Tree-Based Diffusion Schrödinger Bridge with Applications to Wasserstein Barycenters

Maxence Noble[1] Valentin de Bortoli[2]
Arnaud Doucet[3] Alain Durmus[1]

[1]Centre de Mathématiques Appliquées, Ecole Polytechnique
Institut Polytechnique de Paris, France

[2]Département d'Informatique, École Normale Supérieure
CNRS, Université PSL, Paris, France

[3]Department of Statistics, University of Oxford, UK

Outline

Outline

Motivations of Optimal transport

Probability distributions are **everywhere** in machine learning.

Figure: From left to right: Bayesian posterior distribution (supported on $\mathbb{R}^d$), MNIST (supported on $[0,1]^{28\times 28}$) and CELEBA (supported on $[0,1]^{3\times 64\times 64}$).

Motivations of Optimal transport

Probability distributions are **everywhere** in machine learning.

Figure: From left to right: Bayesian posterior distribution (supported on $\mathbb{R}^d$), MNIST (supported on $[0,1]^{28\times 28}$) and CELEBA (supported on $[0,1]^{3\times 64\times 64}$).

- How to compare distributions ?
- How to evaluate similarity between distributions ?
- How to define a proper geometry in the space of distributions ?

Optimal transport (OT) provides tools to answer this question!

Formulation from Kantorovich (1942)

Define $\mathscr{P}^{(2)}$ as the set of probability measures defined on $\mathbb{R}^d \times \mathbb{R}^d$.

Given a **cost** function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$, we aim at solving

$$\pi^\star = \arg\min \left\{ \int c(x_0, x_1) \mathrm{d}\pi(x_0, x_1) : \pi \in \mathscr{P}^{(2)},\ \pi_0 = \mu_0,\ \pi_1 = \mu_1 \right\}$$

Formulation from Kantorovich (1942)

Define $\mathscr{P}^{(2)}$ as the set of probability measures defined on $\mathbb{R}^d \times \mathbb{R}^d$.

Given a **cost** function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$, we aim at solving

$$\pi^\star = \arg\min \left\{ \int c(x_0, x_1) \mathrm{d}\pi(x_0, x_1) : \pi \in \mathscr{P}^{(2)},\ \pi_0 = \mu_0,\ \pi_1 = \mu_1 \right\}$$

- With a **quadratic cost**:

  we obtain the **Wasserstein-2** distance between μ_0 and μ_1

$$W_2(\mu_0, \mu_1) = \inf \left\{ \int \|x_0 - x_1\|^2 \mathrm{d}\pi(x_0, x_1) : \pi \in \mathscr{P}^{(2)},\ \pi_0 = \mu_0,\ \pi_1 = \mu_1 \right\}^{1/2}$$

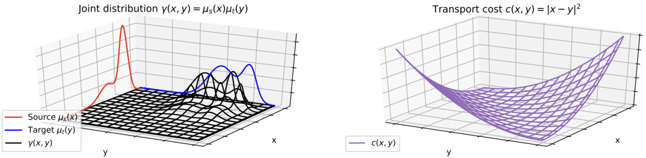


Figure: Illustration from the slides of Rémi Flamary.

Extension to the multimarginal setting

Define $\mathscr{P}^{(\ell+1)}$ as the set of probability measures defined on $(\mathbb{R}^d)^{\ell+1}$.

Given a **cost** function $c : (\mathbb{R}^d)^{\ell+1} \to \mathbb{R}$, a **subset** $\mathsf{S} \subset \{0, \ldots, \ell\}$ and a **family of probability measures** $\{\mu_i\}_{i \in \mathsf{S}} \in (\mathscr{P}(\mathbb{R}^d))^{|\mathsf{S}|}$, we consider the **mOT** problem

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) : \pi \in \mathscr{P}^{(\ell+1)}, \; \pi_i = \mu_i \,, \forall i \in \mathsf{S} \right\}$$

Extension to the multimarginal setting

Define $\mathscr{P}^{(\ell+1)}$ as the set of probability measures defined on $(\mathbb{R}^d)^{\ell+1}$.

Given a **cost function** $c : (\mathbb{R}^d)^{\ell+1} \to \mathbb{R}$, a **subset** $\mathsf{S} \subset \{0, \dots, \ell\}$ and a **family of probability measures** $\{\mu_i\}_{i \in \mathsf{S}} \in (\mathscr{P}(\mathbb{R}^d))^{|\mathsf{S}|}$, we consider the **mOT** problem

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) : \pi \in \mathscr{P}^{(\ell+1)}, \ \pi_i = \mu_i \ , \forall i \in \mathsf{S} \right\}$$

- If $c(x_{0:\ell}) = \sum_{i=1}^{\ell} w_i \|x_0 - x_i\|^2$ with $\{w_i\} \in (\mathbb{R}_+)^{|\mathsf{S}|}$ and $\mathsf{S} = \{1, \dots, \ell\}$:

$$\pi_0^\star = \arg\min \left\{ \sum_{i=1}^{\ell} w_i W_2^2(\nu, \mu_i) : \nu \in \mathscr{P}(\mathbb{R}^d) \right\} ,$$

mOT defines the **Wasserstein barycenter** (Peyré et al., 2019) of the $\{\mu_i\}$.

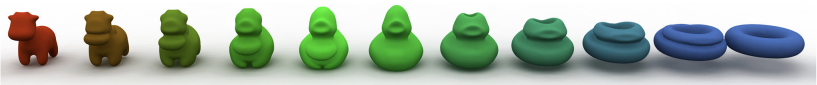


Figure: Illustration from Solomon et al. (2015).

Tree-based OT

Consider an **undirected tree** (*connected acyclic graph*) $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with vertices V (identified with $\{0, \ldots, \ell\}$), edges E and edge weights $\{w_{v,v'}\} \in (\mathbb{R}_+)^{|\mathsf{E}|}$.

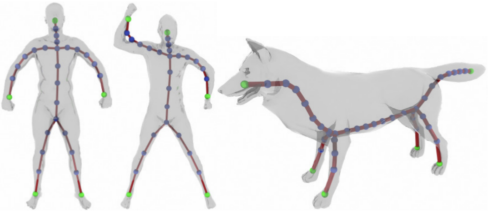


Figure: Illustration from Solomon et al. (2015).

Tree-based OT

Consider an **undirected tree** (*connected acyclic graph*) $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with vertices V (identified with $\{0, \dots, \ell\}$), edges E and edge weights $\{w_{v,v'}\} \in (\mathbb{R}_+)^{|\mathsf{E}|}$.

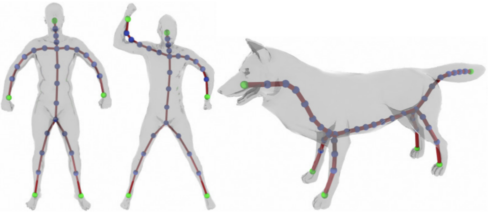


Figure: Illustration from Solomon et al. (2015).

By defining a **quadratic tree-based cost** (Haasler et al., 2021)

$$c(x_{0:\ell}) = \sum_{\{v,v'\}\in\mathsf{E}} w_{v,v'} \|x_v - x_{v'}\|^2 \ ,$$

mOT recovers the **Wasserstein propagation** problem (Solomon et al., 2014)

$$\arg\min \left\{ \sum_{\{v,v'\}\in\mathsf{E}} w_{v,v'} W_2^2(\nu_v, \nu_{v'}) : \{\nu_v\} \in (\mathscr{P}(\mathbb{R}^d))^{|\mathsf{V}|}, \nu_v = \mu_v, \forall v \in \mathsf{S} \right\} \ .$$

It reduces to a Wasserstein barycenter problem when T is a star-shaped tree !

Entropy-regularized OT (EOT)

Solving OT problems faces **computational challenges** in practice (Pele and Werman, 2009), which motivates to consider an **entropic regularization** of OT.

In the **multimarginal** setting, we now aim to solve the **EmOT** problem

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) + \varepsilon \mathrm{KL}(\pi|\nu) : \pi \in \mathscr{P}^{(\ell+1)},\ \pi_i = \mu_i\ , \forall i \in \mathsf{S} \right\}$$

- $\varepsilon > 0$: regularization hyperparameter.
- $\nu \in \mathscr{P}^{(\ell+1)}$: regularization probability measure.
- $\mathrm{KL}(\pi|\nu)$[1]: Kullback-Leibler divergence between π and ν.

[1] $\mathrm{KL}(\pi|\nu) = \int \log(\mathrm{d}\pi/\mathrm{d}\nu)\mathrm{d}\pi$ if $\pi \ll \nu$, $\mathrm{KL}(\pi|\nu) = \infty$ otherwise.

Entropy-regularized OT (EOT)

Solving OT problems faces **computational challenges** in practice (Pele and Werman, 2009), which motivates to consider an **entropic regularization** of OT.

In the **multimarginal** setting, we now aim to solve the **EmOT** problem

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) + \varepsilon \mathrm{KL}(\pi|\nu) : \pi \in \mathscr{P}^{(\ell+1)},\ \pi_i = \mu_i\ , \forall i \in \mathsf{S} \right\}$$

- $\varepsilon > 0$: regularization hyperparameter.
- $\nu \in \mathscr{P}^{(\ell+1)}$: regularization probability measure.
- $\mathrm{KL}(\pi|\nu)$[1]: Kullback-Leibler divergence between π and ν.

The **discrete** state-space counterpart of **EmOT** can be efficiently solved with **Sinkhorn algorithm** (Cuturi, 2013; Knight, 2008; Sinkhorn and Knopp, 1967).

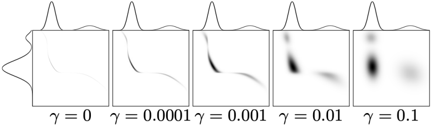


Figure: Illustration from Solomon et al. (2015).

[1] $\mathrm{KL}(\pi|\nu) = \int \log(\mathrm{d}\pi/\mathrm{d}\nu)\mathrm{d}\pi$ if $\pi \ll \nu$, $\mathrm{KL}(\pi|\nu) = \infty$ otherwise.

Schrödinger Bridge Problem

[2] Stochastic Differential Equation

Schrödinger Bridge Problem

Given a **time horizon** $T > 0$, a **reference path measure** $\mathbb{Q}$ and $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$, the ***dynamic* Schrödinger Bridge (SB)** problem amounts to find

$$\mathbb{P}^\star = \mathrm{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{Q}) \ : \ \mathbb{P} \in \mathscr{P}(\mathrm{C}([0,T], \mathbb{R}^d)), \ \mathbb{P}_0 = \mu_0, \ \mathbb{P}_T = \mu_1\}$$

[2]Stochastic Differential Equation

Schrödinger Bridge Problem

Given a **time horizon** $T > 0$, a **reference path measure** $\mathbb{Q}$ and $\mu_0, \mu_1 \in \mathscr{P}(\mathbb{R}^d)$, the ***dynamic* Schrödinger Bridge (SB)** problem amounts to find

$$\mathbb{P}^\star = \operatorname{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{Q}) \,:\, \mathbb{P} \in \mathscr{P}(\mathrm{C}([0,T], \mathbb{R}^d)),\ \mathbb{P}_0 = \mu_0,\ \mathbb{P}_T = \mu_1\}$$

If $\mathbb{Q}$ is associated with the SDE[2] $\mathrm{d}\mathbf{X}_t = -a\mathbf{X}_t\mathrm{d}t + \mathrm{d}\mathbf{B}_t$, with $a \geq 0$, Léonard (2014) states that $\mathbb{P}^\star_{0,T}$ solves the ***static*-SB** problem

$$\operatorname{argmin}\{\mathrm{KL}(\pi|\mathbb{Q}_{0,T}) \,:\, \pi \in \mathscr{P}^{(2)},\ \pi_0 = \mu_0,\ \pi_1 = \mu_1\},$$

and we have

$$\text{static-SB} \iff \text{EOT with } \varepsilon = 2\sinh(aT)/a \text{ if } a > 0 \text{ or } \varepsilon = 2T \text{ if } a = 0\,.$$

Moreover, we have

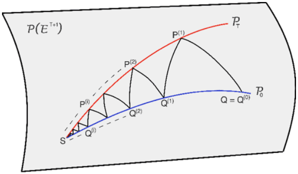
$$\text{static-SB} \iff \text{SB}\,, \text{ since } \mathbb{P}^\star = \mathbb{P}^\star_{0,T} \otimes \mathbb{Q}_{|0,T}\,.$$

[2] Stochastic Differential Equation

Iterative Proportional Fitting (IPF) procedure

The **IPF** procedure[3] (Sinkhorn and Knopp, 1967; Knight, 2008; Peyré et al., 2019; Cuturi and Doucet, 2014) aims at solving **SB** with the iterates $(\mathbb{P}^n)_{n\in\mathbb{N}}$ defined by $\mathbb{P}^0 = \mathbb{Q}$ and for any $n \in \mathbb{N}$

$$\begin{aligned}
\mathbb{P}^{2n+1} &= \operatorname{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n}) \; : \; \mathbb{P}_T = \mu_1\}, \\
&= \mu_1 \otimes (\mathbb{P}^{2n})^R_{|0} \; (\textit{backward}) \; , \\
\mathbb{P}^{2n+2} &= \operatorname{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n+1}) \; : \; \mathbb{P}_0 = \mu_0\} \\
&= \mu_0 \otimes \mathbb{P}^{2n+1}{}_{|0} \; (\textit{forward}) \; ,
\end{aligned}$$



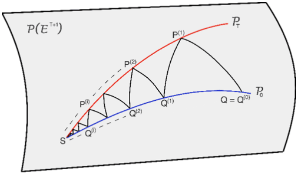
where R is the **time-reversal** operator.

[3] Note that it is the continuous state-space counterpart of Sinkhorn algorithm.

Iterative Proportional Fitting (IPF) procedure

The **IPF** procedure[3] (Sinkhorn and Knopp, 1967; Knight, 2008; Peyré et al., 2019; Cuturi and Doucet, 2014) aims at solving **SB** with the iterates $(\mathbb{P}^n)_{n\in\mathbb{N}}$ defined by $\mathbb{P}^0 = \mathbb{Q}$ and for any $n \in \mathbb{N}$

$$
\begin{aligned}
\mathbb{P}^{2n+1} &= \operatorname{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n}) \;:\; \mathbb{P}_T = \mu_1\}, \\
&= \mu_1 \otimes (\mathbb{P}^{2n})^R_{|0} \; (\textit{backward}) \,, \\
\mathbb{P}^{2n+2} &= \operatorname{argmin}\{\mathrm{KL}(\mathbb{P}|\mathbb{P}^{2n+1}) \;:\; \mathbb{P}_0 = \mu_0\} \\
&= \mu_0 \otimes \mathbb{P}^{2n+1}{}_{|0} \; (\textit{forward}) \,,
\end{aligned}
$$



where R is the **time-reversal** operator.

If $\mathbb{P}$ is associated with $\mathrm{d}\mathbf{X}_t = f_t(\mathbf{X}_t)\mathrm{d}t + \mathrm{d}\mathbf{B}_t$, then under mild assumptions (Haussmann and Pardoux, 1986; Cattiaux et al., 2021), $\mathbb{P}^R$ is associated with

$$
\mathrm{d}\mathbf{Y}_t = \{-f_{T-t}(\mathbf{Y}_t) + \nabla \log p_{T-t}(\mathbf{Y}_t)\}\mathrm{d}t + \mathrm{d}\mathbf{B}_t,
$$

where p_t is the density of $\mathbb{P}_t$ w.r.t. the Lebesgue measure.

[3] Note that it is the continuous state-space counterpart of Sinkhorn algorithm.

Diffusion Schrödinger Bridge (DSB)

De Bortoli et al. (2021) propose a numerical scheme[4], **Diffusion Schrödinger Bridge**, to approximate the IPF iterates by implementing

- an Euler-Maruyama **time discretization** of the forward/backward SDEs,
- an approximation of the scores via 2 **neural networks** (forward/backward).

[4]This algorithm shows great performance for small values of T.

Diffusion Schrödinger Bridge (DSB)

De Bortoli et al. (2021) propose a numerical scheme[4], **Diffusion Schrödinger Bridge**, to approximate the IPF iterates by implementing

- an Euler-Maruyama **time discretization** of the forward/backward SDEs,
- an approximation of the scores via 2 **neural networks** (forward/backward).

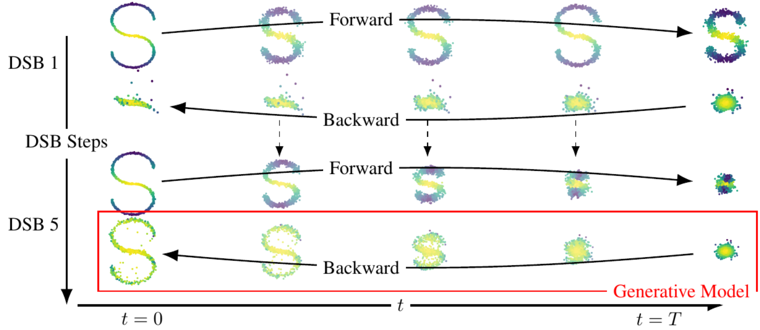


Figure: Illustration from De Bortoli et al. (2021).

[4] This algorithm shows great performance for small values of T.

Extension of EmOT to SB formulation

We recall the regularized OT formulation in the multimarginal setting (**EmOT**)

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) + \varepsilon \mathrm{KL}(\pi|\nu) : \pi \in \mathscr{P}^{(\ell+1)},\ \pi_i = \mu_i \ , \forall i \in \mathsf{S} \right\} .$$

[5] Note that π^0 may not be a probability measure.

Extension of EmOT to SB formulation

We recall the regularized OT formulation in the multimarginal setting (**EmOT**)

$$\pi^\star = \arg\min \left\{ \int c(x_{0:\ell}) \mathrm{d}\pi(x_{0:\ell}) + \varepsilon \mathrm{KL}(\pi|\nu) : \pi \in \mathscr{P}^{(\ell+1)},\ \pi_i = \mu_i\ , \forall i \in \mathsf{S} \right\} .$$

If $\nu \ll \mathrm{Leb}$, then **EmOT** can be rewritten in a **static-SB** fashion, called **mSB**

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

with $(\mathrm{d}\pi^0/\mathrm{dLeb})(x_{0:\ell}) \propto \exp[-c(x_{0:\ell})/\varepsilon](\mathrm{d}\nu/\mathrm{dLeb})(x_{0:\ell})$[5].

Similarly to the bimarginal setting, we obtain

EmOT param. by c, ε and ν $\iff$ **mSB** param. by π^0 .

[5]Note that π^0 may not be a probability measure.

Formulation of Tree-based SB

Consider an **undirected tree** $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with $\mathsf{V} \equiv \{0, \ldots, \ell\}$ and assume that

- S is the set of the leaves of T,
- $c(x_{0:\ell}) = \sum_{\{v,v'\} \in \mathsf{E}} w_{v,v'} \|x_v - x_{v'}\|_2^2$,
- $\mathrm{d}\nu / \mathrm{dLeb}(x_{0:\ell}) = \varphi_r(x_r)$ for some $r \in \mathsf{V}$.

Formulation of Tree-based SB

Consider an **undirected tree** $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with $\mathsf{V} \equiv \{0, \dots, \ell\}$ and assume that

- S is the set of the leaves of T,
- $c(x_{0:\ell}) = \sum_{\{v,v'\} \in \mathsf{E}} w_{v,v'} \|x_v - x_{v'}\|_2^2$,
- $\mathrm{d}\nu / \mathrm{dLeb}(x_{0:\ell}) = \varphi_r(x_r)$ for some $r \in \mathsf{V}$.

Then, $\pi^0 \in \mathscr{P}(\mathbb{R}^d)$ has a **Markovian** factorization along $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$, the directed version of T rooted in r,

$$\pi^0 = \pi_r^0 \bigotimes\nolimits_{(v,v') \in \mathsf{E}_r} \pi_{v'|v}^0 ,$$

where $\pi_{v'|v}^0(\cdot \mid x_v) = \mathrm{N}(x_v, \varepsilon / (2 w_{v,v'}) \mathrm{I}_d)$ and $\pi_r^0 \ll \mathrm{Leb}$ with density φ_r.

Formulation of Tree-based SB

Consider an **undirected tree** $\mathsf{T} = (\mathsf{V}, \mathsf{E})$ with $\mathsf{V} \equiv \{0, \ldots, \ell\}$ and assume that

- S is the set of the leaves of T,
- $c(x_{0:\ell}) = \sum_{\{v,v'\} \in \mathsf{E}} w_{v,v'} \|x_v - x_{v'}\|_2^2$,
- $\mathrm{d}\nu / \mathrm{dLeb}(x_{0:\ell}) = \varphi_r(x_r)$ for some $r \in \mathsf{V}$.

Then, $\pi^0 \in \mathscr{P}(\mathbb{R}^d)$ has a **Markovian** factorization along $\mathsf{T}_r = (\mathsf{V}, \mathsf{E}_r)$, the directed version of T rooted in r,

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

where $\pi^0_{v'|v}(\cdot \mid x_v) = \mathrm{N}(x_v, \varepsilon/(2 w_{v,v'}) \mathrm{I}_d)$ and $\pi_r^0 \ll \mathrm{Leb}$ with density φ_r.

We finally obtain a **tree-based** formulation of the **mSB** problem (**TreeSB**)

$$\pi^\star = \mathrm{argmin}\{\mathrm{KL}(\pi|\pi^0) \, : \, \pi \in \mathscr{P}^{(|\mathsf{V}|)}, \pi_i = \mu_i, \forall i \in \mathsf{S}\}$$
$$\text{with } \pi^0 \text{ param. by } r \text{ and } \varphi_r,$$

and propose to solve it with our algorithm **Tree-Based Diffusion Schrödinger Bridge (TreeDSB)**, which is the **natural extension of DSB**.

Outline

Before starting

Some notation:

- $\mathsf{S} = \{i_0, \ldots, i_{K-1}\}$, and thus $|\mathsf{S}| = K$.
- $k_n = (n-1) \bmod(K)$, $k_n + 1 = n \bmod(K)$.
- $T_{v,v'} = \varepsilon/(2w_{v,v'})$ for any $\{v, v'\} \in \mathsf{E}$.
- $\mathrm{Ext}(\mathbb{P}) = \mathbb{P}_{0,T} \in \mathscr{P}^{(2)}$ for any path measure $\mathbb{P} \in \mathscr{P}(\mathrm{C}([0,T], \mathbb{R}^d))$.

We make the following **assumptions**:

- $\mu_i \ll \mathrm{Leb}$ for any $i \in \mathsf{S}$,
- $r \in \mathsf{S}$ (*optional*),
- $\varphi_r = \mathrm{d}\mu_{i_{K-1}}/\mathrm{dLeb}$ (*optional*).

In practice:

- r may be chosen in $\mathsf{V}\backslash\mathsf{S}$ (only the first iteration of TreeDSB differs),
- if $r \in \mathsf{S}$, the choice of φ_r does not change the solutions of TreeSB[6].

[6] See Proposition 4.2. in Peyré et al. (2019).

Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

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

[7] This is directly obtained by considering branching processes with deterministic time steps.

Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

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

In the **tree-based setting**, we prove that the **dynamic version of mIPF**[7] amounts to **recursively** define path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ as follows.

Proposition 1

- *At step $n = 0$:* $\mathbb{P}^0_{(v,v')|0} \sim (\mathbf{B}_t)_{t\in[0,T_{v,v'}]}$ *for any* $(v,v') \in \mathsf{E}_{k_0}$ *and* $\mathbb{P}^0_{(r,\cdot),0} = \pi^0_r$.

[7] This is directly obtained by considering branching processes with deterministic time steps.

Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

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

In the **tree-based setting**, we prove that the **dynamic version of mIPF**[7] amounts to **recursively** define path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ as follows.

Proposition 1

- *At step $n=0$:* $\mathbb{P}^0_{(v,v')|0} \sim (\mathbf{B}_t)_{t\in[0,T_{v,v'}]}$ *for any* $(v,v') \in \mathsf{E}_{k_0}$ *and* $\mathbb{P}^0_{(r,\cdot),0} = \pi^0_r$.
- *At step $n+1$: consider* $\mathsf{T}_{k_n+1} = (\mathsf{V}, \mathsf{E}_{k_n+1})$ *rooted in* i_{k_n+1} *and* $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. *Let* $(v,v') \in \mathsf{E}_{k_n+1}$.

[7] This is directly obtained by considering branching processes with deterministic time steps.

Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

$$\pi^{n+1} = \operatorname{argmin}\{\mathrm{KL}(\pi|\pi^n) \;:\; \pi \in \mathscr{P}^{(|\mathsf{V}|)},\ \pi_{i_{k_n+1}} = \mu_{i_{k_n+1}}\},$$

In the **tree-based setting**, we prove that the **dynamic version of mIPF**[7] amounts to **recursively** define path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ as follows.

Proposition 1

- *At step $n = 0$:* $\mathbb{P}^0_{(v,v')|0} \sim (\mathbf{B}_t)_{t\in[0,T_{v,v'}]}$ *for any* $(v,v') \in \mathsf{E}_{k_0}$ *and* $\mathbb{P}^0_{(r,\cdot),0} = \pi^0_r$.
- *At step $n+1$: consider* $\mathsf{T}_{k_n+1} = (\mathsf{V}, \mathsf{E}_{k_n+1})$ *rooted in* i_{k_n+1} *and* $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. *Let* $(v,v') \in \mathsf{E}_{k_n+1}$.
  - *If* $(v,v') \in \mathsf{E}_{k_n}\backslash\mathsf{P}$: $\mathbb{P}^{n+1}_{(v,v')} = \pi^{n+1}_v \otimes \mathbb{P}^n_{(v,v')|0}$.

[7] This is directly obtained by considering branching processes with deterministic time steps.

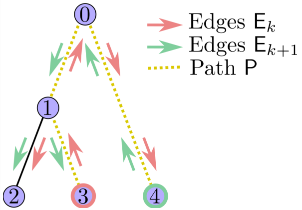
Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

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

In the **tree-based setting**, we prove that the **dynamic version of mIPF**[7] amounts to **recursively** define path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ as follows.

Proposition 1

- *At step $n=0$:* $\mathbb{P}^0_{(v,v')|0} \sim (\mathbf{B}_t)_{t\in[0,T_{v,v'}]}$ *for any* $(v,v')\in\mathsf{E}_{k_0}$ *and* $\mathbb{P}^0_{(r,\cdot),0} = \pi^0_r$.
- *At step $n+1$: consider* $\mathsf{T}_{k_n+1} = (\mathsf{V},\mathsf{E}_{k_n+1})$ *rooted in* i_{k_n+1} *and* $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. *Let* $(v,v')\in\mathsf{E}_{k_n+1}$.
  - *If* $(v,v')\in\mathsf{E}_{k_n}\backslash\mathsf{P}$: $\mathbb{P}^{n+1}_{(v,v')} = \pi^{n+1}_v \otimes \mathbb{P}^n_{(v,v')|0}$.
  - *If* $(v',v)\in\mathsf{P}$: $\mathbb{P}^{n+1}_{(v,v')} = \pi^{n+1}_v \otimes (\mathbb{P}^n_{(v',v)})^R_{|0}$.



[7] This is directly obtained by considering branching processes with deterministic time steps.

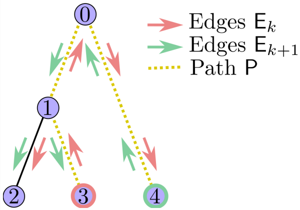
Relying on Benamou et al. (2015), the extension of the **static IPF** procedure in the **multimarginal** setting **(mIPF)** is given by

$$\pi^{n+1} = \operatorname{argmin}\{\mathrm{KL}(\pi|\pi^n) \; : \; \pi \in \mathscr{P}^{(|\mathsf{V}|)}, \; \pi_{i_{k_n+1}} = \mu_{i_{k_n+1}}\},$$

In the **tree-based setting**, we prove that the **dynamic version of mIPF**[7] amounts to **recursively** define path measures $\{\mathbb{P}^n_{(v,v')}\}_{n\in\mathbb{N},(v,v')\in\mathsf{E}_{k_n}}$ as follows.

Proposition 1

- *At step $n=0$:* $\mathbb{P}^0_{(v,v')|0} \sim (\mathbf{B}_t)_{t\in[0,T_{v,v'}]}$ *for any* $(v,v')\in\mathsf{E}_{k_0}$ *and* $\mathbb{P}^0_{(r,\cdot),0} = \pi^0_r$.
- *At step $n+1$: consider* $\mathsf{T}_{k_n+1} = (\mathsf{V},\mathsf{E}_{k_n+1})$ *rooted in* i_{k_n+1} *and* $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. *Let* $(v,v')\in\mathsf{E}_{k_n+1}$.
  - *If* $(v,v')\in\mathsf{E}_{k_n}\backslash\mathsf{P}$: $\mathbb{P}^{n+1}_{(v,v')} = \pi^{n+1}_v \otimes \mathbb{P}^n_{(v,v')|0}$.
  - *If* $(v',v)\in\mathsf{P}$: $\mathbb{P}^{n+1}_{(v,v')} = \pi^{n+1}_v \otimes (\mathbb{P}^n_{(v',v)})^R{}_{|0}$.



We get that $\mathrm{Ext}(\mathbb{P}^n_{(v,v')}) = \pi^n_{v,v'}$ for any $n\in\mathbb{N}$ and any $(v,v')\in\mathsf{E}_{k_n}$!

[7] This is directly obtained by considering branching processes with deterministic time steps.

TreeDSB methodology

In our setting, we consider $2\,|\mathsf{E}|$ **neural networks** to approximate the scores **on each edge** (forward/backward). Then, **our methodology locally acts as DSB**.

TreeDSB methodology

In our setting, we consider $2\,|\mathsf{E}|$ **neural networks** to approximate the scores **on each edge** (forward/backward). Then, **our methodology locally acts as DSB**.

Let $n \in \mathbb{N}$. Assume that we have computed $\mathbb{P}^n$ and want to compute $\mathbb{P}^{n+1}$. Consider the path $\mathsf{P} = \mathrm{path}_{\mathsf{T}_{k_n}}(i_{k_n}, i_{k_n+1})$. Then, for any $(v', v) \in \mathsf{P}$:

(1) *approximately* sample from $\mathbb{P}^n_{(v',v)}$ using E.-M. **time discretization**,

(2) compute an *approximation* of $(\mathbb{P}^n_{(v',v)})^R$ with these samples, using the score-matching technique from De Bortoli et al. (2021).

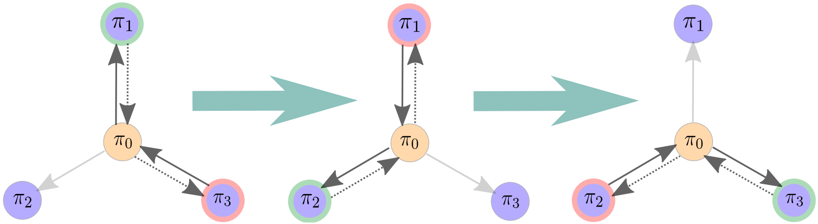


Figure: Illustration of a TreeDSB cycle over a star-shaped tree with 3 leaves.

Outline

Assume that T is a **star-shaped tree** and S $= \{1, \ldots, \ell\}$. Let $\varepsilon > 0$.

We recall the definition of the **entropy-regularized Wasserstein-2 distance**

$$W_{2,\varepsilon}^2(\mu, \nu) = \inf\{\int \|x_1 - x_0\|^2 \mathrm{d}\pi(x_0, x_1) - \varepsilon \mathrm{H}(\pi) : \pi \in \mathscr{P}^{(2)}, \pi_0 = \mu, \pi_1 = \nu\}$$

Assume that T is a **star-shaped tree** and S $= \{1, \ldots, \ell\}$. Let $\varepsilon > 0$.

We recall the definition of the **entropy-regularized Wasserstein-2 distance**

$$W_{2,\varepsilon}^2(\mu, \nu) = \inf\{\int \|x_1 - x_0\|^2 \mathrm{d}\pi(x_0, x_1) - \varepsilon \mathrm{H}(\pi) : \pi \in \mathscr{P}^{(2)}, \pi_0 = \mu, \pi_1 = \nu\}$$

We consider the **doubly-reg. Wasserstein-2 barycenter** problem (**regWB**)

$$\mu_\varepsilon^\star = \arg\min\{\sum_{i=1}^\ell w_i W_{2,\varepsilon/w_i}^2(\mu, \mu_i) + \ell\varepsilon \mathrm{H}(\mu) + \varepsilon \mathrm{KL}(\mu|\mu_0) : \mu \in \mathscr{P}(\mathbb{R}^d)\} ,$$

where $(w_i)_{i\in\{1,\ldots,\ell\}} \in (0, +\infty)^\ell$ and $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$ is a reference measure.

Assume that T is a **star-shaped tree** and S $= \{1, \dots, \ell\}$. Let $\varepsilon > 0$.

We recall the definition of the **entropy-regularized Wasserstein-2 distance**

$$W_{2,\varepsilon}^2(\mu, \nu) = \inf\{\int \|x_1 - x_0\|^2 \mathrm{d}\pi(x_0, x_1) - \varepsilon \mathrm{H}(\pi) : \pi \in \mathscr{P}^{(2)}, \pi_0 = \mu, \pi_1 = \nu\}$$

We consider the **doubly-reg. Wasserstein-2 barycenter** problem (**regWB**)

$$\mu_\varepsilon^\star = \arg\min\{\textstyle\sum_{i=1}^\ell w_i W_{2,\varepsilon/w_i}^2(\mu, \mu_i) + \ell\varepsilon \mathrm{H}(\mu) + \varepsilon \mathrm{KL}(\mu|\mu_0) : \mu \in \mathscr{P}(\mathbb{R}^d)\},$$

where $(w_i)_{i \in \{1,\dots,\ell\}} \in (0, +\infty)^\ell$ and $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$ is a reference measure.

Proposition 2

*Let $\mu_0 \in \mathscr{P}$ such that $\mu_0 \ll \mathrm{Leb}$. Assume that $r = 0$ and $\varphi_r = \mathrm{d}\mu_0/\mathrm{dLeb} > 0$ in **TreeSB**. Also assume that **TreeSB** admits a feasible solution. Then **regWB** has a **unique solution** $\pi_0^\star$, where $\pi^\star$ is the unique solution to **TreeSB**.*

More generally, **TreeSB** is equivalent to a **doubly-regularized formulation of the Wasserstein propagation** problem !

Outline

We compute Wasserstein Barycenters between $K = 3$ probability distributions.

We compare **TreeDSB** with two *state-of-the-art* **regularized OT methods**:

- free-support Wasserstein barycenter (**fsWB**) (Cuturi and Doucet, 2014)
- continuous regularized Wasserstein barycenter (**crWB**) (Li et al., 2020)

TreeDSB setting[8]:

- $T_{v,v'} = K\varepsilon/2$ for any $\{v, v'\} \in$ E,
- μ_0 is a well-chosen Gaussian distribution,
- 50 timesteps in the SDE time discretization,
- the order of the leaves is randomly shuffled between mIPF cycles.

https://github.com/maxencenoble/tree-diffusion-schrodinger-bridge

[8] Further details on the implementation are provided in the paper.

Synthetic 2D datasets (1/3)

Parameters: $\varepsilon = 0.2$ $(T = 0.3)$, 20 mIPF cycles.

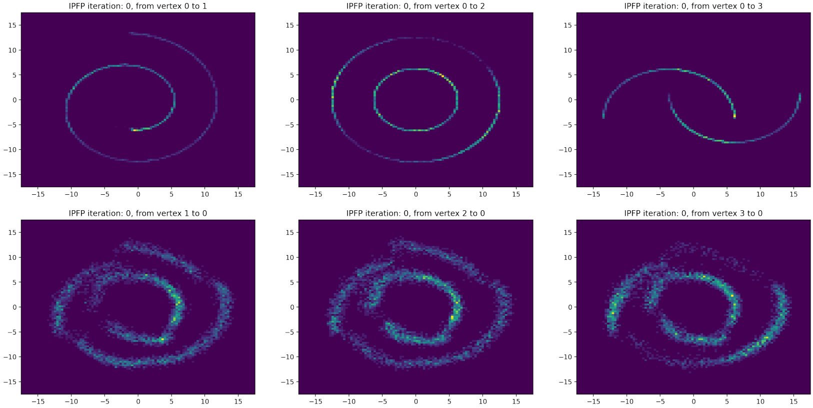


Figure: From left to right: estimated densities (*upper*) and estimated barycenter (*bottom*) for Swiss-roll, circle and moons.

Synthetic 2D datasets (2/3)

Parameters: $\varepsilon = 0.1$ $(T = 0.15)$, 20 mIPF cycles.

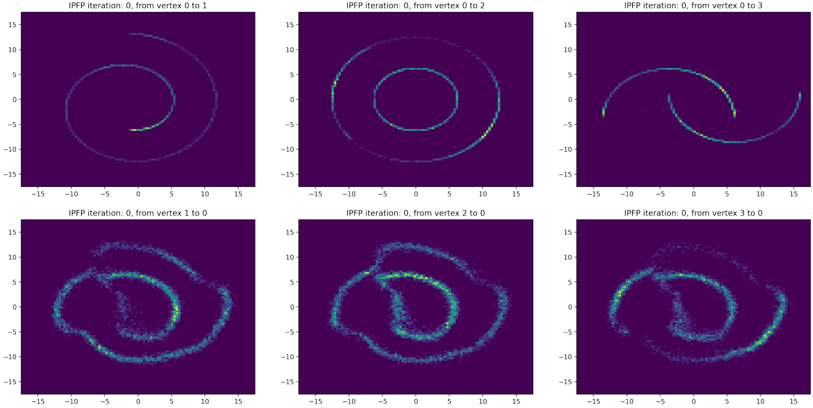


Figure: From left to right: estimated densities (*upper*) and estimated barycenter (*bottom*) for Swiss-roll, circle and moons.

Synthetic 2D datasets (3/3)

Parameters: $\varepsilon = 0.05$ ($T = 0.075$), 35 mIPF cycles.

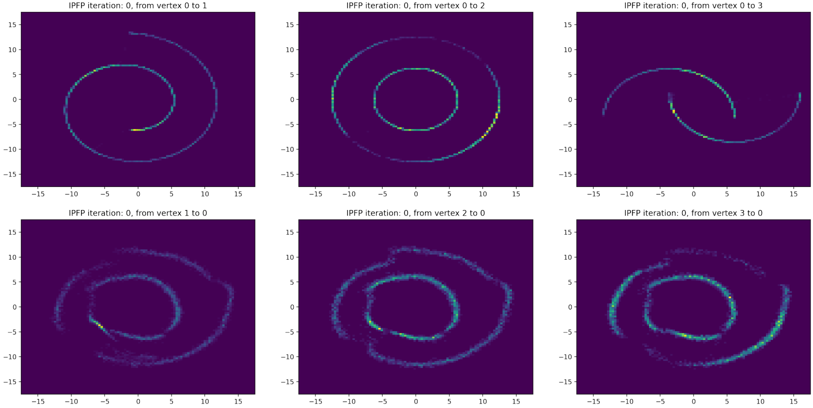


Figure: From left to right: estimated densities (*upper*) and estimated barycenter (*bottom*) for Swiss-roll, circle and moons.

MNIST datasets (1/3)

Parameters: $\varepsilon = 0.5$ ($T = 0.5$), 5 mIPF cycles.

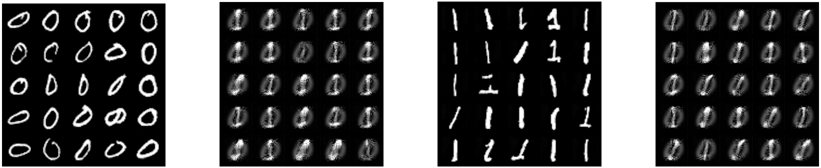


Figure: Reconstructed measures and regularized Wasserstein barycenter obtained from MNIST digits 0 (*left*) and 1 (*right*).

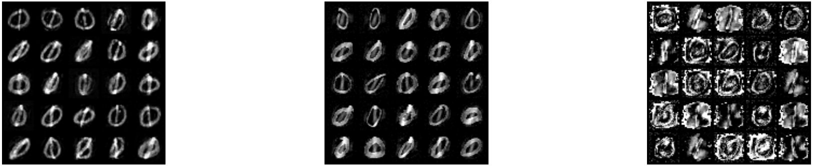


Figure: From left to right: 0-1 Wasserstein barycenter obtained from Fan et al. (2020) (*non-regularized*), Korotin et al. (2021) (*non-regularized*), Li et al. (2020).

MNIST datasets (2/3)

Parameters: $\varepsilon = 0.5$ ($T = 0.75$), 5 mIPF cycles.

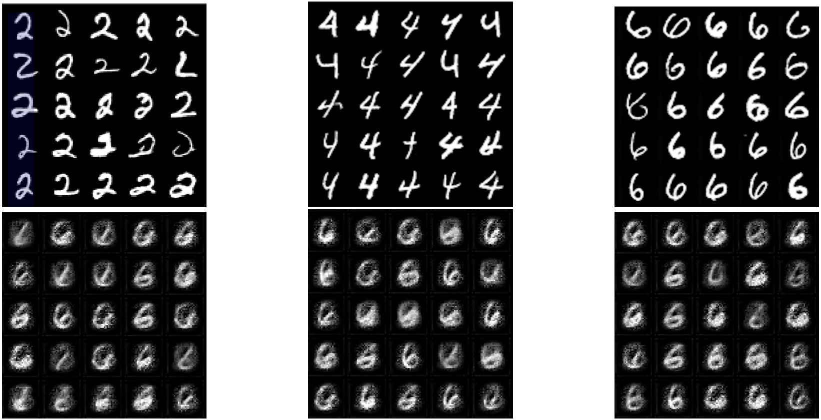


Figure: From left to right: estimated samples (*upper*) and estimated regularized Wasserstein barycenter samples (*bottom*) for MNIST digits 2,4 and 6.

MNIST datasets (3/3)

Parameters: $\varepsilon = 0.5$ ($T = 0.75$), 5 mIPF cycles.

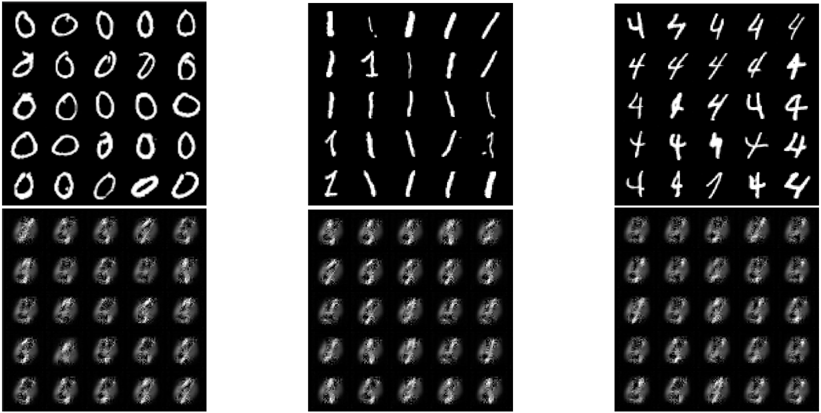


Figure: From left to right: estimated samples (*upper*) and estimated regularized Wasserstein barycenter samples (*bottom*) for MNIST digits 0,1 and 4.

Subset posterior aggregation

Data: **Bayesian posterior** distributions from a **logistic regression** model evaluated on a partition of `wine`[9] dataset ($d = 42$) between 3 subdatasets (splitted **with & without heterogeneity** according to the output).

In theory, the non-regularized Wasserstein barycenter should match the *full data* Bayesian posterior distribution (Srivastava et al., 2018).

Parameters: $\varepsilon = 0.2$ ($T = 0.3$), 10 mIPF cycles.

Evaluation with the **Bures-Wasserstein Unexplained Variance Percentage** (Korotin et al., 2021) between the estimate $\hat{\mu}$ and the full-data posterior $\mu^\star$

$$\mathrm{BW}_2^2\text{-UVP}(\hat{\mu}, \mu^\star) \propto W_2^2(\mathrm{N}(\mathbb{E}[\hat{\mu}], \mathrm{Cov}(\hat{\mu})), \mathrm{N}(\mathbb{E}[\mu^\star], \mathrm{Cov}(\mu^\star))) \ .$$

Method	Without heterogeneity	With heterogeneity
fsWB	$12.95_{\pm 0.35}$	$14.43_{\pm 0.51}$
crWB	$20.66_{\pm 0.71}$	$23.06_{\pm 0.12}$
TreeDSB	$\mathbf{8.69_{\pm 0.12}}$	$\mathbf{8.90_{\pm 0.68}}$

Conclusion

Maxence Noble, Valentin de Bortoli, Arnaud Doucet, Alain Durmus (arxiv preprint, 2023). **Tree-Based Diffusion Schrödinger Bridge with Applications to Wasserstein Barycenters.**

- We introduce **TreeDSB**, a scalable scheme to approximate solutions of **entropy-regularized multimarginal OT** problems defined on general **trees**.
- We prove the **convergence** of this algorithm under mild assumptions.
- We illustrate the efficiency of TreeDSB to compute **Wasserstein barycenters** in several tasks (vision, Bayesian fusion).

Computational limits:

- TreeDSB is unstable when ε, equivalently T, is **too low** (*common EOT limit*),
- **Bias (discretization/learning)** is accumulated along the iterations,
- TreeDSB is **not adapted for a large number of leaves**.

Future work:

- Provide **quantitative convergence** bounds for **mIPF**,
- Rely on recent developments from the **flow matching** community (Lipman et al., 2023; Peluchetti, 2023; Shi et al., 2023).

Tree-Based Diffusion Schrödinger Bridge with Applications to Wasserstein Barycenters

Maxence Noble[1] Valentin de Bortoli[2]
Arnaud Doucet[3] Alain Durmus[1]

[1]Centre de Mathématiques Appliquées, Ecole Polytechnique
Institut Polytechnique de Paris, France

[2]Département d'Informatique, École Normale Supérieure
CNRS, Université PSL, Paris, France

[3]Department of Statistics, University of Oxford, UK

Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard. Time reversal of diffusion processes under a finite entropy condition. *arXiv preprint arXiv:2104.07708*, 2021.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Marco Cuturi and Arnaud Doucet. Fast computation of Wasserstein barycenters. In *International conference on machine learning*, pages 685–693. PMLR, 2014.

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

Jiaojiao Fan, Amirhossein Taghvaei, and Yongxin Chen. Scalable computations of Wasserstein barycenter via input convex neural networks. *arXiv preprint arXiv:2007.04462*, 2020.

Isabel Haasler, Axel Ringh, Yongxin Chen, and Johan Karlsson. Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem. *SIAM Journal on Control and Optimization*, 59(4):2428–2453, 2021.

Ulrich G Haussmann and Etienne Pardoux. Time reversal of diffusions. *The Annals of Probability*, pages 1188–1205, 1986.

Leonid V Kantorovich. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, pages 199–201, 1942.

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

Alexander Korotin, Lingxiao Li, Justin Solomon, and Evgeny Burnaev. Continuous Wasserstein-2 barycenter estimation without minimax optimization. *arXiv preprint arXiv:2102.01752*, 2021.

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

Lingxiao Li, Aude Genevay, Mikhail Yurochkin, and Justin M Solomon. Continuous regularized Wasserstein barycenters. *Advances in Neural Information Processing Systems*, 33: 17755–17765, 2020.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *International Conference on Learning Representations*, 2023.

Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th International Conference on Computer Vision*, pages 460–467. IEEE, 2009.

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

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

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion Schrödinger bridge matching. *arXiv preprint arXiv:2303.16852*, 2023.

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

Justin Solomon, Raif Rustamov, Leonidas Guibas, and Adrian Butscher. Wasserstein propagation for semi-supervised learning. In *International Conference on Machine Learning*, pages 306–314. PMLR, 2014.

Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (ToG)*, 34(4): 1–11, 2015.

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