OPTIMAL PARTITION OF A LARGE LABOR FORCE INTO WORKING PAIRS

ROBERT J. MCCANN[1] AND MAXIM TROKHIMTCHOUK[2]

Abstract. Kremer and Maskin (1996) introduced an idealized model for pairing workers and managers with different skill levels into small teams selected to maximize productivity. They used it to analyze the impact of technological change and widening skill gaps on labor market segregation. The present paper extends their model to a workforce with multidimensional skill-types, continuously distributed, and gives a mathematical analysis of the extension. Pure and mixed notions of optimal pairing are introduced, which play an important role in the formulation and analysis of the model. The existence and uniqueness of such pairings are established using techniques from the theory of optimal transportation and infinite-dimensional linear programming.

Contents

1. Introduction

In the past, several models were proposed to analyze the impact of technological change on labor market segregation. Kremer and Maskin (1996) [13] introduced a matching framework to model within-establishment collaboration between high- and low-skilled labor in hopes of explaining the recent trend of increasing wage inequality between high skill and low skill workers. In this framework workers of different skills are matched with each other to form teams or working pairs. The

---

[1]Department of Mathematics, University of Toronto, Toronto Ontario Canada M5S 3G3.
[2]Department of Mathematics, University of California, Berkeley California USA 94720.
RJM gratefully acknowledges the gracious hospitality of the Université de Marne-la-Vallée, the Mathematical Sciences Research Institute at Berkeley California, and the Institute for Computational Engineering and Sciences at the University of Texas (Austin) during various stages of this work; his research was partially funded by National Science Foundation grant DMS 0354729 and Natural Sciences and Engineering Research Council of Canada Grant 217006-03 RGPIN.

productivity of a pair is supposed to depend on skills of both workers, and the firm assigns workers to teams so as maximize its total production. Kremer and Maskin used this matching framework to argue that increases in the mean and dispersion of the skill distribution in the economy leads to segregation of the labor market. Li and Suen [14] extended this matching framework to the case when the distribution of labor skill types is taken to be continuous and derived several properties of optimal matching. It is however not evident a priori that such a continuous optimal matching problem has a solution or whether it is unique. In this paper, we introduce two rigorous formulations of the continuous multidimensional problem, and give conditions under which these formulations become equivalent, and for which there exists a unique optimal partitioning scheme. In addition we exhibit sufficient conditions for a partition to be optimal. We also confirm several of Li and Suen's results using the rigorous formulation outlined in this paper.

The authors are pleased to express their gratitude to Hao Li, for introducing them to this problem and providing stimulating conversations.

2. Rigorous formulation of the pure pairing problem

We begin with a rigorous definition of the problem. Here and throughout, the space of skill types $\mathbf{X}$ will be a compact Hausdorff space; typically $\mathbf{X} \subseteq \mathbf{R}^n$.

From a mathematical point of view, it is natural to represent the skill distribution of the workforce by a Borel probability measure on the space $\mathbf{X}$. In this context, the pairing of the higher skilled worker, whom we may refer to as a **manager**, with a lower skilled worker, referred to as an **assistant**, is accomplished by a Borel map $f : \mathbf{X} \mapsto \mathbf{X}$ which maps the distribution of managers to the distribution of assistants. This terminology is inherited from the one-dimensional case; the multidimensional generalization allows for workers with complementary strengths in different areas, lending itself to a richer interpretation than the manager-assistant paradigm suggests. Note that even when the skill levels of all workers are known, and the productivity $p(x, y)$ of each working pair is a specified function, presumably depending asymmetrically on the skill level x of the manager and the skill level y of the assistant, it is not obvious which employees should be tasked with which roles: whether it is preferable for the assistant of a talented manager to be more skilled than the manager of a weak assistant, or vice versa. This is one of the questions which such a model is intended to resolve.

Let us denote the known skill distribution of the labor force by the Borel probability measure ω on $\mathbf{X}$, and the unspecified skill distributions of managers by μ and assistants by ν, so that $\mu + \nu = 2\omega$. The above constraints characterize the possible competitors in the optimization problem which we formulate below. The insistence that each manager be assigned a single assistant — rather than sharing a probabilistic combination of many assistants — leads to the notion which we call a pure pairing (in an imperfect analogy with the strategies of game theory). Mathematically, this notion of pure pairing is encoded in the next two definitions. The converse requirement that each assistant have a single manager is not imposed, but sometimes turns out to be a consequence of the net productivity maximization.

Definition 2.1 (Push-forward)**.** Given measurable spaces $(\mathbf{X}, \Sigma)$ and $(\mathbf{X}', \Sigma')$ with a measure μ on $\mathbf{X}$, each measurable map $f : \mathbf{X} \mapsto \mathbf{X}'$ induces a measure on $\mathbf{X}'$, called the **push-forward** of μ through f and denoted $\mu' := f_{\#}\mu$, defined by $\mu'(B') = \mu(f^{-1}(B'))$ for each $B' \in \Sigma'$. Equivalently, for each measurable function

$F : \mathbf{X}' \mapsto \mathbf{R}$,

$$\int_{\mathbf{X}'} F d(f_{\#}\mu) = \int_{\mathbf{X}} (F \circ f)\, d\mu.$$

Definition 2.2 (Pure pairing)**.** From here on we denote the set of Borel probability measures on $\mathbf{X}$ by $\mathscr{P}(\mathbf{X})$. Given a measure $\omega \in \mathscr{P}(\mathbf{X})$, we say that (μ, f) is a **pure pairing** for ω whenever $\mu \in \mathscr{P}(\mathbf{X})$, and $f : \mathbf{X} \mapsto \mathbf{X}$ is a Borel map defined μ-a.e. such that $\mu + f_{\#}\mu = 2\omega$. We denote the set of all pure pairings by $\Gamma_{\omega}^{\text{pure}}$.

For every such pure pairing (μ, f), the expression $\int_{\mathbf{X}} p(x, f(x))\, d\mu$ gives the corresponding total production. Hereafter we will assume $p(x, y)$ is continuous — to ensure above integral can be defined — and nonnegative without loss of generality. Then the maximal possible production is given by

$$K_{\omega} := \sup \Big\{ \int_{\mathbf{X}} p(x, f(x))\, d\mu \,|\, (\mu, f) \in \Gamma_{\omega}^{\text{pure}} \Big\}. \tag{2.1}$$

Of course, it is not clear a priori whether the supremum is attained by any pure pairing. If it is, we call such a pairing **optimal**. We refer to the value K_{ω} of the above supremum as **optimal paired productivity**.

This definition of an optimal pairing generalizes the finite optimization problem nicely, but the non-linear dependence of the supremum and constraints (2.1) on f make it difficult to analyze. Fortunately, mathematical developments in the theory of optimal transportation provide a way to circumvent this difficulty. As in game theory, the Kantorovich approach [11] to Monge's transportation problem [20] introduces a larger class of competitors which convexifies the problem and linearizes the cost, facilitating its analysis. A discrete version of this approach was already explored by Li and Suen in [14].

3. Relaxation: the mixed pairing problem and duality

The similarity between the Monge-Kantorovich and optimal pairing problems will be apparent to those familiar with the former theory, which is reviewed in Villani [27]. In fact the only thing that distinguishes the two problems are the constraints on μ and ν; in the Monge-Kantorovich problem μ and ν would be fixed exogenously, while in the present problem they are allowed to vary, with only their average ω being prescribed. This similarity suggests the definition of **mixed** competitors for the optimal partitioning problem.

Definition 3.1 (Mixed pairing)**.** A probability measure $\gamma \in \mathscr{P}(\mathbf{X} \times \mathbf{X})$ is called a **mixed pairing** of $\omega \in \mathscr{P}(\mathbf{X})$ if its two marginals average to ω, that is, if $\gamma[B \times \mathbf{X}] + \gamma[\mathbf{X} \times B] = 2\omega[B]$ for all Borel sets $B \subseteq \mathbf{X}$. The set of mixed pairings γ of ω is denoted by $\Gamma_{\omega}^{\text{mix}} \subset \mathscr{P}(\mathbf{X}^2)$.

Definition 3.2 (Optimal mixed pairing)**.** We define an **optimal** mixed pairing to be any measure $\gamma_0 \in \Gamma_{\omega}^{\text{mix}}$ which attains the following supremum:

$$J_{\omega} = \sup \Big\{ \int_{\mathbf{X}} p(x, y)\, d\gamma(x, y) \,|\, \gamma \in \Gamma_{\omega}^{\text{mix}} \Big\}. \tag{3.1}$$

We refer to the value of the above supremum as **optimal mixed productivity**.

As a consequence of the Riesz-Markov theorem, $\gamma \in \mathscr{P}(\mathbf{X}^2)$ is a mixed pairing for ω if and only if

$$\int_{\mathbf{X}^2} \{u(x) + u(y)\}\, d\gamma(x,y) = 2\int_{\mathbf{X}} u(x)\, d\omega(x) \qquad \forall u \in C(\mathbf{X}) \tag{3.2}$$

where $C(\mathbf{X})$ is the Banach space of continuous functions on $\mathbf{X}$ equipped with the sup norm. The advantage enjoyed by the mixed pairing problem (3.1) over the pure pairing problem is its linearity in γ on the convex, weak-$*$ compact set $\Gamma_\omega^{\mathrm{mix}}$. A maximizer can therefore be shown to exist using continuity and compactness, and can be characterized using the Kuhn-Tucker conditions which emerge from the dual linear program. We will go on to prove existence of an optimal mixed pairing in this section and to state the dual linear problem. We will also give sufficient conditions for duality and optimality. However, we will postpone the actual proof of the duality theorem for our linear programming problem until the next section.

We begin by proving existence of optimal mixed pairing.

Theorem 3.3 (Existence of optimal mixed pairing)**.** *For any $\omega \in \mathscr{P}(\mathbf{X})$ there exists an optimal mixed pairing.*

Proof. Set $\Gamma_\omega^{\mathrm{mix}}$ is a compact set in a weak-$*$ topology, hence continuous mapping

$$\gamma \to \int_{\mathbf{X}\times\mathbf{X}} p(x,y)\; d\gamma(x,y)$$

obtains its supremum over it. □

Looking at (3.2) and drawing upon analogy with Monge-Kantorovich optimal pairing problem we might suspect that the dual linear problem consists of finding a **feasible potential** — an element of the set

$$\mathcal{U} = \left\{ u \in L^1(\mathbf{X}, d\omega) \;\middle|\; u \text{ is lower semi-continuous, } u(x) + u(y) \geq p(x,y) \right\},$$

which attains the infimum

$$I_\omega = \inf\left\{ 2\int_{\mathbf{X}} u(x)\; d\omega(x) \;\middle|\; u \in \mathcal{U} \right\}. \tag{3.3}$$

Finite dimensional linear programming provides further reason to suspect that this is the right dual problem, see e.g. [14] where duality for the discrete matching problem was derived. Assuming lower semi-continuity of $u \in \mathcal{U}$ will turn out to cost no generality since $p(x,y)$ has been assumed to be continuous.

We proceed to state sufficiency of the Kuhn-Tucker conditions and some aspects of the asserted duality.

Theorem 3.4 (Sufficient conditions for duality and optimality)**.** *With $p : \mathbf{X}^2 \mapsto [0,\infty)$ nonnegative, continuous and $\omega \in \mathscr{P}(\mathbf{X})$, it follows that $I_\omega \geq J_\omega \geq K_\omega$. If for some $\gamma \in \Gamma_\omega^{\mathrm{mix}}$ and some feasible potential $v \in L^1(\mathbf{X}, d\omega)$*

$$2\int_{\mathbf{X}} v(x)\; d\omega = \int_{\mathbf{X}\times\mathbf{X}} p(x,y) d\gamma(x,y), \tag{3.4}$$

then $J_\omega = I_\omega$, with the supremum (3.1) and infimum (3.3) being attained by γ and v respectively.

Proof. First of all it is clear that $J_\omega \geq K_\omega$. This is because any pure pairing (μ, f) is also a mixed pairing of the form $(id \times f)_{\#}\mu$. Furthermore, since for any $\gamma \in \Gamma_\omega^{\mathrm{mix}}$ and any $u \in L^1(\mathbf{X}, d\omega)$ such that $u(x) + u(y) \geq p(x, y)$, we have

$$\int_{\mathbf{X}} 2u(x)\, d\omega = \int_{\mathbf{X}^2} u(x) + u(y)\, d\gamma \geq \int_{\mathbf{X}^2} p(x, y)\, d\gamma,$$

we see immediately that $I_\omega \geq J_\omega$. This proves the first claim. As for the second claim, equality (3.4) together with

$$\int_{\mathbf{X}} 2v(x)\, d\omega \geq I_\omega \geq J_\omega \geq \int_{\mathbf{X}\times\mathbf{X}} p(x, y)\, d\gamma,$$

immediately imply that $I_\omega = J_\omega$ with I_ω attained by v and J_ω attained by γ. $\square$

The value $v(x)$ of the minimizer may be interpreted as the gross marginal utility provided to the company by a worker with skill level x, relative to the utility provided by its other employees. The deviation of this value from the wage commanded by a similar worker in the marketplace would be an appropriate parameter to keep in mind for a company planning to expand its workforce.

Example 3.5. Let us now describe an example which motivates above problem. Suppose skill levels are measured on a scale from 0 to 1 and production function $p : [0, 1] \times [0, 1] \to \mathbf{R}$ is of the form $p(x, y) = x^\alpha y^\beta$, where $\alpha, \beta > 0$. This is the so-called Cobb-Douglas production function. Given that our skill distribution ω is represented by some $L^1([0, 1])$ density, we wish to know how this distribution is split into distribution of managers and assistants. For example, is it possible for managers to have the same skill levels as some assistants? This is the problem studied in [14], where Li and Suen concluded that this is indeed impossible as long as labor distribution is assumed to be concentrated sufficiently close to 1. In the last section we will provide independent confirmation of their results in our setting.

4. Proof of duality theorem

The fact that under suitable conditions linear programs possess dual problems is very well known and goes back to the work of Dantzig [5], von Neumann [29], Kantorovich [10] and Koopmans [12]. There are many books available on the subject; Andersen and Nash [1] for example is particularly relevant to our infinite dimensional setting. For the particular case of optimal transport a version of the duality theorem was originally proved by Kantorovich [10]. Since then the duality theorem for optimal transportation has been proved in a rather general setting [22] [23]. In this section we will derive a duality theorem for our linear programming problem. We could do that using abstract Fenchel-Rockafellar duality theorem, as in for instance [2]; however, we choose to proceed by exploiting Monge-Kantorovich duality instead, which we feel presents a more concrete exposition.

Let us first review Monge-Kantorovich duality. As we mentioned this duality has been much studied and there is extensive literature on it. For a modern treatment with the emphasis on optimal transportation see Rachev and Rüschendorf [21]. See also Villani [28] for a nice and thorough exposition.

There are several ways to state Monge-Kantorovich duality. The following Theorem presents it in a way that is most useful to us in this paper.

Theorem 4.1. *For any two measures $\mu, \nu \in \mathscr{P}(\mathbf{X})$*

$$\sup \left\{ \int_{\mathbf{X}\times\mathbf{X}} p(x,y)\, d\gamma \,\middle|\, \gamma \in \Gamma(\mu,\nu) \right\} = \tag{4.1}$$

$$\inf \left\{ \int_{\mathbf{X}} \phi(x)\, d\mu(x) + \int_{X} \psi(y)\, d\nu(y) \,\middle|\, (\phi,\psi) \in \Phi_p \right\} \tag{4.2}$$

where $\Gamma(\mu,\nu)$ is a set of measures in $\mathscr{P}(\mathbf{X}\times\mathbf{X})$ with x— and y— marginals being μ and ν respectively, and Φ_p is a set of lower semi-continuous pairs in $L^1(d\mu)\times L^1(d\nu)$ satisfying

$$\phi(x) + \psi(y) \geq p(x,y) \quad \textit{for } (x,y) \in \mathbf{X}\times\mathbf{X}. \tag{4.3}$$

Furthermore, both supremum (4.1) *and infimum* (4.2) *are attained. In addition maximizer γ of* (4.1) *and minimizer (ϕ,ψ) of* (4.2) *are characterized by the following*

$$\phi(x) + \psi(y) = p(x,y) \quad \textit{holds for all } (x,y) \in \mathit{spt}\,\gamma. \tag{4.4}$$

Above, sptγ refers to the smallest closed set carrying the mass of γ.

Having reviewed Monge-Kantorovich duality we return to the proof of duality for our pairing problem. The concept that will play an important role in the following is that of symmetrized productivity

$$s(x,y) := \sup\{p(x,y), p(y,x)\}. \tag{4.5}$$

It turns out that optimal mixed productivity with production function $p(x,y)$ is the same as the optimal mixed productivity with symmetrized production function $s(x,y)$. Intuitively, if for some pairing we have a pair of workers with manager of skill x and assistant of skill y, but with

$$p(x,y) < p(y,x),$$

we can always improve paired productivity by making the worker of skill x assistant and worker of skill y manager. The symmetrized production function, however, is symmetric and we can exploit this symmetry together with Monge-Kantorovich duality applied to the production function $s(x,y)$ to obtain a feasible potential which would satisfy Theorem 3.4.

Let us introduce some notation. For any measure $\gamma \in \mathscr{P}(\mathbf{X}\times\mathbf{X})$ by $\gamma \llcorner S$ we mean the measure γ restricted to the set S, i.e. $\gamma\llcorner S[B] = \gamma[B\cap S]$. For any set S by $S^\dagger$ we denote the reflection $S^\dagger = \{(y,x) \in \mathbf{X}\times\mathbf{X} \,|\, (x,y) \in S\}$. Also by $\gamma^\dagger$ we mean a push-forward of γ by a map that interchanges the two coordinates. Notice that the sum of marginals of $\gamma^\dagger$ is the same as that of γ and $\operatorname{spt}(\gamma^\dagger) = (\operatorname{spt}\gamma)^\dagger$.

We proceed to prove that mixed optimal productivity with production function $p(x,y)$ and production function $s(x,y)$ are the same.

Proposition 4.2.

$$\sup \left\{ \int p(x,y)\, d\gamma(x,y) \,\middle|\, \gamma \in \Gamma_\omega^{\text{mix}} \right\} = \sup \left\{ \int s(x,y)\, d\gamma(x,y) \,\middle|\, \gamma \in \Gamma_\omega^{\text{mix}} \right\}.$$

Proof. From (4.5), the expression on the right is greater then the expression on the left. Now let γ be any measure in $\Gamma_\omega^{\text{mix}}$ and

$$S = \left\{ (x,y) \in \mathbf{X}\times\mathbf{X} \,\middle|\, s(x,y) > p(x,y) \right\}.$$

Let $\tilde{\gamma} = (\gamma \llcorner S)^\dagger + \gamma \llcorner S^c$, then $\tilde{\gamma} \in \Gamma^{\mathrm{mix}}_\omega$ and

$$\begin{aligned}\int_{\mathbf{X}\times\mathbf{X}} s\, d\gamma &= \int_S s(x,y)\, d\gamma(x,y) + \int_{S^c} s(x,y)\, d\gamma(x,y) = \\ &= \int_S p(y,x)\, d\gamma(x,y) + \int_{S^c} p(x,y)\, d\gamma(x,y) = \\ &= \int_{\mathbf{X}\times\mathbf{X}} p\, d\tilde{\gamma}.\end{aligned}$$

This equality immediately implies that optimal mixed productivity with production function $p(x,y)$ is at least as big as with production function $s(x,y)$. □

Next we use the symmetry of $s(x,y)$ and Monge-Kantorovich duality to find a feasible potential that satisfies the condition (3.4) of Theorem 3.4, thus proving duality theorem and showing that the dual problem has a minimizer in one shot.

Remark 4.3. Before we proceed, let us point out that optimal mixed pairings are naturally maximizers of Monge-Kantorovich problem. This is because if $\gamma \in \Gamma^{\mathrm{mix}}_\omega$ is an optimal mixed pairing but not a Monge-Kantorovich maximizer, there would be some measure in $\Gamma(\pi^X_\# \gamma, \pi^Y_\# \gamma) \subset \Gamma^{\mathrm{mix}}_\omega$ with higher paired productivity then γ, which is obviously a contradiction. In particular for an optimal mixed pairing γ, the value of the Monge-Kantorovich supremum (4.1) with $\mu = \pi^X_\# \gamma$ and $\nu = \pi^Y_\# \gamma$ is the same as optimal mixed productivity corresponding to γ.

Theorem 4.4. *There exists a feasible potential v that satisfies the conclusion* (3.4) *of Theorem 3.4. In particular $I_\omega = J_\omega$ and the dual problem has a minimizer.*

Proof. Let $\gamma \in \Gamma^{\mathrm{mix}}_\omega$ be an optimal mixed pairing for the production function $s(x,y)$ and set $\tilde{\gamma}$ to be

$$\tilde{\gamma} = \frac{\gamma + \gamma^\dagger}{2},$$

then $\tilde{\gamma} \in \Gamma^{\mathrm{mix}}_\omega$ and is also an optimal mixed pairing for the production function $s(x,y)$ (this is because $s(x,y)$ is symmetric). Also $\tilde{\gamma}$ is a maximizer of the Monge-Kantorovich problem with the production function $s(x,y)$ among measures in $\Gamma(\pi^X_\# \tilde{\gamma}, \pi^Y_\# \tilde{\gamma})$ as was pointed out in Remark 4.3 above. Now we can apply Theorem 4.1 to conclude that there exist functions $(\phi, \psi) \in \Phi_s$ with

$$\phi(x) + \psi(y) = s(x,y) \quad \text{for all } (x,y) \in \operatorname{spt} \tilde{\gamma}.$$

However, $\operatorname{spt} \tilde{\gamma} = \operatorname{spt} \gamma \bigcup \operatorname{spt} \gamma^\dagger$, thus

$$\frac{\phi(x) + \psi(x)}{2} + \frac{\phi(y) + \psi(y)}{2} \geq s(x,y)$$

with equality holding for $(x,y) \in \operatorname{spt} \gamma$. Now if we let

$$v(x) = \frac{\phi(x) + \psi(x)}{2}$$

then v is a feasible potential, since $s(x,y) \geq p(x,y)$, and

$$\int_{\mathbf{X}\times\mathbf{X}} s(x,y)\, d\gamma = \int_{\mathbf{X}\times\mathbf{X}} v(x) + v(y)\, d\gamma = 2 \int_{\mathbf{X}} v(x)\, d\omega(x).$$

Therefore Proposition 4.2 imply that v satisfies conditions of Theorems 3.4. □

At this point we have proved duality theorem and showed that dual problem has a minimizer. We will see later that just as in the case of the Monge-Kantorovich problem, we can take minimizers of the dual problem to be of a special form. This, as in the case of optimal transportation, turns out to be crucial for further analysis of the pairing problem.

5. Review of optimal transportation theory

In this section we review important concepts from the theory of optimal transportation and state some theorems that we will need later to prove existence and uniqueness of optimal pairing. The material in this section is well known. For the relevant discussion see e.g. Rochet [24], Gangbo [6], Gangbo and McCann [8, 9], Caffarelli [3], Knott and Smith [25], McCann [18], Villani [27, 28].

We begin this section by discussing notions of p-convexity and p-cyclical monotonicity.

Definition 5.1 (p-convexity)**.** Fix $u : \mathbf{X} \mapsto \mathbf{R} \cup \{+\infty\}$ not identically infinite.
(i) The **p-subdifferential** of u is the set

$$\partial_p u := \{(x, y) \in \mathbf{X} \times \mathbf{X} \mid u(z) \geq u(x) + p(z, y) - p(x, y) \quad \forall z \in \mathbf{X}\}.$$

(ii) The **p-transform** u^p of u is the function $u^p(y) := \sup\limits_{x \in \mathbf{X}} \{p(x, y) - u(x)\}$.
(iii)The function u is called **p-convex** if $u = u^{p\tilde{p}} := (u^p)^{\tilde{p}}$ where $\tilde{p}(y, x) = p(x, y)$, or equivalently, if there is a set $A \subset \mathbf{X} \times \mathbf{R}$ such that

$$u(x) = \sup_{(y,\lambda) \in A} p(x, y) + \lambda.$$

(iv) A set $S \subset \mathbf{X} \times \mathbf{X}$ is called **p-cyclically monotone** if for each $k \in \mathbf{N}$, chain $(x_1, y_1), \ldots, (x_k, y_k) \in S$, and permutation σ on k letters,

$$\sum_{i=1}^{k} p(x_i, y_i) \geq \sum_{i=1}^{k} p(x_i, y_{\sigma(i)}).$$

Remark 5.2. Notice that $(x, y) \in \partial_p u$ holds if and only if $u(x) + u^p(y) = p(x, y)$, which is also denoted by writing $y \in \partial_p u(x)$.

The relationship between p-convex functions and p-cyclically monotone subsets is made explicit by the following generalization of Rockafellar's theorem (see Rochet [24]):

Theorem 5.3. *A set $S \subset \mathbf{X} \times \mathbf{X}$ is p-cyclically monotone if and only if it is contained in the p-subdifferential of a p-convex function.*

A very useful notion which we will employ later is that of p-contact maps.

Definition 5.4. Suppose $u : \mathbf{X} \to \mathbf{R} \cup \{+\infty\}$ is p-convex. We say that a Borel map $f : \mathbf{A} \to \mathbf{X}$ is a p-contact map for u if $\{x \in \mathbf{A} \mid \{f(x)\} \neq \partial_p u(x)\}$ is of Lebesgue measure zero.

The Monge-Kantorovich optimal plans in general are not pure but the following condition of Gangbo, Carlier, and Ma, Trudinger and Wang, is sufficient to establish purity of optimal plans [7], [4], [16]. This condition can be viewed as a generalization to higher dimensions of the single-crossing criterion of Lorentz [15], Mirrlees [19], and Spence [26], and was called a generalized Spence-Mirrlees condition in [4].

Definition 5.5. (Twist condition) We say that p satisfies a twist condition if for all $x \in \mathbf{X}$ map $y \to \nabla_1 p(x,y)$ is injective. Here and in what follows $\nabla_1 p(x,y) = \left(\frac{\partial p}{\partial x^1}, \dots, \frac{\partial p}{\partial x^n}\right)$ and $\nabla_2 p(x,y) = \left(\frac{\partial p}{\partial y^1}, \dots, \frac{\partial p}{\partial y^n}\right)$.

Remark 5.6. It turns out (see Theorem 5.9 in [28]) that supports of maximizers of Monge-Kantorovich problem are p-cyclically monotone. Since Remark 4.3 tells us that optimal mixed pairings are also maximizers of the Monge-Kantorovich problem, they have p-cyclically monotone supports as well.

Now we state our **assumptions** on p and $\mathbf{X}$:

- **A1:** $\mathbf{X}$ is a compact subset of $\mathbf{R}^n$.
- **A2:** p is non-negative on $\mathbf{X}$.
- **A3:** p is continuously differentiable on $\mathbf{X}$.
- **A4:** p satisfies the twist condition of Definition 5.5.

It is well know that p-convex functions under some assumptions possess regularity properties. In consequent sections we need the following result:

Lemma 5.7. *Under our assumptions on p, p-convex functions are Lipschitz.*

Proof. For the proof we refer to the proof of Theorem 10.26 in [28]. □

In what follows $\mathscr{P}_{ac}(\mathbf{X})$ is a set of Borel measures on X absolutely continuous with respect to Lebesgue measure.

Theorem 5.8. *Let S be a p-cyclically monotone set. Then there exists a p-contact map f such that all $\gamma \in \mathscr{P}(X \times X)$ with $spt\gamma \subseteq S$ and $\pi^X_{\#}\gamma \in \mathscr{P}_{ac}(\mathbf{X})$ are of the form $\gamma = (id \times f)_{\#}(\pi^X_{\#}\gamma)$.*

Proof. The proof of this theorem is essentially contained in the proof of Theorem 5.26 in [28]. We need only to check that for any p-convex function ψ the set of $x \in \mathbf{X}$ such that $\partial_p \psi(x)$ contains more then one element has Lebesgue measure zero. However, this follows easily from differentiability of ψ (see Lemma 5.7) and twist condition 5.5. □

Theorem 5.9 (Uniqueness of optimal transport)**.** *Let f and g be two p-contact maps and $\mu \in \mathscr{P}_{ac}(\mathbf{X})$. If $f_{\#}\mu = g_{\#}\mu$, then $f = g$, μ-almost everywhere.*

Proof. The above uniqueness statement is contained in Theorem 10.27 in [28]. □

6. Existence and uniqueness of optimal pairing

In this section we show that optimal mixed pairing is pure and characterize the possible non-uniqueness. In addition we show that the optimal paring is unique if we assume that the only optimal way to pair managers and assistants who have the same skill distributions is by pairing workers of the same skill level, so that the only optimal transport between identical distributions is the identity map.

Our proof of the fact that optimal mixed pairing is pure is reminiscent of the proof of purity in Gangbo and McCann [9]. Basically we show that union of supports of all optimal mixed pairings is a p-cyclically monotone set. This together with results from optimal transportation will enable us to conclude existence of a pure pairing.

Theorem 6.1 (Existence of optimal pure pairings)**.** *Let p satisfy assumptions **A1-A4** and $\omega \in \mathscr{P}_{ac}(\mathbf{X})$. Then there is a p-contact map f such that all optimal* *mixed pairings γ are of the form $\gamma = (id \times f)_\#(\pi^X_\# \gamma)$. It is unique in the sense that if for some p-contact map g, some optimal mixed pairing γ is of the form $\gamma = (id \times g)_\#(\pi^X_\# \gamma)$, then $f(x) = g(x)$ for $(\pi^X_\# \gamma)$-almost every x.*

Proof. Let us denote the set of optimal mixed pairings in $\Gamma^{\mathrm{mix}}_\omega$ by Γ_0. Let

$$S = \bigcup_{\gamma \in \Gamma_0} \operatorname{spt} \gamma$$

be the union of supports of all optimal mixed pairings. We claim that S is p-cyclically monotone. Indeed, the definition of p-cyclical monotonicity is a condition on finite subsets, thus a union of collection of sets is p-cyclically monotone if and only if the union of any finite subcollection of sets is p-cyclically monotone. However, it is clear that the union of any finite number of supports of optimal mixed pairings is itself the support of a measure that is some convex combination of optimal mixed pairings, hence an optimal mixed pairing itself, and therefore has p-cyclically monotone support by Remark 5.6. See Corollary 2.4 in [9], where a similar argument was used to prove purity of optimal transport.

At this point we established existence of a p-cyclically monotone set S that contains supports of all optimal mixed pairings. By Theorem 5.8, which was discussed in the previous section, we can conclude existence of a p-contact map f such that all optimal mixed pairings γ are of the form $(id \times f)_\#(\pi^X_\# \gamma)$. The uniqueness statement is a consequence of Theorem 5.9. □

Next we derive characterization of non-uniqueness.

Theorem 6.2 (Controlling non-uniqueness)**.** *Suppose p satisfies assumptions **A1**-**A4** and $\omega \in \mathscr{P}_{ac}(\mathbf{X})$, and let f be as in the statement of Theorem 6.1. For any two mixed optimal pairings $\gamma_1, \gamma_2 \in \Gamma^{\mathrm{mix}}_\omega$ with corresponding x-marginals μ_1, μ_2 we have*

$$(\mu_2 - \mu_1)_+ = f_\# \left[(\mu_2 - \mu_1)_-\right], \quad (\mu_2 - \mu_1)_- = f_\# \left[(\mu_2 - \mu_1)_+\right], \tag{6.1}$$

$$|\mu_2 - \mu_1| = f_\# |\mu_2 - \mu_1|. \tag{6.2}$$

Proof. Let γ_1 and γ_2 be two mixed optimal pairings. Theorem 6.1 tells us that γ_1 and γ_2 can be written as $(id \times f)_\# \mu_1$ and $(id \times f)_\# \mu_2$ respectively. Since $(id \times f)_\# \mu_1, (id \times f)_\# \mu_2 \in \Gamma^{\mathrm{mix}}_\omega$ we have the following equality $\mu_1 + f_\# \mu_1 = \mu_2 + f_\# \mu_2 = 2\omega$, in particular $(\mu_2 - \mu_1) + f_\#(\mu_2 - \mu_1) = 0$. Denoting by μ_+ and μ_- positive and negative parts of $\mu_2 - \mu_1$ we see immediately that $\mu_+ + f_\#(\mu_+) = \mu_- + f_\#(\mu_-)$. Let $A \subseteq \mathbf{X}$ denote the set of full measure for μ_+ and zero measure for μ_- provided by the Hahn decomposition and set $B = \mathbf{X} \backslash A$. We observe that

$$\mu_+[\mathbf{X}] \le \mu_+[A] + f_\#(\mu_+)[A] = f_\#(\mu_-)[A] \le \mu_-[\mathbf{X}],$$

However, since $(\mu_+ - \mu_-)[\mathbf{X}] = (\mu_2 - \mu_1)[\mathbf{X}] = 0$, we conclude that $f_\#(\mu_+)[A] = 0$. Similarly we conclude $f_\#(\mu_-)[B] = 0$. In particular, this implies that $\mu_+ = f_\#(\mu_-)$, $\mu_- = f_\#(\mu_+)$, and therefore $|\mu_2 - \mu_1| = f_\# |\mu_2 - \mu_1|$. □

To deduce uniqueness we need to assume that identity is the only optimal matching between two equal measures. In terms of economic interpretation this means the following: given managers and assitants with the same distribution of skills the only optimal way to pair then is to match manager with assistant of the same skill.

Corollary 6.3 (Uniqueness)**.** *Suppose p satisfies assumptions **A1-A4** and $\omega \in \mathscr{P}_{ac}(\mathbf{X})$. If the diagonal $\{(x,x) \in \mathbf{X} \times \mathbf{X}\}$ is p-cyclically monotone, then there is a unique optimal mixed pairing.*

Proof. Let $\gamma_1, \gamma_2 \in \Gamma_\omega^{\mathrm{mix}}$ be two optimal mixed pairings. The above Theorem, equation (6.2) and Theorem 5.9 imply that f is equal to identity $|\mu_2 - \mu_1|$-almost everywhere. But then by (6.1), $(\mu_2 - \mu_1)_+ = (\mu_2 - \mu_1)_-$, which is a contradiction unless $\mu_2 = \mu_1$. $\square$

7. Geometric structure of optimal pairing

In this paper we chose to exploit the Monge-Kantorovich theory to prove existence and uniqueness of optimal pairings. However, we feel it is worthwhile to point out that this optimization problem enjoys a peculiar geometric structure, which could be alternatively used to prove most of the above results. We will also employ this structure to characterize optimal pairing as well as to deduce existence of minimizers of the dual problem that are of special form. Namely we will show that we can take such a minimizer to be

$$\frac{\phi(x) + \phi^s(x)}{2},$$

where s is the symmetrized cost and ϕ is s-convex. Since under our assumptions on p, s-convex functions are Lipschitz (see Lemma 5.7), this implies that minimizers of the dual problem can be taken to be Lipschitz as well.

Let us thus suppose that $\gamma \in \Gamma_\omega^{\mathrm{mix}}$ is optimal with x– and y– marginals μ and ν respectively, and that v is a minimizer of the dual problem. Then duality implies

$$\int v(x)\, d\mu(x) + \int v(y)\, d\nu(y) = 2 \int v(x)\, d\omega(x) = \int p(x,y)\, d\gamma(x,y),$$

and together with $v(x) + v(y) \geq p(x,y)$ we conclude that

$$p(x,y) = v(x) + v(y) \quad \text{for all } (x,y) \in \operatorname{spt}\gamma.$$

The fact that this equality holds on the support of optimal pairings has particular implications for its geometry, which we state in the following Proposition.

Proposition 7.1. Let v be a feasible potential and S_v be the set

$$S_v = \left\{(x,y) \,\middle|\, v(x) + v(y) = p(x,y)\right\}.$$

Then for any k points $(x_1, x_2), \ldots, (x_{2k-1}, x_{2k}) \in S_v$ and for any permutation σ on $2k$ symbols the following holds:

$$\sum_{i=1}^{k} p(x_{2i-1}, x_{2i}) \geq \sum_{i=1}^{k} p(x_{\sigma(2i-1)}, x_{\sigma(2i)}). \tag{7.1}$$

Proof. The proof of this Proposition is very straightforward. Let $(x_1, x_2), \ldots, (x_{2k-1}, x_{2k})$ be k points in S_v, then

$$\begin{aligned}\sum_{i=1}^{k} p(x_{2i-1}, x_{2i}) &= \sum_{i=1}^{k} v(x_{2i-1}) + v(x_{2i}) = \\ &= \sum_{i=1}^{k} v(x_{\sigma(2i-1)}) + v(x_{\sigma(2i)}) \\ &\geq \sum_{i=1}^{k} p(x_{\sigma(2i-1)}, x_{\sigma(2i)}).\end{aligned}$$

□

Definition 7.2. Let us say $S \subset \mathbf{X} \times \mathbf{X}$ is *p-optimal* if for any k points $(x_1, x_2), \ldots, (x_{2k-1}, x_{2k}) \in S$ and for any permutation σ on $2k$ symbols the following inequality holds:

$$\sum_{i=1}^{k} p(x_{2i-1}, x_{2i}) \geq \sum_{i=1}^{k} p(x_{\sigma(2i-1)}, x_{\sigma(2i)}). \tag{7.2}$$

We obtained the existence of p-optimal sets from feasible potentials, but we can ask ourselves whether for any p-optimal set S there exists a feasible potential such that $S \in S_v$. The answer is yes.

Proposition 7.3. If S is a p-optimal set, then there exists a feasible potential v of the form

$$v(x) = \frac{\phi(x) + \phi^s(x)}{2}$$

such that $S \subseteq S_v$. We call feasible potentials of this special form **canonical**.

Proof. Let S be a p-optimal set. We first show that there exists a s-convex ϕ whose s-subdifferential contains $S \cup S^\dagger \subset \partial_s \phi$. To do this we first prove that if S is p-optimal, then $T := S \cup S^\dagger$ is s-optimal.

Indeed, given any $(x_1, x_2) \ldots (x_{2k-1}, x_{2k}) \in S$ and permutation σ there exists a permutation τ such that

$$\begin{aligned}\sum_{i=1}^{k} s(x_{\sigma(2i-1)}, x_{\sigma(2i)}) &= \sum_{i=1}^{k} p(x_{\tau(2i-1)}, x_{\tau(2i)}) \\ &\leq \sum_{i=1}^{k} p(x_{2i-1}, x_{2i}) \\ &\leq \sum_{i=1}^{k} s(x_{2i-1}, x_{2i}).\end{aligned}$$

Thus S is s-optimal. On the other hand $s(x, y) = s(y, x)$, so any sequence $(x_1, x_2) \ldots (x_{2k-1}, x_{2k})$ from $S \cup S^\dagger$ corresponds, after reordering some number of the pairs, to a sequence from S. This shows that $T := S \cup S^\dagger$ is s-optimal.

Since T is s-optimal, it is s-cyclically monotone a fortiori. Now Rochet's generalization Theorem 5.3 of Rockafellar's theorem, applied to the s-cyclically monotone set $T = S \cup S^\dagger$, yields the desired s-convex function $\phi = \phi^{ss}$ with $T \subset \partial_s \phi$.

Now from the definition of ϕ^s we obtain $\phi^s(y) := \sup_{z\in\mathbf{X}}\{s(z,y)-\phi(z)\} \geq s(x,y)-\phi(x)$ for all $(x,y)\in\mathbf{X}\times\mathbf{X}$. Thus

$$\begin{aligned}
v(x)+v(y) &:= \frac{\phi(x)+\phi^s(y)}{2}+\frac{\phi(y)+\phi^s(x)}{2}\\
&\geq \frac{s(x,y)+s(y,x)}{2}\\
&= s(x,y)\\
&\geq p(x,y).
\end{aligned}$$

For $(x,y)\in S$, the first inequality is saturated because both (x,y) and (y,x) lie in $\partial_s\phi$; the second inequality is saturated due to the case $k=1=\sigma(2)$ and $\sigma(1)=2$ in the definition (7.2) of p-optimality of S. $\square$

The above discussion points to the following characterization of optimal pairing.

Theorem 7.4 (Characterization of optimal support)**.** *Let p be continuous. A measure $\gamma\in\Gamma_\omega^{\mathrm{mix}}$ is an optimal mixed pairing if and only if sptγ is p-optimal. Also, γ is an optimal mixed pairing if and only if spt$\gamma\subseteq S_v$ for some feasible potential v. In particular the infimum I_ω is attained by a canonical potential.*

Proof. Suppose $\gamma\in\Gamma_\omega^{\mathrm{mix}}$ is optimal. As we discussed at the beginning of this section, if v is any minimizer of the dual problem, then $\operatorname{spt}\gamma\subseteq S_v$, hence $\operatorname{spt}\gamma$ is p-optimal by Proposition 7.1.

Suppose conversely that $\gamma\in\Gamma_\omega^{\mathrm{mix}}$ has p-optimal support. By Proposition 7.3 we conclude that there exists canonical feasible potential v such that

$$v(x)+v(y)=p(x,y)\quad\text{for all } (x,y)\in\operatorname{spt}\gamma.$$

Therefore, since v is lower semi-continuous and non-negative $(2v(x)\geq p(x,x))$,

$$\int_{\mathbf{X}\times\mathbf{X}} p(x,y)\,d\gamma(x,y)=\int_{\mathbf{X}\times\mathbf{X}} v(x)+v(y)\,d\gamma(x,y)=\int_{\mathbf{X}} v(x)\,d\omega(x)$$

and by Theorem 3.4 we deduce that v is a minimizer of the dual problem and γ is an optimal mixed pairing. $\square$

Remark 7.5. As we discussed at the beginning of this section canonical feasible potentials are Lipschitz due to Lemma 5.7, hence differentiable Lebesgue almost everywhere.

The concepts of p-optimality and canonical potential are clearly analogous to concepts of p-cyclical monotonicity and p-convex function. In fact it is possible to prove existence of optimal pure pairings by pursuing this analogy further and taking an approach similar to the one taken by one of us [17] to prove existence and uniqueness of optimal transport. Thus one could show that there exists a measure γ_0 in $\Gamma_\omega^{\mathrm{mix}}$ with p-optimal support, construct a canonical feasible potential using Proposition 7.3, and then use the fact that equality

$$v(x)+v(y)=p(x,y)$$

holds for all $(x,y)\in\operatorname{spt}\gamma_0$ together with Theorem 3.4 to conclude duality, optimality of γ_0 and v all in one go.

8. Examples and discussion

In this section we revisit several results obtained by Li and Suen in [14] using the mathematical framework we have introduced. Some of the results generalize their work to a multidimensional setting. We begin by deriving a condition that must be satisfied for a worker to be indifferent to being a manager or an assistant. Then we proceed to specialize this condition to the case of a multidimensional Cobb-Douglas type production function and deduce complete segregation of the labor market into managers and assistants in the case of skill distribution concentrated in a narrow skill band. We then derive efficiency of the linear matching with a given slope and point out how one can obtain examples of optimal pairing without role segregation.

The next lemma is multidimensional generalization of Lemma 5.1 in [14]. It essentially states that if in an optimal pairing a worker of skill level z is indifferent to being matched as a manager to assistant of skill level y or as an assistant to a manager of skill level x, then x, y, z must satisfy a condition involving the partial derivatives of p.

Lemma 8.1. *Suppose p is continuously differentiable. Let $S \subset \mathbf{X} \times \mathbf{X}$ be p-optimal. Then for Lebesgue almost all $z \in \pi^X(S) \cap \pi^Y(S)$ we have*

$$\nabla_1 p(z, y) = \nabla_2 p(x, z), \tag{8.1}$$

where $(x, z) \in S$ and $(z, y) \in S$.

Proof. Let v be a canonical feasible potential such that $S_v \supset S$. By Lemma 5.7 and Rademacher theorem v is differentiable Lebesgue almost everywhere. Let $z \in \pi^X(S) \cap \pi^Y(S)$ be a point where v is differentiable and suppose $x, y \in \mathbf{X}$ are such that $(z, y), (x, z) \in S$. Since v is a feasible potential and $S_v \supset S$

$$v(a) + v(b) - p(a, b) \geq 0$$

for all $(a, b) \in \mathbf{X}$ with equality holding for (z, y) and (x, z). Hence we conclude that

$$\nabla v(z) = \nabla_1 p(z, y), \quad \nabla v(z) = \nabla_2 p(x, z),$$

and the lemma follows. □

We need the following lemma to relate the support of the marginal to the projection of the support.

Lemma 8.2 (Projected support is dense in support of projection)**.** *If $\gamma \geq 0$ is a Borel measure on $\mathbf{R}^n \times \mathbf{R}^n$, then $spt\,(\pi^X_\# \gamma) = \overline{\pi^X(spt\,\gamma)}$.*

Proof. We first show that $\pi^X(\operatorname{spt}\gamma)$ is contained in $\operatorname{spt}(\pi^X_\#\gamma)$. Suppose that $x \in \pi^X(\operatorname{spt}\gamma)$, meaning there is a point $(x, y) \in \operatorname{spt}\gamma$. If $B \subset \mathbf{R}^n$ is a neighbourhood of x, then $B \times \mathbf{R}^n$ is a neighbourhood of (x, y), so $(\pi^X_\#\gamma)[B] = \gamma[B \times \mathbf{R}^n] > 0$. Therefore x is in support of $\pi^X_\#\gamma$. Because $\operatorname{spt}\pi^X_\#\gamma$ is closed, $\overline{\pi^X(\operatorname{spt}\gamma)} \subset \operatorname{spt}(\pi^X_\#\gamma)$.

Conversely, suppose $x \in \operatorname{spt}\pi^X_\#\gamma$. Then for every open set $U \in \mathbf{R}^n$ containing x, $\gamma[U \times \mathbf{R}^n] = (\pi^X_\#\gamma)[U] > 0$, hence $U \times \mathbf{R}^n$ intersects the support of γ. Thus U contains a point in $\pi^X(\operatorname{spt}\gamma)$, and we conclude that x belongs to the closure of $\pi^X(\operatorname{spt}\gamma)$. This completes the lemma. □

Let us now restrict our attention to the Cobb-Douglas type production functions of the form

$$p(x_1, ..., x_n, y_1, ..., y_n) = \sum_{i=1}^{n} x_i^{\alpha_i} y_i^{\beta_i} \tag{8.2}$$

on $[0,1]^{2n}$; the case $n = 1$ was discussed in Example 3.5. We take $\alpha_i > \beta_i \geq 1$ for each i, meaning the productivity of each pair is more sensitive to the manager's skill level than to the assistant's. We now specialize the preceding result to deduce and constrain the possibilities of substantial overlap between the skill levels of managers and assistants, represented by the intersection $\operatorname{spt}\mu \cap \operatorname{spt}\nu$.

Theorem 8.3. *Let ω be supported on $[0,1]^n$, and let $p(x,y)$ be as in (8.2) with $\alpha_i > \beta_i \geq 1$ for all i. Suppose $\gamma \in \Gamma_\omega^{\mathrm{mix}}$ is the optimal mixed pairing and μ, ν are the $x-$ and $y-$ marginals of γ respectively. Then for Lebesgue almost every z in $spt\,\mu \bigcap spt\,\nu$, there exist x, y such that $(x,z),(z,y) \in spt\,\gamma$ and the following equality holds for each i:*

$$\left(\frac{z_i}{x_i}\right)^{\alpha_i} \left(\frac{y_i}{z_i}\right)^{\beta_i} = \frac{\beta_i}{\alpha_i}. \tag{8.3}$$

Proof. First notice that equality $\nabla_1 p(z,y) = \nabla_2 p(x,z)$ for our particular p becomes (8.3). Now because π^X is a compact map and the set $\operatorname{spt}\gamma$ is closed and bounded, we have

$$\pi^X(\operatorname{spt}\gamma) = \overline{\pi^X(\operatorname{spt}\gamma)} = \operatorname{spt}\mu,$$

where last equality holds by Lemma 8.2. Similarly, $\pi^Y(\operatorname{spt}\gamma) = \operatorname{spt}\nu$. Thus by Lemma 8.1, the fact that $\operatorname{spt}\gamma$ is p-optimal implies that for almost every point of $\operatorname{spt}\mu \bigcap \operatorname{spt}\nu = \pi^X(\operatorname{spt}\gamma) \bigcap \pi^Y(\operatorname{spt}\gamma)$ our assertion is true. $\square$

We proceed to derive several consequences of this theorem. For the ratio $0 < \beta_i/\alpha_i < 1$, it is clear that (8.3) cannot be satisfied if for some i, x_i, y_i and z_i are all close to 1. Quantifying this observation allows us to establish the following result.

Proposition 8.4 (Role segregation in a narrow skill band)**.** Let γ be an optimal pairing for ω. If for some i, $\pi^{x_i}(\operatorname{spt}\omega)$ is contained in $](\beta_i/\alpha_i)^{\frac{1}{\alpha_i+\beta_i}}, 1]$, then $|\operatorname{spt}\mu \bigcap \operatorname{spt}\nu| = 0$.

Proof. To derive a contradiction, suppose $|\operatorname{spt}\mu \bigcap \operatorname{spt}\nu| \neq 0$. Then Theorem 8.3 provides $z \in \operatorname{spt}\mu \cap \operatorname{spt}\nu$ and $x, y \in \operatorname{spt}\omega$ such that equality (8.3) is true. For the moment set $a := (\beta_i/\alpha_i)^{\frac{1}{\alpha_i+\beta_i}}$. We note that $y_i > a$, $z_i > a$ and $\frac{1}{z_i} \geq 1$, hence $\left(\frac{y_i}{z_i}\right)^{\beta_i} > a^{\beta_i}$. However, equality (8.3) implies that

$$\frac{1}{x_i^{\alpha_i}} \frac{\beta_i}{\alpha_i} = \frac{a^{\alpha_i+\beta_i}}{x_i^{\alpha_i}} < \left(\frac{z_i}{x_i}\right)^{\alpha_i} a^{\beta_i} < \left(\frac{z_i}{x_i}\right)^{\alpha_i} \left(\frac{y_i}{z_i}\right)^{\beta_i} = \frac{\beta_i}{\alpha_i},$$

but that means $x_i > 1$, which is the desired contradiction. $\square$

The preceding proposition illustrates how clustering of skill levels in a narrow range can lead to essentially complete segregation of managers from assistants according to their different skill levels. We now would like to develop a class of examples that illustrate the opposite phenomenon, in which there is a full interval of skill levels where both managers and assistants need to be represented to achieve optimal productivity. To do this we can use Proposition 7.1 in Li and Suen [14], where they have shown that the pairwise matching with constant degree of segregation $\left(\frac{\beta}{\alpha}\right)^{\frac{1}{\alpha+\beta}}$ is optimal if it is feasible. We will confirm their Proposition by first showing that for $p \in C^1(\mathbf{R}\times\mathbf{R})$ of the form $p(x,y) = x^\alpha y^\beta$ there is precisely one line through the origin which is p-optimal in the plane.

Lemma 8.5. *For Cobb-Douglas production function $p(x,y) = x^\alpha y^\beta$ in $C^1([0,1]^2)$, there exists a unique line through the origin which restricted to $[0,1]^2$ is p-optimal. The slope of this line is $\lambda := (\beta/\alpha)^{1/(\alpha+\beta)}$.*

Proof. Suppose that some line S through the origin described by $y(x) = \lambda x$ is p-optimal, then proposition 7.3 gives us existence of canonical feasible potential v such that $S_v \supset S$.

Recall that canonical potential v is Lipschitz. Let now $(x_0, \lambda x_0)$ be a point in S, then notice that $v(x) + v(\lambda x_0) - p(x, \lambda x_0)$ as a function of x attains its infimum at x_0, hence for almost all x_0 it is the case that $v'(x_0) = \frac{\partial p}{\partial x}(x_0, \lambda x_0)$. This in particular implies that $v'(x) = \alpha x^{\alpha+\beta-1}\lambda^\beta$. Integrating both sides and using the fact that $2v(0) = p(0,0) = 0$, we derive that $v(x) = \lambda^\beta \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta}$. Substituting back into equation $v(x) + v(\lambda x) = p(x, \lambda x)$, we obtain

$$\lambda^\beta x^{\alpha+\beta} = \lambda^\beta \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \lambda^{2\beta+\alpha}\frac{\alpha}{\alpha+\beta} x^{\alpha+\beta}.$$

Therefore λ must necessarily be equal to $(\frac{\beta}{\alpha})^{\frac{1}{\alpha+\beta}}$. This proves uniqueness of p-optimal line. At this point to prove existence it suffices to show that $v(x)$ is a feasible potential, since Proposition 7.1 would then imply $S \subset S_v$ is p-optimal. To do this we need only to show that inequality $v(x) + v(y) \geq p(x,y)$ holds for all points $(x,y) \in [0,1]^2$ for $v(x) = (\frac{\beta}{\alpha})^{\frac{\beta}{\alpha+\beta}} \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta}$. To do that it suffices to check that the minimum of $(\frac{\beta}{\alpha})^{\frac{\beta}{\alpha+\beta}} \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + (\frac{\beta}{\alpha})^{\frac{\beta}{\alpha+\beta}} \frac{\alpha}{\alpha+\beta} y^{\alpha+\beta} - x^\alpha y^\beta$ on $[0,1]^2$ is zero, however this is not hard to do. Therefore, as claimed above there is only one line through the origin that is also a p-optimal subset of $[0,1]^2$ and its slope is equal to $(\frac{\beta}{\alpha})^{\frac{1}{\alpha+\beta}}$. □

Now we state Proposition which allow us to construct examples of optimal pairing.

Proposition 8.6. Fix $p(x,y) = x^\alpha y^\beta$. Let $\lambda = (\frac{\beta}{\alpha})^{\frac{1}{\alpha+\beta}}$ and denote by $l_\lambda(x) = \lambda x$ the line through the origin with slope λ. If μ is any measure supported in $[0,1]$ which is absolutely continuous with respect to Lebesgue, then (μ, l_λ) is the unique optimal pairing for $\omega = \mu + (l_\lambda)_\# \mu$.

Proof. By Lemma 8.5 and Theorem 7.4 the measure $(id \times l_\lambda)_\# \mu$ is optimal and to conclude uniqueness we need only then check that the diagonal is a p-cyclically monotone subset of the plane. To do that consider functions $u(x) := \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta}$ and $w(y) := \frac{\beta}{\alpha+\beta} y^{\alpha+\beta}$. It is easy to see that

$$\frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} y^{\alpha+\beta} - x^\alpha y^\beta \geq 0, \tag{8.4}$$

with equality when $x = y$. This implies that $u = w^{\tilde{p}}$ and $w = u^p$. Hence u is p-convex and since equality in (8.4) holds for $x = y$, diagonal $\{(x,x) \in \mathbf{X}\times\mathbf{X}\} \subseteq \partial_p u$ and therefore the diagonal is p-cyclically monotone. Now, since p satisfies assumption of Corollary 6.3, we conclude the proof of the Proposition. □

Corollary 8.7. *If f is any function in $L^1([0,1])$, then $(f\mathcal{L}^1, l_\lambda)$ is an optimal pairing for $\omega := h\mathcal{L}^1$, where $2h(x) = f(x) + f(x/\lambda)/\lambda$.*

Proof. This follows immediately from the fact that $(l_\lambda)_\# f\mathcal{L}^1 = h\mathcal{L}^1$, where $h(x) = f(x/\lambda)/\lambda$. $\square$

The Corollary above immediately provides us with a series of examples in which managers and assistants share same skill levels. The observation contained in the statement of the above Lemma also appears in Li and Suen's work. We refer the reader to [14] for an economic interpretation.

In conclusion we would like to note that our approach not only provides a rigorous foundation for the economic partitioning model discussed in [13, 14], but also uncovers a hidden geometric structure enjoyed by this model. This structure can be exploited, as we have done in the examples, to provide further insight into optimal partitioning.

References

[1] Edward J. Anderson and Peter Nash, *Linear programming in infinite dimensional spaces*, John Wiley and Sons Ltd, 1987.

[2] Haïm Brezis, *Analyse fonctionnelle*, Collection Mathématiques Appliquées pour la Maîtrise. [Collection of Applied Mathematics for the Master's Degree], Masson, Paris, 1983, Théorie et applications. [Theory and applications]. MR MR697382 (85a:46001)

[3] Luis A. Caffarelli, *Allocation maps with general cost functions*, Partial differential equations and applications, Lecture Notes in Pure and Appl. Math., vol. 177, Dekker, New York, 1996, pp. 29–35. MR MR1371577 (97f:49055)

[4] Guillaume Carlier, *Duality and existence for a class of mass transportation problems and economic applications*, Adv. Math. Econom. **5** (2003), 1–21.

[5] George B. Dantzig, *Programming in a linear structure*, Econometrica **17** (1949), 73–74.

[6] Wilfrid Gangbo, *An elementary proof of the polar factorization of vector-valued functions*, Arch. Rational Mech. Anal. **128** (1994), no. 4, 381–399. MR MR1308860 (96a:49048)

[7] ———, *Habilitation thesis*, Université de Metz, 1995.

[8] Wilfrid Gangbo and Robert J. McCann, *Optimal maps in Monge's mass transport problem*, C. R. Acad. Sci. Paris Sér. I Math. **321** (1995), no. 12, 1653–1658. MR MR1367824 (96i:49004)

[9] ———, *The geometry of optimal transportation*, Acta Math. **177** (1996), no. 2, 113–161. MR MR1440931 (98e:49102)

[10] Leonid V. Kantorovich, *On the translocation of masses*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **37** (1942), 199–201. MR MR0009619 (5,174d)

[11] ———, *On a problem of Monge*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **312** (2004), no. Teor. Predst. Din. Sist. Komb. i Algoritm. Metody. 11, 15–16, (Russian) Reprinted from C. R. (Doklady) Acad. Sci. URSS (N.S.) 3 (1948), no. 2. MR MR2117877

[12] Tjalling C. Koopmans, *Exchange ratios between cargoes on various routes*, Scientific papers of Tjalling C. Koopmans, pp. xii+600 pp. (1 plate), Springer-Verlag, New York, 1970. MR MR0264970 (41 #9559)

[13] Michael Kremer and Eric Maskin, *Wage inequality and segregation by skill*, Working Paper 5718, National Bureau of Economic Research, August 1996.

[14] Hao Li and Wing Suen, *Technological change and labor market segregation*, Working paper at `www.chass.utoronto.ca/ haoli/research/kremer.pdf`.

[15] George G. Lorentz, *An inequality for rearrangements*, Amer. Math. Monthly **60** (1953), 176–179. MR MR0052476 (14,626a)

[16] Xi-Nan Ma, Neil S. Trudinger, and Xu-Jia Wang, *Regularity of potential functions of the optimal transportation problem*, Arch. Rational Mech. Anal. **177** (2005), 151–183.

[17] Robert J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. **80** (1995), no. 2, 309–323. MR MR1369395 (97d:49045)

[18] ———, *Exact solutions to the transportation problem on the line*, R. Soc. Lond. Proc. Ser. A Math. Phys. Eng. Sci. **455** (1999), no. 1984, 1341–1380. MR MR1701760 (2000i:90002)

[19] James A. Mirrlees, *An exploration in the theory of optimum income taxation*, Rev. Econom. Stud. **38** (1971), 175–208.

[20] Gaspard Monge, *Mémoire sure la théorie des déblais et de remblais*, Histoire de l'Académie Royale des Sciences de Paris, avec les Mémoires de Mathématique et de Physique pour la même année (1781), 666–704.

[21] Svetlozar T. Rachev and Ludger Rüschendorf, *Mass transportation problems.*, Probability and its Applications (New York), Springer-Verlag, New York, 1998, Applications. MR MR1619171 (99k:28007)

[22] Doraiswamy Ramachandran and Ludger Rüschendorf, *A general duality theorem for marginal problems*, Probab. Theory Related Fields **101** (1995), no. 3, 311–319. MR MR1324088 (96a:60003)

[23] ———, *Duality and perfect probability spaces*, Proc. Amer. Math. Soc. **124** (1996), no. 7, 2223–2228. MR MR1342043 (96m:60003)

[24] Jean-Charles Rochet, *A necessary and sufficient condition for rationalizability in a quasilinear context*, J. Math. Econom. **16** (1987), no. 2, 191–200. MR MR902976 (89e:90016)

[25] Cyril S. Smith and Martin Knott, *Generalising monotonicity*, Distributions with given marginals and moment problems (Prague, 1996), Kluwer Acad. Publ., Dordrecht, 1997, pp. 253–259. MR MR1614679

[26] Michael Spence, *Competitive and optimal responses to signals: An analysis of efficiency and distribution*, J. Econom. Theory **7** (1974), 296–332.

[27] Cédric Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003)

[28] ———, *Optimal transport, old and new*, St. Flour Lecture Notes, Springer, New York, to appear. See also `www.umpa.ens-lyon.fr/~cvillani/Cedrif/B07.StFlour.pdf`.

[29] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, New Jersey, 1944. MR MR0011937 (6,235k)