{n,k}(K-2)}$$

*(b) there exists $\alpha_{n,K-1} \in \mathbb{N}$ such that*

$$1/\bar{c}^{K-1} \cdot (\underline{c}/\bar{c})^{\alpha_{n,K-1}(K-2)} \le \exp[\psi^{n-1}_{i_{K-1}} - \psi^n_{i_{K-1}}] \le 1/\underline{c}^{K-1} \cdot (\bar{c}/\underline{c})^{\alpha_{n,K-1}(K-2)}$$

*where $\{\alpha_{n,k}\}_{n\in\mathbb{N}^*, k\in\{0,\dots,K-1\}}$ is a strictly increasing sequence that can be explicitly defined.*

*Proof.* We prove the result by recursion on $n \in \mathbb{N}^*$.

Take $n = 1$. Let $k \in \{0, \dots, K-2\}$. We define $\alpha_{1,k} = 0$ and directly obtain (a) by **A**5 since $\psi^0_{i_k} = 0$. Let us prove (b). We have by (17)

$$\begin{aligned}\exp[\psi^0_{i_{K-1}} - \psi^1_{i_{K-1}}] &= \frac{\int_{(\mathbb{R}^d)^\ell} \exp[\bigoplus_{k=0}^{K-2} \psi^1_{i_k}]h\mathrm{d}\nu_{-i_{K-1}}}{\int_{(\mathbb{R}^d)^\ell} \exp[\bigoplus_{k=0}^{K-2} \psi^0_{i_k}]h\mathrm{d}\nu_{-i_{K-1}}} \\ &= \frac{\int_{(\mathbb{R}^d)^\ell} \exp[\bigoplus_{k=0}^{K-2}\{\psi^1_{i_k} - \psi^0_{i_k}\} + \bigoplus_{k=0}^{K-2} \psi^0_{i_k}]h\mathrm{d}\nu_{-i_{K-1}}}{\int_{(\mathbb{R}^d)^\ell} \exp[\bigoplus_{k=0}^{K-2} \psi^0_{i_k}]h\mathrm{d}\nu_{-i_{K-1}}} \, .\end{aligned}$$

Using (a) at rank $n = 1$, we have

$$1/\bar{c}^{K-1} \le \exp[\textstyle\bigoplus_{k=0}^{K-2}\{\psi^1_{i_k} - \psi^0_{i_k}\}] \le 1/\underline{c}^{K-1} \, ,$$

and therefore, we obtain (b) by taking $\alpha_{1,K-1} = 0$. Let us assume that the result is verified for some $n \in \mathbb{N}^*$. We have

$$\exp[\psi_{i_0}^n - \psi_{i_0}^{n+1}] = \frac{\int \exp[\bigoplus_{k=1}^{K-1} \psi_{i_k}^n] h \mathrm{d}\nu_{-i_0}}{\int \exp[\bigoplus_{k=1}^{K-1} \psi_{i_k}^{n-1}] h \mathrm{d}\nu_{-i_0}}$$
$$= \frac{\int \exp[\bigoplus_{k=1}^{K-2} \{\psi_{i_k}^n - \psi_{i_k}^{n-1}\} \oplus \{\psi_{i_{K-1}}^n - \psi_{i_{K-1}}^{n-1}\} + \bigoplus_{k=1}^{K-1} \psi_{i_k}^{n-1}] h \mathrm{d}\nu_{-i_0}}{\int \exp[\bigoplus_{k=1}^{K-1} \psi_{i_k}^{n-1}] h \mathrm{d}\nu_{-i_0}}$$

Using (a) and (b) at rank $n$, we have

$$1/\bar{c}^{K-2} \cdot (\underline{c}/\bar{c})^{(K-2)\sum_{k=1}^{K-2} \alpha_{n,k}} \leq \exp[\bigoplus_{k=1}^{K-2} \{\psi_{i_k}^n - \psi_{i_k}^{n-1}\}]$$
$$\leq 1/\underline{c}^{K-2} \cdot (\bar{c}/\underline{c})^{(K-2)\sum_{k=1}^{K-2} \alpha_{n,k}} ,$$
$$\underline{c}^{K-1} \cdot (\underline{c}/\bar{c})^{\alpha_{n,K-1}(K-2)} \leq \exp[\psi_{i_{K-1}}^n - \psi_{i_{K-1}}^{n-1}] \leq \bar{c}^{K-1} \cdot (\bar{c}/\underline{c})^{\alpha_{n,K-1}(K-2)} .$$

Therefore, we obtain

$$\underline{c} \cdot (\underline{c}/\bar{c})^{(K-2)\sum_{k=1}^{K-1} \alpha_{n,k}} \leq \exp[\bigoplus_{k=1}^{K-2} \{\psi_{i_k}^n - \psi_{i_k}^{n-1}\} \oplus \{\psi_{i_{K-1}}^n - \psi_{i_{K-1}}^{n-1}\}]$$
$$\leq \bar{c} \cdot (\bar{c}/\underline{c})^{(K-2)\sum_{k=1}^{K-1} \alpha_{n,k}} ,$$
$$\underline{c} \cdot (\underline{c}/\bar{c})^{(K-2)\sum_{k=1}^{K-1} \alpha_{n,k}} \leq \exp[\psi_{i_0}^n - \psi_{i_0}^{n+1}] \leq \bar{c} \cdot (\bar{c}/\underline{c})^{(K-2)\sum_{k=1}^{K-1} \alpha_{n,k}} .$$

Now, we define $\alpha_{n+1,0} = \sum_{k=1}^{K-1} \alpha_{n,k}$ to obtain (a) for $k = 0$. Consider now $k \in \{1, \ldots, K-2\}$. Following the same steps as above, we recursively define

$$\alpha_{n+1,k} = \sum_{j=0}^{k-1} \alpha_{n+1,j} + \sum_{j'=k+1}^{K-1} \alpha_{n,j'} ,$$

which gives (a) at rank $n+1$. Let us now prove (b) at rank $n+1$. We have

$$\exp[\psi_{i_{K-1}}^n - \psi_{i_{K-1}}^{n+1}] = \frac{\int \exp[\bigoplus_{k=0}^{K-2} \psi_{i_k}^{n+1}] h \mathrm{d}\nu_{-i_{K-1}}}{\int \exp[\bigoplus_{k=0}^{K-2} \psi_{i_k}^{n}] h \mathrm{d}\nu_{-i_{K-1}}}$$
$$= \frac{\int \exp[\bigoplus_{k=0}^{K-2} \{\psi_{i_k}^{n+1} - \psi_{i_k}^n\} + \bigoplus_{k=0}^{K-2} \psi_{i_k}^n] h \mathrm{d}\nu_{-i_k}}{\int \exp[\bigoplus_{k=0}^{K-2} \psi_{i_k}^n] h \mathrm{d}\nu_{-i_{K-1}}} .$$

Using (a) at rank $n+1$, we obtain

$$1/\bar{c}^{K-1} \cdot (\underline{c}/\bar{c})^{(K-2)\sum_{k=0}^{K-2} \alpha_{n+1,k}} \leq \exp[\bigoplus_{k=0}^{K-2} \{\psi_{i_k}^{n+1} - \psi_{i_k}^n\}]$$
$$\leq 1/\underline{c}^{K-1} \cdot (\bar{c}/\underline{c})^{(K-2)\sum_{k=0}^{K-2} \alpha_{n+1,k}} .$$

Therefore, by taking $\alpha_{n+1,K-1} = \sum_{k=0}^{K-2} \alpha_{n+1,k}$, we obtain (b), which concludes the proof. □

Unfortunately, Lemma 22 only yields non-vacuous bounds in the case $K = 2$. Indeed, when $K > 2$, the sequence $\{\alpha_{n,k}\}_{n \in \mathbb{N}^*, k \in \{0,\ldots,K-1\}}$ leads to increase the bounds on the quantities $\exp[\psi_{i_k}^{n-1} - \psi_{i_k}^n]$, which motivates the use of **A**5.

### D.3 Proof of Section 5

For the rest of this section, we consider the multi-marginal Schrödinger bridge problem given by (TreeSB) and establish in Proposition 24 the correspondence with the regularized Wasserstein propagation problem presented in Solomon et al. (2014, 2015). We first state a technical result.

**Lemma 23.** *Let $\varepsilon > 0$. Assume that $\pi^0$ is given by (2), where $r \in \mathsf{V}$ is chosen arbitrarily. Then, for any $\pi \in \mathscr{P}_{\mathsf{T}_r}$, we have*

$$\varepsilon \mathrm{KL}(\pi \mid \pi^0) = \sum_{(v,v') \in \mathsf{E}_r} \{w_{v,v'} \mathbb{E}_{\pi_{v,v'}}[\|X_v - X_{v'}\|^2] - \varepsilon \mathrm{H}(\pi_{v,v'})\}$$
$$+ \varepsilon \sum_{v \in \mathsf{V}} \mathrm{card}(\mathrm{C}_v) \mathrm{H}(\pi_v) + \varepsilon \mathrm{KL}(\pi_r \mid \pi_r^0) ,$$

*where we recall that $\mathrm{C}_v = \{v' \in \mathsf{V} : (v,v') \in \mathsf{E}_r\}$.*

*Proof.* Since $\pi, \pi^0 \in \mathscr{P}_{\mathsf{T}_r}$, we obtain the following decomposition

$$
\begin{aligned}
&\mathrm{KL}(\pi \mid \pi^0) \\
&\quad = \mathrm{KL}(\pi_r \prod_{(v,v')\in\mathsf{E}_r} \pi_{v'|v} \mid \pi_r^0 \prod_{(v,v')\in\mathsf{E}_r} \pi^0_{v'|v}) \\
&\quad = \mathrm{KL}(\pi_r \mid \pi_r^0) + \sum_{(v,v')\in\mathsf{E}_r} \int_{\mathbb{R}^d} \mathrm{KL}(\pi_{v'|v}(\cdot|x_v) \mid \pi^0_{v'|v}(\cdot|x_v))\mathrm{d}\pi_v(x_v) \\
&\quad = \mathrm{KL}(\pi_r \mid \pi_r^0) - \sum_{(v,v')\in\mathsf{E}_r} \int_{\mathbb{R}^d\times\mathbb{R}^d} \log \pi^0_{v'|v} \mathrm{d}\pi_{v,v'} - \sum_{(v,v')\in\mathsf{E}_r} \int_{\mathbb{R}^d} \mathrm{H}(\pi_{v'|v}(\cdot|x_v))\mathrm{d}\pi_v(x_v) \;.
\end{aligned}
$$

We finally obtain the result by using the definition of $\pi^0$ and noticing that $\int_{\mathbb{R}^d} \mathrm{H}(\pi_{v'|v}(\cdot|x_v))\mathrm{d}\pi_v(x_v) = \mathrm{H}(\pi_{v,v'}) - \mathrm{H}(\pi_v)$ for any $(v,v') \in \mathsf{E}_r$. □

**Proposition 24.** *Let $\varepsilon > 0$ and $\mu_0 \in \mathscr{P}$ such that $\mu_0 \ll \mathrm{Leb}$. Assume that $\pi^0$ is given by (2), where $r \in \mathsf{V}$ is chosen arbitrarily, and that $\varphi_r = \mathrm{d}\mu_0/\mathrm{dLeb}$. Also assume **A**2. Then, the set of marginals of the solution to (TreeSB) is exactly the solution to the entropic-regularized Wasserstein Propagation problem (Solomon et al., 2014, 2015) defined by*

$$
\begin{aligned}
\arg\min\{\textstyle\sum_{(v,v')\in\mathsf{E}_r} w_{v,v'} W^2_{2,\varepsilon/w_{v,v'}}(\nu_v,\nu_{v'}) + \varepsilon \sum_{v\in\mathsf{V}} \mathrm{card}(\mathrm{C}_v)\mathrm{H}(\nu_v) + \varepsilon\mathrm{KL}(\nu_r \mid \mu_0) : \quad &\text{(WP)} \\
\{\nu_v\}_{v\in\mathsf{V}} \in \mathscr{P}^{\ell+1},\ \nu_i = \mu_i, \forall i \in \mathsf{S}\} \;, &
\end{aligned}
$$

*where we recall that $\mathrm{C}_v = \{v' \in \mathsf{V} : (v,v') \in \mathsf{E}_r\}$.*

*Proof.* Assume that $\pi^0$ is given by (2), where $r \in \mathsf{V}$ is chosen arbitrarily, and that $\varphi_r = \mathrm{d}\mu_0/\mathrm{dLeb}$. In particular, we have $\pi_r^0 = \mu_0$. Moreover, it is clear that $\pi^0$ verifies **A**1, and **A**3 by Proposition 16.

Let $\{\nu_v\}_{v\in\mathsf{V}} \in \mathscr{P}^{\ell+1}$ and $\{\nu^{(v,v')}\}_{(v,v')\in\mathsf{E}_r} \in (\mathscr{P}^{(2)})^{|\mathsf{E}_r|}$. We define

$$
\begin{aligned}
F(\{\nu_v\}) &= \textstyle\sum_{(v,v')\in\mathsf{E}_r} w_{v,v'} W^2_{2,\varepsilon/w_{v,v'}}(\nu_v,\nu_{v'}) + \varepsilon \sum_{v\in\mathsf{V}} \mathrm{card}(\mathrm{C}_v)\mathrm{H}(\nu_v) + \varepsilon\mathrm{KL}(\nu_r \mid \mu_0) \;, \\
G(\nu_r, \{\nu^{(v,v')}\}) &= \textstyle\sum_{(v,v')\in\mathsf{E}_r} \{w_{v,v'}\mathbb{E}_{\nu^{(v,v')}}[\|X_v - X_{v'}\|^2] - \varepsilon\mathrm{H}(\nu^{(v,v')})\} \\
&\quad + \varepsilon \textstyle\sum_{(v,v')\in\mathsf{E}_r} \mathrm{H}(\nu_v^{(v,v')}) + \varepsilon\mathrm{KL}(\nu_r \mid \mu_0) \;.
\end{aligned}
$$

By definition of the regularized Wasserstein distance given in (3), we have for any $\{\nu_v\}_{v\in\mathsf{V}} \in \mathscr{P}^{\ell+1}$

$$
F(\{\nu_v\}) = \min\{G(\nu_r, \{\nu^{(v,v')}\}) : \nu^{(v,v')} \in \mathscr{P}^{(2)}, \nu_v^{(v,v')} = \nu_v, \nu_{v'}^{(v,v')} = \nu_{v'}, \forall (v,v') \in \mathsf{E}_r\} \;. \tag{26}
$$

In particular, we have $F(\{\pi_v\}) \le G(\pi_r, \{\pi_{v,v'}\})$ for any $\pi \in \mathscr{P}^{(\ell+1)}$. We now prove the result of Proposition 24 in two steps denoted by **Step 1** and **Step 2**.

**Step 1.** Let us not assume **A**2 for now. In this case, we prove in **Step 1.a** and **Step 1.b** that solving (WP) is equivalent to solving a modified version of (TreeSB) given by

$$
\pi^\star = \mathrm{argmin}\{\mathrm{KL}(\pi|\pi^0) \;:\; \pi \in \mathscr{P}_{\mathsf{T}_r},\ \pi_i = \mu_i \;, \forall i \in \mathsf{S}\} \;. \tag{$\mathsf{T}_r$-TreeSB}
$$

Remark that any solution to ($\mathsf{T}_r$-TreeSB) is a solution to (TreeSB), but the converse result may not be true.

**Step 1.a: (WP) $\Longrightarrow$ ($\mathsf{T}_r$-TreeSB).** Consider a solution $\{\nu_v^\star\}_{v\in\mathsf{V}}$ to (WP). For any $(v,v') \in \mathsf{E}_r$, $W^2_{2,\varepsilon/w_{v,v'}}(\nu_v^\star, \nu_{v'}^\star)$ is well defined and thus, there exists $\nu^{(v,v')} \in \Pi(\nu_v^\star, \nu_{v'}^\star)$ such that

$$
\nu^{(v,v')} \in \mathrm{argmin}\{\mathbb{E}_\pi[\|X_v - X_{v'}\|^2] - (\varepsilon/w_{v,v'})\mathrm{H}(\pi) : \pi \in \Pi(\nu_v^\star, \nu_{v'}^\star)\} \;. \tag{27}
$$

Using the gluying lemma, we build the probability measure $\pi^\star = \nu_r^\star \prod_{(v,v')\in\mathsf{E}_r} \nu_{v'|v}^{(v,v')}$ such that (i) $\pi^\star \in \mathscr{P}_{\mathsf{T}_r}$, and (ii) $\pi^\star_{v,v'}$ and $\nu^{(v,v')}$ have the same distribution for any $(v,v') \in \mathsf{E}_r$. In particular, we have $\pi_i^\star = \mu_i$ for any $i \in \mathsf{S}$.

Let us show now that $\pi^\star$ is a solution to ($\mathsf{T}_r$-TreeSB). Let $\pi \in \mathscr{P}_{\mathsf{T}_r}$ such that $\pi_i = \mu_i$ for any $i \in \mathsf{S}$. We have

$$\begin{aligned}
\epsilon \mathrm{KL}(\pi \mid \pi^0) &= G(\pi_r, \{\pi_{v,v'}\}) && \text{(Lemma 23)} \\
&\geq F(\{\pi_v\}) \\
&\geq F(\{\nu_v^\star\}) && \text{(definition of } \nu^\star) \\
&= G(\nu_r^\star, \{\nu^{(v,v')}\}) && \text{(see (27))} \\
&= G(\pi_r^\star, \{\pi^\star_{(v,v')}\}) && \text{(definition of } \pi^\star) \\
&= \epsilon \mathrm{KL}(\pi^\star \mid \pi^0)\ . && \text{(Lemma 23)}
\end{aligned}$$

Therefore, $\pi^\star$ is a solution to ($\mathsf{T}_r$-TreeSB).

**Step 1.b:** ($\mathsf{T}_r$-TreeSB) $\Longrightarrow$ (WP)**.** Consider now a solution $\pi^\star$ to ($\mathsf{T}_r$-TreeSB). Since $\pi^\star \in \mathscr{P}_{\mathsf{T}_r}$, we have $\pi^\star = \pi^\star_r \prod_{(v,v') \in \mathsf{E}_r} \pi^\star_{v'|v}$ and $\pi_i^\star = \mu_i$ for any $i \in \mathsf{S}$.

Let us show that $\{\pi_v^\star\}_{v \in \mathsf{V}}$ is a solution to (WP). Let $\{\nu_v\}_{v \in \mathsf{V}} \in \mathscr{P}^{\ell+1}$ such that $\nu_i = \mu_i$ for any $i \in \mathsf{S}$.

Let $\{\nu^{(v,v')}\}_{(v,v') \in \mathsf{E}_r}$ be a family of probability measures such that $\nu^{(v,v')} \in \mathscr{P}^{(2)}, \nu_v^{(v,v')} = \nu_v, \nu_{v'}^{(v,v')} = \nu_{v'}$ for any $(v,v') \in \mathsf{E}_r$.

Using the gluying lemma, we build the probability measure $\pi = \nu_r \prod_{(v,v') \in \mathsf{E}_r} \nu^{(v,v')}_{v'|v}$, such that (i) $\pi \in \mathscr{P}_{\mathsf{T}_r}$ and (ii) $\pi_{v,v'}$ and $\nu^{(v,v')}$ have the same distribution for any $(v,v') \in \mathsf{E}_r$. We have

$$\begin{aligned}
\varepsilon \mathrm{KL}(\pi \mid \pi^0) &= G(\pi_r, \{\pi_{(v,v')}\}) && \text{(Lemma 23)} \\
&= G(\nu_r, \{\nu^{(v,v')}\}) && \text{(definition of } \pi) \\
&\geq \varepsilon \mathrm{KL}(\pi^\star \mid \pi^0) && \text{(definition of } \pi^\star) \\
&= G(\pi_r^\star, \{\pi^\star_{(v,v')}\})\ . && \text{(Lemma 23)}
\end{aligned}$$

By taking the infimum in the previous inequality over the families $\{\nu^{(v,v')}\}_{(v,v') \in \mathsf{E}_r}$, we obtain by (26) that

$$F(\{\nu_v\}) \geq G(\pi_r^\star, \{\pi^\star_{(v,v')}\}) \geq F(\{\pi_v^\star\}),$$

and therefore, $\{\pi_v^\star\}_{v \in \mathsf{V}}$ is a solution to (WP).

**Step 2.** We now assume **A**2. By Proposition 3, there exists a unique solution $\pi^\star \in \mathscr{P}^{(\ell+1)}$ to (TreeSB) such that we $\pi^0$-a.s. have $(\mathrm{d}\pi^\star / \mathrm{d}\pi^0) = \exp[\bigoplus_{i \in \mathsf{S}} \psi_i^\star]$, where $\{\psi_i^\star\}_{i \in \mathsf{S}}$ are measurable potentials with $\psi^\star : \mathbb{R}^d \to \mathbb{R}$. Since $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$, we also have $\pi^\star \in \mathscr{P}_{\mathsf{T}_r}$, *i.e.*, the potentials $\{\psi_i^\star\}_{i \in \mathsf{S}}$ do not modify the Markovian nature of $\pi^0$. Therefore, $\pi^\star$ is also the unique solution to ($\mathsf{T}_r$-TreeSB). Using the equivalence between ($\mathsf{T}_r$-TreeSB) and (WP) established in **Step 1**, we finally obtain the result of Proposition 24. □

In particular, Proposition 7 directly derives from Proposition 24 by taking $r = i_{K-1}$ and $\mu_0 = \mu_{i_{K-1}}$.

### D.4 Comparison with Haasler et al. (2021)

In their work, Haasler et al. (2021) study the *static* and *discrete-state* counterpart of our approach. Given a state space $\mathsf{X}$ such that $|\mathsf{X}| = n+1$ with $n \in \mathbb{N}$, they establish a correspondence between multi-marginal EOT with a general tree-based cost and discrete-time multi-marginal static Schrödinger bridge, and provide an efficient method to solve these problems. In this section, we provide details on their framework and give a precise comparison between our theory and their results.

To be coherent with the setting of Haasler et al. (2021), we adapt here some of our notation. Let us define $\mathsf{Z}^{(q)} = \mathbb{R}_+^{(n+1)^q}$. For any $q \in \mathbb{N}^*$, the set of probability measures on $\mathsf{X}^q$ is defined as $\mathscr{P}^{(q)} = \{M \in \mathsf{Z}^{(q)} : \langle M, \mathbf{1} \rangle = 1\}$. We denote $\mathscr{P} = \mathscr{P}^{(1)}$. For any tensors $M, P \in \mathsf{Z}^{(q)}$, the Kullback-Leibler divergence between $M$ and $P$ is defined as $\mathrm{KL}(M \mid P) = \langle M \log(M/P) - M + P, \mathbf{1} \rangle$ and the entropy of $M$ is defined as $\mathrm{H}(M) = -\mathrm{KL}(M \mid \mathbf{1})$, where the operations are meant componentwise. In the rest of the section, we consider an undirected tree $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with $|\mathsf{V}| = \ell + 1$ such that $\mathsf{V}$ may be identified with $\{0, \dots, \ell\}$.

**Details on the results of Haasler et al. (2021).** In their paper, the authors consider a cost tensor $C \in \mathsf{Z}^{(\ell+1)}$ that factorizes along $\mathsf{T}$, *i.e.*, for any $\{j_0, \dots, j_\ell\}$ with for any $i \in \{0, \dots, \ell\}$, $j_i \in \{0, \dots, n\}$, we have

$$C_{j_0,\dots,j_\ell} = \textstyle\sum_{(v,v')\in\mathsf{E}} C^{\{v,v'\}}_{j_v,j_{v'}} ,$$

where $C^{\{v,v'\}} \in \mathsf{Z}^{(2)}$ is a cost matrix for transportation between the marginals at vertices $v$ and $v'$, see (Haasler et al., 2021, Eq. (3.1)). In particular, this cost can be seen as the discrete counterpart of the tree-based cost introduced in (1) in the quadratic setting.

Given a subset $\mathsf{S} \subset \mathsf{V}$ with $|\mathsf{S}| = K$ and a set of marginals $\{\mu_i\}_{i\in\mathsf{S}} \in \mathscr{P}^K$, Haasler et al. (2021) study the EOT problem associated to $\mathsf{T}$, see (Haasler et al., 2021, Eq. (2.4)), which is given by

$$\operatorname{argmin}\{\langle C, M\rangle - \varepsilon \mathrm{H}(M) : M \in \mathscr{P}^{(\ell+1)}, \mathrm{proj}_i(M) = \mu_i, \forall i \in \mathsf{S}\} \; . \tag{discrete-EmOT}$$

This problem may be solved with Sinkhorn algorithm (Cuturi, 2013; Knight, 2008; Sinkhorn & Knopp, 1967), for which the authors provide an efficient implementation adapted to the tree-based setting, see (Haasler et al., 2021, Algorithm 3.1). Moreover, they state the convergence of their method in (Haasler et al., 2021, Theorem 3.5), as a direct consequence of the results presented in Luo & Tseng (1992).

In (Haasler et al., 2021, Section 4.2), it is assumed that $\mathsf{S}$ corresponds to the set of the leaves of $\mathsf{T}$, as we do, and it is shown an equivalence between (discrete-EmOT) and the discrete-state static SB problem stated in (Haasler et al., 2021, Eq 4.2), which is given by

$$\begin{aligned}&\operatorname{argmin}\{\textstyle\sum_{(v,v')\in\mathsf{E}_r} \mathrm{KL}(M^{(v,v')} \mid \mathrm{diag}(\nu_v) A^{(v,v')}) : \qquad \text{(discrete-TreeSB)}\\ &M^{(v,v')} \in \mathscr{P}^{(2)}, \{\nu_v\}_{v\in\mathsf{V}} \in \mathscr{P}^{\ell+1}, M^{(v,v')}\mathbf{1} = \nu_v, {M^{(v,v')}}^\top \mathbf{1} = \nu_{v'},\ \nu_i = \mu_i, \forall i \in \mathsf{S}\} \; ,\end{aligned}$$

where $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$ is the directed version of $\mathsf{T}$ rooted in an arbitrary vertex $r \in \mathsf{S}$, and $A^{(v,v')} = \exp(-C^{(v,v')}/\varepsilon) \in \mathsf{Z}^{(2)}$ for any $(v, v') \in \mathsf{E}_r$. Remark that $A^{(v,v')}$ may not necessarily be a transition probability matrix.

Finally, Haasler et al. (2021) provide two main numerical experiments. In (Haasler et al., 2021, Section 5.2), they consider a tree with 15 vertices, 14 edges and 8 leaves, combined to the state-space $\mathsf{X} = \{0, 1\}^{50\times 50}$, and solve the corresponding (discrete-EmOT) problem for the quadratic cost. In (Haasler et al., 2021, Section 6), they apply their methodology to estimate ensemble flows on a hidden Markov chain. Given $\tau \in \mathbb{N}^*$, they consider a tree $\mathsf{T}$ with $\tau$ internal vertices (modeling the distribution of $N$ agents at time $t \in \{1, \dots, \tau\}$), that are linearly linked, and such that each of these vertices is independently linked to $S$ leaves of $\mathsf{T}$ (modeling observations at time $t \in \{1, \dots, \tau\}$). In this setting, the state space is given by $\mathsf{X} = \{1, \dots, 100\}^N$. They solve the formulation (discrete-TreeSB) where the reference measure is chosen as a random walk.

**Comparison with our results.** We now establish remarks on the main differences between our methodology and the work of Haasler et al. (2021).

First of all, the continuous state-space counterpart of (discrete-TreeSB) is given by

$$\operatorname{argmin}\{\mathrm{KL}(\pi \mid \pi^0) : \pi \in \mathscr{P}_{\mathsf{T}_r}, \pi_i = \mu_i, \forall i \in \mathsf{S}\} \; , \tag{28}$$

where $\pi^0$ is a reference measure which factorizes along $\mathsf{T}_r$. In this case, $\pi_{v,v'}$, $\pi_v$ and $\pi^0_{v'|v}$ in (28) respectively correspond to the continuous version of $M^{(v,v')}$, $\nu_v$ and $A^{(v,v')}$ in (discrete-TreeSB). In contrast, our formulation of the multi-marginal Tree Schrödinger Bridge problem given in (TreeSB) is a minimization problem over all probability measures $\pi \in \mathscr{P}^{(\ell+1)}$, and is not restricted to the distributions that admit a Markovian factorization along $\mathsf{T}$ as in (28). Hence, our framework may be considered more general. Remark that under **A**1, **A**2 and **A**3, Proposition 3 states that (TreeSB) admits a unique solution $\pi^\star \ll \pi^0$ such that $(\mathrm{d}\pi^\star/\mathrm{d}\pi^0)$ can be written with potentials. Then, $\pi^\star \in \mathscr{P}_{\mathsf{T}_r}$ since $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$, and (TreeSB) is then equivalent to (28).

Furthermore, (EmOT) is more general than the continuous version of (discrete-EmOT), which we can recover by taking any measure $\nu$ of the form $(\mathrm{d}\nu/\mathrm{dLeb}) = \exp[\bigoplus_{i\in\mathsf{S}} \varphi_i]$ in (EmOT), where $\{\varphi_i\}_{i\in\mathsf{S}}$ is a family of potentials such that $\left|\int_{\mathbb{R}^d} \varphi_i \mathrm{d}\mu_i\right| < \infty$ for any $i \in \mathsf{S}$. As a consequence, our setting allows us to choose the root $r \in \mathsf{V}\backslash\mathsf{S}$ for the SB problem, whereas Haasler et al. (2021) only consider the case where $r \in \mathsf{S}$. In the latter case, we establish in Appendix E that $r$ can be chosen arbitrarily, as stated by (Haasler et al., 2021, Corollary 4.3).

Finally, TreeDSB deeply differs from the framework of Haasler et al. (2021) due its *dynamic* nature. Although we solve the same tree-based static SB problem (up to continuous/discrete state-space consideration), our approach consists in computing dynamic iterates (*i.e.*, path measures) using diffusion-based methods instead of static iterates (*i.e.*, distributions) using Sinkhorn algorithm. This paradigm is at the core of the DSB (De Bortoli et al., 2021) methodology, and offers an efficient approach to tackle high-dimensional settings, where Sinkhorn algorithm would fail.

Here, we present some advantages of the method proposed by Haasler et al. (2021) compared to ours. First, Haasler et al. (2021) may choose any kind of tree-based cost in practice, while our methodology only holds for the quadratic cost. This limitation is shared with all approaches based on the DSB (De Bortoli et al., 2021) methodology. Indeed, since the cost is determined by the reference path measure, we often choose quadratic costs associated with Brownian motions or Ornstein-Uhlenbeck processes. Moreover, Haasler et al. (2021) may consider various inhomogeneous (discrete) state spaces for the vertices of $\mathsf{T}$, as presented in their numerical experiments. In our case, this approach is not compatible with our diffusion-based method. Finally, unlike Haasler et al. (2021), our method is not scalable with the number of vertices or edges in $\mathsf{T}$ due to computational limits. This limitation is common to all multi-marginal approaches which rely on neural networks to parameterize the potential and/or the distributions of the multi-marginal OT method, see Li et al. (2020); Fan et al. (2020); Korotin et al. (2022, 2021) for instance.

## E Further results on TreeSB

**Choice of the root $r$ in (TreeSB).** We recall that the reference measure $\pi^0$ considered in (TreeSB), which is defined in (2), verifies $\pi^0 \in \mathscr{P}_{\mathsf{T}_r}$ for some fixed root $r \in \mathsf{V}$ and $\pi^0_r \ll \mathrm{Leb}$ with density $\varphi_r$. Moreover, we have $\pi^0_{v'|v}(\cdot \mid x_v) = \mathrm{N}(x_v, \varepsilon/(2w_{v,v'})\mathrm{I}_d)$ for any $(v, v') \in \mathsf{E}_r$, and thus, $\pi^0$ is entirely determined by the choice of the root $r$ and the density on the corresponding vertex $\varphi_r$.

As presented in Appendix D.1, we recall that (TreeSB) is equivalent to any multi-marginal Tree-SB problem with a reference measure $\bar{\pi}^0$ given by (15), *i.e.*, $\bar{\pi}^0$ writes as $(\mathrm{d}\bar{\pi}^0/\mathrm{d}\pi^0) = \exp[\bigoplus_{i\in\mathsf{S}} \psi_i^0]$, where $\{\psi_i^0\}_{i\in\mathsf{S}}$ is a family of measurable potentials with $\psi_i^0 : \mathbb{R}^d \to \mathbb{R}$ such that $\left|\int_{\mathbb{R}^d} \psi_i^0 \mathrm{d}\mu_i\right| < \infty$ for any $i \in \mathsf{S}$. In the case where $r$ is chosen as a leaf of $\mathsf{T}$, this result implies that (TreeSB) is unchanged if

(a) $\varphi_r = \mathrm{d}\nu/\mathrm{dLeb}$ where $\nu \in \mathscr{P}$ is such that $\mathrm{KL}(\mu_r|\nu) < \infty$,

(b) $r$ is replaced by $r' \in \mathsf{S}$, as long as $\mathrm{H}(\mu_r) < \infty$ and $\mathrm{H}(\mu_{r'}) < \infty$.

Therefore, under **A**0, the setting chosen in Section 3 is equivalent to any other setting where $r$ is arbitrarily chosen in $\mathsf{S}$ and $\varphi_r = \mathrm{d}\nu/\mathrm{dLeb}$ where $\mathrm{KL}(\mu_r|\nu) < \infty$.

Consider now the case where $r \in \mathsf{S}^{\mathrm{c}}$, *i.e.*, $r$ is not a leaf of $\mathsf{T}$. Then, the choice of $\varphi_r$ can not be made arbitrarily anymore, since it determines a further regularization on the $r$-th marginal of the solution to (TreeSB). In this setting, the sequence defined by (mIPF) is unchanged. Hence, TreeDSB proceeds in the same manner as presented in Section 3, except for the first iteration, which we detail now.

Let us define $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{i_0}}(i_0, r)$, where $\mathsf{T}_{i_0} = (\mathsf{V}, \mathsf{E}_{i_0})$ is the directed version of $\mathsf{T}$ rooted in $i_0$. We recall that first iterate of (mIPF) is defined by

$$\pi^1 = \mathrm{argmin}\{\mathrm{KL}(\pi \mid \pi^0) : \pi \in \mathscr{P}^{(\ell+1)}, \pi_{i_0} = \mu_{i_0}\} \; .$$

Following the proof of Lemma 12, it is clear that

$$\pi^1 = \mu_{i_0} \bigotimes_{(v,v')\in\mathsf{P}} \pi^0_{v'|v} \bigotimes_{(v,v')\in\mathsf{E}_{i_0}\backslash\mathsf{P}} \pi^0_{v'|v} = \mu_{i_0} \bigotimes_{(v,v')\in\mathsf{E}_{i_0}} \pi^0_{v'|v} \; ,$$

where we emphasize that $\mathsf{P} = \{(v, v') \in \mathsf{E}_{i_0} : (v', v) \in \mathsf{E}_r\}$. Therefore, Proposition 2 still applies between $r$ and $i_0$, by considering $r$ instead of $i_{K-1}$. In practice, this means that the first iteration of TreeDSB consists in computing the time reversal of the path measures $\mathbb{P}^0_{(v',v)}$ for any $(v, v') \in \mathsf{P}$.

**Extension of the regularized Wasserstein barycenter problem (regWB).** Consider the regularized Wasserstein-2 barycenter problem defined as follows

$$\mu^\star_\varepsilon = \arg\min\{\textstyle\sum_{i=1}^{\ell} w_i W^2_{2,\varepsilon/w_i}(\mu, \mu_i) + \ell\varepsilon \mathrm{H}(\mu) + \varepsilon \mathrm{KL}(\mu \mid \mu_0) : \mu \in \mathscr{P}\} \ , \qquad (\mu_0\text{-regWB})$$

where $(w_i)_{i\in\{1,\dots,\ell\}} \in (0, +\infty)^\ell$ and $\mu_0 \in \mathscr{P}$ is a reference measure. This formulation admits a further regularization compared to (regWB), which tends to make $\mu^\star_\varepsilon$ closer to $\mu_0$. In particular, given a Wasserstein barycenter problem onto a star-shaped tree, the formulation ($\mu_0$-regWB) may be more adapted than (regWB) if we have an *a priori* on the form of the regularized barycenter. In the case where $\mu_0 = \mathrm{N}(0, \sigma_0^2 \mathrm{I}_d)$, letting $\sigma_0 \to \infty$, we recover the $(\ell\varepsilon, (\ell-1)\varepsilon)$ doubly-regularized Wasserstein barycenter problem (regWB). In the same spirit as Proposition 7, we can derive the following result from Proposition 24, which proves that ($\mu_0$-regWB) can be solved with TreeDSB.

**Proposition 25.** *Let $\varepsilon > 0$ and $\mu_0 \in \mathscr{P}$ such that $\mu_0 \ll \mathrm{Leb}$. Assume* **A**0*. Also assume that $\mathsf{T}$ is a star-shaped tree with central node indexed by 0, and that the reference measure of* (TreeSB) *defined in* (2) *verifies $r = 0$ and $\varphi_r = \mathrm{d}\mu_0/\mathrm{dLeb} > 0$. Under* **A**2*,* ($\mu_0$-regWB) *has a unique solution $\pi^\star_0$, where $\pi^\star$ is the solution to* (TreeSB)*.*

Below, we provide practical guidelines to parameterize $\mu_0$ when it is chosen as a Gaussian distribution.

**Gaussian design of $\mu_0$ in ($\mu_0$-regWB).** Consider an undirected star-shaped tree $\mathsf{T}$ with $K+1$ vertices and leaves $\{1, \dots, K\}$. In order to incorporate the marginal constraints in the penalization brought by $\mu_0$ when it is a Gaussian distribution, we set its mean to $\sum_{i=1}^K \mathbb{E}[\mu_i]/K$ and its diagonal covariance matrix as $\alpha \times (\sum_{i=1}^K \mathrm{diag}(\mathrm{Cov}[\mu_i])^{-1}/K)^{-1}$, where the inverse operation is component-wise and $\alpha$ is a positive hyperparameter. This choice of variance helps to correctly explore the state-space at the very first iteration of TreeDSB, which is key to ensure numerical stability. In this setting, (TreeSB) verifies **A**2 and **A**3, by Proposition 15 and Proposition 16. In particular, we use this approach for two of our experiments: synthetic Gaussian datasets and Bayesian fusion, see Appendix G.

# F Algorithmic techniques

**Time discretization in TreeDSB.** Denote $k_n = (n-1) \bmod (K)$ for any $n \in \mathbb{N}$. Let $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ be a weighted undirected tree and consider the multi-marginal Schrödinger bridge problem (TreeSB) associated to this tree. We recall that for any $\{v, v'\} \in \mathsf{E}$, we define $T_{v,v'} = \varepsilon/(2w_{v,v'})$.

Consider the path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ recursively defined by (a) and (b). By combining Proposition 1, Proposition 2 and results on time reversal theory (Haussmann & Pardoux, 1986), we obtain by recursion that for any $n \in \mathbb{N}$, any $(v, v') \in \mathsf{E}_{k_n}$, $\mathbb{P}^n_{(v,v')}$ is associated with a Stochastic Differential Equation on $[0, T_{v,v'}]$ given by

$$\mathrm{d}\mathbf{X}_t = f^n_{t,v,v'}(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t, \quad \mathbf{X}_0 \sim \pi^n_v \ . \tag{29}$$

Let $N \in \mathbb{N}^*$. In order to sample from the dynamics (29) at iteration $n \in \mathbb{N}$, we consider its Euler-Maruyama discretization on $(N+1)$ time steps,

$$X_{m+1} = X_m + \gamma_{m+1} f^n_{t_m,v,v'}(X_m) + \sqrt{\gamma_{m+1}} Z_{m+1}, \quad X_0 \sim \pi^n_v \ , \tag{30}$$

where $Z_m \sim \mathrm{N}(0, \mathrm{I}_d)$ for any $m \in \{1, \dots, N\}$, $t_m = \sum_{i=1}^m \gamma_i$, and $\{\gamma_m\}_{m=1}^N \in (0, \infty)^N$ is a time schedule such that $\sum_{m=1}^N \gamma_m = T_{v,v'}$. This results in approximating the path measure $\mathbb{P}^n_{(v,v')}$ by the joint distribution $\pi^{n,N}_{(v,v')} \in \mathscr{P}^{(N+1)}$ defined by

$$\pi^{n,N}_{(v,v')} = \pi^n_v \textstyle\bigotimes_{m=0}^{N-1} \pi^{n,N}_{(v,v'),m+1|m} \ ,$$

where $\pi^{n,N}_{(v,v'),m+1|m}(\cdot|x_m) = \mathrm{N}(x_m + \gamma_{m+1} f^n_{t_m,v,v'}(x_m), \gamma_{m+1}\mathrm{I}_d)$ for any $m \in \{0, \dots, N-1\}$. If $N$ is chosen large enough, then $\pi^{n,N}_{(v,v'),m}$ and $\mathbb{P}^n_{(v,v'),t_m}$ have approximately the same distribution for any $m \in \{0, \dots, N\}$. Consequently, $(\mathbb{P}^n_{(v,v')})^R$ is naturally approximated by the joint distribution $\tilde{\pi}^{n,N}_{(v,v')} \in \mathscr{P}^{(N+1)}$ defined by

$$\tilde{\pi}^{n,N}_{(v,v')} = \pi^n_{v'} \bigotimes_{m=0}^{N-1} \pi^{n,N}_{(v,v'),N-m-1|N-m} .$$

If $N$ is chosen large enough, we obtain that

$$\pi^{n,N}_{(v,v'),N-m-1|N-m}(\cdot|x_{N-m}) \\ = \mathrm{N}(x_{N-m} - \gamma_{N-m} f^n_{t_{N-m},v,v'}(x_{N-m}) + \gamma_{N-m} \nabla \log p_{v,v',t_{N-m}}(x_{N-m}), \gamma_{N-m} \mathrm{I}_d) ,$$

where $p_{v,v',t}$ is the density of $\mathbb{P}^n_{(v,v'),t}$ w.r.t. the Lebesgue measure.

Following the construction of our dynamic iterates, we now explain how the sequence $\{\pi^n_{(v,v')}\}_{n \in \mathbb{N}^*, (v,v') \in \mathsf{E}_{k_n}}$ is recursively defined. Let $n \in \mathbb{N}$, $k_n = (n-1) \bmod(K)$. Define the path $\mathsf{P}_n = \mathrm{path}_{\mathsf{T}_{i_{k_n}}}(i_{k_n}, i_{k_n+1})$. Then, for any $(v, v') \in \mathsf{E}_{k_n+1}$,

(a) if $(v, v') \in \mathsf{E}_{k_n} \backslash \mathsf{P}_n$, then $\pi^{n+1,N}_{(v,v')} = \pi^{n+1}_v \bigotimes_{m=0}^{N-1} \pi^{n,N}_{(v,v'),m+1|m}$,

(b) if $(v', v) \in \mathsf{P}_n$, then $\pi^{n+1,N}_{(v,v')} = \pi^{n+1}_v \bigotimes_{m=0}^{N-1} \pi^{n,N}_{(v',v),N-m-1|N-m}$.

These computations may be obtained by considering the sequence given by (mIPF) to solve the multi-marginal Tree-SB problem associated to $\mathsf{T}^{(N)} = (\mathsf{V}^{(N)}, \mathsf{E}^{(N)})$, the $N$-discretized version of $\mathsf{T}$ (see Appendix B) with weights $w^{(N)}_{e_m} = 2\gamma_m/\varepsilon$, which is given by

$$\pi^\star = \mathrm{argmin}\{\mathrm{KL}(\pi|\pi^{0,N}) \, : \, \pi \in \mathscr{P}^{(\mathsf{V}^{(N)})}, \; \pi_i = \mu_i \, , \forall i \in \mathsf{S}\} ,$$

with $\pi^{0,N} = \pi^0_r \bigotimes_{(v,v') \in \mathsf{E}_r} \pi^{0,N}_{(v,v'),1:N|0}$.

To approximate the IPF recursion given by (a) and (b), we use **on each edge** of $\mathsf{T}$ the score-matching approach of De Bortoli et al. (2021), which avoids heavy computations of score approximations. The next proposition is direct adaptation of (De Bortoli et al., 2021, Proposition 3).

**Proposition 26.** *Assume that for any $n \in \mathbb{N}$, any $(v, v') \in \mathsf{E}_{k_n}$ with $k_n = (n-1) \bmod(K)$, we have*

$$\pi^{n,N}_{(v,v'),m+1|m}(\cdot|x_m) = \mathrm{N}(F^n_{m,v,v'}(x_m), \gamma_m \mathrm{I}_d) .$$

*Let $n \in \mathbb{N}$. Consider the path $\mathsf{P}_n = \mathrm{path}_{\mathsf{T}_{i_{k_n}}}(i_{k_n}, i_{k_n+1})$. Let $(v, v') \in \mathsf{E}_{k_n+1}$. Define $p^n = \pi^{n,N}_{(v,v')}$ and $m_N = N - m - 1$. Then, if $(v', v) \in \mathsf{P}_n$, we have*

$$F^{n+1}_{m,v,v'} = \mathrm{argmin}_{\mathrm{F} \in \mathrm{L}^2(\mathbb{R}^d, \mathbb{R}^d)} \\ \mathbb{E}_{p^n_{m_N, m_N+1}}[\|\mathrm{F}(X_{m_N+1}) - (X_{m_N+1} + F^n_{m_N,v',v}(X_{m_N}) - F^n_{m_N,v',v}(X_{m_N+1}))\|^2], \tag{31}$$

*otherwise, we have $F^{n+1}_{m,v,v'} = F^n_{m,v,v'}$.*

In practice, we use two neural networks per edge $\{v, v'\} \in \mathsf{E}$, one for each possible direction of the edge, such that $F_{v,v'}(\theta^n_{v,v'}, m, x) \approx F^n_{m,v,v'}(x)$ and $F_{v',v}(\theta^n_{v',v}, m, x) \approx F^n_{m,v',v}(x)$. For any $\{v, v'\} \in \mathsf{E}$, the parameter $\theta^n_{v,v'}$ is updated at iteration $n$ via the score matching loss defined by (31) in Proposition 26 if $(v, v') \in \mathrm{path}_{T_{i_{k_n}}}(i_{k_n}, i_{k_n+1})$, see Algorithm 1.

# G Additional experimental results and details

The numerical experiments presented in Section 7 are obtained by our own Pytorch implementation, which is inspired from the code[6] provided by De Bortoli et al. (2021). We first provide information on the general setting of our experiments in Appendix G.1, and then give details on each of them in Appendix G.2 along with additional results. We recall that a mIPF cycle is defined as a subset of $K$ consecutive iterations of (mIPF) and that the order of the leaves given by $\{i_0, \dots, i_{K-1}\}$ is randomly shuffled at each new mIPF cycle.

[6] https://github.com/JTT94/diffusion_schrodinger_bridge

### G.1 General experimental setup

**Implementation of Algorithm 1 in practice.** Let $n \in \mathbb{N}$, with $k_n = (n-1) \bmod (K)$, $k_n + 1 = n \bmod (K)$. Consider the path $\mathsf{P}_n = \mathrm{path}_{\mathsf{T}_{i_{k_n}}}(i_{k_n}, i_{k_n+1})$. Assume that we are provided with a dataset $\mathsf{D}_{i_{k_n}}$, which contains $M$ samples from $\pi^n_{i_{k_n}}$. Following Lines 7-9 in Algorithm 1, we apply processes (a) and (b) recursively on the edges $(v, v') \in \mathsf{P}_n$.

(a) Sampling step (Line 7). For any $x_0 \in \mathsf{D}_v$, we sample from the diffusion trajectory (30) given by the Euler Maruyama discretization of $\mathbb{P}^n_{v,v'}$ starting from $x_0$. This gives us $M \times N$ trajectory samples. We then store the last iterate of each trajectory in a new dataset $\mathsf{D}_{v'}$, which thus approximates $\pi^n_{v'}$.

(b) Training step (Lines 8-9). In order to avoid heavy computation, we approximate the *mean-matching* loss (31) by an unbiased estimator obtained by subsampling $b$ elements from the *full* trajectories computed in the sampling process, see (De Bortoli et al., 2021, Eq. (97)-(98)). Here, $b$ refers to the *batch-size* parameter of the neural networks. Then, we perform gradient descent to optimize the parameter $\theta_{v',v}$, which parameterizes the *backward* drift on the edge $(v, v')$.

To avoid any bias issue, the whole trajectories obtained at process (a) are refreshed at a certain frequency over the training iterations of the neural networks by once again simulating the diffusion (30). In our experiments, this refresh occurs each 500 iterations.

**Setting of the time discretization.** The number of time-steps $N$ in the time discretization of the diffusions is chosen to be even and identical for each of the edges of the tree. Let $\{v, v'\} \in \mathsf{E}$. We now give details on the design of the time schedule $\{\gamma_k\}_{k=1}^N$ related to the edge $\{v, v'\}$, see Appendix F. Following De Bortoli et al. (2021), we choose this sequence to be invariant by time reversal and consider $\gamma_k = \gamma_0 + (2k/N)(\bar{\gamma} - \gamma_0)$ for any $k \in \{0, \ldots, N/2\}$ (the rest of the sequence being obtained by symmetry) where $\gamma_0$ is a free parameter and $\bar{\gamma}$ is determined by $\sum_{k=1}^N \gamma_k = T_{v,v'}$. In our experiments, we set $N = 50$ and $\gamma_0 = 10^{-5}$.

**Sampling improvement.** In our code, we implemented the corrector scheme of Song et al. (2021) and the *probability flow*-based sampling approach detailed in (De Bortoli et al., 2021, Section H.3), but did not observe any significant improvement in our experiments using one of these techniques.

**Choice of the architectures of the neural networks.** In the case of the experiments related to synthetic datasets (two-dimensional toy datasets, Gaussian distributions) and to the subset posterior aggregation task, we implement the same architecture as presented in (De Bortoli et al., 2021, Figure 3). We refer to this model as "Basic Model" and detail it in Figure 6. In the "Basic Model", the PositionalEncoding block applies the sine transform described in Vaswani et al. (2017), with output dimension equal to 32, and each MLP Block represents a Multilayer Perceptron Network. In particular, MLPBlock (1a) has shape $(d, 128, \max(256, 2d))$, MLPBlock (1b) has shape $(32, 128, \max(256, 2d))$, and MLPBlock (2) has shape $(2 \times \max(256, 2d), \max(256, 2d), \max(128, d), d)$, where $d$ denotes the dimension of input data. We optimize the networks with ADAM (Kingma & Ba, 2014) with learning rate $10^{-4}$ and momentum 0.9. For each of the networks, we set the batch size to 4,096 and the number of iterations to 10,000 for the synthetic datasets and 15,000 for the subset posterior aggregation task. Our experiments ran on 1 Intel Xeon CPU Gold 6230 20 cores @ 2.1 Ghz CPU.

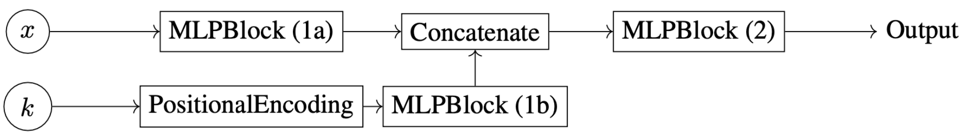


Figure 6: Architecture of the "Basic Model".

In the case of the experiments related to MNIST dataset, we use a reduced UNET architecture based on Nichol & Dhariwal (2021), where we set the number of channels to 64 rather than 128. We implement an exponential moving average of network parameters across training iterations, with rate 0.999. We optimize the networks with ADAM (Kingma & Ba, 2014) with learning rate $10^{-4}$ and momentum 0.9. Finally, we set the batch size to 256 and the number of training iterations to 30,000. Our experiments ran using 1 Nvidia A100.

**Details on regularized state-of the art methods.** We run the fsWB algorithm (Cuturi & Doucet, 2014) with the implementation provided by Flamary et al. (2021). For each experiment, we run 100 Sinkhorn iterations with 1500 samples for each dataset (*i.e.*, the maximum number of samples that it can generate) and set the regularization parameter $\varepsilon$ to its lowest value such that the algorithm is stable. Finally, for sake of fairness with our method, we initialise the barycenter measure with $\pi_r^0$ when solving the problem ($\mu_0$-regWB) for synthetic Gaussian datasets and Bayesian fusion. To run the crWB algorithm (Li et al., 2020), we use the code provided by the authors. We consider the quadratic regularization, which is shown to be empirically more stable than entropic regularization. Following Fan et al. (2020), we choose the potential networks to be fully connected neural networks with 3 hidden layers of shape $(\max(128, 2d), \max(128, 2d), \max(128, 2d))$. The activation functions are ReLu. We optimize the networks with ADAM (Kingma & Ba, 2014) with learning rate $10^{-4}$ for the subset posterior aggregation task and $10^{-3}$ for the Gaussian experiment. Finally, we set the batch size to 4,096 and the number of training iterations to 50,000. We highlight that fsWB and crWB solve a regularized Wasserstein barycenter problem, which does not contain an additional *penalization* term on the entropy of the barycenter, contrary to TreeDSB.

### G.2 Details on the experiments

**Synthetic Gaussian datasets.** For each dimension that we consider, we generate three different triplets of random non-diagonal covariance matrices whose condition number is less than 10. We then run the algorithms on each triplet and aggregate the obtained results. The Gaussian datasets contain 1,500 samples for fsWB, and 10,000 samples for crWB and TreeDSB. We run fsWB with the following settings $(d, \varepsilon) \in \{(2, 0.1), (16, 0.2), (64, 0.5), (128, 1.0), (256, 2.0)\}$. We run TreeDSB for 10 mIPF cycles with regularization parameter $\varepsilon = 0.1$, starting from the central node initialized to a Gaussian distribution $\mu_0$ chosen as detailed in Appendix F with $\alpha = 1$. Thus, we solve the regularized Wasserstein barycenter problem ($\mu_0$-regWB), which contains an additional regularization with respect to $\mu_0$. This choice is justified, since the non-regularized barycenter is known to be a Gaussian distribution, and $\mu_0$ can be seen as an *a priori* for the regularized barycenter. For each of the three settings, we keep the best result among the 30 mIPF iterations. In this setting, TreeDSB and crWB have roughly the same training time.

**Subset posterior aggregation.** When considering a dataset splitted into several subdatasets, a common paradigm in bayesian inference consists in running Monte Carlo Markov Chain methods separately on these subdatasets, and then merge the obtained posteriors to recover the full posterior. The barycenter of these subdataset posteriors is proved to be close to the full data posterior under mild assumptions (Srivastava et al., 2018). In our setting, we consider the posterior aggregation problem for the logistic regression model associated to the `wine` dataset[7] (d = 42) with 3 subdatasets. We consider here two splitting methods: (i) either, data is uniformly splitted between 3 subdatasets with respect to the label distribution, denoted by `wine-homogeneous`, or (ii) data is splitted with some heterogeneity according to a Dirichlet distribution whose parameter is randomly chosen, denoted by `wine-heterogeneous`. Following Korotin et al. (2021), we use the stochastic approximation trick so that the subset posterior samples do not vary consistently from the full posterior in covariance (Minsker et al., 2014). We implement the Unadjusted Langevin Algorithm (ULA) to sample from each subdataset posterior and from the full posterior. In each case, we run ULA for $5.5 \cdot 10^6$ iterations with a well chosen step-size, and obtain 9,900 samples after applying a *burn-in* of order $10\%$ and then a *thinning* of size 500. We provide in Figure 7 some metrics which assess the quality of this sampling process. We recall that the the full posterior samples serve as ground truth in this experiment.

The results presented in Table 2 were computed as follows. For fsWB, we first subsample 1,500 samples out of the 9,900 samples from each posterior, and then run the algorithm with $\varepsilon = 0.5$. We repeat three times this procedure and then aggregate the results. In the case of crWB and TreeDSB, we run the algorithms three times with various seeds. Similarly to the Gaussian setting, we run TreeDSB for 10 mIPF cycles with regularization parameter $\varepsilon = 0.1$. We start from the central node with a Gaussian distribution $\mu_0$ chosen as detailed in Appendix F with $\alpha = 1$, and thus solve the barycenter formulation ($\mu_0$-regWB). For each of the three settings, we keep the best result among the 30 IPF iterations. In this setting, TreeDSB and crWB have roughly the same training time.

[7] `https://archive.ics.uci.edu/ml/datasets/wine`

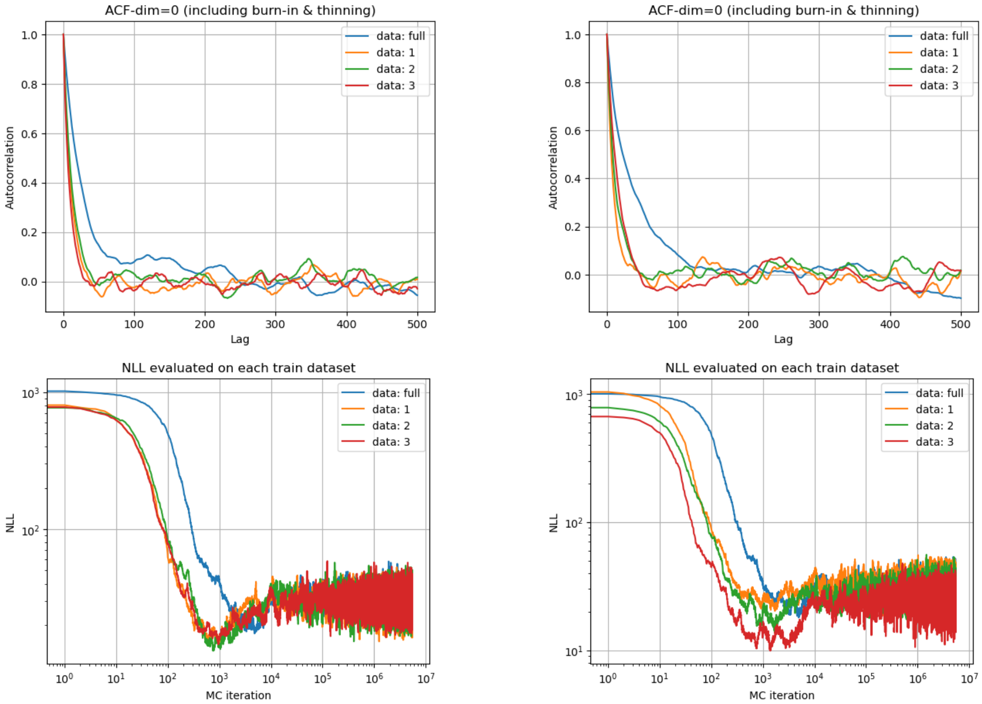


Figure 7: Evaluation of the sampling process for `wine-homogeneous` (left) and `wine-heterogeneous` (right). We display the Autocorrelation function on 500 lags (above) and the evolution over the iterations of ULA of the negative log-likelihood (NLL) evaluated on each training dataset (below). In particular, the samples are decorrelated and the NLL has a satisfying profile.

**Synthetic two-dimensional datasets.** In this setting, we consider three different datasets (*Swiss-roll*, *Circle* and *Moons*) that each contain 10,000 samples. Since we do not have an *a priori* on the shape of the barycenter between these datasets, we consider the regularized Wasserstein barycenter problem (regWB), *i.e.*, $r$ is chosen as a leaf and corresponds to one of the input datasets. We emphasize that this experiment is not intended to demonstrate the superiority of TreeDSB to compute 2D Wasserstein barycenters, but is rather meant to illustrate that (a) the marginals of the leaves are well recovered by the algorithm, see Figure 3, and that (b) the obtained barycenter is consistent when diffusing from the different leaves, see Figure 4. In all our experiments on 2D datasets, we observed that (a) was persistently verified without difficulty. In this section, we rather aim at illustrating (b) by providing additional results which assess the quality of the barycenter obtained by TreeDSB with respect to the choice of the starting leaf $r$ and to the choice of the regularization parameter $\varepsilon$.

To do so, we consider three different choices of regularization in TreeDSB: (i) $\varepsilon = 0.2$ (50 mIPF cycles), see Figure 8, (ii) $\varepsilon = 0.1$ (50 mIPF cycles), see Figure 9 and (iii) $\varepsilon = 0.05$ (60 mIPF cycles), see Figure 10. For each of these settings, we run TreeDSB with the starting leaf $r$ chosen as *Swiss-roll* (first row), *Circle* (second row) or *Moons* (third row), and display the final barycenter obtained by diffusing from *Swiss-roll* (first column), *Circle* (second column) and *Moons* (third column). Note that the vertex 0 always corresponds to the starting leaf, the vertex 1 to the barycenter node and that Figure 4 corresponds to the first row of Figure 9.

We can make the following observations. First, the estimated barycenter is always coherent within each row, which assesses the convergence of our method. Then, for each value of $\varepsilon$, the TreeDSB barycenter is rather consistent between the rows, *i.e.*, the choice of the starting leaf does not have a meaningful impact on our method. Finally, as expected, we observe that the support of the barycenter is less and less diffuse as long as $\varepsilon$ decreases.

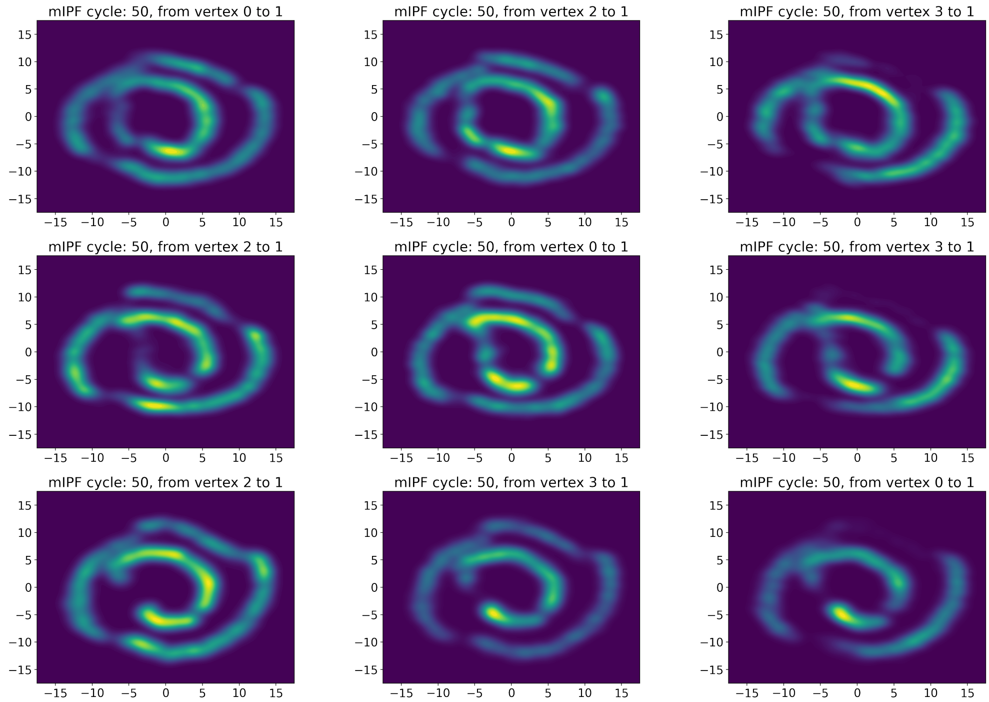


Figure 8: Estimated 2D barycenter obtained by TreeDSB with $\varepsilon = 0.2$ (50 mIPF cycles). First row: starting from *Swiss-roll*. Second row: starting from *Circle*. Third row: starting from *Moons*.

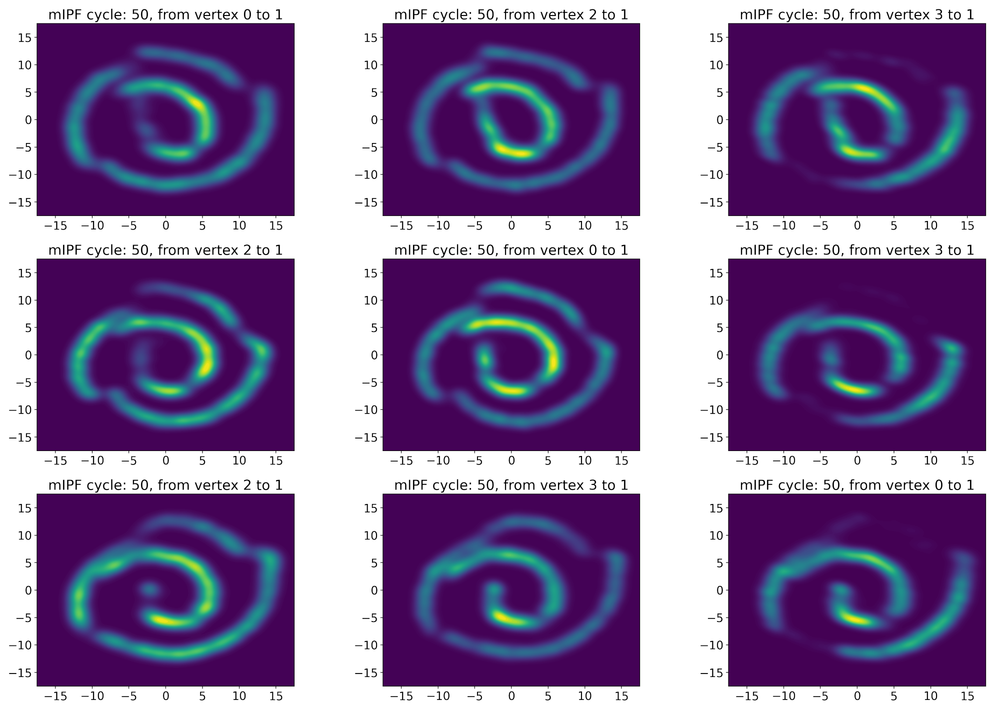


Figure 9: Estimated 2D barycenter obtained by TreeDSB with $\varepsilon = 0.1$ (50 mIPF cycles). First row: starting from *Swiss-roll*. Second row: starting from *Circle*. Third row: starting from *Moons*.

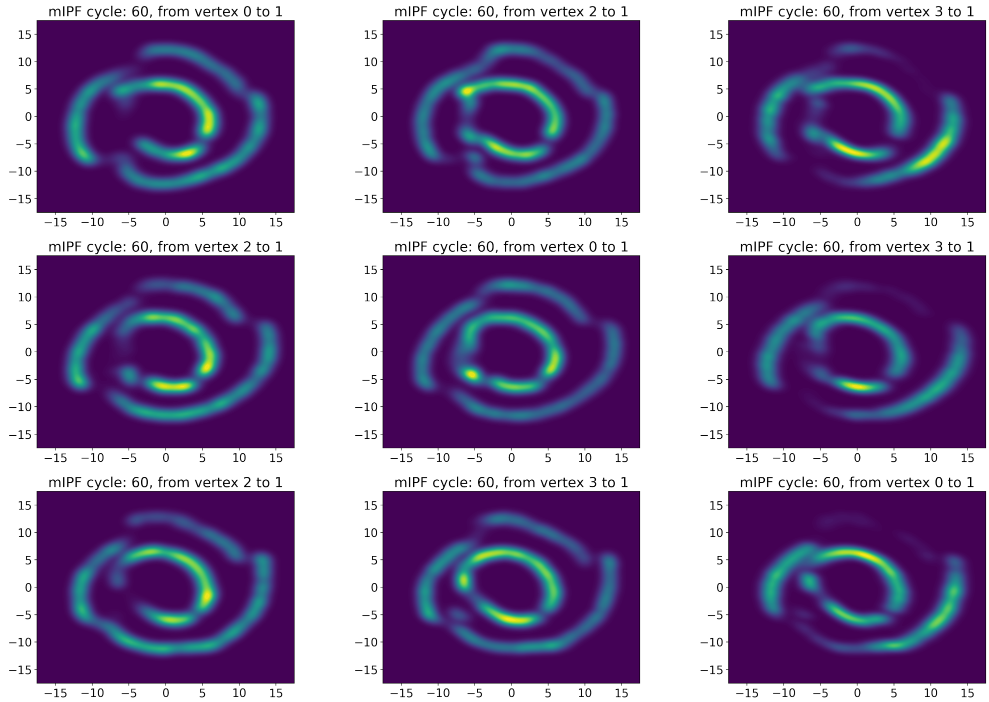


Figure 10: Estimated 2D barycenter obtained by TreeDSB with $\varepsilon = 0.05$ (60 mIPF cycles). First row: starting from *Swiss-roll*. Second row: starting from *Circle*. Third row: starting from *Moons*.

For purpose of illustration, we provide in Figure 11 the barycenter obtained by state-of-the-art two-dimensional *in-sample* methods that are available in POT library (Flamary et al., 2021): (i) non-regularized free-support Wasserstein barycenter (Cuturi & Doucet, 2014), (ii) entropic-regularized free-support Wasserstein barycenter (fsWB) with $\varepsilon = 0.5$ (Cuturi & Doucet, 2014) and (iii) entropic-regularized convolutional Wasserstein barycenter with $\varepsilon = 5.10^{-4}$ (Solomon et al., 2015), which is specifically designed for images. We notably observe that TreeDSB cannot capture the full complexity of the 2D barycenter compared to these methods. We infer that this gap comes from the *dynamic* nature of TreeDSB, since increasing the number of training iterations per IPF iteration or improving the complexity of the neural networks did not bring any significant change in our results. Finally, we recall that the methods (i), (ii) and (iii) do not scale well with dimension, and have to be completely run again when new data samples are available, contrary to TreeDSB.

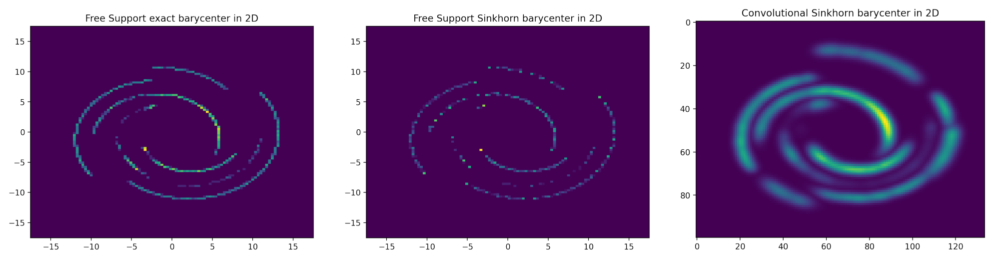


Figure 11: Estimated 2D barycenter obtained by *in-sample* algorithms. From left to right: Cuturi & Doucet (2014) (non-regularized), Cuturi & Doucet (2014) (regularized), Solomon et al. (2015).

**MNIST Wasserstein barycenter.** This setting can be qualified as *high-dimensional*, since the data dimension is $d = 784$. Here, each digit dataset contains 1,000 samples. As in the two-dimensional setting, we do not have an *a priori* on the shape of the barycenter between MNIST digits, and thus consider the formulation (regWB), where the root $r$ is chosen as a leaf. We propose below several experiments to assess the scalability of TreeDSB to this setting.

**Digits 0 and 1.** In Figure 12, we report the results obtained by running TreeDSB on MNIST digits 0 and 1, for 15 mIPF cycles with $\varepsilon = 0.5$, starting from the leaf MNIST-0. We display 25 samples from the reconstructed MNIST-0 marginal (first column), from the reconstructed MNIST-1 marginal (fourth column), from the estimated barycenter by diffusing from MNIST-0 (second column) and diffusing from MNIST-1 (third column). We notably observe that the digits are well recovered and that the barycenter samples are consistent. We draw the reader's attention to the fact that TreeDSB showed numerical unstability with a regularization value $\varepsilon$ lower than 0.5. For purpose of illustration, we display in Figure 13 the Wasserstein barycenter obtained by *non-regularized* methods from Fan et al. (2020) and Korotin et al. (2021), and by the *regularized* approach from Li et al. (2020).

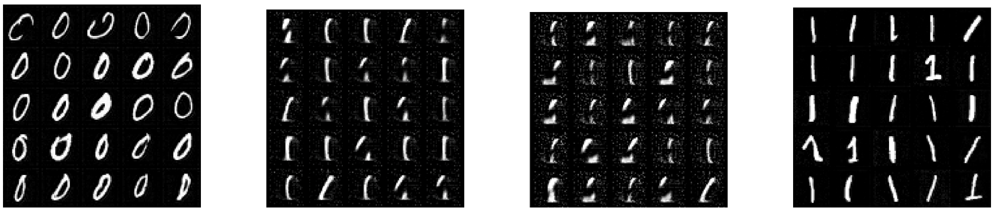

Figure 12: Tree DSB results for MNIST digits 0 and 1, after 15 mIPF cycles with $\varepsilon = 0.5$.

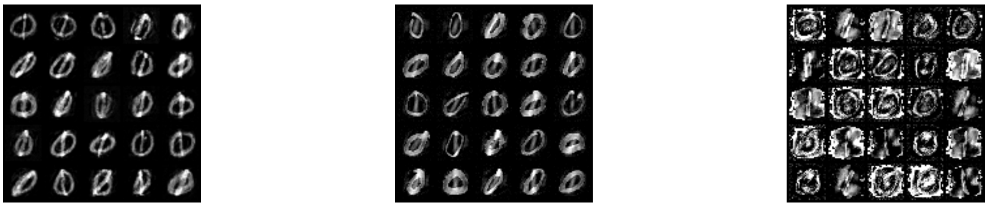

Figure 13: From left to right: Fan et al. (2020), Korotin et al. (2021) and Li et al. (2020).

**Digits 2,4 and 6.** In Figure 14, we report the results obtained by running TreeDSB on MNIST digits 2,4 and 6, for 10 mIPF cycles with $\varepsilon = 0.5$. Here, we consider three settings which differ by the starting leaf $r$ in the algorithm: MNIST-2 (first row), MNIST-4 (second row), or MNIST-6 (third row). For each of these settings, we display 30 samples from the estimated barycenter by diffusing from MNIST-2 (first column), diffusing from MNIST-4 (third column) and diffusing from MNIST-6 (third column). We notably observe a global consistency of the barycenter samples across the various settings. In Figure 15, we report the results obtained by running TreeDSB on MNIST digits 2,4 and 6, for 10 mIPF cycles with $\varepsilon = 0.2$, starting from MNIST-6. We display 30 samples from the reconstructed marginals (first row), from the estimated barycenter (second row) by diffusing from MNIST-2 (first column), diffusing from MNIST-4 (second column) and diffusing from MNIST-6 (third column). As expected, we observe less noisy barycenter samples compared to Figure 14, while still well recovering MNIST digits.

**Digits 0,1 and 4.** In Figure 16, we report the results obtained by running TreeDSB on MNIST digits 0,1 and 4, for 10 mIPF cycles with $\varepsilon = 0.5$. We consider two settings which differ by the starting leaf $r$ in the algorithm: MNIST-0 (second row) and MNIST-1 (first/third rows), for which we display samples from the reconstructed measures (first row). In Figure 17, we report the results obtained by running TreeDSB on MNIST digits 0,1 and 4, for 10 mIPF cycles with $\varepsilon = 0.2$. We consider two settings which differ by the starting leaf $r$ in the algorithm: MNIST-0 (first/second row), for which display samples from the reconstructed measures (first row), and MNIST-1 (third row). For all of these settings, we display 30 samples from the estimated barycenter by diffusing from MNIST-0 (first column), diffusing from MNIST-1 (third column) and diffusing from MNIST-4 (third column). Similarly to the digits 2-4-6, we observe consistency within the barycenter samples, unconditionally to the starting leaf, and less noise as $\varepsilon$ decreases. Note that the reconstructed MNIST digits are less truthful to the original datasets when $\varepsilon$ is low.

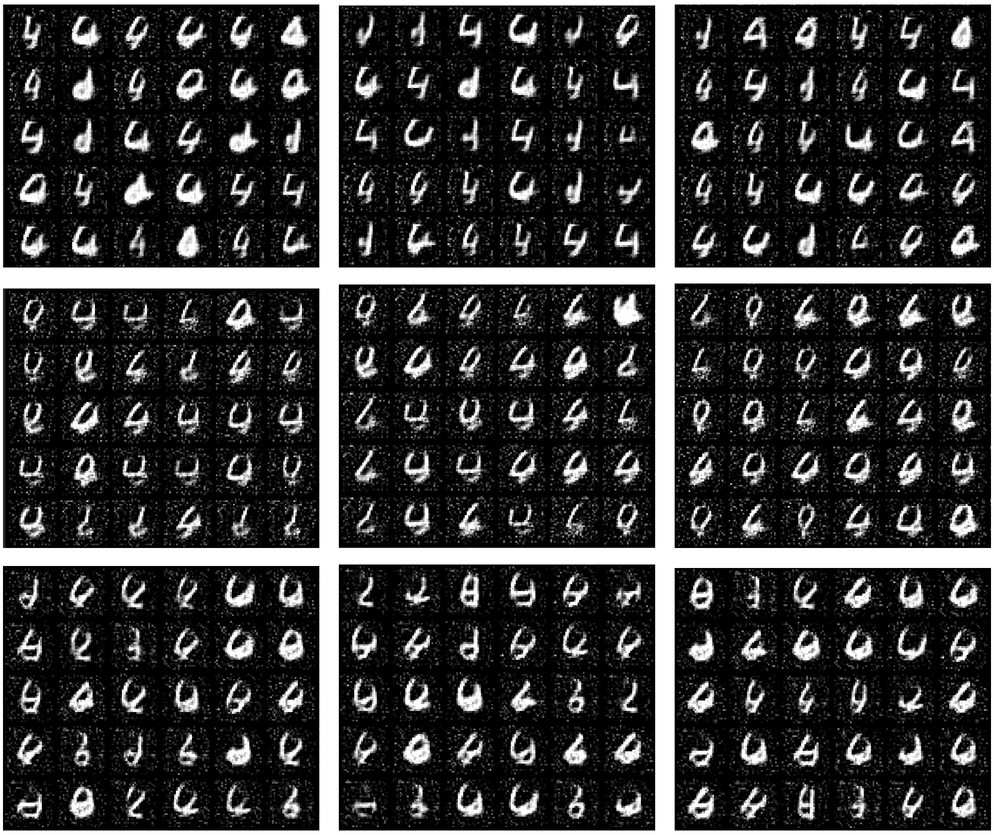

Figure 14: Tree DSB results for MNIST digits 2,4 and 6, after 10 mIPF cycles with $\varepsilon = 0.5$. First row: starting from MNIST-2. Second row: starting from MNIST-4. Third row: starting from MNIST-6.

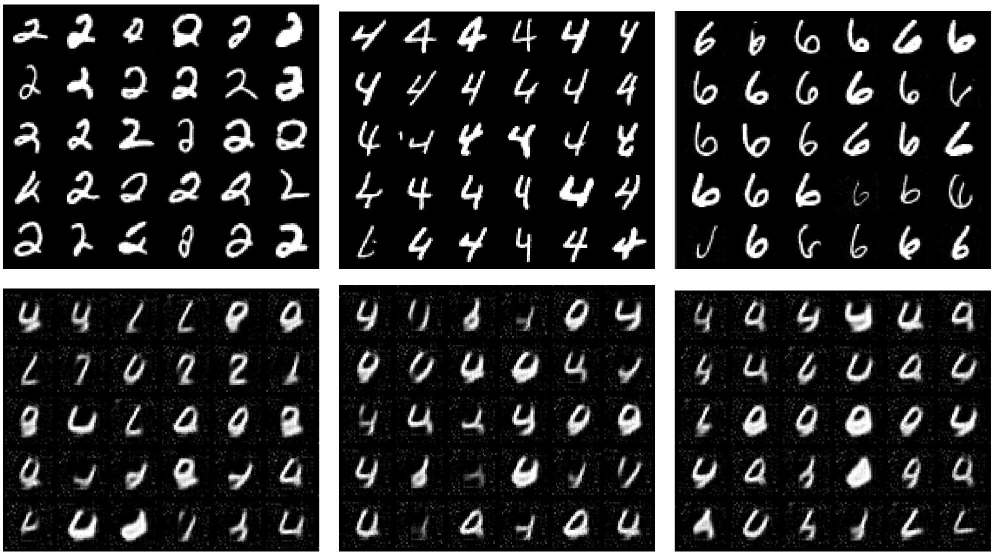

Figure 15: Tree DSB results for MNIST digits 2,4 and 6, after 10 mIPF cycles with $\varepsilon = 0.2$, starting from MNIST-6. First row: samples from the reconstructed marginals. Second row: samples from the estimated barycenter.

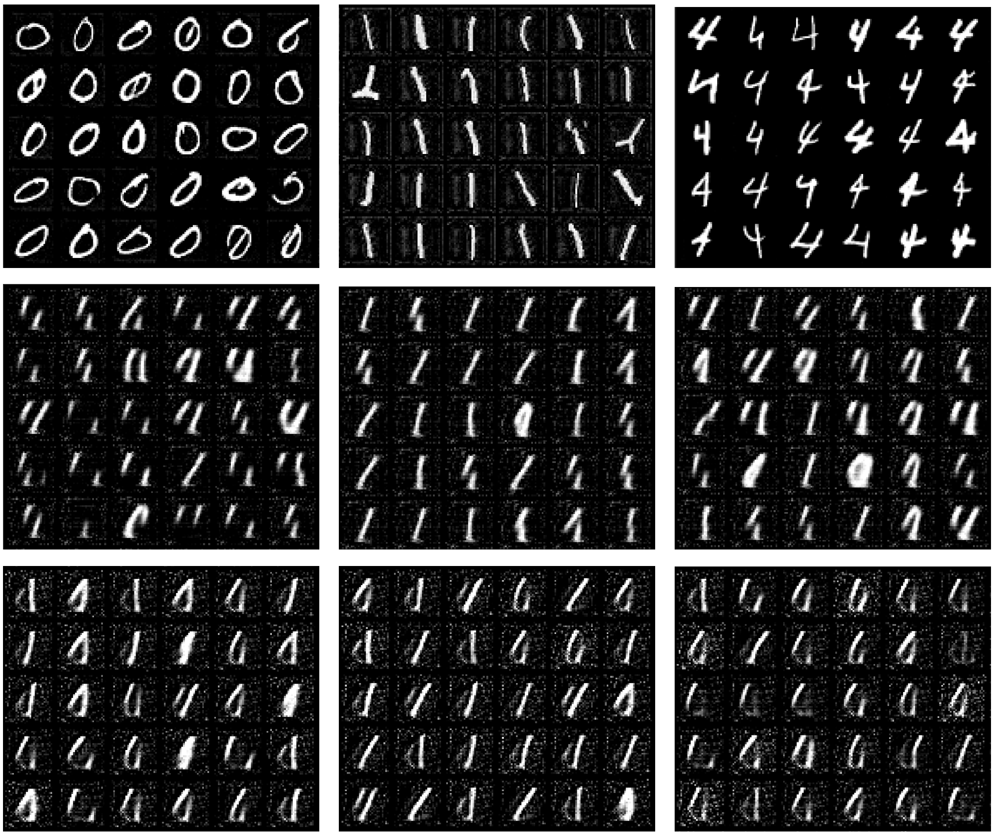

Figure 16: Tree DSB results for MNIST digits 0,1 and 4, after 10 mIPF cycles with $\varepsilon = 0.5$. First row: samples from the reconstructed marginals, starting from MNIST-1. Second row: samples from the estimated barycenter, starting from MNIST-0. Third row: samples from the estimated barycenter, starting from MNIST-1.

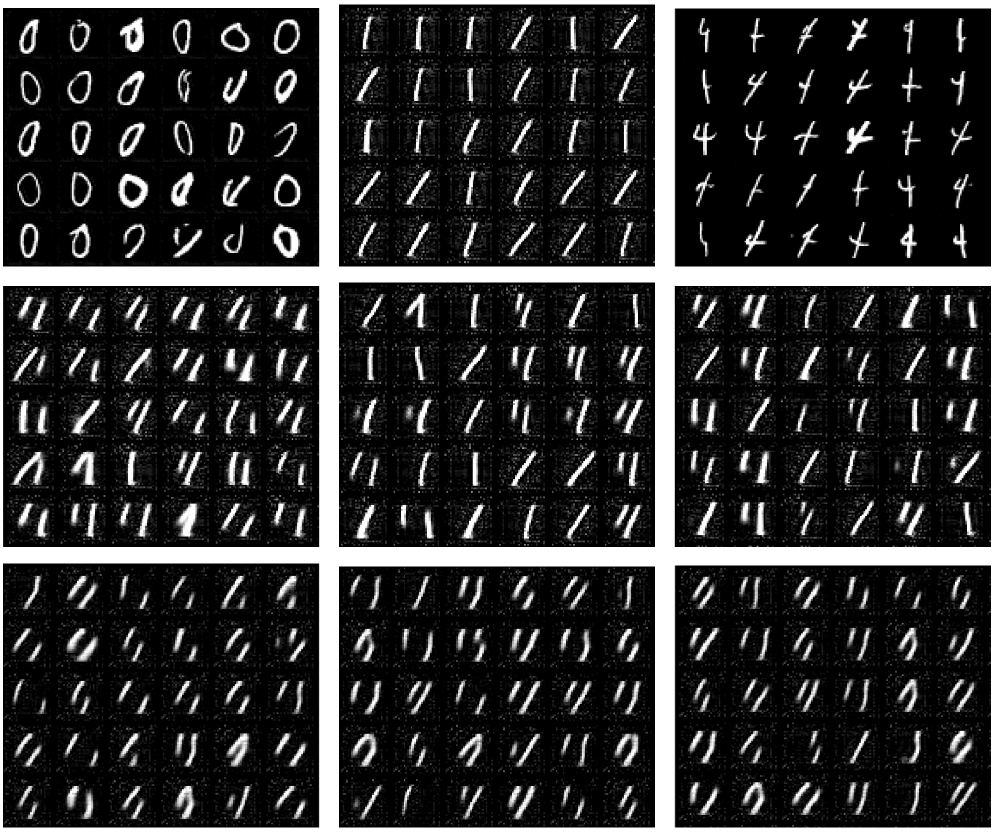

Figure 17: Tree DSB results for MNIST digits 0,1 and 4, after 10 mIPF cycles with $\varepsilon = 0.2$. First row: samples from the reconstructed marginals, starting from MNIST-0. Second row: samples from the estimated barycenter, starting from MNIST-0. Third row: samples from the estimated barycenter, starting from MNIST-1.