Unbalanced Multi-marginal Optimal Transport

Florian Beier · Johannes von Lindheim · Sebastian Neumayer · Gabriele Steidl

Received: 30 July 2021 / Accepted: 24 September 2022 / Published online: 8 October 2022 © The Author(s) 2022

Abstract

Entropy-regularized optimal transport and its multi-marginal generalization have attracted increasing attention in various applications, in particular due to efficient Sinkhorn-like algorithms for computing optimal transport plans. However, it is often desirable that the marginals of the optimal transport plan do not match the given measures exactly, which led to the introduction of the so-called unbalanced optimal transport. Since unbalanced methods were not examined for the multi-marginal setting so far, we address this topic in the present paper. More precisely, we introduce the unbalanced multi-marginal optimal transport problem and its dual and show that a unique optimal transport plan exists under mild assumptions. Furthermore, we generalize the Sinkhorn algorithm for regularized unbalanced optimal transport to the multi-marginal setting and prove its convergence. For cost functions decoupling according to a tree, the iterates can be computed efficiently. At the end, we discuss three applications of our framework, namely two barycenter problems and a transfer operator approach, where we establish a relation between the barycenter problem and the multi-marginal optimal transport with an appropriate tree-structured cost function.

Keywords Entropy regularization · Multi-marginal optimal transport · Sinkhorn algorithm · Unbalanced optimal transport · Wasserstein barycenters

Mathematics Subject Classification 49Q22 · 49Q20 · 49M29 · 65D18 · 37M10

1 Introduction

Over the last decades, optimal transport (OT) has attracted increasing attention in various applications, e.g., barycenter problems [1,4], image matching [48,49], and machine learning [3,29,37]. As the OT minimization problem is numerically difficult to solve, a common approach is to add an entropy regularization term [18]. This enables us to approximately solve the problem using Sinkhorn iterations [46] by exploiting an explicit relation between the solutions to the corresponding primal and dual problems. These iterations can be implemented in parallel on GPUs, which makes even large-scale problems solvable within reasonable time. Recently, a debiased version of the Sinkhorn divergence was investigated in [25,41,44], which has the advantage that it characterizes the weak convergence of measures. However, in many applications, the assumption that the marginal measures are matched exactly appears to be unreasonable. To deal with this issue, unbalanced optimal transport (UOT) [12,13,39] was introduced. Here, the hard marginal constraints are replaced by penalizing the $\phi$-divergences between the given measures and the corresponding marginals. By making minimal modifications, we can use Sinkhorn-like iterations and hence the computational complexity and scalability remain the same. For the unbalanced case, the mentioned debiasing was discussed in [47]. For Gaussian distributions, the corresponding divergence even has a closed form [34].

So far, we have commented on OT between two measures. For certain practical tasks such as matching for teams [10], particle tracking [11], and information fusion [19,24], it is useful to compute transport plans between more than...
two-marginal measures. This is done in the framework of multi-marginal optimal transport (MOT) [42], where again entropy regularization is possible. The problem was tackled numerically for Coulomb cost in [5] and more general repulsive costs in [17,31]. Later, an efficient solution for tree-structured costs using Sinkhorn iterations was established in [32]. More recently, these results were extended to even more general cost functions in [2].

In this paper, we want to combine UOT and MOT by investigating unbalanced multi-marginal optimal transport (UMOT). We can build upon the previous papers, but we will see that this has to be done carefully, since various generalizations that seem to be straightforward at the first glance appear to be not. Inspired by the papers [20] and [47], we formulate Sinkhorn iterations for the UMOT problem. Our approach differs from that in [47] as we cannot rely on the 1-Lipschitz continuity of the ($c, ε$)-transform, which only holds in the two-marginal case. Instead, we establish uniform boundedness of the iterates and then exploit the compactness of the Sinkhorn operator. As in the two-marginal case, we retain the excellent scalability of the algorithm for tree-structured costs. Furthermore, we prove that these iterations are convergent under mild assumptions.

As one possible application, we discuss the computation of regularized UOT barycenters based on the regularized UMOT problem. The OT barycenter problem was first introduced in [1] and then further studied, e.g., in [4,16,21,40] for the balanced setting. As soon as entropic regularization is applied, we usually obtain a blurred barycenter. One possible solution is to use the debiased Sinkhorn divergence instead [33,43]. Similarly as in [32] for the balanced case, we observe that solving an UMOT problem instead of minimizing a sum of UOT “distance” reduces the blur considerably. We also validate this observation theoretically. For this purpose, we show that for tree-structured costs UMOT transport plans are already determined by their two-dimensional marginals. A complementary approach for computing barycenters in an unbalanced setting is based on the Hellinger–Kantorovich distance [14,28].

Furthermore, we provide a numerical UMOT example with a path-structured cost function. We observe that in comparison with a sequence of UOT problems with the same input measures, the UMOT approach has several advantages since the target measure from some optimal UOT plan is not necessarily equal to the source measure of the subsequent UOT plan. For the (single) UMOT plan, this is impossible by construction. Additionally, coupling the problems allows information to be shared between them. Note that there is almost no computational overhead compared to solving the problems sequentially.

Outline of the Paper Section 2 contains the necessary preliminaries. The regularized UMOT problem, in particular the existence and uniqueness of a solution as well as its dual problem, is provided in Sect. 3. In Sect. 4, we derive a Sinkhorn algorithm for solving the regularized UMOT problem and prove its convergence. Then, in Sect. 5, we investigate the barycenter problem with respect to regularized UOT and establish a relation to the regularized UMOT problem with a tree-structured cost function, where the tree is simply star-shaped. Furthermore, we discuss an extension of these considerations to more general tree-structured cost functions. Additionally, we outline an efficient implementation of the Sinkhorn algorithm for tree-structured costs. Numerical examples, which illustrate our theoretical findings, are provided in Sect. 6. Finally, we draw conclusions in Sect. 7.

2 Preliminaries

Throughout this paper, let $X$ be a compact Polish space with associated Borel $σ$-algebra $B(X)$. By $M(X)$, we denote the space of finite signed real-valued Borel measures, which can be identified via Riesz’ representation theorem with the dual space of the continuous functions $C(X)$ endowed with the norm $\|f\|_{C(X)}:=\max_{x\in X}|f(x)|$. Denoting the associated dual pairing by $\langle μ, f \rangle := \int_X f \, dμ$, the space $M(X)$ can be equipped with the weak* topology, i.e., a sequence $(μ_n)_{n\in \mathbb{N}} \subset M(X)$ converges weakly to $μ \in M(X)$, written $μ_n \rightharpoonup μ$, if

$$\int_X f \, dμ_n \rightharpoonup \int_X f \, dμ \quad \text{for all } f \in C(X).$$

The associated dual norm of $μ \in M(X)$, also known as total variation, is given by $TV(μ) = \sup_{\|f\|_{C(X)} \leq 1} \langle μ, f \rangle$. By $M^+(X)$, we denote the subset of nonnegative measures. The support of $μ \in M^+(X)$ is defined as the closed set

$$\text{supp}(μ) := \{x \in X : B \subset X \text{ open, } x \in B \implies μ(B) > 0\}.$$ 

For $μ \in M^+(X)$ and $p \in [1, \infty)$, let $L^p(X, μ)$ be the Banach space (of equivalence classes) of real-valued measurable functions with norm $\|f\|_{L^p(X, μ)} \leq \infty$.

A measure $v \in M(X)$ is called absolutely continuous with respect to $μ$, and we write $v \ll μ$, if for every $A \in B(X)$ with $μ(A) = 0$ we have $v(A) = 0$. For any $μ, v \in M^+(X)$ with $v \ll μ$, the Radon–Nikodym derivative

$$σ_v := \frac{dv}{dμ} \in L^1(X, μ)$$

exists and $v = σ_v μ$. Furthermore, $μ, v \in M(X)$ are called mutually singular, denoted by $μ \perp v$, if there exist two disjoint sets $X_μ, X_v \in B(X)$ such that $X = X_μ \cup X_v$ and
for every $A \in \mathcal{B}(\mathcal{X})$ we have $\mu(A) = \mu(A \cap X_1)$ and $\nu(A) = \nu(A \cap X_1)$. Given $\mu, \nu \in M^+(\mathcal{X})$, there exists a unique Lebesgue decomposition $\mu = \sigma_\mu v + \mu^\perp$, where $\sigma_\mu \in L^1(\mathcal{X}, \nu)$ and $\mu^\perp \perp \nu$. Let $\mathcal{Y}$ be another compact Polish space and $T: \mathcal{X} \to \mathcal{Y}$ be a measurable function, i.e., $T^{-1}(A) \in \mathcal{B}(\mathcal{Y})$ for all $A \in \mathcal{B}(\mathcal{Y})$. Then, the push-forward measure of $\mu$ by $T$ is defined as $T_\# \mu := \mu \circ T^{-1}$.

Let $V$ be a real Banach space with dual $V^*$ and dual pairing $\langle v, x \rangle = vv(x)$, $v \in V^*, x \in V$. For $F: V \to (-\infty, +\infty]$, the domain of $F$ is given by $\text{Dom} F := \{x \in V: F(x) \in \mathbb{R}\}$. If $\text{Dom} F \neq \emptyset$, then $F$ is called proper. By $\Gamma_0(V)$ we denote the set of proper, convex, lower semi-continuous (lsc) functions mapping from $V$ to $(-\infty, +\infty]$. The subdifferential of $F: V \to (-\infty, +\infty]$ at a point $x_0 \in \text{Dom} F$ is defined as

$$\partial F(x_0) := \{q \in V^* : F(x) \geq F(x_0) + \langle q, x - x_0 \rangle\},$$

and $\partial F(x_0) = \emptyset$ if $x_0 \notin \text{Dom} F$. For a function $F: V^n \to (-\infty, +\infty]$, $\partial F$ denotes the subdifferential of $F$ with respect to the $i$th component. The Fenchel conjugate $F^*: V^* \to (-\infty, +\infty]$ is given by

$$F^*(q) = \sup_{x \in V} \{\langle q, x \rangle - F(x)\}.$$  

A nonnegative function $\varphi \in \Gamma_0(\mathbb{R})$ satisfying $\varphi(1) = 0$ and $\varphi|_{(-\infty, 0)} = +\infty$ is called entropy function with recession constant $\varphi_\infty = \lim_{x \to -\infty} \varphi(x)/x$. In this case, $\text{Dom} \varphi^\ast = (-\infty, \varphi_\infty)$. For every $\mu, \nu \in M^+(\mathcal{X})$ with Lebesgue decomposition $\mu = \sigma_\mu v + \mu^\perp$, the $\varphi$-divergence $D_\varphi: M^+(\mathcal{X}) \times M^+(\mathcal{X}) \to [0, +\infty]$ is given in its primal and dual form by

$$D_\varphi(\mu, \nu) := \int_\mathcal{X} \varphi \circ \sigma_\mu \, dv + \varphi' \circ \mu^\perp(\mathcal{X})$$
$$= \sup_{f \in C(\mathcal{X})} \langle \mu, f \rangle - \langle \nu, \varphi \circ f \rangle$$

with the convention $0 \cdot \infty = 0$. The mapping $D_\varphi$ is jointly convex, weakly lsc and fulfills $D_\varphi(\mu, \nu) \geq 0$, see [39, Cor. 2.9]. Furthermore, we have for $t > 0$ that $D_{\varphi^t} = tD_\varphi$. We will use the following properties of $\varphi$-divergences, see also [47].

Example 2.1 (i) Let $\varphi := \iota(1)$, where $\iota_S$ denotes the indicator function of the set $S$, i.e., $\iota_S(x) = 0$ if $x \in S$, and $\iota_S(x) = +\infty$ otherwise. Then, $\varphi^\ast(q) = q$, $\varphi_\infty = +\infty$, and

$$D_\varphi(\mu, \nu) = \begin{cases} 0 & \text{if } \mu = \nu, \\ +\infty & \text{otherwise.} \end{cases}$$

(ii) For $\varphi := \iota_{(0, \infty)}$, we get $\varphi^\ast(q) = \iota_{(-\infty, 0)}$, $\varphi_\infty = +\infty$, and $D_\varphi(\mu, \nu) = 0$.

(iii) Consider the Shannon entropy with $\varphi(x) := x \ln x - x + 1$ and the agreement $0 \ln 0 = 0$. Then, we have that $\varphi^\ast(q) = \exp(q) - 1$, $\varphi_\infty = +\infty$, and the $\varphi$-divergence is the Kullback–Leibler divergence $\text{KL}: M^+(\mathcal{X}) \times M^+(\mathcal{X}) \to [0, +\infty]$. For $\mu, \nu \in M^+(\mathcal{X})$, if the Radon–Nikodym derivative $\sigma_\mu = \frac{d\mu}{d\nu}$ exists, then

$$\text{KL}(\mu, \nu) := \int_\mathcal{X} \ln(\sigma_\mu) \, d\mu + \nu(\mathcal{X}) - \mu(\mathcal{X}),$$

and otherwise, we set $\text{KL}(\mu, \nu) := +\infty$. Note that the KL divergence is strictly convex with respect to the first variable.

(iv) For $\varphi(x) := |x - 1|$, we get $\varphi^\ast(q) = \max(-1, q)$ if $q \leq 1$ and $\varphi^\ast(q) = +\infty$ otherwise, $\varphi_\infty = 1$, and $D_\varphi(\mu, \nu) = \text{TV}(\mu - \nu)$.

3 Unbalanced Multi-Marginal Optimal Transport

Throughout this paper, we use the following abbreviations.

For compact Polish spaces $\mathcal{X}_i \neq \emptyset$, $i = 1, \ldots, N$, and measures $\nu_i \in M^+(\mathcal{X}_i), i = 1, \ldots, N$, we set $\nu := (\nu_1, \ldots, \nu_N)$ and

$$\mathcal{X} := \bigotimes_{i=1}^N \mathcal{X}_i, \quad \nu_i := \bigotimes_{i=1}^N \nu_i, \quad \nu := \bigotimes_{i=1}^N \nu_i.$$

Furthermore, for $p \in [1, \infty]$ and $f_i \in L^p(\mathcal{X}_i, \nu_i), i = 1, \ldots, N$, we write $f := (f_1, \ldots, f_N)$ and

$$L^{p, \times}(\mathcal{X}, \nu) := \bigotimes_{i=1}^N L^p(\mathcal{X}_i, \nu_i),$$

$$f^\oplus := \bigoplus_{i=1}^N f_i, \quad f_i^\oplus := \bigoplus_{\substack{i=1 \atop i \neq j}}^N f_i,$$

where the product space $L^{p, \times}(\mathcal{X}, \nu)$ is equipped with the $L^p$ norm of the components. For example, if $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$, then for $f = (f_1, f_2, f_3) \in L^{p, \times}(\mathcal{X}, \nu)$ we have

$$f^\oplus(x_1, x_2, x_3) = f_1(x_1) + f_3(x_3).$$

If the domains and measures are clear from the context, we abbreviate the associated norms by $\| \cdot \|_p, p \in [1, \infty]$.

Given $\mu_i \in M^+(\mathcal{X}_i), i = 1, \ldots, N$, the measures $\gamma_i \in M^+(\mathcal{X}_i), i = 1, \ldots, N$, are called reference measures for $\mu$

$$\text{KL}(\mu, \gamma) < \infty.$$
Definition 3.1 Let $\varepsilon > 0$. Given a nonnegative cost $c \in C(\mathcal{X})$, fully supported measures $\mu_i \in \mathcal{M}^+(\mathcal{X}_i)$, $i = 1, \ldots, N$, with reference measures $\gamma_i \in \mathcal{M}^+(\mathcal{X}_i)$, and entropy functions $\varphi_i \in \Gamma_0(\mathbb{R})$, $i = 1, \ldots, N$, the associated regularized unbalanced multi-marginal optimal transport problem (UMOT) reads

$$
\text{UMOT}_\varepsilon(\mu) := \inf_{\pi \in \mathcal{M}^+(\mathcal{X})} \int_{\mathcal{X}} c \, d\pi + \varepsilon \text{KL}(\pi, \gamma) + \sum_{i=1}^{N} D_{\varphi_i}(P_{X_i} \pi, \mu_i),
$$

where $P_{X_i} \pi$ is the $i$th marginal of $\pi$.

Note that the full support condition is no real restriction as we can choose $\mathcal{X}_i = \text{supp}(\mu_i)$. Furthermore, we can implicitly incorporate weights for the marginal penalty terms in (2) by rescaling the entropy functions $\varphi_i$.

Remark 3.2 (Regularization and reference measures) Typical choices for the reference measures $\gamma_i$ are

(i) $\gamma_i = \mu_i$ then $\text{KL}(\mu^\circ, \gamma^\circ) = 0$ and we regularize in $\text{UMOT}_\varepsilon$ by $\text{KL}(\pi, \mu^\circ)$.

(ii) $\gamma_i$ the Lebesgue measure on $\mathcal{X}_i \subset \mathbb{R}^d$. Then $\text{KL}(\mu^\circ, \gamma^\circ) < \infty$ is equivalent to $\mu_i$ having a density in a so-called Orlicz space, see [15,41] for details. Furthermore, the regularizer in $\text{UMOT}_\varepsilon$ is the entropy $\text{KL}(\pi, \gamma^\circ) = E(\pi)$ for continuous measures.

(iii) $\gamma_i$ the counting measure if the $\mathcal{X}_i$ are finite. Here, $\text{KL}(\mu^\circ, \gamma^\circ) < \infty$ is equivalent to $\mu$ being positive. Then, the regularizer is the entropy for discrete measures $\text{KL}(\pi, \lambda^\circ) = E(\pi)$.

Definition (2) includes the following special cases:

- If $\varphi_i = \varepsilon_{\{1\}}$ for all $i = 1, \ldots, N$, then we have by Example 2.1 i) the regularized multi-marginal optimal transport (MOT) with hard constraints for the marginals. For $\varepsilon = 0$, we deal with the plain multi-marginal optimal transport (MOT) formulation.

- If $N = 2$, then we are concerned with regularized unbalanced optimal transport (UOT$_\varepsilon$). If $\varphi_1 = \varphi_2 = \varepsilon_{\{1\}}$, we get regularized optimal transport (OT$_\varepsilon$), and if $\varepsilon = 0$, we arrive at the usual optimal transport (OT) formulation.

Regarding existence and uniqueness of minimizers for $\text{UMOT}_\varepsilon$, we have the following proposition.

Proposition 3.3 The $\text{UMOT}_\varepsilon$ problem (2) admits a unique optimal plan.

Proof The problem is feasible due to $\text{KL}(\mu^\circ, \gamma^\circ) < \infty$. Existence follows since $\text{KL}(\cdot, \gamma^\circ)$ and hence the whole functional (2) is coercive, and since all involved terms are lsc. For the uniqueness note that all terms in (2) are convex in $\pi$ and that $\text{KL}$ is moreover strictly convex in its first argument. □

For applications, the dual formulation of $\text{UMOT}_\varepsilon$ is important.

Proposition 3.4 The $\text{UMOT}_\varepsilon$ problem (2) admits the dual representation

$$
\text{UMOT}_\varepsilon(\mu) = \sup_{f \in L^{\infty, \times}(\mathcal{X}, \gamma)} D_\varepsilon^\circ(f) + \varepsilon \gamma^\circ(\mathcal{X}),
$$

where

$$
D_\varepsilon^\circ(f) := -\sum_{i=1}^{N} \int_{\mathcal{X}_i} \varphi_i^\circ(-f_i) \, d\mu_i - \varepsilon \int_{\mathcal{X}} \exp\left(\frac{f^\circ - c}{\varepsilon}\right) \, d\gamma^\circ.
$$

The optimal plan $\hat{\pi} \in \mathcal{M}^+(\mathcal{X})$ for the primal problem (2) is related to any tuple of optimal dual potentials $\hat{f} \in L^{\infty, \times}(\mathcal{X}, \gamma)$ by

$$
\hat{\pi} = \exp\left(\frac{\hat{f}^\circ - c}{\varepsilon}\right) \gamma^\circ.
$$

Furthermore, any pair of optimal dual potentials $\hat{f}, \hat{g} \in L^{\infty, \times}(\mathcal{X}, \gamma)$ satisfies $\hat{f}^\circ = \hat{g}^\circ$. Moreover, if $N - 1$ of the $\varphi_i^\circ, \ i = 1, \ldots, N$, are strictly convex, then it holds $\hat{f} = \hat{g}$.

Proof First, we set $V := L^{\infty, \times}(\mathcal{X}, \gamma)$ and $W := L^\infty(\mathcal{X}, \gamma^\circ)$, and define $A : V \rightarrow W, F \in \Gamma_0(V)$ and $G \in \Gamma_0(W)$ via

$$
A(f) := f^\circ, $$

$$
F(f) := \sum_{i=1}^{N} \int_{\mathcal{X}_i} \varphi_i^\circ \circ f_i \, d\mu_i, $$

$$
G(f) := \varepsilon \int_{\mathcal{X}} \exp\left(\frac{f - c}{\varepsilon}\right) - 1 \, d\gamma^\circ.
$$

Note that $V^*$ and $W^*$ are the respective dual spaces of finitely additive signed measures that are absolutely continuous with respect to $\gamma$ and $\gamma^\circ$, respectively. From [45, Thm. 4] with the choice $g : \mathcal{X} \times \mathbb{R} \rightarrow \mathbb{R}$, the function $g(x, \cdot) := \varepsilon \exp\left(\frac{p - c(x)}{\varepsilon}\right) - 1$, $p \in \Gamma_0(V)$, $q \in \Gamma_0(W)$, and $g^*(x, q) := c(x)q + \varepsilon(\ln(q)q - q + 1)$ if $q \geq 0$ with the convention $0\ln 0 = 0$ and $g^*(x, q) = \infty$ otherwise,
we get that the Fenchel conjugate \( G^* \in \Gamma_0(W^*) \) is given by

\[
G^*(\pi) = \int_X c\sigma d\gamma^\oplus + \varepsilon \ KL(\sigma_\gamma, \gamma^\ominus)
\]

if there exists nonnegative \( \sigma_\gamma \in L^1(X, \gamma^\ominus) \) with \((\pi, f) = \int_X f\sigma d\gamma^\oplus \) for all \( f \in W \) and \( G^*(\pi) = \infty \) for all other \( \pi \in W^* \).

Using the definition of the Fenchel conjugate and (1), the function \( F^* \circ A^* \in \Gamma_0(W^*) \) can be expressed for any such \( \pi \) as

\[
F^*(A^* \pi) = \sup_{f \in L^{\infty, \times}(X, \gamma)} \langle A^* \pi, f \rangle - \sum_{i=1}^N \int_{X^i} \varphi_i^* \circ f_i \ d\mu_i
\]

\[
= \sup_{f \in L^{\infty, \times}(X, \gamma)} \langle \pi, f^\ominus \rangle - \sum_{i=1}^N \int_{X^i} \varphi_i^* \circ f_i \ d\mu_i
\]

\[
= \sup_{f \in L^{\infty, \times}(X, \gamma)} \sum_{i=1}^N \int_{X^i} f_i \ dP_{\mu_i} \pi - \int_{X^i} \varphi_i^* \circ f_i \ d\mu_i
\]

\[
= \sum_{i=1}^N D_{\psi_i}(P_{\mu_i} \pi, \mu_i).
\]

Now, we obtain the assertion by applying the Fenchel–Rockafellar duality relation

\[
\inf_{w \in W^*} \left\{ F^*(A^* w) + G^*(w) \right\} = \sup_{x \in V} \left\{ -F(-x) - G(Ax) \right\},
\]

see [23, Thm. 4.1, p. 61]. Due to the definition of \( G^* \), it suffices to consider elements from \( W^* \) that can be identified with elements in \( L^1(X, \gamma^\ominus) \). Hence, the problem (6) coincides with (2).

The second assertion follows using the optimality conditions. More precisely, let \( \hat{\pi} \) and \( \hat{f} \) be optimal. By [23, Chap. 3, Prop. 4.1], this yields \( A \hat{f} \in \partial G^*(\hat{\pi}) \) which is equivalent to \( \hat{\pi} \in \partial G(A \hat{f}) \). Since \( G \) is Gâteaux-differentiable with \( VG(f) = \exp\left(\frac{c}{\varepsilon}\right)\gamma^\ominus \), we obtain

\[
\hat{\pi} = \exp\left(\frac{\hat{f}^\ominus - c}{\varepsilon}\right)\gamma^\ominus.
\]

Finally, let \( \hat{f}, \hat{g} \in L^{\infty, \times}(X, \gamma) \) be two optimal dual potentials. The second summand is concave in \( f \) and strictly concave in \( \hat{f}^\ominus \). The first summand in (4) is concave. Moreover, if \( N - 1 \) of the \( \varphi_i, i = 1, \ldots, N \), are strictly convex and (4) is strictly concave. Hence, both claims follow. \( \square \)

### 4 Sinkhorn Algorithm for Solving the Dual Problem

In this section, we derive an algorithm for solving the dual problem (3). We prove its convergence under the assumption that for all \( i = 1, \ldots, N \), we have \( \ln(\sigma_\mu) \in L^\infty(X_i, \gamma_i) \), where \( \sigma_\mu \) is the Radon–Nikodym derivative of \( \mu \) with respect to the reference measures \( \gamma_i \) and some mild assumptions on the entropy functions \( \varphi_i \).

First, we introduce two operators that appear in the optimality conditions of the dual problem, namely the \((c, \varepsilon)\)-transform and the anisotropic proximity operator.

**Definition 4.1** For \( j = 1, \ldots, N \), the \((c, \varepsilon)\)-transform \( F^{(c, \varepsilon)}(\cdot) : L^{\infty, \times}(X_i, \gamma_i) \to L^\infty(X_i, \gamma_j) \) is given by

\[
F^{(c, \varepsilon)}(f) = f^{(c, \varepsilon)} \quad \text{(7)}
\]

\[
:= \varepsilon \ln(\sigma_{\mu}) - \varepsilon \ln\left(\int_{X_i} \exp\left(\frac{f - \varepsilon}{\varepsilon}\right) d\gamma^\ominus\right).
\]

This transform was discussed in relation with MOT_\varepsilon in [20], where the following two properties were shown.

**Lemma 4.2** Let \( \ln(\sigma_{\mu}) \in L^\infty(X_i, \gamma_i), i = 1, \ldots, N, \) and \( f \in L^{\infty, \times}(X_i, \gamma_i) \). Then, the following holds:

(i) For every \( j = 1, \ldots, N \) it holds

\[
\|f^{(c, \varepsilon)} + \lambda_{f, j}\|_\infty \leq \|c\|_\infty + \varepsilon \ln(\sigma_{\mu})\|_\infty,
\]

where

\[
\lambda_{f, j} := \varepsilon \ln\left(\int_{X_i} \exp\left(\frac{f - \varepsilon}{\varepsilon}\right) d\gamma^\ominus\right).
\]

In particular, we get that \( F^{(c, \varepsilon, j)} \) has bounded oscillation

\[
\sup_{y \in X_i} f^{(c, \varepsilon, j)}(y) - \inf_{y \in X_i} f^{(c, \varepsilon, j)}(y) < \infty.
\]

(ii) The nonlinear and continuous operator \( F^{(c, \varepsilon, j)} : L^{\infty, \times}(X_i, \gamma_i) \to L^p(X_i, \gamma_j) \) is compact for \( p \in [1, \infty) \), i.e., it maps bounded sets to relatively compact sets.

**Definition 4.3** For any entropy function \( \varphi \in \Gamma_0(\mathbb{R}) \) and \( \varepsilon > 0 \), the anisotropic proximity operator \( \text{aprox}^\varepsilon_{\varphi} : \mathbb{R} \to \mathbb{R} \) is given by

\[
\text{aprox}^\varepsilon_{\varphi}(p) := \arg\min_{q \in \mathbb{R}} \left\{ \varepsilon e^{-\varphi(q) + \varepsilon \varphi(q)} \right\}.
\]

**Remark 4.4** This operator is indeed well defined. Furthermore, it is 1-Lipschitz and can be given in analytic form for various conjugate entropy functions, see [47].
Example 4.5 Let us have a closer look at the functions from Example 2.1.

(i) For \( \varphi = t \mathbb{1}_{\{1\}} \) it holds \( \text{aprox}^t_{\varphi^*}(p) = p \).
(ii) For \( \varphi = t \mathbb{1}_{(0,\infty)} \), we get \( \text{aprox}^t_{\varphi^*}(p) = 0 \).
(iii) For \( \varphi(x) = t(x \ln(x) - x + 1) \) corresponding to the Kullback–Leibler divergence, we have
\[
\text{aprox}^t_{\varphi^*}(p) = \frac{t}{t + e} p.
\]
(iv) For \( \varphi(x) = |x - 1| \) belonging to the TV distance, it holds
\[
\text{aprox}^t_{\varphi^*}(p) = \begin{cases} t, & p < -t, \\ p, & p \in [-t, t], \\ t, & p > t. \end{cases}
\]

Definition 4.6 The \((c, \varepsilon)\)-transform and the anisotropic proximity operator are concatenated to the \(j\)-th Sinkhorn mapping
\[
S^{(c, \varepsilon, \varphi, j)} : L_{\infty}^\infty(\mathcal{X}, \gamma) \to L_{\infty}^\infty(\mathcal{X}_j, \gamma_j)
\]
defined as \( S^{(c, \varepsilon, \varphi, j)}(f) := f^{(c, \varepsilon, \varphi, j)} \) with
\[
f^{(c, \varepsilon, \varphi, j)} := -\text{aprox}^t_{\varphi^*}(-f^{(c, \varepsilon, j)}),
\]
where the operator \( \text{aprox}^t_{\varphi^*} \) is applied pointwise.

Now, we derive the maximizer of \( D^\varepsilon_k \) defined in (4).

Proposition 4.7 Let \( \ln(\sigma_{\mu_i}) \in L_{\infty}^\infty(\mathcal{X}_i, \gamma_i), \ i = 1, \ldots, N \).

Then, it holds for all \( j = 1, \ldots, N \) and \( f \in L_{\infty}^\infty(\mathcal{X}, \gamma) \) that
\[
D^\varepsilon_k(f) = -\sum_{i \neq j} (\mu_i, \varphi^*_j(-f_i)) - \int_{\mathcal{X}_j} \varphi^*_j(-f_j)d\mu_j
\]
with equality if and only if \( f_j = f^{(c, \varepsilon, \varphi, j)} \). Furthermore, \( f \) is a maximizer of \( D^\varepsilon_k \) if and only if \( f_j = f^{(c, \varepsilon, \varphi, j)} \) for all \( j = 1, \ldots, N \).

Proof We fix \( j \in \{1, \ldots, N\} \) and rewrite
\[
D^\varepsilon_k(f) = -\sum_{i \neq j} (\mu_i, \varphi^*_j(-f_i)) - \int_{\mathcal{X}_j} \varphi^*_j(-f_j)d\mu_j
\]
+ \( \varepsilon \exp\left( \frac{f_j}{\varepsilon} \right) \int_{\mathcal{X}_j} \exp\left( \frac{f^{\oplus}_{\gamma} - c}{\varepsilon} \right)d\gamma_j\).

By rearranging the definition (7) of the \((c, \varepsilon)\)-transform, we get that
\[
\exp\left( \frac{f^{(c, \varepsilon, j)}_{\gamma}}{-\varepsilon} \right) = \frac{1}{\sigma_{\mu_j}} \int_{\mathcal{X}_j} \exp\left( \frac{f^{\oplus}_{\gamma} - c}{\varepsilon} \right)d\gamma_j.
\]
Plugging this into the equation above, we get
\[
D^\varepsilon_k(f) = -\sum_{i \neq j} (\mu_i, \varphi^*_j(-f_i)) - \int_{\mathcal{X}_j} \varphi^*_j(-f_j)
+ \varepsilon \exp\left( \frac{f_j - f^{(c, \varepsilon, j)}_{\gamma}}{\varepsilon} \right)d\mu_j.
\]
The integrand on the right-hand side has the form of the functional in (8), so that we obtain
\[
D^\varepsilon_k(f) \leq D^\varepsilon_k(f, \ldots, f_{j-1}, f^{(c, \varepsilon, \varphi, j)}, f_{j+1}, \ldots).
\]
Since the minimization problem (8) admits a unique solution, strict inequality holds if and only if \( f_j \neq f^{(c, \varepsilon, \varphi, j)} \).

By the first part of the proof, the relation \( f_j = f^{(c, \varepsilon, \varphi, j)} \) is equivalent to \( 0 \in \partial_j D^\varepsilon_k(f) \), and the last statement follows if we show that
\[
\partial D^\varepsilon_k(f) = \bigtimes_{i=1}^N \partial_i D^\varepsilon_k(f).
\]
Using \( F_i : L_{\infty}^\infty(\mathcal{X}_i, \gamma_i) \to [-\infty, \infty) \) as well as \( G : L_{\infty}^\infty(\mathcal{X}, \gamma) \to \mathbb{R} \) given by
\[
F_i(f) = -\int_{\mathcal{X}_i} \varphi^*_i( -f ) d\mu_i, \quad i = 1, \ldots, N,
\]
\[
G(f) = -\varepsilon \int_{\mathcal{X}} \exp\left( \frac{f - c}{\varepsilon} \right)d\gamma
\]
respectively, we can decompose \( D^\varepsilon_k \) as \( D^\varepsilon_k = F^\oplus + G \). As \( G \) is Gâteaux-differentiable, we obtain that \( \partial G = \bigtimes_{i=1}^N \partial_i G \).

By continuity of \( G \) and since \( 0 \in \text{dom} \varphi^*_i \), it holds \( 0 \in \text{dom} G \cap \text{dom} F^\oplus \), such that the subdifferentials are additive by [23, Ch. 1, Prop. 5.6]. Thus, using \( \partial F^\oplus = \bigtimes_{i=1}^N \partial_i F \), we obtain
\[
\partial D^\varepsilon_k = \partial G + \partial F^\oplus = \bigtimes_{i=1}^N \partial_i G + \bigtimes_{i=1}^N \partial_i F^\oplus
= \bigtimes_{i=1}^N \partial_i (G + F^\oplus) = \bigtimes_{i=1}^N \partial_i D^\varepsilon_k.
\]
This concludes the proof.
Inspired by the Sinkhorn iterations for MOT$_{\epsilon}$ in [20] and UOT$_{\epsilon}$ in [47], we propose Algorithm 1 for solving UMOT$_{\epsilon}$ in its dual form (3). By Proposition 4.7, every fixed point of the sequence $(f^{r(N)})_{r\in \mathbb{N}}$ generated by Algorithm 1 is a solution of (3).

**Remark 4.8** It holds $\varphi^*(0) = 0$. Hence, we can choose $f(0) = 0$ as an initialization with $D_{\epsilon}^d(f(0)) < -\infty$.

Next, we want to show that the sequence converges. Note that in [20, Thm. 4.7] convergence of the (rescaled) Sinkhorn algorithm was shown by exploiting the property $D_{\epsilon}^d(f_1, \ldots, f_N) = D_{\epsilon}^d(f_1 + \lambda_1, \ldots, f_N + \lambda_N)$ for all $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$ with $\sum_{i=1}^N \lambda_i = 0$, which holds exclusively in the balanced case where $\varphi^*_i(q) = q$. Hence, significant modifications of the proof are necessary. Albeit taking several ideas from [47], our approach differs as we cannot rely on the 1-Lipschitz continuity of the $(\epsilon, \varepsilon)$-transform, which only holds for $N = 2$. Instead, we exploit the compactness of the Sinkhorn operator as in [20], for which we need to establish uniform boundedness of the iterates. To this end, we need the following two lemmata.

**Lemma 4.9** Let $f^{(n)} \in L^{\infty, \times}(\mathbb{X}, \gamma)$, $n \in \mathbb{N}$, satisfy $\|f^{(n)}\|_{\infty} \rightarrow \infty$ and have uniformly bounded oscillations (see Lemma 4.2). Then, it holds $D_{\epsilon}^d(f^{(n)}) \rightarrow -\infty$.

**Proof** Since the entropy functions $\varphi_i$, $i = 1, \ldots, N$, satisfy $\varphi_i(1) = 0$, we have $\varphi_i^\#(x) \geq x$ for all $x \in \mathbb{R}$. Hence, we can estimate

$$-\sum_{i=1}^N \mu_i \varphi_i^\#(f^{(n)}(x)) \leq \langle \mu^\otimes, (f^{(n)})^\otimes \rangle.$$

Since the $\mu_i$ are absolutely continuous with respect to $\gamma_i$ with density $\sigma_{\mu_i}$, we obtain

$$D_{\epsilon}^d(f^{(n)}) \leq \left\langle \mu^\otimes, (f^{(n)})^\otimes \right\rangle - \epsilon \exp \left(\frac{(f^{(n)})^\otimes - c - \epsilon \ln(\prod_{i=1}^N \sigma_{\mu_i})}{\epsilon}\right).$$

Clearly, $(f^{(n)})^\otimes$ has uniformly bounded oscillation. Hence, for $\|f^{(n)}\|_{\infty} \rightarrow \infty$ the integrand diverges to $-\infty$ on a set of positive measure, which yields the assertion. \( \Box \)

**Lemma 4.10** Let $\ln(\sigma_{\mu_i}) \in L^{\infty}(\mathbb{X}_i, \gamma)$, $i = 1, \ldots, N$, and $D_{\epsilon}^d(f(0)) > -\infty$. For the Sinkhorn sequence $(f^{(n)})_{n \in \mathbb{N}} \subseteq L^{\infty, \times}(\mathbb{X}, \gamma)$ generated by Algorithm 1, there exists a constant $M > 0$ and a sequence $(\lambda^{(n)})_{n \in \mathbb{N}}$, $\lambda^{(n)} \in \mathbb{R}^N$, with $\sum_{i=1}^N \lambda_i^{(n)} = 0$ such that

$$\|f^{(n)}_i + \lambda^{(n)}_i\|_{\infty} < M.$$

\( \Box \) Springer
induced by Algorithm 1 satisfies $\|(f^{(n)})^\oplus - \hat{f}\|_p \to 0$ for any optimal solution $\hat{f} \in L^{\infty, \infty}(\mathcal{X}, \gamma)$ of (3) for every $p \in [1, \infty)$. Moreover, if $N - 1$ of the $\phi_i^*$ are strictly convex, then the optimal $\hat{f}$ is unique and $\|f^{(n)} - \hat{f}\|_p \to 0$ for every $p \in [1, \infty)$.

**Proof** First, we show that the sequence $(f^{(n)})_{n\in\mathbb{N}}$ is uniformly bounded. By Lemma 4.10, there exists a sequence $(\lambda_i(n))_{n \in \mathbb{N}} \subseteq \mathbb{R}^N$, with $\sum_{i=1}^{N} \lambda_i(n) = 0$ and $M > 0$ such that

$$\|\mu_i + \lambda_i(n)\|_\infty \leq M$$

(9)

for all $i = 1, \ldots, N$, $n \in \mathbb{N}$. Define $g_i^{(n)} := f_i^{(n)} + \lambda_i(n)$. To obtain uniform boundedness, it suffices to show that $\max_j |\lambda_j^{(n)}|$ is uniformly bounded in $n$. We have for any $q_i \in \text{dom} \phi_i^*$ and $p_i \in \partial \phi_i^*(q_i)$ by the first-order convexity condition and since elements of $\partial \phi_i^*(q_i)$ are nonnegative, that

$$\langle \mu_i, -\phi_i^*(f^{(n)}) \rangle$$

$$\begin{cases} 
= \langle \mu_i, -\phi_i^*(g_i^{(n)}) \rangle \\
\leq \langle \mu_i, -\phi_i^*(g_i^{(n)}) \rangle + \langle p_i, (g_i^{(n)})_i - \lambda_i^{(n)} \rangle + \langle p_i, (q_i - \lambda_i^{(n)}) \rangle \\
\leq \langle \mu_i, -\phi_i^*(g_i^{(n)}) \rangle + \langle p_i, (g_i^{(n)})_i \rangle - \lambda_i^{(n)} + \langle p_i, (q_i) \rangle.
\end{cases}$$

Consequently, we obtain by (9) that

$$\sum_{i=1}^{N} \langle \mu_i, -\phi_i^*(f^{(n)}) \rangle \leq -\sum_{i=1}^{N} \langle \mu_i, \phi_i^{(n)} \rangle + \langle p_i, (M + q_i) - \phi_i^*(q_i) \rangle.$$

$$= k(p_i, q_i)$$

Setting $m_i := \mu_i(\mathcal{X}), i = 1, \ldots, N$, and using that the $\lambda_i^{(n)}$ sum up to zero, we conclude

$$\sum_{i=1}^{N} \langle \mu_i, -\phi_i^*(f^{(n)}) \rangle \leq \sum_{i=1}^{N} m_i p_i \lambda_i^{(n)} + K(p_i, q_i).$$

(10)

First, since $\text{dom} \phi_i^* = (-\infty, (\phi_i^*)_\infty)$ with $\phi_i^*_\infty \geq 0$, we can fix some $q_i \in \text{int} \text{dom} \phi_i^*$ and some $p_i \in \partial \phi_i^*(q_i) \neq \emptyset$. Assume that there exists at least one $i \in \{1, \ldots, N\}$ such that $\langle \lambda_i^{(n)} \rangle_{n \in \mathbb{N}}$ is unbounded. Since the $\lambda_i^{(n)}$ sum to zero, there then also exists at least one such $i \in \{1, \ldots, N - 1\}$. For each of these $i$, we can extract a subsequence $(n_k)_k$ such that either $\lambda_i^{(n_k)} \to -\infty$ or $\lambda_i^{(n_k)} \to \infty$. In the first case, choose some $p_i > 0$ small enough such that $m_N p_N > m_i p_i$. By assumption, $p_i \in \text{int} \text{dom} \phi_i$, such that there exists $q_i \in \partial \phi_i(p_i)$. Since $\phi_i \in \Gamma_0(\mathbb{R})$, it follows that $p_i \in \partial \phi_i^*(q_i)$. Moreover, since

$$\infty > p_i q_i = \phi_i(p_i) + \phi_i^*(q_i)$$

and $p_i \in \text{dom} \phi_i$, we also have $q_i \in \text{dom}(\phi_i^*)$. Similarly, if $\lambda_i^{(n_k)} \to \infty$, choose some $p_i < \infty$ such that $m_N p_N < m_i p_i$ and $q_i \in \text{dom} \phi_i^*$ with $p_i \in \partial \phi_i^*(q_i)$. Now, we have by Proposition 4.7 that $D^p_c(f^{(n)}) \leq D^p_c(f^{(n_k)})$. Since $(f^{(n_k)}) \in (g^{(n_k)}) \oplus$ and the $g^{(n_k)}$, $i = 1, \ldots, N$, are uniformly bounded, the second summand in $D^p_c(f^{(n_k)})$ remains bounded as $k \to \infty$, while the first summand in $D^p_c(f^{(n_k)})$ goes to $-\infty$ by (10) with the above chosen $(q_i, p_i)$. This is a contradiction to $D^p_c(f^{(n)}) \geq -\infty$. Thus, there is $M > 0$ such that for $i \in \{1, \ldots, N\}$ and $n \in \mathbb{N}$ it holds

$$\|f_i^{(n)}\|_p \leq \|f_i^{(n)} + \lambda_i^{(n)}\|_p + \lambda_i^{(n)} \leq M + \hat{M}.$$
for every \( j = 1, \ldots, N \), where \( C \) stands for some unspecified, possibly changing constant. In particular, it holds
\[
j_1^{(n_k+1)} \rightarrow \hat{f}(c, \varepsilon, \varphi, 1) \in L^p(\mathcal{X}_1, \gamma_1).
\]

As all \( \psi^\ast_n \) are lsc, the dominated convergence theorem implies that \( D^p_x(f(n_k)) \rightarrow D^p_x(\hat{f}) \) for any a.e. convergent subsequence \( (f(n_k))_{k \in \mathbb{N}} \) of \( (f(n))_{n \in \mathbb{N}} \) with limit \( \hat{f} \). Due to this continuity property and since \( (D^p_x(f(n)))_n \) is a convergent sequence, we get
\[
D^p_x(\hat{f}(c, \varepsilon, \varphi, 1), \hat{f}_2, \ldots, \hat{f}_N) = \lim_{k \to \infty} D^p_x(f(n_k)) = \lim_{k \to \infty} D^p_x(f(n_k)) = D^p_x(\hat{f}).
\]

Consequently, Proposition 4.7 implies \( \hat{f}_1 = \hat{f}(c, \varepsilon, \varphi, 1) \). In the same way, we obtain
\[
D^p_x(\hat{f}(c, \varepsilon, \varphi, 1), \hat{f}(c, \varepsilon, \varphi, 2), \hat{f}_3, \ldots, \hat{f}_N) = D^p_x(\hat{f}).
\]

Due to \( \hat{f}_1 = \hat{f}(c, \varepsilon, \varphi, 1) \), this gives \( \hat{f}_2 = \hat{f}(c, \varepsilon, \varphi, 2) \). Proceeding iteratively this way, we obtain that \( \hat{f}_i = \hat{f}(c, \varepsilon, \varphi, i) \) for all \( i = 1, \ldots, N \). Hence, Proposition 4.7 implies that \( \hat{f} \) is an optimal dual vector.

If \( N-1 \) of the \( \varphi^\ast_n \) are strictly convex, then the maximizer in (3) is unique and \( (f(n_k))_{n \in \mathbb{N}} \) converges. Otherwise, we obtain convergence of \( (f(n_k))_{n \in \mathbb{N}} \) to \( \hat{f} \in L^p(\mathcal{X}, \varphi) \). Since \( \hat{f} \) is the same for all possible limit points \( \hat{f} \), convergence of \( (f(n_k))_{n \in \mathbb{N}} \) follows. \( \square \)

Remark 4.12 (Relation to UOT in [47]) For entropy functions satisfying the assumptions of Theorem 4.11, our result can be seen as a generalization of the UOT result in Séjourné et al. [47] to the multi-marginal case and non-Lipschitz costs. Notably, the result by Séjourné et al. covers some additional entropy functions.

Finally, we want to remark that all results of this section also hold true if we do not assume that the spaces \( \mathcal{X}_i, i = 1, \ldots, N \), are compact as long as the cost function \( c \) remains bounded.

5 Barycenters and Tree-Structured Costs

In this section, we are interested in the computation of UOT\( _x \) barycenters and their relation to UOM\( T_x \) with tree-structured costs. An undirected graph \( \mathcal{G} = (\mathcal{V}, \mathcal{E}) \) with \( N \) nodes \( \mathcal{V} = \{1, \ldots, N\} \) and edges \( \mathcal{E} \) is a tree if it is acyclic and connected. We write \( e = (j, k) \) if \( e \) joins the nodes \( j \) and \( k \), where we agree that \( j < k \) in order to count edges only once. Let \( \text{deg}(i) \) denote the number of edges in node \( i \) \( \in \mathcal{V} \). A node \( i \) \( \in \mathcal{V} \) is called a leaf if \( \text{deg}(i) = 1 \). By \( \mathcal{N}_j \), we denote the set of neighbors of node \( j \). For a given tree, let \( t = (t_e)_{e \in \mathcal{E}} \) with \( t_e \in [0, 1] \) and \( \sum_{e \in \mathcal{E}} t_e = 1 \). A cost function \( c_t \) is said to be tree-structured, if it is of the form
\[
c_t = \sum_{e \in \mathcal{E}} t_e c_e = \sum_{(j, k) \in \mathcal{E}} t(j, k) c(j, k).
\]

In Sect. 5.1, we consider the case where the tree is star-shaped, i.e., \( \mathcal{E} = \{(1, N), \ldots, (N-1, N)\} \), see Fig. 1 left. General tree-structured costs as, for example, those in Fig. 1 middle and right are addressed in Sect. 5.2. We restrict our attention to \( \mathcal{X}_i, i = 1, \ldots, N \), which are either finite or compact subsets of \( \mathbb{R}^d \). Moreover, all references measures \( \gamma_i \), \( i = 1, \ldots, N \), are counting measures, respectively, Lebesgue measures, so that we regularize exclusively with the entropy from Rem. 3.2 ii) or iii), respectively.

5.1 Barycenters

For the barycenter problem with respect to UOT\( _x \), we introduce an additional finite set \( \mathcal{Y} \) (or compact set \( \mathcal{Y} \subset \mathbb{R}^{d_N+1} \)) and use for \( \xi \in \mathcal{M}^+(\mathcal{Y}) \) the entropy function
\[
\psi := t[1], \quad D_\psi(\cdot, \xi) = t[1],
\]
from Example 2.1 i). Let \( c_t \in C(\mathcal{X}_i \times \mathcal{Y}), i = 1, \ldots, N \), be nonnegative cost functions. The corresponding tree is star-shaped, i.e., given by \( \mathcal{V} = \{1, \ldots, N+1\} \) and \( \mathcal{E} = \{(1, N+1), \ldots, (N, N+1)\} \). To emphasize the dependence of UOT\( _x \) on these functions, we write UOT\( _{x, \psi, \xi} \). We use an analogous notation for UOM\( T_x \). Let
\[
\Delta_N := \left\{ t = (t_i)_{i=1}^N : t_i \geq 0, \sum_{i=1}^N t_i = 1 \right\}
\]
be the \((N-1)\)-dimensional probability simplex. For given barycentric coordinates \( t \in \Delta_N \), the barycenter \( \hat{x} \in \mathcal{M}^+(\mathcal{Y}) \) of \( \mu \) with respect to UOT\( _x \) is given by
\[
\hat{x} := \arg\min_{\xi \in \mathcal{M}^+(\mathcal{Y})} \sum_{i=1}^N t_i \text{UOT}_{x, i, \psi, \xi}(\mu_i, \xi), \tag{13}
\]
\[
\begin{align*}
&= \arg\min_{\pi \in \mathcal{M}^+(\mathbb{X} \times \mathbb{Y})} \sum_{i=1}^{N} \pi^{(i)} \in \mathcal{M}^+(\mathbb{X}_i \times \mathbb{Y}) \int_{\mathbb{X}_i \times \mathbb{Y}} t_i c_i \, d\pi^{(i)} \\
&\quad + \varepsilon \mathcal{E}(\pi^{(i)}) + t_i D_{\psi}(P_{\pi_i^{(i)}, \mu_i}), \\
&\text{subject to } P_{\xi^{(i)}, \pi^{(i)}} = \xi, \ i = 1, \ldots, N.
\end{align*}
\]

Note that by the choice of \( \psi \) the barycentric marginal \( \xi \) is exactly matched. By Proposition 3.3, the involved UOT problems have unique solutions. Moreover, it was shown in [12, Sec. 5.2] that a unique barycenter exists due to the regularization. However, these barycenters do not correspond to a shortest path since UOT\( \varepsilon \) is not a metric.

To establish a relation with the multi-marginal setting, we exploit that the optimal plan of the multi-marginal problem \( \pi \) is readily determined by its marginals. We need the following auxiliary lemma.

**Lemma 5.1** Let \( \pi \in \mathcal{M}^+(\mathbb{X} \times \mathbb{Y}) \) be absolutely continuous with respect to Lebesgue measure, respectively, counting measure, on \( \mathbb{X} \times \mathbb{Y} \) with density \( \sigma_\pi \). If there exists \( a_i \in L^\infty(\mathbb{X}_i \times \mathbb{Y}), \ i = 1, \ldots, N \), such that \( \sigma_\pi = \prod_{i=1}^{N} a_i \), then \( \pi \) is related to its marginals \( \pi_i := P_{\mathbb{X}_i \times \mathbb{Y}, \pi, \mu_i}, i = 1, \ldots, N \), and \( \pi_{N+1} := P_{\mathbb{X}_i \times \mathbb{Y}, \pi} \) via

\[
\sigma_\pi = \frac{\sigma_{\pi_1} \cdots \sigma_{\pi_N}}{\sigma_{\pi_{N+1}}},
\]

where \( \sigma_{\pi_i} \) denotes the density of \( \pi_i \) with respect to the Lebesgue, respectively, counting measure, on \( \mathbb{X}_i \times \mathbb{Y} \).

**Proof** By abuse of notation, we denote the Lebesgue measure, respectively, counting measure, by \( \lambda \). The underlying space becomes clear from the context. By assumption, the marginal densities of \( \pi \) read

\[
\sigma_{\pi_j}(x_j, y) = a_j(x_j, y) \prod_{i \neq j} \int_{\mathbb{X}_i} a_i(x_i, \cdot) \, d\lambda(x_i),
\]

\( j = 1, \ldots, N \). Since \( P_{\mathbb{X}_i \times \mathbb{Y}, \pi} = \pi_{N+1} \), we have \( \sigma_{\pi_{N+1}} = \prod_{i=1}^{N} \int_{\mathbb{X}_i} a_i(x_i, \cdot) \, d\lambda(x_i) \), which finally yields

\[
\sigma_\pi = \prod_{i=1}^{N} a_i = \frac{\sigma_{\pi_1} \cdots \sigma_{\pi_N}}{\prod_{i=1}^{N} \int_{\mathbb{X}_i} a_i(x_i, \cdot) \, d\lambda(x_i))} = \frac{\sigma_{\pi_1} \cdots \sigma_{\pi_N}}{\sigma_{\pi_{N+1}}^\circ}. \]

Now, we can draw the aforementioned connection between the barycenter problem (13) and UOMOT\( \varepsilon \), where

\[
\varphi_{N+1} := (0, \infty),
\]

\[
D_{\varphi_{N+1}}(\cdot, \xi) = 0, \ \xi \in \mathcal{M}^+(\mathbb{Y}),
\]

see Example 2.1ii). Due to the special form of this entropy, the \( (N+1) \)th input of UOMOT\( \varepsilon \) has no effect on the functional. To emphasize this, we use the notation

\[
\text{UOMOT}_{\varepsilon}(\varphi, \psi, \varphi_{N+1})(\mu, \#) \]

\[
:= \min_{\pi \in \mathcal{M}^+(\mathbb{X} \times \mathbb{Y})} \int_{\mathbb{X} \times \mathbb{Y}} c_i \, d\pi + \varepsilon \mathcal{E}(\pi)
\]

\[
+ \sum_{i=1}^{N} D_{\psi}(P_{\pi_i^{(i)}, \mu_i}). \tag{16}
\]

The next theorem establishes the relation between the barycenter problem and this multi-marginal optimal transport problem.

**Theorem 5.2** For \( i = 1, \ldots, N \), let \( \mu_i \in \mathcal{M}^+(\mathbb{X}_i) \) and cost functions \( c_i \in C(\mathbb{X}_i \times \mathbb{Y}) \) be given. Define \( c_i \in C(\mathbb{X} \times \mathbb{Y}) \) by (14), \( \psi \) by (12) and \( \varphi_{N+1} \) by (15). If \( \hat{\pi} \) is optimal for (16) then \( \hat{\pi} \) minimizes the functional \( F(\xi) \) given by

\[
\sum_{i=1}^{N} t_i \text{UOT}_{\varepsilon}^{(c_i, \varphi_i, \varphi_{N+1})}(\mu_i, \#) - \varepsilon(N - 1) \mathcal{E}(\xi). \tag{17}
\]

**Proof** Again, let \( \lambda \) denote the Lebesgue or counting measure, respectively. Set

\[
C = \varepsilon(\mathbb{Y}) \left( \sum_{i=1}^{N} \lambda(\mathbb{X}_i) - (N - 1) \prod_{i=1}^{N} \lambda(\mathbb{X}_i) \right).
\]

Estimating in both directions, we show that

\[
\inf_{\xi \in \mathcal{M}^+(\mathbb{Y})} F(\xi) = \text{UOMOT}_{\varepsilon}(\varphi, \psi, \varphi_{N+1})(\mu, \#) + C.
\]

1. Fix \( \xi \in \mathcal{M}^+(\mathbb{Y}) \) such that optimal plans \( \hat{\pi}^{(i)} \in \mathcal{M}^+(\mathbb{X}_i \times \mathbb{Y}) \) for UOMOT\( \varepsilon \)\( (c_i, \varphi_i, \psi)(\mu_i, \xi) \), \( i = 1, \ldots, N \), exist. Then, we define

\[
\pi := \frac{\sigma_{\hat{\pi}^{(1)}} \cdots \sigma_{\hat{\pi}^{(N)}}}{\sigma_{\hat{\pi}_{N+1}}} \lambda^\circ \in \mathcal{M}^+(\mathbb{X} \times \mathbb{Y}),
\]

which yields \( P_{\mathbb{X}_i \times \mathbb{Y}, \pi} = \hat{\pi}^{(i)} \) and \( P_{\mathbb{X}_i \times \mathbb{Y}, \pi} = \xi \). Consequently, we get

\[
F(\xi) = \sum_{i=1}^{N} \left( \int_{\mathbb{X}_i \times \mathbb{Y}} t_i c_i \, d\hat{\pi}^{(i)} + \varepsilon \mathcal{E}(\hat{\pi}^{(i)}) \right)
\]
\[ + \sum_{i=1}^{N} t_i D_{\psi_i}(P_{X_i}^{(i)}(\hat{\pi}_i), \mu_i) - \varepsilon(N - 1)E(\xi) \]
\[ = \int_{X \times Y} c_i \, d\pi + \varepsilon \left( \sum_{i=1}^{N} E(\hat{\pi}_i) - (N - 1)E(\xi) \right) \]
\[ + \sum_{i=1}^{N} t_i D_{\psi_i}(P_{X_i}^{(i)}(\hat{\pi}_i), \mu_i). \]

Using the definition of \( \pi \), we obtain
\[ \sum_{i=1}^{N} E(\hat{\pi}_i) - (N - 1)E(\xi) \]
\[ = \sum_{i=1}^{N} \left( \int_{X_i \times Y} \ln(\hat{\pi}_i) \, d\hat{\pi}_i + \lambda(\mathcal{X}_i \times Y) \right) \]
\[ - (N - 1) \left( \int_{Y} \ln(\sigma \xi) \, d\xi + \lambda(Y) \right) \]
\[ - \sum_{i=1}^{N} \pi(i) (\mathcal{X}_i \times Y) + (N - 1)\xi(Y) \]
\[ \geq F(\xi). \]

Incorporating \( t_i D_{\psi_i} = D_{\psi_i} \), we obtain
\[ F(\xi) = \int_{X \times Y} c_i \, d\pi + \varepsilon E(\pi) + C \]
\[ + \sum_{i=1}^{N} D_{\psi_i}(P_{X_i}^+, \mu_i). \]

Thus, minimizing the right-hand side over all \( \pi \in \mathcal{M}^+(X \times Y) \), we get
\[ F(\xi) \geq \text{UMOT}_{\mathcal{E}_e^*(\mathcal{E})}(\mu, *) + C, \]
such that minimizing the left-hand side over all \( \xi \in \mathcal{M}^+(Y) \) yields the desired estimate.

2. Next, we show the converse estimate. Let \( \hat{\pi} \in \mathcal{M}^+(X \times Y) \) be the optimal plan and \( (\hat{f}, \hat{g}) \in L^\infty(\mathcal{X}) \times L^\infty(\mathcal{Y}) \) be optimal dual potentials for \( \text{UMOT}_{\mathcal{E}_e^*(\mathcal{E})}(\mu, *) \). For \( i = 1, \ldots, N \), define \( a_i \in L^\infty(\mathcal{X}_i \times Y) \) by
\[ a_i := \exp\left( \frac{\hat{f}_i - t_i c_i}{\varepsilon} \right), \quad i = 2, \ldots, N. \]

The definition of \( c_i \) together with (5) yields
\[ \sigma_\hat{\pi} = \exp\left( \frac{\hat{f} + \hat{g} - c_1}{\varepsilon} \right) = \prod_{i=1}^{N} a_i. \]

Then, Lemma 5.1 implies
\[ \hat{\pi} = (\sigma_\hat{\pi}_1 \cdots \sigma_\hat{\pi}_N / \sigma_\xi^{N-1}) \lambda^\circ, \]
where \( \hat{\pi}_i := P_{X_i \times Y \# \hat{\pi}}, i = 1, \ldots, N, \) and \( \xi := P_{Y \# \hat{\pi}}. \) Similarly as for the previous considerations, this results in
\[ \text{UMOT}_{\mathcal{E}_e^*(\mathcal{E})}(\mu, *) + C \]
\[ = \sum_{i=1}^{N} \left( \int_{X_i \times Y} t_i c_i \, d\hat{\pi}_i + \varepsilon E(\hat{\pi}_i) \right) \]
\[ + \sum_{i=1}^{N} t_i D_{\psi_i}(P_{X_i \# \hat{\pi}_i}, \mu_i) - \varepsilon(N - 1)E(\xi) \]
\[ \geq F(\xi). \]

Minimizing the right-hand side over all \( \xi \in \mathcal{M}^+(Y) \) yields
\[ \text{UMOT}_{\mathcal{E}_e^*(\mathcal{E})}(\mu, *) + C \geq \inf_{\xi \in \mathcal{M}^+(Y)} F(\xi) \]
and thus the desired equality. As a direct consequence, we get \( P_{\hat{\pi} \# \hat{\pi}} \) minimizes (17). This concludes the proof. \( \Box \)

Remark 5.3 The previous result directly generalizes to UMOT problems that also regularize the marginal on \( Y \). More precisely, let \( \mu_{N+1} \in \mathcal{M}^+(Y) \) and \( \varphi_{N+1} \in \Gamma_0(\mathbb{R}) \) be an arbitrary entropy function. If \( \hat{\pi} \) is the optimal plan for \( \text{UMOT}_{\mathcal{E}_e^*(\mathcal{E})}(\mu, \mu_{N+1}) \), then \( P_{\hat{\pi} \# \hat{\pi}} \) solves
\[ \min_{\xi \in \mathcal{M}^+(Y)} \sum_{i=1}^{N} t_i \text{UOT}_{\mathcal{E}_e^*(\mathcal{E})}(\varphi_{N+1})(\mu_i, \xi) \]
\[ - \varepsilon(N - 1)E(\xi) + D_{\psi_{N+1}}(\xi, \mu_{N+1}). \]

Remark 5.4 (Comparison of formulations (13) and (16)) The proof of Theorem 5.2 reveals that the barycenter \( \hat{\xi} \) in (13) is “over-regularized,” since it appears as the marginal measure of \( \pi(i) \) in each of the \( N \) entropy terms \( E(\pi(i)) \). On the other hand, the proposed multi-marginal approach does not involve these \( N - 1 \) superfluous regularizers. This ensures that the minimizer \( \hat{\xi} \) is less “blurred” compared to the original UOT barycenter, which is favorable for most applications. A numerical illustration of this behavior is given in Sect. 6. We will see in Subsect. 5.3 that for tree-structured costs
the computation of optimal transport plans for the multi-
marginal case has the same complexity per iteration as for the “barycentric” problems.

Furthermore, the computation of barycenters with an addi-
tional penalty term as outlined in Rem. 5.3 is possible with the Sinkhorn-type algorithm detailed in Sect. 5.3. In con-
trast, we are unaware of an efficient algorithm to solve the corresponding pairwise coupled formulation.

On the other hand, we only obtain the equivalence of the “pairwise coupled” formulation (13) and the multi-
marginal approach for the choices made in (12) and (15). These enforce that all marginals of the plans in (13) coincide with the barycenter, which is not necessary in general. Although this generalization comes at the cost of a nested optimization problem, the pairwise coupled formulation is thus more flexi-
able than UMOT.

**Remark 5.5 (Barycenters and MOT)** By Theorem 5.2, for-

mula (17) with \( \varepsilon = 0 \) and \( \varphi_i = \{1\}, i = 1, \ldots, N \), our MOT formulation with the cost

$$
c_i(x_1, \ldots, x_{N+1}) = \sum_{i=1}^N t_i c_i(x_i, x_{N+1})
$$

is equivalent to the OT barycenter problem. There is another reformulation of the OT barycenter problem via the McCann interpolation using the cost function

$$
c(x_1, \ldots, x_N) := \min_y \sum_{i=1}^N t_i c_i(x_i, y)
$$

if a unique minimizer exists, see [10]. To the best of our knowledge, there is no similar reformulation for our setting.

### 5.2 General Tree-Structured Costs

In the above barycenter problem, we have considered UMOT with a tree-structured cost function, where the tree was just star-shaped. In the rest of this section, we briefly discuss an extension of the UMOT problem to costs of the form (11), where \( G = (V, E) \) is a general tree graph with cost functions \( c_{(j,k)} \in C(X_j \times X_k) \) for all \( (j, k) \in E \). For the balanced case, this topic was addressed in [32, Prop. 4.2].

For a disjoint decomposition

$$
V = V \cup U, \quad V \cap U = \emptyset,
$$

where \( V \) contains only leaves, and measures \( \mu_v \in \mathcal{M}^+(X_v), v \in V \), we want to find measures \( \mu_u \in \mathcal{M}^+(X_u), u \in U \), that solve the problem

$$
\inf_{\{\mu_u\}_{u \in U}} \sum_{(j,k) \in E} t_{(j,k)} \text{UOT}_{\epsilon/|E|}(c_{(j,k)}, \varphi_j, \varphi_k)(\mu_j, \mu_k).
$$

Again, we assume that the unknown marginals \( \mu_u \) are exactly matched, i.e., \( \varphi_u = \{1\}, u \in U \).

**Example 5.6** For the barycenter problem (13), we have \( V = \{1, \ldots, N\} \) and the tree is star-shaped, meaning that \( \mathcal{E} = \{(j, N+1) : j = 1, \ldots, N\} \).

Fig. 1 left. In Fig. 1 middle, we have an H-shaped tree with \( N = 7 \), edge set \( \mathcal{E} = \{(1,2), (2,3), (2,4), (4,6), (5,6), (6,7)\} \), and we consider problem (18) with \( V = \{1, 3, 5, 7\} \). Finally, Fig. 1 right shows a line-shaped tree with \( N = 4 \), edge set \( \mathcal{E} = \{(1,2), (2,3), (3,4)\} \) and \( V = \{1,4\} \). This graph is related to a so-called multiple barycenter problem, and its solution was discussed for the balanced case, e.g., in [9].

In general, it is unclear how to solve problem (18) using Sinkhorn iterations. Therefore, we propose to solve a related multi-marginal problem

$$
\text{UOT}_{\epsilon/|E|}(c_{(j,v)} \in V, (\mu_v)_{v \in U}) = \left(\mu_v\right)_{v \in \mathcal{V}}
$$

where again \( \varphi_v := \{0, \infty\} \) for all \( u \in U \) and \( t_v = t_e \) if \( e \in \mathcal{E} \) joins \( v \) with some other node (indeed well defined for leaves).

Then, we can prove in analogy to Lemma 5.1 that the optimal plan \( \tilde{\pi} \) is related to its marginals \( \tilde{\sigma}_e \) and \( \tilde{\mu}_u := P_{\mu_u \# \tilde{\pi}} \) by

$$
\tilde{\sigma}_e = \frac{\prod_{e \in \mathcal{E}} \tilde{\sigma}_{e}}{\prod_{u \in U} \tilde{\mu}_u^{\deg(u) - 1}}.
$$

Furthermore, we can show similarly as in the proof of Theorem 5.2 the following corollary.

**Corollary 5.7** Under the above assumptions, if \( \tilde{\pi} \) is the optimal plan in (19), then the \( \tilde{\mu}_u := P_{\mu_u \# \tilde{\pi}} \), \( u \in U \), solve

$$
\inf_{\mu_u \in \mathcal{M}^+(X_u)} \sum_{(j,k) \in \mathcal{E}} t_{(j,k)} \text{UOT}_{\epsilon/|E|}(c_{(j,k)}, \varphi_j, \varphi_k)(\mu_j, \mu_k)
$$

$$
-\varepsilon \sum_{u \in U} (\deg(u) - 1) E(\mu_u).
$$

### 5.3 Efficient Sinkhorn Iterations for Tree-Structured Costs

Throughout this section, let \( X_1, \ldots, X_N \) be finite subsets of \( \mathbb{R}^d \) of size \( M_i := |X_i| > 0 \), \( i = 1, \ldots, N \). Furthermore, let \( \mu_i \in \mathcal{M}^+(X_i) \) be positive measures that are identified with vectors in \( \mathbb{R}^{M_i} \). Hence, the reference measures \( \gamma_i \) are chosen as the counting measure. Recall that the Sinkhorn mapping for a cost function \( c \in \mathbb{R}^{M_1 \times \cdots \times M_N} \) is the concatenation of the approx operator and the multi-marginal \((c, \varepsilon)\)-transform. As the former is applied pointwise, its computational cost is negligible. Hence, it suffices to discuss the efficient implementa-
tion of the multi-marginal \((c, \varepsilon)\)-transform. For vectors
and matrices, we denote pointwise multiplication by \( \odot \) and pointwise division by \( \oslash \). Set
\[
K := \exp(-c/\varepsilon) \in \mathbb{R}_+^{M_1 \times \cdots \times M_N}.
\]

For efficiency reasons, we perform computations in the exp-domain, i.e., instead of the Sinkhorn iterates \((f^{(n)})_{n \in \mathbb{N}} \) in Algorithm 1 we consider
\[
\begin{align*}
  u^{(n)}(i) &:= u_1^{(n)} \odot \cdots \odot u_N^{(n)} \\
  u_i^{(n)} &:= \exp\left(\frac{f^{(n)}(i)}{\varepsilon}\right), \quad i = 1, \ldots, N.
\end{align*}
\]

Convergence of Algorithm 1 implies convergence of \( u^{(n)} \) to some \( \bar{u} \in \mathbb{R}_+^{M_1 \times \cdots \times M_N} \). By Proposition 3.4, the optimal plan \( \bar{\pi} \) is given by \( \bar{\pi} = K \odot \bar{u} \). For \( n \equiv j \mod N \) the Sinkhorn updates in the exp-domain can be written as
\[
\begin{align*}
  u_j^{(n)} &= \exp\left(-\frac{1}{\varepsilon}\text{aprox}_{\phi}\left(-\varepsilon \ln \left(\frac{u_j^{(n-1)} \odot \mu_j}{P_{X_j}(\bar{K} \odot u^{(n-1)})}\right)\right)\right),
\end{align*}
\]
where division has to be understood componentwise. Note that in this context \( P_{X_j}(\cdot) \) corresponds to summing over all but the \( j \)-th dimension. Although the involved expression \( \exp(-\frac{1}{\varepsilon}\text{aprox}_{\phi}\left(-\varepsilon \ln (\cdot)\right)) \) appears to be complicated, it simplifies for all the entropies from Example 4.5, see also [12].

As recently discussed for the balanced case in [32], multi-marginal Sinkhorn iterations can be computed efficiently if the cost function decouples according to a tree, see also [2] for a wider class of cost functions. In this section, we generalize the approach for tree-structured costs to the unbalanced setting. As in the balanced case, computing the projections \( P_{X_j}(K \odot u^{(n)}) \), \( j = 1, \ldots, N \), is the computational bottleneck of the Sinkhorn algorithm. Fortunately, the Sinkhorn iterations reduce to standard matrix-vector multiplications in our particular setting.

Consider a tree \( G = (V, E) \) as in Subsect. 5.2 and corresponding cost functions
\[
c = \bigoplus_{(j,k) \in E} c(j,k), \quad c(j,k) \in \mathbb{R}_{\geq 0}^{M_j \times M_k}.
\]

Then, it holds \( K_{i_1, \ldots, i_N} = \prod_{(j,k) \in E} K_{i_j,k}^{(j,k)} \), where
\[
K^{(j,k)} := \exp(-c^{(j,k)}(i,j)/\varepsilon) \in \mathbb{R}_+^{M_j \times M_k}.
\]

The next result, c.f. [32, Thm. 3.2], is the main ingredient for an efficient implementation of Algorithm 1 for solving UMOTs with tree-structured cost functions.

**Theorem 5.8** The projection onto the \( j \)-th marginal of \( K \odot u \) is given by
\[
P_{X_j}(K \odot u) = u_j \odot \bigoplus_{l \in N_j} \alpha_{(j,l)}.
\]

Here, the \( \alpha_{(j,k)} \) are computed recursively for \( (j,k) \in E \) := \( \{(v,w) \mid (v, w) \in E \text{ or } (w, v) \in E\} \) starting from the leaves by
\[
\alpha_{(j,k)} = K^{(j,k)}(u_k \odot \bigoplus_{l \in N_j \setminus \{j\}} \alpha_{(k,l)}),
\]
with the usual convention that the empty product is 1.

First, we traverse the tree \( G \) by a pre-order depth-first search with respect to a root \( v_0 \). This results in a strict ordering of the nodes, which is encoded in the list \( V_- \). Every node \( k \in V \) except the root has a unique parent, denoted by \( p(k) \). We denote by \( V_+ \) the reversed list \( V_- \). Note that the order in which we update the vectors \( (\alpha_{(j,k)})_{(j,k) \in E} \) and potentials \( (u_j)_{j \in V} \) in Algorithm 2 fits to the underlying recursion in (20). Furthermore, the computational complexity of Algorithm 2 is linear in \( N \). More precisely, \( 2(N-1) \) matrix-vector multiplications are performed to update every \( u_j \) once, which is in alignment with the two-marginal case. In particular, solving \( N \) decoupled problems has the same complexity per iteration with the disadvantage that the marginals of the obtained transport plans do not necessarily fit to each other. Although we discussed Algorithm 2 mainly for computing barycenters, it can also be applied without free marginals, see Sect. 6.3 for a numerical example.

### 6 Numerical Examples

In this section, we present three numerical examples, where the first two confirm our theoretical findings from Section 5. Part of our Python implementation is built upon the POT toolbox [27].

From now on, all measures are of the form \( \sum_{k=1}^m \mu^k \delta_{x_k} \) with support points \( x_k \in \mathbb{R}^d \), \( k = 1, \ldots, m \). We always use the cost functions \( c(x,y) = \|x - y\|^2 \) with corresponding cost matrices \( c = c(x_j, x_k)^{m, m}_{j,k=1} \). All reference measures are the counting measure, see Remark 3.2iii), i.e., we exclusively deal with entropy regularization.

#### 6.1 Barycenters of 1D Gaussians

We start with computing the barycenter for two simple measures \( \mu_1, \mu_2 \), which are produced by sampling truncated normal distributions \( \mathcal{N}(0.2, 0.05) \) and \( 2\mathcal{N}(0.8, 0.08) \) on \([0, 1]\) on a uniform grid. Clearly, this choice makes an
unbalanced approach necessary. As before, we denote the discrete spaces and those of the barycenter by $X_1 = X_2 = \mathcal{Y}$. To approximate the marginals, we use the Shannon entropy functions $\varphi = \varphi_1 = \varphi_2$ so that $D_{\varphi} = KL$. First, we solve the barycenter problem (13), which reads for $t_1 = 1 - t$ and $t_2 = t \in (0, 1)$ as

$$
\hat{\xi} = \arg\min_{\xi} \sum_{i=1}^{2} t_i \min\langle \varphi, \pi^{(i)} \rangle + \varepsilon E(\pi^{(i)})
+ t_i \text{KL}(P_{X_1} \pi^{(i)}, \mu_i)
\text{subject to } P_{X_1} \pi^{(1)} = P_{Y} \pi^{(2)} = \xi.
$$

The resulting barycenter for $\varepsilon = 0.005$ is computed using the Sinkhorn iterations described in [12, Sec. 5.2] and is shown in Fig. 2 for different $t \in (0, 1)$ together with the marginals $\mu_i := P_{X_1} \hat{\pi}^{(i)}, i = 1, 2$.

We compare these barycenters with the marginal $P_{Y} \hat{\pi}$ of the corresponding optimal plan $\hat{\pi}$ for the multi-marginal problem

$$
\arg\min_{\pi \in M^+((X \times Y) \times X)} \langle c, \pi \rangle + \varepsilon E(\pi)
+ \sum_{i=1}^{2} t_i \text{KL}(P_{X_1} \pi, \mu_i)
$$

computed by Algorithm 2. The resulting marginals are provided in Fig. 2. As explained in Remark 5.4, the barycenters $\hat{\xi}$ appear smoothed compared to $P_{X_1} \hat{\pi}$. As already mentioned, we do not have relations with shortest paths due to the missing metric structure.

6.2 H-tree Shaped Cost Functions

Next, we turn our attention to the “interpolation” of four gray value images of size $100 \times 100$ considered as probability measures $\mu_1, v = 1, 3, 5, 7$, along a tree that is H-shaped, see Fig. 1. The images are depicted in the four corners of Fig. 3a.

In this example, the measures corresponding to the inner nodes with $u = 2, 4, 6$ have to be computed. For this purpose, we choose $\varepsilon = 4 \cdot 10^{-4}$ and $D_{\psi}(\cdot, \mu_v) = 0.05 KL(\cdot, \mu_v)$.

Comparison with MOT $\varepsilon$ As the measures have the same mass, we can compare our proposed UMOT $\varepsilon$ approach with the balanced MOT $\varepsilon$ method. Equal costs $c$ as well as equal weights are assigned to the edges. Note that the total cost decouples according to the H-shaped tree. The obtained results for MOT $\varepsilon$ and UMOT $\varepsilon$ are depicted in Figs. 3a and b, respectively. For UMOT $\varepsilon$, the corners contain the marginals $P_{v} \hat{\pi} = P_{X_1} \hat{\pi}$ instead of the given measures. As the mass in the different image regions is different, mass is transported between them in the MOT $\varepsilon$ interpolation. In contrast, only a minimal amount of mass is transported between the images regions for UMOT $\varepsilon$, where the mass difference is compensated by only approximately matching the prescribed marginals. This behavior can be controlled by adjusting the weights in the $\varphi$-divergences.

While MOT becomes numerically unstable for smaller $\varepsilon$ than $4 \cdot 10^{-4}$, the UMOT problem remains solvable for smaller values of $\varepsilon$. In our numerical experiments, which are not reported here, this led to less blurred images.

Comparison with Multiple Star Graph Barycenters Next, we provide a heuristic comparison with an alternative interpolation approach. Instead of computing three interpolating measures simultaneously, we solve three individual barycenter problems based on star-shaped graphs with leaves corresponding to $\mu_v$, $v = 1, 3, 5, 7$. This is done both with the UOT $\varepsilon$ and corresponding UMOT $\varepsilon$ approach. More precisely, we solve (13) and (16) three times with weights $1/12 \cdot (5, 5, 1, 1), 1/4 \cdot (1, 1, 1, 1)$ and $1/12 \cdot (1, 1, 5, 5)$, respectively. These weights have been derived from the solution of the balanced H-graph problem with Dirac measures at the leaves, which is easy to obtain from a linear system corresponding to the first-order optimality conditions. The results are provided in Figs. 3c and d. Noteworthy, both interpolations have an even less pronounced mass transfer between the different image structures. However, the computed images

---

**Algorithm 2 Sinkhorn Iterations for Tree-Structured Costs**

**Input:** Tree $G = (\mathcal{V}, \mathcal{E})$ with $N$ nodes, discrete measures $\mu_i, i = 1, \ldots, N$, and cost function $c$ decoupling according to $G$

Choose a root $v_0$ and initialize $u_j = 1$ for all $j \in \mathcal{V}$

$\mathcal{V} \leftarrow \text{PreOrderDepthFirstSearch}(\mathcal{E}, \mathcal{V}, v_0)$

for $j \in \mathcal{V}_{\infty} \setminus \{v_0\}$ do

Initialize $a_{(j, i)}$ according to 20

while Sinkhorn not converged do

for $j \notin \mathcal{V}_{\infty}$ do

if $j \neq v_0$ then

Update $a_{(j, p(j))}$ according to 20

Set $u_j \leftarrow \exp\left(-\frac{1}{\varepsilon} \text{approx}^{\varepsilon}(\pi, \psi, \mu_j)\right)$

for $j \in \mathcal{V}_{\infty} \setminus \{v_0\}$ do

Update $a_{(p(j), j)}$ according to 20

Output: Optimal dual potentials $(\nu_1, \ldots, \nu_N)$ and vectors $(a_{(j, k)})_{(j, k) \in \mathcal{E}}$
Fig. 2 Comparison between the unbalanced barycenter \( \hat{\xi} \) of two Gaussians (left) and the marginal \( P_{Y \# \hat{\pi}} \) of the optimal transport plan for the corresponding multi-marginal problems (right) with different \( t \in (0, 1) \).

The input measures \( \mu_i, i = 1, 2 \) and the corresponding marginals \( \tilde{\mu}_i, i = 1, 2 \), of the respective transport plans are shown as well.
Fig. 3 Different “interpolations” between the four given images at the corners of Fig. 3a. Top: H-tree structured cost function. Bottom: Star-shaped decomposition approach. A common color coding is used for all images look considerably smoother than their counterparts in Fig. 3b. Again, as expected, we observe that the UOT results in Fig. 3d are more blurred than the corresponding UMOT interpolations in Fig. 3c. As they are all marginals of multi-marginal transport plans, the images in Figs. 3a and b have the same mass. In contrast, the images in Figs. 3c and d do not necessarily have the same mass as they are marginals of different transport plans. Hence, depending on the application, one or the other approach might be preferable.

Note that in order to compute the multiplications with the matrices \( K \) in Algorithm 2, we exploit the fact that the Gaussian kernel convolution \( K \) is separable in the two spatial dimensions and can be performed over the rows and columns of the images one after another, such that we never actually store the matrix \( K \in \mathbb{R}^{10000\times10000} \). Consequently, we cannot use stabilizing absorption iterations as proposed in [12].

6.3 Particle Tracking and Transfer Operators

Finally, we investigate whether computing a single joint UMOT solution can be beneficial compared to computing several UOT plans sequentially, e.g., for particle tracking. To this end, we create a time series of five images. Each image has size 100 \( \times \) 100 pixels and consists of “dots” by sampling uniform noise, setting all values above a small threshold to zero, and applying a Gaussian filter with small variance. One time step corresponds to shifting the image by two pixels downward filling with the small constant background value from the top, which results in images \( \mu_{\text{clean},i}, i = 1, \ldots, 5 \). We modify this time series of five images by adding dots randomly for every time step in a similar manner. Consequently, the data consist of a drift component and some random dots popping up and disappearing again. The resulting data \( \mu_i, i = 1, \ldots, 5 \), are shown in Fig. 4. As we want to extract only the drift component, we apply UMOT(\( \mu \)) for a line-tree-structured cost function with the same costs \( c_i \) along the path, regularization parameter \( \varepsilon = 10^{-4} \), and \( D_{\phi_i}(\cdot, \mu_i) = 7 \cdot 10^{-4} \text{TV}((\cdot, \mu_i), i = 1, \ldots, 5 \). We expect that the hard thresholding of the corresponding approx-operators for \( \text{TV}(\cdot, \mu_i), i = 1, \ldots, 5 \), is particularly well suited for removing the noise in our example, see also [8,26]. The resulting marginals of the optimal transport plan \( P_{X_{\mu \#}} \hat{\pi} \) are shown in Fig. 4. Indeed, the method manages to remove most of the noise dots.
Transfer Operators For our next comparison, we use OT-related transfer operators, which have recently been discussed in [36]. We abuse notation in this section by sometimes identifying empirical measures with their vectors or matrices of weights for simplicity. In a nutshell, assuming a discrete, two-marginal setting of probability measures with optimal transport matrix $\pi$ and left marginal vector $\mu_\ell = (\mu_{\ell k})_{k=1}^m$, we can define a row-stochastic transition matrix $K$ by setting

$$K := \text{diag} \left( \mu_{\ell}^{-1} \right) \pi.$$

This concept allows us to propagate other measures than $\mu_\ell$ forward in time by multiplication with $K^T$. Note that there is a continuous analog in terms of Frobenius–Perron operators, Markov kernels, and the disintegration theorem; see [6,7,30,35,38] for details.
Now, we compute the marginal \( \tilde{\pi}_{1,5} := P_{X_1 \times X_5} \tilde{\pi} \) of the optimal UMOT\(_\varepsilon\)(\(\mu\)) transport plan \( \tilde{\pi} \). Using the marginal \( \tilde{\mu}_1 := P_{X_1} \tilde{\pi}_{1,5} = P_{X_1} \tilde{\pi} \), we get the transfer operator

\[
K_{\text{UMOT}} = \text{diag} \left( \tilde{\mu}_1^{-1} \right) \tilde{\pi}_{1,5}.
\]

Then, we propagate the first clean image \( \mu_{\text{clean}, 1} \) by this transfer operator, i.e., we compute \( K_{\text{UMOT}}^T \mu_{\text{clean}, 1} \). The result is shown in Fig. 5.

For comparison, we also compute \( N-1 \) successive UOT\(_\varepsilon\)(\(\mu_1, \mu_{i+1} \)) plans \( \tilde{\pi}^{(i)}, i = 1, \ldots, N \). Denoting the marginals by \( \tilde{\mu}_i := P_{X_i} \tilde{\pi}^{(i)}, i = 1, \ldots, N \), we consider the transfer kernel

\[
K_{\text{UOT}} = \prod_{i=1}^{N} \text{diag} \left( \tilde{\mu}_i^{-1} \right) \tilde{\pi}^{(i)}.
\]

Then, we transfer the clean first image by this operator, i.e., we compute \( K_{\text{UOT}}^T \mu_{\text{clean}, 1} \). The obtained results are shown in Fig. 5. Note that similarly as described in the previous subsection, the computations can be carried out using separable convolutions without ever storing the large matrix \( K \).

As we wanted to extract the drift behavior using only the noisy images, the propagated images should be compared to \( \mu_{\text{clean}, 5} \), i.e., the last image of the first row in Fig. 4, which is just a shifted version of the first image. In some sense, we can interpret this image as the propagation using the “true” transfer operator. There are considerably less artifacts visible in the UOT\(_\varepsilon\) propagated image compared to the UOT\(_\varepsilon\) version. This is particularly pronounced in the middle left part of the images. As an error measure, we also computed the squared Euclidean distances between the propagated images and the ground truth, which are 2.98 and 6.48 for the UOT\(_\varepsilon\) and UOT\(_\varepsilon\) version, respectively.

From an intuitive point of view, the results are not surprising. If we are only provided with a single pair of images from the sequence, it is much harder to judge which points correspond to noise than for a whole sequence of images. Note that a single Sinkhorn iteration for the coupled UOT\(_\varepsilon\) problem has the same computational complexity as for all of the decoupled UOT\(_\varepsilon\) problems together. Hence, the UOT\(_\varepsilon\) approach appears to be better suited for this application, as it incorporates more information about the problem without significant additional computational cost.

### 7 Conclusions

In this paper, we introduced a regularized unbalanced multi-marginal optimal transport framework, abbreviated UMOT\(_\varepsilon\), bridging the gap between regularized unbalanced optimal transport and regularized multi-marginal optimal transport. We outlined how the Sinkhorn algorithm can be adapted to solve UMOT\(_\varepsilon\) efficiently for tree-structured costs. For this case, we have also shown how UMOT\(_\varepsilon\) provides alternative solutions of barycenter-like problems with desirable properties, such as improved sharpness. In the future, we plan to examine UMOT\(_\varepsilon\) in connection with particle cluster tracking methods, e.g., following the ideas in [36]. Furthermore, it would be interesting to examine a reformulation of the regularized unbalanced barycenter problem as UMOT problem using a cost function similar to that in Remark 5.5. Additionally, we want to investigate the theoretical relation between the two interpolation approaches from Sect. 6.2. Finally, we are interested in UOT\(_\varepsilon\) for measures having the same moments up to a certain order and for measures living on special manifolds such as, e.g., torus or spheres, see also [22].

### Acknowledgements

The datasets generated during and analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request. Funding by the DFG under Germany’s Excellence Strategy – The Berlin Mathematics Research Center MATH+ (EXC-2046/1, Projektnummer: 390685689) and by the DFG Research Training Group DAEDALUS (RTG 2433) is acknowledged.

### Funding

Open Access funding enabled and organized by Projekt DEAL.

### Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

### References

1. Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. SIAM J. Math. Anal. 43(2), 904–924 (2011)
2. Altschuler, J.M., Boix-Adsera, E.: Polynomial-time algorithms for multimarginal optimal transport problems with decomposable structure. arXiv:2008.03006, (2020)
3. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In Proc. of Machine Learning, volume 70, pages 214–223. PMLR, (2017)
4. Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative Bregman projections for regularized transportation problems. SIAM J. Sci. Comput. 37(2), A1111–A1138 (2015)
5. Benamou, J.-D., Carlier, G., Nenna, L.: A numerical method to solve multi-marginal optimal transport problems with Coulomb cost. In: Splitting Methods in Communication, Imaging, Science, and Engineering, pp. 577–601. Springer, Cham (2016)
6. Boyarsky, A., Góra, P.: Laws of Chaos. Probability and its Applications. Birkhäuser, Boston (1997)
7. Brin, M., Stuck, G.: Introduction to Dynamical Systems. Cambridge University Press, UK (2002)
8. Caffarelli, L.A., McCann, R.J.: Free boundaries in optimal transport and Monge-Ampère obstacle problems. Ann. of Math. (2), 171(2), 673–730 (2010)
9. Caillé, C.: Asymptotical estimates for some algorithms for data and image processing: a study of the Sinkhorn algorithm and a numerical analysis of total variation minimization. PhD Thesis, École Polytechnique, Paris, (2020)
10. Carlier, G., Eckeland, I.: Matching for teams. Econ. Theory 42(2), 397–418 (2010)
11. Chen, Y., Karlsson, J.: State tracking of linear ensembles via optimal mass transport. IEEE Contr. Syst. Lett. 2(2), 260–265 (2018)
12. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Scaling algorithms for unbalanced optimal transport problems. Math. Comp. 87(314), 2563–2609 (2018)
13. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: Unbalanced optimal transport: dynamic and Kantorovich formulations. J. Funct. Anal. 274(11), 3090–3123 (2018)
14. Chung, N.-P., Phung, M.-N.: Barycenters in the Hellinger-Kantorovich space. Appl. Math. Optim, to appear
15. Clason, C., Lorenz, D., Mahler, H., Wirth, B.: Entropic regularization of continuous optimal transport problems. J. Math. Anal. Appl. 494(1), 124432 (2021)
16. Cohen, S., Kumar, K.S.S., Deisenroth, M.P.: Sliced multi-marginal optimal transport. arXiv:2102.07115, (2021)
17. Colombo, M., De Pascale, L., Di Marino, S.: Multimarginal optimal transport maps for one-dimensional repulsive costs. Ann. Math. Statist. 87(314), 2563–2609 (2018)
18. Cuturi, M.: Sinkhorn distances: Lightspeed computation of optimal transport. In Advances in Neural Information Processing Systems 26, pages 2292–2300. Curran Associates, Inc., (2013)
19. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In Proc. of Machine Learning Research 32(2), 685–693. PMLR, (2014)
20. Di Marino, S., Gerolin, A.: An optimal transport approach for the Schrödinger bridge problem and convergence of Sinkhorn algorithm. J. Sci. Comput. 85, 27 (2020)
21. Dyurechenskii, P., Dvinskikh, D., Gasnikov, A., Uribe, C., Nedić, A.: Decentralize and randomize: Faster algorithm for Wasserstein barycenters. In Advances in Neural Information Processing Systems 31, pages 10760–10770. Curran Associates, Inc., (2018)
22. Ehler, M., Gräf, M., Neumayer, S., Steidl, G.: Curve based approximation of manifolds by discrepancy minimization. Foundations of Computational Mathematics, (2021)
23. Eckeland, I., Temam, R.: Convex Analysis and Variational Problems. SIAM, Philadelphia (1999)
24. Elvander, F., Haasler, I., Jakobsson, A., Karlsson, J.: Multi-marginal optimal transport using partial information with applications in robust localization and sensor fusion. Signal Process. 171, 107474 (2020)
25. Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S., Trouvè, A., Peyré, G.: Interpolating between optimal transport and MMD using Sinkhorn divergences. In Proc. of Machine Learning Research 89, 2681–2690. PMLR, (2019)
26. Figalli, A.: The optimal partial transport problem. Arch. Ration. Mech. Anal. 195(2), 533–560 (2010)
27. Flamary, R., Courty, N.: POT Python Optimal Transport library. https://github.com/PythonOT/POT, (2017), Accessed: 03.03.2021
28. Frießecke, G., Matthes, D., Schmitzer, B.: Barycenters for the Hellinger-Kantorovich distance over . SIAM J. Math. Anal. 53(1), 62–110 (2021)
29. Frogner, C., Zhang, C., Mobahi, H., Araya, M., Poggio, T.A.: Learning with a Wasserstein loss. In Advances in Neural Information Processing Systems 28, pages 2053–2061. Curran Associates, Inc., (2015)
30. Froyland, G.: An analytic framework for identifying finite-time coherent sets in time-dependent dynamical systems. Phys. D 250, 1–19 (2013)
31. Gerolin, A., Kausamo, A., Rajala, T.: Multi-marginal entropy-transport with repulsive cost. Calc. Var. Partial Differ. Equ. 59(3), 90 (2020)
32. Haasler, I., Ringh, A., Chen, Y., Karlsson, J.: Multimarginal optimal transport with a tree-structured cost and the Schrödinger bridge problem. SIAM J. Control and Optimiz. 59(4), 2428–2453 (2021)
33. Janati, H., Cuturi, M., Gramfort, A.: Debiased Sinkhorn barycenters. In Proc. of Machine Learning Research 119, 4692–4701.PMLR, (2020)
34. Janati, H., Muzellec, B., Peyré, G., Cuturi, M.: Entropic optimal transport between unbalanced Gaussian measures has a closed form. In Advances in Neural Information Processing Systems 33, pages 10468–10479. Curran Associates, Inc., (2020)
35. Klus, S., Nüske, F., Koltai, P., Wu, H., Kevrekidis, I., Schütte, C., Noé, F.: Data-driven model reduction and transfer operator approximation. J. Nonlinear Sci. 28(3), 985–1010 (2018)
36. Koltai, P., von Lindheim, J., Neumayer, S., Steidl, G.: Transfer operators from optimal transport plans for coherent set detection. Phys. D 426, 132980 (2021)
37. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In Proc. of Machine Learning Research 37, 957–966. PMLR, (2015)
38. Lasota, A., Mackey, M.: Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics. Applied Mathematical Sciences, vol. 97. Springer, New York (1994)
39. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new Helly-Kantorovich distance between positive measures. Invent. Math. 211(3), 969–1117 (2018)
40. Luise, G., Salzo, S., Pontil, M., Clielli, C.: Sinkhorn barycenters with free support via Frank-Wolfe algorithm. In: Advances in Neural Information Processing Systems 32, pages 9322–9333. Curran Associates, Inc., (2019)
41. Neumayer, S., Steidl, G.: From optimal transport to discrepancy. To appear in Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging, (2020)
42. Pass, B.: Multi-marginal optimal transport: theory and applications. ESAIM Math. Model. Numer. Anal. 49(6), 1771–1790 (2015)
43. Quang, M.H.: Entropic regularization of Wasserstein distance between infinite-dimensional Gaussian measures and Gaussian processes. J. Theor. Probab., (2022)
44. Ramdas, A., Trillos, N.G., Cuturi, M.: On Wasserstein two-sample testing and related families of nonparametric tests. Entropy, 19(2), (2017)
45. Rockafellar, R.T.: Integrals which are convex functionals. Pacific J. Math. 274, 132980 (2021)
46. Sinkhorn, R.T.: Integrals which are convex functionals. Pacific J. Math. 24, 525–539 (1968)
47. Séjourné, T., Feydy, J., Vialard, F.-X., Trouvè, A., Peyré, G.: Sinkhorn divergences for unbalanced optimal transport. arXiv:1910.12958, (2019)
48. Wang, W., Slepčev, D., Basu, S., Ozolek, J., Rohde, G.: A linear optimal transportation framework for quantifying and visualizing variations in sets of images. Int. J. Comput. Vis. 101, 254–269 (2013)
49. Zhu, L., Yang, Y., Haker, S., Tannenbaum, A.: An image morphing technique based on optimal mass preserving mapping. IEEE Trans. Image Process. 16, 1481–95 (2007)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Florian Beier studied Mathematics at the Technical University Berlin. After receiving his M.Sc. in 2021, he started his Ph.D. under the supervision of Gabriele Steidl. His main research interest is optimal transport.

Johannes von Lindheim received his B.Sc. degree in Mathematics in 2016 and M.Sc. degree in 2018 at Freie Universität Berlin. He is working as a Ph.D. student in the group of Prof. Gabriele Steidl at Technische Universität Berlin. His current interests are computational optimal transport and spectral methods in dynamical systems.

Sebastian Neumayer received his Ph.D. in Mathematics from Technische Universität Berlin. Currently, he is a Postdoctoral researcher at the Biomedical Imaging Group at EPFL Lausanne. His research interests include numerical analysis, optimization, inverse problems, and mathematics of machine learning, all with applications in imaging.

Gabriele Steidl received her Ph.D. and habilitation in Mathematics from the University of Rostock (Germany). She held a position as an Assistant Professor at the TU Darmstadt (Germany, 1993–1996) and as a Full Professor at the University of Mannheim (Germany, 1996–2010). From 2011 to 2020, she was a Professor at the TU Kaiserslautern (Germany) and a consultant of the Fraunhofer Institute for Industrial Mathematics. Since 2020, she has been a Professor at the Department of Mathematics at the TU Berlin. She worked as a postdoc at the University of Debrecen (Hungary), the University of Zürich, and was a Visiting Professor at the ENS Paris/Cachan, the Université Paris East Marne-la-Vallée and the Sorbonne. She is the Editor-in-Chief of the SIAM Journal of Imaging Sciences and SIAM Fellow. Her research interests include harmonic analysis, optimization, inverse problems, and machine learning with applications in image processing.