Equitable and Optimal Transport with Multiple Agents

Meyer Scetbon* CREST, ENSAE, Inst. Polytechnique de Paris

Laurent Meunier* Facebook AI Research Univ. Paris-Dauphine

Jamal Atif CNRS, LAMSADE, Univ. Paris-Dauphine

Marco Cuturi Google Brain CREST, ENSAE

Abstract

We introduce an extension of the Optimal Transport problem when multiple costs are involved. Considering each cost as an agent, we aim to share equally between agents the work of transporting one distribution to another. To do so, we minimize the transportation cost of the agent who works the most. Another point of view is when the goal is to partition equitably goods between agents according to their heterogeneous preferences. Here we aim to maximize the utility of the least advantaged agent. This is a fair division problem. Like Optimal Transport, the problem can be cast as a linear optimization problem. When there is only one agent, we recover the Optimal Transport problem. When two agents are considered, we are able to recover Integral Probability Metrics defined by α-Hölder functions, which include the widely-known Dudley metric. To the best of our knowledge, this is the first time a link is given between the Dudley metric and Optimal Transport. We provide an entropic regularization of that problem which leads to an alternative algorithm faster than the standard linear program.

1 Introduction

Optimal Transport (OT) has gained interest last years in machine learning with diverse applications in neuroimaging (Janati et al., 2020), generative models (Arjovsky et al., 2017; Salimans et al., 2018), supervised learning (Courty et al., 2016), word embeddings (Alvarez-Melis et al., 2018), reconstruction cell trajectories (Yang et al., 2020; Schiebinger et al., 2019) or adversarial examples (Wong et al., 2019). The key to use OT in these applications lies in the gain of computation efficiency thanks to regularizations that smoothes the OT problem. More specifically, when one uses an entropic penalty, one recovers the so called Sinkhorn distances (Cuturi, 2013). In this paper, we introduce a new family of variational problems extending the optimal transport problem when multiple costs are involved with various applications in fair division of goods/work and operations research problems.

Fair division (Steinhaus, 1949) has been widely studied by the artificial intelligence (Lattimore et al., 2015) and economics (Moulin, 2004) communities. Fair division consists in partitioning diverse resources among agents according to some fairness criteria. One of the standard problems in fair division is the fair cake-cutting problem (Dubins and Spanier, 1961; Brandt et al., 2016). The cake is an heterogeneous resource, such as a cake with different toppings, and the agents have heterogeneous preferences over different parts of the cake, i.e., some people prefer the chocolate toppings, some prefer the cherries, others just want a piece as large as possible. Hence, taking into account these preferences, one might share the cake equitably between the agents. A generalization of this problem, for which achieving fairness constraints is more challenging, is when the splitting involves several heterogeneous cakes, and where the agents have linked preferences over the different parts of the cakes. This problem has many variants such as the cake-cutting with two cakes (Cloutier et al., 2010), or the Multi Type Resource Allocation (Mackin and Xia, 2015; Wang et al., 2019). In all these models it is assumed that there is only one indivisible unit per type of resource available in each cake, and once an agent choose it, he or she has to take it all. In this setting, the cake can be seen as a set where each element of the set represents a type of resource, for instance each element of the cake represents a topping. A natural relaxation of these problems is when a divisible quantity of each type of resources is available. We introduce EOT (**E**quitable and **O**ptimal **T**ransport), a formulation that solves both the cake-cutting and the cake-cutting with two cakes problems in this setting.

---

* Authors contributed equally. Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021, San Diego, California, USA. PMLR: Volume 130.

Our problem expresses as an optimal transportation problem. Hence, we prove duality results and provide fast computation based on Sinkhorn algorithm. As interesting properties, some Integral Probability Metrics (IPMs) (Müller, 1997) as Dudley metric (Dudley et al., 1966), or standard Wasserstein metric (Villani, 2003) are particular cases of the EOT problem.

Contributions. In this paper we introduce EOT an extension of Optimal Transport which aims at finding an equitable and optimal transportation strategy between multiple agents. We make the following contributions:

- In Section 3, we introduce the problem and show that it solves a fair division problem where heterogeneous resources have to be shared among multiple agents. We derive its dual and prove strong duality results. As a by-product, we show that EOT is related to some usual IPMs families and in particular the widely known Dudley metric.
- In Section 4, we propose an entropic regularized version of the problem, derive its dual formulation, obtain strong duality. We then provide an efficient algorithm to compute EOT. Finally we propose other applications of EOT for Operations Research problems.

2 Related Work

Optimal Transport. Optimal transport aims to move a distribution towards another at lowest cost. More formally, if c is a cost function on the ground space $\mathcal{X}\times\mathcal{Y}$, then the relaxed Kantorovich formulation of OT is defined for μ and ν two distributions as

$$\mathrm{W}_c(\mu,\nu) := \inf_{\gamma}\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\gamma(x,y)$$

where the infimum is taken over all distributions γ with marginals μ and ν. Kantorovich theorem states the following strong duality result under mild assumptions (Villani, 2003)

$$\mathrm{W}_c(\mu,\nu) = \sup_{f,g}\int_{\mathcal{X}} f(x)d\mu(x) + \int_{\mathcal{Y}} g(y)d\nu(y)$$

where the supremum is taken over continuous bounded functions satisfying for all x,y, $f(x)+g(y)\leq c(x,y)$. The question of considering an optimal transport problem when multiple costs are involved has already been raised in recent works. For instance, (Paty and Cuturi, 2019) proposed a robust Wasserstein distance where the distributions are projected on a k-dimensional subspace that maximizes their transport cost. In that sense, they aim to choose the most expensive cost among Mahalanobis square distances with kernels of rank k. In articles (Li et al., 2019; Sun et al., 2020), the authors aim to learn a cost given observed matchings by inversing the optimal transport problem (Dupuy et al., 2016). In (Petrovich et al., 2020) the authors study "feature-robust" optimal transport, which can be also seen as a robust cost selection for optimal transport. In articles (Genevay et al., 2017; Scetbon and Cuturi, 2020), the authors learn an adversarial cost to train a generative adversarial network. Here, we do not aim to consider a worst case scenario among the available costs but rather consider that the costs work together in order to split equitably the transportation problem among them at lowest cost.

Entropic relaxation of OT. Computing exactly the optimal transport cost requires solving a linear program with a supercubic complexity ($n^3\log n$) (Tarjan, 1997) that results in an output that is *not* differentiable with respect to the measures' locations or weights (Bertsimas and Tsitsiklis, 1997). Moreover, OT suffers from the curse of dimensionality (Dudley, 1969; Fournier and Guillin, 2015) and is therefore likely to be meaningless when used on samples from high-dimensional densities. Following the line of work introduced by Cuturi (2013), we propose an approximated computation of our problem by regularizing it with an entropic term. Such regularization in OT accelerates the computation, makes the problem differentiable with regards to the distributions (Feydy et al., 2018) and reduces the curse of dimensionality (Genevay et al., 2018). Taking the dual of the approximation, we obtain a smooth and convex optimization problem under a simplicial constraint.

Fair Division. Fair division of goods has a long standing history in economics and computational choice. A classical problem is the fair cake-cutting that consists in splitting the cake between N individuals according to their heterogeneous preferences. The cake $\mathcal{X}$, viewed as a set, is divided in $\mathcal{X}_1,\dots,\mathcal{X}_N$ disjoint sets among the N individuals. The utility for a single individual i for a slice S is denoted $V_i(S)$. It is often assumed that $V_i(\mathcal{X})=1$ and that V_i is additive for disjoint sets. There exists many criteria to assess fairness for a partition $\mathcal{X}_1,\dots,\mathcal{X}_N$ such as proportionality ($V_i(\mathcal{X}_i)\geq 1/N$), envy-freeness ($V_i(\mathcal{X}_i)\geq V_i(\mathcal{X}_j)$) or equitability ($V_i(\mathcal{X}_i)=V_j(\mathcal{X}_j)$). The cake-cutting problem has applications in many fields such as dividing land estates, advertisement space or broadcast time. An extension of the cake-cutting problem is the cake-cutting with two cakes problem (Cloutier et al., 2010) where two heterogeneous cakes are involved. In this problem, preferences of the agents can be coupled over the two cakes. The slice of one cake that an agent prefers might be influenced by the slice of the other cake that he or she might also obtain. The goal is to find a partition of the cakes that satisfies fairness conditions for the agents sharing the cakes. Cloutier et al. (2010) studied the envy-freeness partitioning. Both the cake-cutting and the cake-cutting with two cakes problems assume that there is only one indivisible unit of supply per element $x \in \mathcal{X}$ of the cake(s). Therefore sharing the cake(s) consists in obtaining a paritition of the set(s). In this paper, we show that EOT is a relaxation of the cutting cake and the cake-cutting with two cakes problems, when there is a divisible amount of each element of the cake(s). In that case, cakes are no more sets but distributions that we aim to divide between the agents according to their coupled preferences.

Integral Probability Metrics. In our work, we make links with some integral probability metrics. IPMs are (semi-)metrics on the space of probability measures. For a set of functions $\mathcal{F}$ and two probability distributions μ and ν, they are defined as

$$\text{IPM}_{\mathcal{F}}(\mu,\nu) = \sup_{f\in\mathcal{F}} \int f d\mu - \int f d\nu.$$

For instance, when $\mathcal{F}$ is chosen to be the set of bounded functions with uniform norm less or equal than 1, we recover the Total Variation distance (Steerneman, 1983) (TV). They recently regained interest in the Machine Learning community thanks to their application to Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) where IPMs are natural metrics for the discriminator (Dziugaite et al., 2015; Arjovsky et al., 2017; Mroueh and Sercu, 2017; Husain et al., 2019). They also helped to build consistent two-sample tests (Gretton et al., 2012; Scetbon and Varoquaux, 2019). However when a closed form of the IPM is not available, exact computation of IPMs between discrete distributions may not be possible or can be costful. For instance, the Dudley metric can be written as a Linear Program (Sriperumbudur et al., 2012) which has at least the same complexity as standard OT. Here, we show that the Dudley metric is in fact a particular case of our problem and obtain a faster approximation thanks to the entropic regularization.

3 Equitable and Optimal Transport

Notations. Let $\mathcal{Z}$ be a Polish space, we denote $\mathcal{M}(\mathcal{Z})$ the set of Radon measures on $\mathcal{Z}$. We call $\mathcal{M}_+(\mathcal{Z})$ the sets of positive Radon measures, and $\mathcal{M}_+^1(\mathcal{Z})$ the set of probability measures. We denote $\mathcal{C}^b(\mathcal{Z})$ the vector space of bounded continuous functions on $\mathcal{Z}$. Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. We denote for $\mu \in \mathcal{M}(\mathcal{X})$ and $\nu \in \mathcal{M}(\mathcal{Y})$, $\mu \otimes \nu$ the tensor product of the measures μ and ν, and $\mu \ll \nu$ means that ν dominates μ. We denote $\Pi_1 : (x,y) \in \mathcal{X}\times\mathcal{Y} \mapsto x$ and

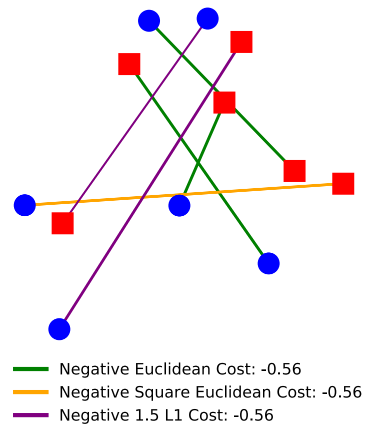

Figure 1: Equitable and optimal division of the resources between $N = 3$ different negative costs (i.e. utilities) given by EOT. Utilities have been normalized. Blue dots and red squares represent the different elements of resources available in each cake. We consider the case where there is exactly one unit of supply per element in the cakes, which means that we consider uniform distributions. Note that the partition between the agents is equitable (i.e. utilities are equal) and proportional (i.e. utilities are larger than $1/N$).

$\Pi_2 : (x,y) \in \mathcal{X}\times\mathcal{Y} \mapsto y$ respectively the projections on $\mathcal{X}$ and $\mathcal{Y}$, which are continuous applications. For an application g and a measure μ, we denote $g_\sharp\mu$ the pushforward measure of μ by g. For $\mathcal{X}$ and $\mathcal{Y}$ two Polish spaces, we denote $\text{LSC}(\mathcal{X}\times\mathcal{Y})$ the space of lower semi-continuous functions on $\mathcal{X}\times\mathcal{Y}$, $\text{LSC}^+(\mathcal{X}\times\mathcal{Y})$ the space of non-negative lower semi-continuous functions on $\mathcal{X}\times\mathcal{Y}$ and $\text{LSC}_*^-(\mathcal{X}\times\mathcal{Y})$ the set of negative bounded below lower semi-continuous functions on $\mathcal{X}\times\mathcal{Y}$. We also denote $\text{C}^+(\mathcal{X}\times\mathcal{Y})$ the space of non-negative continuous functions on $\mathcal{X}\times\mathcal{Y}$ and $\text{C}_*^-(\mathcal{X}\times\mathcal{Y})$ the set of negative continuous functions on $\mathcal{X}\times\mathcal{Y}$. Let $N \geq 1$ be an integer and denote $\Delta_N^+ := \{\lambda \in \mathbb{R}_+^N \text{ s.t. } \sum_{i=1}^N \lambda_i = 1\}$, the probability simplex of $\mathbb{R}^N$. For two positive measures of same mass $\mu \in \mathcal{M}_+(\mathcal{X})$ and $\nu \in \mathcal{M}_+(\mathcal{Y})$, we define the set of couplings with marginals μ and ν:

$$\Pi_{\mu,\nu} := \{\gamma \text{ s.t. } \Pi_{1\sharp}\gamma = \mu \ , \ \Pi_{2\sharp}\gamma = \nu\} \ .$$

We introduce the subset of $(\mathcal{M}_+^1(\mathcal{X})\times\mathcal{M}_+^1(\mathcal{Y}))^N$ representing marginal decomposition:

$$\Upsilon_{\mu,\nu}^N := \Big\{(\mu_i,\nu_i)_{i=1}^N \text{ s.t. } \sum_i \mu_i = \mu, \sum_i \nu_i = \nu \text{ and } \forall i,\ \mu_i(\mathcal{X}) = \nu_i(\mathcal{Y})\Big\}.$$

We also define the following subset of $\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})^N$ corresponding to the coupling decomposition:

$$\Gamma_{\mu,\nu}^N := \Big\{(\gamma_i)_{i=1}^N \text{ s.t. } \Pi_{1\sharp}\sum\gamma_i = \mu \ , \ \Pi_{2\sharp}\sum\gamma_i = \nu\Big\}.$$

3.1 Primal Formulation

Consider a fair division problem where several agents aim to share two sets of resources, $\mathcal{X}$ and $\mathcal{Y}$, and assume that there is a divisible amount of each resource $x \in \mathcal{X}$ (resp. $y \in \mathcal{Y}$) that is available. Formally, we consider the case where resources are no more sets but rather distributions on these sets. Denote μ and ν the distribution of resources on respectively $\mathcal{X}$ and $\mathcal{Y}$. For example, one might think about a situation where agents want to share fruit juices and ice creams and there is a certain volume of each type of fruit juices and a certain mass of each type of ice creams available. Moreover each agent defines his or her paired preferences for each couple $(x, y) \in \mathcal{X} \times \mathcal{Y}$. Formally, each person i is associated to an upper semi-continuous mapping $u_i : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^+$ corresponding to his or her preference for any given pair (x, y). For example, one may prefer to eat chocolate ice cream with apple juice, but may prefer pineapple juice when it comes with vanilla ice cream. The total utility for an individual i and a pairing $\gamma_i \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$ is then given by $V_i(\gamma_i) := \int u_i d\gamma_i$. To partition fairly among individuals, we maximize the minimum of individual utilities.

From a transport point of view, let assume that there are N workers available to transport a distribution μ to another one ν. The cost of a worker i to transport a unit mass from location x to the location y is $c_i(x, y)$. To partition the work among the N workers fairly, we minimize the maximum of individual costs.

These problems are in fact the same where the utility u_i, defined in the fair division problem, might be interpreted as the opposite of the cost c_i defined in the transportation problem, i.e. for all i, $c_i = -u_i$. The two above problem motivate the introduction of EOT defined as follows.

Definition 1 (Equitable and Optimal Transport)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. Let $\mathbf{c} := (c_i)_{1\leq i\leq N}$ be a family of bounded below lower semi-continuous cost functions on $\mathcal{X}\times\mathcal{Y}$, and $\mu \in \mathcal{M}_+^1(\mathcal{X})$ and $\nu \in \mathcal{M}_+^1(\mathcal{Y})$. We define the equitable and optimal transport primal problem:*

$$\mathrm{EOT}_{\mathbf{c}}(\mu, \nu) := \inf_{(\gamma_i)_{i=1}^N \in \Gamma_{\mu,\nu}^N} \max_i \int c_i d\gamma_i \ . \tag{1}$$

We prove along with Theorem 1 that the problem is well defined and the infimum is attained. Lower-semi continuity is a standard assumption in OT. In fact, it is the weakest condition to prove Kantorovich duality (Villani, 2003, Chap. 1). Note that the problem defined here is a linear optimization problem and when $N = 1$ we recover standard optimal transport. Figure 1 illustrates the equitable and optimal transport problem we consider. Figure 5 in Appendix D shows an illustration with respect to the transport viewpoint in the exact same setting, i.e. $c_i = -u_i$. As expected, the couplings obtained in the two situations are not the same.

We now show that in fact, EOT optimum satisfies equality constraints in case of constant sign costs, i.e. total utility/cost of each individual are equal in the optimal partition. See Appendix A.2 for the proof.

Proposition 1 (EOT solves the problem under equality constraints)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. Let $\mathbf{c} := (c_i)_{1\leq i\leq N} \in \mathrm{LSC}^+(\mathcal{X} \times \mathcal{Y})^N \cup \mathrm{LSC}_*^-(\mathcal{X} \times \mathcal{Y})^N$, $\mu \in \mathcal{M}_+^1(\mathcal{X})$ and $\nu \in \mathcal{M}_+^1(\mathcal{Y})$. Then the following are equivalent:*

- $(\gamma_i^*)_{i=1}^N \in \Gamma_{\mu,\nu}^N$ *is solution of Eq. (1),*
- $(\gamma_i^*)_{i=1}^N \in \underset{(\gamma_i)_{i=1}^N \in \Gamma_{\mu,\nu}^N}{\mathrm{argmin}} \left\{t \text{ s.t. } \forall i \ \int c_i d\gamma_i = t\right\}$.

Moreover,

$$\mathrm{EOT}_{\mathbf{c}}(\mu, \nu) = \min_{(\gamma_i)_{i=1}^N \in \Gamma_{\mu,\nu}^N} \left\{t \text{ s.t. } \forall i \ \int c_i d\gamma_i = t\right\} .$$

This property highly relies on the sign of the costs. For instance if two costs are considered, one always positive and the other always negative, then the constraints cannot be satisfied. When the cost functions are non-negatives, EOT refers to a transportation problem while when the costs are all negatives, costs become utilities and EOT refers to a fair division problem. The two points of view are concordant, but proofs and interpretations rely on the sign of the costs.

3.2 An Equitable and Proportional Division

When the cost functions considered c_i are all negatives, EOT become a fair division problem where the utility functions are defined as $u_i := -c_i$. Indeed according to Proposition 1, EOT solves

$$\max_{(\gamma_i)_{i=1}^N \in \Gamma_{\mu,\nu}^N} \left\{t \text{ s.t. } \forall i, \ \int u_i d\gamma_i = t\right\} .$$

Recall that in our model, the total utility of the agent i is given by $V_i(\gamma_i) := \int u_i d\gamma_i$. Therefore EOT aims to maximize the total utility of each agent i while ensuring that they are all equal. Let us now analyze which fairness conditions the partition induced by EOT verifies. Assume that the utilities are normalized, i.e., $\forall i$, there exists $\gamma_i \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y})$ such that $V_i(\gamma_i) = 1$. For example one might consider the cases where $\forall i$, $\gamma_i = \mu \otimes \nu$ or $\gamma_i \in \mathrm{argmin}_{\gamma\in\Pi_{\mu,\nu}} \int c_i d\gamma$. Then any solution $(\gamma_i^*)_{i=1}^N \in \Gamma_{\mu,\nu}^N$ of EOT satisfies:

- **Proportionality**: for all i, $V_i(\gamma_i^*) \geq 1/N$,

- **Equitablity**: for all i,j, $V_i(\gamma_i^*) = V_j(\gamma_j^*)$.

Proportionality is a standard fair division criterion for which a resource is divided among N agents, giving each agent at least $1/N$ of the heterogeneous resource by his/her own subjective valuation. Therefore here, this situation corresponds to the case where the normalized utility of each agent is at least $1/N$. Moreover, an equitable division is a division of an heterogeneous resource, in which each partner is equally happy with his/her share. Here this corresponds to the case where the utility of each agent are all equal.

The problem solved by EOT is a fair division problem where heterogeneous resources have to be shared among multiple agents according to their preferences. This problem is a relaxation of the two cake-cutting problem when there are a divisible amount of each item of the cakes. In that case, cakes are distributions and EOT makes a proportional and equitable partition of them. Details are left in Appendix A.2.

Fair Cake-cutting. Consider the case where the cake is an heterogeneous resource and there is a certain divisible quantity of each type of resource available. For example chocolate and vanilla are two types of resource present in the cake for which a certain mass is available. In that case, each type of resource in the cake is pondered by the actual quantity present in the cake. Up to a normalization, the cake is no more the set $\mathcal{X}$ but rather a distribution on this set. Note that for the two points of view to coincide, it suffices to assume that there is exactly the same amount of mass for each type of resources available in the cake. In that case, the cake can be represented by the uniform distribution over the set $\mathcal{X}$, or equivalently the set $\mathcal{X}$ itself. When cakes are distributions, the fair cutting cake problem can be interpreted as a particular case of EOT when the utilities of the agents do not depend on the variable $y \in \mathcal{Y}$. In short, we consider that utilities are functions of the form $u_i(x,y) = v_i(x)$ for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. The normalization of utilities can be cast as follows: $\forall i,\ V_i(\mu) = \int v_i(x)d\mu(x) = 1$. Then Proposition 1 shows that the partition of the cake made by EOT is proportional and equitable. Note that for EOT to coincide with the classical cake-cutting problem, one needs to consider that the uniform masses of the cake associated to each type of resource cannot be splitted. This can be interpreted as a Monge formulation (Villani, 2003) of EOT which is out of the scope of this paper.

3.3 Optimality of EOT

We next investigate the coupling obtained by solving EOT. In the next proposition, we show that under the same assumptions of Proposition 1, EOT solutions are optimal transportation plans. See Appendix A.3 for the proof.

Proposition 2 (EOT realizes optimal plans)**.** *Under the same conditions of Proposition 1, for any $(\gamma_i^*)_{i=1}^N \in \Gamma_{\mu,\nu}^N$ solution of Eq. (1), we have for all $i \in \{1,\dots,N\}$*

$$\gamma_i^* \in \operatorname*{argmin}_{\gamma\in\Pi_{\mu_i^*,\nu_i^*}} \int c_i d\gamma \qquad \text{where}\quad \mu_i^* := \Pi_{1\sharp}\gamma_i^*\ ,\ \nu_i^* := \Pi_{2\sharp}\gamma_i^*\ , \tag{2}$$

and

$$\begin{aligned}\mathrm{EOT}_{\mathbf{c}}(\mu,\nu) = &\min_{(\mu_i,\nu_i)_{i=1}^N\in\Upsilon_{\mu,\nu}^N} t\\ &\text{s.t. } \forall i\ \mathrm{W}_{c_i}(\mu_i,\nu_i) = t\ .\end{aligned} \tag{3}$$

Given the optimal matchings $(\gamma_i^*)_{i=1}^N \in \Gamma_{\mu,\nu}^N$, one can easily obtain the partition of the agents of each marginals. Indeed for all i, $\mu_i^* := \Pi_{1\sharp}\gamma_i^*$ and $\nu_i^* := \Pi_{2\sharp}\gamma_i^*$ represent respectively the portion of the agent i from distributions μ and ν.

Remark 1 (Utilitarian and Optimal Transport)**.** *To contrast with* EOT*, an alternative problem is to maximize the sum of the total utilities of agents, or equivalently minimize the sum of the total costs of agents. This problem can be cast as follows:*

$$\inf_{(\gamma_i)_{i=1}^N\in\Gamma_{\mu,\nu}^N} \sum_i \int c_i d\gamma_i \tag{4}$$

Here one aims to maximize the total utility of all the agents, while in EOT *we aim to maximize the total utility per agent under egalitarian constraint. The solution of (4) is not fair among agents and one can show that this problem is actually equal to* $\mathrm{W}_{\min_i(c_i)}(\mu,\nu)$*. Details can be found in Appendix C.1.*

3.4 Dual Formulation

Let us now introduce the dual formulation of the problem and show that strong duality holds under some mild assumptions. See Appendix A.4 for the proof.

Theorem 1 (Strong Duality)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. Let $\mathbf{c} := (c_i)_{i=1}^N$ be bounded below lower semi-continuous costs. Then* strong duality *holds, i.e. for $(\mu,\nu) \in \mathcal{M}_+^1(\mathcal{X}) \times \mathcal{M}_+^1(\mathcal{Y})$:*

$$\mathrm{EOT}_{\mathbf{c}}(\mu,\nu) = \sup_{\substack{\lambda\in\Delta_N^+\\ (f,g)\in\mathcal{F}_{\mathbf{c}}^\lambda}} \int f d\mu + \int g d\nu \tag{5}$$

where $\mathcal{F}_{\mathbf{c}}^\lambda := \{(f,g) \in \mathcal{C}^b(\mathcal{X}) \times \mathcal{C}^b(\mathcal{Y}) \text{ s.t. } \forall i \in \{1,...,N\},\ f \oplus g \leq \lambda_i c_i\}$.

This theorem holds under the same hypothesis and follows the same reasoning as the one in (Villani, 2003,

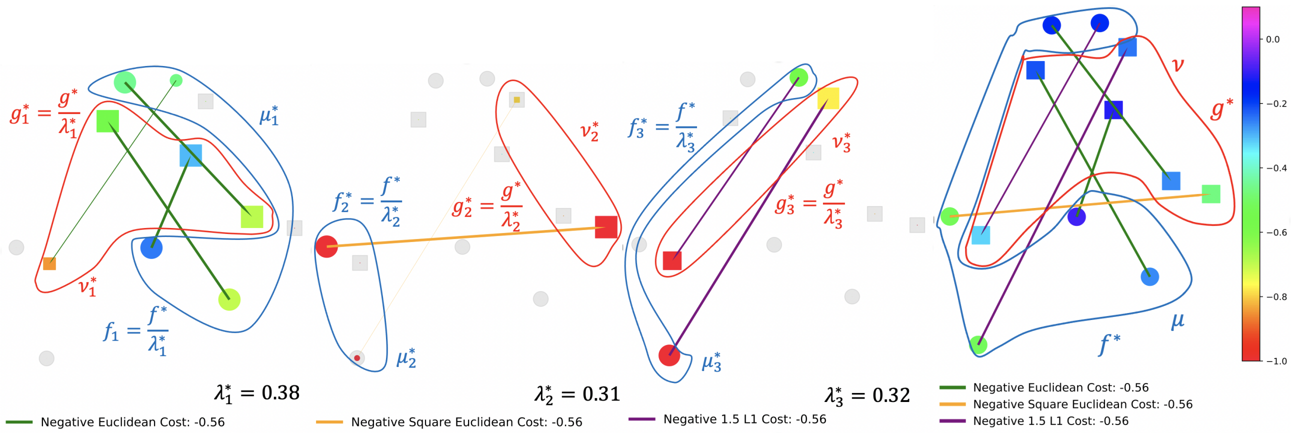

Figure 2: *Left, middle left, middle right*: the size of dots and squares is proportional to the weight of their representing atom in the distributions μ_k^* and ν_k^* respectively. The utilities f_k^* and g_k^* for each point in respectively μ_k^* and ν_k^* are represented by the color of dots and squares according to the color scale on the right hand side. The gray dots and squares correspond to the points that are ignored by agent k in the sense that there is no mass or almost no mass in distributions μ_k^* or ν_k^*. *Right*: the size of dots and squares are uniform since they correspond to the weights of uniform distributions μ and ν respectively. The values of f^* and g^* are given also by the color at each point. Note that each agent gets exactly the same total utility, corresponding exactly to EOT. This value can be computed using dual formulation (5) and for each figure it equals the sum of the values (encoded with colors) multiplied by the weight of each point (encoded with sizes).

Theorem 1.3). While the primal formulation of the problem is easy to understand, we want to analyse situations where the dual variables also play a role. For that purpose we show in the next proposition a simple characterisation of the primal-dual optimality in case of constant sign cost functions. See Appendix A.5 for the proof.

Proposition 3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be compact Polish spaces. Let $\mathbf{c} := (c_i)_{1\leq i\leq N} \in \mathrm{C}^+(\mathcal{X}\times\mathcal{Y})^N \cup \mathrm{C}^-_*(\mathcal{X}\times\mathcal{Y})^N$, $\mu\in\mathcal{M}^1_+(\mathcal{X})$ and $\nu\in\mathcal{M}^1_+(\mathcal{Y})$. Let also $(\gamma_k)_{k=1}^N \in \Gamma^N_{\mu,\nu}$ and $(\lambda, f, g) \in \Delta_n^+ \times \mathcal{C}^b(\mathcal{X})\times\mathcal{C}^b(\mathcal{Y})$. Then Eq. (5) admits a solution and the following are equivalent:*

- *$(\gamma_k)_{k=1}^N$ is a solution of Eq. (1) and (λ, f, g) is a solution of Eq. (5).*
- *1. $\forall i\in\{1,...,N\},\ f\oplus g\leq \lambda_i c_i$*
  2. $\forall i,j\in\{1,...,N\}\ \int c_i d\gamma_i = \int c_j d\gamma_j$
  3. $f\oplus g = \lambda_i c_i\quad \gamma_i$-a.e.

Remark 2. *It is worth noting that when we assume that $\mathbf{c} := (c_i)_{1\leq i\leq N} \in \mathrm{C}^+_*(\mathcal{X}\times\mathcal{Y})^N \cup \mathrm{C}^-_*(\mathcal{X}\times\mathcal{Y})^N$, then we can refine the second point of the equivalence presented in Proposition 3 by adding the following condition: $\forall i\in\{1,...,N\}\ \lambda_i\neq 0$.*

Given two distributions of resources represented by the measures μ and ν, and N utility functions denoted $(u_i)_{i=1}^N$, we want to find an *equitable* and *stable* partition among the agents in case of *transferable utilities*. Let k be an agent. We say that his or her utility is transferable when once $x\in\mathcal{X}$ and $y\in\mathcal{Y}$ get matched, he or she has to decide how to split his or her associated utility $u_k(x,y)$. She or he divides $u_k(x,y)$ into a quantity $f_k(x)$ which can be seen as the utility of having x and $g_k(y)$ for having y. Therefore in that problem we ask for $(\gamma_k, f_k, g_k)_{k=1}^N$ such that

$$u_k(x,y) = f_k(x) + g_k(y)\quad \gamma_k\text{-a.e.} \tag{6}$$

Moreover, for the partition to be *stable* (Sotomayor and Roth, 1990), we want to ensure that, for every agent k, none of the resources $x\in\mathcal{X}$ and $y\in\mathcal{Y}$ that have not been matched together for this agent would increase their utilities, $f_k(x)$ and $g_k(y)$, if there were matched together in the current matching instead. Formally we ask that for $k\in\{1,\dots,N\}$ and all $(x,y)\in\mathcal{X}\times\mathcal{Y}$,

$$f_k(x) + g_k(y) \geq u_k(x,y)\ . \tag{7}$$

Indeed if there exist k, x and y such that $u_k(x,y) > f_k(x)+g_k(y)$, then x and y will not be matched together in the share of the agent k and he can improve his utility for both x and y by matching x with y.

Finally we aim to share equitably the resources among the agents which boils down to ask

$$\forall i,j\in\{1,...,N\}\ \int u_i d\gamma_i = \int u_j d\gamma_j \tag{8}$$

Thanks to Proposition 3, finding $(\gamma_k, f_k, g_k)_{k=1}^N$ satisfying (6), (7) and (8) can be done by solving Eq. (1)

and Eq. (5). Indeed let $(\gamma_k)_{k=1}^N$ an optimal solution of Eq. (1) and (λ, f, g) an optimal solution of Eq. (5). Then by denoting for all $k = 1, \dots, N$, $f_k = \frac{f}{\lambda_k}$ and $g_k = \frac{g}{\lambda_k}$, we obtain that $(\gamma_k, f_k, g_k)_{k=1}^N$ solves the *equitable* and *stable* partition problem in case of *transferable utilities*. Note that again, we end up with equality constraints for the optimal dual variables. Indeed, for all $i, j \in \{1, \dots, N\}$, at optimality we have $\int f_i + g_i d\gamma_i = \int f_j + g_j d\gamma_j$. Figure 2 illustrates this formulation of the problem with dual potentials. Figure 7 in Appendix D shows the dual solutions with respect to the transport viewpoint in the exact same setting, i.e. $c_i = -u_i$. Once again, the obtained solutions differ.

3.5 Link with other Probability Metrics

In this section, we provide some topological properties on the object defined by the EOT problem. In particular, we make links with other known probability metrics, such as Dudley and Wasserstein metrics and give a tight upper bound.

When $N = 1$, recall from the definition (1) that the problem considered is exactly the standard OT problem. Moreover any EOT problem with $k \leq N$ costs can always be rewritten as a EOT problem with N costs. See Appendix C.2 for the proof. From this property, it is interesting to note that, for any $N \geq 1$, EOT generalizes standard Optimal Transport.

Optimal Transport. Given a cost function c, if we consider the problem EOT with N costs such that, for all i, $c_i = N \times c$ then, the problem $\mathrm{EOT}_{\mathbf{c}}$ is exactly W_c. See Appendix C.2 for the proof.

Now we have seen that all standard OT problems are sub-cases of the EOT problem, one may ask whether EOT can recover other families of metrics different from standard OT. Indeed we show that the EOT problem recovers an important family of IPMs with supremum taken over the space of α-Hölder functions with $\alpha \in (0, 1]$. See Appendix A.6 for the proof.

Proposition 4. *Let $\mathcal{X}$ be a Polish space. Let d be a metric on $\mathcal{X}^2$ and $\alpha \in (0, 1]$. Denote $c_1 = 2 \times \mathbf{1}_{x \neq y}$, $c_2 = d^\alpha$ and $\mathbf{c} := (c_1, (N-1) \times c_2, ..., (N-1) \times c_2) \in \mathrm{LSC}(\mathcal{X} \times \mathcal{X})^N$ then for any $(\mu, \nu) \in \mathcal{M}_+^1(\mathcal{X}) \times \mathcal{M}_+^1(\mathcal{X})$*

$$\mathrm{EOT}_{\mathbf{c}}(\mu, \nu) = \sup_{f \in B_{d^\alpha}(\mathcal{X})} \int_{\mathcal{X}} f d\mu - \int_{\mathcal{X}} f d\nu \quad (9)$$

where $B_{d^\alpha}(\mathcal{X}) := \{f \in C^b(\mathcal{X})\colon \|f\|_\infty + \|f\|_\alpha \leq 1\}$ and $\|f\|_\alpha := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d^\alpha(x,y)}$.

Dudley Metric. When $\alpha = 1$, then for $(\mu, \nu) \in \mathcal{M}_+^1(\mathcal{X}) \times \mathcal{M}_+^1(\mathcal{X})$, we have

$$\mathrm{EOT}_{\mathbf{c}}(\mu, \nu) = \mathrm{EOT}_{(c_1, d)}(\mu, \nu) = \beta_d(\mu, \nu)$$

where β_d is the *Dudley Metric* (Dudley et al., 1966). In other words, the Dudley metric can be interpreted as an equitable and optimal transport between the measures with the trivial cost and a metric d. We acknowledge that Chizat et al. (2018) made a link between Unbalanced Optimal Transport and the "flat metric", an IPM close to the Dudley metric, defined on the space $\{f\colon \|f\|_\infty \leq 1,\ \|f\|_1 \leq 1\}$.

Weak Convergence. When d is an unbounded metric on $\mathcal{X}$, it is well known that W_{d^p} with $p \in (0, +\infty)$ metrizes a convergence a bit stronger than weak convergence (Villani, 2003, Chap. 7). A sufficient condition for Wasserstein distances to metrize weak convergence on the space of distributions is that the metric d is bounded. In contrast, metrics defined by Eq. (9) do not require such assumptions and $\mathrm{EOT}_{(\mathbf{1}_{x \neq y}, d^\alpha)}$ metrizes the weak convergence of probability measures (Villani, 2003, Chap. 1-7).

For an arbitrary choice of costs $(c_i)_{1 \leq i \leq N}$, we obtain a tight upper control of EOT and show how it is related to the OT problem associated to each cost involved. See Appendix A.7 for the proof.

Proposition 5. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces. Let $\mathbf{c} := (c_i)_{1 \leq i \leq N}$ be a family of nonnegative lower semi-continuous costs. For any $(\mu, \nu) \in \mathcal{M}_+^1(\mathcal{X}) \times \mathcal{M}_+^1(\mathcal{Y})$*

$$\mathrm{EOT}_{\mathbf{c}}(\mu, \nu) \leq \left(\sum_{i=1}^N \frac{1}{\mathrm{W}_{c_i}(\mu, \nu)} \right)^{-1} \quad (10)$$

Proposition 5 means that the minimal cost to transport all goods under the constraint that all workers contribute equally is lower than the case where agents share equitably and optimally the transport with distributions μ_i and ν_i respectively proportional to μ and ν, which equals the harmonic sum written in Equation (10).

Example. *Applying the above result in the case of the Dudley metric recovers the following inequality (Sriperumbudur et al., 2012, Proposition 5.1)*

$$\beta_d(\mu, \nu) \leq \frac{\mathrm{TV}(\mu, \nu) \mathrm{W}_d(\mu, \nu)}{\mathrm{TV}(\mu, \nu) + \mathrm{W}_d(\mu, \nu)}.$$

4 Entropic Relaxation

In their original form, as proposed by Kantorovich Kantorovich (1942), Optimal Transport distances are not a natural fit for applied problems: they minimize a network flow problem, with a supercubic complexity $(n^3 \log n)$ (Tarjan, 1997). Following the work of Cuturi (2013), we propose an entropic relaxation of EOT, obtain its dual formulation and derive an efficient algorithm to compute an approximation of EOT.

4.1 Primal-Dual Formulation

Let us first extend the notion of Kullback-Leibler divergence for positive Radon measures. Let $\mathcal{Z}$ be a Polish space, for $\mu,\nu\in\mathcal{M}_+(\mathcal{Z})$, we define the generalized Kullback-Leibler divergence as $\text{KL}(\mu||\nu)=\int\log\frac{d\mu}{d\nu}d\mu+\int d\nu-\int d\mu$ if $\mu\ll\nu$, and $+\infty$ otherwise. We introduce the following regularized version of EOT.

Definition 2 (Entropic relaxed primal problem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces, $\mathbf{c}:=(c_i)_{1\leq i\leq N}$ a family of bounded below lower semi-continuous costs lower semi-continuous costs on $\mathcal{X}\times\mathcal{Y}$ and $\boldsymbol{\varepsilon}:=(\varepsilon_i)_{1\leq i\leq N}$ be non negative real numbers. For $(\mu,\nu)\in\mathcal{M}_+^1(\mathcal{X})\times\mathcal{M}_+^1(\mathcal{Y})$, we define the* EOT regularized primal problem*:*

$$\text{EOT}_{\mathbf{c}}^{\boldsymbol{\varepsilon}}(\mu,\nu):=\inf_{\gamma\in\Gamma_{\mu,\nu}^N}\max_i\int c_i d\gamma_i + \sum_{j=1}^N\varepsilon_j\text{KL}(\gamma_j||\mu\otimes\nu)$$

Note that here we sum the generalized Kullback-Leibler divergences since our objective is function of N measures in $\mathcal{M}_+(\mathcal{X}\times\mathcal{Y})$. This problem can be compared with the one from standard regularized OT. In the case where $N=1$, we recover the standard regularized OT. For $N\geq 1$, the underlying problem is $\sum_{i=1}^N\varepsilon_i-$strongly convex. Moreover, we prove the essential property that as $\boldsymbol{\varepsilon}\to 0$, the regularized problem converges to the standard problem. See Appendix C.3 for the full statement and the proof. As a consequence, entropic regularization is a consistent approximation of the original problem we introduced in Section 3.1. Next theorem shows that strong duality holds for lower semi-continuous costs and compact spaces. This is the basis of the algorithm we will propose in Section 4.2. See Appendix A.8 for the proof.

Theorem 2 (Duality for the regularized problem)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two compact Polish spaces, $\mathbf{c}:=(c_i)_{1\leq i\leq N}$ a family of bounded below lower semi-continuous costs on $\mathcal{X}\times\mathcal{Y}$ and $\boldsymbol{\varepsilon}:=(\varepsilon_i)_{1\leq i\leq N}$ be non negative numbers. For $(\mu,\nu)\in\mathcal{M}_+^1(\mathcal{X})\times\mathcal{M}_+^1(\mathcal{Y})$, strong duality holds:*

$$\text{EOT}_{\mathbf{c}}^{\boldsymbol{\varepsilon}}(\mu,\nu)=\sup_{\lambda\in\Delta_N^+}\sup_{\substack{f\in\mathcal{C}_b(\mathcal{X})\\ g\in\mathcal{C}_b(\mathcal{Y})}}\int f d\mu+\int g d\nu - \sum_{i=1}^N\varepsilon_i\left(\int e^{\frac{f(x)+g(y)-\lambda_i c_i(x,y)}{\varepsilon_i}}d\mu(x)d\nu(y)-1\right) \tag{11}$$

and the infimum of the primal problem is attained.

As in standard regularized optimal transport there is a link between primal and dual variables at optimum. Let γ^* solving the reguralized primal problem and (f^*,g^*,λ^*) solving the dual one:

$$\forall i,\ \gamma_i^*=\exp\left(\frac{f^*+g^*-\lambda_i^*c_i}{\varepsilon_i}\right)\cdot\mu\otimes\nu$$

.

4.2 Proposed Algorithms

Algorithm 1 Projected Alternating Maximization

Input: $\mathbf{C}=(C_i)_{1\leq i\leq N}$, a, b, ε, L_λ
Init: $f^0\leftarrow\mathbf{1}_n$; $g^0\leftarrow\mathbf{1}_m$; $\lambda^0\leftarrow(1/N,...,1/N)\in\mathbb{R}^N$
for $k=1,2,...$ **do**
  $K^k\leftarrow\sum_{i=1}^N K_i^{\lambda_i^{k-1}}$,
  $c_k\leftarrow\langle f^{k-1},K^k g^{k-1}\rangle$, $f^k\leftarrow\frac{c_k a}{K^k g^{k-1}}$,
  $d_k\leftarrow\langle f^k,K^k g^{k-1}\rangle$, $g^k\leftarrow\frac{d_k b}{(K^k)^T f^k}$,
  $\lambda^k\leftarrow\text{Proj}_{\Delta_N^+}\left(\lambda^{k-1}+\frac{1}{L_\lambda}\nabla_\lambda F_{\mathbf{C}}^{\varepsilon}(\lambda^{k-1},f^k,g^k)\right)$.
end
Result: λ,f,g

We can now present algorithms obtained from entropic relaxation to approximately compute the solution of EOT. Let $\mu=\sum_{i=1}^n a_i\delta_{x_i}$ and $\nu=\sum_{j=1}^m b_j\delta_{y_j}$ be discrete probability measures where $a\in\Delta_n^+$, $b\in\Delta_m^+$, $\{x_1,...,x_n\}\subset\mathcal{X}$ and $\{y_1,...,y_m\}\subset\mathcal{Y}$. Moreover for all $i\in\{1,\dots,N\}$ and $\lambda>0$, define $\mathbf{C}:=(C_i)_{1\leq i\leq N}\in(\mathbb{R}^{n\times m})^N$ with $C_i:=(c_i(x_k,y_l))_{k,l}$ the N cost matrices and $K_i^\lambda:=\exp(-\lambda C_i/\varepsilon)$. Assume that $\varepsilon_1=\dots=\varepsilon_N=\varepsilon$. Compared to the standard regularized OT, the main difference here is that the problem contains an additional variable $\lambda\in\Delta_N^+$. When $N=1$, one can use Sinkhorn algorithm. However when $N\geq 2$, we do not have a closed form for updating λ when the other variables of the problem are fixed. In order to enjoy from the strong convexity of the primal formulation, we consider instead the dual associated with the equivalent primal problem given when the additional trivial constraint $\mathbf{1}_n^T\left(\sum_i P_i\right)\mathbf{1}_m=1$ is considered. In that the dual obtained is

$$\widehat{\text{EOT}}_{\mathbf{C}}^{\boldsymbol{\varepsilon}}(a,b)=\sup_{\substack{\lambda\in\Delta_N^+\\ f\in\mathbb{R}^n,\ g\in\mathbb{R}^m}}\langle f,a\rangle+\langle g,b\rangle - \varepsilon\left[\log\left(\sum_i\langle e^{\mathbf{f}/\varepsilon},K_i^{\lambda_i}e^{\mathbf{g}/\varepsilon}\rangle\right)+1\right]$$

We show that the new objective obtained above is smooth w.r.t (λ,f,g). See Appendix C.4 for the proof. One can apply the accelerated projected gradient ascent (Beck and Teboulle, 2009; Tseng, 2008) which enjoys an optimal convergence rate for first order methods of $\mathcal{O}(k^{-2})$ for k iterations.

It is also possible to adapt Sinkhorn algorithm to our problem. See Algorithm 1. We denoted by $\text{Proj}_{\Delta_N^+}$ the orthogonal projection on Δ_N^+ (Shalev-Shwartz and Singer, 2006), whose complexity is in $\mathcal{O}(N\log N)$. The smoothness constant in λ in the algorithm is $L_\lambda=\max_i\|C_i\|_\infty^2/\varepsilon$. In practice Alg. 1 gives better results than the accelerated gradient descent. Note

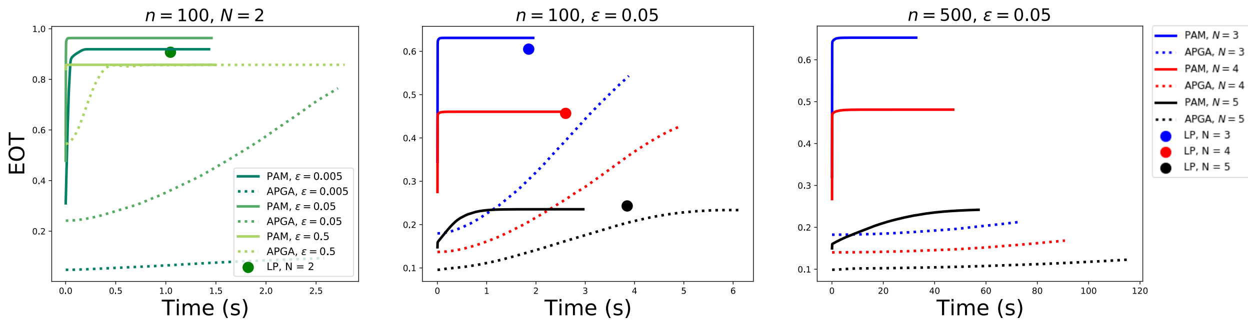

Figure 3: Comparison of the time-accuracy tradeoffs between the different proposed algorithms. *Left:* we consider the case where the number of days is $N = 2$, the size of support for both measures is $n = m = 100$ and we vary ε from 0.005 to 0.5. *Middle:* we fix $n = m = 100$ and the regularization $\varepsilon = 0.05$ and we vary the number of days N from 3 to 5. *Right:* the setting considered is the same as in the figure in the middle, however we increase the sample size such that $n = m = 500$. Note that in that case, **LP** is too costly to be computed.

that the proposed algorithm differs from the Sinkhorn algorithm in many points and therefore the convergence rates cannot be applied here. Analyzing the rates of a *projected* alternating maximization method is, to the best of our knowledge, an unsolved problem. Further work will be devoted to study the convergence of this algorithm. We illustrate Algorithm 1 by showing the convergence of the regularized version of EOT towards the ground truth when $\varepsilon \to 0$ in the case of the Dudley Metric. See Figure 8 in Appendix D.

5 Other applications of EOT

Minimal Transportation Time. Assume there are N internet service providers who propose different debits to transport data across locations, and one needs to transfer data from multiple servers to others, the fastest as possible. We assume that $c_i(x, y) \geq 0$ corresponds to the transportation time needed by provider i to transport one unit of data from a server x to a server y. For instance, the unit of data can be one Megabit. Then $\int c_i d\gamma_i$ corresponds the time taken by provider i to transport $\mu_i = \Pi_{1\sharp}\gamma_i$ to $\nu_i = \Pi_{2\sharp}\gamma_i$. Assuming the transportation can be made in parallel and given a partition of the transportation task $(\gamma_i)_{i=1}^N$, $\max_i \int c_i d\gamma_i$ corresponds to the total time of transport the data $\mu = \Pi_{1\sharp} \sum \gamma_i$ to the locations $\nu = \Pi_{2\sharp} \sum \gamma_i$ according to this partition. Then EOT, which minimizes $\max_i \int c_i d\gamma_i$, is finding the fastest way to transport the data from μ to ν by splitting the task among the N internet service providers. Note that at optimality, all the internet service providers finish their transportation task at the same time (see Proposition 1).

Sequential Optimal Transport. Consider the situation where an agent aims to transport goods from some stocks to some stores in the next N days. The cost to transport one unit of good from a stock located at x to a store located at y may vary across the days. For example the cost of transportation may depend on the price of gas, or the daily weather conditions. Assuming that he or she has a good knowledge of the daily costs of the N coming days, he or she may want a transportation strategy such that his or her daily cost is as low as possible. By denoting c_i the cost of transportation the i-th day, and given a strategy $(\gamma_i)_i^N$, the maximum daily cost is then $\max_i \int c_i d\gamma_i$, and EOT therefore finds the cheapest strategy to spread the transport task in the next N days such that the maximum daily cost is minimized. Note that at optimality he or she has to spend the exact same amount everyday.

In Figure 3 we aim to simulate the Sequential OT problem and compare the time-accuracy trade-offs of the proposed algorithms. Let us consider a situation where one wants to transport merchandises from $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ to $\nu = \frac{1}{m}\sum_{j=1}^m \delta_{y_j}$ in N days. Here we model the locations $\{x_i\}$ and $\{y_j\}$ by drawing them independently from two Gaussian distributions in $\mathbb{R}^2$: $\forall i,\ x_i \sim \mathcal{N}\left(\binom{3}{3}, \left(\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right)\right)$ and $\forall j,\ y_j \sim \mathcal{N}\left(\binom{4}{4}, \left(\begin{smallmatrix}1 & -.2\\ -.2 & 1\end{smallmatrix}\right)\right)$. We assume that everyday there is wind modeled by a vector $w \sim \mathcal{U}(B(0,1))$ where $B(0,1)$ is the unit ball in $\mathbb{R}^2$ that is perfectly known in advance. We define the cost of transportation on day i as $c_i(x, y) = \|y - x\| - 0.7\langle w_i, y - x\rangle$ to model the effect of the wind on the transportation cost. In the following figures we plot the estimates of EOT obtained from the proposed algorithms in function of the runtime for various sample sizes n, number of days N and regularizations ε. **PAM** denotes Alg. 1, **APGA** denotes Alg. 2 (See Appendix C.4), **LP** denotes the linear program which solves exactly the primal formulation of the EOT problem. Note that when **LP** is computable (i.e. $n \leq 100$), it is therefore the ground truth. We show that in all the settings, **PAM** performs better than **APGA** and provides very high accuracy with order of magnitude faster than LP.

Acknowledgments

The authors would like to thank V. Do and Y. Chevaleyre for fruitful discussions. We gratefully acknowledge support from "Chaire d'excellence de l'IDEX Paris Saclay".

References

David Alvarez-Melis, Stefanie Jegelka, and Tommi S Jaakkola. Towards optimal transport with global invariances. *arXiv preprint 1806.09277*, 2018.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1): 183–202, 2009.

Dimitris Bertsimas and John N Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.

Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D Procaccia. *Handbook of computational social choice*. Cambridge University Press, 2016.

Haim Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media, 2010.

Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.

John Cloutier, Kathryn L Nyman, and Francis Edward Su. Two-player envy-free multi-cake division. *Mathematical Social Sciences*, 59(1):26–37, 2010.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

Lester E Dubins and Edwin H Spanier. How to cut a cake fairly. *The American Mathematical Monthly*, 68 (1P1):1–17, 1961.

Richard M. Dudley. The speed of mean Glivenko-Cantelli convergence. *Annals of Mathematical Statistics*, 40(1):40–50, 1969.

Richard Mansfield Dudley et al. Weak convergence of probabilities on nonseparable metric spaces and empirical measures on euclidean spaces. *Illinois Journal of Mathematics*, 10(1):109–126, 1966.

Paul Dupuis and Richard S Ellis. *A weak convergence approach to the theory of large deviations*, volume 902. John Wiley & Sons, 2011.

Arnaud Dupuy, Alfred Galichon, and Yifei Sun. Estimating matching affinity matrix under low-rank constraints. *arXiv preprint arXiv:1612.09585*, 2016.

Gintare Karolina Dziugaite, Daniel M Roy, and Zoubin Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. *arXiv preprint arXiv:1505.03906*, 2015.

Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-Ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. *arXiv preprint arXiv:1810.08278*, 2018.

Nicolas Fournier and Arnaud Guillin. On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162 (3-4):707–738, 2015.

Aude Genevay, Gabriel Peyré, and Marco Cuturi. Learning generative models with sinkhorn divergences, 2017.

Aude Genevay, Lénaic Chizat, Francis Bach, Marco Cuturi, and Gabriel Peyré. Sample complexity of sinkhorn divergences. *arXiv preprint arXiv:1810.02733*, 2018.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(Mar):723–773, 2012.

Hisham Husain, Richard Nock, and Robert C Williamson. A primal-dual link between gans and autoencoders. In *Advances in Neural Information Processing Systems*, pages 413–422, 2019.

Hicham Janati, Thomas Bazeille, Bertrand Thirion, Marco Cuturi, and Alexandre Gramfort. Multi-subject meg/eeg source imaging with sparse multi-task regression. *NeuroImage*, page 116847, 2020.

Leonid Kantorovich. On the transfer of masses (in russian). *Doklady Akademii Nauk*, 37(2):227–229, 1942.

Tor Lattimore, Koby Crammer, and Csaba Szepesvári. Linear multi-resource allocation with semi-bandit feedback. In *Advances in Neural Information Processing Systems*, pages 964–972, 2015.

Ruilin Li, Xiaojing Ye, Haomin Zhou, and Hongyuan Zha. Learning to match via inverse optimal transport. *J. Mach. Learn. Res.*, 20:80–1, 2019.

Erika Mackin and Lirong Xia. Allocating indivisible items in categorized domains. *arXiv preprint arXiv:1504.05932*, 2015.

Hervé Moulin. *Fair division and collective welfare*. MIT press, 2004.

Youssef Mroueh and Tom Sercu. Fisher gan. In *Advances in Neural Information Processing Systems*, pages 2513–2523, 2017.

Alfred Müller. Integral probability metrics and their generating classes of functions. *Advances in Applied Probability*, 29(2):429–443, 1997.

Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1):127–152, 2005.

François-Pierre Paty and Marco Cuturi. Subspace robust wasserstein distances. *arXiv preprint arXiv:1901.08949*, 2019.

Mathis Petrovich, Chao Liang, Yanbin Liu, Yao-Hung Hubert Tsai, Linchao Zhu, Yi Yang, Ruslan Salakhutdinov, and Makoto Yamada. Feature robust optimal transport for high-dimensional data. *arXiv preprint arXiv:2005.12123*, 2020.

Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving GANs using optimal transport. In *International Conference on Learning Representations*, 2018. URL https://openreview.net/forum?id=rkQkBnJAb.

M. Scetbon and G. Varoquaux. Comparing distributions: ℓ_1 geometry improves kernel two-sample testing, 2019.

Meyer Scetbon and Marco Cuturi. Linear time sinkhorn divergences using positive features, 2020.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Shai Shalev-Shwartz and Yoram Singer. Efficient learning of label ranking by soft projections onto polyhedra. *Journal of Machine Learning Research*, 7(Jul): 1567–1599, 2006.

Maurice Sion. On general minimax theorems. *Pacific J. Math.*, 8(1):171–176, 1958. URL https://projecteuclid.org:443/euclid.pjm/1103040253.

Marilda Sotomayor and Alvin Roth. Two-sided matching: A study in game-theoretic modelling and analysis. *Econometric Society Monographs*, (18), 1990.

Bharath K Sriperumbudur, Kenji Fukumizu, Arthur Gretton, Bernhard Schölkopf, Gert RG Lanckriet, et al. On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6: 1550–1599, 2012.

Ton Steerneman. On the total variation and hellinger distance between signed measures; an application to product measures. *Proceedings of the American Mathematical Society*, 88(4):684–688, 1983.

H. Steinhaus. Sur la division pragmatique. *Econometrica*, 17:315–319, 1949. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907319.

Haodong Sun, Haomin Zhou, Hongyuan Zha, and Xiaojing Ye. Learning cost functions for optimal transport. *arXiv preprint arXiv:2002.09650*, 2020.

Robert E. Tarjan. Dynamic trees as search trees via euler tours, applied to the network simplex algorithm. *Mathematical Programming*, 78(2):169–177, 1997.

Paul Tseng. On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*, 1, 2008.

Cédric Villani. *Topics in optimal transportation*. Number 58. American Mathematical Soc., 2003.

Haibin Wang, Sujoy Sikdar, Xiaoxi Guo, Lirong Xia, Yongzhi Cao, and Hanpin Wang. Multi-type resource allocation with partial preferences. *arXiv preprint arXiv:1906.06836*, 2019.

Weiran Wang and Miguel A. Carreira-Perpinan. Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application, 2013.

Eric Wong, Frank R. Schmidt, and J. Zico Kolter. Wasserstein adversarial examples via projected sinkhorn iterations, 2019.

Karren Dai Yang, Karthik Damodaran, Saradha Venkatachalapathy, Ali C Soylemezoglu, GV Shivashankar, and Caroline Uhler. Predicting cell lineages using autoencoders and optimal transport. *PLoS computational biology*, 16(4):e1007828, 2020.