

SCALING ALGORITHMS FOR UNBALANCED OPTIMAL TRANSPORT PROBLEMS

LÉNAÏC CHIZAT, GABRIEL PEYRÉ, BERNHARD SCHMITZER,
AND FRANÇOIS-XAVIER VIALARD

ABSTRACT. This article introduces a new class of fast algorithms to approximate variational problems involving unbalanced optimal transport. While classical optimal transport considers only normalized probability distributions, it is important for many applications to be able to compute some sort of relaxed transportation between arbitrary positive measures. A generic class of such “unbalanced” optimal transport problems has been recently proposed by several authors. In this paper, we show how to extend the now classical entropic regularization scheme to these unbalanced problems. This gives rise to fast, highly parallelizable algorithms that operate by performing only diagonal scaling (i.e., pointwise multiplications) of the transportation couplings. They are generalizations of the celebrated Sinkhorn algorithm. We show how these methods can be used to solve unbalanced transport, unbalanced gradient flows, and to compute unbalanced barycenters. We showcase applications to 2-D shape modification, color transfer, and growth models.

1. INTRODUCTION

Optimal transport (OT) is a standard way to lift a metric defined on some “ground” space X to a metric on probability distributions (positive Radon measures with unit mass) $\mathcal{M}_+(X)$. Initially formulated by Monge [53] as a nonconvex optimization problem over transport maps, its modern formulation as a linear program is due to Kantorovitch [40], and it has been revitalized thanks to the groundbreaking work of Brenier [15]. We refer to the monographs [67, 76] for a more detailed background on the theory of OT.

While initially developed by theoreticians, OT is now becoming popular in applied fields, and we refer, for instance, to applications for color manipulation in image processing [61], reflectance interpolation in computer graphics [12], image retrieval in computer vision [66] and statistical inference in machine learning [74]. A key limitation of classical OT is that it requires the input measures to be normalized to unit mass, which is a problematic assumption for many applications that require either to handle arbitrary positive measures and/or to allow for only partial displacement of mass. All these applications might indeed benefit from OT algorithms that can handle mass variation (creation or destruction) as well as mass transportation.

While many proposals have been made to account for these “unbalanced” optimal transport problems, they used to be tailored for specific applications. There

Received by the editor October 25, 2016, and, in revised form, May 22, 2017.

2010 *Mathematics Subject Classification.* Primary 90C25; Secondary 65K10, 68U10.

Key words and phrases. Optimal transport, Wasserstein distance, unbalanced transport, Bregman projections, Wasserstein barycenters.

have been several recent works (reviewed below) that try to put all these proposals (and some new ones) into a common generic framework. These emerging theoretical advances call for algorithms that extend existing fast OT methods to these new unbalanced problems. It is precisely the goal of this paper to show how a popular numerical approach to OT—namely entropic regularization—does extend in a very natural and efficient way to solve a variety of unbalanced problems, including unbalanced OT, barycenters and gradient flows.

1.1. Previous works.

1.1.1. *Algorithms for optimal transport.* The Kantorovitch formulation of OT as a linear program, when restricted to sums of Diracs, can be directly tackled using simplex or interior point methods. For the special case of optimal linear assignment (i.e., for sums of the same number of uniform-mass Diracs) one can also use combinatorial algorithms such as the Hungarian method [18] or the auction algorithm [11]. The time complexity of these algorithms is roughly cubic in the number of Diracs, and hence they do not scale to very large problems.

In the specific case of the squared Euclidean cost, it is possible to make use of the geodesic structure of the OT distance, and reparametrize this as a convex optimization problem, as proposed by Benamou and Brenier [7]; see also [56] for a discussion on the use of first order nonsmooth optimization schemes. In [68] a multi-scale algorithm is developed that consistently leverages the structure of geometric transport problems to accelerate linear program solvers.

A last class of approaches deals with semidiscrete problems, when one measure has a density, and the second one is a weighted sum of Dirac masses. This problem, introduced by Alexandrov and Pogorelov as a theoretical tool, can be solved with geometric tools when using the squared Euclidean cost [4] (see also [52] for the development of efficient algorithms using methods from computational geometry and [46] for 3-D computations).

1.1.2. *Entropic transport.* These computational methods thus cannot cope with large scale problems with arbitrary transportation costs. A recent class of approaches, initiated and revitalized by the paper of Marco Cuturi [26], proposes to compute an approximate transport coupling using entropic regularization. This idea has its origins in many different fields, most notably it is connected with Schrödinger's problem in statistical physics [45, 70] and with the iterative scaling algorithm by Sinkhorn [72] (also known as IPFP [29]) which, given a square matrix with positive entries, aims at finding two vectors of positive numbers—so-called scalings—that make it a bistochastic matrix after multiplying rows and columns by these vectors. This entropic smoothing can be interpreted as a strictly convex barrier for positivity, but its main computational advantage is that it leads to very simple closed form expressions for all steps of the algorithm, which would not be possible when using different regularization functionals. Several follow-up articles [8, 27] to [26] have shown that the same strategy can also be used to tackle the computation of barycenters for the Wasserstein distance (as initially formulated by [2]), and for solving OT problems on geometric domains (such as regular grids or triangulated meshes) using convolution and the heat diffusion kernel [73]. Some theoretical properties of this regularization are studied in [20], including the Γ -convergence of the regularized problem toward classical transport when regularization vanishes.

1.1.3. *Optimization with Bregman divergences.* The success of this entropic regularization scheme is tightly linked with use of the Kullback-Leibler (KL) divergence as a natural Bregman divergence [14] for the optimization on the space of positive Radon measures. Not only is this divergence quite natural, but it also leads to simple formulas for the computation of projectors and so-called proximal operators (see below for a definition) for many functions typically involved in OT. The most simple algorithm, which is actually at the heart of Sinkhorn’s iterations, is the iterative projection on affine subspaces for the KL divergence. A refined version of these iterations, which works for arbitrary convex sets (not just affine spaces) is the so-called Dykstra’s algorithm [30], which can be interpreted (just like iterative projections) as an iterative block-coordinates minimization on a dual problem. Dykstra’s algorithm is known to converge when used in conjunction with Bregman divergences [5, 22] (see [58] for details on the underlying idea for sums of two arbitrary functions). Many other first order proximal methods for Bregman divergences exists. The most simple one is the proximal point algorithm [31], but most proximal splitting schemes have been extended to this setting, such as, for instance, ADMM [78], primal-dual schemes [23] and forward-backward [54].

The algorithm we propose in this article can be seen as special instance of Dykstra’s iterations, but with an extremely simple (both conceptually and algorithmically) structure, which we refer to as a “scaling” algorithm. It extends Sinkhorn’s iterations to more complex problems. This structure is due to the fact that the functions involved in OT problems make use of the marginals of the couplings that are being optimized.

1.1.4. *Unbalanced transport: from theory to numerics.* There has been a large number of proposals to extend OT methods to arbitrary “unbalanced” positive measures. Let us, for instance, quote the Kantorovitch-Rubinstein dual-Lipschitz norms [38], optimal partial transport [33] and geodesic computations with source terms [49, 59]. Most of these approaches, and much more, can be seen as special instances of a generic class of OT-like problems, that have been proposed independently in [47] and [24]. These unbalanced problems can be formulated in several ways, that are equivalent (under some restrictive conditions on the cost): a geodesic (dynamic) formulation with a source term [25, 41, 48], a static formulation with two semicouplings [24] and a static formulation with approximate marginal constraints [47]. From a numerical perspective, the last formulation (approximate marginals constraints) is the most simple to handle, since, as we show in the following section, it only involves a minor modification of the initial linear program, that has to be turned into a convex problem involving φ -divergences (see Definition 2.2). This is the one that we consider in this article. Note that the use of such a relaxed formulation, in conjunction with entropic smoothing has been introduced, without proof of convergence, in [35] for application in machine learning.

1.1.5. *Gradient flows.* Beyond OT problems and barycenter problems, a popular use of Wasserstein distances is to study gradient flows, following the formalism of “minimizing movements” detailed in [3]. This corresponds to discrete implicit stepping (i.e., proximal maps) for the Wasserstein distance (instead of more common Euclidean or Hilbertian metrics). In some cases, these time-discrete flows can be shown to converge to continuous flows that solve a suitable PDE, as the temporal stepsize tends to 0. The most famous example is the gradient flow of the entropy for

the Wasserstein metric, which solves the diffusion equation [39]. Nonlinear PDEs are considered for example in [55]. An application to imaging can be found in [17]. Another use of these implicit steps is to construct minimizing flows of nonsmooth functionals, for instance, to model crowd motions [50].

A large variety of dedicated numerical schemes has been proposed for spatial discretization and solving of these time-discrete flows such as, for instance, finite differences [16], finite volumes [21] and Lagrangian schemes [9, 10]. A bottleneck of these schemes is the high computational complexity due to the resolution of an OT-like problem at each time step. The use of entropic regularization has been proposed recently in [58] and studied theoretically in [20]. A chief advantage of this approach is that each step can be solved efficiently on a regular grid with fast Sinkhorn iterations involving only Gaussian convolutions, but this comes at the price of additional diffusivity introduced by the approximation, which makes it unsuitable to capture sharp features of solutions.

It is possible to extend these gradient flows by replacing the Wasserstein distance by more general unbalanced distances. This allows to define flows over arbitrary positive measures, hence involving creation and destruction of mass, which is crucial to model growth phenomena, such as, for instance, the Hele-Shaw model of tumor evolution [57]. An analysis of such flows based on a splitting scheme has been recently provided in [37]. It is one of the goals of this article to propose a versatile algorithm, based on iterative scalings, to approximate numerically these unbalanced flows.

1.2. Contributions and outline. The main contribution of this article is to define a class of iterative scaling algorithms to solve the entropic approximation of a variety of unbalanced optimal transport problems. First, in Section 2, we exhibit a common structure to most optimization problems related to OT: a nonnegativity constraint, a linear transport cost and convex functions acting on the marginals of the optimized couplings. For solving the entropic regularization of these problems, we then introduce in Section 3 a generic “scaling” algorithm which is a direct generalization of Sinkhorn’s algorithm. In a continuous setting, we show under some assumptions that the iterates are well-defined and correspond to alternating optimization on the dual and we prove linear convergence in a particular key case. In a discrete setting, we show in Section 4 that this algorithm converges as soon as the dual problem is well-posed. We also give a simple description of the algorithm, introduce a stabilization scheme to reach very small values of the regularization parameter and sketch a generalization of this algorithm. Finally, in Section 5, we showcase the application of these methods to 1-D and 2-D unbalanced optimal transport, generalizations of Wasserstein barycenters and gradient flows with growth. The main advantages of our approach are that it is simple (it only involves matrix multiplication and pointwise elementary operations), quite generic (it applies to most known OT-like problems), enjoys fast convergence (linear convergence is observed, and shown in particular cases) and is highly parallelizable. The code to reproduce the results of this article is available online¹. Note also that an efficient numerical implementation of the scaling algorithm developed in this paper is studied further in [69]; see Section 4.4.4 for a discussion.

¹<https://github.com/lchizat/optimal-transport>

1.3. Notation. The space of nonnegative finite Radon measures² on a (Hausdorff) topological space T is denoted by $\mathcal{M}_+(T)$, and the vector space it generates by $\mathcal{M}(T)$. For a measure on a product space $\mu \in \mathcal{M}(X \times Y)$, $P_{\#}^X \mu$ (or sometimes $P_{\#}^1 \mu$ if $X = Y$) denotes its first marginal and $P_{\#}^Y \mu$ (or $P_{\#}^2 \mu$) its second marginal.

We denote by dt, dx, \dots the reference measures on some measurable spaces T, X, \dots . They are not assumed to be the Lebesgue measure (but most of the time, they are probability measures). If (T, dt) is a measured space, we denote by $L^1(T)$ and $L^\infty(T)$ the usual spaces of equivalence classes of measurable functions from T to \mathbb{R} which are, respectively, absolutely integrable and essentially bounded. When there is a subscript $+$ appended to the name of a space of functions, it refers to the cone of nonnegative functions in that space.

Divergence functionals (or φ -divergences) are denoted by a calligraphic letter \mathcal{D} when they act on measures (as defined in Definition 2.2), by straight letters D when they act on functions (as defined in (5.2)) and with an overline \overline{D} when they act on two real numbers (as in (5.2)). More generally, functionals on measures are denoted by calligraphic letters $(\mathcal{F}, \mathcal{G}, \dots)$ and functionals on functions by straight capital letters (F, G, \dots) .

The same convention holds for the Kullback-Leibler divergence (which is just a particular case of divergence functionals) thus denoted by \mathcal{KL} , KL or $\overline{\text{KL}}$ depending on the nature of its arguments. Moreover, we generalize the definition of KL to families of functions as follows: if (X, dx) and (Y, dy) are two measured spaces and $r, s : X \times Y \rightarrow \mathbb{R}^n$ are families of measurable functions, the KL -divergence is defined as

$$(1.1) \quad \text{KL}(r|s) \stackrel{\text{def.}}{=} \sum_i \int_{X \times Y} \left[\log \left(\frac{r_i(x, y)}{s_i(x, y)} \right) r_i(x, y) - r_i(x, y) + s_i(x, y) \right] dx dy$$

(with $0 \log(0/0) = 0$) if $r, s \geq 0$ a.e. and $(s_i(x, y) = 0) \Rightarrow (r_i(x, y) = 0)$ a.e., and ∞ otherwise.

The generalization of some notations to families of functions is often implicit. For instance, if $(r_i)_{i=1}^n$ and $(s_i)_{i=1}^n$ are two families of functions, we write

$$\langle r, s \rangle \stackrel{\text{def.}}{=} \sum_{i=1}^n \int_{X \times Y} r_i(x, y) s_i(x, y) dx dy$$

and the projection operators are defined componentwise, for $i \in \{1, \dots, n\}$, as

$$(1.2) \quad (P_{\#}^X r)_i(x) = \int_Y r_i(x, y) dy \quad \text{and} \quad (P_{\#}^Y r)_i(y) = \int_X r_i(x, y) dx.$$

If X and Y are finite spaces (i.e., contain only a finite number of points), we represent functions on X by vectors denoted by bold letters \mathbf{a} and functions on $X \times Y$ by matrices denoted by capital bold letters \mathbf{A} . In this context the notations $\mathbf{a} \odot \mathbf{b}$ and $\mathbf{a} \oslash \mathbf{b}$ denote, respectively, entrywise multiplication and entrywise division with convention $0/0 = 0$ between vectors.

The conjugate of a convex function f is denoted by f^* and its subdifferential is denoted by ∂f . Some reminders on convex analysis are given in Appendix A.1. The indicator of some convex set C is

$$\iota_C(a) = \begin{cases} 0 & \text{if } a \in C, \\ +\infty & \text{otherwise.} \end{cases}$$

²Borel measures which are inner regular. If T is Polish, all Borel measures are inner regular.

Finally, the proximal operator plays a central role in this article. In full generality, if E a set, $D : E \times E \rightarrow \mathbb{R}_+ \cup \{\infty\}$ is a divergence which measures “closeness” between points in E , it is defined for $F : E \rightarrow \mathbb{R} \cup \{\infty\}$ and $x \in E$ as

$$\text{prox}_F^D(x) = \underset{y \in E}{\operatorname{argmin}} F(y) + D(y|x).$$

2. UNIFYING FORMULATION OF TRANSPORT-LIKE PROBLEMS

We review below a variety of variational problems related to optimal transport that can all be recast as a generic variational problem, involving transport and functions on the marginals. The notion of “divergence” functionals, introduced in optimal transport by [47], makes this unification possible and is defined in the following.

2.1. Divergence functionals. Divergences are functionals which, by looking at the pointwise “ratio” between two measures, give a sense of how close they are. They have nice analytical and computational properties and are built from *entropy functions*.

Definition 2.1 (Entropy function). A function $\varphi : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ is an entropy function if it is lower semicontinuous, convex, $\operatorname{dom} \varphi \subset [0, \infty[$ and satisfies the following feasibility condition: $\operatorname{dom} \varphi \cap]0, \infty[\neq \emptyset$. The speed of growth of φ at ∞ is described by

$$\varphi'_\infty = \lim_{x \uparrow +\infty} \varphi(x)/x \in \mathbb{R} \cup \{\infty\}.$$

If $\varphi'_\infty = \infty$, then φ grows faster than any linear function and φ is said *superlinear*. Any entropy function φ induces a φ -divergence (also known as Csiszár divergence) as follows.

Definition 2.2 (Divergences). Let φ be an entropy function. For $\mu, \nu \in \mathcal{M}(T)$, let $\frac{d\mu}{d\nu}\nu + \mu^\perp$ be the Lebesgue decomposition³ of μ with respect to ν . The divergence \mathcal{D}_φ is defined by

$$\mathcal{D}_\varphi(\mu|\nu) \stackrel{\text{def.}}{=} \int_T \varphi\left(\frac{d\mu}{d\nu}\right) d\nu + \varphi'_\infty \mu^\perp(T)$$

if μ, ν are nonnegative and ∞ otherwise.

The proof of the following Proposition can be found in [47, Thm 2.7].

Proposition 2.3. *If φ is an entropy function, then \mathcal{D}_φ is jointly 1-homogeneous, convex and weakly* lower semicontinuous in (μ, ν) .*

The Kullback-Leibler divergence, also known as the relative entropy, plays a central role in this article.

Example 2.4. The Kullback-Leibler divergence $\mathcal{KL} \stackrel{\text{def.}}{=} \mathcal{D}_{\varphi_{\text{KL}}}$ is the divergence associated to the entropy function φ_{KL} , given by

$$(2.1) \quad \varphi_{\text{KL}}(s) = \begin{cases} s \log(s) - s + 1 & \text{for } s > 0, \\ 1 & \text{for } s = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

³The Lebesgue decomposition Theorem asserts that, given ν, μ admits a unique decomposition as the sum of two measures $\mu^s + \mu^\perp$ such that μ^s is absolutely continuous with respect to ν and μ^\perp and ν are singular.

When restricted to densities on a measured space, it is also the Bregman divergence [14] associated to (minus) the entropy.

The following divergences will also be considered as examples.

Example 2.5. The total variation distance $\text{TV}(\mu|\nu) \stackrel{\text{def.}}{=} |\mu - \nu|_{\text{TV}}$ is the divergence associated to:

$$(2.2) \quad \varphi_{\text{TV}}(s) = \begin{cases} |s - 1| & \text{for } s \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Example 2.6. The equality constraint $\iota_{\{=\}}(\mu|\nu)$ which is 0 if $\mu = \nu$ and ∞ otherwise, is the divergence associated to $\varphi = \iota_{\{1\}}$.

Example 2.7. One can generalize the latter and define a “range constraint”, denoted $\text{RG}_{[\alpha,\beta]}(\mu|\nu)$, as the divergence which is zero if $\alpha\nu \leq \mu \leq \beta\nu$ with $0 \leq \alpha \leq \beta \leq \infty$, and ∞ else. This is the divergence associated to $\varphi = \iota_{[\alpha,\beta]}$.

2.2. Balanced optimal transport. Using divergences, the classical “balanced” optimal transport problem can be defined as follows.

Definition 2.8 (Balanced optimal transport). Let X and Y be Hausdorff topological spaces, let $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous function. The optimal transport problem between $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$ is

$$(2.3) \quad \inf_{\gamma \in \mathcal{M}_+(X \times Y)} \int_{X \times Y} c \, d\gamma + \iota_{\{=\}}(P_{\#}^X \gamma|\mu) + \iota_{\{=\}}(P_{\#}^Y \gamma|\nu).$$

Proposition 2.9. *If c is lower bounded and (2.3) is feasible, then the infimum is attained.*

The proof of this result, as well as a general theory of optimal transport can be found in [77]. If one interprets $c(x, y)$ as the cost of transporting a unit mass from x to y , the problem is then to find the cheapest way to move a mass initially distributed according to μ toward the distribution ν . If γ is additionally constrained to be of the form $(\text{id} \times T)_{\#}\mu$ for some map $T : X \rightarrow Y$, then (2.3) is called the Monge problem.

Example 2.10. An important special case is when $X = Y$ and c is the power of a distance on X . Then, the optimal cost (i.e., the value of (2.3)) is itself the power of a distance on the space of probability measures on X . This distance is often referred to as “Wasserstein” distance and denoted by W , although this denomination is disputed. We refer to [77, Chap. 6, Bibliographical Notes] for a discussion of the historical context.

2.3. Unbalanced optimal transport. Standard optimal transport only allows meaningful comparison of measures with the same total mass: whenever $\mu(X) \neq \nu(Y)$, there is no feasible γ in (2.3). Several propositions have been made to circumvent this limitation by defining “unbalanced” transport problems in various dynamic and static formulations (see Section 1.1). The following formulation with relaxed marginal constraints is best suited for the numerical schemes presented in this article.

Definition 2.11 (Unbalanced optimal transport [47, Problem 3.1]). Let X and Y be Hausdorff topological spaces, let $c : X \times Y \rightarrow [0, \infty]$ be a lower semicontinuous

function and let $\mathcal{D}_{\varphi_1}, \mathcal{D}_{\varphi_2}$ be two divergences over X and Y , as in Definition 2.2. For $\mu \in \mathcal{M}_+(X)$ and $\nu \in \mathcal{M}_+(Y)$, the unbalanced soft-marginal transport problem is

$$(2.4) \quad \inf_{\gamma \in \mathcal{M}_+(X \times Y)} \int_{X \times Y} c \, d\gamma + \mathcal{D}_{\varphi_1}(P_{\#}^X \gamma | \mu) + \mathcal{D}_{\varphi_2}(P_{\#}^Y \gamma | \nu).$$

Proposition 2.12. *Assume that (2.4) is feasible. If φ_1 and φ_2 are superlinear, then the infimum is attained. This is also the case if c has compact sublevel sets and $\varphi'_{1\infty} + \varphi'_{2\infty} + \inf c > 0$.*

The proof of this result as well as a thorough study of duality properties of this problem can be found in [47]. Note that (2.3) is a particular case of (2.4) when the domains of the entropy functions φ_i are the singleton $\{1\}$. More generally, if the functions φ_i admit unique minima at 1, (2.4) can be viewed as a relaxation of the initial problem (2.3) where the “hard” marginal constraints are replaced by “soft” constraints, penalizing the deviation of the marginals of γ from μ and ν . Now we discuss a specific case of particular interest.

Definition 2.13 (Wasserstein-Fisher-Rao distance). Take $X = Y$, let d be a distance on X and $\lambda \geq 0$. For the cost

$$(2.5) \quad c(x, y) = -\log(\cos_+^2(d(x, y)))$$

with $\cos_+ : z \mapsto \cos(z \wedge \frac{\pi}{2})$ and $\mathcal{D}_{\varphi_1} = \mathcal{D}_{\varphi_2} = \lambda \mathcal{KL}$, we define $\text{WFR}_\lambda(\mu, \nu)$ as the square root of the minimum in (2.4) (as a function of the measures $(\mu, \nu) \in \mathcal{M}_+(X)^2$). We simply write $\text{WFR} \stackrel{\text{def.}}{=} \text{WFR}_1$ for $\lambda = 1$.

As shown in [47], WFR defines a distance on $\mathcal{M}_+(X)$, which is equal to the *Wasserstein-Fisher-Rao* geodesic distance introduced simultaneously and independently in [25, 41, 47] (it is named *Hellinger-Kantorovich* in [47]). An alternative static formulation is given in [24]. Other important cases include:

- The *Gaussian Hellinger-Kantorovich* distance GHK is obtained by taking the cost $c = d^2$ (d still a metric) and $\mathcal{D}_{\varphi_i} = \lambda \text{KL}$ with $\lambda > 0$. It has been introduced in [47] where it is also shown that when X is a geodesic space, WFR is the geodesic distance generated by GHK.
- The *optimal partial transport* problem, which is obtained by taking a cost function bounded from below by -2λ and $\mathcal{D}_{\varphi_i} = \lambda \text{TV}$, with $\lambda > 0$. Originally, optimal partial transport refers to a problem where the marginal constraints are replaced by a constraint on the total mass of the marginals (and domination constraints); here λ is the Lagrange multiplier associated to the mass constraint (see [19]).
- With the divergence RG as in Example 2.7, one imposes a range constraint on the marginal:

$$\text{RG}(P_{\#}^X \gamma | \mu) < +\infty \quad \Leftrightarrow \quad \alpha \mu \leq P_{\#}^X \gamma \leq \beta \mu$$

(and similarly for $P_{\#}^Y \gamma$).

The following result is a minor extension of results from [24, 47] and is useful if one wishes to vary the parameter λ in numerical applications.

Proposition 2.14. *WFR $_\lambda$ defines a distance if $0 \leq \lambda \leq 1$ (degenerate if $\lambda = 0$). The upper bound on λ is necessary when X is a geodesic space of diameter greater than $\pi/2$.*

Proof. The case $\lambda = 0$ is trivial. If $\lambda > 0$, when dividing (2.4) by λ , one obtains the new cost $-2 \log(\cos_+^{1/\lambda}(d(x, y)))$. But the function $f : d \mapsto \arccos(\cos^{1/\lambda}(d))$ defined on $[0, \pi/2]$ is increasing, positive, satisfies $\{0\} = f^{-1}(0)$ and for $x \in]0, \pi/2[$ it holds that

$$f'' = \frac{1}{\lambda} \frac{\cos^{1/\lambda - 2}}{(1 - \cos^{2/\lambda})^2} \left(1 - \frac{1}{\lambda} \sin^2 - \cos^{2/\lambda} \right).$$

From the convexity inequality $\text{sign}[(1 - X)^{1/\lambda} - 1 + \frac{1}{\lambda}X] = \text{sign}(\frac{1}{\lambda} - 1)$ it follows that f is strictly concave on $[0, \pi/2]$ if $\lambda < 1$ and strictly convex if $\lambda > 1$. Thus, if $\lambda \leq 1$, $f \circ d$ still defines a distance on X and consequently WFR_λ , too.

If X is a geodesic space of diameter greater than $\pi/2$, take $(x_1, x_2) \in X^2$ such that $d(x_1, x_2) = \pi/2$. From [47, Corollary 8.3], $(\mathcal{M}(X), \text{WFR}_1)$ is itself a geodesic space. Consequently, there exists a midpoint $\mu \in \mathcal{M}_+(X)$, i.e., such that

$$\text{WFR}_1(\delta_{x_1}, \mu) = \text{WFR}_1(\delta_{x_2}, \mu) = \frac{1}{2} \text{WFR}_1(\delta_{x_1}, \delta_{x_2}).$$

From [47, Theorem 8.6] and the characterization of geodesics in $\text{Cone}(X)$ it holds $0 < d(x, x_i) < \pi/2$ for μ a.e. $x \in X$. This implies that for $\lambda > 1$, and μ a.e. x , $-\log(\cos_+^{1/\lambda}(d(x_i, x))) < -\log(\cos_+(d(x_i, x)))$. Thus $(\text{WFR}_\lambda^2/\lambda)(\delta_{x_i}, \mu) < \text{WFR}_1^2(\delta_{x_i}, \mu)$. But, for $\lambda \geq 1$, $(\text{WFR}_\lambda^2/\lambda)(\delta_{x_1}, \delta_{x_2}) = 2 \text{KL}(0|1) = 2$ this leads to

$$\text{WFR}_\lambda(\delta_{x_1}, \mu) + \text{WFR}_\lambda(\delta_{x_2}, \mu) < \text{WFR}_\lambda(\delta_{x_1}, \delta_{x_2})$$

and the triangle inequality property is lost. □

2.4. Barycenter problem and extensions. The problem of finding an “average” measure $\sigma \in \mathcal{M}_+(Y)$ which minimizes the sum of the (possibly unbalanced) transport cost toward every measure of a family $(\mu_k)_{k=1}^n \in \mathcal{M}_+(X)^n$ is of theoretical and practical interest (see Section 1.1). To formalize this problem, consider a family of costs functions $(c_k)_{k=1}^n$ on $X \times Y$ and a family of divergences $(\mathcal{D}_{\varphi_{k,1}}, \mathcal{D}_{\varphi_{k,2}})_{k=1}^n$. The problem is to solve

$$\inf_{\sigma \in \mathcal{M}_+(Y)} \inf_{\substack{(\gamma_k)_{k=1}^n \in \\ \mathcal{M}_+(X \times Y)^n}} \sum_{k=1}^n \left(\int_{X \times Y} c_k \, d\gamma_i + \mathcal{D}_{\varphi_{k,1}}(P_{\#}^X \gamma_i | \mu_i) + \mathcal{D}_{\varphi_{k,2}}(P_{\#}^Y \gamma_i | \sigma) \right).$$

By exchanging the infima, this is equivalent to

(2.6)

$$\inf_{\substack{(\gamma_k)_{k=1}^n \in \\ \mathcal{M}_+(X \times Y)^n}} \sum_{k=1}^n \int_{X \times Y} c_k \, d\gamma_k + \sum_{k=1}^n \mathcal{D}_{\varphi_{i,1}}(P_{\#}^X \gamma_k | \mu_k) + \inf_{\sigma \in \mathcal{M}_+(Y)} \sum_{k=1}^n \mathcal{D}_{\varphi_{k,2}}(P_{\#}^Y \gamma_k | \sigma).$$

Note that while the object of interest is the minimizer σ and not the family of couplings, we will see in Section 5.2 that the computation of σ is a byproduct of the “scaling” algorithm defined below.

The two following examples are specific instances of so-called Fréchet means, defined in any complete metric space (E, d) as solutions to

$$\operatorname{argmin}_{\sigma \in E} \sum_k \alpha_k d(\mu_k, \sigma)^2,$$

where $(\alpha_k)_{k=1}^n$ is a family of nonnegative weights and $(\mu_k)_{k=1}^n \in E^n$.

Example 2.15 (Wasserstein barycenters). Let $X = Y$ and let c be the quadratic cost $(x, y) \mapsto d(x, y)^2$ for a metric d and define $c_k = \alpha_k c$. Let all the divergences be the equality constraints. Then (2.6) is a formulation of the Fréchet means in the Wasserstein space, also known as Wasserstein barycenters (see [1]).

Example 2.16 (WFR barycenters). Let c be as in (2.5), define $c_k = \alpha_k c$ and let $\mathcal{D}_{\varphi_{k,1}} = \mathcal{D}_{\varphi_{k,1}} = \alpha_k \mathcal{KL}$ for all $k \in \{1, \dots, n\}$. Then (2.6) is a formulation of Fréchet means for the WFR distance.

2.5. Gradient flows.

2.5.1. *General outline.* Given a metric space (T, d) and some lower semicontinuous function $\mathcal{G} : T \rightarrow \mathbb{R} \cup \{\infty\}$ with compact sublevel sets, a time-discrete gradient flow with step size $\tau > 0$ starting from an initial point $\mu_0 \in T$ corresponds to the computation of a sequence $(\mu_k^\tau)_{k \in \mathbb{N}}$ via

$$\mu_0^\tau \stackrel{\text{def.}}{=} \mu_0, \quad \mu_{k+1}^\tau \in \operatorname{argmin}_{\mu \in T} \mathcal{G}(\mu) + \frac{d^2(\mu_k^\tau, \mu)}{2\tau}.$$

This scheme, which is a particular case of De Giorgi’s *minimizing movements*, generalizes the notion of implicit Euler steps to approximate gradient flows in Euclidean spaces, which are recovered when $(T, d) = (\mathbb{R}^n, |\cdot|)$. From a theoretical point of view, the object of interest is the limit trajectory when $\tau \rightarrow 0$ of the piecewise constant interpolation

$$\mu^\tau(t) \stackrel{\text{def.}}{=} \mu_k^\tau \text{ for all } t \in [k\tau, (k+1)\tau].$$

Initiated by [39], the study of such flows when T is the space of probability measures endowed with the Wasserstein metric $d = W$ (see Section 2.2) has led to considerable advances in the theoretical study of PDEs and their numerical resolution (see Section 1.1). The recent introduction of “unbalanced transport” metrics paves the way for further applications, since it is now possible to consider the whole space of nonnegative measures endowed, for instance, with the metric WFR, as considered in [37].

2.5.2. *Minimization problem.* For gradient flows based on an optimal transport metric, such as W , WFR, or GHK, each step requires to solve, after swapping the two infima, a problem of the form

$$(2.7) \quad \inf_{\gamma \in \mathcal{M}_+(X \times X)} \int_{X \times Y} c \, d\gamma + \mathcal{D}_{\varphi_1}(P_{\#}^1 \gamma | \mu_k^\tau) + \inf_{\mu \in \mathcal{M}_+(X)} (\mathcal{D}_{\varphi_2}(P_{\#}^2 \gamma | \mu) + 2\tau \mathcal{G}(\mu)),$$

where φ_1, φ_2 are entropy functions and \mathcal{G} is a lower semicontinuous functional which we also assume convex. This problem, as well as variants, involving so-called “splitting” techniques (see [37]) fit into the framework developed below. In particular, we show in Section 5.3 that the computation of μ_{k+1} is a byproduct of the minimization of (2.7) with the algorithm defined below.

2.5.3. *Flows associated to WFR.* The distance WFR was only introduced recently, so a sound theoretical analysis of the corresponding gradient flows and their limit PDEs is not yet available (and is not the subject of the present article). Meanwhile, we present here heuristic arguments in a smooth setting (see also [41, Section 3.2]) which lead to the evolution equation

$$(2.8) \quad \partial_t \mu = \operatorname{div}(\mu \nabla \mathcal{G}'(\mu)) - 4\mu \mathcal{G}'(\mu)$$

for the limit trajectory of $\mu^\tau(t)$ as $\tau \rightarrow 0$, where $\mathcal{G}'(\mu)$ is, if it exists, the unique function such that $\frac{d}{d\varepsilon}\mathcal{G}(\mu + \varepsilon\chi)|_{\varepsilon=0} = \int_X \mathcal{G}'(\mu)d\chi$ for all perturbations admissible perturbation χ .

Heuristic argument. When restricted to measures μ with positive density of Sobolev regularity, WFR is the weak Riemannian metric associated to the tensor

$$g(\mu)(\delta\mu, \delta\mu) \stackrel{\text{def.}}{=} \inf_{v,\alpha} \int_X |v(x)|^2 \mu(x) dx + \frac{1}{4} \int_X |\alpha(x)|^2 \mu(x) dx,$$

where, for a small variation $\delta\mu$, one searches over decompositions $\delta\mu = -\text{div}(v\mu) + \alpha\mu$ into displacement (given by the velocity field $v \in L^2(X, \mu)^d$) and growth (given by the rate of growth $\alpha \in L^2(X, \mu)$) (see [24]). A new step μ_{k+1}^τ is given from μ_k^τ through the resolution of

$$\mu_{k+1}^\tau \in \underset{\mu \in \mathcal{M}_+(X)}{\text{argmin}} \mathcal{G}(\mu) + \frac{1}{2\tau} \text{WFR}^2(\mu, \mu_k^\tau).$$

Searching the minimizer in the form $\mu = \mu_k^\tau + \tau(-\text{div}(v\mu_k^\tau) + \alpha\mu_k^\tau)$ with unknown (v, α) , this can be rewritten, in first order of $(\tau v, \tau\alpha)$, as

$$\inf_{v,\alpha} \mathcal{G}(\mu_k^\tau) + \int_X \left(\tau \mathcal{G}'(\mu_k^\tau)(\alpha\mu_k^\tau - \text{div}(v\mu_k^\tau)) + \frac{\tau^2}{2\tau} \left(|v(x)|^2 + \frac{1}{4}\alpha(x)^2 \right) \mu_k^\tau(x) \right) dx.$$

The first order optimality conditions yield $v = -\nabla\mathcal{G}'(\mu)$ and $\alpha = -4\mathcal{G}'(\mu)$. One thus obtains

$$\frac{1}{\tau}(\mu - \mu_k^\tau) = \text{div}(\mu\nabla\mathcal{G}'(\mu)) - 4\mu\mathcal{G}'(\mu)$$

which yields (2.8) as $\tau \rightarrow 0$. □

2.6. Generic formulation. Consider two convex and lower semicontinuous functions \mathcal{F}_1 and \mathcal{F}_2 defined on $\mathcal{M}_+(X)^n$ and $\mathcal{M}_+(Y)^n$, respectively. The variety of problems reviewed above can be seen as special cases of

$$(2.9) \quad \min_{\gamma \in \mathcal{M}_+^n(X \times Y)} \mathcal{J}(\gamma), \quad \text{where } \mathcal{J}(\gamma) \stackrel{\text{def.}}{=} \langle c, \gamma \rangle + \mathcal{F}_1(P_{\#}^X \gamma) + \mathcal{F}_2(P_{\#}^Y \gamma),$$

where $\langle c, \gamma \rangle = \sum_{k=1}^n \int c_k d\gamma_i$ is a short notation for the total transport cost of a family of couplings $(\gamma_k)_{k=1}^n$ w.r.t. a family of cost functions $(c_k)_{k=1}^n$ and the projection operators act elementwise. More precisely, they are recovered with the following choices of functions:

- Balanced OT (2.3): $\mathcal{F}(\sigma) = \iota_{\{\cdot\}}(\sigma|\mu)$;
- Unbalanced OT (2.4): $\mathcal{F}(\sigma) = \mathcal{D}_\varphi(\sigma|\mu)$;
- Barycenters (2.6): $\mathcal{F}(\sigma) = \inf_{\omega \in \mathcal{M}^n(X)} \{ \sum_{k=1}^n \mathcal{D}_{\varphi_k}(\sigma_k|\omega_k) + \iota_D(\omega) \}$, where D is the set of families of measures for which all components are equal,

$$D \stackrel{\text{def.}}{=} \{(\omega_1, \dots, \omega_n) \in \mathcal{M}^n(X) : \omega_i = \omega_j \text{ for } i, j = 1, \dots, n\};$$

- Gradient flows (2.7): $\mathcal{F}(\sigma) = \inf_{\omega \in \mathcal{M}(X)} \{ \mathcal{D}_\varphi(\sigma|\omega) + 2\tau\mathcal{G}(\omega) \}$.

It is remarkable that, even if \mathcal{F} is sometimes defined through an auxiliary minimization problem, it can still be handled efficiently in some cases by the scaling algorithm detailed below. For instance, functions of the form

$$(2.10) \quad \mathcal{F}(\sigma) = \inf_{\omega \in \mathcal{M}^n(X)} \sum_{k=1}^n \mathcal{D}_{\varphi_k}(\sigma_k|\omega_k) + \mathcal{G}(\omega),$$

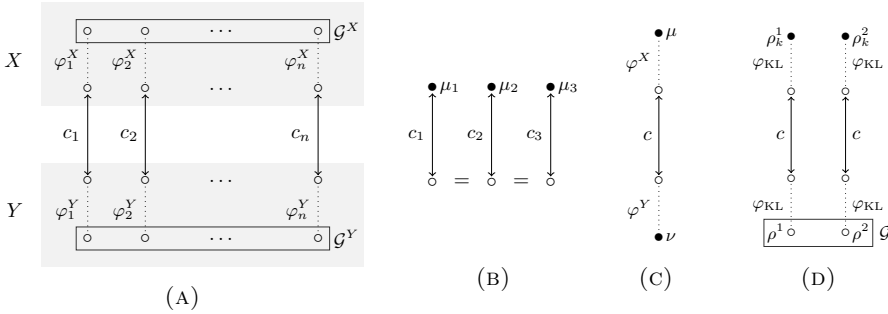


FIGURE 1. Representation of some problems of the form (2.9). Each circle represents a measure (filled with black if is fixed by an equality constraint); the terms in the functional \mathcal{J} (see (2.9)) are represented by arrows (for transport terms), dotted lines (for divergence terms) and rectangles (for the \mathcal{G} terms). (A) With \mathcal{F}_1 and \mathcal{F}_2 of the form (2.10); (B) Balanced barycenter; (C) Unbalanced optimal transport; (D) WFR gradient flow of 2 species with interaction.

where \mathcal{G} is convex and $(\varphi_k)_n$ is a family of entropy functions, are still convex and can model a great variety of problems. A graphical interpretation of these problems is suggested in Figure 1 along with some examples.

3. ENTROPIC REGULARIZATION AND ITERATIVE SCALING ALGORITHM

In this section, we introduce and describe an iterative scaling algorithm which solves a regularized version of the generic variational problem (2.9). This analysis is carried out in a continuous setting.

3.1. Entropic regularization. Following several recent articles, discussed in Section 1.1, we consider an approximation of problem (2.9) where an entropy term is used in place of the nonnegativity constraint on γ . Let dx and dy be reference probability measures on X and Y and assume that the product measure $dx dy$ is the reference measure on $X \times Y$ (one could consider more general reference measures on $X \times Y$ but this would require to invoke the concept of disintegration of measures). We define (minus) the entropy of a family of couplings $(\gamma_k)_{k=1}^n$ as

$$\mathcal{H}(\gamma) \stackrel{\text{def.}}{=} \sum_{k=1}^n \int_{X \times Y} r_k (\log(r_k) - 1) dx dy$$

if each γ_k admits a nonnegative density r_k w.r.t. $dx dy$ (with the convention $0 \log 0 = 0$) and ∞ otherwise. Note that each term is, up to a constant, the Kullback-Leibler divergence \mathcal{KL} of γ_i with respect to $dx dy$, as defined in Example 2.4. Given a small regularization parameter $\varepsilon > 0$, we consider

$$(3.1) \quad \min_{\gamma \in \mathcal{M}^n(X \times Y)} \mathcal{J}(\gamma) + \varepsilon \mathcal{H}(\gamma).$$

Recalling that $\mathcal{J}(\gamma) = \langle c, \gamma \rangle + \mathcal{F}_1(P_{\#}^X \gamma) + \mathcal{F}_2(P_{\#}^Y \gamma)$, this can be rewritten, up to a constant, as

$$(3.2) \quad \min_{\gamma \in \mathcal{M}^n(X \times Y)} \mathcal{F}_1(P_{\#}^X \gamma) + \mathcal{F}_2(P_{\#}^Y \gamma) + \varepsilon \sum_{k=1}^n \mathcal{KL}(\gamma_k | e^{-c_k/\varepsilon} dx dy).$$

Adding the entropy term can be interpreted in the following ways, detailed for $n = 1$ for simplicity.

3.1.1. *Interpretation 1.* The resolution of (3.1) can be understood as a regularization scheme, since the term $\varepsilon \mathcal{H}(\gamma)$ makes the problem strictly convex. In the limit $\varepsilon \rightarrow 0$, the unique solution to (3.1) should converge to a solution of the original problem. This is shown in the case of balanced transport (2.3) in [20, 44]. In our setting, a general result is beyond the scope of this article. Nevertheless, the finite dimensional case is insightful: if $dx dy$ has full support, then one recovers the minimizer of \mathcal{J} of minimal entropy.

Proposition 3.1. *Consider the case where $n = 1$ (one coupling) and X and Y are finite spaces. Assume that \mathcal{J} is convex, lower semicontinuous and that $A \stackrel{\text{def.}}{=} \operatorname{argmin}_{\gamma \ll dx dy} \mathcal{J}(\gamma)$ is nonempty. Let $(\varepsilon_k)_{k \in \mathbb{N}}$ be a sequence of strictly positive real numbers converging to 0. Then the sequence $(\gamma_k)_{k \in \mathbb{N}}$ of minimizers of (3.1) is well-defined and converges to $\operatorname{argmin}_{\gamma \in A} \mathcal{H}(\gamma)$.*

Proof. Since the objective function in (3.1) is coercive, strictly convex, lower semicontinuous and its domain is nonempty, there exists a unique minimizer γ_k of (3.1) associated to ε_k for all $k \in \mathbb{N}$. Now, let $\bar{\gamma} \in A$. For all $k \in \mathbb{N}$, the optimality of γ_k implies

$$\gamma_k \in \{ \gamma \in \mathcal{M}_+(X \times Y) ; \mathcal{H}(\gamma) \leq \mathcal{H}(\bar{\gamma}) \},$$

where $\mathcal{H}(\bar{\gamma})$ is finite since it is a finite sum of real numbers. As \mathcal{H} is coercive and lower semicontinuous, this set is compact. This implies the existence of cluster points for (γ_k) : let γ^* be one of them. One has that γ^* belongs to A since for all k , $|\mathcal{J}(\gamma_k) - \mathcal{J}(\bar{\gamma})| \leq 2\varepsilon_k \mathcal{H}(\bar{\gamma}) \rightarrow 0$ and $\mathcal{J}(\gamma^*) \leq \liminf_{k \rightarrow \infty} \mathcal{J}(\gamma_k)$. Moreover, as $\mathcal{H}(\gamma^*) \leq \mathcal{H}(\bar{\gamma})$ and $\bar{\gamma}$ is arbitrarily chosen in A , it holds that $\gamma^* \in \operatorname{argmin}_{\gamma \in A} \mathcal{H}(\gamma)$. By strict convexity, this cluster point is unique and $\gamma_k \rightarrow \gamma^*$. \square

3.1.2. *Interpretation 2.* A second way to interpret (3.1) is as a proximal step for the Kullback-Leibler divergence. Indeed, by denoting $\gamma^{(0)} \stackrel{\text{def.}}{=} dx dy$ and $\gamma^{(1)}$ the solution to (3.1), one has $\gamma^{(1)} = \operatorname{prox}_{\mathcal{J}/\varepsilon}^{\text{KL}}(\gamma^{(0)})$. It is possible to iterate this relation and define

$$\gamma^{(\ell+1)} \stackrel{\text{def.}}{=} \operatorname{prox}_{\mathcal{J}/\varepsilon}^{\text{KL}}(\gamma^{(\ell)})$$

which is the so-called proximal-point algorithm for the KL divergence, known to converge to the solution of (2.9) under the same conditions as in Proposition 3.1 (see [31]). It is a simple exercise to show, that for balanced optimal transport performing ℓ proximal steps with stepsize ε corresponds to a single step with stepsize ε/ℓ . While this precise relation is not true for the more general transport problems discussed here, we still find in practice that with small enough ε a single iteration yields sufficient accuracy (see Section 4.3 for numerical handling of small ε values).

3.1.3. *Interpretation 3.* Beyond computational aspects, for the specific case of standard OT with quadratic cost, the entropic regularization admits several interpretations as a stochastic version of OT. For instance, it is obtained by considering an optimal matching problem where there is (a specific model of) unknown heterogeneity in the preferences within each class to be matched [36]. What is more, it amounts to computing the law of motion of the (indistinguishable) particles of a gas which follow a Brownian motion, conditionally to the observation of its density at times $t = 0$ and $t = 1$, the so-called Schrödinger bridge problem [45].

3.1.4. *Choice of the reference measure.* The choice of dx and dy is critical for obtaining the weak convergence to a solution of (2.9) when $\varepsilon \rightarrow 0$. An obvious necessary condition is that the support of $dx dy$ should contain the support of an optimizer of the unregularized problem (2.9). For instance, when solving an unbalanced optimal transport problem between μ and ν , the choice $dx = \mu/\mu(X)$ and $dy = \nu/\nu(Y)$ is not suitable if one of the divergences is not superlinear because $\mathcal{D}_\varphi(\sigma|\mu) < \infty$ does not imply $\sigma \ll \mu$. In this case, the support of the reference measure $dx dy$ should contain the sets (well-defined under the hypotheses of Proposition 2.12)

$$\{(x, \operatorname{argmin}_y c(x, y)); x \in X\} \quad \text{and} \quad \{(\operatorname{argmin}_x c(x, y), y); y \in Y\}.$$

Also, for the barycenter (2.6) or the gradient flow problems (2.7), one does not have access in general to a reference measure according to which an optimizer is absolutely continuous. For numerics, the choice of the reference measures then corresponds to the choice of a discretization grid on which we find approximate solutions to the original problem.

3.2. **Reformulation using densities and duality.** From the definition of the entropy \mathcal{H} , any feasible γ of the generic problem (3.1) admits an L^1 density with respect to the reference measure $dx dy$. Accordingly, for the convenience of the analysis, we reformulate (3.2) as a variational problem on measurable functions:

$$(P_\varepsilon) \quad \min_{r \in L^1(X \times Y)^n} F_1(P_\#^X r) + F_2(P_\#^Y r) + \varepsilon \operatorname{KL}(r|K),$$

where $F_1(s) \stackrel{\text{def.}}{=} \mathcal{F}_1(s dx)$, $F_2(s) \stackrel{\text{def.}}{=} \mathcal{F}_2(s dy)$, $K \in L^1_+(X \times Y)^n$ is defined componentwise by

$$(3.3) \quad K_k(x, y) \stackrel{\text{def.}}{=} e^{-\frac{c_k(x, y)}{\varepsilon}}$$

for $k \in \{1, \dots, n\}$, and with the convention $\exp(-\infty) = 0$, the projection operator acts on each component of r and KL is the sum of the Kullback-Leibler divergences on each component (see the notations in Section 1.3). In this section, we only make the following general assumptions on the objects involved in (P_ε) .

Assumptions 1.

- (i) (X, dx) and (Y, dy) are probability spaces (i.e., measured spaces with unit total mass). The product space $X \times Y$ is equipped with the product measure $dx dy \stackrel{\text{def.}}{=} dx \otimes dy$;
- (ii) $F_1 : L^1(X)^n \rightarrow \mathbb{R} \cup \{\infty\}$ and $F_2 : L^1(Y)^n \rightarrow \mathbb{R} \cup \{\infty\}$ are weakly lower semicontinuous, convex and proper functionals;
- (iii) $K \in L^1_+(X \times Y)^n$ and $\varepsilon > 0$.

We begin with a general duality result, which is an application of Fenchel-Rockafellar Theorem (see Appendix A.1). It is similar to duality results that can be found in the literature on entropy minimization [13] but the functional we consider is more general.

Theorem 3.2 (Duality). *The dual problem of (P_ε) is*

$$(D_\varepsilon) \quad \sup_{(u,v) \in L^\infty(X)^n \times L^\infty(Y)^n} -F_1^*(-u) - F_2^*(-v) - \varepsilon \langle e^{(u \oplus v)/\varepsilon} - 1, K \rangle,$$

where $u \oplus v : (x, y) \mapsto u(x) + v(y)$. Strong duality holds, i.e., $\min(P_\varepsilon) = \sup(D_\varepsilon)$ and the minimum of (P_ε) is attained for a unique $r = (r_k)_{k=1, \dots, n} \in L^1(X \times Y)^n$. Moreover, u and v maximize (D_ε) if and only if

$$(3.4) \quad \begin{cases} -u \in \partial F_1(P_{\#}^X r) \\ -v \in \partial F_2(P_{\#}^Y r) \end{cases} \quad \text{and} \quad r_k(x, y) = e^{\frac{u_k(x)}{\varepsilon}} K_k(x, y) e^{\frac{v_k(y)}{\varepsilon}} \text{ for } k = 1, \dots, n.$$

Proof. In this proof, the spaces L^∞ and $(L^\infty)^*$ are endowed with the strong and the weak* topology, respectively. As $L^1(X \times Y)^n$ can be identified with a subset of the topological dual of $L^\infty(X \times Y)^n$, the function $\text{KL}(\cdot|K)$ can be extended on $(L^\infty(X \times Y)^n)^*$ as $G(r) = \text{KL}(r|K)$ if $r \in L^1(X \times Y)^n$ and $+\infty$ otherwise. Its convex conjugate $G^* : L^\infty(X \times Y)^n \rightarrow \mathbb{R}$ is

$$G^*(w) = \sum_{k=1}^n \int_{X \times Y} (e^{w_k(x,y)} - 1) K_k(x, y) dx dy = \langle e^w - 1, K \rangle$$

and is everywhere continuous for the strong topology [63, Theorem 4] (this property relies on the finiteness of $dx dy$ and the boundedness of K). The linear operator $A : L^\infty(X)^n \times L^\infty(Y)^n \rightarrow L^\infty(X \times Y)^n$ defined by $A(u, v) : (x, y) \mapsto u(x) + v(y)$ is continuous and its adjoint is defined on $(L^\infty(X \times Y)^n)^*$ (identified with a subset of $\mathcal{M}(X \times Y)^n$) by $A^*(r) = (P_{\#}^X r, P_{\#}^Y r)$. Since F_1 and F_2 are convex, lower semicontinuous, proper and since G^* is everywhere continuous on $L^\infty(X \times Y)^n$, strong duality and the existence of a minimizer for (P_ε) is given by Fenchel-Rockafellar theorem (see Appendix A.1). More explicitly, this theorem states that

$$\sup_{(u,v) \in L^\infty(X)^n \times L^\infty(Y)^n} -F_1^*(-u) - F_2^*(-v) - \varepsilon G^*(A(u, v)/\varepsilon)$$

and

$$\min_{r \in (L^\infty(X \times Y)^n)^*} F_1(P_{\#}^X r) + F_2(P_{\#}^Y r) + \varepsilon G(r)$$

are equal, the latter being exactly (P_ε) since G is infinite outside of $L^1(X \times Y)^n$. It states also that if (u, v) maximizes (D_ε) , then any minimizer of (3.1) satisfies $r \in \partial G^*(A(u, v)/\varepsilon)$ and the expression for the subdifferential of G^* is an application of the result in Appendix A.2. Finally, uniqueness of the minimizer for (P_ε) comes from the strict convexity of G . \square

3.3. Scaling algorithm. The specific splitting of the problem (P_ε) makes it suitable for the well-known Dykstra algorithm (see Section 1.1 for more background on this algorithm). Since the functions F_1 and F_2 operate on the marginals only, Dykstra’s iterations take a very simple form which is related to the celebrated Sinkhorn algorithm. They are obtained as an alternating maximization on (D_ε) .

Let \mathcal{K} and \mathcal{K}^T be the linear operators defined, for $a : X \rightarrow [0, \infty]^n$ and $b : Y \rightarrow [0, \infty]^n$ measurable, for $k = 1, \dots, n$ as

$$(3.5) \quad (\mathcal{K}b)(x)_k \stackrel{\text{def.}}{=} \int_Y K_k(x, y) b_k(y) dy \quad \text{and} \quad (\mathcal{K}^T a)(y)_k \stackrel{\text{def.}}{=} \int_X K_k(x, y) a_k(x) dx.$$

Given $v^{(0)} \in L^\infty(Y)^n$, alternating maximizations on the dual problem (D_ε) define, for $\ell = 0, 1, 2, \dots$, the iterates

$$(3.6) \quad \begin{cases} u^{(\ell+1)} = \operatorname{argmax}_{u \in L^\infty(X)^n} -F_1^*(-u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, \mathcal{K} e^{\frac{v^{(\ell)}}{\varepsilon}} \rangle_X, \\ v^{(\ell+1)} = \operatorname{argmax}_{v \in L^\infty(Y)^n} -F_2^*(-v) - \varepsilon \langle e^{\frac{v}{\varepsilon}}, \mathcal{K}^T e^{\frac{u^{(\ell+1)}}{\varepsilon}} \rangle_Y, \end{cases}$$

where we used the fact that, by Fubini-Tonelli, one has

$$(3.7) \quad \langle e^{(u \oplus v)/\varepsilon}, K \rangle_{X \times Y} = \langle e^{\frac{u}{\varepsilon}}, \mathcal{K} e^{\frac{v}{\varepsilon}} \rangle_X = \langle e^{\frac{v}{\varepsilon}}, \mathcal{K}^T e^{\frac{u}{\varepsilon}} \rangle_Y.$$

Conditions which ensure the existence of these iterates are given later in Theorem 3.8; for now, remark that by strict convexity, they are uniquely defined when they exist. Given an initialization $b^{(0)} \in L_+^\infty(Y)$, the main iterations of this paper are defined as follows for $\ell = 0, 1, 2, \dots$:

Scaling iterations:

$$(S) \quad a^{(\ell+1)} \stackrel{\text{def.}}{=} \frac{\operatorname{prox}_{F_1/\varepsilon}^{\text{KL}}(\mathcal{K}b^{(\ell)})}{\mathcal{K}b^{(\ell)}}, \quad b^{(\ell+1)} \stackrel{\text{def.}}{=} \frac{\operatorname{prox}_{F_2/\varepsilon}^{\text{KL}}(\mathcal{K}^T a^{(\ell+1)})}{\mathcal{K}^T a^{(\ell+1)}}$$

where the proximal operator for the KL divergence is defined for F_1 (and similarly for F_2) as

$$\operatorname{prox}_{F_1/\varepsilon}^{\text{KL}}(z) \stackrel{\text{def.}}{=} \operatorname{argmin}_{\substack{s: X \rightarrow \mathbb{R}_+^n \\ \text{measurable}}} F_1(s) + \varepsilon \text{KL}(s|z).$$

The following proposition shows that, as long as they are well-defined, these iterates are related to the alternate dual maximization iterates.

Proposition 3.3. *Define $a^{(0)} = \exp(v^{(0)})$, let $(a^{(\ell)}, b^{(\ell)})$ be the scaling iterates (S) and $(u^{(\ell)}, v^{(\ell)})$ the alternate dual maximization iterates defined in (3.6). If for all $\ell \in \mathbb{N}$, either $\log a^{(\ell)} \in L^\infty(X)^n$ and $\log b^{(\ell)} \in L^\infty(Y)^n$, or $u^{(\ell)} \in L^\infty(X)^n$ and $v^{(\ell)} \in L^\infty(Y)^n$, then*

$$(a^{(\ell)}, b^{(\ell)}) = (e^{u^{(\ell)}/\varepsilon}, e^{v^{(\ell)}/\varepsilon}).$$

The proof of this proposition makes use of the following lemma.

Lemma 3.4. *Let (T, dt) be a measure space and $v \in L_+^1(T)$. For any $u : T \rightarrow \mathbb{R}$ measurable, if $\text{KL}(u|v) < \infty$, then $u \in L_+^1(T)$.*

Proof. Without loss of generality, one can assume v positive since for dt -a.e. t , if $v(t) = 0$, then $u(t) = 0$. The subgradient inequality at 1 gives, for all $s \in \mathbb{R}$, $s \leq \varphi_{\text{KL}}(s) + e^1 - 1$. Consequently,

$$\int_T u dt = \int_T (u(t)/v(t))v(t) dt \leq \int_T (\varphi_{\text{KL}}(u(t)/v(t)) + e^1 - 1) v(t) dt < \infty. \quad \square$$

Proof of Proposition 3.3. Suppose that $v^{(\ell)} \in L^\infty(Y)^n$ and that $b^{(\ell)} = e^{v^{(\ell)}/\varepsilon}$. One has $\mathcal{K}b^{(\ell)} \in L^1(X)^n$ and from Lemma 3.4, one can compute $\text{prox}_{F_1/\varepsilon}^{\text{KL}}(\mathcal{K}b^{(\ell)})$ in $L^1(X)^n$. The Fenchel-Rockafellar theorem gives (see the proof of Theorem 3.2 for more details):

$$\sup_{u \in L^\infty(X)^n} -F_1^*(-u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, \mathcal{K}e^{\frac{v^{(\ell)}}{\varepsilon}} \rangle = \min_{s \in L^1(X)^n} F_1(s) + \varepsilon \text{KL}(s | \mathcal{K}e^{\frac{v^{(\ell)}}{\varepsilon}})$$

and the optimality conditions state that u^* maximizes the problem on the right if and only if the minimizer $s^* \stackrel{\text{def.}}{=} \text{prox}_{F_1/\varepsilon}^{\text{KL}}(\mathcal{K}e^{\frac{v^{(\ell)}}{\varepsilon}})$ of the problem on the left belongs to the subdifferential of $u \mapsto \langle e^{\frac{u}{\varepsilon}}, \mathcal{K}e^{\frac{v^{(\ell)}}{\varepsilon}} \rangle$ at the point u^* . That is (see Appendix A.2) if and only if for dx almost every $x \in X$,

$$s^*(x) = e^{u^*(x)/\varepsilon} \cdot (\mathcal{K}e^{v^{(\ell)}/\varepsilon})(x).$$

Thus, if $\varepsilon \log a^{(\ell+1)}$ belongs to $L^\infty(X)^n$ or if $u^* \in L^\infty(X)^n$ exists, then $u^* = \varepsilon \log a^{(\ell+1)}$. The rest of the proof is done by induction. □

3.4. Existence of the iterates for integral functionals. Our next step is to give conditions on F_1 and F_2 that guarantee the existence of the scaling iterates (S) and an equivalence with alternate maximization on the dual (3.6). This is provided by Theorem 3.8, where it is required that F_1 and F_2 are integral functionals, as we define now.

Definition 3.5 (Normal integrands and integral functionals [65]). A function $f : X \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is called a *normal integrand* if its epigraphical mapping $X \ni x \mapsto \text{epi } f(x, \cdot)$ is closed-valued and measurable. A convex integral functional is a function $F : L^1(X)^n \rightarrow \mathbb{R} \cup \{\infty\}$ of the form

$$F(s) = I_f(s) \stackrel{\text{def.}}{=} \int_X f(x, s(x)) dx,$$

where f is a normal integrand and $f(x, \cdot)$ is convex for all $x \in X$. In this paper, F is an *admissible* integral functional if, moreover, for all $x \in X$, $f(x, \cdot)$ takes nonnegative values, has a domain which is a subset of $[0, \infty]^n$ and if there exists $s \in L^1(X)^n$ such that $I_f(s) < \infty$.

The concept of normal integrands allows to deal conveniently with measurability issues. For finite dimensional problems (when X and Y have a finite number of points), integral functionals are simply sums of pointwise lower semicontinuous functions. The following proposition shows that for such functionals, conjugation and subdifferentiation can be performed pointwise.

Proposition 3.6. *If F is an admissible integral functional associated to the convex normal integrand f , then F is convex and weakly lower semicontinuous, f^* is also a normal convex integrand, $F^* = I_{f^*}$ and*

$$\begin{aligned} \partial F(s) &= \{u; u(x) \in \partial f(x, s(x)), dx \text{ a.e.}\}, \\ \partial F^*(u) &= \{s; s(x) \in \partial f^*(x, u(x)), dx \text{ a.e.}\}, \end{aligned}$$

where conjugation and subdifferentiation on f are w.r.t. the second variable.

Proof. This property can be found in [64] under the assumption of existence of a feasible point $s^* \in L^1(X)^n$ for I_f and a feasible point $u^* \in L^\infty(X)^n$ for I_{f^*} . Our admissibility criterion requires the existence of s^* and one has

$$I_{f^*}(0) = \int_X f^*(x, 0) dx = - \int_X \inf_{s \in \mathbb{R}^n} f(x, s) dx < \infty$$

since $\inf_s f(x, s) \in [0, f(x, s^*(x))]$. \square

We now prove a general result on the “separability” of the KL proximal operator. The Kullback-Leibler divergence between two vectors s and z in \mathbb{R}^n is defined as

$$(3.8) \quad \overline{\text{KL}}(s|z) \stackrel{\text{def.}}{=} \sum_{k=1}^n s_k \log \left(\frac{s_k}{z_k} \right) - s_k + z_k$$

if $(z_k = 0) \Rightarrow (s_k = 0)$, and $+\infty$ otherwise, with the convention $0 \log(0/0) = 0$. While we use a separate notation for the sake of clarity, this definition is actually consistent with the definition of KL if one interprets s and z as vector densities on a space (X, dx) which is a singleton with unit mass.

Proposition 3.7. *Let $s : X \rightarrow \mathbb{R}^n$ be measurable. If $F = I_f$ is an admissible integral functional, then for almost all $x \in X$,*

$$\left(\text{prox}_{\frac{1}{\varepsilon} F}^{\text{KL}}(s) \right)(x) = \text{prox}_{\frac{1}{\varepsilon} f(x, \cdot)}^{\overline{\text{KL}}}(s(x)).$$

Proof. The problem which defines the proximal operator is that of minimizing $I_f(z) + \text{KL}(z|s) \stackrel{\text{def.}}{=} I_g(z)$ over measurable functions $z : X \rightarrow \mathbb{R}^n$ with

$$g : (x, z) \in X \times \mathbb{R}^n \mapsto f(x, z) + \overline{\text{KL}}(z|s(x)).$$

The function $(x, z) \mapsto \overline{\text{KL}}(z|s(x))$ is a convex normal integrand by [65, Prop. 14.30 and 14.45c]. Thus g is itself a normal convex integrand, as the sum of normal convex integrands [65, Prop. 14.44]. Then a minimization interchange result [65, Thm. 14.60] states that minimizing I_g is the same as minimizing g pointwise. \square

By Proposition 3.6, if F_1 and F_2 are admissible integral functionals then F_1^* and F_2^* are also integral functionals. So the alternating optimization on (D_ε) can be relaxed to the space of measurable functions and still have a meaning:

$$(3.9) \quad \begin{cases} u^{(\ell+1)} = \operatorname{argmax}_{u: X \rightarrow \mathbb{R}^n} -I_{f_1^*}(-u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, \mathcal{K} e^{\frac{v^{(\ell)}}{\varepsilon}} \rangle_X, \\ v^{(\ell+1)} = \operatorname{argmax}_{v: Y \rightarrow \mathbb{R}^n} -I_{f_2^*}(-v) - \varepsilon \langle e^{\frac{v}{\varepsilon}}, \mathcal{K}^T e^{\frac{u^{(\ell+1)}}{\varepsilon}} \rangle_Y. \end{cases}$$

The following theorem gives existence, uniqueness of this iterates and a precise relation with the scaling iterates (S).

Theorem 3.8. *Let F_1 and F_2 be admissible integral functionals as in Definition (3.5) associated to the normal integrands f_1 and f_2 . Assume that for all $x \in X$ and $y \in Y$, there exists points s_1 and s_2 with strictly positive coordinates such that $f_1(x, s_1) < \infty$ and $f_2(y, s_2) < \infty$ and that K takes positive values. Define $a^{(0)} = 1$ and let $(a^{(\ell)}, b^{(\ell)})$ be the scaling iterates (S). Then, with initialization $v^{(0)} = 0$, the iterates $(u^{(\ell)}, v^{(\ell)})$ in (3.9) are uniquely well-defined and for all $\ell \in \mathbb{N}$ one has $(a^{(\ell)}, b^{(\ell)}) = (e^{\frac{u^{(\ell)}}{\varepsilon}}, e^{\frac{v^{(\ell)}}{\varepsilon}})$.*

Proof. Suppose that $v^{(\ell)} : Y \rightarrow \mathbb{R}^n$ is a well-defined measurable function and that $b^{(\ell)} = e^{v^{(\ell)}/\varepsilon}$. As K is positive, $\mathcal{K}b^{(\ell)}$ is positive dx a.e. Let $a^{(\ell+1)}$ be computed with (S). Thanks to Proposition 3.7, the proximal operator can be decomposed as pointwise optimization problems and our assumptions allows to apply Fenchel-Rockafellar (see Appendix A.1) in the case where both problems reach their optima

$$\max_{u \in \mathbb{R}^n} -f_1^*(x, -u) - \varepsilon \langle e^{u/\varepsilon}, \mathcal{K}b^{(\ell)} \rangle = \min_{s \in \mathbb{R}^n} f_1(x, s) + \varepsilon \overline{\text{KL}}(s|\mathcal{K}b^{(\ell)}(x))$$

with the relation between optimizers: $e^{u/\varepsilon} = s/\mathcal{K}b^{(\ell)}(x)$. This formula guarantees that the function of pointwise maximizers $x \mapsto u(x)$ is measurable since the function of pointwise minimizers $x \mapsto s(x)$ is uniquely well-defined and measurable by [65, Thm. 14.37]. Indeed, the minimized function is a strictly convex normal integrand because it is shown in Appendix A.2 that $\overline{\text{KL}}(\cdot|\mathcal{K}b^{(\ell)}(x))$ is a normal integrand and sum of normal integrands are normal [65, Prop. 14.44]. This shows that $a^{(\ell+1)} = e^{\frac{u^{(\ell+1)}}{\varepsilon}}$ and one concludes by induction. \square

3.5. Convergence analysis. This Section gives a fixed point result and a convergence result for a particular case. The finite dimensional case is postponed to the next section.

The following proposition sheds some light on the name “scaling” given to iterations (S). It comes from the fact that these iterations allow to recover a solution to (P_ε) by multiplying the kernel K with positive functions, interpreted as scalings.

Proposition 3.9. *Under the assumptions of Theorem 3.8, if the scaling iterations (S) admit a fixed point (a, b) such that $\log a \in L^\infty(X)^n$ and $\log b \in L^\infty(Y)^n$, then $(\varepsilon \log a, \varepsilon \log b)$ is the unique solution of (D_ε) and the function r defined for each $k = 1, \dots, n$ by $r_k(x, y) = a_k(x)K_k(x, y)b_k(y)$ is the unique solution of (P_ε) .*

Proof. As a consequence of Proposition 3.7, on can write the optimality condition of a fixed point of (3.9) for almost every $x \in X$ as

$$u_k(x) = \varepsilon \log \left(\frac{a_k(x) \cdot \mathcal{K}b_k(x)}{\mathcal{K}b_k(x)} \right)$$

for some $u(x) \in -\partial f_1(x, a(x) \cdot \mathcal{K}b(x))$. Thus $-\varepsilon \log(a) \in \partial I_{f_1}(P_\#^X r)$ because $P_\#^X r = a \cdot \mathcal{K}b$ (the dot denotes componentwise multiplication). Similar derivations for b show that the couple $(\varepsilon \log a, \varepsilon \log b)$ and r satisfies the primal dual optimality conditions (3.4). \square

Sinkhorn’s algorithm (the special case of the scaling iterations (S) obtained when $F_1 = \iota_{\{p\}}(s)$ and $F_2 = \iota_{\{q\}}(s)$ are the convex indicators of equality constraints) is known to converge at a linear rate. This property is usually shown (see, for instance, [34]) by using the contraction property of the operator \mathcal{K} for the Hilbert metric (a projective metric on the cone of nonnegative functions). Using a similar approach, based on the related *Thompson* metric, we now show the linear convergence of the scaling iterates (S) in another special case, which is central in unbalanced optimal transport: when F_1 and F_2 are KL divergences with respect to fixed densities.

An approach involving the Hilbert metric would still be possible, but the use of the Thompson metric allows for a very short proof and a convergence rate which does not depend on a bound on the cost. This metric is defined as follows.

Definition 3.10 (Thompson metric on L_+^∞ [75]). For $r, s \in L_+^\infty(X)$, let $M(r/s) \stackrel{\text{def.}}{=} \inf \{ \alpha \geq 0 ; r \leq \alpha s \}$ (or ∞ if that set is empty). The Thompson metric is defined by

$$d(r, s) = \max \{ \log M(r/s), \log M(s/r) \} .$$

Key properties of this metric are:

- if $T : L^\infty(X) \rightarrow L^\infty(Y)$ is a z -homogeneous order-preserving (or order-reversing) operator, then $d(Tx, Ty) \leq |z| \cdot d(x, y)$;
- the equivalence relation $r \sim s \Leftrightarrow d(r, s) < +\infty$ generates a partition of $(L^\infty(X), d)$ and each part is a complete metric space. In particular, the set $\{s \in L_+^\infty(X) ; \log s \in L^\infty(X)\}$ endowed with the Thompson metric form a complete metric space.

Theorem 3.11. *Let p and q be such that $\log p \in L^\infty(X)$ and $\log q \in L^\infty(Y)$ and define*

$$F_1 = \lambda_1 \text{KL}(\cdot|p) \quad \text{and} \quad F_2 = \lambda_2 \text{KL}(\cdot|q)$$

for some $\lambda_1, \lambda_2 > 0$. Assume that the kernel K is lower bounded by a positive real number. If $(a^{(\ell)}, b^{(\ell)})$ are the scaling iterates (S) initialized with $b^{(0)} = 1$, then $(x, y) \mapsto a^{(\ell)}(x)K(x, y)b^{(\ell)}(y)$ converges at a linear rate (for the Thompson metric) to a solution of (P_ε) .

Proof. Let $z_i \stackrel{\text{def.}}{=} \lambda_i / (\lambda_i + \varepsilon) \in]0, 1[$ for $i \in \{1, 2\}$ and let $b^{(0)} = 1$. Following a simple application of Proposition 3.7 (or see Table 1) the iterates in this specific case read

$$a^{(\ell+1)} = \left(\frac{p}{\mathcal{K}b^{(\ell)}} \right)^{z_1} \quad \text{and} \quad b^{(\ell+1)} = \left(\frac{q}{\mathcal{K}a^{(\ell+1)}} \right)^{z_2} .$$

Our assumptions are such that $d(b^{(1)}, b^{(0)})$ and $d(a^{(1)}, a^{(0)})$ are finite (this is direct since the logarithms of $b^{(0)}$, K , p and q are bounded). Using the properties of the Thompson metric, it holds that

$$d(a^{(\ell+1)}, a^{(\ell)}) = z_1 \cdot d(\mathcal{K}b^{(\ell)}, \mathcal{K}b^{(\ell-1)}) \leq z_1 \cdot d(b^{(\ell)}, b^{(\ell-1)}) ,$$

since \mathcal{K} is an order preserving linear operator. In a similar fashion, we obtain $d(b^{(\ell+1)}, b^{(\ell)}) \leq z_2 \cdot d(a^{(\ell)}, a^{(\ell)})$. Thus the application $a^{(\ell)} \mapsto a^{(\ell+1)}$ is contractant of factor $z_1 \cdot z_2 < 1$ and $(a^{(\ell)}, b^{(\ell)})_\ell$ converges linearly to some $(a^*, b^*) \in L_+^\infty(X) \times L_+^\infty(Y)$ which is a fixed point of the iterations. Moreover, $|\log a^*|$ is bounded because $d(a^{(0)}, a^*) < \infty$, and so is $|\log b^*|$. The result follows then by Proposition 3.9. \square

4. ALGORITHM FOR DISCRETE MEASURES

In this section, we take a closer look at the scaling iterations (S) in the specific case of discrete finite spaces. We give a convergence result, we adapt the notation to obtain an algorithm that can be implemented in a straightforward manner and also discuss how to numerically stabilize the algorithm for small values of the regularization parameter ε , to reach a higher precision.

4.1. Convergence. In finite dimension⁴, the scaling iterations (S) converge in general. Note that the given convergence rate is pessimistic compared to that observed in practice (which is linear, see Figure 5).

⁴when X and Y are finite sets, or when the reference measures are finite sums of atoms

Theorem 4.1. *Assume that X and Y are finite sets and that K takes positive values. Let $(a^{(\ell)}, b^{(\ell)})$ denote the scaling iterates (S), $(u^{(\ell)}, v^{(\ell)}) \stackrel{\text{def.}}{=} (\varepsilon \log a^{(\ell)}, \varepsilon \log b^{(\ell)})$ the associated dual iterates and $-I$ the dual functional (D_ε). Then one has*

$$\varepsilon \text{KL}(r^*|r^{(\ell)}) \leq I(u^{(\ell)}, v^{(\ell)}) - \inf I \leq O(1/\ell),$$

where $r^{(\ell)}(x, y) = a^{(\ell)}(x)K(x, y)b^{(\ell)}(y)$.

Proof. By Proposition 3.3, $(u^{(\ell)}, v^{(\ell)})$ are the iterates of an alternate maximization on the dual problem (D_ε). Moreover, (minus) the dual functional has the structure

$$I(u, v) = F^*(-(u, v)/\varepsilon) + \varepsilon G^*(A(u, v)/\varepsilon),$$

where $G : r \mapsto \text{KL}(r|K)$ so that $G^*(z) = \langle e^z - 1, K \rangle$ is continuously differentiable, A is defined as $A(u, v)(x, y) = u(x) + v(y)$, and $F^*(u, v) = F_1^*(u) + F_2^*(v)$ is separable. Let us check that (i) G^* is Lipschitz on the sublevel sets of I and (ii) I admits minimizers, so that [6, Thm. 3.9] applies and proves $I(u^{(\ell)}, v^{(\ell)}) - \inf I \leq O(1/\ell)$.

In order to check (i), take any feasible point s for F . The Hölder inequality $F(s) + F^*(-(u, v)/\varepsilon) \geq \langle s, -(u, v)/\varepsilon \rangle$ gives

$$\varepsilon G^*(A(u, v)/\varepsilon) \leq I(u, v) + \langle (u, v)/\varepsilon, s \rangle + F(s).$$

As a consequence, the components of $A(u, v)$ are upper bounded on the set of dual variables (u, v) satisfying $I(u, v) \leq I(u^{(1)}, v^{(0)})$ and so I is uniformly Lipschitz on this set. In order to check (ii), remark that a primal minimizer r^* exists and since K takes positive values, one can build a minimizer by simply inverting the primal-dual relationship $(u^*, v^*) \in A^{-1}(\varepsilon \log r^*/K)$.

It remains to prove the first part of the inequality of the Theorem. We follow the strategy developed in [32, Thm. 1]. By direct computations (standard for Bregman divergences), one has for any pair (z, z^*) in the domain of G^* , by defining $r(z)(x, y) = K(x, y) \exp(z(x, y))$,

$$\varepsilon(G^*(z) - G^*(z^*) - \langle \nabla G^*(z^*), z - z^* \rangle) = \varepsilon \text{KL}(r(z^*)|r(z))$$

which is similar to a strict convexity estimate. Also, for any $m \in \partial F^*(-(u^*, v^*)/\varepsilon)$, one has by definition of the subdifferential

$$F^*(-(u^{(\ell)}, v^{(\ell)})/\varepsilon) - F^*(-(u^*, v^*)/\varepsilon) \geq -\langle m, (u^{(\ell)} - u^*, v^{(\ell)} - v^*) \rangle/\varepsilon.$$

By optimality, it holds $0 \in \partial I(u^*, v^*)$ so one can chose $m = \varepsilon A^* \nabla G^*(A(u^*, v^*)/\varepsilon)$. Then one poses $z = A(u^{(\ell)}, v^{(\ell)})/\varepsilon$ and $z^* = A(u^*, v^*)/\varepsilon$ in the first inequality, and add it to the second to obtain the result. \square

4.2. Numerical algorithm for discrete measures. We now adopt a more “implementation oriented” point of view, which leads to algorithms which are straightforward to implement. For the sake of clarity, we focus on the case of a single unknown coupling (i.e., $n = 1$, the extension to $n > 1$ merely requires to add an index).

Assume that $X = \{x_1, \dots, x_I\}$ and $Y = \{y_1, \dots, y_J\}$ are discrete finite spaces of cardinal I and J , respectively, endowed with the reference probability measures dx and dy described by nonnegative vectors summing to one

$$dx \stackrel{\text{def.}}{=} (dx(\{x_i\}))_{i=1}^I \in \mathbb{R}_+^I \quad \text{and} \quad dy \stackrel{\text{def.}}{=} (dy(\{y_j\}))_{j=1}^J \in \mathbb{R}_+^J.$$

In this context, one identifies functions $a : X \rightarrow \mathbb{R} \cup \{\infty\}$ or $b : Y \rightarrow \mathbb{R} \cup \{\infty\}$ to vectors $\mathbf{a} = (a(x_i))_{i=1}^I$ and $\mathbf{b} = (b(y_j))_{j=1}^J$ and functions on the product space $r : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$ to matrices $\mathbf{R} = (r(x_i, y_j))_{i,j}$ (denoted with capital letters

for a clearer distinction between vectors and matrices). This notation allows for a simple expression for the projections (1.2) through

$$P_{\#}^X \mathbf{R} = \mathbf{R} \, \mathbf{d}\mathbf{y} \quad \text{and} \quad P_{\#}^Y \mathbf{R} = \mathbf{R}^T \, \mathbf{d}\mathbf{x},$$

where $(\cdot)^T$ denotes matrix transposition and $\mathbf{R} \, \mathbf{d}\mathbf{y}$ denotes the standard matrix-vector product. Finally, the notation \odot denotes the entrywise product and \oslash the entrywise division with convention $0/0 = 0$.

Suppose we are given a cost function c described by the $I \times J$ matrix \mathbf{C} and two convex, lower semicontinuous and lower bounded functions F_1 and F_2 whose domain is nonempty and included in \mathbb{R}_+^I and \mathbb{R}_+^J , respectively. The Gibbs kernel (3.3) is then given by $\mathbf{K} = (e^{-\mathbf{C}_{i,j}/\varepsilon})_{i,j}$ and, by simply replacing continuous variables by their discrete counterparts, the main variational problem (P_ε) reads

$$(4.1) \quad \min_{\mathbf{R} \in \mathbb{R}^{I \times J}} F_1(\mathbf{R} \, \mathbf{d}\mathbf{y}) + F_2(\mathbf{R}^T \, \mathbf{d}\mathbf{x}) + \varepsilon \sum_{ij} \overline{\text{KL}}(\mathbf{R}_{i,j} | \mathbf{K}_{i,j}),$$

where $\overline{\text{KL}}$ is the pointwise Kullback-Leibler divergence (defined in (3.8)). Applying the operators \mathcal{K} and \mathcal{K}^T defined in (3.5) amounts to

$$\mathcal{K} \mathbf{b} = \mathbf{K}(\mathbf{b} \odot \mathbf{d}\mathbf{y}) \quad \text{and} \quad \mathcal{K}^T \mathbf{a} = \mathbf{K}^T(\mathbf{a} \odot \mathbf{d}\mathbf{x}).$$

For computing the scaling iterates (S), all the information we need about F_1 and F_2 can be condensed by the specification of functions $\text{proxdiv}_{F_1} : \mathbb{R}_+^I \times \mathbb{R}_+ \rightarrow \mathbb{R}_+^I$ and $\text{proxdiv}_{F_2} : \mathbb{R}_+^J \times \mathbb{R}_+ \rightarrow \mathbb{R}_+^J$ defined as

$$(4.2) \quad \text{proxdiv}_{F_i} : (\mathbf{s}, \varepsilon) \mapsto \text{prox}_{F_i/\varepsilon}^{\text{KL}}(\mathbf{s}) \oslash \mathbf{s}.$$

Given this function, which is easy to compute in many cases (see, e.g., Table 1), the scaling algorithm is straightforward to implement.

Algorithm 1. Scaling algorithm

```

1: function SCALINGALGO( $\text{proxdiv}_{F_1}, \text{proxdiv}_{F_2}, \mathbf{K}, \mathbf{d}\mathbf{x}, \mathbf{d}\mathbf{y}, \varepsilon$ )
2:    $\mathbf{b} \leftarrow \mathbb{1}_J$ 
3:   repeat
4:      $\mathbf{a} \leftarrow \text{proxdiv}_{F_1}(\mathbf{K}(\mathbf{b} \odot \mathbf{d}\mathbf{y}), \varepsilon)$ 
5:      $\mathbf{b} \leftarrow \text{proxdiv}_{F_2}(\mathbf{K}^T(\mathbf{a} \odot \mathbf{d}\mathbf{x}), \varepsilon)$ 
6:   until stopping criterion
7:   return  $(\mathbf{a}_i \mathbf{K}_{ij} \mathbf{b}_j)_{ij}$  ▷ The primal optimizer
8: end function

```

By virtue of Theorem 4.1, Algorithm 1 is guaranteed to stop and to return an approximate minimizer of (4.1) if one uses a consistent stopping criterion.

4.3. Log-domain stabilization. For small values of ε the entries of $\mathbf{K} = e^{-\mathbf{C}/\varepsilon}$, $\mathbf{a}^{(\ell)} = e^{\mathbf{u}^{(\ell)}/\varepsilon}$ and $\mathbf{b}^{(\ell)} = e^{\mathbf{v}^{(\ell)}/\varepsilon}$ may become both very small and very large, leading to numerically imprecise values of their mutual products and overflow of the numerical range. On the other hand, executing the algorithm in the log domain (i.e., storing the logarithms of the entries of \mathbf{K} , $\mathbf{a}^{(\ell)}$ and $\mathbf{b}^{(\ell)}$) is not completely satisfying because then the computation of $\mathcal{K}\mathbf{b}^{(\ell)}$ and $\mathcal{K}\mathbf{a}^{(\ell)}$ is not a simple matrix/vector

product and the algorithm is considerably slowed down. We suggest a middle way, using a redundant parametrization of the iterates as follows:

$$(4.3) \quad \mathbf{a}^{(\ell)} = \tilde{\mathbf{a}}^{(\ell)} \odot \exp(\tilde{\mathbf{u}}^{(\ell)}/\varepsilon), \quad \mathbf{b}^{(\ell)} = \tilde{\mathbf{b}}^{(\ell)} \odot \exp(\tilde{\mathbf{v}}^{(\ell)}/\varepsilon).$$

The idea is to keep $\tilde{\mathbf{a}}^{(\ell)}$ and $\tilde{\mathbf{b}}^{(\ell)}$ close to 1 and to absorb the extreme values of $\mathbf{a}^{(\ell)}$ and $\mathbf{b}^{(\ell)}$ into the log-domain via $\tilde{\mathbf{u}}^{(\ell)}$ and $\tilde{\mathbf{v}}^{(\ell)}$ from time to time.

Consider the stabilized kernels $\tilde{\mathbf{K}}^{(\ell)}$ whose entries are given by

$$\tilde{\mathbf{K}}_{ij}^{(\ell)} = \exp((\tilde{\mathbf{u}}_i^{(\ell)} + \tilde{\mathbf{v}}_j^{(\ell)} - \mathbf{C}_{i,j})/\varepsilon)$$

and remark that it holds, after direct computations,

$$\mathcal{K}\mathbf{b}^{(\ell)} = e^{-\frac{\tilde{\mathbf{u}}^{(\ell)}}{\varepsilon}} \odot \tilde{\mathbf{K}}^{(\ell)}(\tilde{\mathbf{b}}^{(\ell)} \odot \mathbf{d}\mathbf{y}) \quad \text{and} \quad \mathcal{K}^T\mathbf{a}^{(\ell)} = e^{-\frac{\tilde{\mathbf{v}}^{(\ell)}}{\varepsilon}} \odot (\tilde{\mathbf{K}}^{(\ell)})^T(\tilde{\mathbf{a}}^{(\ell)} \odot \mathbf{d}\mathbf{x}).$$

The scaling iterates (S) then read

$$\tilde{\mathbf{a}}^{(\ell+1)} = \text{prox}_{\frac{1}{\varepsilon}F_1}^{\text{KL}}(e^{-\frac{\tilde{\mathbf{u}}^{(\ell)}}{\varepsilon}} \odot \mathbf{s}_1) \odot \mathbf{s}_1, \quad \tilde{\mathbf{b}}^{(\ell+1)} = \text{prox}_{\frac{1}{\varepsilon}F_2}^{\text{KL}}(e^{-\frac{\tilde{\mathbf{v}}^{(\ell)}}{\varepsilon}} \odot \mathbf{s}_2) \odot \mathbf{s}_2,$$

where $\mathbf{s}_1 = \tilde{\mathbf{K}}^{(\ell)}(\tilde{\mathbf{b}}^{(\ell)} \odot \mathbf{d}\mathbf{y})$ and $\mathbf{s}_2 = (\tilde{\mathbf{K}}^{(\ell)})^T(\tilde{\mathbf{a}}^{(\ell)} \odot \mathbf{d}\mathbf{x})$.

For computing these iterates, all the information we need about F_1 and F_2 can be condensed by the specification of “stabilized proxdiv” functions $\text{proxdiv}_{F_1} : \mathbb{R}^I \times \mathbb{R}^I \times \mathbb{R}_+ \rightarrow \mathbb{R}^I$ and $\text{proxdiv}_{F_2} : \mathbb{R}^J \times \mathbb{R}^J \times \mathbb{R}_+ \rightarrow \mathbb{R}^J$ defined as

$$(4.4) \quad \text{proxdiv}_{F_i} : (\mathbf{s}, \mathbf{u}, \varepsilon) \mapsto \text{prox}_{F_i/\varepsilon}^{\text{KL}}(e^{-\frac{\mathbf{u}}{\varepsilon}} \odot \mathbf{s}) \odot \mathbf{s}.$$

We adopt the same notation as in (4.2) because it is just the special case when $\mathbf{u} = 0$. The main numerical algorithm thus obtained is displayed in Algorithm 2.

4.3.1. *Computing proxdiv.* The issue of extreme numerical values is not completely remedied by the absorption steps in Algorithm 2 since the proxdiv operation still involves the potentially extreme factor $e^{-\mathbf{u}/\varepsilon}$. In practice, however, we find that for many problems proxdiv can be computed without evaluating the exponential $e^{-\mathbf{u}/\varepsilon}$ and the formula remains numerically stable in the limit of small ε . Several examples for this are given in Section 5.

4.3.2. *Frequency of absorptions.* In practice, we recommend to run stabilized iterations and check for the extreme values of $(\tilde{\mathbf{a}}^{(\ell)}, \tilde{\mathbf{b}}^{(\ell)})$ every couple of iterations. When they exceed a given threshold, an absorption step is performed.

4.4. **Comments on implementation.** We wish to emphasize the simplicity of Algorithm 1 and Algorithm 2. When an optimization problem of the form (2.9) is given, one just has to: (i) choose the reference measures (which also determines a discretization grid in practice), (ii) determine the functions F_i by going to the space of densities, and (iii) find a way to efficiently compute proxdiv_{F_i} or proxdiv_{F_i} (in many cases, this operator has a closed form, or can be computed with a few parallelizable iterations). Let us briefly discuss some details on the practical implementation of Algorithm 2.

Algorithm 2. Scaling algorithm with stabilization

```

1: function SCALINGALGO2(proxdivF1, proxdivF2, C, dx, dy,  $\varepsilon$ )
2:   ( $\tilde{\mathbf{b}}$ , u, v)  $\leftarrow$  ( $\mathbb{1}_J$ , 0I, 0J)
3:    $\tilde{\mathbf{K}}_{ij} \leftarrow \exp(-\mathbf{C}_{ij}/\varepsilon)$  ▷ for all  $i, j$ .
4:   repeat
5:      $\tilde{\mathbf{a}} \leftarrow \text{proxdiv}_{F_1}(\tilde{\mathbf{K}}(\tilde{\mathbf{b}} \odot \mathbf{dy}), \mathbf{u}, \varepsilon)$ 
6:      $\tilde{\mathbf{b}} \leftarrow \text{proxdiv}_{F_2}(\tilde{\mathbf{K}}^T(\tilde{\mathbf{a}} \odot \mathbf{dx}), \mathbf{v}, \varepsilon)$ 
7:     if a component of  $|\log \tilde{\mathbf{a}}|$  or  $|\log \tilde{\mathbf{b}}|$  is “too big” then
8:       (u, v)  $\leftarrow$  (u +  $\varepsilon \log \tilde{\mathbf{a}}$ , v +  $\varepsilon \log \tilde{\mathbf{b}}$ )
9:        $\tilde{\mathbf{K}}_{ij} \leftarrow \exp((\mathbf{u}_i + \mathbf{v}_j - \mathbf{C}_{i,j})/\varepsilon)$  ▷ for all  $i, j$ .
10:       $\tilde{\mathbf{b}} \leftarrow \mathbb{1}_J$ 
11:    end if
12:  until stopping criterion
13:  return ( $\tilde{\mathbf{a}}_i \tilde{\mathbf{K}}_{i,j} \tilde{\mathbf{b}}_j$ ) $i,j$  ▷ The primal optimizer
14: end function

```

4.4.1. *Computing the matrix multiplications.* The size of the matrix \mathbf{K} is $I \times J$ so the matrix-vector multiplication step or even merely storing \mathbf{K} in memory can quickly become intractable as the sizes of X and Y increase. For special problems it is possible to avoid dense matrix multiplication (and storage). For instance, when X, Y are Cartesian grids in \mathbb{R}^d and $c(x, y) = |x - y|^2$ is the squared Euclidean distance, then \mathbf{K} is the separable Gaussian kernel: multiplying by \mathbf{K} can be done by successive “1-D convolutions”. We use this trick in Section 5, when mentioned. For more general geometric surfaces, \mathbf{K} can also be approximated by the heat kernel [73]. These methods however cannot be combined with Algorithm 2, as the “stabilized kernels” $\tilde{\mathbf{K}}$ lose the particular structure.

4.4.2. *Gradually decreasing ε .* When solving for a very small ε , most of the entries of \mathbf{K} are below machine precision, so one need to first “estimate” the dual variables (\mathbf{u}, \mathbf{v}) by performing several iterations with higher values of ε . Reduction of ε should be performed between lines 8 and 9 in Algorithm 2: after line 8, (\mathbf{u}, \mathbf{v}) are “approximations” of the dual variable of the unregularized problem so one can change ε and start solving for a different ε with (\mathbf{u}, \mathbf{v}) as a starting point. We use this heuristic in Section 5 (when mentioned) as follows: starting from $\varepsilon = 1$, after every 100 iteration we perform an absorption step and divide ε by factor chosen so that the final value ε is reached after 10 divisions. Then we run the standard Algorithm 2 until the desired convergence criterion is met.

Note that gradually decreasing ε has also been proposed to asymptotically solve the unregularized OT problem [71] (but ε should be of the order $1/\log(\ell)$ which is too slow to be of practical interest) and, heuristically, to accelerate the convergence of Sinkhorn’s iterations [42], in the same fashion as for interior point methods, but theoretical justification is an open problem.

4.4.3. *Multiple couplings.* The general optimization problem in Section 3 involved n couplings, potentially $n > 1$. For simplicity, throughout Section 4 we focussed on the case $n = 1$. However, the extension to $n > 1$ is rather simple. In particular, the variables $\mathbf{a}, \mathbf{b}, \mathbf{u}, \mathbf{v}$ of the algorithm lie in $\mathbb{R}^{n \times I}$ or $\mathbb{R}^{n \times J}$, the kernel \mathbf{K} is a n -family of $I \times J$ matrices, the entrywise operations (multiplication, division) are still

performed entrywise and the matrix vector multiplications are performed “coupling by coupling”, e.g., for $k \in \{1, \dots, n\}$ and $j \in \{1, \dots, J\}$:

$$(\mathbf{K} \mathbf{b})_{k,i} = \sum_j \mathbf{K}_{k,i,j} \mathbf{b}_{k,j}.$$

4.4.4. *More tricks in [69].* An efficient numerical implementation of Algorithm 2 is studied further in [69]. In addition to the log-domain stabilization and gradually decreasing ε , it is proposed to approximate \mathbf{K} by a sparse matrix, obtained by adaptive truncation. Thus, one can avoid storing of and multiplication by the dense kernel matrix, while keeping the inflicted truncation error negligible. This is more flexible than, for instance, the Gaussian convolution trick and can easily be extended to more general cost functions. In addition, this can be directly combined with the log-domain stabilization and therefore allows to solve larger problems with small regularization parameter (and hence, with little entropic blur). We choose, however, not to use these additional tricks Section 5 in order to display results which are easily reproducible.

4.5. **Generalization: more spaces and pushforward operators.** Scaling algorithms similar to Algorithm (S) can be formulated for solving problems of more general form than (P_ε) . There can be more than 2 functionals, more than 2 spaces involved and the projection operators $P_\#^X$ and $P_\#^Y$ can be replaced by more general linear operators, such as pushforwards of functions t which are not necessarily projections (i.e., not of the form $t(x, y) = x$). Several examples of such extensions can be found in [8] for the special case of classical optimal transport. Let us sketch this extension in the discrete setting and for the case of $n = 1$ (one “coupling”) so as to remain simple and to stick close to implementation concerns. For brevity, we limit ourselves to giving the “scaling” form of the alternate maximization on the dual and an example, without proof.

Let $(X^k, d\mathbf{x}^k)_{k=1}^N$ and $(Z, d\mathbf{z})$ be finite measured spaces of respective cardinalities $(I_k)_{k=1}^N$ and L and let $(t^k : Z \rightarrow X_k)_{k=1}^N$ be surjective maps. The space Z plays the role of $X \times Y$ in the previous discussions, but in this generalization the structure of a product space is lost. For conciseness, in the notations, the maps $(t^k)_k$ act on indices of points instead of points, with an obvious meaning. Given $k \in \{1, \dots, N\}$, $\mathbf{R} \in \mathbb{R}^L$ and $\mathbf{u} \in \mathbb{R}^{I_k}$ the pushforward operator $t_{\#}^k$ and its adjoint $(t_{\#}^k)^*$ read

$$t_{\#}^k \mathbf{R} = \left(\sum_{l \in (t^k)^{-1}(i)} \mathbf{R}_l \cdot d\mathbf{z}_l / d\mathbf{x}_i^k \right)_{i=1}^{I_k} \quad \text{and} \quad (t_{\#}^k)^* \mathbf{u} = (\mathbf{u}_{t^k(l)})_{l=1}^L.$$

Given a nonnegative vector $\mathbf{K} \in \mathbb{R}^L$ and N convex, proper, lower semicontinuous functions $F_k : \mathbb{R}^{I_k} \rightarrow \mathbb{R} \cup \{\infty\}$, the generalization of (4.1) is (up to a constant)

$$\min_{\mathbf{R} \in \mathbb{R}^L} \sum_{k=1}^N F_k(t_{\#}^k \mathbf{R}) + \varepsilon \sum_{l=1}^L \mathbf{R}_l \cdot (\log(\mathbf{R}_l / \mathbf{K}_l) - 1) \cdot d\mathbf{z}_l$$

with the convention $0 \log(0/0) = 0$, and the dual reads

$$(4.5) \quad \sup_{(\mathbf{u}^k)_{k=1}^N \in \mathbb{R}^{\sum_k I_k}} - \sum_{k=1}^N F_k^*(-\mathbf{u}^k) - \varepsilon \sum_{l=1}^L \exp\left(\frac{1}{\varepsilon} \sum_{k=1}^N \mathbf{u}_{t^k(l)}^k\right) \cdot \mathbf{K}_l \cdot d\mathbf{z}_l.$$

In order to perform alternate maximization on the dual as before, one needs a “disintegration” relation, in the spirit of (3.7). To this end, we define, for $(\mathbf{a}^n)_{n=1}^N \in$

$\mathbb{R}^{\sum_n I_n}$ and $k \in \{1, \dots, N\}$ the operator \mathcal{K}^k as

$$\mathcal{K}^k((\mathbf{a}^n)_{n \neq k}) \stackrel{\text{def.}}{=} \left(\sum_{l \in (t^k)^{-1}(i)} \left(\prod_{n \neq k} \mathbf{a}_{t^n(l)}^n \right) \cdot \mathbf{K}_l \cdot d\mathbf{z}_l / d\mathbf{x}_i^k \right)_{i=1}^{I_k}.$$

With those operators, the rightmost term of (4.5) can be computed in a “marginalized” way, using the relation

$$\langle \mathbf{a}^k, \mathcal{K}^k((\mathbf{a}^n)_{n \neq k}) \rangle_{d\mathbf{x}^k} = \sum_{l=1}^L \left(\prod_{n=1}^N \mathbf{a}_{t^n(l)}^n \right) \cdot \mathbf{K}_l \cdot d\mathbf{z}_l$$

valid for $k \in \{1, \dots, N\}$. The key feature for obtaining this relation is the fact that $((t^k)^{-1}(i))_{i=1}^{I_k}$ forms a partition of Z , and this explains why the scaling algorithm generalizes naturally to linear operators which are “pushforward”. It is now simple, at least formally, to define the generalization of the scaling algorithm displayed in Algorithm 3, by writing the alternate optimization on the dual problem, and taking again the dual, in the spirit of Proposition 3.3.

Algorithm 3. Generalized scaling algorithm

- 1: **function** GENERALSCALINGALGO($(\text{proxdiv}_{F_k}, \mathcal{K}^k, t^k)_{k=1}^N, \mathbf{K}, \varepsilon$)
 - 2: $\mathbf{a}^k \leftarrow \mathbb{1}_{I_k}$ ▷ for all $k = 1, \dots, N$.
 - 3: **repeat**
 - 4: **for** $k = 1, \dots, N$ **do**
 - 5: $\mathbf{a}^k \leftarrow \text{proxdiv}(\mathcal{K}^k((\mathbf{a}^n)_{n \neq k}), \varepsilon)$
 - 6: **end for**
 - 7: **until** stopping criterion
 - 8: **return** $(\mathbf{K}_l \cdot \prod_{k=1}^N \mathbf{a}_{t^k(l)}^k)_{l=1}^L$ ▷ The primal optimizer
 - 9: **end function**
-

As a simple illustration, consider, in the setting of equation (4.1), an extension where is added a function of the total mass

$$\min_{\mathbf{R} \in \mathbb{R}^{I \times J}} F_1(P_{\#}^X \mathbf{R}) + F_2(P_{\#}^Y \mathbf{R}) + F_3(m_{\#} \mathbf{R}) + \varepsilon \text{KL}(\mathbf{R} | \mathbf{K}),$$

where $m : (x, y) \rightarrow \{0\}$ ($m_{\#}$ returns the total mass) and $F_3 : \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper, lower semicontinuous and convex function. Applying the reasoning above, and after some rearrangement (here Z is still a product space and thus it is convenient to store \mathbf{R} as a matrix), one obtains Algorithm 4. If F_3 is the indicator of equality with a positive real number, this algorithm solves the partial optimal transport problem [19, 33] but more general F_3 can be considered such as, for instance, a range constraint on the total mass.

5. APPLICATIONS

Throughout this section, we detail how to use the scaling iterations (S) for solving the problems discussed in Section 2. We first analyze the properties of the functionals on marginals F_i , then we derive the iterations in a continuous setting, and finally show numerical experiments. We extend the definition of the operator proxdiv (defined in (4.4) in the discrete setting) to the continuous setting as follows: for $s \in L^1(X)$, $u \in L^\infty(X)$ and $\varepsilon > 0$,

$$(5.1) \quad \text{proxdiv}_F(s, u, \varepsilon) = \text{prox}_{F/\varepsilon}^{\text{KL}}(s e^{-\frac{u}{\varepsilon}}) / s$$

Algorithm 4. Scaling algorithm with a function on the total mass

```

1: function MASSSCALINGALGO( $(\text{proxdiv}_{F_k})_{k=1}^3, \mathbf{K}, \mathbf{dx}, \mathbf{dy}, \varepsilon$ )  $\triangleright \mathbf{K}$  is a matrix.
2:    $\mathbf{b} \leftarrow \mathbb{1}_J$ 
3:    $z \leftarrow 1$ 
4:   repeat
5:      $\mathbf{a} \leftarrow \text{proxdiv}_{F_1}(z \cdot \mathbf{K}(\mathbf{b} \odot \mathbf{dy}))$ 
6:      $\mathbf{b} \leftarrow \text{proxdiv}_{F_2}(z \cdot \mathbf{K}^T(\mathbf{a} \odot \mathbf{dx}))$ 
7:      $z \leftarrow \text{proxdiv}_{F_3}((\mathbf{a} \odot \mathbf{dx})^T \mathbf{K}(\mathbf{b} \odot \mathbf{dy}))$ 
8:   until convergence
9:   return  $z \cdot (\mathbf{a}_i \mathbf{K}_{i,j} \mathbf{b}_j)_{i,j}$   $\triangleright$  The primal optimizer
10: end function

```

with the convention $0/0 = 0$. Moreover, in order to avoid the heavy notation $\mathcal{D}_\varphi(ax|b dx)$, we denote by straight letters the divergences between functions, i.e., for a, b measurable nonnegative functions on X and φ a nonnegative entropy function (see Definition 2.1):

$$(5.2) \quad \mathcal{D}_\varphi(a|b) \stackrel{\text{def.}}{=} \int_X \overline{\mathcal{D}}_\varphi(a(x)|b(x)) dx \quad \text{with} \quad \overline{\mathcal{D}}_\varphi(a|b) = \begin{cases} b \cdot \varphi(a/b) & \text{if } b > 0, \\ a \cdot \varphi'_\infty & \text{otherwise} \end{cases}$$

with the convention $0 \times \infty = 0$. Some properties of these divergences between functions are studied in Appendix A.2.

All reported runtimes were obtained with an implementation in Julia, on a standard laptop with CPU clock rate 2.5 GHz.

5.1. Balanced and unbalanced optimal transport.

5.1.1. *Derivation of the algorithm.* The basic framework of classical and unbalanced optimal transport has been recalled in Sections 2.2 and 2.3. Assume that we are given marginals $\mu \in \mathcal{M}_+(X)$, $\nu \in \mathcal{M}_+(\nu)$ and a cost function $c : X \times Y \rightarrow \mathbb{R} \cup \{\infty\}$. By defining $p = d\mu/dx$ and $q = d\nu/dy$, the marginal functionals involved in the regularized problem are

$$(5.3) \quad F_1(s_1) = \mathcal{D}_{\varphi_1}(s_1|p) \quad \text{and} \quad F_2(s_2) = \mathcal{D}_{\varphi_2}(s_2|q).$$

as in (5.2) above. As shown in Appendix A.2, if φ_1 and φ_2 are a nonnegative entropy function (Definition 2.1), then F_1 and F_2 are admissible integral functionals (Definition 3.5). In order to compute the associated proxdiv operator, let us apply Proposition 3.7 in this precise case.

Proposition 5.1. *Let φ be a nonnegative entropy function and $(s, p) \in L^1_+(X)^2$ such that $0 \in \text{dom } \varphi$ or $s(x) = 0 \Rightarrow p(x) = 0$ a.e. Let $F(s) = \mathcal{D}_\varphi(s|p)$. Then $\text{prox}_{F/\varepsilon}^{\text{KL}}(s)$ is not empty and is the singleton s^* satisfying for a.e. $x \in X$,*

$$\begin{cases} 0 = s^*(x) & \text{if } s(x) = 0, \\ 0 = \varepsilon \log(s^*(x)/s(x)) + \varphi'_\infty & \text{if } p(x) = 0 \text{ and } s(x) > 0, \\ 0 \in \varepsilon \log(s^*(x)/s(x)) + \partial\varphi(s^*(x)/p(x)) & \text{otherwise.} \end{cases}$$

Proof. It is the pointwise optimality conditions associated to Proposition 3.7. \square

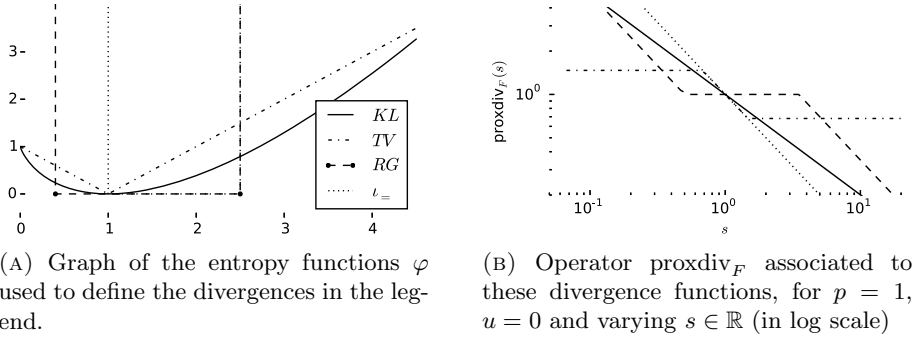


FIGURE 2. Divergences and proxdiv operators for the examples of Table 1

This formula allows to compute explicitly the proxdiv operators of the examples introduced in Section 2.1, as listed in Table 1. These entropy functions as well as the associated proxdiv operators are displayed in Figure 2. Note that, in Table 1 the first line corresponds to standard Sinkhorn iterations and these iterations are recovered in the second and third line by letting $\lambda \rightarrow +\infty$ and by setting $\alpha = \beta = 1$ in the fourth line. In the context of the log-domain stabilization (Section 4.3), all four proxdiv operators remain stable in the limit of small ε : either proxdiv is independent of u , only a regularized exponential $e^{-u/(\lambda+\varepsilon)}$ must be evaluated, or extreme values are cut off by thresholding.

TABLE 1. Some divergence functionals and the associated prox and proxdiv operators for functions s and u defined on X . All operators are acting pointwise and $\lambda > 0$, $0 \leq \alpha \leq \beta$ are real parameters (see Section 2.1 for the definitions).

F	$\text{prox}_{F/\varepsilon}^{\text{KL}}(s)$	$\text{proxdiv}_F(s, u, \varepsilon)$
$\iota_{\{=\}}(\cdot p)$	p	p/s
$\lambda \text{KL}(\cdot p)$	$s^{\frac{\varepsilon}{\varepsilon+\lambda}} \cdot p^{\frac{\lambda}{\varepsilon+\lambda}}$	$(p/s)^{\frac{\lambda}{\lambda+\varepsilon}} \cdot e^{-u/(\lambda+\varepsilon)}$
$\lambda \text{TV}(\cdot p)$	$\min \left\{ s \cdot e^{\frac{\lambda}{\varepsilon}}, \max \left\{ s \cdot e^{-\frac{\lambda}{\varepsilon}}, p \right\} \right\}$	$\min \left\{ e^{\frac{\lambda-u}{\varepsilon}}, \max \left\{ e^{-\frac{\lambda+u}{\varepsilon}}, p/s \right\} \right\}$
$\text{RG}_{[\alpha, \beta]}$	$\min \{ \beta p, \max \{ \alpha p, s \} \}$	$\min \{ \beta p/s, \max \{ \alpha p/s, e^{-u/\varepsilon} \} \}$

5.1.2. *Numerical examples for $X = [0, 1]$.* Let $X = Y$ be the discretization of the interval $[0, 1]$ into $I = J = 1000$ uniform samples and $\mathbf{dx} = \mathbf{dy} = \frac{1}{I} \mathbf{1}_I$ (the discretized Lebesgue measure). Let \mathbf{p}, \mathbf{q} be the (discrete) marginals displayed on Figures 3a–3b. We solve the discrete entropic regularized problem (4.1) using the stabilized scaling Algorithm 2 for marginal functions F_i of the form (5.3) with several choices of divergences listed in Figure 3. The algorithm was stopped after 10^3 iterations with a runtime of approximately 30 seconds.

For the optimal solution $\mathbf{R} \in \mathbb{R}^{I \times J}$, obtained after convergence, Figure 3 displays its projections $\mathbf{R} \mathbf{dy}$ and $\mathbf{R}^T \mathbf{dx}$ on the domain X and Figure 4 illustrates the entries

of \mathbf{R} which are greater than 10^{-10} , which can be interpreted as the approximate support of the optimizers. Thanks to the stabilization of Algorithm 2, the parameter ε could be made extremely small, until the plan is quasi-deterministic. Here we set it to $\varepsilon = 10^{-7}$ for the support to be clearly visible in Figure 4. Remark that for the TV case (orange support), the straight segments correspond to a density of 1 on the diagonal: this is the plan with minimal entropy among the (nonunique) optimal plans, in accordance with Proposition 3.1.

Figure 5 displays the primal-dual gap $(P_\varepsilon) - (D_\varepsilon)$ as a function of the iteration number ℓ when running Algorithm 1 for solving unbalanced optimal transport problems with $\varepsilon = 0.01$. More precisely, we display

$$P_\varepsilon(r^{(\ell)}) - D_\varepsilon(u^{(\ell)}, v^{(\ell)}),$$

where P_ε and D_ε are the primal and dual functionals (where set constraints are replaced by an exponential function of the distance to the set) as in Theorem 3.2 and $r^{(\ell)}, u^{(\ell)}, v^{(\ell)}$ are the primal and dual iterates as in Theorem 4.1. The marginals are the same as shown in Figures 3a–3b and the cost is quadratic. The discretization is $I = J = 500$ except for the thin black line where $I = J = 1000$. This plot leads to 3 observations which generalize the conclusion of Theorem 3.11: (i) convergence in function values is empirically linear, (ii) convergence is faster when the divergences are multiplied by a small weight, (iii) convergence speed is insensitive to dimension of the problem.

5.1.3. *Numerical examples for $X = Y = [0, 1]^2$.* Let $X = Y$ be the discretization of the 2-dimensional domain $[0, 1]^2$ into $I = J = 200 \times 200$ uniform samples and $\mathbf{dx} = \mathbf{dy}$ be the discrete Lebesgue measures. Let p, q be the densities displayed together in Figure 6a and c be the quadratic cost $c(x, y) = |y - x|^2$. We solve the discrete entropic regularized problem (4.1) using Algorithm 1 for several unbalanced optimal transport problems. We apply the “separable kernel” method (see Section 4.4) to accelerate the algorithm. Since this cannot be combined with the log-stabilization, the regularization parameter ε has been fixed to the rather high value $\varepsilon = 10^{-4}$ in order to avoid numerical issues. Figure 6 shows the marginals of the optimal coupling \mathbf{R} and Figure 7 illustrates the resulting transport plan: points with the same color correspond roughly to the same mass particle before and after transport. The algorithm was stopped after 10^3 iterations and the running time was of 35 seconds approximately.

5.1.4. *Color transfer.* In general it is difficult to display optimal transport maps for three-dimensional problems. An interesting application which allows intuitive visualization is color transfer: a classical task in image processing where the goal is to impose the color histogram of one image onto another image. Optimal transport between histograms has proven useful for problems of this sort such as contrast adjustment [28] and color transfer via 1-D transportation [60]. Indeed, optimal transport produces a correspondence between histograms which minimizes the total amount of color “distortion” (where the notion of distortion is specified by the cost function) and thus maintains maximal visual consistency.

In our experiments we represent colors in the three-dimensional “CIE-Lab” space (one coordinate for luminance and two for chrominance), resized to fit into a cuboid $X = Y = [0, 1]^3$, discretized into $I = J = 64 \times 32 \times 32$ uniform bins and we choose the quadratic cost $c(x, y) = |x - y|^2$. The anisotropic discretization of X accounts

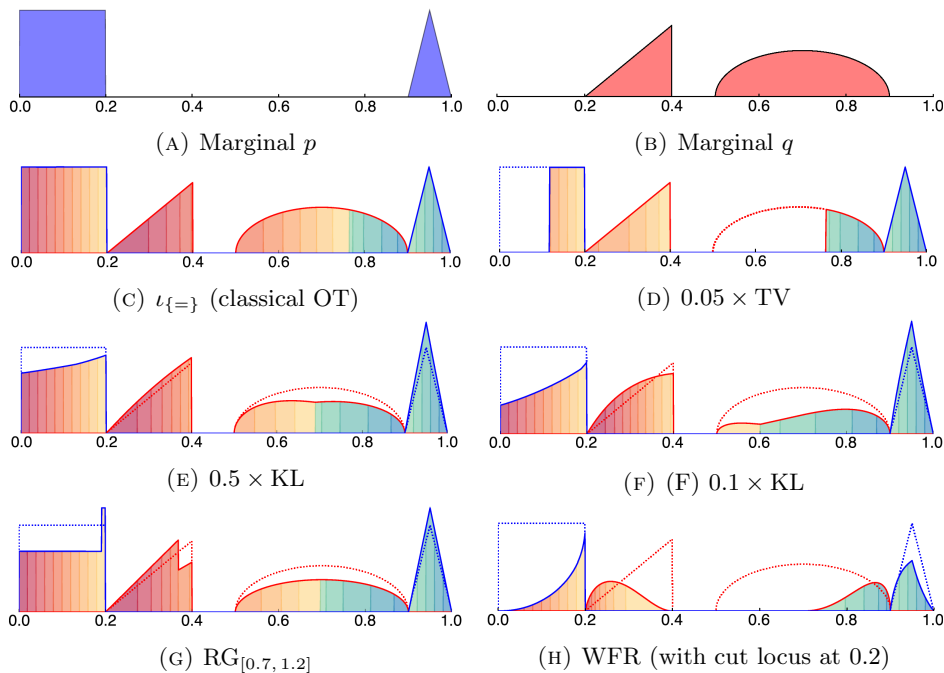


FIGURE 3. (A)–(B) Input marginals. (C)–(H) Marginals of the optimal plan \mathbf{R} displayed together for several divergences (specified in the caption). The color shows the location of the same subset of mass before and after transportation. The cost is the quadratic cost $c(x, y) = |y - x|^2$ except for (H) which is a computation of the optimal plan for WFR that is $F_1 = F_2 = \text{KL}$ and c as in (2.5) (with a spatial rescaling so that $c(x, y) = \infty \Leftrightarrow |y - x| \geq 0.2$).

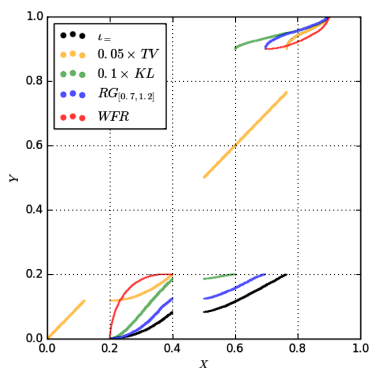


FIGURE 4. Support of the optimal plans $\{(x_i, y_j) ; \mathbf{R}_{i,j} > 10^{-10}\}$ for all the examples displayed in Figure 3, except (E). Orange and black are superimposed at the top.

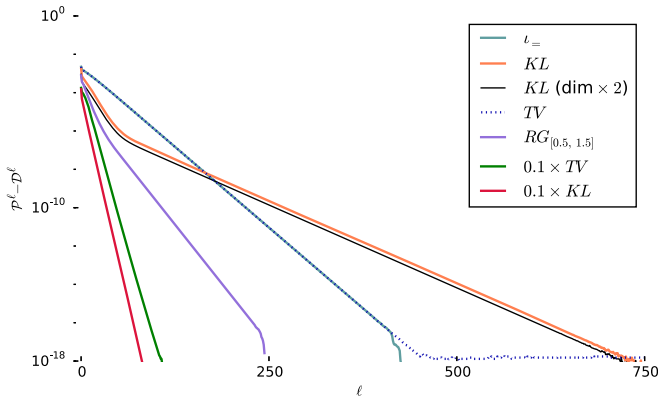


FIGURE 5. Primal dual gap as a function of the iterations for Algorithm 1 applied to unbalanced optimal transport problems.

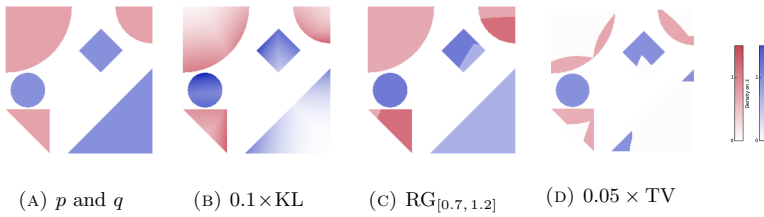


FIGURE 6. Marginals of the optimal plan \mathbf{R} for several unbalanced optimal transport problems (quadratic cost, divergence specified in the captions).

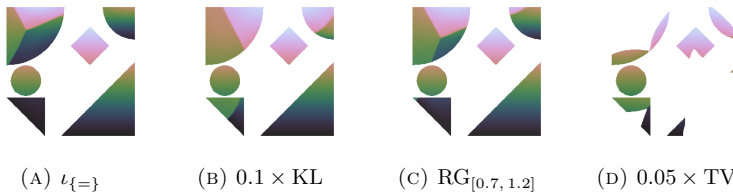


FIGURE 7. Representation of the transport map for the experiments of Figure 6.

for the fact that the eye is more sensitive to variations in luminance than variations in chrominance.

Let $\Omega \subset \mathbb{R}^2$ be the image domain. An image is described by a function $g : \Omega \rightarrow X$ and its color histogram is the pushforward of the Lebesgue measure on Ω by g . Let $g_X : \Omega \rightarrow X$ and $g_Y : \Omega \rightarrow Y (= X)$ be two images and let p, q be the densities of the associated color histograms with respect to the reference measures $dx = dy = \mathbb{1}_I$ which gives unit mass to each point of X . We run Algorithm 1 with the “separable kernel” method (see Section 4.4) to obtain an (unbalanced) optimal transport plan \mathbf{R} . An approximate transport map T is then computed according to the barycentric

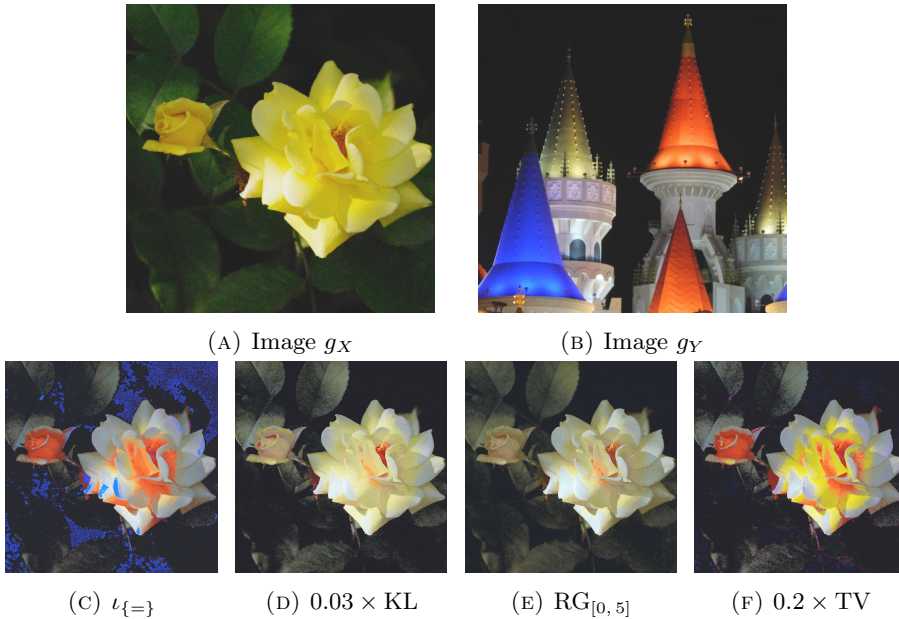


FIGURE 8. A challenging color transfer experiment where the colors of the image g_Y are transferred to the image g_X . In all cases $F_1 = \iota_{\{=\}}(\cdot | p)$ and F_2 is the divergence with respect to q specified in the caption. Note that in (F) some colors are not “displaced”. The parameters for F_2 are chosen arbitrarily.

projection (as already used, for instance, in [73])

$$T(x_i) = \mathbf{T}_i \quad \text{where} \quad \mathbf{T} \stackrel{\text{def.}}{=} (\mathbf{R} \cdot \mathbf{y}^1, \mathbf{R} \cdot \mathbf{y}^2, \mathbf{R} \cdot \mathbf{y}^3) / \mathbf{R} \cdot \mathbf{1}_J,$$

where $\mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2, \mathbf{y}^3) \in \mathbb{R}^{J \times 3}$ is the vector of coordinates of the points in $Y (= X)$. The modified image is finally obtained as $T \circ g_X$.

On Figure 8, we display the color transfer between very dissimilar images, computed with parameter $\varepsilon = 0.002$. The algorithm was stopped after 2000 iterations and the running time was approximately 160 seconds. This intentionally challenging example is insightful as it exhibits a strong effect of the choice of the divergence. There are no quantitative measures for the quality of a transformed image, but the application of unbalanced optimal transport allows to select the “right amount of colors” in the target histogram so as to match the modes of the initial histogram and yields meaningful results.

5.2. Unbalanced barycenters.

5.2.1. *Well-posedness.* Barycenters and related problems have been defined in Section 2.4. Assume we are given a family of n measures $(p_k dx)_{k=1}^n$, families of entropies $(\varphi_{1,k})_{k=1}^n$ and $(\varphi_{2,k})_{k=1}^n$ and families of cost functions $(c_k)_{k=1}^n$. For a clearer picture, let us focus on the case where

$$\varphi_{2,k} = \alpha_k \cdot \lambda \cdot \varphi \quad \text{for} \quad k = 1, \dots, n,$$

where φ is a nonnegative entropy function, $(\alpha_k)_{k=1}^n \in]0, \infty[^n$ are weights and $\lambda > 0$ is a (redundant) parameter. It is also convenient to slightly modify (for this section only) the definition of the KL divergence given in (1.1) by introducing weights as

$$\text{KL}(r|s) \stackrel{\text{def.}}{=} \sum_k \alpha_k \cdot \text{KL}(r_k|s_k).$$

No theoretical aspect is affected by this small change, but the definition of the kernel has to be adapted, for each component, as

$$K_k = \exp(-c_k/(\alpha_k \varepsilon))$$

so that one still has $\langle c, r \rangle + \varepsilon \text{KL}(r|1) = \varepsilon \text{KL}(r|K) + \text{const.}$ By equation (2.6), the barycenter problem with entropic regularization corresponds to defining

$$F_1(r) = \sum_{k=1}^n \mathcal{D}_{\varphi_{k,1}}(r_k dx | p_k dx) \quad \text{and} \quad F_2(s) = \inf_{\sigma \in \mathcal{M}_+(Y)} \lambda \sum_{k=1}^n \alpha_k \mathcal{D}_{\varphi}(s_k dy | \sigma)$$

(for all $r \in L^1(X)^n$ and $s \in L^1(Y)^n$) in (P_ε) . It is direct to see that the proximal operator of F_1 for the KL divergence (according to our specific definition of KL) can be computed componentwise as

$$(5.4) \quad \text{prox}_{F_1}^{\text{KL}}(r) = (\text{prox}_{\mathcal{D}_{\varphi_{1,1}}(\cdot|p_1)}^{\text{KL}}(r_1), \dots, \text{prox}_{\mathcal{D}_{\varphi_{n,1}}(\cdot|p_n)}^{\text{KL}}(r_n)),$$

so we can use the results from the previous section (recall that D denotes divergences between functions as in (5.2)). Let us turn our attention to the function F_2 . As it is assumed $\varphi(0) \geq 0$, F_2 is not changed by taking the infimum over σ of the form $h \cdot dy$ for $h \in L^1(Y)$. So one can express F_2 for $s \in L^1(Y)^n$, with notation of (5.2) as

$$(5.5) \quad F_2(s) = \min_{h \in L^1(Y)} \lambda \sum_{k=1}^n \alpha_k \mathcal{D}_{\varphi}(s_k | h) = \min_{h \in L^1(Y)} \lambda \sum_{k=1}^n \alpha_k \int_X \bar{D}_{\varphi}(s_k(x) | h(x)) dx.$$

Proposition 5.2. *If $\varphi'_\infty > 0$, then F_2 is an admissible integral functional in the sense of Definition 3.5 and for all $s \in L^1(Y)^n$, there exists a minimizer $h \in L^1(Y)$. Moreover, if $r^* \in L^1(X \times Y)^n$ minimizes (P_ε) , then the associated minimizer h^* is a pointwise minimizer of (5.5) at the point $s = P_{\#}^Y r^*$.*

Proof. Let $G(s, h)$ be the function on the right side, which is an admissible integral functional (see Proposition A.2 in Appendix A.2). Let us verify the assumptions of the “reduced minimization Theorem” [64, Corollary 3B]. Assumption (i) is satisfied because $\varphi'_\infty > 0$ guarantees the growth condition [64, 2R] and assumption (ii) is guaranteed by the fact that if $u \in L^1(Y)$ and $\mathcal{D}_{\varphi}(u|v) < \infty$, then $v \in L^1(Y)$ (this is proven by adapting slightly the proof of Lemma 3.4, using the—at least linear—growth of φ). Thus the cited corollary applies. \square

5.2.2. *Derivation of the algorithm.* According to Proposition 3.7, computing the prox and proxdiv operators requires to solve, for each point $y \in Y$, a problem of the form

$$(5.6) \quad \min_{(\tilde{s}, h) \in \mathbb{R}^n \times \mathbb{R}} \sum_{k=1}^n \alpha_k (\varepsilon \cdot \overline{\text{KL}}(\tilde{s}_k | s_k) + \lambda \cdot \bar{D}_{\varphi}(\tilde{s}_k | h)).$$

If φ is smooth, first order optimality conditions for (5.6) are simple to obtain. The next proposition deals with the general case where more care is needed.

Proposition 5.3. *Let $(s_i)_{i=1}^n \in \mathbb{R}_+^n$. Assume that there exists a feasible candidate (\tilde{s}^0, h^0) for (5.6) such that $(s_i > 0) \Rightarrow (\tilde{s}_i^0 > 0)$. A candidate (\tilde{s}, h) is a solution of (5.6) if and only if*

- $(s_i = 0) \Leftrightarrow (\tilde{s}_i = 0)$ and
- *there exists $b \in \mathbb{R}^n$ such that $\sum_{k=1}^n \alpha_k b_k = 0$ and, for all $k \in \{1, \dots, n\}$, $(a_k, b_k) \in \partial D_\varphi(\tilde{s}_k|h)$ with $a_k \stackrel{\text{def.}}{=} \frac{\varepsilon}{\lambda} \log \frac{s_k}{\tilde{s}_k}$ if $s_k > 0$ and $b_k \in \partial_2 D_\varphi(0, h)$ otherwise.*

Proof. First assume that $s_i > 0$ for all $i \in \{1, \dots, n\}$. The positivity assumption on the feasible point implies that the sum of the subdifferential is the subdifferential of the sum by continuity of KL for positive arguments. Moreover, a minimizer necessarily satisfies $\tilde{s}_i > 0$ for all i . Consequently, the subderivative of the function in (5.6) at $((\tilde{s}_i)_n, h)$ is the set of vectors in \mathbb{R}^{n+1} of the form

$$\begin{pmatrix} \vdots \\ \varepsilon \alpha_i \log(\tilde{s}_i/s_i) + \lambda \alpha_i a_i \\ \vdots \\ \lambda \sum_k \alpha_k b_k \end{pmatrix}$$

with $(a_i, b_i) \in \partial D_\varphi(\tilde{s}_i|h)$. Writing the first order optimality condition yields the second condition of the proposition. Now for all i such that s_i is null, set $\tilde{s}_i = 0$ (this is required for feasibility) and do the reasoning above by withdrawing the variables \tilde{s}_i which have been fixed. □

Note that once the optimal h is found (possibly with the help of the optimality conditions of Proposition 5.3), determining the optimal values for \tilde{s} can be done componentwise as in (5.4) with the help of Proposition 5.1. In Table 2 we provide formulas for h for some examples (proofs can be found in Appendix A.3), for the subsequent computation of \tilde{s}_k (and the proxdiv step) we refer the reader to Table 1 (where of course one must replace p by h).

5.2.3. *Numerical experiments.* Geodesics in Wasserstein and Wasserstein–Fisher–Rao space can be computed as weighted barycenters between their endpoints. In Figure 9 geodesics for the Wasserstein–Fisher–Rao distance on $X = Y = [0, 1]$ for different cut-loci d_{\max} and the Wasserstein distance (corresponding to $d_{\max} = \infty$) are compared and the influence of entropy regularization is illustrated. The interval $[0, 1]$ was discretized into $I = J = 512$ uniform samples. \mathbf{dx} and \mathbf{dy} were chosen to be the discretized Lebesgue measure.

In Figure 10, we display Fréchet means-like experiments for a family of 4 given marginals where $X = Y$ is the segment $[0, 1]$ (discretized as above). The (discrete) densities of the marginals $(\mathbf{p}_k)_{k=1}^4$ consist each of the sum of three bumps (centered near the points $x = 0.1, x = 0.5$ and $x = 0.9$). These computations were performed with Algorithm 2 which was stopped after 1500 iterations (running time of 30 seconds approximately) and with $\varepsilon = 10^{-5}$. We observe that relaxing the marginal constraints (Figures 10c–10f) allows us to conserve this structure in three bumps in contrast to classical optimal transport (Figure 10b).

Figure 11 and 12 display barycenters for the Wasserstein and the GHK metric (defined in Section 2.3) between three densities on $[0, 3]^2$ discretized into 200×200 samples. Computations were performed using Algorithm 1 and the “separable kernel” method (see Section 4.4) which was stopped after 1500 iterations (running

TABLE 2. Expression for the minimizer h of (5.6) as a function of $s \in \mathbb{R}^n$, where $I_+ = \{k; s_k > 0\}$ (proofs in Appendix A.3). For the implicit equations of cases (iii) and (iv), an exact solution can be given quickly because computing $\log h$ consists in finding the root of a piecewise linear nondecreasing function (with at most $2n$ pieces), which is guaranteed to change its sign.

	D_φ	Formula for h as a function of $s \in \mathbb{R}^n$
(i)	$\iota_{\{=\}}$	$h = (\prod_k s_k^{\alpha_k})^{\frac{1}{\sum_k \alpha_k}}$
(ii)	λ KL	$h = \left(\frac{\sum_k \alpha_k s_k^{\frac{\varepsilon}{\varepsilon+\lambda}}}{\sum_k \alpha_k} \right)^{\frac{\varepsilon+\lambda}{\varepsilon}}$
(iii)	λ TV	if $\sum_{k \notin I_+} \alpha_k \geq \sum_{k \in I_+} \alpha_k$ then $h = 0$ otherwise solve: $\sum_{k \notin I_+} \alpha_k + \sum_{k \in I_+} \alpha_k \max\left(-1, \min\left(1, \frac{\varepsilon}{\lambda} \log \frac{h}{s_i}\right)\right) = 0$
(iv)	$\text{RG}_{[\beta_1, \beta_2]}$	if $s_k = 0$ for some k then $h = 0$ otherwise solve: $\sum_k \alpha_k \left[\beta_2 \min\left(\log \frac{\beta_2 h}{s_k}, 0\right) + \beta_1 \max\left(\log \frac{\beta_1 h}{s_k}, 0\right) \right] = 0$

time of 70 seconds approximately) with $\varepsilon = 9.10^{-4}$. The barycenter coefficients are the following:

$$\begin{aligned}
 & (0, 1, 0) \\
 & (1, 3, 0)/4 \quad (0, 3, 1)/4 \\
 & (2, 2, 0)/4 \quad (1, 2, 1)/4 \quad (0, 2, 2)/4 \\
 & (3, 1, 0)/4 \quad (2, 1, 1)/4 \quad (1, 1, 2)/4 \quad (0, 1, 3)/4 \\
 & (1, 0, 0) \quad (3, 0, 1)/4 \quad (2, 0, 2)/4 \quad (1, 0, 3)/4 \quad (0, 0, 1)
 \end{aligned}$$

The input densities have a similar global structure: each is made of three distant “shapes” of varying mass. The comparison between Figures 11 and 12 lead to a similar remark than for Figure 10: relaxing the strict marginal constraints allows us to maintain the global “structure” of the input densities.

5.3. Gradient flows and evolution of densities. The basic framework of gradient flows has been briefly laid out in Section 2.5. This section details the application of Algorithm 2 for solving them. As the transition from the measures formulation to the density formulation and further to the algorithm with pointwise optimality conditions was carefully detailed in Sections 5.1 and 5.2, we skip some of these intermediate steps here.

5.3.1. In the Wasserstein space. Scaling algorithms for solving Wasserstein gradient flows are not new [58], but our framework allows to simplify the derivation of the algorithm and the stabilized Algorithm 2 allows to use much smaller regularization parameter ε yielding sharper and more precise flows. Given a convex, lower semicontinuous function on measures \mathcal{G} with compact sublevel sets, each step

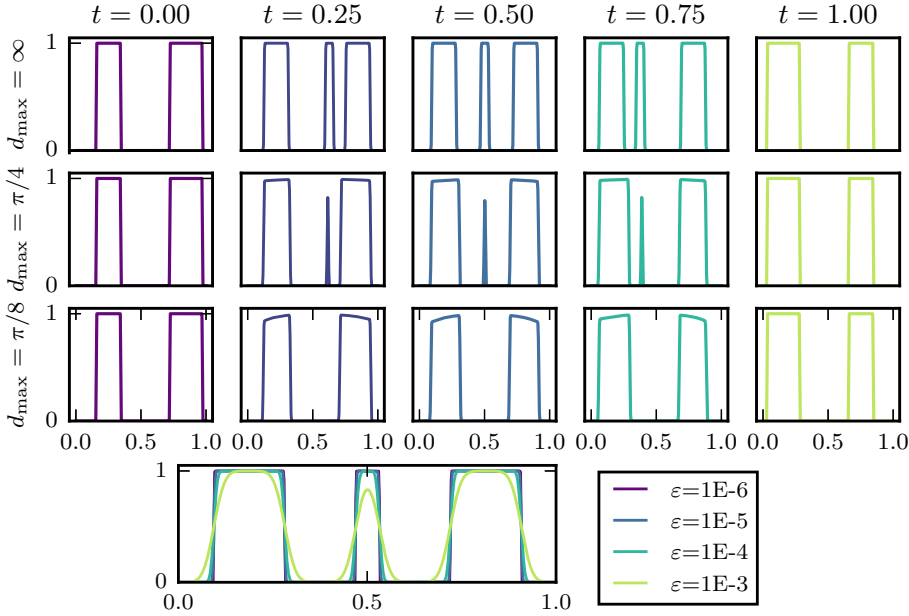


FIGURE 9. Top: Geodesics in Wasserstein distance ($d_{\max} = \infty$) and Wasserstein–Fisher–Rao distance (d_{\max} gives the cut-locus distance) as weighted (unbalanced) barycenters between endpoints. For $d_{\max} = \infty$ the mass difference between left and right blocks must be compensated by transport. As d_{\max} is reduced, the mass difference is increasingly compensated by growth and shrinkage. In all experiments $\varepsilon = 10^{-6}$. Bottom: Midpoint of the Wasserstein geodesic for various ε .

requires to find the minimizer of

$$\min_{\gamma \in \mathcal{M}_+(X \times X)} \langle c, \gamma \rangle + \iota_{\{=\}}(P_{\#}^1 \gamma | \mu_k^\tau) + 2\tau \mathcal{G}(P_{\#}^2 \gamma),$$

where $c : (x, y) \mapsto |y - x|^2$ is the quadratic cost. This directly fits in our framework by choosing $F_1(s) = \iota_{\{=\}}(s dx | \mu_k^\tau)$ and $F_2(s) = 2\tau \mathcal{G}(s dx)$.

Example 5.4. If the energy is given by the relative entropy $\mathcal{G} = \mathcal{KL}(\cdot | p dx)$ for some reference measure $p dx$, then the proximal step for F_2 is given by

$$\text{prox}_{\frac{1}{\varepsilon} F_2}^{\text{KL}}(q) = \operatorname{argmin}_{s: X \rightarrow \mathbb{R}} (\varepsilon \text{KL}(s|q) + 2\tau \text{KL}(s|p)) = q^{\frac{\varepsilon}{\varepsilon+2\tau}} \cdot p^{\frac{2\tau}{\varepsilon+2\tau}}.$$

This problem is identical to an unbalanced transport problem where one of the marginal is fixed and the other is controlled by a \mathcal{KL} divergence. The difference is, that we want to solve a whole sequence of such problems, each time taking the second marginal of the optimal coupling and plugging it into the next problem as constraint for the first marginal. The gradient flow associated to this example allows us to recover the heat flow when $p dx$ is the Lebesgue measure on $X = \mathbb{R}^d$.

Example 5.5. Some models of crowd motion with congestion [51] can be approximately simulated by computing a sequence of measures (μ_k^τ) from an initial

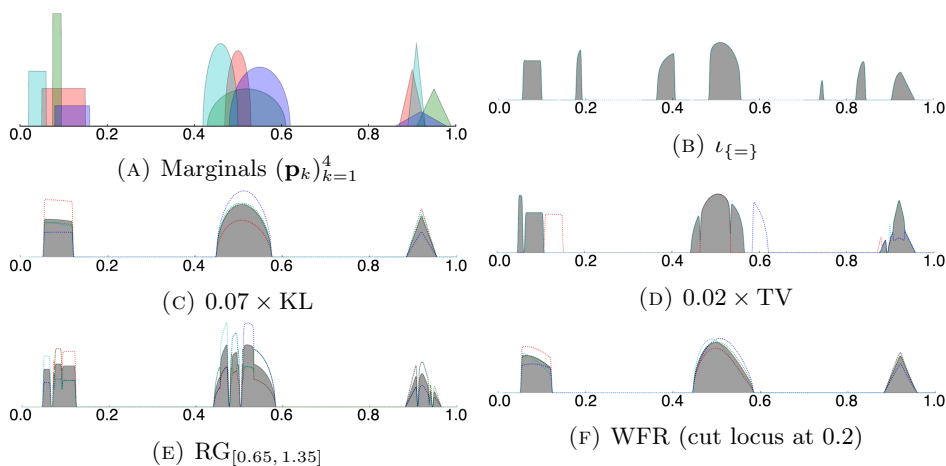


FIGURE 10. Illustration of barycenter-like problems on $X = Y = [0, 1]$. Except for (F), the function F_1 is the equality constraint with respect to the densities $(\mathbf{p}_k)_{k=1}^4$, the cost is $c(x, y) = |y - x|^2$, the weights are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and the function F_2 is of the type (5.5) with the divergence specified in the legend. (F) represents the Fréchet mean for the WFR metric. The dotted lines display the second marginal of the optimal plans.

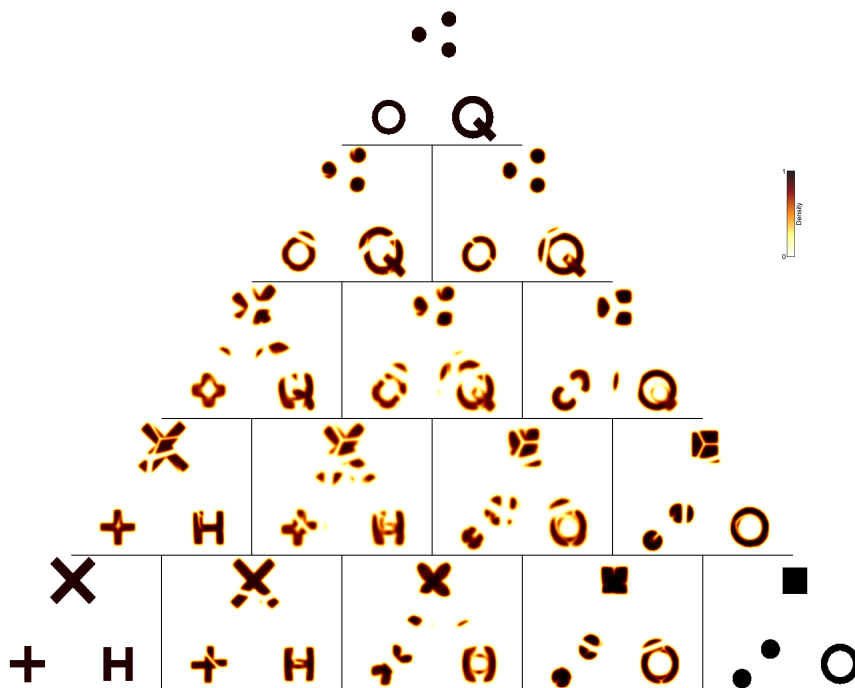


FIGURE 11. Wasserstein barycenters with entropic smoothing

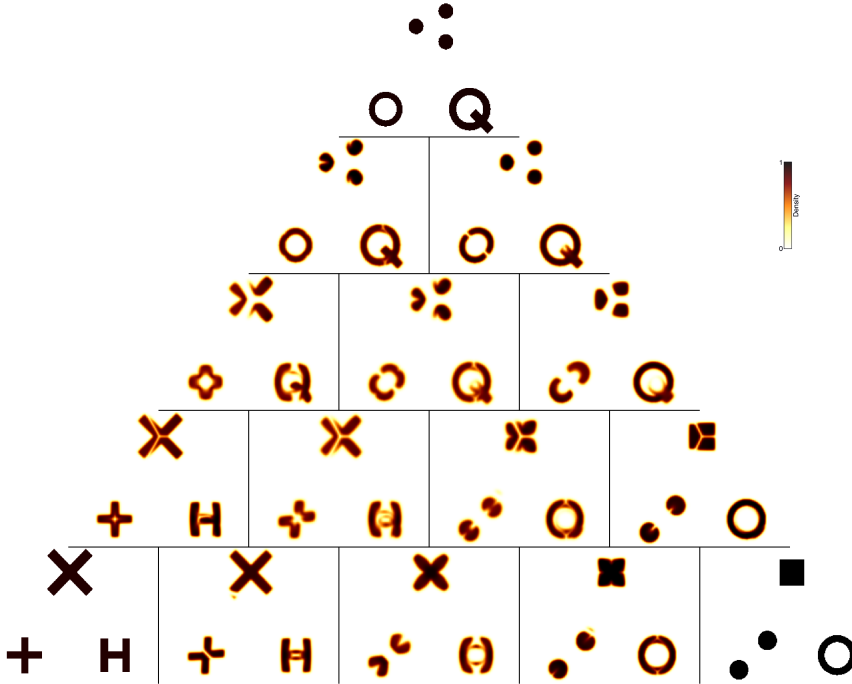


FIGURE 12. GHK barycenters with entropic smoothing (the definition of the GHK metric is similar to Wasserstein but the marginal constraints are replaced by KL divergences, see Section 2.3).

measure μ_0 , where for all $k \in \mathbb{N}$, μ_{2k+1}^τ is obtained from μ_{2k}^τ by performing a free evolution during a time τ (this requires another algorithm) and then defining μ_{2k+2}^τ as the Wasserstein projection of μ_{2k+1}^τ (which we write as $p_{2k+1} dx$) onto the set of measures with densities smaller than 1. The second step can be performed by setting $F_1(s) = \nu_{\{=\}}(s|p_{2k+1})$ and $F_2(s) = \nu_{\leq 1}(s)$. The proximal operator of F_2 at a point s is given by $\min\{1, s\}$ and thus p_{2k+2} can be obtained from p_{2k+1} with the stabilized Algorithm 2 with

$$\text{proxdiv}_{F_1}(s, u, \varepsilon) = \frac{p_{2k+1}}{s} \quad \text{and} \quad \text{proxdiv}_{F_2}(s, u, \varepsilon) = \min\left\{\frac{1}{s}, e^{-u/\varepsilon}\right\}.$$

5.3.2. WFR *gradient flows*. For WFR gradient flows, each step requires to solve

$$\inf_{\substack{\mu \in \mathcal{M}_+(X) \\ \gamma \in \mathcal{M}_+(X \times X)}} \langle c, \gamma \rangle + \mathcal{KL}(P_{\#}^1 \gamma | \mu_k^\tau) + \mathcal{KL}(P_{\#}^2 \gamma | \mu) + 2\tau \mathcal{G}(\mu),$$

where c is the cost given in (2.5). By denoting p_k the density of μ_k^τ with respect to dx , one defines the functions

$$F_1(s) = \text{KL}(s|p_k) \quad \text{and} \quad F_2(s) = \inf_{p \in L^1(X)} \text{KL}(s|p) + 2\tau G(p),$$

where $G(p) \stackrel{\text{def.}}{=} \mathcal{G}(pdx)$. If G is an integral functional $G(s) = \int_X g_x(s(x))dx$, the proximal operator is given pointwise by

$$\inf_{(\tilde{s}, p) \in \mathbb{R}^2} \varepsilon \text{KL}(\tilde{s}|s) + \text{KL}(\tilde{s}|p) + 2\tau g_x(p)$$

for which the first order optimality conditions read

$$\begin{cases} 0 = \varepsilon \log(\tilde{s}/s) + \log(\tilde{s}/p), \\ (\tilde{s}/p - 1)/(2\tau) \in \partial g_x(p), \end{cases}$$

and $\tilde{s} = 0$ if $s = 0$ or $p = 0$. In many cases, this system can be easily solved, as in the following example.

Example 5.6. One of the simplest functional which generates nontrivial gradient flows is

$$\mathcal{G}(\mu) = -\alpha \cdot \mu(X) + \iota_{\leq 1}(\mathrm{d}\mu/\mathrm{d}x),$$

where $\alpha \in]0, \infty[$. Since the distance WFR measures both the displacement and the rate of growth, one can interpret the gradient flow of \mathcal{G} as describing the evolution of a density of cells (a tumor, say) which have a tendency to multiply—hence increase the total mass—but which density cannot exceed 1. One can solve the time discretized gradient flow with Algorithm 2 by choosing $F_2(s) = \inf_p \mathrm{KL}(s|p) - 2\alpha\tau p + \iota_{\leq 1}(p)$. With the optimality conditions above, one obtains

$$\mathrm{proxdiv}_{F_2}(s, u, \varepsilon) = \begin{cases} (e^u(1 - 2\tau\alpha))^{-\frac{1}{\varepsilon}} & \text{if } s \leq e^{\frac{u}{\varepsilon}}(1 - 2\tau\alpha)^{\frac{1+\varepsilon}{\varepsilon}}, \\ (s e^u)^{-\frac{1}{1+\varepsilon}} & \text{otherwise.} \end{cases}$$

Moreover, given an optimal coupling r , one has $\mu_{k+1}^\tau = \min\left(\frac{P_{\#}^2 r}{1-2\tau\alpha}, 1\right) \mathrm{d}x$.

A numerical illustration is given in Figure 13 where we used Algorithm 2 (stopped after 500 iterations) for solving each step, with the following parameters: X is the segment $[0, 1]$ discretized into 3000 uniformly spaced samples, the initial density p_0 is the black line in Figure 13b, $\tau = 0.006$, $\alpha = 1$ and $\varepsilon = 10^{-8}$. The running time was 315 seconds. Notice how, by using a very small value for ε , the “smoothing” effect of the entropy disappears: the contours of the free-boundary which evolve with time remain sharp.

5.3.3. WFR *Gradient flows with multiple species.* The generic form (2.9) also includes gradient flows with multiple species with a mutual interaction (with $n > 1$, similar to the barycenter problem). Such systems have been theoretically studied in [43, 79]. Here we consider this class of problems in order to illustrate the versatility of the algorithm and it is not our purpose to make a link with the theory of PDEs. Let us consider a simple example which is a direct extension of Example 5.6. Consider the following functional:

$$\mathcal{G}(\mu^a, \mu^b) = -\alpha \cdot \mu^a - \alpha \cdot \mu^b + \iota_{\leq 1}(\mathrm{d}(\mu^a + \mu^b)/\mathrm{d}x).$$

In this model, one has two species which have a tendency to grow in mass (with the same incentive $\alpha > 0$, for simplicity of the algorithm), and their sum cannot exceed the reference measure. The corresponding F_2 is given by

$$F_2(s^a, s^b) = \inf_{(r^a, r^b) \in \mathbb{L}^1(X)^2} \mathrm{KL}(s^a|r^a) + \mathrm{KL}(s^b|r^b) - 2\alpha\tau \int_X (r^a + r^b) \mathrm{d}x + \iota_{\leq 1}(r^a + r^b).$$

The optimality conditions yield

$$\mathrm{proxdiv}_{F_2}(s, u, \varepsilon) = (e^{-\frac{u^a}{\varepsilon}}, e^{-\frac{u^b}{\varepsilon}}) / \beta(s^a e^{-\frac{u^a}{\varepsilon}}, s^b e^{-\frac{u^b}{\varepsilon}}),$$

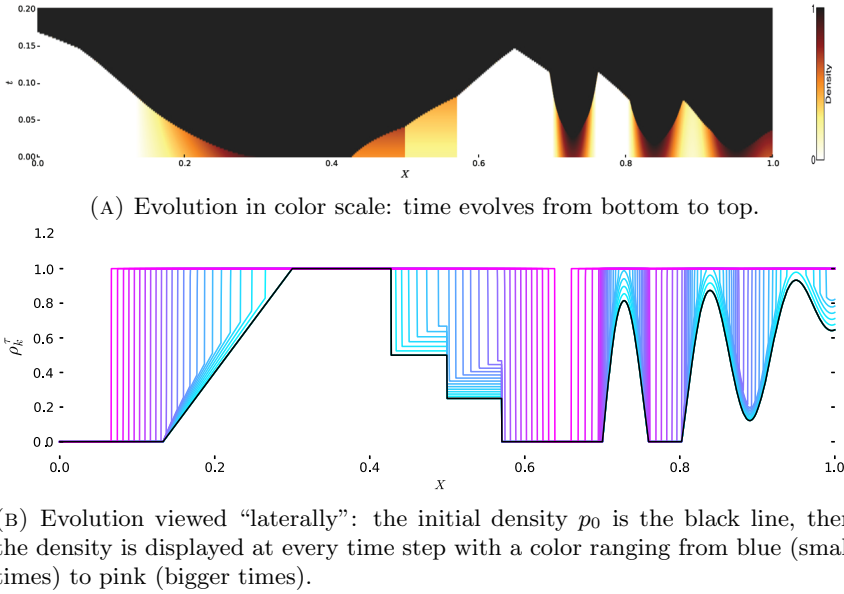


FIGURE 13. Evolution of the density with respect to time for the growth model of Example 5.6.

where $\beta(x, y) \stackrel{\text{def.}}{=} \max \left\{ (x + y)^{\frac{1}{1+\varepsilon}}, (1 - 2\tau\alpha)^{\frac{1}{\varepsilon}} \right\}$. Moreover, given an optimal pair of couplings r^a, r^b , the next densities are given by

$$(p_{k+1}^a, p_{k+1}^b) = (P_{\#}^2 r^a, P_{\#}^2 r^b) / \beta(P_{\#}^2 r^a, P_{\#}^2 r^b).$$

Notice that in this model, as in Example 5.6, if the domain X is compact and the initial densities are not null, a steady state is reached in finite time, where the sum of the two densities is constant and equal to 1.

Some initial densities and the associated final steady state are shown in Figure 14. For this illustration, we started with input densities p^a and p^b on the segment $[0, 1]$ discretized into 3000 uniform samples, as displayed in Figure 14a (where the red density p^b is layered over p^a). We computed the evolution with Algorithm 2 for solving each step of the discretized gradient flow, with the parameters $\tau = 0.004$, $\alpha = 1$ and $\varepsilon = 10^{-7}$.

Note that although the incentive of growing mass α is the same for the two species, the resulting interaction is nontrivial; for instance, the small amount of blue mass is pushed to the right by the action of the expanding red mass. This behavior is explained by the fact that for the WFR metric, it requires less effort (i.e., the distance is smaller) to add a given amount of mass to a high density than to a small one.

APPENDIX A

A.1. Reminders on convex analysis. Let E and E^* be *topologically paired* vector spaces, i.e., vector spaces assigned with locally convex Hausdorff topology such that all continuous linear functionals on each space can be identified with the elements of the other. The pairing between those spaces is the bilinear form

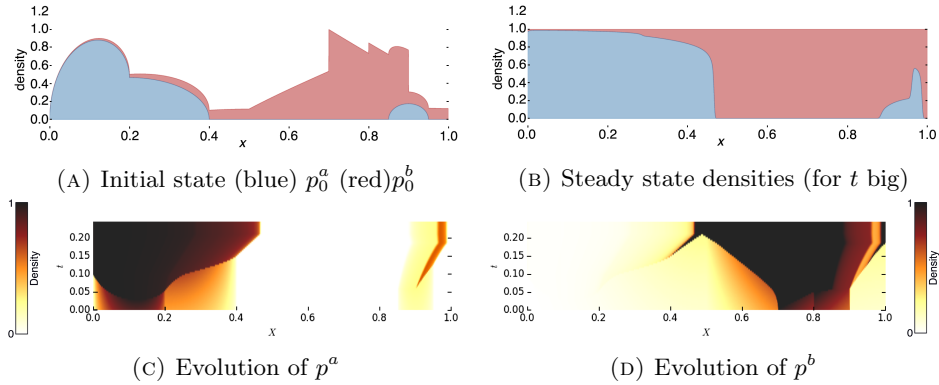


FIGURE 14. Top row: Initial state and steady state (density p^b is layered over p^a); Bottom row: Time evolution (time evolves from bottom to top) with density in color scale.

$\langle \cdot, \cdot \rangle : E \times E^* \rightarrow \mathbb{R}$. The convex conjugate of a function $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ is defined for each $y \in E^*$ by

$$f^*(y) \stackrel{\text{def.}}{=} \sup_{x \in E} \langle x, y \rangle - f(x).$$

The subdifferential operator is defined at a point $x \in E$ as

$$\partial f(x) \stackrel{\text{def.}}{=} \{y \in E^*; f(x') - f(x) \geq \langle y, x' - x \rangle \text{ for all } x' \in E\}$$

and is empty if $f(x) = \infty$. Those definitions admit their natural counterparts for functions defined on E^* .

Theorem A.1 (Fenchel-Rockafellar [62]). *Let (E, E^*) and (F, F^*) be two couples of topologically paired spaces. Let $A : E \rightarrow F$ be a continuous linear operator and $A^* : F^* \rightarrow E^*$ its adjoint. Let f and g be lower semicontinuous and proper convex functions defined on E and F , respectively. If there exists $x \in \text{dom } f$ such that g is continuous at Ax , then*

$$\sup_{x \in E} -f(-x) - g(Ax) = \min_{y^* \in F^*} f^*(A^*y^*) + g^*(y^*)$$

and the min is attained. Moreover, if there exists a maximizer $x \in E$, then there exists $y^* \in F^*$ satisfying $Ax \in \partial g^*(y^*)$ and $A^*y^* \in \partial f(-x)$.

A.2. Properties of divergence functionals. Here we collect a few results on divergences functionals when they are defined on functions as in (5.2) (as opposed to Section 2.1 where they are defined between measures).

Proposition A.2. *Let φ be a nonnegative entropy function as in Definition 2.1 and (X, dx) a measured space. Then $(u, v) \in L^1(X)^2 \mapsto D_\varphi(u|v)$ is an admissible integral functional (in the sense of Definition 3.5) which is positively 1-homogeneous, convex and weakly lower semicontinuous. Moreover, $D_\varphi^* = \iota_{B_\varphi}$ with $B_\varphi = \{(a, b) \in \mathbb{R}^2; b \leq -\varphi^*(a)\}$.*

Proof. As a preliminary, note that for $(u, v) \in \mathbb{R}^2$,

$$\bar{D}^*(u|v) = \sup_{a, b \in \mathbb{R}_+} \begin{cases} b(u \cdot a/b + v - \varphi(a/b)) & \text{if } b > 0, \\ a(u - \varphi'_\infty) & \text{otherwise} \end{cases} = \begin{cases} 0 & \text{if } v \leq \varphi^*(u), \\ \infty & \text{otherwise.} \end{cases}$$

Now, the function $(x, u, v) \in X \times \mathbb{R}^2 \mapsto \overline{D}_\varphi(u, v)$ defined in (5.2) is a normal integrand since it does not depend on x . Moreover, there exists feasible points for D_φ (take $v \in L^1(X)$ and $u = \alpha v$ where $\alpha \in \text{dom } \varphi$) and for the integral functional associated to \overline{D}_φ^* (given its expression above). The conclusion follows by [64, Theorem 3C]. \square

Proposition A.3. *Let (X, dx) be a measured space, $v \in L^1_+(X)$ and φ a nonnegative entropy function as in Definition 2.1. Then $D_\varphi(\cdot|v)$ is a proper, weakly lower semicontinuous convex function on $L^1(X)$ and its convex conjugate is given, for $a \in L^\infty(X)$, by*

$$D_\varphi^*(a|v) \stackrel{\text{def.}}{=} \int_X \varphi^*(a(x))v(x)dx + \int_X \iota_{\leq \varphi'_\infty}(a(x))dx,$$

where φ^* is the convex conjugate of φ .

Moreover, the subdifferential $\partial D_\varphi(\cdot|v)$ at a point $u \in L^1(X)$ is the set of functions $a \in L^\infty(X)$ such that $\varphi'_\infty - a$ is nonnegative and such that, for a.e. x , where $v(x) > 0$, $a(x) \in \partial\varphi(u(x)/v(x))$.

Similarly, the subdifferential $\partial D_\varphi^*(\cdot|v)$ at a point $a \in L^\infty(X)$ bounded above by φ'_∞ is the set of nonnegative functions $u \in L^1(X)$ such that, for a.e. x , $u(x) \in \partial\varphi^*(a(x))v(x)$ if $v(x) > 0$ and $u(x) = 0$ if $v(x) = 0$ and $a(x) < \varphi'_\infty$.

Proof. By [65, Proposition 14.45c], $(x, u) \in X \times \mathbb{R} \mapsto \overline{D}_\varphi(u|v(x))$ is a normal integrand. Then [64, Theorem 3C] and Corollaries apply and conjugation and subdifferentiation can be performed pointwise. \square

A.3. Proof of the iterates for the barycenter problems. We explain below how to derive the expression for h which are given in Table 2, by applying Proposition 5.3. Assume that $(s_i)_{i=1}^n \in \mathbb{R}^n \geq 0$ is given.

A.3.1. Case $D_\varphi = \iota_{\{=\}}$. This case is simple because solving (5.6) boils down to solving the one-dimensional problem $\min_h \sum \alpha_k \overline{\text{KL}}(h|s_k)$, which is direct with first order optimality conditions.

A.3.2. Case $D_\varphi = \lambda \text{KL}$. First note that the assumption of Proposition 5.3 is satisfied and that $h = 0$ if and only if for all k , $s_k = 0$ (otherwise, the joint subdifferential is empty). Since φ is smooth, its joint subdifferential is a singleton $\partial \overline{\text{KL}}(\tilde{s}|h) = \{(\log(\tilde{s}/h), 1 - \tilde{s}/h)\}$ if $\tilde{s}, h > 0$. Also, since $\overline{\text{KL}}(0|h) = h + \iota_{]0, \infty[}(h)$, one has $\partial_2 \overline{\text{KL}}(0, h) = \{1\}$ if $h > 0$. Thus, optimality conditions in Proposition 5.3 yields the system

$$\begin{cases} \log \frac{\tilde{s}_k}{h} = \frac{\varepsilon}{\lambda} \log \frac{s_k}{\tilde{s}_k} & \text{if } s_k > 0, \\ \tilde{s}_k = 0 & \text{if } s_k = 0, \\ \sum \alpha_k (1 - \frac{\tilde{s}_k}{h}) = 0. \end{cases}$$

A.3.3. Case $D_\varphi = \lambda \text{TV}$. By Proposition A.2, one has $\overline{D}_{\varphi_{\text{TV}}}(\tilde{s}|h) = \sup_{(a,b) \in B} a \cdot x + b \cdot y$ with $B = \{(a, b) \in \mathbb{R}^2; a \leq 1, b \leq 1, a + b \leq 0\}$. The set of points in B at which this supremum is attained is easy to see graphically and gives the set

$\partial\overline{D}_{\varphi_{\text{TV}}}(\tilde{s}|h)$. With the notations of Proposition 5.3, one has with $a_k = \frac{\varepsilon}{\lambda} \log \frac{s_k}{\tilde{s}_k}$,

$$\begin{aligned} (1) \quad \tilde{s}_k > h > 0 &\Leftrightarrow -b_k = a_k = 1, & (2) \quad h > \tilde{s}_k > 0 &\Leftrightarrow -b_k = a_k = -1, \\ (3) \quad \tilde{s}_k = h > 0 &\Leftrightarrow -b_k = a_k \in [-1, 1], & (4) \quad h > \tilde{s}_k = 0 &\Leftrightarrow b_k = 1, \\ (5) \quad \tilde{s}_k > h = 0 &\Leftrightarrow a_k = 1 \text{ and } b_k \leq -1, & (6) \quad \tilde{s}_k = h = 0 &\Leftrightarrow b_k \leq 1. \end{aligned}$$

Let us first deal with the case $h = 0$ (cases (5) and (6)). Condition $\sum \alpha_k b_k = 0$ from Proposition 5.3 says that it is the case if and only if $\sum_{k \notin I_+} \alpha_k \geq \sum_{k \in I_+} \alpha_k$. Now assume that $h > 0$. If $\tilde{s}_k > 0$ (cases (1), (2) and (3)) then b_k can be expressed as $\max(-1, \min(1, \frac{\varepsilon}{\lambda} \log \frac{h}{\tilde{s}_k}))$ otherwise $b_k = 1$. The implicit expression given for h is thus the condition $\sum \alpha_k b_k = 0$.

A.3.4. *Case* $D_\varphi = \text{RG}_{[\beta_1, \beta_2]}$. In this case, \overline{D}_φ is the support function of $B = \{(a, b) \in \mathbb{R}^2; \text{ for } i \in \{1, 2\}, b \leq -\beta_i \cdot a\}$. With the notations of Proposition 5.3, one has with $a_k = \frac{\varepsilon}{\lambda} \log \frac{s_k}{\tilde{s}_k}$,

$$\begin{aligned} (1) \quad 0 < \beta_1 h < \tilde{s}_k < \beta_2 h &\Leftrightarrow a_k = b_k = 0, & (2) \quad 0 < \beta_1 h = \tilde{s}_k &\Leftrightarrow b_k = -\beta_1 a_k, \\ (3) \quad 0 < \beta_2 h = \tilde{s}_k &\Leftrightarrow b_k = -\beta_2 a_k, & (4) \quad 0 = h = \tilde{s}_k &\Leftrightarrow (b_k, a_k) \in B. \end{aligned}$$

If $s_k = 0$ for some $k \in \{1, \dots, n\}$ then $h = 0$ (this is the only feasible point). Otherwise, $h > 0$ and the condition $\sum \alpha_k b_k = 0$ gives the implicit equation.

ACKNOWLEDGMENTS

The work of Bernhard Schmitzer was supported by the French National Research Agency (ANR) as part of the ‘Investissements d’avenir’, program-reference ANR-10-LABX-0098 via the Fondation Sciences Mathématiques de Paris. The work of Bernhard Schmitzer and Gabriel Peyré was supported by the European Research Council (ERC project SIGMA-Vision).

REFERENCES

- [1] M. Agueh and G. Carlier, *Barycenters in the Wasserstein space*, SIAM J. Math. Anal. **43** (2011), no. 2, 904–924, DOI 10.1137/100805741. MR2801182
- [2] L.-P. Saumier, B. Khouider, and M. Agueh, *Optimal transport for particle image velocimetry*, Commun. Math. Sci. **13** (2015), no. 1, 269–296, DOI 10.4310/CMS.2015.v13.n1.a13. MR3238148
- [3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient flows in metric spaces and in the space of probability measures*, 2nd ed., Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2008. MR2401600
- [4] F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica **20** (1998), no. 1, 61–76, DOI 10.1007/PL00009187. MR1483422
- [5] H. H. Bauschke and A. S. Lewis, *Dykstra’s algorithm with Bregman projections: a convergence proof*, Optimization **48** (2000), no. 4, 409–427, DOI 10.1080/02331930008844513. MR1811866
- [6] A. Beck, *On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes*, SIAM J. Optim. **25** (2015), no. 1, 185–209, DOI 10.1137/13094829X. MR3300410
- [7] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math. **84** (2000), no. 3, 375–393, DOI 10.1007/s002110050002. MR1738163
- [8] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput. **37** (2015), no. 2, A1111–A1138, DOI 10.1137/141000439. MR3340204
- [9] J.-D. Benamou, G. Carlier, Q. Mérigot, and É. Oudet, *Discretization of functionals involving the Monge-Ampère operator*, Numer. Math. **134** (2016), no. 3, 611–636, DOI 10.1007/s00211-015-0781-y. MR3555350

- [10] Jean-David Benamou, Guillaume Carlier, and Maxime Laborde, *An augmented lagrangian approach to wasserstein gradient flows and applications*, (2015).
- [11] D. P. Bertsekas and D. A. Castañón, *The auction algorithm for the transportation problem*, Ann. Oper. Res. **20** (1989), no. 1-4, 67–96, DOI 10.1007/BF02216923. Network optimization and applications. MR1015946
- [12] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich, *Displacement interpolation using Lagrangian mass transport*, ACM Transactions on Graphics (SIGGRAPH ASIA'11) **30** (2011), no. 6.
- [13] J. M. Borwein and A. S. Lewis, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim. **29** (1991), no. 2, 325–338, DOI 10.1137/0329017. MR1092730
- [14] L. M. Brègman, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming* (Russian), Ž. Vyčisl. Mat. i Mat. Fiz. **7** (1967), 620–631. MR0215617
- [15] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417, DOI 10.1002/cpa.3160440402. MR1100809
- [16] M. Burger, J. A. Carrillo, and M.-T. Wolfram, *A mixed finite element method for nonlinear diffusion equations*, Kinet. Relat. Models **3** (2010), no. 1, 59–83, DOI 10.3934/krm.2010.3.59. MR2580954
- [17] M. Burger, M. Franek, and C.-B. Schönlieb, *Regularized regression and density estimation based on optimal transport*, Appl. Math. Res. Express. AMRX **2** (2012), 209–253. MR2982774
- [18] R. Burkard, M. Dell’Amico, and S. Martello, *Assignment problems*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009. MR2488749
- [19] L. A. Caffarelli and R. J. McCann, *Free boundaries in optimal transport and Monge-Ampère obstacle problems*, Ann. of Math. (2) **171** (2010), no. 2, 673–730, DOI 10.4007/annals.2010.171.673. MR2630054
- [20] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer, *Convergence of entropic schemes for optimal transport and gradient flows*, SIAM J. Math. Anal. **49** (2017), no. 2, 1385–1418, DOI 10.1137/15M1050264. MR3635459
- [21] J. A. Carrillo, A. Chertock, and Y. Huang, *A finite-volume method for nonlinear nonlocal equations with a gradient flow structure*, Commun. Comput. Phys. **17** (2015), no. 1, 233–258, DOI 10.4208/cicp.160214.010814a. MR3372289
- [22] Y. Censor and S. Reich, *The Dykstra algorithm with Bregman projections*, Commun. Appl. Anal. **2** (1998), no. 3, 407–419. MR1626725
- [23] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal-dual algorithm*, Math. Program. **159** (2016), no. 1-2, Ser. A, 253–287, DOI 10.1007/s10107-015-0957-3. MR3535925
- [24] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Unbalanced optimal transport: Geometry and kantovich formulation*, Preprint 1508.05216, Arxiv, 2015.
- [25] L. Chizat, B. Schmitzer, G. Peyré, and F.-X. Vialard, *An interpolating distance between optimal transport and Fischer-Rao*, Preprint 1506.06430, Arxiv, 2015.
- [26] M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport*, Advances in Neural Information Processing Systems (NIPS) 26, 2013, pp. 2292–2300.
- [27] M. Cuturi and A. Doucet, *Fast computation of Wasserstein barycenters*, Proceedings of ICML, vol. 32, 2014.
- [28] J. Delon, *Movie and video scale-time equalization application to flicker reduction*, IEEE Transactions on Image Processing **15** (2006), no. 1, 241–248.
- [29] W. E. Deming and F. F. Stephan, *On a least squares adjustment of a sampled frequency table when the expected marginal totals are known*, Ann. Math. Statistics **11** (1940), 427–444. MR0003527
- [30] R. L. Dykstra, *An algorithm for restricted least squares regression*, J. Amer. Statist. Assoc. **78** (1983), no. 384, 837–842. MR727568
- [31] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Math. Oper. Res. **18** (1993), no. 1, 202–226, DOI 10.1287/moor.18.1.202. MR1250114
- [32] J. M. Fadili and G. Peyré, *Total variation projection with first order schemes*, IEEE Trans. Image Process. **20** (2011), no. 3, 657–669, DOI 10.1109/TIP.2010.2072512. MR2799177

- [33] A. Figalli, *The optimal partial transport problem*, Arch. Ration. Mech. Anal. **195** (2010), no. 2, 533–560, DOI 10.1007/s00205-008-0212-7. MR2592287
- [34] J. Franklin and J. Lorenz, *On the scaling of multidimensional matrices*, Linear Algebra Appl. **114/115** (1989), 717–735, DOI 10.1016/0024-3795(89)90490-4. MR986904
- [35] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio, *Learning with a Wasserstein loss*, Preprint 1506.05439, Arxiv, 2015.
- [36] A. Galichon and B. Salanić, *Matching with trade-offs: Revealed preferences over competing characteristics*, Tech. report, Preprint SSRN-1487307, 2009.
- [37] T. O. Gallouët and L. Monsaingeon, *A JKO splitting scheme for Kantorovich-Fisher-Rao gradient flows*, SIAM J. Math. Anal. **49** (2017), no. 2, 1100–1130, DOI 10.1137/16M106666X. MR3628313
- [38] L. G. Hanin, *Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces*, Proc. Amer. Math. Soc. **115** (1992), no. 2, 345–352, DOI 10.2307/2159251. MR1097344
- [39] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker-Planck equation*, SIAM J. Math. Anal. **29** (1998), no. 1, 1–17, DOI 10.1137/S0036141096303359. MR1617171
- [40] L. Kantorovitch, *On the translocation of masses*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **37** (1942), 199–201. MR0009619
- [41] S. Kondratyev, L. Monsaingeon, and D. Vorotnikov, *A new optimal transport distance on the space of finite Radon measures*, Tech. report, Pre-print, 2015.
- [42] J.J. Kosowsky and A.L. Yuille, *The invisible hand algorithm: Solving the assignment problem with statistical physics*, Neural Networks **7** (1994), no. 3, 477–490.
- [43] Maxime Laborde, *On some non linear evolution systems which are perturbations of wasserstein gradient flows*, arXiv preprint arXiv:1506.00126 (2015).
- [44] C. Léonard, *From the Schrödinger problem to the Monge-Kantorovich problem*, J. Funct. Anal. **262** (2012), no. 4, 1879–1920, DOI 10.1016/j.jfa.2011.11.026. MR2873864
- [45] C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1533–1574, DOI 10.3934/dcds.2014.34.1533. MR3121631
- [46] B. Lévy, *A numerical algorithm for L_2 semidiscrete optimal transport in 3D*, ESAIM Math. Model. Numer. Anal. **49** (2015), no. 6, 1693–1715, DOI 10.1051/m2an/2015055. MR3423272
- [47] M. Liero, A. Mielke, and G. Savaré, *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, SIAM J. Math. Anal. **48** (2016), no. 4, 2869–2911, DOI 10.1137/15M1041420. MR3542003
- [48] Matthias Liero, Alexander Mielke, and Giuseppe Savaré, *Optimal transport in competition with reaction: the Hellinger-Kantorovich distance and geodesic curves*, ArXiv e-prints (2015).
- [49] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon, *A generalized model for optimal transport of images including dissipation and density modulation*, ESAIM Math. Model. Numer. Anal. **49** (2015), no. 6, 1745–1769, DOI 10.1051/m2an/2015043. MR3423274
- [50] B. Maury, A. Roudneff-Chupin, and F. Santambrogio, *A macroscopic crowd motion model of gradient flow type*, Math. Models Methods Appl. Sci. **20** (2010), no. 10, 1787–1821, DOI 10.1142/S0218202510004799. MR2735914
- [51] B. Maury, A. Roudneff-Chupin, F. Santambrogio, and J. Venel, *Handling congestion in crowd motion modeling*, Netw. Heterog. Media **6** (2011), no. 3, 485–519, DOI 10.3934/nhm.2011.6.485. MR2826756
- [52] Q. Mérigot, *A multiscale approach to optimal transport*, Comput. Graph. Forum **30** (2011), no. 5, 1583–1592.
- [53] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l’Académie Royale des Sciences (1781), 666–704.
- [54] Q. V. Nguyen, *Forward-backward splitting with Bregman distances*, arXiv:1505.05198 (2015).
- [55] F. Otto, *The geometry of dissipative evolution equations: the porous medium equation*, Comm. Partial Differential Equations **26** (2001), no. 1-2, 101–174, DOI 10.1081/PDE-100002243. MR1842429
- [56] N. Papadakis, G. Peyré, and E. Oudet, *Optimal transport with proximal splitting*, SIAM J. Imaging Sci. **7** (2014), no. 1, 212–238, DOI 10.1137/130920058. MR3158785

- [57] B. Perthame, F. Quirós, and J. L. Vázquez, *The Hele-Shaw asymptotics for mechanical models of tumor growth*, Arch. Ration. Mech. Anal. **212** (2014), no. 1, 93–127, DOI 10.1007/s00205-013-0704-y. MR3162474
- [58] G. Peyré, *Entropic approximation of Wasserstein gradient flows*, SIAM J. Imaging Sci. **8** (2015), no. 4, 2323–2351, DOI 10.1137/15M1010087. MR3413589
- [59] B. Piccoli and F. Rossi, *Generalized Wasserstein distance and its application to transport equations with source*, Arch. Ration. Mech. Anal. **211** (2014), no. 1, 335–358, DOI 10.1007/s00205-013-0669-x. MR3182483
- [60] F. Pitié, A.C. Kokaram, and R. Dahyot, *Automated colour grading using colour distribution transfer*, Computer Vision and Image Understanding **107** (2007), no. 1, 123–137.
- [61] J. Rabin and N. Papadakis, *Convex color image segmentation with optimal transport distances*, Scale space and variational methods in computer vision, Lecture Notes in Comput. Sci., vol. 9087, Springer, Cham, 2015, pp. 256–269, DOI 10.1007/978-3-319-18461-6_21. MR3394935
- [62] R. T. Rockafellar, *Duality and stability in extremum problems involving convex functions*, Pacific J. Math. **21** (1967), 167–187. MR0211759
- [63] R. T. Rockafellar, *Integrals which are convex functionals*, Pacific J. Math. **24** (1968), 525–539. MR0236689
- [64] R. T. Rockafellar, *Integral functionals, normal integrands and measurable selections*, Nonlinear operators and the calculus of variations, Springer, 1976, pp. 157–207.
- [65] R.T. Rockafellar and R. J-B. Wets, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009.
- [66] Y. Rubner, C. Tomasi, and L.J. Guibas, *The earth mover’s distance as a metric for image retrieval*, International Journal of Computer Vision **40** (2000), no. 2.
- [67] F. Santambrogio, *Optimal transport for applied mathematicians*, Progress in Nonlinear Differential Equations and their Applications, vol. 87, Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling. MR3409718
- [68] B. Schmitzer, *A sparse multiscale algorithm for dense optimal transport*, J. Math. Imaging Vision **56** (2016), no. 2, 238–259, DOI 10.1007/s10851-016-0653-9. MR3535020
- [69] Bernhard Schmitzer, *Stabilized sparse scaling algorithms for entropy regularized transport problems*, <https://arxiv.org/abs/1610.06519>, 2016.
- [70] E. Schrödinger, *Über die Umkehrung der Naturgesetze*, Sitzungsberichte Preuss. Akad. Wiss. Berlin. Phys. Math. **144** (1931), 144–153.
- [71] Meisam Sharify, Stéphane Gaubert, and Laura Grigori, *Solution of the optimal assignment problem by diagonal scaling algorithms*, arXiv preprint arXiv:1104.3830 (2011).
- [72] R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879, DOI 10.1214/aoms/1177703591. MR0161868
- [73] J. Solomon, F. de Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas, *Convolutional Wasserstein distances: Efficient optimal transportation on geometric domains*, ACM Transactions on Graphics (Proc. SIGGRAPH 2015) **34** (2015), no. 4, 66:1–66:11.
- [74] J. Solomon, R.M. Rustamov, L. Guibas, and A. Butscher, *Wasserstein propagation for semisupervised learning*, Proc. ICML 2014, 2014.
- [75] A. C. Thompson, *On certain contraction mappings in a partially ordered vector space.*, Proc. Amer. Math. soc. **14** (1963), 438–443, DOI 10.2307/2033816. MR0149237
- [76] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR1964483
- [77] C. Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009. Old and new. MR2459454
- [78] H. Wang and A. Banerjee, *Bregman alternating direction method of multipliers*, preprint arXiv:1306.3203 (2014).
- [79] J. Zinsl and D. Matthes, *Transport distances and geodesic convexity for systems of degenerate diffusion equations*, Calc. Var. Partial Differential Equations **54** (2015), no. 4, 3397–3438, DOI 10.1007/s00526-015-0909-z. MR3426082

CEREMADE, CNRS, UNIVERSITÉ PARIS-DAUPHINE, INRIA PROJECT TEAM MOKAPLAN,
FRANCE

Email address: chizat@ceremade.dauphine.fr

CNRS AND DMA, ÉCOLE NORMALE SUPÉRIEURE, INRIA PROJECT TEAM MOKAPLAN, FRANCE

Email address: gabriel.peyre@ens.fr

CEREMADE, CNRS, UNIVERSITÉ PARIS-DAUPHINE, INRIA PROJECT TEAM MOKAPLAN,
FRANCE

Email address: schmitzer@ceremade.dauphine.fr

CEREMADE, CNRS, UNIVERSITÉ PARIS-DAUPHINE, INRIA PROJECT TEAM MOKAPLAN,
FRANCE

Email address: vialard@ceremade.dauphine.fr