Matching Equilibria with Weak Optimal Transport: Theory, Algorithms, and Experiments

SUBMISSION 320

We numerically study the matching between workers and firms in the setting of [Choné and Kramarz, 2022] where firms choose their size and aggregate their employees' skills to produce output. To allow for skill aggregation within firms and for endogenous choice of size, we rely on weak forms of optimal transport developed in [Gozlan et al., 2017] and [Choné et al., 2022], where the transport cost is allowed to be nonlinear in the transport plan and mass is transported from the worker's space to the firms' space through normalized or unnormalized kernels.

In this paper, we develop mirror descent algorithms to solve the primal and dual versions of weak optimal transport problems with normalized and unnormalized kernels. The main numerical challenge lies in the transportation constraints that are costly to project onto. We derive an upper bound of the error which we use to monitor the convergence of the algorithms.

We run experiments to check the consistency of the algorithms in cases where the transport plan or the value of the optimum are characterized as closed-form solutions or through differential equation. In the case where workers have multidimensional skills, we check theoretical predictions about how the wages and the sorting of workers into firms vary as the proportion of specialist workers in the economy increases.

Manuscript submitted for review to the 24nd ACM Conference on Economics & Computation (EC'23).

1 INTRODUCTION

This paper provides numerical methods to compute optimal matching between workers and firms in a setting where workers differ in their skills and firms (i) endogenously choose their size; (ii) aggregate their employees' skills to produce output; and (iii) differ in their exogenous production technology, i.e., in the way they transform their employees' aggregate skills into output. The firms' ability to aggregate the skills of their employees is what makes labor markets different from most markets where buyers generally cannot aggregate the characteristics of consumption goods.[1]

Considering the two distributions that represent firms' heterogeneity in technology and workers' heterogeneity in skills, [Choné and Kramarz, 2022] express this many-to-one matching problem in the framework of optimal transport (OT) theory. In the primal version of the problem, the objective is to maximize total output in the economy. Yet the output produced by a firm obtains by applying its production function to the total skills of its employees. Because production functions are generally nonlinear, the produced output is not the sum of the output that would be produced by each individual employee separately. The objective of the primal problem, therefore, cannot be expressed as the integral of a variable (a "transport cost") that would depend on each individual firm and each individual worker.

For this fundamental reason, the ability of firms to aggregate their employees' skills requires to extend the standard OT framework in two directions. First, it requires to allow the transport cost between a firm with technology x and workers with skills y to depend non linearly on the measure $\pi^x(dy)$ of the employees' skills. This idea has been formalized by [Alibert et al., 2019, Gozlan et al., 2017] and given rise to the notion of weak optimal transport, hereafter abbreviated as WOT. In the WOT framework, the measure or "kernel" $\pi^x(dy)$ is constrained to be a probability measure. Second, in our matching problem, the total mass of the $\pi^x(dy)$ represents the number of workers employed by a firm with technology x. At a competitive equilibrium, this number may vary across firms. In particular, more productive firms tend to recruit more employees. To allow the sizes of firms to be endogenously determined in equilibrium, [Choné et al., 2022] generalized the WOT framework and introduced the notion of weak optimal transport with unnormalized kernel, henceforth WOTUK, where the kernels $\pi^x(dy)$ are positive measures that can have any nonnegative mass.

In this paper, we present algorithms to solve the WOT and WOTUK problems numerically, when the considered measures are discrete. If the production function is concave, weak optimal transport (WOT) defines a convex optimization problem over the transportation polytope [Paty and Cuturi, 2020]. Nevertheless, algorithms to compute WOT when the measures are discrete have only been proposed in the special case of quadratic barycentric WOT [Cazelles et al., 2021] (see subsection 2.2 for a precise definition). In their recent preprint, [Korotin et al., 2022] propose to use neural networks to approximate the WOT problem, but do not provide guarantees for their optimization procedure.

The paper is organized as follows. In Sections 2 and 3, we recall theoretical results for the primal and dual versions of the WOT and WOTUK problems while providing economic interpretations in our matching context. We present numerical algorithms in Section 4 and provide an upper bound for the approximation of the primal objective. In Section 5, we show the convergence of the algorithm and compute the numerical guarantee on an economic

[1] Two cars do not provide the same utility as a single, more powerful one.

example. In Section 6, we confront the algorithm to various theoretical predictions. Section 7 concludes and present avenues for future work.

2 FROM OPTIMAL TRANSPORT TO WEAK OPTIMAL TRANSPORT

2.1 Optimal Transport and Matching

A classic problem in labor economics [Eeckhout and Kircher, 2018, Heckman and Scheinkman, 1987, Kelso and Crawford, 1982] is to understand the matching between workers and firms, i.e. to explain why workers work in their employing firms, and conversely, why firms hire some employees and not others. Optimal transport has been used in the economics literature [Galichon, 2016, Lindenlaub, 2017] to model workers-to-firms matching.

Firms differ in technologies and workers differ in skills. Let $\mathcal{X} \subset \mathbb{R}^p$, $p \geq 1$, denote the set of firms' types (or technologies) and likewise, let $\mathcal{Y} \subset \mathbb{R}^q$, $q \geq 1$, denote the set of workers' types in the economy. Given a probability distribution of firm types $\mu \in \mathscr{P}(\mathcal{X})$ and a probability distribution of worker types $\nu \in \mathscr{P}(\mathcal{Y})$, a coupling $\pi \in \mathscr{P}(\mathcal{X} \times \mathcal{Y})$ of μ and ν represents the matching between firms and workers, in the sense that $\pi(A \times B)$, for $A \subset \mathcal{X}, B \subset \mathcal{Y}$ Borel sets, is the proportion of firms whose type is in A that employ a worker whose type is in B. The primal problem is to maximize the total output in the economy:

$$\mathrm{OT}(\mu,\nu) \overset{\text{def}}{=} \sup_{\pi \in \Pi(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{Y}} F(x,y)\, pi(dx,dy) \tag{1}$$

where $\Pi(\mu,\nu)$ is the set of all couplings between μ and ν, i.e.

$$\Pi(\mu,\nu) = \Big\{\pi \in \mathscr{P}(\mathcal{X}\times\mathcal{Y}),\, \forall A \subset \mathcal{X}, B \subset \mathcal{Y} \text{ Borel}, \pi(A\times\mathcal{Y}) = \mu(A), \pi(\mathcal{X}\times B) = \nu(B)\Big\},$$

and $F : \mathcal{X}\times\mathcal{Y} \longrightarrow \mathbb{R}$ is the production function, i.e. $F(x,y)$ is the output (in \$) produced by a worker of type $y \in \mathcal{Y}$ working in a firm of type $x \in \mathcal{X}$.

Problem (1) corresponds to the definition of the Kantorovich [1942] problem in optimal transport with cost function $-F$. It admits the following dual formulation:

$$\mathrm{OT}(\mu,\nu) = \inf_{\substack{\chi\in C(\mathcal{X}),\varphi\in C(\mathcal{Y})\\ \chi\oplus\varphi\geq F}} \int \chi\, d\mu + \int \varphi\, d\nu \tag{2}$$

where for $C(\mathcal{X})$ (resp. $C(\mathcal{Y})$) is the set of real continuous functions over $\mathcal{X}$ (resp. over $\mathcal{Y}$), and $\chi \oplus \varphi \in C(\mathcal{X}\times\mathcal{Y})$ is the function $\chi\oplus\varphi : (x,y) \mapsto \chi(x) + \varphi(y)$. In the labor market context, $\chi(x)$ and $\varphi(y)$ represent respectively the profit of firms with type x and the wage of workers with type y.

2.2 Weak Optimal Transport

The Kantorovich problem (1) can be rewritten as:

$$\sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} \left[\int_{\mathcal{Y}} F(x,y)\, \pi^x(dy)\right] \mu(dx) \tag{3}$$

where $(\pi^x)_{x\in\mathcal{X}} \subset \mathscr{P}(\mathcal{Y})$ is the (μ-almost surely unique) probability kernel that allows to disintegrate π with respect to μ as $\pi(dx,dy) = \mu(dx)\pi^x(dy)$. In other words, π^x is the law of $Y|X = x$ when $(X,Y) \sim \pi$.

Gozlan et al. [2017] introduce the weak optimal transport problem as

$$\mathrm{WOT}(\mu,\nu) \overset{\text{def}}{=} \sup_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} \mathcal{F}(x,\pi^x)\, \mu(dx) \tag{4}$$

where $\mathcal{F} : \mathcal{X} \times \mathscr{P}(\mathcal{Y}) \to \mathbb{R}$, i.e. $\mathcal{F}(x, p)$ now denotes the production (in \$) of a firm of type $x \in \mathcal{X}$ hiring employees with distribution $p \in \mathscr{P}(\mathcal{Y})$. The classic Kantorovich problem is the special case of WOT where $\mathcal{F}(x, p) = \int_{\mathcal{Y}} F(x, y)\, p(dy)$.

In our economic context, the probability kernel π^x represents the distribution of types among the workers hired by firms with type x. The economic interpretation of the reworded problem (3) is that the output produced by a firm of type $x \in \mathcal{X}$ is the sum of the output produced by its employees, $\int F(x, y)\, \pi^x(dy)$. In particular, the production of a firm of type $x \in \mathcal{X}$ depends linearly on the distribution of its employees' types, $\pi^x \in \mathscr{P}(\mathcal{Y})$. Choné and Kramarz [2022] relax this restriction and allow firms to aggregate the skills of their employees in more general way. The production of a firm of type $x \in \mathcal{X}$ depends non-linearly on the distribution $\pi^x \in \mathscr{P}(\mathcal{Y})$ of its employees' types.

Barycentric WOT problem. We will say that the WOT problem (4) is barycentric when $\mathcal{F}(x, p)$ only depends on the barycenter of p, that is when $\mathcal{F}(x, p) = F\left(x, \int y\, p(dy)\right)$ for some function $F : \mathcal{X} \times \mathrm{conv}(\mathcal{Y}) \to \mathbb{R}$, where $\mathrm{conv}(\mathcal{Y})$ is the convex hull of $\mathcal{Y}$.[2] In the economic context, the barycentric specification is valid if the production of a firm depends on the distribution of its employees' types, $p \in \mathscr{P}(\mathcal{Y})$, only through their aggregate skills, $\int y\, p(dy)$.

2.3 Duality

Just like the Kantorovich problem (1) admits the dual formulation (2), the WOT problem (4) also admits a dual formulation under some assumptions on $\mathcal{F}$. For the WOT problem to be convex, and hence hope for strong duality to hold, we require that $p \mapsto \mathcal{F}(x, p)$ is convex for all $x \in \mathcal{X}$. We refer to [Gozlan et al., 2017, Section 9] for the technical assumptions and details.

Under these assumptions, the WOT problem (4) admits the following dual formulation [Gozlan et al., 2017, Theorem 9.5]:

$$\mathrm{WOT}(\mu, \nu) = \inf_{\varphi \in C(\mathcal{Y})} \int_{\mathcal{X}} R_{\mathcal{F}}(\varphi)\, d\mu + \int_{\mathcal{Y}} \varphi\, d\nu \tag{5}$$

where $R_{\mathcal{F}}(\varphi)(x) = \sup_{p \in \mathscr{P}(\mathcal{Y})} \mathcal{F}(x, p) - \int \varphi\, dp$.

This dual formulation can in turn be interpreted in our economic framework: $\varphi(y)$ represents the wage of the worker with type $y \in \mathcal{Y}$. Given a wage function φ, a firm of type $x \in \mathcal{X}$ hires workers according to a probability distribution $p \in \mathscr{P}(\mathcal{Y})$ chosen to maximize profit defined as output $\mathcal{F}(x, p)$ minus wage bill $\int \varphi(y)\, p(dy)$. $R_{\mathcal{F}}(\varphi)(x)$ is therefore the maximum profit the firm type $x \in \mathcal{X}$ can attain given the wage function φ, so that $\int_{\mathcal{X}} R_{\mathcal{F}}(\varphi)\, d\mu$ is the total profit in the economy. The wages are then chosen so as to minimize the sum of the profits and of the wages.

When the cost function $\mathcal{F}$ is barycentric, i.e. when $\mathcal{F}(x, p) = F\left(x, \int y\, p(dy)\right)$ for some $F : \mathcal{X} \times \mathrm{conv}(\mathcal{Y}) \to \mathbb{R}$ [Gozlan et al., 2017, Proof of Theorem 2.11] prove another dual formulation:

$$\mathrm{WOT}(\mu, \nu) = \inf_{\substack{\psi \in C(\mathrm{conv}(\mathcal{Y})) \\ \text{convex, Lipschitz}}} \int_{\mathcal{X}} Q_F(\psi)\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu \tag{6}$$

[2] The particular case where $F(x, y) = \|x - y\|^2$ is called the quadratic barycentric WOT problem.

where $Q_F(\psi)(x) = \sup_{y \in \mathrm{conv}(\mathcal{Y})} F(x,y) - \psi(y)$. [Gozlan et al., 2017, Proof of Theorem 2.11] gives a way to construct a minimizer $\psi_\star$ of the dual (6) from a minimizer $\varphi_\star$ of the more general dual problem (5), by simply taking for $\psi_\star$ the largest convex function that is smaller than $\varphi_\star$, i.e.:

$$\psi_\star : z \mapsto \inf_{\substack{p \in \mathscr{P}(\mathcal{Y}) \\ \int y\, p(dy) = z}} \int_{\mathcal{Y}} \varphi_\star \, dp. \tag{7}$$

The convexity of dual minimizers in the barycentric case is easy to interpret in our economic setting. Here the output produced by a firm depends only on the aggregate skill of its employees. If the wage function is $\varphi_\star$, $\psi_\star(z)$ given by (7) represents the lowest wage bill that a firm must spend to achieve the aggregate skill $z = \int y\, p(dy)$. The convexity of the wage thus directly results from the firms' ability to aggregate the skills of their employees.

3 WEAK OPTIMAL TRANSPORT WITH UNNORMALIZED KERNEL (WOTUK)

3.1 From WOT to WOTUK

Choné et al. [2022] relax the assumption that π^x in (3) and (4) is a probability measure. They allow π^x to be a positive measure. Denoting by $\mathscr{M}(\mathcal{Y})$ the set of positive measures over $\mathcal{Y}$, they introduce the weak optimal transport problem with unnormalized kernel (WOTUK) as

$$\mathrm{WOTUK}(\mu, \nu) \stackrel{\mathrm{def}}{=} \sup_{\substack{q \in \mathscr{M}(\mathcal{Y})^{\mathcal{X}} \\ \int q^x \, \mu(dx) = \nu}} \int_{\mathcal{X}} \mathcal{F}(x, q^x) \, \mu(dx) \tag{8}$$

where $\mathcal{F} : \mathcal{X} \times \mathscr{M}(\mathcal{Y}) \to \mathbb{R}$. The constraint $\int q^x \, \mu(dx) = \nu$ expresses that the unnormalized kernel q transports μ onto ν. Choné et al. [2022] connect the WOTUK problem (8) to a WOT problem as follows. Letting

$$\Pi(\ll \mu, \nu) \stackrel{\mathrm{def}}{=} \{\pi \in \Pi(\eta, \nu),\ \eta \in \mathscr{P}(\mathcal{X}), \eta \ll \mu\},$$

denote the set of probability measure over $\mathcal{X}$ that are absolutely continuous with respect to μ, they show that

$$\mathrm{WOTUK}(\mu, \nu) = \sup_{\substack{\eta \in \mathscr{P}(\mathcal{X}) \\ \eta \ll \mu}} \sup_{\pi \in \Pi(\eta, \nu)} \int \mathcal{F}\left(x, \frac{d\eta}{d\mu}(x) \pi^x\right) \mu(dx) \tag{9}$$

$$= \sup_{\pi \in \Pi(\ll \mu, \nu)} \int \mathcal{F}\left(x, \frac{d\pi_1}{d\mu}(x) \pi^x\right) \mu(dx) \tag{10}$$

where $\pi^x \in \mathscr{P}(\mathcal{Y})$ is the unique disintegration of π with respect to η, i.e. such that $\pi(x,y) = \eta(dx)\pi^x(dy)$, and π_1 is the first marginal of π. At given η, we thus go back to the WOT problem studied in Section 2. Instead of constraining the first marginal of π to be μ, the WOTUK problem only imposes that the first marginal is absolutely continuous with respect to μ. Choné et al. [2022] show that the density of η with respect to μ is nothing else than the mass of q^x, i.e., $\frac{d\eta}{d\mu}(x) = \frac{d\pi_1}{d\mu}(x) = q^x(\mathcal{Y})$.

In the economic setting of [Choné and Kramarz, 2022], $q^x(\mathcal{Y})$ represents the number of employees (i.e., the size) of firms with type x. Allowing q^x to be an unnormalized positive measure instead of a probability measure avoids having to impose that all firms have the same size. In contrast to earlier literature, firms' sizes are unknowns to be (optimally) determined rather than given parameters.

Conical WOTUK problem. The conical WOTUK problem corresponds to the case where

$$\mathcal{F}(x,q) = F\left(x, \int_{\mathcal{Y}} y\, q(dy)\right)$$

for some $F : \mathcal{X} \times \mathrm{cone}(\mathcal{Y}) \to \mathbb{R}$, where the conical hull of $\mathcal{Y}$ is given by

$$\mathrm{cone}(\mathcal{Y}) \overset{\text{def}}{=} \left\{ \sum_{i=1}^{n} \lambda_i y_i\,,\ \lambda_1, \ldots, \lambda_n \in \mathbb{R}_+, y_1, \ldots, y_n \in \mathcal{Y}, n \geq 1 \right\}.$$

In [Choné and Kramarz, 2022], a firm's output depends on the conical combination of its employees' types, $\int y\, q^x(dy)$. The combination is said to be "conical" because the mass of q^x is not necessarily equal to one. In other words, the aggregate skill of the workers hired by a firm is not their average skills as in the WOT setting, but their average skills scaled by the positive factor $q^x(\mathcal{Y})$ that represents the number of employees.

3.2 Duality

The WOTUK problem (8) admits dual formulations that are similar to those of the dual WOT (5) and (6). The main difference with the results of subsection 2.3 lies in replacing $\mathscr{P}(\mathcal{Y})$ by $\mathscr{M}(\mathcal{Y})$ and $\mathrm{conv}(\mathcal{Y})$ by $\mathrm{cone}(\mathcal{Y})$.

Under some technical assumptions on $\mathcal{F}$, detailed in [Choné et al., 2022], the theorem 3.2 in the same reference proves that the WOTUK problem (8) admits the following dual formulation:

$$\mathrm{WOTUK}(\mu, \nu) = \inf_{\varphi \in C_b(\mathcal{Y})} \int_{\mathcal{X}} K_{\mathcal{F}}(\varphi)\, d\mu + \int_{\mathcal{Y}} \varphi\, d\nu \tag{11}$$

where $K_{\mathcal{F}}(\varphi)(x) = \sup_{m \in \mathscr{M}(\mathcal{Y})} \mathcal{F}(x,m) - \int \varphi\, dm$.

Similarly, [Choné et al., 2022, Theorem 5.1] proves that the conical WOTUK problem admits the dual formulation:

$$\mathrm{WOTUK}(\mu, \nu) = \inf_{\substack{\psi \in C(\mathrm{cone}(\mathcal{Y}))\ \text{convex,} \\ \text{positively homogeneous}}} \int_{\mathcal{X}} J_F(\psi)\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu \tag{12}$$

where $J_F(\psi)(x) = \sup_{y \in \mathrm{cone}(\mathcal{Y})} F(x,y) - \psi(y)$. They show that a minimizer $\psi_\star$ of the dual problem (12) is derived from a minimizer $\varphi_\star$ of the more general dual problem (11) by taking for $\psi_\star$ the largest convex and positively homogeneous function that is smaller than $\varphi_\star$, i.e.:

$$\psi_\star : z \mapsto \inf_{\substack{m \in \mathscr{M}(\mathcal{Y}) \\ \int y\, dm(y) = z}} \int_{\mathcal{Y}} \varphi_\star\, dm. \tag{13}$$

In the economic setting of [Choné and Kramarz, 2022], a dual optimizer φ is a wage function: $\varphi(y)$ represents the wage paid to a worker of type $y \in \mathcal{Y}$. As $R_{\mathcal{F}}(\varphi)(x)$ and $Q_F(\psi)(x)$ in subsection 2.3, $K_{\mathcal{F}}(\varphi)(x)$ and $J_F(\psi)(x)$ are two forms for the profit function, i.e., for the maximal profit that firms of each type $x \in \mathcal{X}$ achieve under the wage functions φ or ψ.

4 ALGORITHMS

In this section, we only consider the case of discrete measures $\mu \in \mathscr{P}(\mathcal{X})$ and $\nu \in \mathscr{P}(\mathcal{Y})$. We will write $\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$ where $n \geq 1$ is the number of firm types, $x_1, \ldots, x_n \in \mathcal{X}$ are the firm types and $a \in \mathbb{R}^n$ represents the proportion of the firm types in the economy ($a > 0$ and $\sum_{i=1}^{n} a_i = 1$). Likewise, we will write $\nu = \sum_{j=1}^{m} b_j \delta_{y_j}$ where $m \geq 1$ is the number of worker types, $y_1, \dots, y_m \in \mathcal{Y}$ are the worker types and b represents the proportions of the worker types in the population ($b > 0$ and $\sum_{j=1}^m b_j = 1$). Since we consider here maximization problems, we will require that the cost function $\mathcal{F}$ is concave and differentiable with respect to its second argument.

4.1 The primal problems

A matching $\pi \in \Pi(\mu, \nu)$ is now represented by a matrix $P \in \mathbb{R}^{n\times m}$ such that P_{ij} represents the proportion of the firm type x_i which is matched with the worker type y_j. For the WOT problem (4), the marginal constraints on P translate into:

$$\Pi(\mu, \nu) = \{P \in \mathbb{R}_+^{n\times m},\, P\mathbf{1} = a, P^\top \mathbf{1} = b\}.$$

For the WOTUK problem (9), the set of constraints is simply

$$\Pi(\ll \mu, \nu) = \{P \in \mathbb{R}_+^{n\times m},\, P^\top \mathbf{1} = b\}$$

because as explained in Section 3.1 the first marginal $P\mathbf{1} = \eta$ is unconstrained and η is an unknown variable to be determined. The difference between the WOT problem (4) and the WOTUK problem (9) lies in the constraint set only. The WOTUK problem corresponds to the WOT problem where the first marginal constraint has been removed, which establishes an interesting link with the unbalanced optimal transport theory [Chizat et al., 2018].

Let us now write the objective for the WOT and WOTUK problems in the discrete setting. The disintegration π_{x_i} representing the workers hired by the firm of type x_i writes $\frac{1}{a_i} \sum_{j=1}^m P_{ij}\delta_{y_j}$. For simplicity, we make the following change of notations: we define $\widetilde{\mathcal{F}} : \mathcal{X}\times\mathbb{R}^m$ by

$$\widetilde{\mathcal{F}} : (x, p) \mapsto \mathcal{F}\left(x, \sum_{j=1}^m p_j \delta_{y_j}\right).$$

Note that $\widetilde{\mathcal{F}}$ depends on $y_1, \dots, y_m$ and that $\frac{\partial \widetilde{\mathcal{F}}}{\partial p_j}(x,p) = \left\langle \delta_{y_j}, \nabla_2 \mathcal{F}\left(x, \sum_{k=1}^m p_k \delta_{y_k}\right)\right\rangle$ where $\langle \cdot, \cdot \rangle : \mathcal{M}(\mathcal{X}) \times \mathcal{M}(\mathcal{X})^*$ is the duality bracket. We will use the notation $P_{i:} = (P_{ij})_{1\le j\le m}$ for $1 \le i \le n$.

With these notations, the objective of the WOT problem (4) and of the WOTUK problem (9) writes:

$$f(P) \overset{\text{def}}{=} \sum_{i=1}^n a_i \mathcal{F}\left(x_i, \frac{1}{a_i}\sum_{j=1}^m P_{ij}\delta_{y_j}\right) = \sum_{i=1}^n a_i \widetilde{\mathcal{F}}\left(x_i, \frac{P_{i:}}{a_i}\right).$$

Since the constraints sets $\Pi(\mu, \nu)$ and $\Pi(\ll \mu, \nu)$ are convex and $\mathcal{F}$ and $\widetilde{\mathcal{F}}$ are convex in their second argument, both the discrete WOT and WOTUK problems are convex (although not strictly) optimization problems. In order to solve them, we propose to apply a mirror ascent on P (using the Kullback-Leibler divergence). The gradient of the total output writes:

$$\frac{\partial f}{\partial P_{ij}}(P) = \left[\nabla_2 \widetilde{\mathcal{F}}\left(x_i, \frac{P_{i:}}{a_i}\right)\right]_j.$$

After each gradient step, the resulting matching P should be projected (for the KL divergence) onto $\Pi(\mu, \nu)$ (for the WOT problem) or $\Pi(\ll \mu, \nu)$ (for the WOTUK problem). For the WOT problem, we have to solve $\min_{Q\in\Pi(\mu,\nu)} \mathrm{KL}(Q|P)$. This problem is equivalent to the entropic OT problem (20) with cost function $-\log P$ and regularization strength $\varepsilon = 1$, and can therefore be efficiently solved using the Sinkhorn algorithm [Cuturi, 2013]. For the WOTUK problem, we have to solve $\min_{Q\in\Pi(\ll\mu,\nu)} \mathrm{KL}(Q|P)$ which admits the following closed-form solution (see a proof in Appendix B.1): $Q_\star = P \odot b / P^\top \mathbf{1}$ where $\odot$ and $/$ are the elementwise multiplication and division.

Numerical guarantee. Along the mirror ascent iterations over P, we can monitor the convergence by looking at the gap

$$\mathcal{G}(P) \overset{\text{def}}{=} \sup_{Q \in K} f(Q) - f(P)$$

where $K = \Pi(\mu, \nu)$ for the WOT problem and $K = \Pi(\ll \mu, \nu)$ for the WOTUK problem. By definition, $\mathcal{G}(P) \geq 0$ and by the concavity of f, $\mathcal{G}(P) \leq \overline{\mathcal{G}}(P)$, where

$$\overline{\mathcal{G}}(P) = \sup_{Q \in K} \langle \nabla f(P), Q - P \rangle. \tag{14}$$

In the WOT case, $K = \Pi(\mu, \nu)$, the upper bound $\overline{\mathcal{G}}(P)$ on $\mathcal{G}(P)$ corresponds to an optimal transport problem (with cost matrix $-\nabla f(P)$) and can either be computed exactly or be itself upper bounded using the solution $Q_\star$ of an entropic OT problem (efficiently solved using the Sinkhorn algorithm).

In the WOTUK case, $K = \Pi(\ll \mu, \nu)$, the upper bound $\overline{\mathcal{G}}(P)$ on $\mathcal{G}(P)$ admits the following closed form solution: $Q_{\star ij} = b_j$ if $i = \arg\max_{1 \leq k \leq n} [\nabla f(P)]_{kj}$ and $Q_{\star ij} = 0$ otherwise.

The algorithm stops when $\overline{\mathcal{G}}(P) \leq \varepsilon f(P)$ for a tolerance $\varepsilon > 0$. We summarize the mirror ascent method for the WOT and WOTUK problems in Algorithm 1.

Algorithm 1 Mirror Ascent Algorithm for WOT and WOTUK (primal)

```
Input Stepsize $\gamma > 0$, tolerance $\varepsilon$
Initialize $P = ab^\top$
while $\overline{\mathcal{G}}(P) > \varepsilon f(P)$ do
  $P \leftarrow P \exp(\gamma \nabla f(P))$
  For WOT:
    $P \leftarrow \text{Sinkhorn}(a, b, \text{kernel} = P)$
  For WOTUK:
    $P \leftarrow P \odot b / P^\top \mathbf{1}$
end while
Return $P$
```

4.2 The dual problems

In the discrete setting, the general dual for WOT (5) writes:

$$\min_{\varphi \in \mathbb{R}^m_+} \langle b, \varphi \rangle + \max_{\substack{P \in \mathbb{R}^{n \times m}_+ \\ P\mathbf{1} = a}} \sum_{i=1}^{n} a_i \widetilde{\mathcal{F}}\left(x_i, \frac{P_{i:}}{a_i}\right) - \mathbf{1}^\top P \varphi \tag{15}$$

and likewise for the general dual for the WOTUK problem (11):

$$\min_{\varphi \in \mathbb{R}^m_+} \langle b, \varphi \rangle + \max_{P \in \mathbb{R}^{n \times m}_+} \sum_{i=1}^{n} a_i \widetilde{\mathcal{F}}\left(x_i, \frac{P_{i:}}{a_i}\right) - \mathbf{1}^\top P \varphi. \tag{16}$$

Let us define

$$h : (\varphi, P) \mapsto \sum_{i=1}^{n} a_i \widetilde{\mathcal{F}}\left(x_i, \frac{P_{i:}}{a_i}\right) - \mathbf{1}^\top P \varphi.$$

To solve problems (15) and (16), we propose to run a mirror descent on φ (with the Kullback-Leibler divergence). The objective is itself a maximization problem (over P). The envelope theorem yields the gradient of the objective, provided the optimal P is given. At each gradient step on φ, we therefore propose to run a mirror ascent on P at fixed φ. For the WOTUK problem (16), no projection are needed, while for the WOT problem (15), we need to project P onto $\{P \in \mathbb{R}_+^{n\times m}, P\mathbf{1} = a\}$ during the ascents. The proof in Appendix B.1 directly adapts to this case, and the projection amounts to reweighting the rows of P.

We can construct dual minimizers for the barycentric WOT (6) and conical WOTUK (12) from a solution $\varphi_\star$ of (15) and (16) respectively, using the results given in (7) and (13) respectively. In the discrete setting, these linear programs respectively write:

$$\psi_\star : z \mapsto \min_{\substack{p\in\mathbb{R}_+^m \\ \sum_{j=1}^m p_j=1 \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi_\star\rangle$$

and for WOTUK

$$\psi_\star : z \mapsto \min_{\substack{p\in\mathbb{R}_+^m \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi_\star\rangle.$$

Since we may be interested in differentiating those functions $\psi_\star$, we rather compute the dual problems of the above linear programs (see a proof in the Appendix B.2):

$$\psi_\star : z \mapsto \max_{\substack{\lambda\in\mathbb{R}^q,\mu\in\mathbb{R} \\ \forall j, \langle\lambda, y_j\rangle+\mu\le\varphi_{\star j}}} \langle \lambda, z\rangle + \mu \quad \text{and} \quad \psi_\star : z \mapsto \max_{\substack{\lambda\in\mathbb{R}^q \\ \forall j, \langle\lambda, y_j\rangle\le\varphi_{\star j}}} \langle \lambda, z\rangle.$$

We summarize the mirror ascent method for the WOT and WOTUK problems in Algorithm 2.

Algorithm 2 Mirror Descent Algorithm for WOT and WOTUK (dual)

```
Input $\gamma_1 > 0$, $\gamma_2 > 0$, $K_1, K_2 \in \mathbb{N}$
Initialize $\varphi \in \mathbb{R}_+^m$ and $P = ab^\top$
for $k_1 = 0$ to $K_1$ do
  for $k_2 = 0$ to $K_2$ do
    $P \leftarrow P \exp(\gamma_1 \nabla_P h(\varphi, P))$
    For WOT:
      $P \leftarrow \mathrm{diag}(a/P\mathbf{1})\, P$
  end for
  $\varphi \leftarrow \varphi \exp(-\gamma_2 [b - P^\top \mathbf{1}])$
end for
Use a linear programming solver to compute $\psi(z)$ for $z \in \mathcal{Y}$:
For WOT:
      $\psi(z) = \max_{\lambda\in\mathbb{R}^q,\mu\in\mathbb{R};\ \forall j, \langle\lambda, y_j\rangle+\mu\le\varphi_j} \langle \lambda, z\rangle + \mu.$
For WOTUK:
      $\psi(z) = \max_{\lambda\in\mathbb{R}^q;\ \forall j, \langle\lambda, y_j\rangle\le\varphi_j} \langle \lambda, z\rangle.$
Return Dual variables $\varphi_j, \psi(y_j)$ for $1 \le j \le m$.
```

5 CONVERGENCE OF THE ALGORITHM

All the experiments have been run on a Google Colab Notebook using JAX. (A Notebook is provided as supplementary material.) We used the following packages: OTT [Cuturi et al., 2022] for entropic optimal transport and POT [Flamary et al., 2021] for exact optimal transport.

Workers have two-dimensional skills and firms' technologies differ in two dimensions, i.e., $p = q = 2$. Firms' production function are assumed to be Constant Elasticity of Substitution:

$$F(x, y) = \frac{z}{\eta} \left[(1-\alpha)\, y_1^{\sigma} + \alpha\, y_2^{\sigma}\right]^{\eta/\sigma},$$

where $x = (z, \alpha)$ is a firm's type and $y = (y_1, y_2)$ is a worker' type. In the simulations, we take $\eta = 1/2$ and $\sigma = -1$.

The set of firms' types is $\mathcal{X} = \{\, (z, \alpha) \in [0, \infty] \times [0, 1] \,\}$, where z represents the productivity of the firm type and α its technical intensity in skill 2, i.e., how important that skill i is in its production function.

The set of worker types is $\mathcal{Y} = \{(y_1, y_2) \in \mathbb{R}_+^2\}$, where y_i, $i \in \{1, 2\}$, represents the proficiency of the worker type in skill i. A worker's global quality can be represented by the Euclidian norm of the skill vector (y_1, y_2). A worker's skill profile, defined as his comparative advantage in skill 2 over skill 1, is given by $\theta = \arctan(y_2/y_1) \in [0, \pi/2]$, where $\theta = 0$ represents an expert worker in skill 1, $\theta = \pi/2$ an expert worker in skill 2, and $\theta = \pi/4$ a generalist worker.

Figures 1, 5 and 7 provide examples of discrete distributions μ and ν over $\mathcal{X}$ and $\mathcal{Y}$.

5.1 One-dimensional heterogeneity

We first assume that workers (y_1, y_2) have the same overall quality and differ only in their skill profiles θ, i.e., they all have the same "quality". Similarly, we assume that firms all have the total factor productivity z, i.e., they differ only in their technical intensities in each skills. In Scenario A represented on Figurefig:one:dim:few:specialist, consider firms of same z and with α uniformly distributed between 0 and 1 and workers distributed according to a Beta distribution on the positive quarter of the unit circle. The discrete distributions involves $n = m = 200$ different mass points.

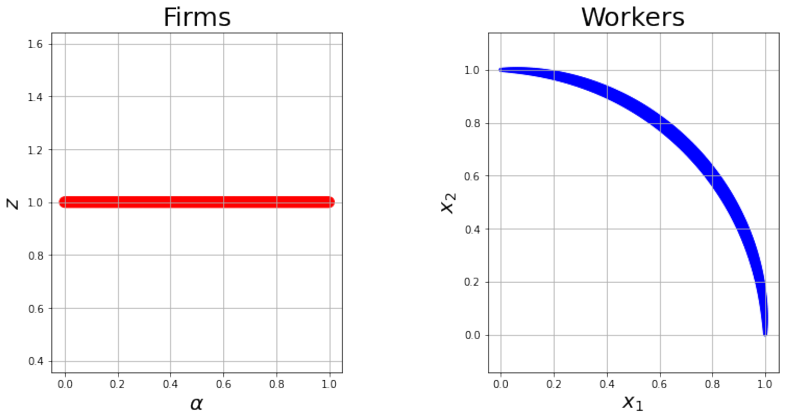


Fig. 1. Scenario A: Firms have same TFP and workers have same quality (each dot represents a Dirac mass and its size represents its weight)

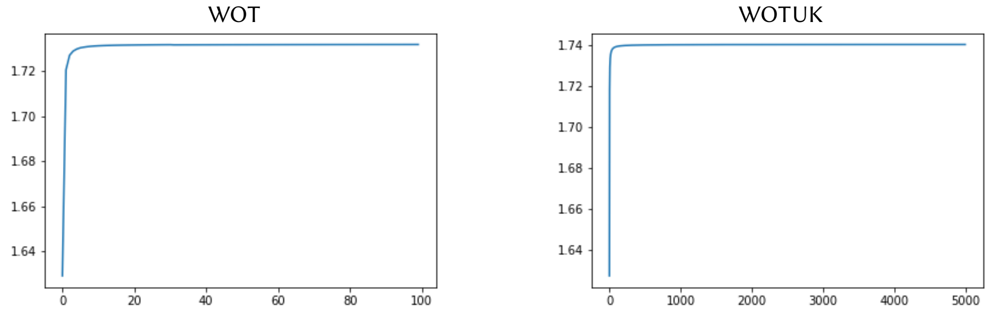


Fig. 2. Total output as a function of the number of iterations for WOT and WOTUK with $n = m = 200$ firms and workers. (Distributions μ and ν are represented on Figure 1)

Empirically, the algorithm converges in a few iterations, as Figure 2 shows. We can accelerate convergence by playing with step sizes. In Figure 3, the variable step size γ_t is proportional to $1/\sqrt{t}$.

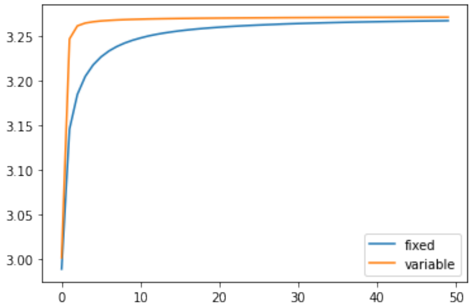


Fig. 3. Total output as a function of the number of iterations for WOT with $n = m = 200$ firms and workers (Distributions μ and ν are represented on Figure 1)

Numerical guarantee. We have found that the quantity $\overline{\mathcal{G}}(P)$ given by (14) is an upper bound of the error in the algorithm. As explained in Sectino 4.1, the computation of $\overline{\mathcal{G}}$ is straightforward in the WOTUK case. For the WOT case, we use the Sinkhorn algorithm with $\varepsilon = 10^{-4}$ to approximate from above $\overline{\mathcal{G}}$ without high computation times. Determining theoretically the rate at which $\overline{\mathcal{G}}$ converges to zero seems a difficult task. In practice, Figure 4 shows convergence in a few iterations.

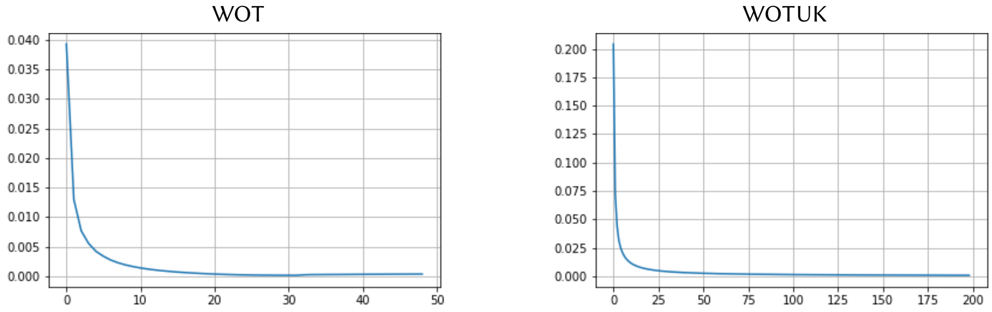


Fig. 4. Numerical guarantee $\overline{\mathcal{G}}$ as a function of the number of iterations ran by our algorithm for WOT and WOTUK with 200 firms and workers (Distributions μ and ν are represented on Figure 1)

5.2 Two-dimensional heterogeneity

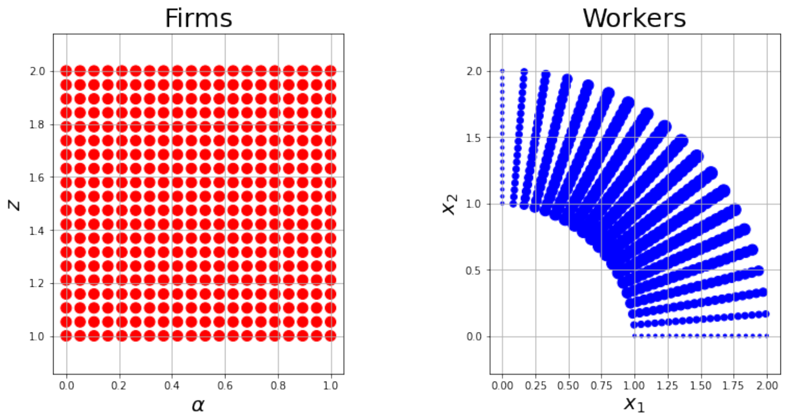


Fig. 5. Scenario C. Each dot represents a Dirac mass and its size represents its weight ($n = m = 400$ workers' and firms' types).

We now assume that workers differ in both quality and skill profiles and firms differ in both TFP and technical intensities. We keep the same marginals for the α and θ as in Scenario A, and take firms' TFP z and workers' qualities $\sqrt{y_1^2 + y_2^2}$ as uniformly distributed on $[1, 2]$. In all the discrete versions of the problem that we consider, we take the number of workers' and firms' types as equal, $n = m$, see Figure 5 with $n = m = 400$. Figure 6 displays the values of the objective as a function of the number of firms and workers. We run the algorithm for 200 iterations for WOT and 100 iterations for WOTUK. We check that the WOTUK objective is higher than the WOT, consistently with the former problem being less constrained than the latter.

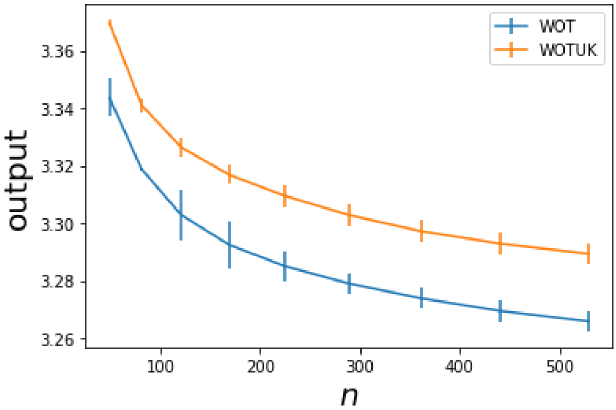


Fig. 6. Total output for WOT and WOTUK as a function of the number of firm's types (equal to the number of workers' types). "Error bars" based on upper bound $\overline{\mathcal{G}}$ (Distributions μ and ν are represented on Figure 5)

6 OUR ALGORITHM AGAINST THEORETICAL PREDICTIONS

In this section, we check the consistency of our algorithm by comparing simulations against theoretical results. The specifications considered in Subsections 6.1, 6.2 and 6.3 are conical in the sense that a firm's output depends only on the sum of its employees' skills. Subsection 6.4 considers a one-dimensional non-conical problem.

6.1 How does the proportion of specialist workers in the economy affect the matching equilibrium?

[Choné and Kramarz, 2022] have shown that as the proportion of specialist workers in the economy increases the wage schedule becomes flatter and specialized firms tend to specialize further, a phenomenon they call "polarization". To illustrate this phenomenon, we consider the two following situations:

- Scenario A: Specialist workers are relative rare in the economy and hence will earn high wages at a competitive equilibrium (the distributions of firms and workers' types are represented on Figure 1);
- Scenario B: There are many specialist workers in the economy. The firms' technologies are the same as in Scenario A. See Figure 7.

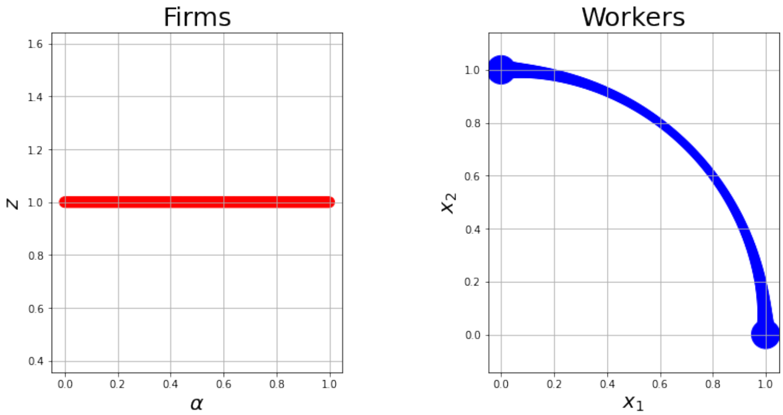
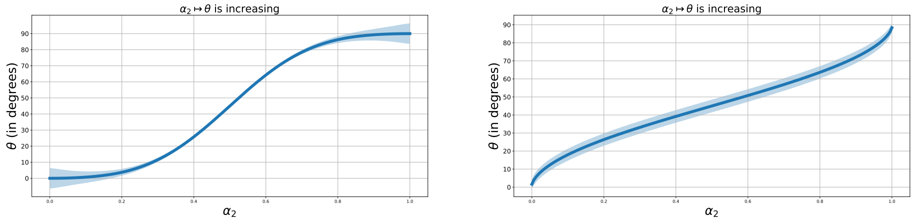


Fig. 7. Each dot represents a Dirac mass, and its size represents its weight. Here, all firm types have the same weight while there are more specialist workers in the economy than generalists (Scenario B)

Figure 8 shows the aggregate skill profile of employees (the ratio of skill 2 over skill 1) as a function of the technical intensity in skill 2 of their employing firm. Given the discrete nature of the workers' and firms' distributions μ and ν, the matching between firms and workers is not perfectly pure: the transport plan is not exactly a Monge map $\theta(\alpha)$, i.e., a firm of type α_i hires workers of different skill profiles θ_i. The dark blue lines represent the mean of the skill profiles of the workers employed by a firm of type α, and the light blue area contains values of θ within one standard deviation from the mean.

As predicted by [Choné and Kramarz, 2022], the firms whose technology is very intensive in skill 2 (i.e., high α) use more skill 2 relative to skill 1 in Scenario A compared to Scenario B. In other words, firms are able to specialize to their "core business" in Scenario A. This is because in that scenario (with many specialist workers), the salary tends to become linear (see Figure 9) and firms freely adjust the proportion of specialists they hire to achieve

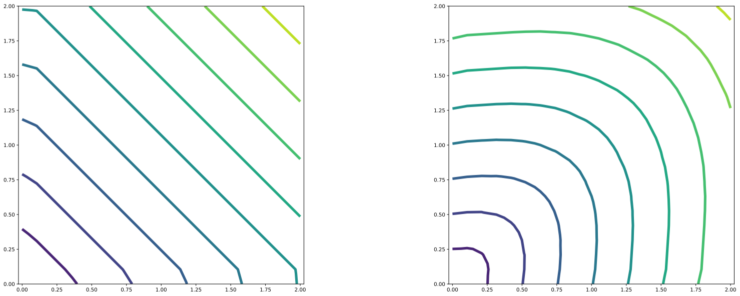


(a) Scenario A

(b) Scenario B

Fig. 8. Mapping between $\alpha \mapsto \theta$ under WOTUK.

their optimal mix of skills. In contrast, in Scenario B, the workers' salary is strictly convex, implying that specialist workers are expensive (compared to generalists) and hence it is too costly for firms to hire the specialists they would need to take full advantage of their technology.



(a) When there are many specialists workers in the economy, the salary becomes linear.

(b) When there are many generalist workers in the economy, the salary is strictly convex.

Fig. 9. The isowage curves $\psi(x_1, x_2)$ = constant in both Scenarios.

6.2 Endogenizing firms' sizes: Comparing WOT and WOTUK

Considering the above scenarios A and C, we now compare outcomes of interest at the optimum of the WOT and WOTUK problems. Recall that in the WOT problem all firms must have the same number of employees, while in the WOTUK problem firms freely choose their number of employees. We already checked in Section 5 that the objective of the primal problem is higher under WOTUK than under WOT.

[Choné and Kramarz, 2022] show that the number of employees is not uniquely defined at an optimum of the WOT and WOTUK problems. They establish, however, the uniqueness of the firm-aggregate skill $\int yq^x(dy)$ at a competitive equilibrium. We define hereafter the firms' sizes as the Euclidian norm of the firm-aggregate skill, and we denote the size as $\Lambda^d(\alpha)$.

In Scenario A, Figure 10 shows that the sorting maps that describe the matching between workers and firms are very close under WOT and WOTUK. By contrast, the figure shows that firms' sizes greatly differ at the optima of the two problems. In the WOT problem, the kernel (q^x) is constrained to be a family of probability measures, and because all workers' skill vectors have norm one in this scenario, the firms' sizes are one in that case. By contrast, we observe that under WOTUK specialist firms are bigger than generalist firms.

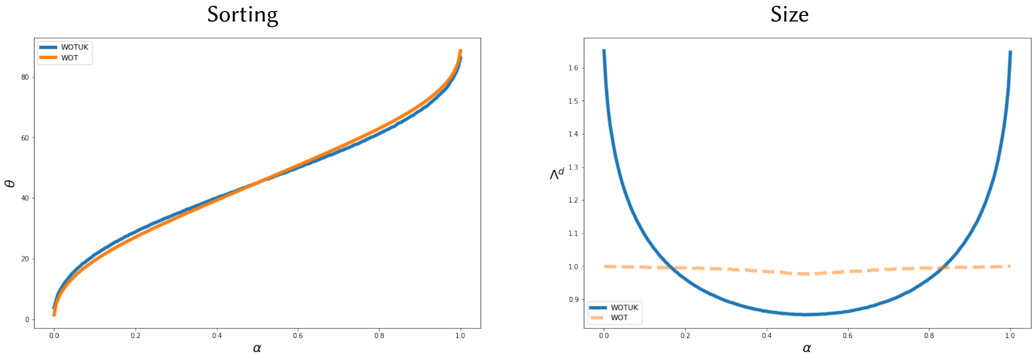


Fig. 10. Sorting and firm sizes under WOT and WOTUK (Scenario A, Fig. 1)

In Scenario C, firms and workers differ in two dimensions. We compute the average size of firms with given TFP, z, and with given technical intensity in skill 2, α. Figure 11 shows that both under WOT and WOTUK firms with higher total factor productivity have greater size, but the effect of TFP on size is much stronger when firms are free to adjust their number of employees. Contrary to what happens in Scenario A, firms with different technical intensities α can differ in size in Scenario C because in that scenario they can pick workers of different qualities. But the ability of firms to decide how many workers they hire greatly exacerbates their heterogeneity in size.

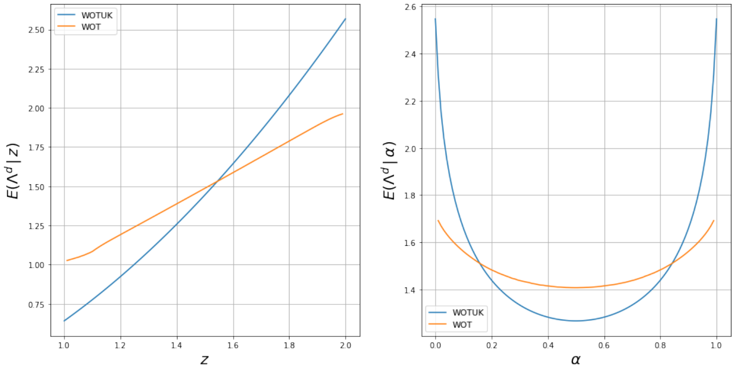


Fig. 11. Firms' sizes as a function of Total Factor Productivity z and technical intensity α (Scenario C, Fig. 5)

6.3 Closed-form solution for a one-dimensional conical WOTUK problem

We consider the conical example presented in Corollary 1.1 of [Choné et al., 2022]. We let $\mu = \nu$ be the uniform measure on $[0, 1]$ and we take the production function: $F(x, y) := x\, y^{\eta}$. These authors exhibit multiple solutions to the WOTUK problem but show that the value of the objective (total output) is unique and given by

$$O(\eta) = \frac{C^{\eta}}{2} \cdot \frac{1}{2 + a_0\, \eta},$$

where $C = (2 - \eta)/(1 - \eta)$ and $a_0 = 1/(1 - \eta)$. We sample the uniform laws with 200 points and run our algorithm for 30 iterations. Figure 12 compares the theoretical and numerical values of the objective.

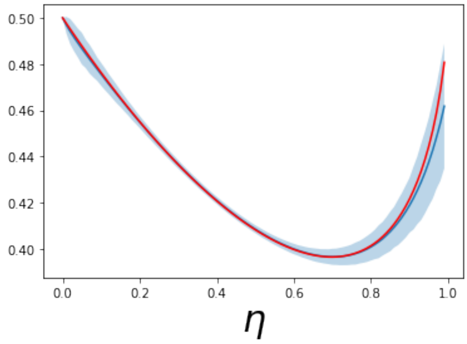


Fig. 12. In red, the theoretical value of the objective $O(\eta)$ for WOTUK when μ and ν are uniform. In blue, the output of the algorithm with in light blue the "confidence interval" obtained from $\overline{\mathcal{G}}$

The WOT problem has been shown (see Theorem 1.3 of [Backhoff-Veraguas and Pammer, 2022]) to be stable in the sense that the solution to the discretized problem tends to the solution of the problem associated with the continuous distribution when we refine the grid. The property is not yet known for WOTUK. However, we also try our algorithm in the WOTUK case for different numbers n of sampling points. Here η is fixed at 0.5 and we run 30 iterations. Figure 13 suggests that the objective of the discretized problem tends to the objective of the continuous one as n increases.

6.4 Pure solution of a one-dimensional non-conical WOTUK problem

In this subsection, we try our algorithm on a one-dimensional, non-conical, WOTUK problem for which the optimal transport plan is pure and unique. We consider the problem of maximizing

$$\int \left(\int k(x, y)\, q^x(dy) \right)^{\eta} \mu(dx), \tag{17}$$

over the kernels (q^x) that satisfy $\int q^x(dy)\, \mu(dx) = \nu(dy)$ and $0 < \eta < 1$. Hereafter, we take $\mu = \nu$ as the uniform distribution on $[0, 1]$. The above specification reflects a non-conical production function. Here the firm's types x may represent quality of capital (e.g., of machine) or management style. The types y represents workers' qualities. An employee's task is produced from the interaction of the employee's and firm's types, with that interaction being assumed to be log-supermodular, i.e., to satisfy $\partial^2 \ln k/\partial x \partial y > 0$. In our numerical application, we adopt $k(x, y) = \exp(xy)$. Then, the tasks produced by each employee are

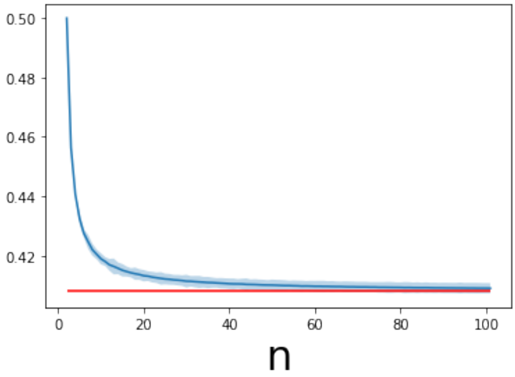


Fig. 13. In red, the theoretical value of the objective $O(\eta = 0.5)$ for WOTUK when μ and ν are uniform. In blue, the output of the algorithm (in light blue the "confidence interval" obtained from $\overline{\mathcal{G}}$). n number of iterations.

aggregated. Finally, output is produced from the firm-aggregated task under decreasing returns to scale ($\eta < 1$). While in the conical setting the firm's type x would be interacted only with a firm-level aggregate (namely the sum of the employees' skills), the non-conical specification (17) involves interactions of x with each of the individual employees' types y separately.

From Theorem 4.2 of [Choné et al., 2022], we know that the optimal transport plan is unique and is of the form:

$$q^x(dy) = N(x)\,\delta_{T(x)}(dy),$$

where $T(x)$ is a map from $[0,1]$ into $[0,1]$ and $N(x)$ is a map from $[0,1]$ to $\mathbb{R}_+$. In other words, a firm of type x hires $N(x)$ employees, all with the same type $y = T(x)$. The firm-level aggregate task is therefore given by

$$\int e^{xy}\,q^x(dy) \;=\; N(x)\,e^{xT(x)}.$$

The equilibrium condition $q\mu = \nu$ yields a relation between the sorting map $T(x)$ and the size of the firms $N(x)$. When μ and ν are uniform on $[0,1]$, this relation is simply $N(x) = T'(x)$. We thus look for T increasing and sufficiently smooth that maximizes

$$\int_0^1 N(x)^\eta\,e^{xT(x)}\,dx \;=\; \int_0^1 (T'(x))^\eta\,e^{xT(x)}\,dx \;=\; \int_0^1 H(T(x),T'(x))\,dx.$$

The Euler-Lagrange equation

$$\frac{\partial H}{\partial T}(T(x),T'(x)) = \frac{d}{dx}\left[\,\frac{\partial H}{\partial T'}(T(x),T'(x))\,\right]$$

yields in this case

$$\eta(1-\eta)T'' + (1-\eta)x(T')^2 - \eta TT' = 0. \tag{18}$$

For $\eta = .5$, the unique solution is $T(x) = x$. For other values of η, we solve numerically the differential equation (18) with SciPy[3] and compare the results to those found with our algorithm. Figure 14 shows that the numerical solutions for the sorting map $T(x)$ found by the two methods perfectly coincide.

[3] See [Virtanen et al., 2020].

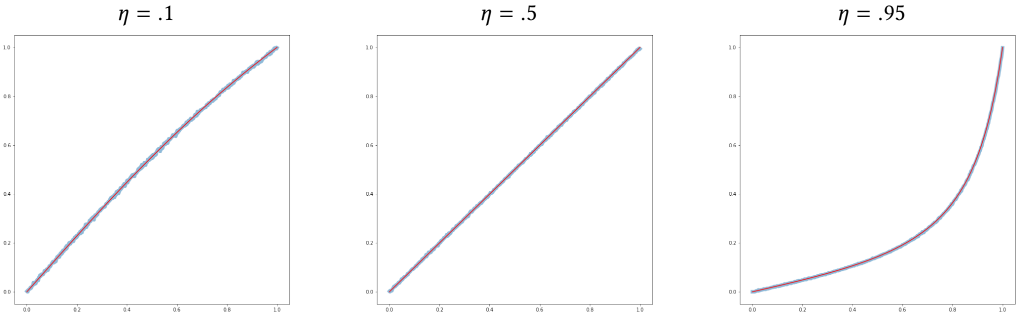


Fig. 14. In red, the solution of the equation (18). In light blue, the mapping obtained by our algorithm after 100,000 iterations for the uniform distributions μ and ν discretized with 500 points

As already mentioned, if $\eta = .5$, all firms have the same size $N(x) = 1$. By contrast, when the returns to scale are close to one ($\eta = .95$), Figure 15 shows that the size of firms is strongly increasing in their type. Using the equality between wage and marginal productivity, the wage of a worker $y = T(x)$ employed in firm x can be recovered as

$$w(T(x)) = \eta e^{\eta x T(x)} N(x)^{\eta - 1},$$

where $N(x) = T'(x)$. We find that the wage function, as the firm's size, is strongly increasing and convex in the firm's type. Hence, with log-supermodular worker-firm interactions, the matching of workers and firms generates "superstar firms" (see [Rosen, 1981]) that hire the best workers and are both very large and very productive.

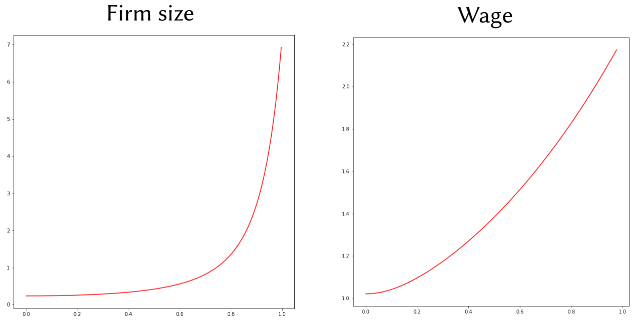


Fig. 15. Superstars: Firm's size $N(x)$ and worker's wage $w(y)$ are strongly increasing for $\eta = .95$

7 CONCLUSION AND FUTURE WORK

Over the past few years, optimal transport (OT) has found applications to various domains such as graphics [Bonneel et al., 2016, Solomon et al., 2015], imaging [Cuturi and Peyré, 2016, Rabin and Papadakis, 2015], generative models [Arjovsky et al., 2017, Salimans et al., 2018], biology [Hashimoto et al., 2016, Schiebinger et al., 2019], NLP [Alaux et al., 2019, Grave et al., 2019], finance [Acciaio et al., 2020, Beiglböck et al., 2013, Galichon et al., 2014] and economics [Galichon, 2016, Galichon and Salanié, 2015, Lindenlaub, 2017]. The key in making the optimal transport approach work in these applications lies in the different forms of regularization added to the classical optimal transport problem. Most of these reguralizations, including the celebrated entropic regularization [Cuturi, 2013], correspond to penalized versions of the Monge-Kantorovich problem, possibly with relaxed [Chizat, 2017, Figalli, 2010] or tightened [Beiglböck et al., 2013, Paty et al., 2020] constraints.

None of the above variants allow to model workers-to-firms matching with aggregation of workers' skills within employing firms and endogenous choice of size by firms. For this, we need to rely on weaker forms of optimal transport.[4] In this paper, we have proposed algorithms to compute solutions to weak optimal transport with normalized or unnormalized kernel problems in the discrete setting, both in their primal and dual formulations.

Assuming exogenous labor supply, the matching problem studied in this paper is equivalent to finding competitive equilibria in a pure exchange economy with a continuum of agents and commodities.[5] In our context, the agents and commodities are respectively the firms and the workers' skill sets. [Choné and Kramarz, 2022] make labor supply endogenous by allowing workers and firms to trade (possibly at some cost) stand-alone skills on separate markets. In practice, this "unbundling" of skills is made possible by new technologies, increased access to outsourcing, and by online platforms where buyers and sellers can trade specialized tasks. In equilibrium, workers choose how much skills to unbundle (given the wages offered by employing firms) while wages are determined from the firms' demand for work (given the workers' unbundling choices). Developing algorithms to solve for such equilibria is an important avenue for future research.

REFERENCES

Beatrice Acciaio, Mathias Beiglboeck, and Gudmund Pammer. Weak transport for non-convex costs and model-independence in a fixed-income market. arXiv preprint arXiv:2011.04274, 2020.

Jean Alaux, Edouard Grave, Marco Cuturi, and Armand Joulin. Unsupervised hyper-alignment for multilingual word embeddings. In International Conference on Learning Representations, 2019.

J-J Alibert, Guy Bouchitté, and Thierry Champion. A new class of costs for optimal transport planning. European Journal of Applied Mathematics, 30(6):1229–1263, 2019.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. Proceedings of the 34th International Conference on Machine Learning, 70:214–223, 2017.

Julio Backhoff-Veraguas and Gudmund Pammer. Stability of martingale optimal transport and weak optimal transport. The Annals of Applied Probability, 32(1):721–752, 2022.

Mathias Beiglböck, Pierre Henry-Labordere, and Friedrich Penkner. Model-independent bounds for option pricesa mass transport approach. Finance and Stochastics, 17(3):477–501, 2013.

Nicolas Bonneel, Gabriel Peyré, and Marco Cuturi. Wasserstein barycentric coordinates: histogram regression using optimal transport. ACM Transactions on Graphics, 35(4):71:1–71:10, 2016.

Elsa Cazelles, Felipe Tobar, and Joaquin Fontbona. Streaming computation of optimal weak transport barycenters. arXiv preprint arXiv:2102.13380, 2021.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: geometry and Kantorovich formulation. Journal of Functional Analysis, 274(11):3090–3123, 2018.

Lénaïc Chizat. Unbalanced optimal transport: models, numerical methods, applications. PhD thesis, 2017.

Philippe Choné and Francis Kramarz. Matching workers' skills and firms' technologies: From bundling to unbundling. CEPR Working Paper 17645, 2022.

Philippe Choné, Nathael Gozlan, and Francis Kramarz. Weak optimal transport with unnormalized kernels. arXiv preprint arXiv: preprint: 2203.16227, 2022.

Marco Cuturi. Sinkhorn distances: lightspeed computation of optimal transport. In Advances in Neural Information Processing Systems 26, pages 2292–2300, 2013.

[4] WOT is connected to classic problems in other fields, such as entropic transport and martingale transport, see Appendix A.

[5] [Khan and Yannelis, 1991] examine such equilibria but rule out product aggregation

Marco Cuturi and Gabriel Peyré. A smoothed dual approach for variational Wasserstein problems. SIAM Journal on Imaging Sciences, 9(1):320–343, 2016.

Marco Cuturi, Laetitia Meng-Papaxanthos, Yingtao Tian, Charlotte Bunne, Geoff Davis, and Olivier Teboul. Optimal transport tools (ott): A jax toolbox for all things wasserstein, 2022.

Jan Eeckhout and Philipp Kircher. Assortative matching with large firms. Econometrica, 86(1):85–132, 2018.

Alessio Figalli. The optimal partial transport problem. Archive for Rational Mechanics and Analysis, 195 (2):533–560, 2010.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. Journal of Machine Learning Research, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Alfred Galichon. Optimal Transport Methods in Economics. Princeton University Press, 2016.

Alfred Galichon and Bernard Salanié. Cupids invisible hand: Social surplus and identification in matching models. Available at SSRN 1804623, 2015.

Alfred Galichon, Pierre Henry-Labordère, and Nizar Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. Annals of Applied Probability, 24(1): 312–336, 2014.

Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali. Kantorovich duality for general transport costs and applications. Journal of Functional Analysis, 273(11):3327–3405, 2017. ISSN 0022-1236. doi: https://doi.org/10.1016/j.jfa.2017.08.015. URL https://www.sciencedirect.com/science/article/pii/S0022123617303294.

Edouard Grave, Armand Joulin, and Quentin Berthet. Unsupervised alignment of embeddings with wasserstein procrustes. 2019.

Tatsunori Hashimoto, David Gifford, and Tommi Jaakkola. Learning population-level diffusions with generative RNNs. In International Conference on Machine Learning, pages 2417–2426, 2016.

James Heckman and Jose Scheinkman. The importance of bundling in a gorman-lancaster model of earnings. The Review of Economic Studies, 54(2):243–255, 1987.

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

Alexander S Kelso and Vincent P Crawford. Job matching, coalition formation, and gross substitutes. Econometrica: Journal of the Econometric Society, pages 1483–1504, 1982.

M Ali Khan and Nicholas C Yannelis. Equilibria in markets with a continuum of agents and commodities. In Equilibrium theory in infinite dimensional spaces, pages 233–248. Springer, 1991.

Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport, 2022.

Ilse Lindenlaub. Sorting multidimensional types: Theory and application. The Review of Economic Studies, 84(2):718–789, 2017.

François-Pierre Paty, Alexandre d'Aspremont, and Marco Cuturi. Regularity as regularization: Smooth and strongly convex brenier potentials in optimal transport. In Silvia Chiappa and Roberto Calandra, editors, Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, volume 108 of Proceedings of Machine Learning Research, pages 1222–1232. PMLR, 26–28 Aug 2020. URL https://proceedings.mlr.press/v108/paty20a.html.

François-Pierre Paty and Marco Cuturi. Regularized optimal transport is ground cost adversarial. In Hal Daumé III and Aarti Singh, editors, Proceedings of the 37th International Conference on Machine Learning, volume 119 of Proceedings of Machine Learning Research, pages 7532–7542. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/paty20a.html.

Julien Rabin and Nicolas Papadakis. Convex color image segmentation with optimal transport distances. In International Conference on Scale Space and Variational Methods in Computer Vision, pages 256–269. Springer, 2015.

Sherwin Rosen. The economics of superstars. The American economic review, 71(5):845–858, 1981.

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.

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.

Justin Solomon, Fernando De Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. Convolutional Wasserstein distances: efficient optimal transportation on geometric domains. ACM Transactions on Graphics, 34(4):66:1–66:11, 2015.

Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. Nature methods, 17(3):261–272, 2020.

Alan Geoffrey Wilson. The use of entropy maximizing models, in the theory of trip distribution, mode split and route split. Journal of Transport Economics and Policy, pages 108–126, 1969.

A OTHER EXAMPLES OF WOT PROBLEMS

Entropic optimal transport. The entropic optimal transport (EOT) problem [Wilson, 1969] is a variant of the Kantorovich problem in which an entropic regularization term is added:

$$\mathscr{S}_\varepsilon(\mu,\nu) \overset{\text{def}}{=} \inf_{\pi\in\Pi(\mu,\nu)} \iint_{\mathcal{X}\times\mathcal{Y}} c\,d\pi + \varepsilon\,\mathrm{KL}(\pi|\mu\otimes\nu) \tag{19}$$

$$= \inf_{\pi\in\Pi(\mu,\nu)} \int c\,d\pi + \varepsilon \int \log\frac{d\pi}{d\mu\otimes\nu}\,d\pi \tag{20}$$

where $c \in C(\mathcal{X}\times\mathcal{Y})$ is the cost function, $\varepsilon > 0$ is the regularization strength, and $\mathrm{KL}(\cdot|\cdot)$ is the relative entropy (or Kullback-Leibler divergence). Considering the disintegration $(\pi^x)_{x\in\mathcal{X}}$ of π with respect to μ, and noting that $\frac{d\pi}{d\mu\otimes\nu}(x,y) = \frac{d\pi^x}{d\nu}(y)$, problem (20) rewrites:

$$\inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} \left[\int_{\mathcal{Y}} \left(c(x,y) + \varepsilon\log\frac{d\pi^x}{d\nu}(y)\right) d\pi^x(y)\right] d\mu(x)$$

which corresponds to the WOT problem (4) with

$$\mathcal{F}(x,p) = \int_{\mathcal{Y}} \left(c(x,y) + \varepsilon\log\frac{dp}{d\nu}(y)\right) p(dy)$$

$$= \int_{\mathcal{Y}} c(x,y)\,p(dy) + \varepsilon\,\mathrm{KL}(p|\nu).$$

Martingale optimal transport. The martingale optimal transport (MOT) problem [Beiglböck et al., 2013] is a variant of the Kantorovich problem in which the optimal transport plan is constrained to be a martingale:

$$\sup_{\substack{\pi\in\Pi(\mu,\nu)\\ \mu\,\text{a.e.},\int y\,d\pi^x(y)=x}} \iint_{\mathcal{X}\times\mathcal{Y}} F(x,y)\,d\pi(x,y).$$

Up to the fact that $\mathcal{F}$ is now allowed to take value $+\infty$, this problem corresponds to the WOT problem (4) with $\mathcal{F}(x,p) = \int c(x,y)\,p(dy) - \iota\left(\mu\,\text{a.e.}, \int y\,p(dy) = x\right)$ where $\iota(a)$ equals 0 if the assertion a is true, and $+\infty$ if a is false.

B PROOFS

B.1 Proof for the projection onto $\Pi(\ll\mu,\nu)$

Projecting a matrix $P \in \mathbb{R}_+^{n\times m}$ onto $\Pi(\ll\mu,\nu)$ means solving the following optimization problem:

$$\min_{Q\in\Pi(\ll\mu,\nu)} \mathrm{KL}(Q|P) = \min_{\substack{Q\in\mathbb{R}_+^{n\times m}\\ Q^\top 1=b}} \mathrm{KL}(Q|P) = \min_{\substack{Q\in\mathbb{R}^{n\times m}\\ Q^\top 1=b}} \sum_{i,j} Q_{ij}\left(\log\frac{Q_{ij}}{P_{ij}} - 1\right)$$

where we can drop the non-negativity constraint over Q for it is already constrained by the log in the objective.

The Lagrangian of the problem is

$$L(Q,\lambda) = \sum_{i,j} Q_{ij}\left(\log\frac{Q_{ij}}{P_{ij}} - 1\right) + \langle\lambda, b - Q^\top \mathbf{1}\rangle$$

where $\lambda \in \mathbb{R}^m$ is the Lagrange multiplier for the constraint $Q^\top \mathbf{1} = b$. The problem is convex, so the first-order condition is sufficient for optimality. So the solution should verify

$$\log \frac{Q_{ij}}{P_{ij}} = \lambda_j$$

hence

$$Q_{ij} = \frac{P_{ij} b_j}{\sum_{i'} P_{i'j}}.$$

B.2 Proof for the dual formulation of $\psi_\star$

Let us first compute the dual problem corresponding to

$$\min_{\substack{p \in \mathbb{R}_+^m \\ \sum_{j=1}^m p_j = 1 \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi \rangle.$$

One has:

$$\begin{aligned}
\min_{\substack{p \in \mathbb{R}_+^m \\ \sum_{j=1}^m p_j = 1 \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi \rangle &= \min_{p \in \mathbb{R}_+^m} \langle p, \varphi \rangle + \sup_{\lambda \in \mathbb{R}^q, \mu \in \mathbb{R}} \left\langle \lambda, z - \sum_{j=1}^m p_j y_j \right\rangle + \mu \left(1 - \sum_{j=1}^m p_j \right) \\
&= \sup_{\lambda \in \mathbb{R}^q, \mu \in \mathbb{R}} \langle \lambda, z \rangle + \mu + \inf_{p \in \mathbb{R}_+^m} \sum_{j=1}^m p_j \left(\varphi_j - \mu - \langle \lambda, y_j \rangle \right) \\
&= \sup_{\substack{\lambda \in \mathbb{R}^q, \mu \in \mathbb{R} \\ \forall j, \langle \lambda, y_j \rangle + \mu \leq \varphi_j}} \langle \lambda, z \rangle + \mu
\end{aligned}$$

where we have swapped the $\min$ and the $\max$ using the strong duality theorem for linear programs.

Likewise, the dual of

$$\min_{\substack{p \in \mathbb{R}_+^m \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi \rangle$$

will be the same as before but without $\mu \in \mathbb{R}$, because we have dropped the associated constraint, i.e.

$$\min_{\substack{p \in \mathbb{R}_+^m \\ \sum_{j=1}^m p_j y_j = z}} \langle p, \varphi \rangle = \sup_{\substack{\lambda \in \mathbb{R}^q \\ \forall j, \langle \lambda, y_j \rangle \leq \varphi_j}} \langle \lambda, z \rangle.$$