

# Composite Sorting\*

Job Boerma

University of Wisconsin-Madison

Aleh Tsyvinski

Yale University

Ruodu Wang

University of Waterloo

Zhenyuan Zhang

Stanford University

March 2023

## Abstract

We propose a new tractable general framework for sorting – composite sorting. Composite sorting comprises of (1) distinct workers being assigned to the same job and (2) a given worker type simultaneously being part of both positive and negative sorting. We show that composite sorting generically arises in a class of sorting models when fixed investments can mitigate the variable costs of mismatch. We develop a complete characterization of equilibrium sorting as well as the corresponding equilibrium wages. Wages exhibit a local hierarchical structure meaning the relative wages depend solely on sorting patterns within narrow skill groups. Using this framework, we study within-job wage dispersion and demonstrate that quantitatively composite sorting may explain a sizable portion of wage dispersion within occupations in the United States.

**JEL-Codes:** J01, D31, C78

**Keywords:** Sorting, Assignment

---

\*We thank Carter Braxton, Hector Chade, Rasmus Lentz, Ilse Lindenlaub, Paolo Martellini, Luigi Pistaferri, Tommaso Porzio, Fabien Postel-Vinay, and Alexander Zimin for comments.

# 1 Introduction

Since [Koopmans and Beckmann \(1957\)](#) and [Becker \(1973\)](#), sorting models have taken a prominent position in economics. An important insight of this literature is when the output function is supermodular or submodular, sorting is positive or negative, implying that very similar or identical worker types are assigned to the same job type. [Sattinger \(1993\)](#), [Chiappori and Salanié \(2016\)](#), [Chade, Eeckhout, and Smith \(2017\)](#), and [Eeckhout \(2018\)](#) provide comprehensive reviews of this literature.<sup>1</sup>

Our paper provides the first complete characterization of optimal sorting and wages in a general sorting model with an output function that is neither supermodular nor submodular. We establish that equilibrium sorting is significantly richer than in the canonical settings yet we characterize it fully. We refer to the sorting pattern that emerges as composite sorting, which has two key features. First, distinct workers are sorted to the same job which enables us to study wage dispersion within jobs. Second, the fact that a given worker type can simultaneously be in both positive and negative sorting allows us to generate intricate sorting patterns to bring to the data. We quantitatively demonstrate our model using American Community Survey data. Our model can explain 32 percent of the overall wage dispersion and about 50 percent of the wage dispersion at the top and the bottom of the wage distribution. In contrast, with either supermodular or submodular costs, the solution is positive or negative one-to-one sorting with no dispersion in wages.<sup>2</sup>

We consider a sorting model with heterogeneous workers and heterogeneous jobs. Mismatch is the difference between the skill of the worker and the difficulty of the job. When the job's difficulty exceeds the worker's skill level, that is, a worker is underqualified, mismatch decreases output. The second type of mismatch (as in [Lise and Postel-Vinay \(2020\)](#)) arises when the worker's ability surpasses the job's demands, that is, a worker is overqualified, resulting in a utility loss for a worker and a corresponding loss in the joint surplus. Following [Stigler \(1939\)](#) and [Laffont and Tirole \(1986, 1991\)](#), firms can incur fixed costs to reduce the variable costs of mismatch. Firms thus mitigate the adverse effects of mismatch by investing in technologies or providing amenities. An example is investing in an industrial cobot (collaborative robot) that enhances the output of a low-skill welder. The result of the technology decision is that the effective output function is concave in the extent of underqualification. Similarly, employers may invest in amenities to decrease the disutility of workers (as in [Rosen \(1986\)](#)) when they are overqualified. An example is investing in a premium truck with advanced comfort and drivers support system for a

---

<sup>1</sup>[Antràs and Rossi-Hansberg \(2009\)](#) and [Costinot and Vogel \(2015\)](#) provide an overview of the significance of assignment models in international trade.

<sup>2</sup>To be precise, wage dispersion in this case can be positive but solely due to the discretization of occupations.

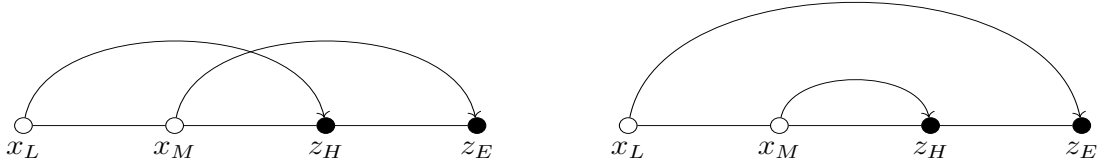


Figure 1: Pairings Between Workers and Jobs Do Not Intersect

Figure 1 illustrates that under an optimal sorting, pairings between workers and jobs do not intersect. Positive sorting, in the left panel, pairs a low-skill worker and a high-skill job  $(x_L, z_H)$ , and a medium-skill worker with an exceptional skill job  $(x_M, z_E)$ , resulting in two medium-size mismatches. For concave costs, however, it is preferable to have one large mismatch  $(x_L, z_E)$  and one small mismatch  $(x_M, z_H)$ . It is optimal to have pairs that do not intersect as in the right panel.

high-skill trucker. The result of investment in amenities is that the effective surplus is concave in the extent of overqualification. In sum, the effective output function is concave in mismatch. The key characteristic of a generic concave function in mismatch is that it is neither supermodular nor submodular. We are able to provide a complete characterization of this general sorting problem with neither supermodular nor submodular costs by analyzing it using the tools of optimal transport theory.

We now turn to the main characteristics of an optimal assignment. First, an optimal assignment maximizes the number of perfect pairs, which refer to pairs without mismatch. When mismatch costs are concave, it is preferable to have a combination of one pair with small mismatch and another with significant mismatch, as opposed to having two pairs with moderate mismatch. A combination of a perfect pair and a pair with significant mismatch exemplifies this.

The second feature of an optimal assignment is that pairings between workers and jobs do not intersect. Consider workers with low and medium skills  $\{x_L, x_M\}$  and jobs that require high and exceptional skills  $\{z_H, z_E\}$ . Positive sorting assigns the low-skill worker to the high-skill job  $(x_L, z_H)$ , and a medium-skill worker to the exceptional skill job  $(x_M, z_E)$ . This results in two medium-size mismatch costs. Visualizing the pairing as arcs connecting a worker with a job reveals that the pairs intersect as in the left panel of Figure 1. For concave costs, in contrast, it is preferable to have one large mismatch (between a low-skill worker and a very high-skill job  $(x_L, z_E)$ ) and one small mismatch (between a medium-skill worker and a high-skill job  $(x_M, z_H)$ ). Thus, it is optimal to have pairs that do not intersect as in the right panel of Figure 1.

This principle of no intersecting pairs gives rise to two additional characteristics of an optimal assignment. First, the assignment problem can be decomposed into independent problems by a layer mechanism. Layers are constructed by first designing a measure of underqualification that evaluates the cumulative deficiency of worker skills compared to the cumulative demands of jobs at a given skill level.

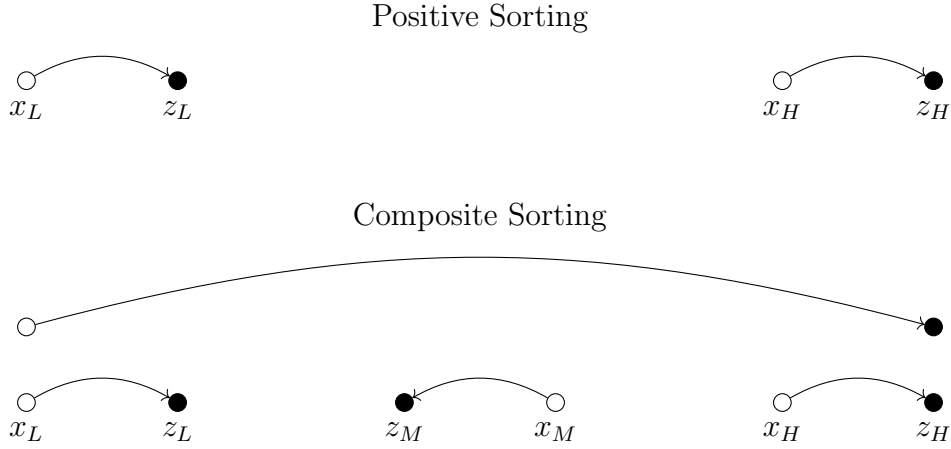


Figure 2: An Example of Composite Sorting

Figure 2 explains composite sorting with four workers (circles) and four jobs (dots). In the top panel, workers are optimally assigned to jobs in their skill group which results in two positively sorted pairs:  $(x_L, z_L)$  and  $(x_H, z_H)$ . Adding an identical low-skill worker and an identical high-skill job does not allow to reduce the initial mismatch and results in a pair  $(x_L, z_H)$ . The medium worker-job group  $\{x_M, z_M\}$  can be used to rematch the low-skill worker and the high-skill job with the medium-skill job and worker forming pairs  $(x_L, z_M)$  and  $(x_M, z_H)$ , but such rematching is not optimal due to the concavity of the costs. In the bottom panel, we hence also have two negatively sorted pairs  $(x_L, z_H)$  and  $(x_M, z_M)$  which delivers composite sorting (observe that worker  $x_L$  is assigned to distant job types; not to  $z_M$ ).

A layer contains all the workers and jobs in a particular slice of this measure of underqualification. An assignment in each layer optimally pairs workers and jobs within a given slice, regardless of all workers and jobs in other layers. The full optimal assignment combines the independent optimal assignments in each layer. Second, we characterize an optimal assignment within a given layer by formulating a Bellman equation. By leveraging the property of no intersecting pairs, it is straightforward to construct this recursive characterization.

The main result of our paper is composite sorting – the optimal assignment sorts identical workers into different jobs, some positively and some negatively. The intuition for composite sorting can be described through an example. Suppose a low worker-job group  $\{x_L, z_L\}$  is far from a high worker-job group  $\{x_H, z_H\}$  as in the top panel of Figure 2. When there is a significant gap between the groups, workers are optimally assigned to jobs within their skill groups resulting in two pairs  $(x_L, z_L)$  and  $(x_H, z_H)$ . This results in two small mismatches and positive sorting. Suppose one identical low-skill worker  $x_L$  and one identical high-skill job  $z_H$  as well a medium-skill worker-job group  $\{x_M, z_M\}$  are added as in the bottom panel of Figure 2. Adding an identical low-skill worker and an identical high-skill job does not allow to reduce the initial mismatched pairs  $(x_L, z_L)$  and  $(x_H, z_H)$ , and so the added low-skill worker is assigned to the added high-skill job forming a pair  $(x_L, z_H)$ . The medium worker-job group  $\{x_M, z_M\}$

can, in principle, be used to rematch the added low-skill worker and the added high-skill job with the medium-skill job and worker forming pairs  $(x_L, z_M)$  and  $(x_M, z_H)$ . However, such rematching is not optimal, because due to concavity of the costs, two medium mismatches  $(x_L, z_M)$  and  $(x_M, z_H)$  are worse than one small mismatch within the medium-skill group  $(x_M, z_M)$  and one large mismatch for  $(x_L, z_H)$ . As a result, the medium-skill group is paired together and this results in the negative sorting for the added group. We emphasize that the optimal assignment sorts the same worker type to very different jobs, with some positive and some negative sorting. The low-skill worker  $x_L$  is paired positively with the low-skill job  $z_L$  (as a part of the sorting  $(x_L, z_L)$  and  $(x_H, z_H)$ ) but the same worker type is also paired negatively to the distant high-skill job  $z_H$  as a part of the sorting  $((x_L, z_H)$  and  $(x_M, z_M))$ . Thus, the optimal full assignment is not one-to-one because the same worker type is assigned to different jobs as a part of simultaneously positive and negative sorting. The problem thus delivers composite sorting.

For general distributions, composite sorting may be very rich with very distinct worker types being assigned to the same job and may exhibit various local and global intervals of positive and negative sorting. Our theory offers a complete characterization of composite sorting.

We determine equilibrium wages and firm values by characterizing the solution to the dual planning problem. Our contribution here is to characterize the dual for mismatched workers and jobs by demonstrating that the shadow cost of mismatch exhibits a local hierarchical structure. For a given skill group, the relative wages are determined locally, that is, they depend only on information within the narrowly defined local skill group and do not depend on any other groups or skills. The hierarchical structure aggregates the local relative wages to the wages for the larger groups preserving the relative wages implied by the local structure. Finally, we extend the construction of the wages and the firm values to perfectly paired workers.

Having characterized the equilibrium theoretically, we quantitatively illustrate the theory. An important implication of our results is frictionless wage dispersion within jobs. We investigate to what extent our model can generate qualitative and quantitative patterns of wage dispersion within 497 occupations in the American Community Survey between 2010 and 2017.

We find that the model can deliver sizable wage dispersion within occupations with high and low mean wages. Composite sorting can account for 46 percent of wage dispersion within jobs at the bottom of the distribution, and 58 percent at the top. At the top, a high-skill surgeon performs the same job as a lower-skill surgeon assisted by a robotic surgical system. This leads to a wide range of salaries within the occupation, encompassing both low- and high-skill surgeons. A similar pattern holds at the bottom of the distribution. The model can account for 19 percent of the wage dispersion in the middle

of the distribution. The main reason is that the log-wage profile is relatively flat in this range and, even though there may be substantial skill dispersion within a particular occupation, it does not give rise to substantial wage dispersion. Overall, composite sorting accounts for 32 percent of the wage dispersion in the sample. In contrast, one-to-one sorting models (such as models with positive or negative sorting arising from submodular or supermodular costs) would deliver no wage dispersion within jobs.

**Literature.** We now briefly discuss additional relevant literature. [Kremer and Maskin \(1996\)](#) study sorting of heterogeneous workers with a form of technology choice where there is selection in managerial and worker roles. [Porzio \(2017\)](#) uses optimal transport to extend their framework with a continuous technological investment choice for each worker and manager which scales the strength of the complementary between them. [Anderson and Smith \(2022\)](#) depart from focusing on the conditions for assortative matching by developing comparative statics predictions for the sorting patterns without having to solve for optimal sorting. [Fagereng, Guiso, and Pistaferri \(2022\)](#) empirically document the patterns on non-assortative pairing in wealth. Another recent development in the sorting literature is to study assignment models with multiple agents combined together ([Kremer \(1993\)](#), [Chiappori, McCann, and Pass \(2017\)](#), [Chade and Eeckhout \(2018\)](#), and [Boerma, Tsyvinski, and Zimin \(2021\)](#)).

Another general alternative that results in imperfect assortative matching is the search and matching literature (for example, [Shimer and Smith \(2000\)](#), [Postel-Vinay and Robin \(2002\)](#), [Cahuc, Postel-Vinay, and Robin \(2006\)](#), [Eeckhout and Kircher \(2010\)](#), [Lise and Robin \(2017\)](#), [Bagger and Lentz \(2019\)](#)). Specifically, this approach generates wage dispersion within jobs due to the search frictions. Our work instead generates wage dispersion solely due to sorting in a frictionless environment.

Our paper uses results from the optimal transport literature (see [Galichon \(2018\)](#) for a comprehensive overview of applications of optimal transport theory to solve economic problems) and, more specifically, from the literature with concave distance costs started by [Gangbo and McCann \(1996\)](#) and [McCann \(1999\)](#). In the optimal transport literature the idea of the perfect pairs is referred to as “mass stays in place if it can” ([Gangbo and McCann \(1996\)](#); [Villani \(2003\)](#)). [Villani \(2009\)](#) states that the non-intersecting rule first appears in [Monge \(1781\)](#). This non-intersecting property is central to the literature on optimal transport with concave distance costs as well as to algorithmic sorting problems with distance costs ([Aggarwal, Barnoy, Khuller, Kravets, and Schieber \(1995\)](#); [Werman, Peleg, Melter, and Kong \(1986\)](#)). [Aggarwal, Barnoy, Khuller, Kravets, and Schieber \(1995\)](#) proposed the first combinatorial algorithm that can be used to solve for an optimal assignment within a layer when the cost of mismatch is linear. The Bellman equation in our papers adopts a recursive algorithm developed by [Nechaev, Sobolevski,](#)

and Valba (2013), designed to model statistical properties of polymer chains. The idea of layering was introduced in Delon, Salomon, and Sobolevski (2012a). Their work also develops a more computationally efficient version of the Bellman equation building on Aggarwal, Barnoy, Khuller, Kravets, and Schieber (1995). We provide a novel, concise proof of this more efficient Bellman equation. Our construction of the dual solution and characterization of its local hierarchical structure are also mathematical contributions within the optimal transport literature.

## 2 Model

We study an environment in which workers with heterogeneous skills sort into heterogeneous jobs, where the mismatch between their skills and the job difficulty leads to output losses. Firms reduce the extent to which mismatch penalizes surplus by reducing variable mismatch costs with fixed cost investments.

### 2.1 Environment

The economy is populated by risk-neutral workers and jobs. The workers differ in their skills which are indexed by a real number  $x \in X \subseteq \mathbb{R}$ . The set of worker skills  $X$  contains a finite number  $n$  of types  $x_1 < x_2 < \dots < x_n$ . Workers are distributed according to the cumulative distribution function  $F(x)$ .

Jobs differ in their difficulty which is indexed by a real number  $z \in Z \subseteq \mathbb{R}$ . The set of job intensities  $Z$  contains a finite number  $m$  of types  $z_1 < z_2 < \dots < z_m$ . Jobs are distributed according to cumulative distribution function  $G(z)$ . Both the worker and the job distributions have finite support.

**Technology.** Firms produce a single good. Production requires one worker for each job. A firm with a job  $z$  that employs a worker with skill  $x$  produces output according to

$$p(x, z) = g(x) + h(z) - \gamma_p \max(z - x, 0), \quad (1)$$

with  $\gamma_p \geq 0$ . There are three terms in this technology specification. The first term  $h(z)$  where  $h'(z) > 0$  reflects that a more difficult job  $z$  produces more output and is thus more valuable, independent of the worker that fulfills the job. The second term  $g(x)$  with  $g'(x) > 0$  reflects that a more skilled worker contributes more to production, independent of the job performed. The third term  $\gamma_p \max(z - x, 0)$  captures the idea that a worker with a skill  $x$  that is lower than the job demand  $z$  causes a loss of output. It is costly to have workers perform tasks for which they have limited talent.

**Preferences.** The utility of a worker  $x$  working a job  $z$  earning wage  $w$  is given by  $w - \kappa(x, z)$ , where  $\kappa(x, z)$  is the disutility from work which depends on the worker-job pair. Following Lise and Postel-Vinay

(2020), disutility from work is positive when the worker is overqualified for the job. We assume the utility cost of being overqualified is given by  $\kappa(x, z) = \gamma_\kappa \max(x - z, 0)$ , where the parameter  $\gamma_\kappa \geq 0$ . The specification of the disutility cost function parallels the one we use to capture the output loss due to inadequacy of worker skills compared to the job difficulties in the production function (1). The difference between these specifications is governed by the parameters  $\gamma_p$  and  $\gamma_\kappa$ . The utility cost of being overqualified introduces an excess of skills to cause a loss of joint surplus without causing a loss of output. In equilibrium, overqualified workers are compensated for their utility cost, and therefore are paid more in a given job than workers whose skills exactly match the job's difficulties (as in Rosen (1986)).

**Technology Choice.** Mismatch is the distance between worker  $x$  and job  $z$ . The environment so far assumes that the costs of mismatch are exogenous. We now endogenize the cost of mismatch. A firm can reduce the exogenous cost of mismatch by making a fixed investment. In case a worker is underqualified, the firm can make technology investments to reduce the cost of mismatch in output. When workers are overqualified they experience a disutility cost which can be reduced by providing amenities. The main insight of Stigler (1939) and Laffont and Tirole (1986, 1991) for our setting is that the technology selection results in an output function with a concave cost of mismatch. We proceed to formally define these technology choices.

We model technology choice as a firm making fixed investments (for example, by purchasing better equipment) to reduce variable costs associated with mistakes and delays caused by underqualified workers. Specifically, firms choose the variable cost  $\gamma_p$  of production mismatch, which comes at an associated fixed cost  $\Psi_p(\gamma_p)$ . By decreasing the variable costs  $\gamma_p$ , the firm increases its fixed costs, or  $\Psi'_p < 0$ , where  $\Psi''_p > 0$ . To highlight the key forces that arise by introducing technology selection into the assignment model, we use the specification  $\Psi_p(\gamma_p) = \frac{B_p}{\eta_p} \gamma_p^{-\eta_p}$ , where parameters  $B_p$  and  $\eta_p$  are both strictly positive.<sup>3</sup> The effective output created by worker  $x$  on job  $z > x$  is then:

$$y(x, z) = \max_{\gamma_p \geq 0} g(x) + h(z) - \gamma_p(z - x) - \frac{B_p}{\eta_p} \gamma_p^{-\eta_p}. \quad (2)$$

A firm makes a technology investment when a worker is underqualified to fulfill the job, when  $x < z$ . The extent of investment increases in the distance between the worker skill and the job difficulty. Using the first-order condition, the optimal technology choice is  $\gamma_p = \left(\frac{z-x}{B_p}\right)^{-\frac{1}{1+\eta_p}}$ . Firms choose a low variable cost of production mismatch if the worker is less qualified, that is, when the distance between skills and demands ( $z - x$ ) is large, or if the cost of technology investment  $B_p$  is low.

---

<sup>3</sup>We generalize our approach to any convex cost function in the Technical Appendix.



The effective output of worker  $x$  in job  $z$  is, using the technology investment decision, given by:

$$y(x, z) = g(x) + h(z) - \rho_p(z - x)^{\zeta_p} \quad (3)$$

for underqualified workers  $x < z$ , with implied parameters  $\zeta_p = \frac{\eta_p}{1+\eta_p} \in (0, 1)$  and  $\rho_p = B_p^{\frac{1}{1+\eta_p}} / \zeta_p$ . The distinctive feature of this function is that the effective output is concave in the distance between worker skill and job difficulty. The technology choice transforms a production function with linear mismatch costs into an effective output function with strictly concave mismatch costs.

Firms can similarly reduce the extent to which overqualification penalizes worker utility by providing amenities. We model amenities analogously to technology choice.<sup>4</sup> If workers are underqualified,  $x < z$ , their employers want to invest in technology so that the costs of having underqualified employees are reduced. Similarly, when workers are overqualified,  $x > z$ , their employers want to invest in amenities to lower the disutility from working. Firms choose the variable disutility cost due to mismatch  $\gamma_\kappa$  at an associate fixed cost  $\Psi_\kappa(\gamma_\kappa) = \frac{B_\kappa}{\eta_\kappa} \gamma_\kappa^{-\eta_\kappa}$ . The optimal provision of amenities is thus characterized by  $\gamma_\kappa = \left(\frac{x-z}{B_\kappa}\right)^{-\frac{1}{1+\eta_\kappa}}$ . Amenities increase with worker skills and decrease with the cost of providing amenities,  $B_\psi$ . As a result, the effective output of overqualified workers is  $y(x, z) = g(x) + h(z) - \rho_\kappa(z - x)^{\zeta_\kappa}$ .

**Effective Output.** We can write the effective output of overqualified and underqualified workers as:

$$y(x, z) = g(x) + h(z) - \begin{cases} \rho_p(z - x)^{\zeta_p} & \text{if } z \geq x \\ \rho_\kappa(x - z)^{\zeta_\kappa} & \text{if } z < x, \end{cases} \quad (4)$$

where  $\zeta_p$  and  $\zeta_\kappa$  are strictly positive and strictly less than one. We use effective output function (4) to define the cost of mismatch between worker  $x$  and job  $z$  as  $c(x, z) = g(x) + h(z) - y(x, z)$ , or

$$c(x, z) = \begin{cases} \rho_p(z - x)^{\zeta_p} & \text{if } z \geq x \\ \rho_\kappa(x - z)^{\zeta_\kappa} & \text{if } z < x, \end{cases} \quad (5)$$

that is, maximal output of worker  $x$  and job  $z$  minus effective output  $y(x, z)$ . Thus, the sections above give a microfoundation for a mismatch cost function that is concave in the discrepancy between worker  $x$  and job  $z$ . We plot the mismatch cost function in Figure 3.

The key observation is that the effective output function is neither supermodular nor submodular. For both  $z > x$  and  $z < x$ , the cross-derivatives of the output function are negative. If the cross-derivative is negative for all  $(x, z)$  the production function is submodular. To see that our production function is

---

<sup>4</sup>Alternative frameworks that incorporate the decision to provide amenities are presented in [Hwang, Mortensen, and Reed \(1998\)](#), [Lang and Majumdar \(2004\)](#), and [Morchio and Moser \(2021\)](#).

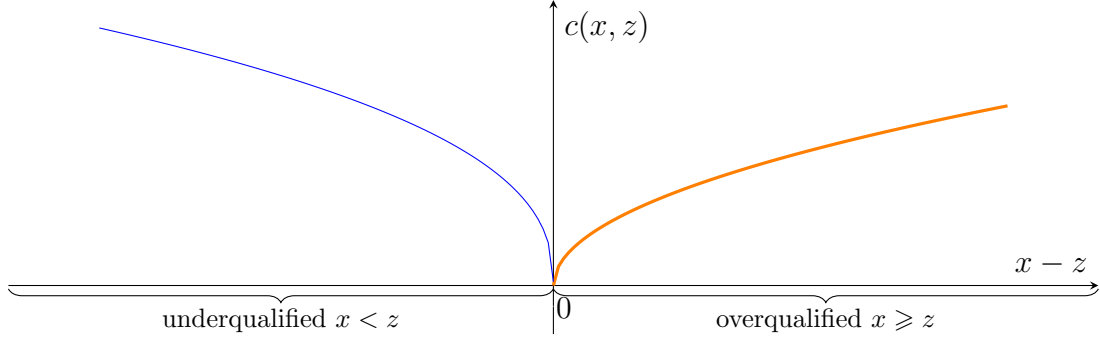


Figure 3: The Cost of Mismatch

Figure 3 illustrates the cost of mismatch  $c(x, z)$ , which is concave in the discrepancy between the worker skill  $x$  and the job difficulty  $z$ . The result of the firm technology decision is that the mismatch cost function is concave in the extent of underqualification ( $z - x$ ). The result of investment in amenities is that the mismatch cost function is concave in the extent of overqualification ( $x - z$ ).

not submodular, consider pairs  $(a, b)$  and  $(b, a)$  such that  $b > a$ . Pairing worker  $a$  positively with job  $a$  and worker  $b$  with job  $b$  gives zero mismatch. The original configuration  $(a, b)$  and  $(b, a)$ , however, has mismatch and hence has lower output than the combination of  $(a, a)$  and  $(b, b)$ . Submodularity, however, requires that the combined output of  $(a, b)$  and  $(b, a)$  is higher than the combined output of  $(a, a)$  and  $(b, b)$ . Thus, the output function is not submodular. In sum, even though the cross-derivative is negative for all  $z > x$  and  $z < x$ , the output function is neither supermodular nor submodular.

**Assignment.** An assignment pairs workers and jobs. Formally, an assignment is a probability measure  $\pi$  over workers and jobs. Given a worker distribution  $F$  and a distribution of jobs  $G$ , the set of feasible assignment functions is  $\Pi := \Pi(F, G)$  which is the set of probability measures on the product space  $X \times Z$  such that the marginal distributions of  $\pi$  onto  $X$  and  $Z$  are  $F$  and  $G$ . For an assignment  $\pi$ , we denote the support of this assignment by  $\Gamma_\pi = \{(x, z) : \pi(\{(x, z)\}) > 0\} \subseteq \mathbb{R}^2$ . Feasibility of an assignment function is equivalent to labor market clearing; that is, all workers and jobs are sorted.

## 2.2 Equilibrium

To define an equilibrium for our economy, we first specify the firm and worker problem. A firm with job  $z$  employs worker  $x$  to maximize profits taking the wage schedule  $w$  as given. The firm problem is:

$$v(z) = \max_{x \in X} y(x, z) - w(x). \quad (6)$$

Taking firm compensation  $v$  as given, worker  $x$  chooses to fulfill job  $z$  to maximize their wage income:

$$w(x) = \max_{z \in Z} y(x, z) - v(z). \quad (7)$$

An equilibrium is a wage schedule  $w$ , firm value function  $v$ , and feasible assignment function  $\pi$ , such that firms solve their profit maximization problem (6), workers solve the worker problem (7), and a feasibility constraint is satisfied,  $\int y(x, z) d\pi = \int w(x) dF + \int v(z) dG$ , which states that the total quantity of output produced,  $\int y(x, z) d\pi$ , equals the total quantity of output distributed to workers and jobs.

## 2.3 Planning Problem

We solve a planning problem to characterize the equilibrium assignment. The planning problem is to choose an assignment to maximize aggregate output:

$$\max_{\pi \in \Pi} \int y(x, z) d\pi \tag{8}$$

The planning problem to maximize output is equivalent, in terms of choosing an optimal assignment, to a planning problem that minimizes the cost of mismatch. When we characterize the optimal assignment, we can thus equivalently represent the planning problem as:

$$\min_{\pi \in \Pi} \int c(x, z) d\pi, \tag{9}$$

where  $c(x, z)$  represents the cost of mismatch (5). This is an optimal transport problem in the formulation of Kantorovich (1942) with the cost function that is neither supermodular nor submodular.

**Dual Problem.** The solution to the planning problem is an assignment, which we argue is an equilibrium assignment. To obtain the equilibrium wage schedule and firm value function, we solve a dual problem. The dual formulation for the assignment planning problem is to choose functions  $w$  and  $v$  that solve:

$$\min \int w(x) dF + \int v(z) dG, \tag{10}$$

subject to the constraint  $w(x) + v(z) \geq y(x, z)$  for any  $(x, z) \in X \times Z$ . The Monge-Kantorovich duality states that the values from (8) and (10) are the same, or  $\max_{\pi \in \Pi} \int y(x, z) d\pi = \min \int w(x) dF + \int v(z) dG$ .

In Appendix A.1, we prove the following relation between the primal and dual solutions.

**Lemma 1.** Suppose we have some assignment  $\pi \in \Pi$  and some functions  $w$  and  $v$  such that  $w(x) + v(z) \geq y(x, z)$  for any  $(x, z) \in X \times Z$  and that  $w(x) + v(z) = y(x, z)$  for any  $(x, z) \in \Gamma_\pi$ , then assignment  $\pi$  is an optimal assignment and  $(w, v)$  is an optimal dual pair.

## 3 Optimal Sorting

We now characterize an optimal assignment of workers to jobs. A necessary condition for an optimal assignment is that aggregate output does not increase by a bilateral exchange of workers between jobs.

For any two pairs in an optimal assignment  $(x, z)$  and  $(\hat{x}, \hat{z})$ , the following inequality should be satisfied:<sup>5</sup>

$$y(x, z) + y(\hat{x}, \hat{z}) \geq y(x, \hat{z}) + y(\hat{x}, z). \quad (11)$$

We next use this optimality condition to demonstrate the main features of an optimal assignment.

**Maximal Number of Perfect Pairs.** An optimal assignment maximizes the number of pairs that are perfectly sorted, i.e., the number of pairs with zero cost of mismatch between workers and jobs, or  $x = z$ . In other words, whenever feasible, an optimal assignment assigns worker  $x$  to job  $z = x$ .

The rationale for why an optimal assignment maximizes the number of perfect pairs is as follows. When mismatch costs are concave, it is preferable to have a pair with small mismatch and a pair with significant mismatch than to have two pairs with moderate mismatch. A perfect pair is a demonstration of this concept since it has zero mismatch. Specifically, let an optimal assignment contain pairs  $(x, z)$  and  $(x', z')$  but  $x' = z$  where  $x$  and  $z'$  do not equal  $z$ . Suppose  $x < x' = z < z'$ , that is, the value  $z = x'$  is in between the values  $x$  and  $z$ . The original cost of mismatch is two medium-size mismatches in pairs  $(x, z)$  and  $(x', z')$ . Consider reshuffling to form a perfect pair  $(x', z)$  and the remaining pairing  $(x, z')$ . The cost of the mismatch of the reshuffled pairs is that of a large mismatch  $(x, z')$  and a zero mismatch for the perfect pair  $(x', z)$ . By concavity of the costs, this is lower than two medium-size mismatches.<sup>6</sup>

The output loss due to mismatch is strictly reduced by assigning worker  $x$  to job  $z'$  and by perfectly assigning worker  $x'$  to job  $z$ , which contradicts the optimality condition (11). We summarize this idea in Lemma 2.

**Lemma 2.** *Maximal Perfect Pairs.* Let  $F \wedge G$  denote the common component of the worker distribution  $F$  and the job distribution  $G$ . Any optimal assignment  $\pi$  between workers consists of perfect pairings on the support of  $F \wedge G$  and an optimal assignment between workers  $F - F \wedge G$  and jobs  $G - F \wedge G$ .

We present a proof in Appendix A.2.

Lemma 2 shows that workers and jobs which are part of the common component are positively sorted. Since workers in the common component are perfectly paired, better workers are paired with better jobs. Lemma 2 implies that in analyzing the sorting problem between the remaining workers and jobs we can restrict attention to assignments between worker and job distributions where the common parts of the distributions are removed. For brevity, we label the remaining worker distribution  $F$  and the remaining job distribution  $G$ .

---

<sup>5</sup>This is an instance of the notion of cyclical monotonicity (see, e.g., Definition 1.36 in Santambrogio (2015)).

<sup>6</sup>A strictly convex cost of mismatch implies that the optimal sorting is positive which, in general, conflicts with maximal perfect pairing.



Figure 4: Intersecting and Non-Intersecting Pairs

Figure 4 illustrates intersecting and non-intersecting pairs. The left panel shows an example of intersecting pairs as the arcs of pairs  $(x, z)$  and  $(x', z')$  intersect. The right panel shows an example of non-intersecting pairs as the arcs corresponding to pairs  $(x, z)$  and  $(x', z')$  do not intersect. Lemma 3 shows that under an optimal assignment arcs never intersect.


**No Intersecting Pairs.** The second feature of an optimal assignment is that pairings between workers and jobs do not intersect. Consider pairs  $(x, z)$  and  $(x', z')$  and illustrate their pairings by arcs. In Figure 4 we show intersecting and non-intersecting pairs. The left panel shows an example of intersecting pairs as the arcs of pairs  $(x, z)$  and  $(x', z')$  intersect. The right panel shows an example of non-intersecting pairs as the arcs corresponding to pairs  $(x, z)$  and  $(x', z')$  do not intersect. We say that the arcs  $(x, z)$  and  $(x', z')$  do not intersect if and only if the intervals  $(x, z)$  and  $(x', z')$  are either disjoint or one interval is a subset of the other interval.<sup>7</sup> Lemma 3 establishes that under an optimal assignment arcs never intersect.

**Lemma 3. No Intersecting Pairs.** Let  $\pi$  be an optimal assignment. For any two pairs  $(x, z)$  and  $(x', z')$  in the support  $\Gamma_\pi$ , their arcs do not intersect.

The proof is in Appendix A.3 and we now outline the main idea.

We prove Lemma 3 by considering the two unique configurations with intersecting pairs.<sup>8</sup> The first configuration is  $x < z' < z < x'$ . Since the mismatch cost increases in the distance between worker and job, an improvement is to instead pair the closer points  $(x, z')$  and  $(x', z)$  as it reduces the mismatch cost for each worker and hence the total mismatch costs. The second configuration is  $x < x' < z < z'$ . In this case, we utilize the concavity of the cost of mismatch to obtain a contradiction to the optimality condition (11). We reshuffle so that we have one large mismatch  $(x, z')$  and one small mismatch  $(x', z)$ . By concavity, the cost of this is smaller than two medium-size mismatches. The output loss due to mismatch strictly reduces by assigning worker  $x$  to job  $z'$  and worker  $x'$  to job  $z$ , which contradicts the optimality condition (11).

<sup>7</sup>When referring to an interval  $(x, z)$ , we do not require that the worker skills and job difficulties are ordered: we mean the set of real numbers lying between  $z$  and  $x$  on the real line.

<sup>8</sup>There are six distinct orderings of workers  $x$ 's (white circles) and jobs  $z$ 's (black dots) to consider, which can be represented as: . The first four configurations do not contain any intersections. The final two scenarios are discussed in the main text.

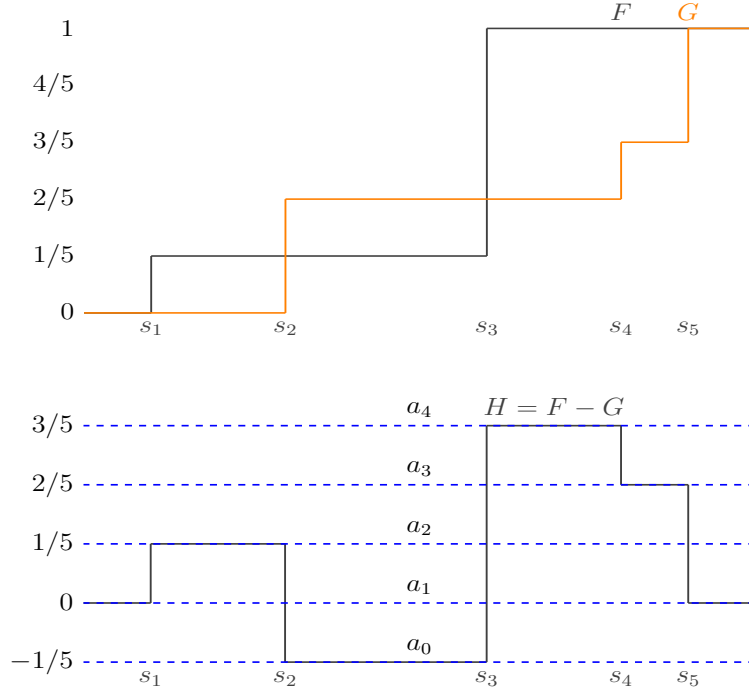


Figure 5: Measure of Underqualification

Figure 5 illustrates the construction of the measure of underqualification given an example of the worker distribution  $F$  and the job distribution  $G$  in the top panel. Distribution  $F$  has one worker at skill level  $s_1$  and four workers at skill level  $s_3$ . Distribution  $G$  has two jobs each at skill level  $s_2$  and  $s_5$  and a single job at  $s_4$ . The measure of underqualification  $H$  measures the extent to which workers up to skill level  $s$  outweigh the jobs requiring skill levels up to  $s$  as  $H = F - G$ .

**Layering.** The observation that an optimal assignment does not contain intersections enables us to next show that the optimal assignment problem can be decomposed into a series of independent problems.

To see how the assignment problem is decomposed into different layers, let  $H$  be the difference between the cumulative distribution function of workers  $F$  and the cumulative distribution function of jobs  $G$ . The measure of underqualification  $H$  measures the extent to which workers up to skill level  $s$  outweigh the jobs requiring skill levels up to  $s$  since  $H(s) := F(s) - G(s)$ .

Figure 5 illustrates the measure of underqualification. The top panel gives an example of the worker distribution  $F$  and the job distribution  $G$ . Distribution  $F$  has one worker at skill level  $s_1$  and four workers at skill level  $s_3$ . Distribution  $G$  has two jobs each at skill level  $s_2$  and  $s_5$  and a single job at  $s_4$ . The underqualification measure between  $s_1$  and  $s_2$  is  $H = \frac{1}{5}$  as there is one worker at  $s_1$  and no jobs less than  $s_2$ . Between  $s_2$  and  $s_3$ , the underqualification measure is  $H = -\frac{1}{5}$  as there are two jobs and only one worker compared to level  $s_3$ . The similar intuition holds for the other points yielding the measure of underqualification  $H \in \{a_v\}_{v=0}^4 = \{-\frac{1}{5}, 0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}\}$  at skill levels  $\{s_i\}_{i=1}^5$ . Generally, an underqualification measure has jumps at a finite number of skill levels  $\{s_i\}$  and takes a finite number of

values in  $\{a_0, a_1, \dots, a_L\}$  if the distributions of workers and jobs are discrete.

To illustrate the construction of layers we plot the blue dashed lines which indicate the different values of the measure of underqualification in the bottom panel of Figure 4. Layers are thus confined between two successive blue lines. For example, the top layer (between  $a_3 = \frac{2}{5}$  and  $a_4 = \frac{3}{5}$ ) contains a worker at skill level  $s_3$  and a job at skill level  $s_4$ . Similarly, the layer between  $a_1 = 0$  and  $a_2 = \frac{1}{5}$  consists of one worker each at skill levels  $s_1$  and  $s_3$ , and one job each at skill levels  $s_2$  and  $s_5$ .

We observe that each layer consists of an alternating configuration of workers and jobs, that is, either  $x_1 < z_1 < x_2 < z_2 < \dots < x_n < z_n$  or  $z_1 < x_1 < z_2 < x_2 < \dots < z_n < x_n$ . We define an alternating assignment problem as an assignment problem between  $n$  workers and  $n$  jobs. Workers and jobs are arranged in increasing order, and alternating such that every worker skill level is followed by a job difficulty level, except for the last one. Let  $F_\ell$  and  $G_\ell$  be the measure of the workers and the jobs in each layer.<sup>9</sup>

After providing a decomposition into layers with alternating configurations, Lemma 4 shows how to solve the full assignment problem using the solutions to the assignment problem within each layer.

**Lemma 4.** *Layering.* Let  $\pi_\ell$  be an optimal assignment between the worker distribution  $F_\ell$  and the job distribution  $G_\ell$  for the layer  $\ell \in \{1, \dots, L\}$ . Then, an optimal assignment between workers  $F$  and jobs  $G$  is their sum,  $\pi := \sum_\ell \pi_\ell$ .

This result is proved in Appendix A.4.

Lemma 4 shows that an optimal assignment between workers and jobs is equal to the sum of optimal assignments in each layer. This observation breaks the original problem down into  $L$  independent problems. In each of the layers, there is an alternating configuration of workers and jobs. We use this observation to characterize an optimal assignment for our full problem by aggregating the solutions to the assignment problems for each layer.

The intuition for Lemma 4 is illustrated in Figure 6 for the case where all points have equal weight. Let us consider two points  $x_0$  and  $z_0$  that are matched under an optimal assignment depicted by a solid arc. There is the same number of workers and jobs between these two points. Suppose that instead there exists at least one worker or a job that needs to be paired with a job or worker outside of this range. In the figure, worker  $x_1$  does not have a pair in the interval between  $x_0$  and  $z_0$  and has to be connected by a dashed arc. This dashed arc intersects the solid arc  $(x_0, z_0)$  which contradicts the no-intersecting pairs principle in Lemma 3.

---

<sup>9</sup>The formal definition is given with the proof in Appendix A.4.

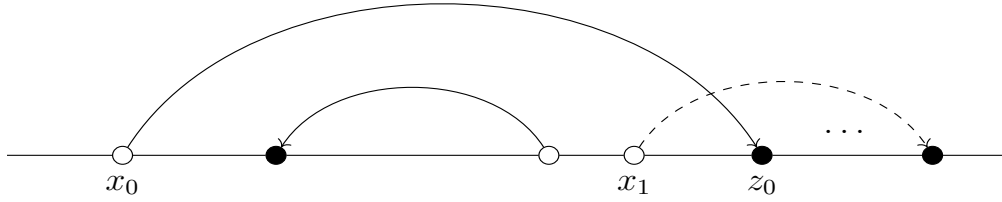


Figure 6: Layering

Figure 6 shows the intuition for Lemma 4 when all points have equal weights. Consider two points  $x_0$  and  $z_0$  that are paired under an optimal assignment as depicted by a solid arc. There is the same number of workers and jobs between these two points. Suppose that instead there is one worker or a job that needs to be paired with a job or worker outside of this range. In the figure, worker  $x_1$  cannot be paired in the interval between  $x_0$  and  $z_0$  and has to be connected by the dashed arc. This dashed pair intersects the solid pair  $(x_0, z_0)$  which contradicts the no-intersecting pairs principle.

**Sorting Within a Layer.** By leveraging the property of no intersecting pairs, it becomes straightforward to construct a recursive characterization for an optimal assignment within a particular layer. In particular, for each layer, we adopt the recursive algorithm developed by [Nechaev, Sobolevski, and Valba \(2013\)](#), designed to model statistical properties of polymer chains. The complete optimal solution is then obtained by combining the optimal solutions of each individual layer.

The optimal assignment problem for a given layer is an alternating assignment problem. An optimal assignment within a layer matches one worker with precisely one job.<sup>10</sup> For notational convenience, we order workers and jobs within each layer by their skill levels. Let there be  $n$  workers and  $n$  jobs in a given layer, and we denote the skill levels by  $s_1 < s_2 < \dots < s_{2n-1} < s_{2n}$ .

We employ a Bellman equation to calculate the minimum total cost of mismatch. Rather than using a Bellman equation to solve a dynamic program, we apply the Bellman equation to solve a static assignment program. The dynamic component of the Bellman equation is that we consider assignment problems with an increasing number of skill levels. We start by solving all assignment problems between two consecutive elements – the assignment problem between one worker and one job. That is, we consider assignments between skill levels  $s_i$  and  $s_{i+1}$ , for each  $i$ . Using the solutions from the previous step, we proceed to solve all assignment problems between four consecutive elements (two workers and two jobs). That is, we solve for an optimal assignment between  $s_i, s_{i+1}, s_{i+2}, s_{i+3}$ , for each  $i$ . Then we consider all optimal assignment problems between three workers and three jobs. We continue increasing the size of the optimal assignment problem until we solve the assignment problem between  $n$  workers and  $n$  jobs, which is the

<sup>10</sup>An equivalent form of Birkhoff’s Theorem ([Birkhoff, 1946](#)) states that for this setting, the set of Kantorovich transport plans is precisely that of mixtures of one-to-one transport maps. Since the cost is linear with respect to mixtures, there exists a one-to-one optimal assignment.



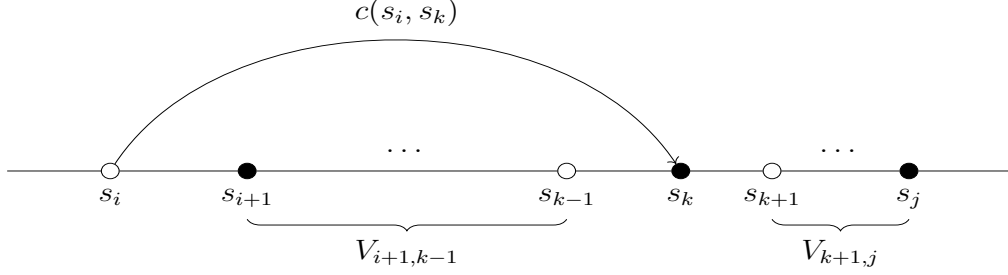


Figure 7: Bellman Equation

Figure 7 illustrates the Bellman equation for an optimal sorting. Consider an optimal assignment problem between workers and jobs with skill levels in  $[s_i, s_j]$ . We can pair the leftmost skill  $s_i$  with any skill  $s_k$  such that  $k - i$  is odd, as illustrated by the pair  $(s_i, s_k)$ . Upon pairing  $s_i$  with  $s_k$ , the planner remains to optimally pair the other workers and jobs in  $[s_{i+1}, s_{k-1}]$ , and all workers and jobs with skill levels in  $[s_{k+1}, s_j]$ . There are no pairings between these two segments because this violates the property of no intersecting pairs.

original alternating assignment problem for the layer.

By the property of no intersecting pairs, the formulation of the Bellman equation directly arises. We denote the minimum cost of mismatch when sorting all workers and jobs with skill levels between  $s_i$  and  $s_j$  by the value  $V_{i,j}$ , where  $j > i$ . The difference  $j - i$  has to be odd so that there are equal numbers of workers and jobs between  $s_i$  and  $s_j$ . Considering an optimal assignment of workers and jobs with skill levels in  $[s_i, s_j]$ , the planner can pair the leftmost  $s_i$  with any  $s_k$  such that  $k - i$  is odd. Upon pairing  $s_i$  with  $s_k$ , the planner remains to optimally pair the other workers and jobs in  $[s_{i+1}, s_{k-1}]$ , and all workers and jobs with skill levels in  $[s_{k+1}, s_j]$ . There are no pairings between these two segments because this violates the property of no intersecting pairs. In this way, we can use results from previous steps of assignment and thus obtain a Bellman equation:

$$V_{i,j} = \min_{k \in \{i+1, i+3, \dots, j\}} (c(s_i, s_k) + V_{i+1, k-1} + V_{k+1, j}) \quad (12)$$

with boundary conditions  $V_{i+1, i} = 0$  for all  $i$ .<sup>11</sup>

Once the Bellman equation (12) is solved, we construct an optimal assignment. Starting from  $V_{1, 2n_\ell}$ , the optimal pairing of skill  $s_1$  is given by skill  $s_k$  that solves equation (12). Then two corresponding continuation values,  $V_{2, k-1}$  and  $V_{k+1, 2n_\ell}$ , are evaluated to determine optimal pairings for skill  $s_2$  and for

<sup>11</sup>The boundary conditions are invoked at either end of the choice interval. When  $k = i + 1$ , the minimum cost of mismatch is  $c(s_i, s_{i+1}) + V_{i+2, j}$ , the cost of pairing the first worker to the first job, together with optimally sorting all skill levels from  $s_{i+2}$  to  $s_j$ . When  $k = j$ , the minimum cost is  $c(s_i, s_j) + V_{j+1, j-1}$ , the cost of pairing the first worker to the last job, together with optimally matching all intermediate skill levels between  $s_{i+1}$  and  $s_{j-1}$ .

The boundary conditions are also used when  $j - i = 1$ , in which case there is a unique choice necessarily at the bounds of the choice set. This is an important step, as the Bellman equation (12) is solved by induction over all  $j - i$  odd, starting with  $j - i = 1$ .

skill  $s_{k+1}$ , respectively. This process of finding an optimal assignment continues recursively until the full assignment is constructed.

Our three-step approach to characterizing the optimal sorting is as follows. First, we maximize perfect pairs. This always leads to positive sorting between identical workers and jobs and enables us to withdraw them from further analysis. Second, we construct a measure of underqualification to decompose the assignment problem into a sequence of independent problems – layers. Third, we characterize an optimal assignment for each layer using the Bellman algorithm. The optimal allocation is thus given by the following proposition.

**Proposition 1.** An optimal assignment between workers  $F$  and jobs  $G$  sums an optimal assignment in each layer, where the optimal assignment in each layer attains  $V_{1,2n_\ell}$  in the Bellman equation (12).

In Appendix A.5, we introduce a novel, more concise proof of a computationally efficient version of the Bellman equation in Delon, Salomon, and Sobolevski (2012a). Our proof also applies in the case of asymmetric cost functions that we consider.

In Appendix A.6, we derive an additional explicit optimal sorting when the mismatch power indices  $\zeta_p$  and  $\zeta_\kappa$  are close to one. Specifically, we show that the optimal assignment within each layer is positive. The implication is that for mismatch power indices close to one, the solution can be directly constructed by constructing the measure of underqualification, and constructing the positive sorting by layer. Juillet (2020) shows that this layered positive assignment is the limit of some optimal assignments as  $\zeta \rightarrow 1^-$ . Our results proves the existence of a threshold beyond which the layered positive assignment is optimal for our environment and is thus applicable away from the limit.

## 4 Examples

Having described how to characterize the optimal sorting, we illustrate various elements of composite sorting through examples.

**Positive and Negative Sorting.** Our assignment problem, which features neither a supermodular nor a submodular output function, can deliver simultaneously positive and negative sorting. Importantly, the sorting pattern, rather than being dictated by the shape of the production function alone as in the classic optimal assignment problems, depends on the distributions of worker skills and job difficulties.

To see that an optimal assignment can feature both positively and negatively assorted pairings and sort identical workers into different jobs, consider an assignment problem with two workers and two jobs.

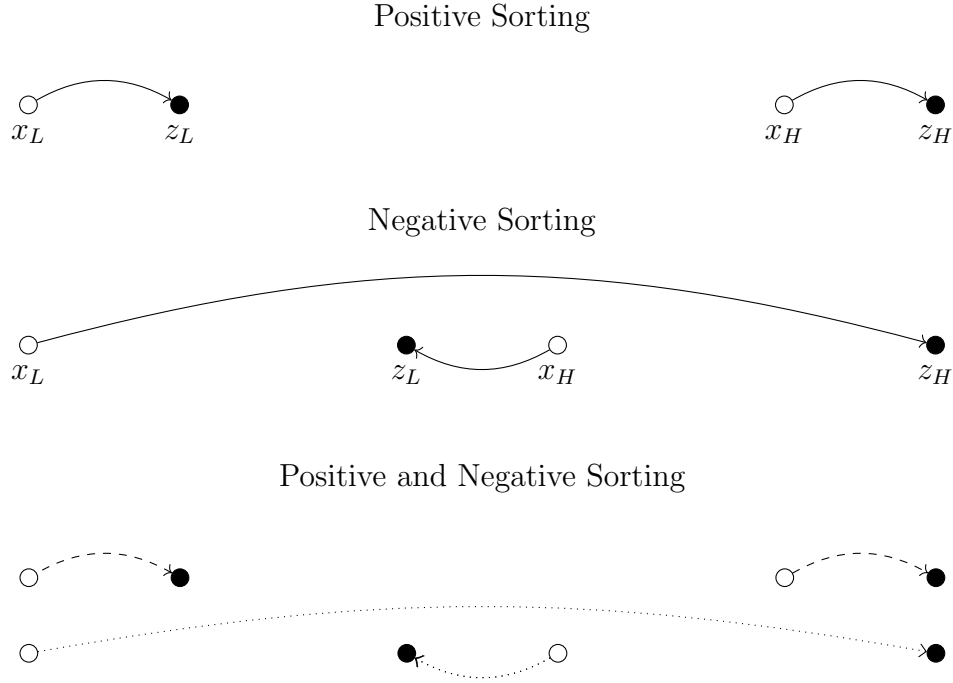


Figure 8: Positive and Negative Sorting

Figure 8 explains composite sorting for an example with two workers (indicated by the circles) and two jobs (indicated by the dots). The top panel illustrates the case of positive sorting, and the middle panel illustrates the case of negative sorting. The bottom panel shows that the equilibrium assignment between workers and jobs may simultaneously feature positive and negative sorting by combining the two prior cases.

Worker skills are given by  $x_L$  and  $x_H$  and jobs are given by  $z_L$  and  $z_H$  satisfying  $x_L < z_L < x_H < z_H$ . Let the absolute distance between  $x_i$  and  $z_j$  be  $d_{ij} := |x_i - z_j|$ .

To explain the presence of positive sorting, consider the following configuration of distances:  $d_{LL}^\zeta + d_{HH}^\zeta \leq d_{LH}^\zeta + d_{HL}^\zeta$  with the specific cost function  $c(x, z) = |x - z|^\zeta$  with  $\zeta \in (0, 1)$ . This case is presented in the top panel of Figure 8, where the workers are given by circles and the jobs are given by dots. The low-skill worker  $x_L$  and the low-skill job  $z_L$  as well as the high-skill worker  $x_H$  and the high-skill job  $z_H$  are very close to each other, while the skill gap between the low-skill job  $z_L$  and high-skill worker  $x_H$  is large. In this case, it is optimal for the low-skill worker to fulfill the low skill job and for the high-skill worker to fulfill the high-skill job and, hence, the optimal assignment is positive. The optimal assignment is visualized using an arrow between the low-skill worker and low-skill job, and an arrow between the high-skill worker and high-skill job. In other words, when there are worker-job skill groups that are far apart, it is optimal to sort within groups.

To explain the presence of negative sorting, consider the opposite configuration of distances where  $d_{LL}^\zeta + d_{HH}^\zeta > d_{LH}^\zeta + d_{HL}^\zeta$ . In this case, the high-skill worker and the low-skill job are close to each other,

while the distance between the low-skill worker and the low-skill job as well as the distance between the high-skill worker and the high-skill job is relatively large. Since the mismatch cost is concave in the distance between the worker skills and the job demands, it is optimal to pair the high-skill worker with the low-skill job since having one small mismatch and one large mismatch is better than having two medium-sized mismatches. This case is presented in the middle panel of Figure 8.

To see that an assignment between workers and jobs may simultaneously feature positive and negative sorting, we combine the two prior cases. This case is displayed in the bottom panel of Figure 8. The assignment problem features four workers (circles) and four jobs (dots). In this case, we first split the assignment problem into layers. The top layer contains the worker and job types that we considered in the top panel, implying that their optimal sorting is positive. Turning to the bottom layer, we see that the worker and job types in the bottom layer are given by the worker and job types that we considered in the middle panel, which implies that their optimal sorting is negative. By Lemma 4, a sum of optimal assignments of each of the two layers is optimal for the combined problem. In conclusion, the equilibrium assignment simultaneously features positive and negative sorting.

We emphasize the optimal assignment not only features positive and negative sorting simultaneously but also assigns the same worker type to distant job types, and thus delivers the two defining features of composite sorting. Consider the bottom panel in Figure 8. The low-skill worker is paired positively with the low-skill job (dashed arc), but the same worker type is also paired negatively with the very distant high-skill job (dashed dotted arc). While the optimal assignment within each layer is a one-to-one assignment, the full optimal assignment is not one-to-one because the same worker type is assigned to different jobs in different layers. Thus, sorting is composite – an identical worker type is assigned to multiple job types as part of both positive and negative pairings.<sup>12</sup>

**Reflecting Binomial Distributions.** To illustrate layering, we consider the following example. Let the skill levels be given by  $\{0, 1, 2, 3, 4\}$ , the numbers of workers with these skill levels are given by  $\{16, 32, 24, 8, 1\}$ , and the numbers of jobs with these skill levels are given by  $\{1, 8, 24, 32, 16\}$ . This corresponds to an economy where both worker skills and job difficulties are distributed following binomial distributions, where we denote the worker skill distribution by  $B(n, p)$  and the job distribution by  $B(n, 1 - p)$  with  $n = 4$  and  $p = \frac{1}{3}$ . In this case, the two binomial distributions are called reflecting as they have parameters  $p$  and  $1 - p$ .

We determine the solution to this sorting problem using the theory in Section 3. The first step is to

---

<sup>12</sup>In equilibrium, workers of a type  $x$  that is assigned to multiple job types are indifferent between these job types since the worker receives identical compensation  $w(x)$ .

maximize the number of perfect pairs. At skill level 0, there are 1 job and 16 workers. We perfectly pair this job with one of these 16 workers. Analogously, we perfectly pair  $\{8, 24, 8, 1\}$  workers at each skill level in  $\{1, 2, 3, 4\}$ . After maximizing perfect pairings, we are left with 15 workers of skill level 0 and 24 workers of skill level 1 together with 24 remaining jobs of skill level 3, and 15 jobs of skill level 4.

Next, we apply the layer analysis to the remaining workers and jobs. The difference between the worker and job distribution, the measure of underqualification, is presented in the top panel of Figure 9. In this case, we split the assignment problem into two layers. The top layer consists of 24 workers ( $39 - 15$ ) of skill level 1 and 24 jobs of skill level 3. The bottom layer consists of 15 workers of skill level 0 and 15 jobs of skill level 4. Within each layer there is only one feasible sorting. Across the layers, there is negative sorting, meaning that the best job is paired with the worst worker (bottom layer) and the worst job is paired with the top remaining worker (top layer). This negative sorting pattern within each layer is represented by the top two arcs in the bottom panel of Figure 9. The circular arcs in the bottom of the second panel of Figure 9 represent the optimal perfect pairing between workers and jobs to the extent that their distributions overlap.

Summarizing, the assignment problem with these reflecting binomial distributions simultaneously features both positive and negative sorting. Furthermore, we note that, as in the first example, the same job type is assigned to distinct worker types. For example, among the 32 jobs with skill level 3, 8 of them are matched perfectly with workers of skill level 3, while 24 of them are matched with workers of relatively low-skill level 1. In sum, the structure of the assignment is twofold: positive sorting of identical pairs and, for the remaining workers and jobs, pairing two points in each layer leading to negative sorting.<sup>13</sup>

**Reflecting Binomial Mixture Distributions.** We now develop a more general example that provides a baseline intuition for the quantitative analysis in Section 6.

We take the previous case with the reflecting worker skill distribution and the job distribution parameterized by  $p = \frac{1}{3}$  and  $n = 4$ , and mix these distributions with a reflecting binomial worker and job distribution with  $\hat{p} = 1$  and  $n = 4$  in the ratio of 3 : 1. Again, we consider skill levels  $\{0, 1, 2, 3, 4\}$  where now the number of jobs with these skill levels is  $\{28, 8, 24, 32, 16\}$ , and the number of workers with these skill levels is  $\{16, 32, 24, 8, 28\}$ .

We first maximize perfect pairs. For example, we sort 16 jobs with skill level 0 to all workers with skill level 0. Similarly, we perfectly pair  $\{8, 24, 8, 16\}$  jobs with skill levels  $\{1, 2, 3, 4\}$ . After maximizing perfect pairs, we are left with 24 workers of skill level 1 and 12 workers of skill level 4 together with 12

---

<sup>13</sup>In fact, this result applies more broadly to any pair of worker and job distributions that lead to a unimodal measure of underqualification  $H$ .

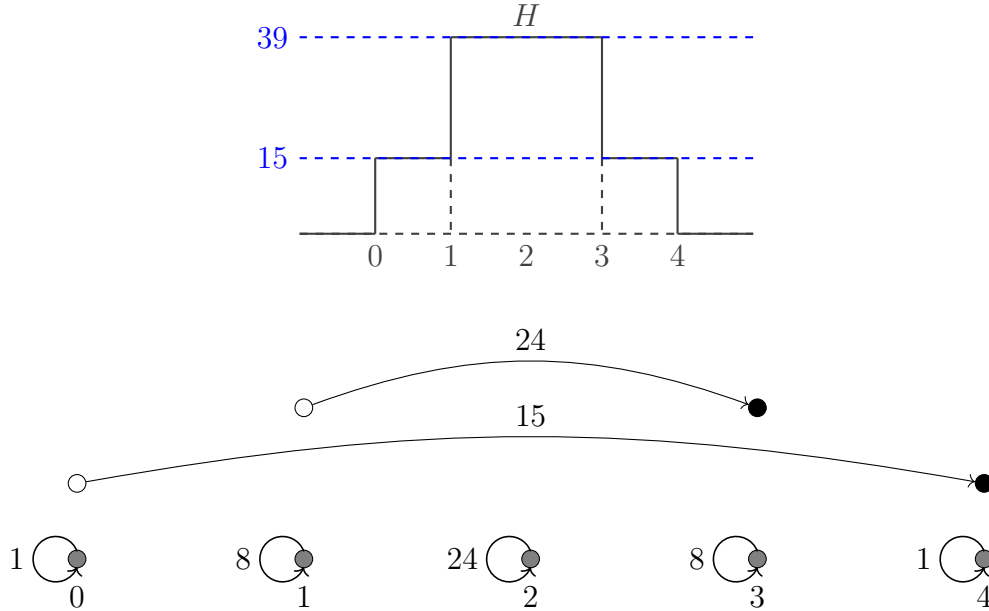


Figure 9: Sorting with Reflecting Binomial Distributions

Figure 9 explains composite sorting for an example with reflecting binomial distributions for workers (indicated by the hollow circles) and jobs (indicated by the dots). The top panel illustrates the measure of underqualification. The bottom panel visualizes the optimal assignment, where an arrow indicates that a given worker type is paired with a given job type. The numbers next to the arrows indicate the number of workers of this type that are assigned to the corresponding job type.

remaining jobs of skill level 0, and 24 jobs of skill level 3.

Second, we decompose the problem into independent layers. The measure of underqualification  $H$  after normalization<sup>14</sup> at the respective skill levels is thus given by  $\{-12, 12, 12, -12, 0\}$ , and is depicted in Figure 10. The top layer consists of 12 workers of skill level 1 and 12 jobs of skill level 3. The bottom layer contains four distinct elements: 12 jobs each at skill levels 0 and 3 and 12 workers each at skill levels 1 and 4.

Third, we characterize the optimal assignment for each layer independently. As in the case of the two reflecting binomial distributions, there is only one feasible sorting in the top layer. The problem for the bottom layer is identical to the example of positive and negative sorting where worker skills are represented by  $x_L$  and  $x_H$ , while jobs are represented by  $z_L$  and  $z_H$  satisfying  $z_L < x_L < z_H < x_H$ . The assumed distributions of skill levels imply that  $d_{LL} = d_{HH} = 1$ ,  $d_{LH} = 2$  and  $d_{HL} = 4$ . In this case, the sorting within the top layer is positive since  $2 \leq 2^\zeta + 4^\zeta$  for all  $\zeta \in (0, 1)$ .

The full optimal assignment is shown in Figure 10. Consider the job with skill difficulty 3. This job is paired with workers of skill 1, 3, and 4. Similarly, the workers with skill type 1 are paired with jobs of

<sup>14</sup>That is, multiplied by the total number of workers or jobs, equal to 108.

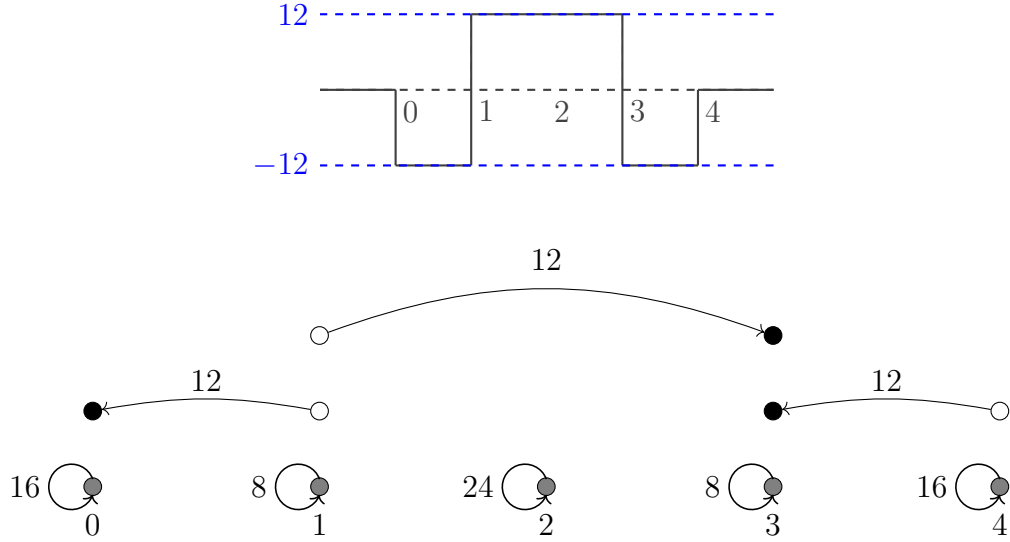


Figure 10: Sorting with Reflecting Binomial Mixture Distributions

Figure 10 explains composite sorting for an example with reflecting binomial mixture distributions for workers (indicated by the hollow circles) and jobs (indicated by the dots). The top panel illustrates the measure of underqualification. The bottom panel visualizes the optimal assignment, where an arrow indicates that a given worker type is paired with a given job type. The numbers next to the arrows indicate the number of workers of this type that are assigned to the corresponding job type.

type 0, 1, and 3. We emphasize that, even without perfect pairing (the bottom row of the bottom panel, circular arcs), the job with skill difficulty 3 is a part of both positive and negative pairing.

## 5 Wages and Firm Values

We determine the equilibrium wages and firm values by characterizing the solution to the dual planning problem (10). The presentation is divided into two parts. The first part characterizes the dual for mismatched workers and jobs by constructing the local hierarchical structure which delivers the shadow cost of mismatch. The second part then constructs the dual for both the mismatched and perfectly paired workers. Our construction of the dual solution and characterization of its local hierarchical structure are interesting not only from the economic perspective, but also are new to the optimal transport literature.

**Mismatched Workers and Jobs.** We consider a problem with  $n$  workers with skill levels in a set  $X$  and  $n$  jobs with difficulty levels in a set  $Z$  with potentially multiple workers at each skill level and multiple jobs at each difficulty level, where  $X$  and  $Z$  are disjoint after the removal of perfect pairs. We denote by  $S = X \cup Z$  the set of all skill levels.

Suppose that an optimal assignment  $\pi$  consists of  $n$  worker-job pairs  $(x_i, z_i) \in \Gamma_\pi$ . Our goal is to first construct a function  $\phi : S \rightarrow \mathbb{R}$  such that for each worker  $x$  and every job  $z$ ,  $\phi(x) - \phi(z) \leq c(x, z)$ ,

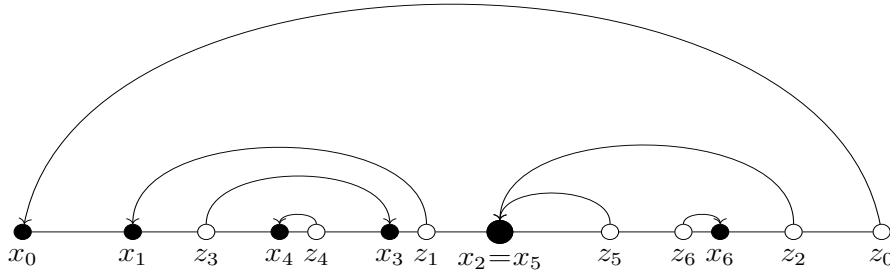


Figure 11: Local Hierarchical Structure

Figure 11 illustrates our local hierarchical structure. A lower skill group has skills in the interval between  $x_1$  and  $z_1$  while a higher skill group has skills in the interval between  $x_2$  and  $z_2$ . The local hierarchical structure implies that the relative shadow cost of mismatch for all skills within either the first or the second group is determined exclusively within the group (local). Wages are determined hierarchically within groups. In the lower skill group, wages are first determined for the innermost pair  $(x_4, z_4)$  which contains no nested skill groups, then for the intermediate pair  $(x_3, z_3)$ , and finally for the outer pair  $(x_1, z_1)$ .

which holds with equality if the assignment  $\pi$  pairs worker  $x$  to job  $z$ .<sup>15</sup> This gives the dual potentials  $\phi$  and  $\psi$ , yielding a shadow cost  $\phi$  on  $X \cup Z$  of mismatch for the worker and the job via the connection  $\psi(z) = -\phi(z)$  for all mismatched jobs  $z \in Z$ .

We propose a construction of the dual solution that is specific to the model of composite sorting. This mechanism builds on a local hierarchical structure for the shadow prices of mismatch. We show that locality and hierarchy are direct consequences of the absence of intersecting pairs. The main implication of this structure is that the relative wages are determined locally – for any two skills  $s$  and  $s'$  in a given skill group, their relative wages  $\phi(s) - \phi(s')$  can be computed only based on information in this skill group.<sup>16</sup> Next, the mechanism aggregates this local structure by constructing suitable level shifts that preserve the local structure of relative wages within groups while ensuring consistency across all groups.<sup>17</sup>

<sup>15</sup>Equivalently, we construct worker wage function  $\phi : X \rightarrow \mathbb{R}$  and firm value function  $\psi : Z \rightarrow \mathbb{R}$  such that for each worker  $x$  and job  $z$ ,  $\phi(x) + \psi(z) \leq c(x, z)$ , where the equality holds if the optimal assignment  $\pi$  matches  $x$  with  $z$ , that is  $(x, z) \in \Gamma_\pi$ . Since the intersection of the set of workers and the set of jobs is empty these formulations are equivalent.

<sup>16</sup>Since the optimality of a dual potential is invariant constant shifts, the relative difference  $\phi(s) - \phi(s')$  contains all information of the solution to the dual problem.

<sup>17</sup>A local-global structure for the primal problem appears in [McCann \(1999\)](#) where he argues that the points of nondifferentiability in the dual problem can be used to determine local regions for the primal problem. Instead, we characterize a local-global structure for the dual solution directly using the primal problem. The primary technical difficulty in our construction is to ensure consistency at every scale from local to global via aggregation of the local relative wages.

The local region generally does not appear in the case of convex costs. For instance, when all worker skills are lower than each job difficulty, the optimal assignment for a convex cost is the positive sorting, which allows for no local regions (except for the whole set) since any two pairs intersect. In contrast, for a concave cost, every pair forms a local region regardless of how the worker skills and jobs are located. Therefore, our findings on the hierarchical structure for the dual problem are specific to concave costs.



Figure 11 illustrates our local hierarchical structure. The lower skill group has skills in the interval between  $x_1$  and  $z_1$ . The higher skill group has skills in the interval between  $x_2$  and  $z_2$ . The relative shadow cost of mismatch for all skills within either the first or the second group is determined solely within each group. That is, the wage determination is local by group. Wage determination is also hierarchical within groups. In the lower skill group, the wage is first determined for the innermost pair  $(x_4, z_4)$  which contains no nested skill groups, then for the intermediate pair  $(x_3, z_3)$ , and finally for the outer pair  $(x_1, z_1)$ . In other words, hierarchical structure implies that we move sequentially from lower-level pairs to higher-level pairs. At each stage, wages depend only on information from the skill group nested within the progressively larger group. For the higher skill group, the relative wage for the pair  $(x_2, z_2)$  is thus constructed from the relative wages for both pairs  $(x_5, z_5)$  and  $(x_6, z_6)$  which are, in turn, determined locally. Finally, the wages for the outermost pair  $(x_0, z_0)$  are constructed using the relative wages for the first and the second skill group. The formal definition of the local hierarchical mechanism is analogous to this example but is notationally more involved. We present a condensed statement of the theorem here and include the complete definition of the mechanism, the full statement of the theorem, and the proof in the Appendix.

**Theorem 2.** Given an optimal assignment, the local hierarchical mechanism constructs an optimal dual pair  $(\phi, \psi)$  where  $\psi = -\phi$ . Within each skill group, relative wages are determined locally: for any two points  $s$  and  $s'$  in a skill group,  $\phi(s) - \phi(s')$  depends only on the pairings within the group.

We next use the dual functions to define worker wages and firm values for these non-overlapping segments of the worker distribution and the job distribution. We let worker wages be  $w(x) = g(x) - \phi(x)$  and firm values be  $v(z) = h(z) - \psi(z)$ . The first observation is that the assignment  $\pi$  which solves the primal mismatch cost minimization problem (9) also solves the primal output maximization problem (8). Moreover,  $w(x) + v(z) \geq y(x, z)$  holds for all  $(x, z)$  with equality if  $(x, z) \in \Gamma_\pi$ , where  $y(x, z) = g(z) + f(x) - c(x, z)$ . By Lemma 1, it follows that  $(w, v)$  is a dual optimizer for the output maximization problem. In sum, given the shadow mismatch cost  $(\phi, \psi)$  for the minimization problem without overlapping parts, the dual pair  $(w, v)$  for the maximization problem without overlapping parts is obtained.

**Incorporating Perfectly Paired Workers and Jobs.** Up to this point, we have addressed the issue of determining worker wages and firm values in the output maximization problem where there is no overlap between the distributions of workers and jobs. Next, we discuss how these wage and value functions can be used to construct worker wages and firm values for the problem with overlapping distributions.

Before we formalize our construction, an informal description is helpful. We start with an equilibrium

of only mismatched workers and jobs. We then add perfectly matched firms and determine what income each firm could maximally generate given mismatched workers and wages  $\tilde{w}$  that is obtained in the discussion above. We next introduce perfectly matched workers and pass on the new information, the assessed value of all jobs, to both mismatched and perfectly matched workers. We then determine what wage income both the imperfectly and perfectly paired workers would generate given the compensation required by firms. As a result, we obtain the required wages for both mismatched and perfectly paired workers. We pass on this information, the compensation required by workers, further to all firms such that all firms now have information about all workers. Finally, we determine what income all firms would generate given all workers and their required compensation. The dual values are the required compensation of workers and firms in the final step. We next formalize this sequential construction.

After removing the common parts of distributions, denote the sets of remaining workers and jobs as  $I$  and  $J$ . We start by defining the first auxiliary firm decision problem, which is the problem of a firm restricted to choosing an employee among only mismatched workers  $x \in I$ , that is, the workers who are not in perfect pairs. Formally, a firm with job  $z \in I \cup J$  solves:  $\tilde{v}(z) := \max_{x \in I} y(x, z) - \tilde{w}(x)$ . We denote by  $\tilde{v}$  the firm mismatch compensation, the maximum profits firms can attain given a set of mismatched workers and wage schedule  $\tilde{w}$ . We use firm mismatch compensation to define an auxiliary worker decision problem, which is the problem of a worker  $x \in I \cup J$  choosing any job, including the perfect pairs, given the firm mismatch compensation schedule  $\tilde{v}$ :

$$\hat{w}(x) := \max_{z \in I \cup J} y(x, z) - \tilde{v}(z). \quad (13)$$

We set up a second auxiliary firm decision problem, which is the problem of a mismatched job  $z \in J$  now choosing an employee among all workers (perfectly aligned and mismatched) subject to auxiliary wage schedule  $\hat{w}$ :

$$\hat{v}(z) := \max_{x \in I \cup J} y(x, z) - \hat{w}(x). \quad (14)$$

We refer to  $\hat{v}$  as mismatched firm compensation, since it represents the profits of firm type  $z \in J$  that is mismatched in equilibrium.

Equilibrium wages are formulated using auxiliary wage income for mismatched workers,  $w(x) = \hat{w}(x)$  for all  $x \in I$ , as well as mismatched firm compensation,  $v(z) = \hat{v}(z)$  for all  $z \in J$ , to set  $w(x) = -v(x)$  for all  $x \in J$ . Equilibrium firm values  $v$  are then given by  $v(z) = -w(z)$  for every job  $z \in I \cup J$ . Theorem 3 shows that the wage function  $w$  and the firm value function  $v$  indeed solve the dual problem for the full

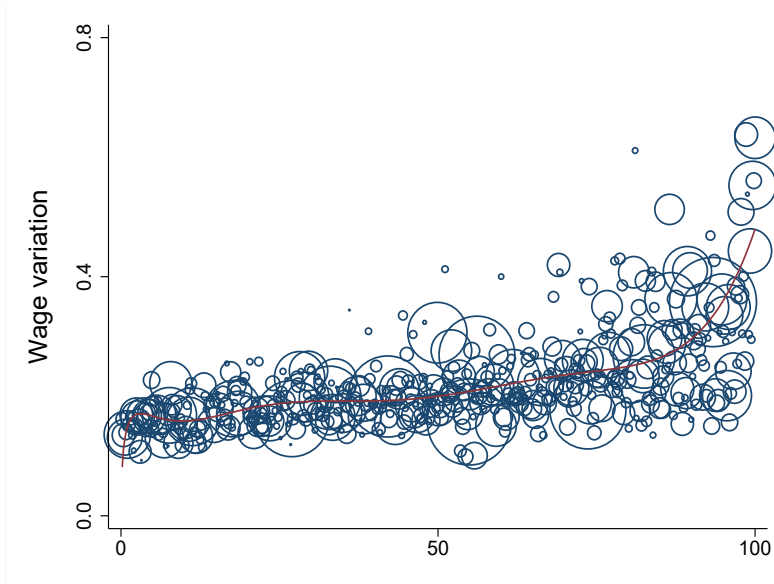


Figure 12: Wage Dispersion by Occupation Rank

Figure 12 displays wage dispersion within occupation. On the horizontal axis, we rank occupations by the average wages earned in each occupation. For every occupation, we calculate the dispersion in logarithmic wages within that occupation, where the size of the circle indicates the share of employment within the occupation. The data pattern is summarized by the red solid line.

assignment problem between worker distribution  $F$  and job distribution  $G$ .<sup>18</sup>

**Theorem 3.** The constructed functions  $(w, v)$  are a dual solution for the sorting problem between worker distribution  $F$  and job distribution  $G$ , that is,  $w(x) + v(z) \geq y(x, z)$ , where equality holds everywhere with respect to  $\Gamma_\pi$ .

The proof of Theorem 3, as well as a formal analysis of the above mechanism, is in Appendix A.9.

## 6 Quantitative Analysis

The distinguishing feature of the model is frictionless wage dispersion within occupations. We explore the ability of the model to generate qualitative and quantitative patterns of wage dispersion within occupations.

**Data.** Our main data source is the American Community Survey. We consider all individuals between 25 and 60 years of age between 2010 and 2017. The final sample includes about 6.7 million individuals. We select this period to work with a single Standard Occupational Classification (SOC) code, which yields 497 distinct and consistent occupations. Our measure of labor income is wage and salary income before

<sup>18</sup>We can extend the domain of wages  $w$  and values  $v$  to  $K = X \setminus (I \cup J)$  by setting  $w(x) = \max_{z \in I \cup J} y(x, z) - v(z)$  for  $x \in K$  and  $v(z) = -w(z)$  for  $z \in K$ .

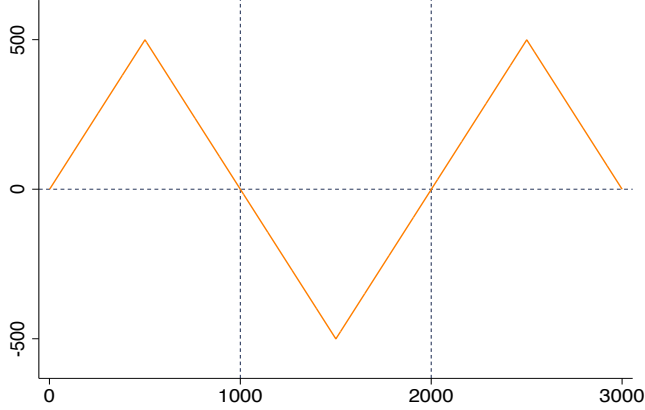


Figure 13: Measure of Underqualification

Figure 13 shows the measure of underqualification for our quantitative analysis. An increasing measure of underqualification indicates mismatched workers, while a decreasing measure of underqualification indicates mismatched jobs.

taxes over the past 12 months.<sup>19</sup>

Figure 12 shows the dispersion of individual wages within occupation. On the horizontal axis, we rank occupations by the average wages earned in each occupation. For every occupation, we calculate the dispersion in logarithmic wages within that occupation, where the size of the circle indicates the share of employment within the occupation. The salient data pattern is captured by the red solid line which is a fractional polynomial fit. Wage dispersion within occupations is sizable, with the average wage dispersion equal to 0.24 log points. Second, the wage dispersion within occupations is relatively flat at the bottom 80 percent of occupations, but rapidly increases for the top fifth of occupations.

**Analysis.** We next calibrate the technology. We choose simple parameter values for the cost of mismatch:  $\zeta_p = \zeta_\kappa = \frac{1}{2}$  and  $\rho_p = \rho_\kappa = 1$ . These parameters, together with the distributions of workers and jobs, determine the equilibrium assignment. We discipline the choice for the production value of a worker independent of their job  $g(x)$  to exactly replicate earnings data in the United States. That is, we calibrate the function  $g$  to exactly replicate the wages at each percentile in the data.<sup>20</sup>

We now expand on the example of reflecting binomial mixture distributions that we discussed in Section 4 to develop an understanding of the quantitative exploration. First, in Figure 13, we plot a continuous analog of the measure of underqualification in that example. It follows from our theoretical analysis of Section 3 that this measure is an important determinant of an optimal assignment. There are

<sup>19</sup>Labor income includes wages, salaries, commissions, cash bonuses, tips, and other money income received from an employer. We drop individuals with wages below a threshold to focus on workers who are attached to the labor market. This minimum is equal to the federal minimum wage times 52 weeks at 40 hours per week.

<sup>20</sup>We do not need to calibrate the function  $h$  as all our results are invariant to it. In principle, the function could be directly inferred using a measure of the value of a job.

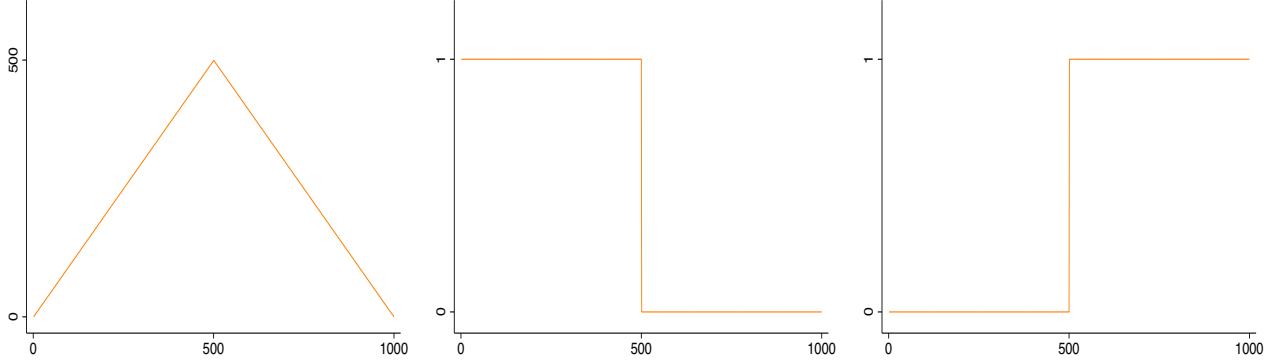


Figure 14: Mismatched Workers and Jobs in the First Region

Figure 14 shows the measure of underqualification and the underlying distributions of mismatched workers and jobs for the first region in our quantitative analysis. The measure of underqualification linearly increases and then linearly decreases (left panel). The underlying distribution of mismatched workers is a uniform distribution with skills between 0 and 500 (middle panel). The underlying distribution of mismatched jobs is uniform with skills between 500 and 1000 (right panel).

two peaks and one trough – an increasing measure of underqualification indicates mismatched workers; a decreasing measure of underqualification indicates mismatched jobs. Figure 13 thus indicates mismatched workers with skill difficulty levels between 0 and 500 as well as mismatched workers between 1500 and 2500. Similarly, Figure 13 indicates mismatched jobs with skill levels between 500 and 1500 as well as mismatched jobs with skill difficulties between 2500 and 3000. To explain the equilibrium assignment, we divide the measure of underqualification into three symmetric regions. The regions are given by 0 to 1000, 1000 to 2000, and 2000 to 3000.

Let us consider the first region given by skills between 0 and 1000. The analysis for the other two regions is analogous.<sup>21</sup> In this region, the measure of underqualification linearly increases and then linearly decreases, as we plot in the left panel of Figure 14. The underlying distribution of the mismatched workers is plotted in the middle panel: a uniform distribution of workers with skills between 0 and 500. The underlying distribution of the mismatched jobs is uniform between 500 and 1000 and is plotted in the right panel.

To understand the optimal assignment, consider the workers  $x_1$  and  $x_2$  (where  $x_1 < x_2$ ) and the jobs  $z_1$  and  $z_2$  (where  $z_1 < z_2$ ). By the principle of no intersecting pairs, it follows that we optimally pair worker  $x_1$  with job  $z_2$  and worker  $x_2$  with job  $z_1$ . There are only two possible pairings. The first is

<sup>21</sup>The intuition for separation into three regions is similar to the example of reflecting binomial mixture distributions and is verified formally using the Bellman algorithm. In general, separation may not be optimal and a complete characterization of sorting for arbitrary distributions and any concave cost function is constructed using the Bellman recursion (12). Given the separation of the measure of underqualification it follows that our equilibrium sorting is not sensitive to the technological parameter values.

positive assortative matching:  $(x_1, z_1)$  and  $(x_2, z_2)$ . The second is negative assortative matching:  $(x_1, z_2)$  and  $(x_2, z_1)$ . Positive assortative matching here leads to intersections, contradicting the property of no intersecting pairs, and, hence, is not optimal. Negative assortative matching is the only non-intersecting configuration and is thus optimal. Since this reasoning holds for any two workers and two jobs in this region it follows that all workers and jobs are negatively sorted. Alternatively, we could have established the optimal assignment directly from the principle of layering. Since each layer within the region contains a single worker and a single job, the optimal assignment pairs them.

The optimal pairing features significant variation in mismatch. The mismatch varies from almost zero for the worker at skill (slightly below) 500 and the job (slightly above) 500 to the mismatch of 1000 for the worker at skill 0 and the job at skill 1000. In contrast, if sorting were positive, there would be no variation in mismatch – mismatch for each worker-job pair would equal 500.

Negative sorting in this region implies that more valuable jobs feature larger fixed investments. The worker with the highest skills is paired with the job with the lowest skills demands meaning that the mismatch  $(z - x)$  between the worker and the job is small. Since mismatch is small and technology choice increases in mismatch, the extent of investment is small. However, the worker with the lowest skills is paired with the most valuable job. To ensure that the value of this job is not significantly diminished, a large investment is made. An example would be the case of welding (SOC 51-4120, mean wage of 45 thousand dollars) and the adoption of collaborative welding robots (cobots).<sup>22</sup> A low-skill welder (with high mismatch between the skill and the job) is assisted by a cobot which ensures the quality of the output. A high-skill welder (with no mismatch between the skill and the job) instead completes the job without robotic assistance. In the right region of Figure 14, the intuition is similar for the case of surgeons (Physicians and Surgeons SOC 29-1060 with a mean wage of 213 thousand dollars is the top-ranked occupation). A low-skill surgeon is assisted by a robotic surgical system while a high-skill surgeon operates without robotic assistance.<sup>23</sup> The welding company or a surgical practice makes investments in technology in order to produce valuable output with minimal mismatch costs. In the middle region, workers are more qualified than the job requires and firms invest in amenities for more skilled employees to reduce their utility cost of mismatch. Consider the case of semi-truck drivers (SOC 53-3030, mean wage 47 thousand dollars). A high-skill trucker drives a new truck with a quiet cabin, enhanced sleep space, and driver assistance technologies such as adaptive cruise control and digital mirrors. A low-skill

---

<sup>22</sup>A welding cobot consists of a robotic arm with welding instruments attached. A high-end cobot costs about 100 thousand dollars versus 5 thousand dollars for a basic industrial welding machine.

<sup>23</sup>A surgical robot consists of a control console and robotic arms with surgical instruments and costs about 2 million dollars.

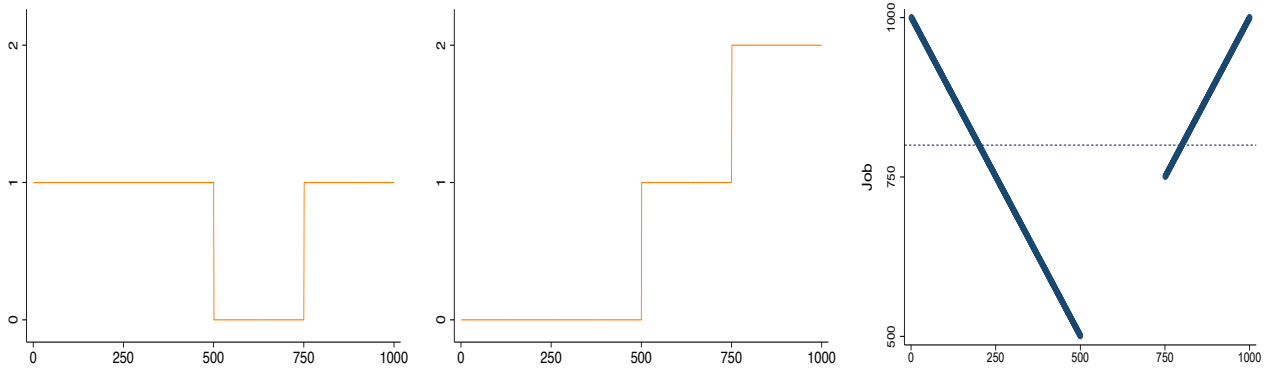


Figure 15: Mismatched and Perfectly Paired Workers and Jobs

Figure 15 shows the distribution of workers (left panel) and the distribution of jobs (middle panel) for the first region in our quantitative analysis after adding perfectly paired workers and jobs into the setting described in Figure 14. Specifically, we add 250 workers and 250 jobs with identical skill levels between 750 and 1000. The optimal sorting (right panel) shows that two distinct workers can work in the same job: one positively sorted worker in a perfect pair, and one negatively sorted worker.

trucker drives a basic model without such features.<sup>24</sup> A high-skill computer user support specialist (SOC 15-1150, mean wage 64 thousand dollars) has an option to work from home using investments in remote access and control software. A low-skill computer support specialist instead works on-site in person.<sup>25</sup>

Second, we add common parts to the distributions of workers and jobs. In particular, we add 250 workers and 250 jobs with identical skill levels between 750 and 1000. Following Section 3, an optimal assignment pairs these workers and jobs to maximize the number of perfect pairs. We illustrate the resulting new distributions of workers and jobs in Figure 15. Thus, we now have jobs without wage dispersion with skill difficulty between 500 and 750 and jobs with wage dispersion and skill difficulty between 750 and 1000. Two distinct workers work in each of these jobs: one positively sorted worker in a perfect pair, and one negatively sorted worker. Consider a job with skill demand 800 as indicated by the horizontal dashed line in the right panel of Figure 15. The horizontal dashed line intersects the assignment at two points. At the right intersection, the worker is perfectly paired and has skill 800. There is no mismatch for this worker. At the left intersection, the worker’s skill is instead 200 and the worker is mismatched. No additional technology investment is necessary for a worker who is perfectly matched, while a worker with lower skills benefits from technology. Consider the example of welders. The welder with skill 200 is assisted by a cobot to ensure that the output is not diminished due to skill

<sup>24</sup>A new high-end Kenworth Signature truck costs about 250 thousand dollars versus an older basic model that can be purchased for about 25 thousand dollars or less.

<sup>25</sup>For more empirical and theoretical discussion on the provision and valuation of amenities, see, amongst others, Gronberg and Reed (1994), Mas and Pallais (2017), Sorkin (2018), Wiswall and Zafar (2018), Morchio and Moser (2021), and Sockin (2022).

mismatch. The perfectly paired welder with skill 800 does not need a cobot and instead works with a standard manual welding machine. The skill is a normalization that represents workers' rank. We now explain how the results on sorting by skills can be linked to sorting by wages. To convert skills into wages we use a one-to-one ranking between workers' positions in the skill distribution and their corresponding position in the earnings distribution. As an illustration, a worker located at the lowest point of the skill distribution (skill level 0) would earn a salary of 16 thousand dollars, which is the lowest possible wage level. Similarly, a worker situated in the middle of the skill distribution earns a wage of 47 thousand dollars (the median of the earnings distribution).

The key attribute of our model is its ability to create wage dispersion within the same job. The analysis above concluded that there is dispersion of skills in the same job and thus there is corresponding dispersion of wages within jobs as well. On the other hand, a model that delivers either positive or negative sorting cannot result in any variation in skill levels within a particular job, and as a consequence, does not create any wage dispersion within that job. Furthermore, any model that pairs only one worker to each job does not generate any wage dispersion within jobs.

In our example above, the same job is performed by a high-skill welder and a low-skill welder assisted by a cobot. It should be emphasized that this situation is different from a scenario where a high-skill welder is assigned to a more complex job (for example, underwater welding requiring specialized skills) and the low-skill welder is assigned to a simpler job (for example, assembling an industrial boiler). This would result in wage dispersion not within the same job, but rather across two different jobs: the complex and the simple job.

By combining the optimal assignment within the three regions, we find that the model delivers sizable wage dispersion in occupations with high and low mean wages. Consider a high-skill occupation such as surgeon. There is a large wage variation at the top of the income distribution. A high-skill surgeon earns a multiple of the wages of a low-skill surgeon. According to our model, a low-skill surgeon assisted by a robotic surgical system performs the same job as a high-skill surgeon. As a result, the salary variation in this profession is significant, since it encompasses low-skill surgeons as well as high-skill surgeons. A similar intuition holds for the bottom of the wage distribution. The wage distribution exhibits a lower level of dispersion at the bottom compared to the top, and the model accounts for a sizable part of this reduced level of dispersion. In contrast, the model only accounts for a moderate portion of wage dispersion in the middle of the income distribution. The reason is that the log-wage profile is relatively flat in this range. As a result, even though there may be substantial skill dispersion within a particular occupation, it does not give rise to substantial wage dispersion.



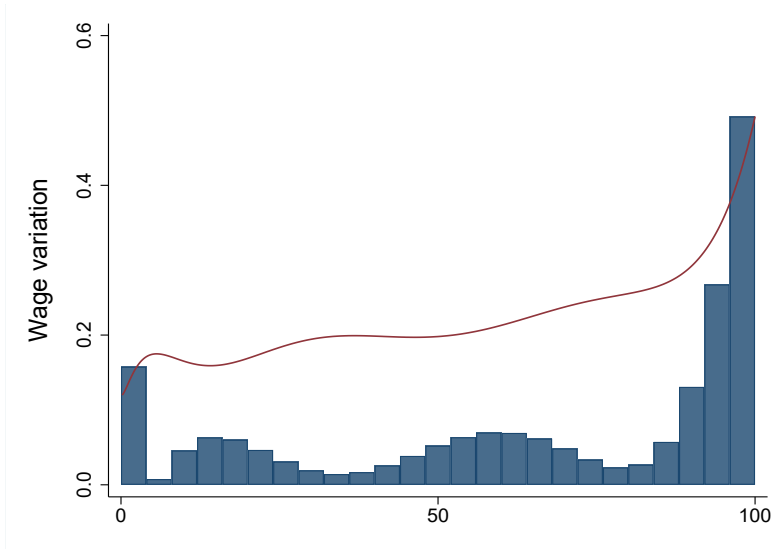


Figure 16: Wage Variation by Occupation in Model and Data

Figure 16 displays wage dispersion within occupation for the model and for the data. On the horizontal axis, we rank occupations by the average wages earned in each occupation. The red solid line is a fractional polynomial fit that captures the salient data patterns (Figure 12). The blue bars represent the variation in logarithmic wages averaged across model occupations. Our model explains 28 percent of the squared deviation and 32 percent of the absolute deviation of wages across the sample.

Our model explains 28 percent of the squared deviation and 32 percent of the absolute deviation of wages across the sample. For the middle of the distribution (20th to 80th percentiles), the model accounts for 12 percent of the squared deviation and 19 percent of the absolute deviation. In contrast, any one-to-one sorting model (for example, a model with positive or negative sorting arising from submodular or supermodular costs) would yield zero wage dispersion within occupations. For the bottom of the distribution (0th to 20th percentiles), the model accounts for 20 percent of the squared deviation and 46 percent of the absolute deviation. For the top of the distribution (80th to 100th percentiles), the model accounts for 57 percent of the squared deviation and 58 percent of the absolute deviation.

## 7 Conclusion

We provide a complete solution to an assignment problem with heterogeneous workers and heterogeneous jobs when the underlying technology is neither supermodular nor submodular. Our analysis introduces a general tractable framework – composite sorting – that yields both multiple workers sorted to the same job and workers being a part of both positive and negative sorting. We also show that wages have a new local hierarchical structure with relative wages being determined by narrow skill groups and aggregated to determine wages at different scales. We illustrate the composite sorting framework quantitatively to argue that it may account for a sizable portion of wage dispersion within jobs.

## References

- AGGARWAL, A., A. BARNOY, S. KHULLER, D. KRAVETS, AND B. SCHIEBER (1995): “Efficient Minimum Cost Matching and Transportation Using the Quadrangle Inequality,” *Journal of Algorithms*, 19(1), 116–143.
- ANDERSON, A., AND L. SMITH (2022): “The Comparative Statics of Sorting,” Discussion paper, University of Wisconsin-Madison.
- ANTRÀS, P., AND E. ROSSI-HANSBERG (2009): “Organizations and Trade,” *Annual Review of Economics*, 1(1), 43–64.
- BAGGER, J., AND R. LENTZ (2019): “An Empirical Model of Wage Dispersion with Sorting,” *Review of Economic Studies*, 86(1), 153–190.
- BASS, R. F., AND D. KHOSHNEVISAN (1995): “Laws of the Iterated Logarithm for Local Times of the Empirical Process,” *Annals of Probability*, 23(1), 388–399.
- BECKER, G. S. (1973): “A Theory of Marriage: Part I,” *Journal of Political Economy*, 81(4), 813–846.
- BIRKHOFF, G. (1946): “Tres Observaciones Sobre el Algebra Lineal,” *Universidad Nacional de Tucumán Revista Series A*, 5, 147–154.
- BOERMA, J., A. TSYVINSKI, AND A. P. ZIMIN (2021): “Sorting with Team Formation,” Discussion paper, NBER Working Paper No. 29290.
- CAHUC, P., F. POSTEL-VINAY, AND J.-M. ROBIN (2006): “Wage Bargaining with On-The-Job Search: Theory and Evidence,” *Econometrica*, 74(2), 323–364.
- CHADE, H., AND J. EECKHOUT (2018): “Matching Information,” *Theoretical Economics*, 13(1), 377–414.
- CHADE, H., J. EECKHOUT, AND L. SMITH (2017): “Sorting through Search and Matching Models in Economics,” *Journal of Economic Literature*, 55(2), 493–544.
- CHIAPPORI, P.-A., R. J. MCCANN, AND B. PASS (2017): “Multi-to One-Dimensional Optimal Transport,” *Communications on Pure and Applied Mathematics*, 70(12), 2405–2444.
- CHIAPPORI, P.-A., AND B. SALANIÉ (2016): “The Econometrics of Matching Models,” *Journal of Economic Literature*, 54(3), 832–61.

- COSTINOT, A., AND J. VOGEL (2015): “Beyond Ricardo: Assignment Models in International Trade,” *Annual Review of Economics*, 7(1), 31–62.
- CSÖRGŐ, M., Z. SHI, AND M. YOR (1999): “Some Asymptotic Properties of the Local Time of the Uniform Empirical Process,” *Bernoulli*, 5(6), 1035–1058.
- DELON, J., J. SALOMON, AND A. SOBOLEVSKI (2012a): “Local Matching Indicators for Transport Problems with Concave Costs,” *SIAM Journal on Discrete Mathematics*, 26(2), 801–827.
- (2012b): “Minimum-Weight Perfect Matching for Non-Intrinsic Distances on the Line,” *Journal of Mathematical Sciences*, 181(6), 782–791.
- EECKHOUT, J. (2018): “Sorting in the Labor Market,” *Annual Review of Economics*, 10, 1–29.
- EECKHOUT, J., AND P. KIRCHER (2010): “Sorting and Decentralized Price Competition,” *Econometrica*, 78(2), 539–574.
- FAGERENG, A., L. GUISO, AND L. PISTAFERRI (2022): “Assortative Mating and Wealth Inequality,” Discussion paper, NBER Working Paper No. 29903.
- GALICHON, A. (2018): *Optimal Transport Methods in Economics*. Princeton University Press.
- GANGBO, W., AND R. J. MCCANN (1996): “The Geometry of Optimal Transportation,” *Acta Mathematica*, 177(2), 113–161.
- GRONBERG, T. J., AND W. R. REED (1994): “Estimating Workers’ Marginal Willingness to Pay for Job Attributes Using Duration Data,” *Journal of Human Resources*, pp. 911–931.
- HWANG, H.-S., D. T. MORTENSEN, AND W. R. REED (1998): “Hedonic Wages and Labor Market Search,” *Journal of Labor Economics*, 16(4), 815–847.
- JUILLET, N. (2020): “On a Solution to the Monge Transport Problem on the Real Line Arising from the Strictly Concave Case,” *SIAM Journal on Mathematical Analysis*, 52(5), 4783–4805.
- KANTOROVICH, L. V. (1942): “On the Translocation of Masses,” in *Dokl. Akad. Nauk. USSR*, vol. 37, pp. 227–229.
- KHOSHNEVISAN, D. (1992): “Level Crossings of the Empirical Process,” *Stochastic Processes and their Applications*, 43(2), 331–343.

- KOOPMANS, T. C., AND M. BECKMANN (1957): “Assignment Problems and the Location of Economic Activities,” *Econometrica*, 25(1), 53–76.
- KREMER, M. (1993): “The O-Ring Theory of Economic Development,” *Quarterly Journal of Economics*, 108(3), 551–575.
- KREMER, M., AND E. MASKIN (1996): “Wage Inequality and Segregation by Skill,” Discussion paper, NBER Working Paper No. 5718.
- LAFFONT, J.-J., AND J. TIROLE (1986): “Using Cost Observation to Regulate Firms,” *Journal of Political Economy*, 94(3), 614–641.
- (1991): “The Politics of Government Decision-Making: A Theory of Regulatory Capture,” *Quarterly Journal of Economics*, 106(4), 1089–1127.
- LANG, K., AND S. MAJUMDAR (2004): “The Pricing of Job Characteristics when Markets do not Clear: Theory and Policy Implications,” *International Economic Review*, 45(4), 1111–1128.
- LISE, J., AND F. POSTEL-VINAY (2020): “Multidimensional Skills, Sorting, and Human Capital Accumulation,” *American Economic Review*, 110(8), 2328–76.
- LISE, J., AND J.-M. ROBIN (2017): “The Macrodynamics of Sorting between Workers and Firms,” *American Economic Review*, 107(4), 1104–1135.
- MAS, A., AND A. PALLAIS (2017): “Valuing Alternative Work Arrangements,” *American Economic Review*, 107(12), 3722–3759.
- MCCANN, R. J. (1999): “Exact Solutions to the Transportation Problem on the Line,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1984), 1341–1380.
- MONGE, G. (1781): “Mémoire sur la Théorie des Déblais et des Remblais,” *Histoire de l’Académie Royale des Sciences de Paris*, pp. 666–704.
- MORCHIO, I., AND C. MOSER (2021): “The Gender Pay Gap: Micro Sources and Macro Consequences,” Discussion paper, Columbia University.
- NECHAEV, S. K., A. SOBOLEVSKI, AND O. V. VALBA (2013): “Planar Diagrams from Optimization for Concave Potentials,” *Physical Review E*, 87(1), 1–9.

- PEGON, P., F. SANTAMBROGIO, AND D. PIAZZOLI (2015): “Full Characterization of Optimal Transport Plans for Concave Costs,” *Discrete & Continuous Dynamical Systems*, 35(12), 6113.
- PORZIO, T. (2017): “Cross-Country Differences in the Optimal Allocation of Talent and Technology,” Discussion paper, Columbia University.
- POSTEL-VINAY, F., AND J.-M. ROBIN (2002): “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 70(6), 2295–2350.
- RACHEV, S. T., AND L. RÜSCHENDORF (1998): *Mass Transportation Problems: Volume I: Theory*, vol. 1. Springer Science.
- ROSEN, S. (1986): “The Theory of Equalizing Differences,” *Handbook of Labor Economics*, 1, 641–692.
- SANTAMBROGIO, F. (2015): *Optimal Transport for Applied Mathematicians*. Birkhäuser.
- SATTINGER, M. (1993): “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature*, 31(2), 831–880.
- SHIMER, R., AND L. SMITH (2000): “Assortative Matching and Search,” *Econometrica*, 68(2), 343–369.
- SOCKIN, J. (2022): “Show Me the Amenity: Are Higher-Paying Firms Better All Around?,” Discussion paper, University of Pennsylvania.
- SORKIN, I. (2018): “Ranking Firms Using Revealed Preference,” *Quarterly Journal of Economics*, 133(3), 1331–1393.
- STIGLER, G. (1939): “Production and Distribution in the Short Run,” *Journal of Political Economy*, 47(3), 305–327.
- VILLANI, C. (2003): *Topics in Optimal Transportation*, vol. 58. American Mathematical Society.
- (2009): *Optimal Transport: Old and New*. Springer.
- WERMAN, M., S. PELEG, R. MELTER, AND T. Y. KONG (1986): “Bipartite Graph Matching for Points on a Line or a Circle,” *Journal of Algorithms*, 7(2), 277–284.
- WISWALL, M., AND B. ZAFAR (2018): “Preference for the Workplace, Investment in Human Capital, and Gender,” *Quarterly Journal of Economics*, 133(1), 457–507.

# Composite Sorting

## Online Appendix

Job Boerma, Aleh Tsyvinski, Ruodu Wang, and Zhenyuan Zhang

March 2023

### A Proofs

In this appendix, we formally prove the results in the main text.

#### A.1 Proof to Lemma 1

Suppose the assignment  $\pi$  and the functions  $(w, v)$  satisfy our assumptions above. Then it holds that  $\int y(x, z) d\pi = \int w(x) dF + \int v(z) dG$ . By Monge-Kantorovich duality, we must have that the maximum for the primal problem is attained by  $\pi$  and the minimum for the dual problem is attained by  $(w, v)$ , as required.

#### A.2 Proof to Lemma 2

We show that a perfect pairing is made when feasible. By contradiction, suppose an optimal assignment contains pairings  $(x, z)$  and  $(x', z')$  when  $x' = z$ .<sup>26</sup>

By symmetry, it suffices to consider two cases. Consider first the case  $x \leq z' < z = x'$ . Since the cost of mismatch  $\bar{c}$  is strictly increasing,  $c(x, z') + c(x', z) = c(x, z') < c(x, z) \leq c(x, z) + c(x', z')$  where the equality follows since  $c(x', z) = 0$ . Thus, the cost of mismatch when making the perfect pairing is strictly lower than under the optimal configuration, which is a contradiction.

Second, we consider the case where  $x < x' = z < z'$ . In this case, the cost of mismatch  $\bar{c}$  is given by  $c(x, z') + c(x', z) = \bar{c}(z' - x)$  since  $c(x', z) = 0$ . To arrive at a contradiction, choose some weight  $\lambda \in (0, 1)$  to scale the maximum distance such that  $z - x = (1 - \lambda)(z' - x)$ . Since the total distance is given by  $(z - x) + (z' - x') = (z' - x)$ , we also have  $z' - x' = \lambda(z' - x)$ . Since the cost of mismatch is strictly concave, we use strict concavity and add the two previous equations to obtain  $\bar{c}(z - x) + \bar{c}(z' - x') > \bar{c}(z' - x)$ .

---

<sup>26</sup>If either  $x = x'$  or  $z = z'$ , a perfect pairing is naturally made since  $x' = z$ . We thus restrict our attention to the cases where  $x \neq x'$  and  $z \neq z'$ .

The output loss due to mismatch can be strictly reduced by assigning worker  $x$  to job  $z'$  and by perfectly assigning worker  $x'$  to job  $z$ , which is a contradiction.<sup>27</sup>

### A.3 Proof to Lemma 3

To establish the result, we show that if two pairings  $(x, z)$  and  $(x', z')$  under an optimal assignment intersect, then the support of the assignment is not optimal.

By symmetry, it suffices to consider two cases.<sup>28</sup> First, consider the case  $x < z' < z < x'$ . Since the cost function is increasing,  $c(x, z') + c(x', z) = \bar{c}(z' - x) + \underline{c}(x' - z) < \bar{c}(z - x) + \underline{c}(x' - z') = c(x, z) + c(x', z')$ . The output loss due to mismatch is strictly reduced by assigning worker  $x$  to job  $z'$  and worker  $x'$  to job  $z$ , which is a contradiction.

Second, consider the case  $x < x' < z < z'$ . In this case, the cost of mismatch  $\bar{c}$  is given by  $c(x, z') + c(x', z) = \bar{c}(z' - x) + \bar{c}(z - x')$ . To arrive at a contradiction, choose some weight  $\lambda \in (0, 1)$  to average the minimum and maximum distance such that

$$z - x = (1 - \lambda)(z' - x) + \lambda(z - x').$$

Since the total distance is given by  $(z - x) + (z' - x') = (z' - x) + (z - x')$ , we moreover obtain

$$z' - x' = \lambda(z' - x) + (1 - \lambda)(z - x').$$

Since the cost of mismatch is strictly concave, we can use strict concavity and add the two previous equations to obtain  $\bar{c}(z - x) + \bar{c}(z' - x') > \bar{c}(z' - x) + \bar{c}(z - x')$ . The output loss due to mismatch can be strictly reduced by assigning worker  $x$  to job  $z'$  and worker  $x'$  to job  $z$ , which is a contradiction.

### A.4 Layering and Proof to Lemma 4

To prove Lemma 4, we make use of the following result due to Villani (2009). We repeat the result here for completeness.

<sup>27</sup>Pegon, Santambrogio, and Piazzoli (2015) further show that Lemma 2 extends with continuous worker and job distributions in higher dimensions. In these environments, the solution is partially characterized by two parts. Overlapping sets of the distributions are perfectly paired, and the remaining mass is assigned using Monge maps. In our economy featuring distributions with atoms the remaining mass is not assigned using one-to-one assignments as identical worker types fulfill different jobs. We highlight this feature of optimal assignments in the first example in Section 4.

<sup>28</sup>There is a total of six distinct scenarios to consider, which can diagrammatically be represented by the following pairings: . The first four scenarios do not contain intersecting pairs. The fifth scenario is the first scenario in the proof, while the sixth scenario is the second scenario in the proof.

**Lemma 5.** *Stability of Optimal Assignment.* Let  $c(x, z)$  be a continuous non-negative cost function, and  $\{F_n\}_{n \in \mathbb{N}}, \{G_n\}_{n \in \mathbb{N}}$  be sequences of distributions of workers and jobs, respectively. Suppose  $F_n \rightarrow F, G_n \rightarrow G$  weakly for some  $F, G$ ,<sup>29</sup> and let  $\pi_n$  be an optimal assignment between  $F_n$  and  $G_n$ . If  $\pi_n \rightarrow \pi$  in distribution, then  $\pi$  is an optimal assignment between  $F$  and  $G$ .

To establish Lemma 4, we first formally define  $F_\ell$  and  $G_\ell$ , the measures of the workers and the jobs in each layer. To obtain different layers for a general underqualification measure, we identify the skill levels where underqualification increases, and the skill levels where underqualification decreases. Recall that an underqualification measure  $H$  takes a finite number of values in  $a_0 < a_1 < \dots < a_L$ . Underqualification increases from  $a_\ell$  to  $a_{\ell+1}$  at a skill level  $s$  if  $H(s_-) \leq a_\ell < a_{\ell+1} \leq H(s)$ , where  $s_-$  represents the limit from the left. Analogously, the measure of underqualification decreases from  $a_{\ell+1}$  to  $a_\ell$  at a skill level  $s$  if  $H(s_-) \geq a_{\ell+1} > a_\ell \geq H(s)$ . The set of skill levels where underqualification increases is denoted by  $X_\ell^\uparrow := \{s: H \text{ increases from } a_{\ell-1} \text{ to } a_\ell \text{ at skill level } s\}$  for all  $1 \leq \ell \leq L$ . Similarly, the set of skill levels where underqualification decreases is denoted by  $X_\ell^\downarrow := \{s: H \text{ decreases from } a_\ell \text{ to } a_{\ell-1} \text{ at skill level } s\}$ . Define the discrete measures for all layers  $1 \leq \ell \leq L$  by

$$F_\ell := (a_\ell - a_{\ell-1}) \sum_{x \in X_\ell^\uparrow} \delta_x \quad \text{and} \quad G_\ell := (a_\ell - a_{\ell-1}) \sum_{x \in X_\ell^\downarrow} \delta_x$$

where  $\delta_x$  is the Dirac measure at  $x$ . It follows that  $F = \sum F_\ell$  and  $G = \sum G_\ell$ . Note that  $F_\ell$  and  $G_\ell$  are generally not probability measures as they are not required to have total measure one.

Having defined the measure of workers and jobs in each layer, we next observe that the worker and job distributions are supported on disjoint sets and on a finite set of skills  $\{s_j\}_{1 \leq j \leq S}$ . We smooth both the discrete distributions of workers and the discrete distribution of jobs by replacing each atom in the worker and job distribution at level  $s_j$  by a uniform distribution on  $[s_j, s_j + \varepsilon]$  with the same mass for every  $1 \leq j \leq S$ , where  $\varepsilon$  is small enough such that the intervals  $[s_j, s_j + \varepsilon]$  for all  $1 \leq j \leq S$  do not intersect. In Figure A.1, we provide an illustration of this procedure using the underqualification measure  $H$  of Figure 5. We denote the smoothed measure of workers by  $F_\varepsilon$  and the smoothed measure of jobs by  $G_\varepsilon$ , and the corresponding underqualification measure by  $H_\varepsilon$ . An optimal assignment given worker measure  $F_\varepsilon$  and job measure  $G_\varepsilon$  is given by  $\pi_\varepsilon$ . Since the mismatch cost  $c$  is continuous, by stability of the optimal transport,  $\pi_\varepsilon \rightarrow \pi$  weakly where  $\pi$  is the optimal matching between workers  $F$  and jobs  $G$ .

Consider the support  $A_\varepsilon^\ell := H_\varepsilon^{-1}((a_{\ell-1}, a_\ell))$  for all layers  $1 \leq \ell \leq L$  and define the smoothed worker distribution  $F_\varepsilon^\ell := F_\varepsilon|_{A_\varepsilon^\ell}$  and the smoothed job distribution  $G_\varepsilon^\ell := G_\varepsilon|_{A_\varepsilon^\ell}$  for every layer such that

<sup>29</sup>This means that  $F_n \rightarrow F$  on continuity points of  $F$  and respectively for  $G$ .



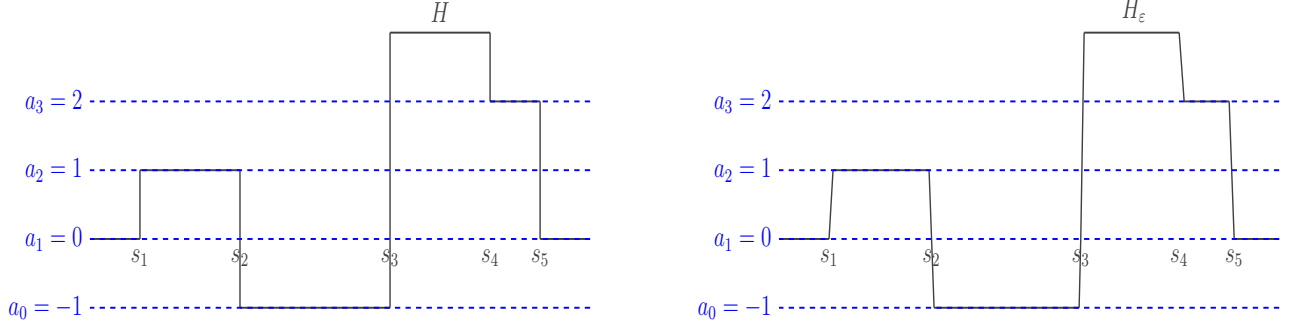


Figure A.1: Smoothed Measure of Underqualification  $H_\varepsilon$ .

Figure A.1 illustrates the smoothing of the measure of underqualification  $H$  displayed in the left panel. The corresponding smoothed measure of underqualification  $H_\varepsilon$  is presented in the right panel.

$F_\varepsilon = \sum F_\varepsilon^\ell$  and  $G_\varepsilon = \sum G_\varepsilon^\ell$ .<sup>30</sup> Moreover, let an optimal assignment between workers  $F_\varepsilon^\ell$  and jobs  $G_\varepsilon^\ell$  in layer  $\ell$  be denoted by  $\pi_\varepsilon^\ell$ . Clearly, this assignment  $\pi_\varepsilon^\ell$  is supported on the set  $(A_\varepsilon^\ell)^2$ .

Next, we establish that the sum of optimal assignments across layers  $\sum \pi_\varepsilon^\ell$  is an optimal assignment between the smoothed worker distribution  $F_\varepsilon$  and the smoothed job distribution  $G_\varepsilon$ . Let  $\pi_\varepsilon$  be some optimal assignment between  $F_\varepsilon$  and  $G_\varepsilon$ . By cyclical monotonicity, the assignment  $\pi_\varepsilon$  is concentrated on a support  $\Gamma_\varepsilon$  that satisfies the property of no intersecting pairs. Since the smoothed distributions  $F_\varepsilon$  and  $G_\varepsilon$  are both atomless, this implies that any pairing  $(x, z) \in \Gamma_\varepsilon$  where  $x < z$  satisfies  $F_\varepsilon([x, z]) = G_\varepsilon([x, z])$ . In turn, by the definition of the measure of underqualification  $H$ , this implies  $H_\varepsilon(x) = H_\varepsilon(z)$  meaning that  $x$  and  $z$  are both part of the same layer  $A_\varepsilon^\ell$ . As a result, it follows that the support of the assignment  $\pi_\varepsilon$  is contained in the union of the support of all layers, or  $\Gamma_\varepsilon \subseteq \cup (A_\varepsilon^\ell)^2$ . Since all the supports  $\{A_\varepsilon^\ell\}$  are disjoint, the assignment  $\pi_\varepsilon|_{(A_\varepsilon^\ell)^2}$  transports between  $F_\varepsilon^\ell$  and  $G_\varepsilon^\ell$ . Since  $\pi_\varepsilon^\ell$  is an optimal assignment between workers  $F_\varepsilon^\ell$  and jobs  $G_\varepsilon^\ell$ , it follows that the cost of mismatch for layer  $\ell$  is greater under the assignment  $\pi_\varepsilon|_{(A_\varepsilon^\ell)^2}$ , that is,  $\int c d\pi_\varepsilon|_{(A_\varepsilon^\ell)^2} \geq \int c d\pi_\varepsilon^\ell$ . By summing over all layers  $1 \leq \ell \leq L$ , we can write that

$$\int c d\pi_\varepsilon = \sum_{1 \leq \ell \leq L} \int c d\pi_\varepsilon|_{(A_\varepsilon^\ell)^2} \geq \sum_{1 \leq \ell \leq L} \int c d\pi_\varepsilon^\ell = \int c d\left(\sum_{1 \leq \ell \leq L} \pi_\varepsilon^\ell\right).$$

Since  $\sum \pi_\varepsilon^\ell$  is a feasible assignment between the smoothed distributions  $F_\varepsilon$  and  $G_\varepsilon$ , and the mismatch cost is below the minimum mismatch cost, it follows that  $\sum \pi_\varepsilon^\ell$  must be an optimal assignment.

To conclude the proof it follows from our construction and the stability of the optimal assignment

<sup>30</sup>The choice of an open or closed interval  $(a_{\ell-1}, a_\ell)$  does not matter because the inverse of the boundary points is negligible with respect to the measure  $F_\varepsilon + G_\varepsilon$ .

that  $\pi_\varepsilon^\ell \rightarrow \pi^\ell$  and  $\sum \pi_\varepsilon^\ell \rightarrow \pi$  weakly. Thus,

$$\pi = \lim_{\varepsilon \rightarrow 0} \sum_{1 \leq \ell \leq L} \pi_\varepsilon^\ell = \sum_{1 \leq \ell \leq L} \lim_{\varepsilon \rightarrow 0} \pi_\varepsilon^\ell = \sum_{1 \leq \ell \leq L} \pi^\ell.$$

## A.5 Representation of Bellman Algorithm

We can follow [Delon, Salomon, and Sobolevski \(2012b\)](#) to further simplify solving the Bellman equation. Rather than considering all potential pairings  $k$  as in equation (12), it suffices to compare only two alternatives. Specifically, the minimal mismatch cost can be described by the simple recursive equation:

$$V_{i,j} = \min(c(s_i, s_j) + V_{i+1,j-1}, V_{i,j-2} + V_{i+2,j} - V_{i+2,j-2}), \quad (\text{A.1})$$

where  $j - i$  is odd so that the assignment problem contains an equal mass of workers and jobs. We now illustrate Bellman equation (A.1). Consider, say,  $V_{3,8}$  which is the cost of an optimal pairing for the skills in  $[s_3, s_8]$ . It is given as the minimum between two alternatives. The first alternative is  $c(s_3, s_8) + V_{4,7}$ . This is the cost of pairing  $s_3$  and  $s_8$  which is given by  $c(s_3, s_8)$  and the minimal cost of pairing in  $[s_4, s_7]$  given by the value  $V_{4,7}$ . The second alternative is  $V_{3,6} + V_{5,8} - V_{5,6}$ .

We use the Bellman equation (A.1), with initial conditions, to obtain the minimal cost of mismatch  $V_{i,j}$  for every pair  $(i, j)$  such that  $1 \leq i < j \leq 2n$ .<sup>31</sup>

**Proposition 4.** An optimal assignment between workers  $F$  and jobs  $G$  sums an optimal assignment in each layer, where the optimal assignment in each layer attains  $V_{1,2n_\varepsilon}$  in the Bellman equation (A.1).

**Proof to Proposition 4.** The proposition is proved in two steps. First, we show the minimum aggregate cost of mismatch is described by the Bellman equation. Given the Bellman equation, we then characterize the assignment that attains the minimum cost of mismatch as discussed in the main text.

**Cost of Mismatch.** We prove by induction that each mismatch cost  $V_{i,j}$  in (A.1) represents the minimal cost of mismatching between the distribution of workers  $F|_{[s_i, s_j]}$  and the distribution of jobs  $G|_{[s_i, s_j]}$ . Let us first check the base cases  $|j - i| \leq 3$ . Recall the initial conditions  $V_{i,i-1} = 0$  and  $V_{i+2,i-1} = -c(s_i, s_{i+1})$ .

First,

$$\begin{aligned} V_{i,i+1} &= \min(c(s_i, s_{i+1}) + V_{i+1,i}, V_{i,i-1} + V_{i+2,i+1} - V_{i+2,i-1}) \\ &= \min(c(s_i, s_{i+1}), c(s_i, s_{i+1})) = c(s_i, s_{i+1}). \end{aligned}$$

---

<sup>31</sup>For this Bellman formulation, the initial conditions are  $V_{i,i-1} = 0$  and  $V_{i+2,i-1} = -c(s_i, s_{i+1})$ . To observe the role of initial conditions consider the assignment problems of size two and four. When  $j - i = 1$ , we obtain  $V_{i,i+1} = \min(c(s_i, s_{i+1}) + V_{i+1,i}, V_{i,i-1} + V_{i+2,i+1} - V_{i+2,i-1}) = c(s_i, s_{i+1})$ . When  $j - i = 3$ , we obtain  $V_{i,i+3} = \min(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3} - V_{i+2,i+1}) = \min(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3})$ . We note that both are indeed identical to the simple specification in equation (12).

This is the correct cost of matching as there exists only a unique assignment with one worker and one job. Second,

$$\begin{aligned} V_{i,i+3} &= \min(c(s_i, s_{i+3}) + V_{i+1,i+2}, V_{i,i+1} + V_{i+2,i+3} - V_{i+2,i+1}) \\ &= \min(c(s_i, s_{i+3}) + c(s_{i+1}, s_{i+2}), c(s_i, s_{i+1}) + c(s_{i+2}, s_{i+3})). \end{aligned}$$

This is the correct cost of matching since there are only two possibilities for an assignment in this case, match either the first and the fourth or the first and the second.

Suppose we know  $V_{i,j}$  is the minimum cost of mismatch for every  $|j - i| < 2n - 1$ . For the induction step, we show this also holds at  $|j - i| = 2n - 1$ , or the complete alternating assignment problem, where:

$$V_{1,2n} = \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2}).$$

Denote by  $F_t = F|_{[3,2n-2]}$  and  $G_t = G|_{[3,2n-2]}$  the worker and job distributions on a thin support, and we let  $F_a = F + F_t$  and  $G_a = G + G_t$  denote the addition of the original worker and job distribution and their respective distributions on the thin support. In order to prove that the Bellman equation (A.1) holds, we show that both directions hold, that is,  $V_{1,2n} \leq \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2})$  and  $V_{1,2n} \geq \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2})$ .

We establish the  $\leq$  direction by showing  $V_{1,2n} \leq c(s_1, s_{2n}) + V_{2,2n-1}$  as well as  $V_{1,2n} \leq V_{1,2n-2} + V_{3,2n} - V_{3,2n-2}$ . The first inequality follows immediately. Take the optimal assignment in the interval  $[s_2, s_{2n-1}]$ , and pair the remaining elements,  $s_1$  and  $s_{2n}$ , which induces a cost of mismatch  $c(s_1, s_{2n}) + V_{2,2n-1}$ . Since this is a feasible assignment for all workers and jobs on the interval  $[s_1, s_{2n}]$  the minimum mismatch cost on this interval is below the mismatch cost of the constructed assignment, or  $V_{1,2n} \leq c(s_1, s_{2n}) + V_{2,2n-1}$ .

To prove  $V_{1,2n} + V_{3,2n-2} \leq V_{1,2n-2} + V_{3,2n}$ , consider the assignment problem between workers  $F_a$  and jobs  $G_a$ . By Lemma 4, the minimal cost of mismatch is  $V_{1,2n} + V_{3,2n-2}$ . On the other hand, combining the two optimal assignments between  $F|_{[1,2n-2]}$  and  $G|_{[1,2n-2]}$  as well as between  $F|_{[3,2n]}$  and  $G|_{[3,2n]}$  gives a feasible assignment between workers  $F_c$  and jobs  $G_c$ . This shows that the optimal assignment cost between workers  $F_c$  and jobs  $G_c$  is less than  $V_{1,2n-2} + V_{3,2n}$ , which concludes the proof of the  $\leq$  direction.

Next, we establish the  $\geq$  direction, or  $V_{1,2n} \geq \min(c(s_1, s_{2n}) + V_{2,2n-1}, V_{1,2n-2} + V_{3,2n} - V_{3,2n-2})$ . It suffices to establish that if  $V_{1,2n} < c(s_1, s_{2n}) + V_{2,2n-1}$  then  $V_{1,2n} + V_{3,2n-2} \geq V_{1,2n-2} + V_{3,2n}$ .<sup>32</sup> Suppose  $V_{1,2n} < c(s_1, s_{2n}) + V_{2,2n-1}$ . This implies that an optimal assignment between workers  $F$  and jobs  $G$  does not contain the pairing  $(s_1, s_{2n})$ . Suppose that  $s_1$  is paired with  $s_j$  where  $j < 2n$ . The absence of

---

<sup>32</sup>Logically, we prove  $V \geq \min(A, B)$  by showing that  $V \geq B$  or  $V \geq A$ . We show this by proving the statement that if  $V < A$ , then  $V \geq B$ .

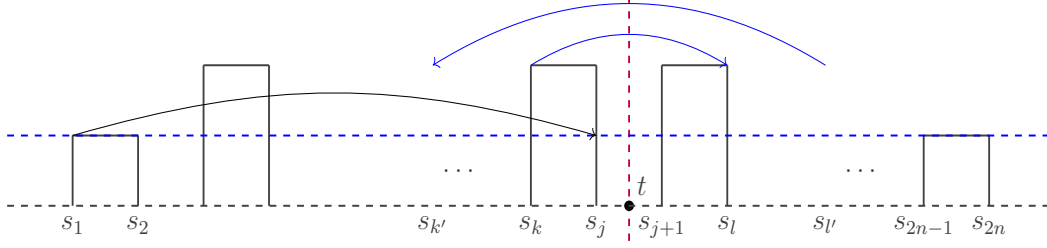


Figure A.2: Layer Decomposition with  $F_a$  and  $G_a$

The blue dashed line indicates the layer decomposition given a worker distribution  $F_a$  and a job distribution  $G_a$ . The blue arcs indicate a nested pair in the optimal assignment  $\pi_t$  between a worker distribution  $F_t$  and a job distribution  $G_t$ . These blue arcs will intersect the arc corresponding to the pair  $(s_1, s_j)$  in the optimal assignment between distributions  $F$  and  $G$ , which gives a contradiction.

intersecting pairs implies that the optimal assignment is concentrated on square subsets to the left and right of skill  $s_j$ , that is,  $\text{supp } \pi \subseteq [s_1, s_j]^2 \cup [s_{j+1}, s_{2n}]^2$ . We denote the skill threshold in between  $s_j$  and  $s_{j+1}$  by  $t = (s_j + s_{j+1})/2$ .

We next claim that an optimal assignment  $\pi_t$  between the thin distribution of workers  $F_t$  and the thin distribution of jobs  $G_t$  has no pairing with a corresponding interval that contains the point  $t$ . By contradiction, suppose there does exist a pair  $(s_k, s_l)$  crossing the threshold  $t$  in  $\pi_t$ . Since the masses to the left and the right of  $t$  are equal, there also exists a pairing  $(s_{k'}, s_{l'})$  that contains  $(s_k, s_l)$ . This is illustrated by the blue arrows in Figure A.2. Consider an optimal assignment  $\pi_a$  between the added distribution workers  $F_a$  and jobs  $G_a$ . By Lemma 4,  $\pi_a = \pi + \pi_t$ , and hence assignment  $\pi_a$  contains the pairings  $(s_1, s_j)$ ,  $(s_k, s_l)$ , and  $(s_{k'}, s_{l'})$ . This contradicts the property of no intersecting pairs. Since the optimal assignment  $\pi_t$  does not contain a pair whose interval contains  $t$ , it follows that  $\pi_t$  is concentrated on the support  $[s_1, s_j]^2 \cup [s_{j+1}, s_{2n-2}]^2$ .

We combine the optimal assignment between workers  $F$  and jobs  $G$  with the optimal assignment between thinned distributions for workers  $F_t$  and jobs  $G_t$  to obtain the result. We construct a feasible assignment  $\pi_{a,1}$  for the problem on  $[s_1, s_{2n-2}]$  by combining the optimal assignment  $\pi$  on  $[s_1, s_j]$  with the optimal assignment  $\pi_t$  on the interval  $[s_{j+1}, s_{2n-2}]$ , so that  $\pi_{a,1} = \pi|_{[s_1, s_j]^2} + \pi_t|_{[s_{j+1}, s_{2n-2}]^2}$ . Analogously, we construct a feasible assignment  $\pi_{a,2}$  for the problem on  $[s_3, s_{2n}]$  by combining the optimal assignment  $\pi$  on  $[s_{j+1}, s_{2n}]$  with the optimal assignment  $\pi_t$  on the interval  $[s_3, s_j]$  which generates  $\pi_{a,2} = \pi_t|_{[s_3, s_j]^2} + \pi|_{[s_{j+1}, s_{2n}]^2}$ . Since the constructed assignments are feasible for the assignment problems on  $[s_1, s_{2n-2}]$  and  $[s_3, s_{2n}]$ , respectively, we have  $\mathcal{C}(\pi_{a,1}) \geq V_{1,2n-2}$  and  $\mathcal{C}(\pi_{a,2}) \geq V_{3,2n}$ . Finally, we observe that the addition of  $\pi_{a,1}$  and  $\pi_{a,2}$  solves the assignment problem with worker distribution  $F_a$  and job distribution  $G_a$ , such that  $\pi_{a,1} + \pi_{a,2} = \pi_a$ . These together yield  $V_{1,2n-2} + V_{3,2n} \leq \mathcal{C}(\pi_{a,1}) + \mathcal{C}(\pi_{a,2}) = \mathcal{C}(\pi_a) = V_{1,2n} + V_{3,2n-2}$ ,

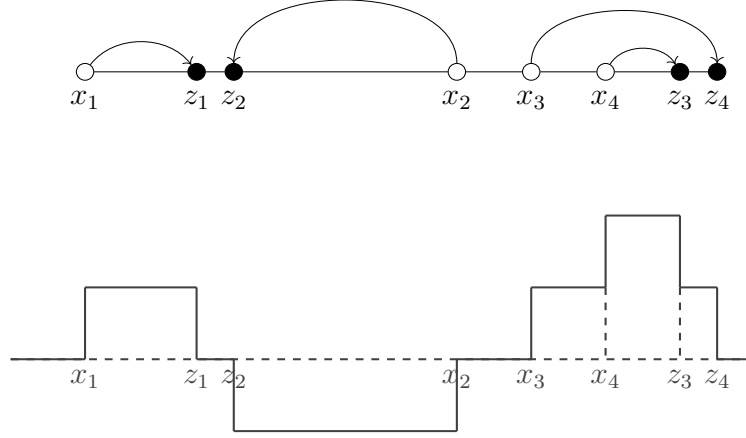


Figure A.3: Layered Positive Assignment

Figure A.3 illustrates the layered positive assignment with four workers and four jobs. The layered positive assignment is presented in the top panel. We use the measure of underqualification  $H$  in the bottom panel to construct this assignment. The decomposition yields two layers with only one worker and one job  $(x_4, z_3)$  and  $(x_2, z_2)$  for which the assignment is immediate. The third layer contains skill levels  $(x_1, z_1, x_3, z_4)$ . The alternating assignment pairs  $(x_1, z_1)$  and  $(x_3, z_4)$ .

completing the  $\geq$  direction for (A.1). This concludes the proof.

## A.6 Other Examples

In this appendix, we provide additional illustrations of our theory.

**Close to Linear Mismatch Costs.** This example emphasizes the dependence of the optimal sorting on the cost function. In particular, we consider the assignment problem when  $\zeta$  is close to 1, that is, when the cost of mismatch is close to linear in the distance between the worker skill and the job.<sup>33</sup>

First, we maximize the number of perfect pairings. Second, we decompose the assignment problem into layers  $0 \leq \ell \leq L$ . Third, when  $\zeta$  is close to one, the optimal assignment within each layer is simple. Specifically, we show in Proposition 5 below that the optimal assignment within each layer is positive sorting which we denote by  $\pi_\ell^+$ .

The simplicity of the optimal assignment within layer enables us to provide a simple characterization of the overall solution in this case. The solution to the full assignment problem is given by the sum of the positive assignments within each layer. We refer to this assignment as the layered positive assignment denoted by  $\pi^+ = \sum \pi_\ell^+$ .

Figure A.3 illustrates the layered positive assignment with four workers and four jobs. The optimal

<sup>33</sup>The linear case of mismatch ( $\zeta = 1$ ) is the original formulation of the optimal transport problem due to Monge (1781) and has been well studied (Rachev and Rüschendorf (1998); Villani (2003)). See a discussion below.

assignment is given by the top panel. To construct this assignment, we use the measure of underqualification  $H$  in the bottom panel. The measure of underqualification induces three independent layers. The decomposition yields two layers with only one worker and one job  $(x_4, z_3)$  and  $(x_2, z_2)$  for which there is no other option but to pair them. The third layer contains skill levels  $(x_1, z_1, x_3, z_4)$ . The assignment positively pairs  $(x_1, z_1)$  and  $(x_3, z_4)$ .

We summarize the lessons from this subsection in Proposition 5.

**Proposition 5.** Given a worker distribution  $F$  and a job distribution  $G$ , there exists  $\bar{\zeta} < 1$  such that for any  $\zeta \in [\bar{\zeta}, 1]$ , the layered positive assignment  $\pi^+$  is optimal.

*Proof.* After maximizing perfect pairs, by Lemma 2, we can restrict attention to assignments between worker and job distributions that are supported on disjoint sets. This means that the distributions  $F$  and  $G$  are supported on a finite set  $S$ , and we denote by  $\delta$  the smallest pairwise distance between elements in  $S$ , and by  $D$  the largest pairwise distance between elements.<sup>34</sup>

We show there exists  $0 < \bar{\zeta} < 1$  such that for any  $\zeta \in [\bar{\zeta}, 1]$ , the layered positive assignment  $\pi$  is optimal with respect to the mismatch cost  $c(x, z)$ . To prove the result, consider  $\bar{\zeta}$  such that for any  $\delta \leq \delta' \leq D' \leq D$ :

$$2^{1-\bar{\zeta}}(D' - \delta')^{\bar{\zeta}} \leq \delta'^{\bar{\zeta}} + D'^{\bar{\zeta}}. \tag{A.2}$$

Such a  $\bar{\zeta}$  exists because equation (A.2) with  $D' - \delta' \leq \delta$  is vacuously true while otherwise both sides of (A.2) are uniformly continuous in  $\bar{\zeta}$  on  $\{(\delta', D') : \delta \leq \delta' \leq D' \leq D, D' - \delta' > \delta\}$  and “ $<$ ” holds uniformly when  $\bar{\zeta} = 1$ . Consider  $\zeta \in [\bar{\zeta}, 1]$ . It suffices to prove that the optimal assignment within a layer does not contain any nested arc for the mismatch cost  $|x - z|^\zeta$ .

By the principle of layering Lemma 4, we decompose both measures  $F_n$  and  $G_n$  into layers. Here we assume without loss of generality that the lowest skill worker comes before the lowest skill job:  $x_1 < z_1$ . On each layer there are  $2k$  equal masses on the skill levels  $x_1 < z_1 < \dots < x_k < z_k$  in  $S$  that are at least  $\delta$  apart. The maximum distance within this layer is exceeded by  $D > z_k - x_1$ . Let  $\{x_j\}_{1 \leq j \leq k}$  be the locations of mass on the layer for workers  $F_n$ , and let  $\{z_j\}_{1 \leq j \leq k}$  be the locations of mass on the layer for jobs  $G_n$ . By contradiction, suppose that the optimal assignment within this layer instead does contain a nested arc, so it holds for some  $x_1 \leq u < v < s < t \leq z_k$  that

$$(t - u)^\zeta + (s - v)^\zeta \leq (t - s)^\zeta + (v - u)^\zeta.$$

---

<sup>34</sup>This proof can be extended to the case of continuous distributions when  $F$  and  $G$  are compactly supported with the measure of underqualification  $H := F - G$  satisfying that both  $H$  and  $-H$  have finitely many local maxima and those maxima are strictly above zero.

Since by concavity of the function  $x^\zeta$  for  $x \geq 0$ ,

$$(t-s)^\zeta + (v-u)^\zeta \leq 2 \left( \frac{1}{2}(t-s) + \frac{1}{2}(v-u) \right)^\zeta = 2^{1-\zeta} ((t-u) - (s-v))^\zeta.$$

Putting  $t-u = D'$  and  $s-v = \delta'$  in (A.2) leads to a contradiction. Hence the optimal assignment within this layer does not contain any nested arc for  $\zeta \in [\bar{\zeta}, 1]$ .  $\square$

The implication of the proposition is that for mismatch power values close to one, the solution can be directly constructed by constructing the measure of underqualification, and constructing the positive alternating assignment by layer.<sup>35</sup> While this assignment generates positive sorting within each layer, we emphasize this does not imply positive sorting overall as we demonstrated in the example of two binomial distributions. In particular, there is negative sorting since  $x_3$  is sorted with  $z_4$  but  $x_4$  is sorted with  $z_3$ .

**Linear Cost of Mismatch.** The last example naturally begs the question what is optimal when the cost of mismatch is linear in the distance between worker skills and job difficulties. The linear case is exactly the specification of the mismatch cost prior to the technology choice in Section 2.1. This formulation is in fact the original formulation of the optimal transport problem due to Monge (1781) and has been well studied (Rachev and Rüschendorf (1998); Villani (2003)). When the cost of mismatch is linear, an optimal assignment is not unique. To illustrate this, we argue that both the Bellman equation and positive sorting are optimal.

To see that the Bellman equation also delivers the optimal assignment in the case of linear costs of mismatch, we note that the derivation of the Bellman equation did not require that the cost of mismatch was strictly concave, only that the cost of mismatch was concave. As a result, the same approach as before characterizes an optimal assignment.

Another solution to the assignment problem with linear costs of mismatch is positive sorting. In this case, as with convex costs of the distance, the production function is supermodular, and it follows that a positively sorted assignment is optimal.

## A.7 Local Hierarchical Algorithm

We propose a new algorithm specifically tailored to the model of composite sorting, which has two distinct merits. First, it is more efficient than the existing generic algorithms, as shown in Appendix B.3. Second, and more importantly, this new algorithm reveals a hierarchical structure of the dual potential functions,

---

<sup>35</sup>Juillet (2020) calls the layered positive assignment an excursion coupling, and shows that the layered positive assignment is the limit of some optimal couplings as  $\zeta \rightarrow 1^-$ . We complement their result by proving the existence of a threshold  $\bar{\zeta}$  beyond which the layered positive assignment is optimal for our environment.

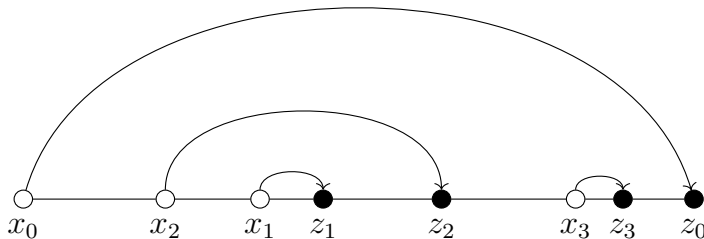


Figure A.4: Subpairs and Processing the Pairs

Figure A.4 illustrates the definition of the subpairs and the recursive computation structure in a setting of 4 workers and 4 jobs. For instance, The pair  $(x_1, z_1)$  is a subpair of  $(x_2, z_2)$  but not of  $(x_0, z_0)$ . We process the pairs in the order  $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 0$  to compute local dual optimizers  $\phi_{[x_1, z_1]}$ ,  $\phi_{[x_2, z_2]}$ ,  $\phi_{[x_3, z_3]}$ , and  $\phi_{[x_0, z_0]}$ . Note that the order of processing pairs 2 and 3 is irrelevant.

highlighting an implication of the absence of intersecting pairs to dual optimizers. This hierarchical structure means that a local dual optimizer within each region, which is the interval between two points in a pair, is computed from the local dual optimizers on its subregions. This hierarchical structure leads to the following separation property of the global dual optimizer  $\phi$ : for any two points  $s$  and  $s'$  in one region,  $\phi(s) - \phi(s')$  can be computed only based on points within this region, and thus it does not depend on points outside. We will describe this structure in detail and our algorithm below.

Recall that in our setting, we consider a problem with  $n$  workers with skill levels in  $X$  and  $n$  jobs with difficulty levels in  $Z$ , where  $X$  and  $Z$  are disjoint sets. We denote by  $S = X \cup Z$  the set of all skill levels. In what follows, when we mention an interval  $[x, z]$ , we meant for the set of points between  $x$  and  $z$ , regardless of whether  $x < z$  or  $z < x$ .

Our algorithm relies on recursive computations of  $\phi$  constrained on smaller subsets of  $S$ . To explain such a recursive procedure, we need some preparation by introducing the notion of subpairs. A pair  $(x, z)$  is called a subpair of the pair  $(x_0, z_0)$  if  $(x, z)$  is a non-nested pair inside the interval  $[x_0, z_0]$  that is not equal to  $(x_0, z_0)$ . In Figure A.4, the pairs  $(x_2, z_2)$  and  $(x_3, z_3)$  are subpairs of the pair  $(x_0, z_0)$ . The pair  $(x_1, z_1)$  is a subpair of  $(x_2, z_2)$  but not of  $(x_0, z_0)$  as it is nested in  $(x_2, z_2)$ . For  $(x, z) \in \Gamma_\pi$ , let  $X_{[x, z]}$  denote the set of all points in  $X$  between  $x$  and  $z$  inclusive of the boundary, and similarly for  $Z_{[x, z]}$ .

We process each pair  $(x_0, z_0) \in \Gamma_\pi$  sequentially in a certain order described below to get a local dual optimizer on  $[x_0, z_0]$ , that is, a function  $\phi_{[x_0, z_0]}$  such that for any  $(x, z) \in X_{[x_0, z_0]} \times Z_{[x_0, z_0]}$ ,  $\phi_{[x_0, z_0]}(x) - \phi_{[x_0, z_0]}(z) \leq c(x, z)$  which holds with equality when  $(x, z) \in \Gamma_\pi$ . We observe that this property is preserved if  $\phi_{[x_0, z_0]}$  is shifted by any constant  $a \in \mathbb{R}$ .

A simple illustrative example is given in Figure A.4. We start with the pair  $(x_1, z_1)$  and found a local dual optimizer  $\phi_{[x_1, z_1]}$ . Here, the local dual optimizer is any function  $\phi_{[x_1, z_1]}$  satisfying  $\phi_{[x_1, z_1]}(x_1) -$



$\phi_{[x_1, z_1]}(z_1) = c(x_1, z_1)$  as the only pair contained in  $[x_1, z_1]$  is  $(x_1, z_1)$ . We then process the pair  $(x_2, z_2)$ . A local dual optimizer on  $[x_2, z_2]$  is a function  $\phi_{[x_2, z_2]}$  satisfying  $\phi_{[x_2, z_2]}(x) - \phi_{[x_2, z_2]}(z) \leq c(x, z)$  for  $(x, z) \in \{(x_1, z_1), (x_2, z_2), (x_1, z_2), (x_2, z_1)\}$  with equality holding for  $(x, z) = (x_1, z_1)$  and  $(x, z) = (x_2, z_2)$ . We proceed with the subpair  $(x_3, z_3)$  and finally the remaining pair  $(x_0, z_0)$ .

Below is a full description of the algorithm, where  $\phi$ ,  $\phi_i$ ,  $x_i$  and  $z_i$  are local variables which vary across each iteration, and  $\phi_{[x, z]}$  for  $(x, z) \in \Gamma_\pi$  are global variables which are the output of the algorithm.

1. Pick any pair  $(x_0, z_0) \in \Gamma_\pi$  that has not been processed such that all subpairs of  $(x_0, z_0)$  have been processed. Let  $(x_1, z_1), \dots, (x_p, z_p)$  be the subpairs ordered in dictionary order<sup>36</sup> with potential functions  $\phi_i := \phi_{[x_i, z_i]}$  for  $i = 1, \dots, p$ .
2. If  $p = 0$ , then let  $\phi(z_0) = 0$  and  $\phi(x_0) = c(x_0, z_0)$ .
3. If  $p \geq 1$ , then continue with the following sub-steps.

(a) If  $p > 1$ , let  $(\beta_2, \dots, \beta_p) \in \mathbb{R}^{p-1}$  be a solution to the inequality system

$$\max(c_{00} - c_{0n} - c_{m0}, -c_{mn}) + c_{nn} \leq \sum_{k=n+1}^m \beta_k \leq \min(c_{0m} + c_{n0} - c_{00}, c_{nm}) - c_{mm} \quad (\text{A.3})$$

for all  $1 \leq n < m \leq p$ , where  $c_{ij} = c(x_i, z_j)$  for  $i, j \in \{1, \dots, p\}$ . We show the existence of such  $(\beta_2, \dots, \beta_p)$  in Lemma 6.

(b) For  $i = 1, \dots, p$ , let

$$\phi(s) = \phi_i(s) + \sum_{k=i+1}^p \beta_k + \phi_p(x_p) - \phi_i(x_i) \quad (\text{A.4})$$

for  $s \in X_{[x_i, z_i]}$  or  $s \in Z_{[x_i, z_i]}$ . The above sum is 0 if  $i = p$ .

(c) Define  $\phi(x_0)$  and  $\phi(z_0)$  by the following equations

$$\phi(z_0) = \begin{cases} \max_{i \in \{1, \dots, p\}} (\phi(x_i) - c(x_i, z_0)) & \text{if } x_1 \neq x_0, \\ \min_{i \in \{1, \dots, p\}} (\phi(z_i) + c(x_0, z_i)) - c(x_0, z_0) & \text{elsewhere;} \end{cases} \quad (\text{A.5})$$

$$\phi(x_0) = \phi(z_0) + c(x_0, z_0). \quad (\text{A.6})$$

4. Let  $\phi_{[x_0, z_0]}$  be equal to  $\phi$ .
5. Return to step 1 with the next pair to process, or terminate if all pairs have been processed.

<sup>36</sup>A vector  $(a_1, \dots, a_d)$  is smaller than a vector  $(b_1, \dots, b_d)$  in dictionary order if either  $(a_1, \dots, a_d) = (b_1, \dots, b_d)$  or there exists  $p = 1, \dots, d$  such that  $a_i = b_i$  for  $i = 1, \dots, p-1$  and  $a_{p+1} < b_{p+1}$ .

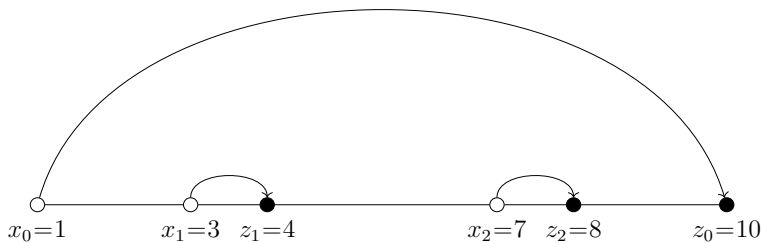


Figure A.5: Illustration of the Algorithm

Figure A.5 illustrates the construction of the dual potential  $\phi$  in the example where  $x_0 = 1$ ,  $x_1 = 3$ ,  $x_2 = 7$ ,  $z_0 = 10$ ,  $z_1 = 4$ ,  $z_2 = 8$ , with the cost function given by  $c(x, z) = \sqrt{|x - z|}$ . The pairs are processed in the order of  $(x_1, z_1)$ ,  $(x_2, z_2)$ ,  $(x_0, z_0)$ . The output of the algorithm is  $\phi(x_1) = 5 - 2\sqrt{3}$ ,  $\phi(z_1) = 4 - 2\sqrt{3}$ ,  $\phi(x_2) = 1$ ,  $\phi(z_2) = 0$ ,  $\phi(x_0) = 4 - \sqrt{3}$ , and  $\phi(z_0) = 1 - \sqrt{3}$ .

We note from step 3(b) that for  $s, s' \in X_{[x_i, z_i]} \cup Z_{[x_i, z_i]}$ , we have  $\phi(s) - \phi(s') = \phi_i(s) - \phi_i(s')$ . This means that after many iterations, the value of  $\phi(s) - \phi(s')$  does not change, and therefore it depends only on points in the region  $[x_i, z_i]$ .

There is some flexibility in the above algorithm in choosing the order of processing the pairs in step 1 and in solving for  $(\beta_2, \dots, \beta_p)$  in (A.3) in step 3(a). This flexibility is natural as the dual optimizers are not unique in general. The order of processing the pairs does not affect the output of the algorithm because each  $\phi_{[x_0, z_0]}$  only depends on the local dual optimizers of its subpairs, which are all processed before this pair. A default order is to always choose the unprocessed pair  $(x_0, z_0)$  with the smallest  $x_0$  satisfying the condition in step 1. On the other hand, the choice of  $(\beta_2, \dots, \beta_p)$  does affect the output of the algorithm. As a default,  $(\beta_2, \dots, \beta_p)$  can be chosen as the solution of (A.3) which is the smallest in dictionary order.<sup>37</sup> In this way, we obtain a unique output of the algorithm. Nevertheless, in the next result, we will show that a dual potential  $\phi$  is obtained from the algorithm with arbitrary choices of  $(\beta_2, \dots, \beta_p)$  in each iteration.

We illustrate the construction of the dual potential  $\phi$  in a simple example where  $X = \{1, 3, 7\}$ ,  $Z = \{4, 8, 10\}$ , and  $x_0 = 1$ ,  $x_1 = 3$ ,  $x_2 = 7$ ,  $z_0 = 10$ ,  $z_1 = 4$ ,  $z_2 = 8$ , with the cost function given by  $c(x, z) = \sqrt{|x - z|}$ . By the principles established in Section 3, the optimal assignment  $\pi$  pairs  $x_i$  with  $z_i$  for all  $i = \{0, 1, 2\}$  as displayed in Figure A.5. Our goal is to construct a function  $\phi : S \rightarrow \mathbb{R}$  such that  $\phi(x_i) - \phi(z_j) \leq c(x_i, z_j)$  for all  $i$  and  $j$  in  $\{0, 1, 2\}$  with equality holding when  $i = j$ .

Following step 1 of the algorithm, we first process the pairs  $(x_1, z_1)$  and  $(x_2, z_2)$ , as they do not have any subpair. Define  $\phi_1(x_1) = \phi_2(x_2) = 1$  and  $\phi_1(z_1) = \phi_2(z_2) = 0$ , which are the local dual optimizers on  $[x_1, z_1]$  and  $[x_2, z_2]$ , respectively. We now process the pair  $(x_0, z_0)$ . According to (A.3), the second step

<sup>37</sup>Note that such a smallest solution always exists since  $(\beta_2, \dots, \beta_p)$  satisfying (A.3) lies in a compact region.

is to find  $\beta_2 \in \mathbb{R}$  such that

$$\max(c_{00} - c_{01} - c_{20}, -c_{21}) + c_{11} \leq \beta_2 \leq \min(c_{02} + c_{10} - c_{00}, c_{12}) - c_{22}. \quad (\text{A.7})$$

Plugging in the values of  $x_i, z_i$ , we get

$$\max(c_{00} - c_{01} - c_{20}, -c_{21}) + c_{11} = 4 - 2\sqrt{3} \approx 0.536; \quad \min(c_{02} + c_{10} - c_{00}, c_{12}) - c_{22} = \sqrt{3} - 1 \approx 0.732.$$

Following the default choice of choosing the smallest  $\beta_2$ , we set  $\beta_2 = 4 - 2\sqrt{3}$ .<sup>38</sup> Following step 3(b) of the algorithm, we have

$$\begin{aligned} \phi(x_2) &= \phi_2(x_2) + 0 + \phi_2(x_2) - \phi_2(x_2) = c_{22} = 1; \\ \phi(z_2) &= \phi_2(z_2) + 0 + \phi_2(x_2) - \phi_2(x_2) = 0; \\ \phi(x_1) &= \phi_1(x_1) + \beta_2 + \phi_2(x_2) - \phi_1(x_1) = c_{22} + \beta_2 = 5 - 2\sqrt{3}; \\ \phi(z_1) &= \phi_1(z_1) + \beta_2 + \phi_2(x_2) - \phi_1(x_1) = c_{22} - c_{11} + \beta_2 = 4 - 2\sqrt{3}. \end{aligned}$$

Further, step 3(c) yields

$$\begin{aligned} \phi(z_0) &= \max(\phi(x_1) - c_{10}, \phi(x_2) - c_{20}) = 1 - \sqrt{3}; \\ \phi(x_0) &= \phi(z_0) + c_{00} = 4 - \sqrt{3}. \end{aligned}$$

One can verify numerically that the function  $\phi$  defined above satisfies the conditions of a dual optimizer.

## A.8 Proof to Theorem 2

We prove Theorem 2 in two parts. First, we prove there exists a solution  $(\beta_2, \dots, \beta_p)$  to (A.3), in Lemma 6. Second, we prove that the function  $\phi$  defined in (A.4)-(A.6) is indeed a local dual optimizer on  $S_{[x_0, z_0]}$ .

**Lemma 6.** Suppose  $(x_1, z_1), \dots, (x_p, z_p)$  are ordered subpairs of pair  $(x_0, z_0)$  in the optimal assignment  $\pi$ . Define  $c_{ij} := c(x_i, z_j)$ . Then the system of inequalities, where for all  $1 \leq n < m \leq p$ :

$$\max(c_{00} + c_{nn} - c_{0n} - c_{m0}, c_{nn} - c_{mm}) \leq \sum_{k=n+1}^m \beta_k \leq \min(c_{0m} + c_{n0} - c_{00} - c_{mm}, c_{nm} - c_{mm}) \quad (\text{A.8})$$

admits a solution  $(\beta_2, \dots, \beta_p)$ .

---

<sup>38</sup>Any value in between  $4 - 2\sqrt{3}$  and  $\sqrt{3} - 1$  will produce a dual potential. That the left-hand side of (A.7) is no more than the right-hand side of (A.7) is not a coincidence. We will prove that the system of inequalities (A.3) always admits a solution in Lemma 6 in Appendix A.7. The set of solutions is always non-empty, although, intuitively, the larger  $|S|$  is, the less freedom for the dual optimizer one has.

*Proof of Lemma 6.* We use Farkas' Lemma to prove our existence result, specifically, to have a necessary and sufficient condition for a system of linear inequalities to have a solution. We state Farkas' Lemma for completeness.

**Lemma 7.** *Farkas.* Let  $A$  be a  $d_1 \times d_2$  matrix,  $b \in \mathbb{R}^{d_1}$ , and let  $x = (x_1, \dots, x_{d_2})^\top$  be a set of real-valued variables. Then the system  $Ax \geq b$  allows a set of solutions if and only if for any  $y \in [0, \infty)^{d_1}$  such that  $y^\top A = 0$ , it holds  $y^\top b \leq 0$ .

We aim to show that equation (A.8) admits a solution  $(\beta_2, \dots, \beta_p) \in \mathbb{R}^{p-1}$ . We observe that we can think of (A.8) equivalently as the following set of inequalities:

$$\begin{aligned} \sum_{k=n+1}^m \beta_k &\geq c_{00} + c_{nn} - c_{0n} - c_{m0} \\ \sum_{k=n+1}^m \beta_k &\geq c_{nn} - c_{mn} \\ - \sum_{k=n+1}^m \beta_k &\geq c_{mm} - c_{nm} \\ - \sum_{k=n+1}^m \beta_k &\geq c_{00} + c_{mm} - c_{0m} - c_{n0} \end{aligned}$$

for all  $1 \leq n < m \leq p$ . All inequalities implied by (A.8) are thus linear in the variables  $(\beta_2, \dots, \beta_p)$ . Matrix  $A$  is given by columns with values  $(-1, 0, +1)$ , while vector  $b$  is governed by the costs  $c$ .

By Lemma 7 it suffices to prove the following.<sup>39</sup> For any set of non-negative weights  $(\lambda_{mn}^+, \lambda_{mn}^-, \omega_{mn}^+, \omega_{mn}^-)$ ,  $1 \leq n < m \leq p$  on each of the inequalities above such that

$$\sum_{1 \leq n < m \leq p} (\lambda_{mn}^+ + \omega_{mn}^+) \sum_{k=n+1}^m \beta_k = \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \omega_{mn}^-) \sum_{k=n+1}^m \beta_k, \text{ for all } (\beta_2, \dots, \beta_p), \quad (\text{A.9})$$

it holds that

$$\begin{aligned} \sum_{1 \leq n < m \leq p} \left( \lambda_{mn}^- (c_{00} + c_{nn} - c_{0n} - c_{m0}) + \omega_{mn}^- (c_{nn} - c_{mn}) \right) \\ \leq \sum_{1 \leq n < m \leq p} \left( \lambda_{mn}^+ (c_{0m} + c_{n0} - c_{00} - c_{mm}) + \omega_{mn}^+ (c_{nm} - c_{mm}) \right). \end{aligned} \quad (\text{A.10})$$

We start by simplifying equations (A.9) and (A.10). We first simplify equation (A.9). Since (A.9) has to hold for all  $(\beta_2, \dots, \beta_p)$ , we note that the coefficient on each  $\beta_k$  has to equal zero. For each  $2 \leq k \leq p$ ,

---

<sup>39</sup>Equation (A.9) is the analog of  $y^\top A = 0$  in the statement of Farkas' Lemma. Specifically, we use  $y^\top A = 0$  if and only if  $y^\top Az = 0$  for all  $z \in \mathbb{R}^{d_1}$ . Applied to our setting, where  $\beta$  takes the position of  $x$  in Farkas' Lemma, this states that the weighted sum of all left-hand sides in the system of inequalities equals zero. Equation (A.10) below is similarly the analog of  $y^\top b \leq 0$  in the statement of Farkas' Lemma.

equating the coefficients for some  $\beta_k$  requires

$$\sum_{m,n} (\lambda_{mn}^+ + \omega_{mn}^+) = \sum_{m,n} (\lambda_{mn}^- + \omega_{mn}^-), \quad (\text{A.11})$$

where we sum over all  $(m, n)$  satisfying  $1 \leq n < k \leq m \leq p$ , that is, we sum over all equations where  $\beta_k$  appears. Furthermore, subtracting equation (A.11) evaluated at  $k$  from equation (A.11) evaluated at  $k + 1$  yields:

$$\sum_{k < m \leq p} (\lambda_{mk}^+ + \omega_{mk}^+) - \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) = \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+) - \sum_{1 \leq n < k} (\lambda_{kn}^- + \omega_{kn}^-), \quad (\text{A.12})$$

for all  $2 \leq k < p$ .

We next simplify (A.10). Rearranging (A.10) by collecting terms by coefficients in front of  $c_{ij}$  leads to the equivalent form:

$$\begin{aligned} & \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \lambda_{mn}^+) c_{00} + \sum_{1 < m \leq p} (\lambda_{m1}^- + \omega_{m1}^-) c_{11} + \sum_{1 \leq n < p} (\lambda_{pn}^+ + \omega_{pn}^+) c_{pp} \\ & \quad + \sum_{k=2}^{p-1} \left( \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) + \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+) \right) c_{kk} \\ & \leq \sum_{1 < m \leq p} \lambda_{m1}^- c_{01} + \sum_{1 \leq n < p} \lambda_{pn}^+ c_{0p} + \sum_{1 \leq n < p} \lambda_{pn}^- c_{p0} + \sum_{1 < m \leq p} \lambda_{m1}^+ c_{10} \\ & \quad + \sum_{k=2}^{p-1} \left( \sum_{1 \leq n < k} \lambda_{kn}^+ + \sum_{k < m \leq p} \lambda_{mk}^- \right) c_{0k} + \left( \sum_{1 \leq n < k} \lambda_{kn}^- + \sum_{k < m \leq p} \lambda_{mk}^+ \right) c_{k0} \\ & \quad + \sum_{1 \leq n < m \leq p} \omega_{mn}^- c_{mn} + \sum_{1 \leq n < m \leq p} \omega_{mn}^+ c_{nm}, \end{aligned} \quad (\text{A.13})$$

where the left-hand side of the inequality collects all ‘‘diagonal’’ elements, and the right-hand side collects all other elements.

Our next step in proving that equation (A.13) indeed holds, is to show that both sides of equation (A.13) represent transport costs of an assignment between a measure of workers  $\tilde{F}$  and a measure of jobs  $\tilde{G}$ . Specifically, consider the assignment problem between a measure  $\tilde{F}$  and a measure  $\tilde{G}$ , satisfying:

$$\begin{aligned} \tilde{F} = & \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \lambda_{mn}^+) \delta_{x_0} + \sum_{1 < m \leq p} (\lambda_{m1}^- + \omega_{m1}^-) \delta_{x_1} + \sum_{1 \leq n < p} (\lambda_{pn}^+ + \omega_{pn}^+) \delta_{x_p} \\ & + \sum_{k=2}^{p-1} \left( \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) + \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+) \right) \delta_{x_k}, \end{aligned} \quad (\text{A.14})$$

and, similarly,

$$\begin{aligned} \tilde{G} = & \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \lambda_{mn}^+) \delta_{z_0} + \sum_{1 < m \leq p} (\lambda_{m1}^- + \omega_{m1}^-) \delta_{z_1} + \sum_{1 \leq n < p} (\lambda_{pn}^+ + \omega_{pn}^+) \delta_{z_p} \\ & + \sum_{k=2}^{p-1} \left( \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) + \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+) \right) \delta_{z_k}. \end{aligned} \quad (\text{A.15})$$

Both measures may not be probability measures, but they do have the same total mass.

The fact that the left-hand side of (A.13) represents a transport cost between workers  $\tilde{F}$  and jobs  $\tilde{G}$  is evident. Under this assignment each worker type is assigned to an identically indexed job, which have identical masses by construction of the worker distribution  $\tilde{F}$  in (A.14) and the job distribution  $\tilde{G}$  in (A.15). To establish the same on the right-hand side requires work. Consider first the worker  $x$  marginal on the right-hand side of (A.13).

1. The mass on  $x_0$  is

$$\sum_{1 < m \leq p} \lambda_{m1}^- + \sum_{1 \leq n < p} \lambda_{pn}^+ + \sum_{k=2}^{p-1} \left( \sum_{1 \leq n < k} \lambda_{kn}^+ + \sum_{k < m \leq p} \lambda_{mk}^- \right) = \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \lambda_{mn}^+).$$

2. Using equation (A.11) with  $k = 2$ , the mass on  $x_1$  is

$$\sum_{1 < m \leq p} \lambda_{m1}^+ + \sum_{1 < m \leq p} \omega_{m1}^+ = \sum_{1 < m \leq p} (\lambda_{m1}^- + \omega_{m1}^-).$$

3. For  $2 \leq k < p$ , using (A.12) and grouping terms, the mass on  $x_k$  is

$$\sum_{1 \leq n < k} \lambda_{kn}^- + \sum_{k < m \leq p} \lambda_{mk}^+ + \sum_{1 \leq n < k} \omega_{k,n}^- + \sum_{k < m \leq p} \omega_{mk}^+ = \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) + \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+).$$

4. Using equation (A.11) with  $k = p$ , the mass on  $x_p$  is

$$\sum_{1 \leq n < p} (\lambda_{pn}^- + \omega_{pn}^-) = \sum_{1 \leq n < p} (\lambda_{pn}^+ + \omega_{pn}^+).$$

Combining these four terms we see that the  $x$ -marginal of the right-hand side of (A.13) corresponds with that of (A.14). We proceed to show that the same is true for the distribution of jobs.

1. The mass on  $z_0$  is

$$\sum_{1 \leq n < p} \lambda_{pn}^- + \sum_{1 < m \leq p} \lambda_{m1}^+ + \sum_{k=2}^{p-1} \left( \sum_{1 \leq n < k} \lambda_{kn}^- + \sum_{k < m \leq p} \lambda_{mk}^+ \right) = \sum_{1 \leq n < m \leq p} (\lambda_{mn}^- + \lambda_{mn}^+),$$

where the equality follows by simple accounting.

2. The mass on  $z_1$  is

$$\sum_{1 < m \leq p} \lambda_{m1}^- + \sum_{1 < m \leq p} \omega_{m1}^- = \sum_{1 < m \leq p} (\lambda_{m1}^- + \omega_{m1}^-).$$

3. For  $2 \leq k < p$ , the mass on  $z_k$  is

$$\sum_{1 \leq n < k} \lambda_{kn}^+ + \sum_{k < m \leq p} \lambda_{mk}^- + \sum_{k < m \leq p} \omega_{mk}^- + \sum_{1 \leq n < k} \omega_{kn}^+ = \sum_{k < m \leq p} (\lambda_{mk}^- + \omega_{mk}^-) + \sum_{1 \leq n < k} (\lambda_{kn}^+ + \omega_{kn}^+).$$

4. Finally, the mass on  $z_p$  is given by

$$\sum_{1 \leq n < p} \lambda_{pn}^+ + \sum_{1 \leq n < p} \omega_{pn}^+ = \sum_{1 \leq n < p} (\lambda_{pn}^+ + \omega_{pn}^+).$$

We have thus proved the marginal distributions on both sides of the costs (A.13) are the worker distribution  $\tilde{F}$  and the job distribution  $\tilde{G}$ .

Why is the left-hand side of equation (A.13) the optimal transportation cost between the worker distribution  $\tilde{F}$  and job distribution  $\tilde{G}$ ? To characterize an optimal assignment between the constructed measures  $\tilde{F}$  and  $\tilde{G}$ , we decompose the corresponding measure of underqualification  $\tilde{H} := \tilde{F} - \tilde{G}$  into layers. By the definition of the worker measure  $\tilde{F}$  in equation (A.14) and the job measure  $\tilde{G}$  in equation (A.15), we know that for each  $k$  we have  $\tilde{F}(x_k) = \tilde{G}(z_k)$ . This means each layer  $\ell$  will consist of a subset  $S \subseteq \{0, \dots, p\}$  and the distributions within the layer  $F_\ell$  and  $G_\ell$  will be uniform on  $\{x_k\}_{k \in S}$  and  $\{z_k\}_{k \in S}$  respectively. From the assumption of the theorem we recall that the optimal assignment  $\pi$  pairs  $x_k$  with  $z_k$  for every  $k$  in the optimal assignment problem with uniform distributions on  $\{x_k\}_{0 \leq k \leq p}$  and  $\{z_k\}_{0 \leq k \leq p}$ . Since a restriction of an optimal assignment is also optimal on the restricted marginals, we know that an optimal assignment between  $F_\ell$  and  $G_\ell$  matches  $x_k$  to  $z_k$  for each  $k \in S$ . After adding the layers, the same holds for an optimal assignment between  $\tilde{F}$  and  $\tilde{G}$  by the principle of layering. Therefore, the pairs  $\{(x_k, z_k)\}$  are paired under an optimal assignment between  $\tilde{F}$  and  $\tilde{G}$ . This establishes the inequality (A.13), hence we finally conclude (A.8) has a solution.  $\square$

Next, we continue to prove the second part of the result, that the function  $\phi$  defined in (A.4)-(A.6) is indeed a local dual optimizer on  $S_{[x_0, z_0]}$ .

First, we record the following simple observation on our construction of  $\phi$  in the case of overlapping masses.

We will extensively make use of Lemma 8.

**Lemma 8.** Suppose  $h : [0, \infty) \rightarrow \mathbb{R}$  is concave. Then for  $0 \leq x \leq y$  and  $a > 0$  we have

$$h(x + a) - h(x) \geq h(y + a) - h(y).$$

*Proof.* From concavity, it follows that

$$h(y) + h(x + a) \geq \frac{(y - x)h(y + a) + ah(x)}{y + a - x} + \frac{(y - x)h(x) + ah(y + a)}{y + a - x} = h(x) + h(y + a),$$

completing the proof.  $\square$

*Proof of Theorem 2.* Suppose  $(x_1, z_1), \dots, (x_p, z_p)$  are ordered subpairs of pair  $(x_0, z_0)$  in the optimal assignment  $\pi$ , and that  $\phi_i$  are dual potentials on  $S_{[x_i, z_i]} := X_{[x_i, z_i]} \cup Z_{[x_i, z_i]}$  for all  $1 \leq i \leq p$ . Define  $c_{ij} = c_{i,j} := c(x_i, z_j)$ . We first prove that, with the possibilities of multiple workers on the same skill level and multiple jobs on the same difficulty level, our  $\phi$  in (A.4)-(A.6) is well-defined.

1. Suppose that  $x_n = x_{n+1}$  or  $z_n = z_{n+1}$  for some  $1 \leq n < p$ . Then any solution  $(\beta_2, \dots, \beta_p)$  to the system of inequalities (A.8) satisfies

$$\max(c_{00} - c_{0n} - c_{n+1,0}, -c_{n+1,n}) + c_{nn} \leq \beta_{n+1} \leq \min(c_{0,n+1} + c_{n0} - c_{00}, c_{n,n+1}) - c_{n+1,n+1}.$$

As a consequence,

$$c_{nn} - c_{n+1,n} \leq \beta_{n+1} \leq c_{n,n+1} - c_{n+1,n+1},$$

and hence we must have  $\beta_{n+1} = c_{nn} - c_{n+1,n+1}$ . In particular, the  $\phi$  defined in (A.4) satisfies  $\phi(x_n) = \phi(x_{n+1})$  in the case  $x_n = x_{n+1}$ , and  $\phi(z_n) = \phi(z_{n+1})$  in the case  $z_n = z_{n+1}$ .

2. Suppose that  $z_0 = z_1$  and  $x_0 \neq x_1$ . We first prove that

$$\max_{i \in \{1, \dots, p\}} (\phi(x_i) - c(x_i, z_0)) = \phi(x_1) - c(x_1, z_0).$$

Indeed, for any  $i \in \{1, \dots, p\}$ , it holds that

$$\phi(x_1) - c(x_1, z_0) = \phi(x_1) - c(x_1, z_1) = \phi(z_1) \geq \phi(x_i) - c(x_i, z_1) = \phi(x_i) - c(x_i, z_0).$$

Therefore, by (A.5),

$$\phi(z_0) = \phi(x_1) - c(x_1, z_0) = \phi(x_1) - c(x_1, z_1) = \phi(z_1).$$

3. Suppose that  $x_0 = x_1$ . Similarly as in step 2 above, we have

$$\phi(x_0) = \phi(z_1) + c(x_0, z_1) = \phi(z_1) + c(x_1, z_1) = \phi(x_1).$$

4. Suppose that  $x_0 = x_1$  and  $z_0 = z_1$ . We need to show that

$$\max_{i \in \{1, \dots, p\}} (\phi(x_i) - c(x_i, z_0)) = \min_{i \in \{1, \dots, p\}} (\phi(z_i) + c(x_0, z_i)) - c(x_0, z_0). \quad (\text{A.16})$$

This is because the left-hand side of (A.16) is equal to  $\phi(z_1)$  by step 3 and the right-hand side of (A.16) is equal to  $\phi(x_1) - c_{00}$  by step 2. Since  $c_{00} = c_{11} - \phi(x_1) - \phi(z_1)$ , this proves (A.16).



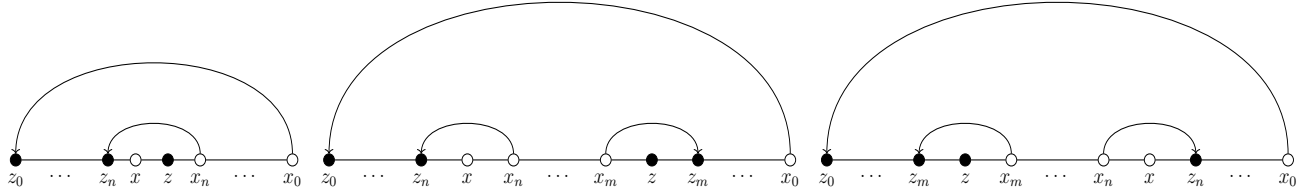


Figure A.6: Three Cases for the proof to Theorem 2

Figure A.6 illustrates the three different cases that we consider in the proof of Theorem 2. The first case is when  $n = m$ , the second case is when  $m > n$ , and the third case is when  $n > m$ . In each case, worker  $x \in (x_n, z_n)$  and job  $z \in (x_m, z_m)$ .

We prove that  $\phi$  is a local dual optimizer on the domain  $S_{[x_0, z_0]} \setminus \{x_0, z_0\}$ . Suppose that both  $x \in X_{[x_n, z_n]}$  and  $z \in Z_{[x_m, z_m]}$ . The equality  $\phi(x) - \phi(z) = c(x, z)$  when  $(x, z)$  is a worker-job pair is obvious because the same condition is satisfied by  $\phi_i$  for all  $1 \leq i \leq p$ . Our goal is to prove  $\phi(x) - \phi(z) \leq c(x, z)$  when worker  $x$  and job  $z$  are not paired.

We consider three cases.

*Case I:*  $n = m$ . This follows immediately since  $\phi_n$  is a dual potential on  $S_{[x_n, z_n]}$ .

*Case II:*  $n < m$ . Observe

$$\begin{aligned}
\phi(x) - \phi(z) &= (\phi(x) - \phi(z_n)) - \sum_{k=n}^m (\phi(x_k) - \phi(z_k)) + \sum_{k=n+1}^m (\phi(x_{k-1}) - \phi(z_k)) + (\phi(x_m) - \phi(z)) \\
&\leq c(x, z_n) - \sum_{k=n}^m c_{kk} + \sum_{k=n+1}^m (\beta_k + c_{kk}) + c(x_m, z) \\
&= c(x, z_n) - c_{nn} + c(x_m, z) + \sum_{k=n+1}^m \beta_k \\
&\leq c(x, z_n) - c_{nn} + c_{nm} - c_{mm} + c(x_m, z).
\end{aligned}$$

The first inequality follows since both  $\phi(x) - \phi(z_n) \leq c(x, z_n)$  and  $\phi(x_m) - \phi(z) \leq c(x_m, z)$  follow from the dual potential within the same part, and  $\phi(x_k) - \phi(z_k) = c_{kk}$  follows by the dual potential within the same part for paired workers and jobs, and finally  $\beta_k + c_{kk} = \phi(x_{k-1}) - \phi(z_k)$  by the definition of the constants  $a_i$  in statement of the theorem. The final inequality is implied by the upper bound on  $\sum_{k=n+1}^m \beta_k$  from (A.8). To show  $\phi(x) - \phi(z) \leq c(x, z)$ , it suffices to prove

$$c(x, z_n) + c(x_m, z) - c(x, z) \leq c_{nn} + c_{mm} - c_{nm}. \quad (\text{A.17})$$

Note that  $c(x, z_n) \leq c(x_n, z_n)$  and  $c(x_m, z) \leq c(x_m, z_m)$ . There are four cases:

1.  $z_n \leq x_n \leq z_m \leq x_m$ . Then  $c(x, z) \geq c(x_n, z_m)$  and the claim follows.

2.  $x_n \leq z_n \leq z_m \leq x_m$ . In this case  $z_n - x \leq z_n - x_n$ . Applying Lemma 8 with  $a = z - z_n$  yields  $c(x, z_n) - c(x, z) \leq c(x_n, z_n) - c(x_n, z) \leq c(x_n, z_n) - c(x_n, z_m)$ . Using that  $c(x_m, z) \leq c(x_m, z_m)$  equation (A.17) follows.
3.  $z_n \leq x_n \leq x_m \leq z_m$ . In this case  $z - x_m \leq z_m - x_m$ . Applying Lemma 8 with  $a = x_m - x_n$  yields  $c(x_m, z) - c(x, z) \leq c(x_m, z) - c(x_n, z) \leq c(x_m, z_m) - c(x_n, z_m)$ , where the first inequality follows from  $c(x_n, z) \leq c(x, z)$ . Further using  $c(x, z_n) \leq c(x_n, z_n)$ , equation (A.17) follows.
4.  $x_n \leq z_n \leq x_m \leq z_m$ . In this case, the configuration between  $x_n$  and  $z_n$  is identical to Case 2, and the configuration between  $x_m$  and  $z_m$  is identical to Case 3. We apply Lemma 8 exactly as in Case 2 and in Case 3. First,  $c(x, z_n) - c(x, z) \leq c(x_n, z_n) - c(x_n, z)$ . Second,  $c(x_m, z) - c(x_n, z) \leq c(x_m, z_m) - c(x_n, z_m)$ . Summing the two inequalities delivers inequality (A.17).

This completes the proof of (A.17) for Case II.

*Case III:  $n > m$ .* Observe that

$$\begin{aligned}
\phi(x) - \phi(z) &= (\phi(x) - \phi(z_n)) + \sum_{k=m+1}^{n-1} (\phi(x_k) - \phi(z_k)) - \sum_{k=m+1}^n (\phi(x_{k-1}) - \phi(z_k)) + (\phi(x_m) - \phi(z)) \\
&\leq c(x, z_n) + \sum_{k=m+1}^{n-1} c_{kk} - \sum_{k=m+1}^n (\beta_k + c_{kk}) + c(x_m, z) \\
&= c(x, z_n) - c_{nn} + c(x_m, z) - \sum_{k=m+1}^n \beta_k \\
&\leq c(x, z_n) - c_{nn} + c_{nm} - c_{mm} + c(x_m, z).
\end{aligned}$$

The first inequality follows since both  $\phi(x) - \phi(z_n) \leq c(x, z_n)$  and  $\phi(x_m) - \phi(z) \leq c(x_m, z)$  follow from the dual potential within the same part, and  $\phi(x_k) - \phi(z_k) = c_{kk}$  follows by the dual potential within the same part for paired workers and jobs, and finally  $\beta_k + c_{kk} = \phi(x_{k-1}) - \phi(z_k)$  by the definition of the constants  $a_i$  in statement of the theorem. The final inequality is implied by the upper bound on  $\sum_{k=n+1}^m \beta_k$  from (A.8). The rest follows similarly as in Case II.

It then remains to check that  $\phi$  is a dual potential on  $S_{[x_0, z_0]}$ , that is, to show that

1.  $\phi(x_0) - \phi(z_0) = c(x_0, z_0)$ ;
2. For  $x \in X_{[x_0, z_0]} \setminus \{x_0\}$ ,  $\phi(x) - \phi(z_0) \leq c(x, z_0)$ ;
3. For  $z \in Z_{[x_0, z_0]} \setminus \{z_0\}$ ,  $\phi(x_0) - \phi(z) \leq c(x_0, z)$ .

In view of (A.5) and (A.6), it remains to prove

$$\min_{1 \leq i \leq p} (\phi(z_i) + c(x_0, z_i)) - \max_{1 \leq i \leq p} (\phi(x_i) - c(x_i, z_0)) \geq c(x_0, z_0). \tag{A.18}$$

Equivalently, it suffices to show that

$$\phi(x) - \phi(z) \leq c(x_0, z) + c(x, z_0) - c_{0,0} \quad (\text{A.19})$$

for all  $x \in X_{[x_0, z_0]} \setminus \{x_0\}$  and  $z \in Z_{[x_0, z_0]} \setminus \{z_0\}$ . Recall that  $\phi(x_{i-1}) - \phi(z_i) = \beta_i + c_{ii}$  for  $2 \leq i \leq p$  by the definition of the constants in the statement of the theorem. Again, we have the three cases of Figure A.6 to consider to show that the sufficient condition (A.19) is satisfied.

*Case I:  $n = m$ .* By symmetry, we may assume  $x_0 < z_0$ . Consider first the case  $z_n \leq x_n$ . Since  $\phi_n$  is a dual potential on  $S_{[x_n, z_n]}$ , we have  $\phi(x) - \phi(z) \leq c(x, z)$ . Since  $\pi$  is optimal,

$$c(x, z) + c(x_0, z_0) \leq c(x_n, z_n) + c(x_0, z_0) \leq c(x_n, z_0) + c(x_0, z_n) \leq c(x, z_0) + c(x_0, z),$$

where the first inequality follows since  $(x, z) \in (x_n, z_n)$ , the second follows by optimality, and the third one follows since  $z_n \leq x_n$  and  $x_0 < z_0$ . Using the above inequality, we write:

$$\phi(x) - \phi(z) \leq c(x, z) \leq c(x_0, z) + c(x, z_0) - c_{0,0},$$

verifying (A.19).

Next, we consider the case where  $x_n < z_n$ . By the property of no intersecting pairs, it follows that

$$c(x_n, z_n) + c(x_0, z_0) \leq c(x_0, z_n) + c(x_n, z_0). \quad (\text{A.20})$$

Furthermore, two applications of Lemma 8 yield both

$$c(x_0, z_n) - c(x_0, z) \leq c(x_n, z_n) - c(x_n, z) \quad (\text{A.21})$$

when  $a = z_n - z$  and

$$c(x_n, z_0) - c(x, z_0) \leq c(x_n, z_n) - c(x, z_n). \quad (\text{A.22})$$

when  $a = z_0 - z_n$ . Summing up equations (A.20), (A.21), and (A.22) yields

$$c(x_0, z_0) + c(x, z_n) + c(x_n, z) \leq c(x_0, z) + c(x, z_0) + c(x_n, z_n).$$

Since  $\pi$  is optimal, we have

$$\phi(x) - \phi(z) \leq c(x, z) \leq c(x, z_n) + c(x_n, z) - c(x_n, z_n) \leq c(x_0, z) + c(x, z_0) - c_{0,0},$$

as desired by (A.19).

*Case II:  $n < m$ .* Similar to the other Case II discussed above, we obtain

$$\begin{aligned}\phi(x) - \phi(z) &\leq c(x, z_n) - c_{nn} + c(x_m, z) + \sum_{k=n+1}^m \beta_k \\ &\leq c(x, z_n) - c_{nn} + c_{0m} + c_{n0} - c_{00} - c_{mm} + c(x_m, z),\end{aligned}\tag{A.23}$$

where the second inequality follows from the upper bound in (A.8). To prove equation (A.19) it suffices to show that the upper bound in the previous equation is below the upper bound in equation (A.19), or

$$c(x, z_n) + c(x_m, z) - c(x_0, z) - c(x, z_0) \leq c(x_n, z_n) + c(x_m, z_m) - c(x_n, z_0) - c(x_0, z_m).\tag{A.24}$$

We observe that the arcs  $(x, z_0)$  and  $(x_n, z_n)$  intersect, but that the arcs  $(x, z_n)$  and  $(x_n, z_0)$  do not cross. By Lemma 3, describing that intersecting pairs are never optimal, we thus have

$$c(x, z_n) + c(x_n, z_0) \leq c(x, z_0) + c(x_n, z_n).\tag{A.25}$$

Similarly, the arcs  $(x_0, z)$  and  $(x_m, z_m)$  intersect, but the arcs  $(x_0, z_m)$  and  $(x_m, z)$  do not intersect. Thus, by the property of no intersecting pairs,

$$c(x_0, z_m) + c(x_m, z) \leq c(x_0, z) + c(x_m, z_m).\tag{A.26}$$

Summing up (A.25) and (A.26) yields (A.24) and hence completes the proof of Case II.

*Case III:  $n > m$ .* We have

$$\begin{aligned}\phi(x) - \phi(z) &\leq c(x, z_n) - c_{nn} + c(x_m, z) - \sum_{k=m+1}^n \beta_k \\ &\leq c(x, z_n) - c_{nn} - c_{00} - c_{mm} + c_{0m} + c_{n0} + c(x_m, z),\end{aligned}\tag{A.27}$$

where the last step follows from (A.8). Inequality (A.27) coincides with inequality (A.23) in Case II. Then (A.19) follows in exactly the same way from (A.24) as in Case II.  $\square$

Finally, we emphasize that our construction relies on the concavity of the mismatch cost function  $c(x, z)$  in two respects. First, the property of no intersecting pairs is essential for our induction structure. Second, (A.17) requires concavity as well.

## A.9 Proof to Theorem 3

In this appendix, we prove Theorem 3. We make use of Lemma 9 and Lemma 10, which we prove first.

**Lemma 9.** *Triangle Inequality.* For all  $x, y, z \in \mathbb{R}$ , it holds that  $c(x, y) + c(y, z) \geq c(x, z)$ .

*Proof.* Our cost of mismatch takes the form  $c(x, z) = h(z - x)$  where  $h$  is strictly concave and increasing on  $[0, \infty)$ , strictly concave and decreasing on  $(-\infty, 0]$ , satisfying  $h(0) = 0$ .

The case where  $x = z$  is trivial, so we focus our attention to the case where  $x \neq z$ . By symmetry, we assume  $x < z$  without loss of generality. If  $z - y \geq z - x > 0$ , then necessarily  $c(y, z) \geq c(x, z)$  and hence  $c(x, y) + c(y, z) \geq c(x, z)$ . The same argument applies when  $y - x \geq z - x > 0$ . In the remaining scenario where both  $(z - y)$  and  $(y - x)$  are in  $[0, z - x)$ , we have by concavity of  $h$  that<sup>40</sup>

$$c(x, y) + c(y, z) = h(y - x) + h(z - y) \geq \frac{y - x}{z - x} h(z - x) + \frac{z - y}{z - x} h(z - x) = h(z - x) = c(x, z),$$

where the inequality follows since  $y - x = \frac{y - x}{z - x} \times (z - x) + \frac{z - y}{z - x} \times 0$ . This completes the proof.  $\square$

In the main text, we established the connection between the dual optimizers for the cost minimization problem and the dual optimizers for the output maximization problem. In this appendix, we exploit this connection to simplify the exposition to the proof of Theorem 3. Specifically, we use that we can equivalently characterize the dual functions  $(\phi, \psi)$  for the overlapping segments of the worker and the job distribution, with the understanding that we can obtain wages and job values using  $w(x) = g(x) - \phi(x)$  and  $v(z) = h(z) - \psi(z)$ , where  $y(x, z) = g(x) + h(z) - c(x, z)$ .

To formulate Theorem 3 in terms of the dual potentials for the mismatch cost minimization problem, we need to describe our sequential construction of the functions. For the interpretation of these objects we refer the reader to the main text. We define sequentially the dual maps, analogous to our previous definitions (13) and (14). Starting from  $\tilde{\phi} = g - \tilde{w}$ , where  $\tilde{w}$  are the dual values for mismatched workers  $x \in I$ , let

$$\tilde{\psi}(z) := \min_{x \in I} c(x, z) - \tilde{\phi}(x) \quad \text{and} \quad \hat{\phi}(x) := \min_{z \in I \cup J} c(x, z) - \tilde{\psi}(z), \quad (\text{A.28})$$

where we recall that  $J$  is the set of mismatched jobs. Moreover, let

$$\hat{\psi}(z) := \min_{x \in I \cup J} c(x, z) - \hat{\phi}(x),$$

$\phi(x) = \hat{\phi}(x)$  for  $x \in I$ ,  $\psi(z) = \hat{\psi}(z)$  for  $z \in J$ , and set  $\phi(x) = -\psi(x)$  for  $x \in J$  and  $\psi(z) = -\phi(z)$  for  $z \in I$ . Finally, we define for  $x \in K$

$$\phi(x) = \min_{z \in I \cup J} c(x, z) - \psi(z) \quad (\text{A.29})$$

---

<sup>40</sup>The interval is bounded below by zero because when  $z - y \geq z - x$  is not true, then  $y - x > 0$ , and similarly when  $y - x \geq z - x$  is not true then  $z - y > 0$ .

and  $\psi(z) = -\phi(z)$  for  $z \in K$ . It is easy to check that with these definitions,  $\phi = g - w$  and  $\psi = f - v$ , with  $w, v$  given in Theorem 3.

To prove the result, we first define  $c$ -conjugate functions.

For  $\phi : I \rightarrow \mathbb{R}$ , we define the  $c$ -conjugate function for all jobs  $z \in J$  as

$$\phi^c(z) := \min_{x \in I} c(x, z) - \phi(x). \quad (\text{A.30})$$

Denote by  $\bar{c}(z, x) = c(x, z)$  and we further define for  $x \in I$

$$\phi^{c\bar{c}}(x) = (\phi^c)^{\bar{c}}(x) = \min_{z \in J} \bar{c}(z, x) - \phi^c(z) = \min_{z \in J} c(x, z) - \phi^c(z). \quad (\text{A.31})$$

Given these definitions, the following statements follow:<sup>41</sup>

1.  $\phi^{c\bar{c}} \geq \phi$ .

This follows since for each  $x \in I$  and  $z \in J$ ,  $\phi(x) + \phi^c(z) \leq c(x, z)$  or  $\phi(x) \leq c(x, z) - \phi^c(z)$  by the definition of the  $c$ -conjugate function. By taking the infimum  $z \in J$  this leads to  $\phi^{c\bar{c}}(x) \geq \phi(x)$  by the definition (A.31).

2. If  $\phi = \psi^{\bar{c}}$  for some  $\psi$ , then  $\phi = \phi^{c\bar{c}}$ .

First, we observe that  $\phi = \psi^{\bar{c}}$  naturally implies  $\phi^c = \psi^{\bar{c}c}$ . To see this, note that  $\phi^c = \psi^{\bar{c}c} \geq \psi$  which follows from the previous statement. This inequality, by uniformly decreasing from  $\phi^c$  to  $\psi$ , together with the definition (A.31), implies we uniformly increase the conjugate, or  $\phi^{c\bar{c}} = (\phi^c)^{\bar{c}} \leq \psi^{\bar{c}} = \phi$ . We establish  $\phi = \phi^{c\bar{c}}$  by combining  $\phi^{c\bar{c}} \leq \phi$  with Statement 1.

3. If  $(\phi, \psi)$  is an optimal dual pair, then so is  $(\phi, \phi^c)$ .

Suppose  $(\phi, \psi)$  is a dual pair, then  $\phi(x) + \psi(z) \leq c(x, z)$ . It holds by the definition in (A.30) that  $\phi(x) + \phi^c(z) \leq c(x, z)$  as well as  $\phi^c(z) \geq \psi(z)$ . Since  $\phi^c(z) \geq \psi(z)$  and  $(\phi, \phi^c)$  is a dual solution, it follows that if  $(\phi, \psi)$  is a solution to the dual maximization problem, then so is  $(\phi, \phi^c)$ .

**Lemma 10.**  $\hat{\phi}(x) + \tilde{\psi}(z) \leq c(x, z)$  for all workers  $x$  and jobs  $z$  such that  $x, z \in I \cup J$ , and equality holds for  $(x, z) \in \Gamma_\pi$ .

*Proof.* That  $\hat{\phi}(x) + \tilde{\psi}(z) \leq c(x, z)$  follows from definition of the dual optimizer for workers  $\hat{\phi}(x)$  in (13). Next, we prove  $\hat{\phi}(x) + \tilde{\psi}(z) = c(x, z)$  for all workers and jobs  $(x, z) \in \Gamma_\pi$ .

To prove that  $\hat{\phi}(x) + \tilde{\psi}(z) = c(x, z)$  for workers and jobs  $(x, z) \in \Gamma_\pi$ , we fix some pair  $(x, z) \in \Gamma_\pi$ . Since  $(\tilde{\phi}, \tilde{\phi}^c)$  is a dual solution to the assignment problem between remaining workers and jobs,

---

<sup>41</sup>See Chapter 1 of Santambrogio (2015) for further details.

$\tilde{\phi}(x) + \tilde{\phi}^c(z) = c(x, z)$  for all  $(x, z) \in \Gamma_\pi$ . Given the definition of the dual value for jobs  $z \in I \cup J$  in (A.28) we obtain that  $\tilde{\psi}(z) = \min_{x \in I} c(x, z) - \tilde{\phi}(x) = \tilde{\phi}^c(z)$  and hence that  $\tilde{\phi}(x) + \tilde{\psi}(z) = c(x, z)$  for all  $(x, z) \in \Gamma_\pi$ . To conclude the proof it remains to show that  $\hat{\phi}(x) = \tilde{\phi}(x)$  for every worker  $x \in I$ .

We next show that  $\hat{\phi}(x) = \tilde{\phi}(x)$  for every worker  $x \in I$ . Since we replaced, without loss of generality, the dual potential  $\tilde{\phi}$  by the  $c$ -transform  $\tilde{\phi}^{c\bar{c}}$ ,<sup>42</sup> it follows from the definition of the  $c$ -transform that for all  $x \in I$ :

$$\tilde{\phi}(x) = \tilde{\phi}^{c\bar{c}}(x) = \min_{z \in J} c(x, z) - \tilde{\phi}^c(z).$$

Further, since  $z \in J$ , by definition of the dual potential for jobs  $\tilde{\psi}(z) = \min_{x \in I} c(x, z) - \tilde{\phi}(x) = \tilde{\phi}^c(z)$ , where the second equality follows from the definition of the  $c$ -transform. We substitute this relationship into the previous expression for  $\tilde{\phi}(x)$  to write

$$\tilde{\phi}(x) = \min_{z \in J} c(x, z) - \tilde{\phi}^c(z) = \min_{z \in J} c(x, z) - \tilde{\psi}(z).$$

We can use the definition of the dual optimizers (13) to write that for all workers  $x \in I$ :

$$\hat{\phi}(x) = \min \left( \min_{z \in J} c(x, z) - \tilde{\psi}(z), \min_{z \in I} c(x, z) - \tilde{\psi}(z) \right) = \min \left( \tilde{\phi}(x), \min_{z \in I} c(x, z) - \tilde{\psi}(z) \right)$$

where the first equality follows by splitting the sets in (A.30) and the second equality follows from the equation above.

Hence, we want to show for  $(x, z) \in I$  the infimum is attained by  $\tilde{\phi}(x)$ . We show  $c(x, z) \geq \tilde{\psi}(z) + \tilde{\phi}(x)$ . This follows since the dual optimizer for all jobs is defined as  $\tilde{\psi}(z) := \min_{x \in I} c(x, z) - \tilde{\phi}(x)$  for all  $z \in I \cup J$ .  $\square$

Having established the two claims, we next prove Theorem 3.

*Proof to Theorem 3.* The proof is divided in three parts. We first show the inequality holds on  $I \cup J$ . To do so, we consider four cases:

1. If  $x \in I$  and  $z \in J$ , we have

$$\phi(x) + \psi(z) = \phi(x) - \phi(z) = \hat{\phi}(x) - \max_{x \in I \cup J} (\hat{\phi}(x) - c(x, z)) \leq c(x, z)$$

because in the final step we subtract the maximum, but a feasible deduction is  $\hat{\phi}(x) - c(x, z)$ .

---

<sup>42</sup>We can always improve on the original  $\tilde{\phi}$  by doing a double  $c$ -conjugate transform by Statement 1 that is weakly better in terms of the dual maximization problem, see Remark 1.13 in Santambrogio (2015).

2. If  $x, z \in I$ , then by Lemma 9 we have

$$\phi(x) + \psi(z) = \phi(x) - \phi(z) = \hat{\phi}(x) - \hat{\phi}(z) = \min_{y \in I \cup J} (c(x, y) - \tilde{\psi}(y)) - \min_{y \in I \cup J} (c(z, y) - \tilde{\psi}(y)) \leq c(x, z)$$

where the final equality follows by (13). The concluding inequality is obtained as follows. Suppose  $y_0$  attains the infimum for the second term, the same  $y_0$  may not attain the infimum for the first term but is feasible, so the left hand side will be lower than when both terms are evaluated at  $y_0$ , or  $\min_{y \in I \cup J} (c(x, y) - \tilde{\psi}(y)) - \min_{y \in I \cup J} (c(z, y) - \tilde{\psi}(y)) \leq c(x, y_0) - c(z, y_0)$ . We combine the right-hand side with the triangle inequality of Lemma 9 to write  $c(x, z) + c(z, y_0) \geq c(x, y_0)$  or  $c(x, z) \geq c(x, y_0) - c(z, y_0)$  to obtain the inequality.

3. If  $x \in J$  and  $z \in I$ .

$$\phi(x) + \psi(z) = \max_{y \in I \cup J} (\hat{\phi}(y) - c(y, x)) - \hat{\phi}(z).$$

We next want to show that this expression is less than  $c(x, z)$ . This is equivalent to showing that  $\hat{\phi}(y) - c(y, x) \leq c(x, z) + \hat{\phi}(z)$  for all  $y \in I \cup J$ . To establish this, fix  $y$ , and evaluate:

$$\hat{\phi}(y) - \hat{\phi}(z) = \min_{w \in I \cup J} (c(y, w) - \tilde{\psi}(w)) - \min_{w \in I \cup J} (c(z, w) - \tilde{\psi}(w))$$

where the equality follows from the definition of  $\hat{\phi}$  in equation (13). Let  $w_0$  be the value that attains the infimum in the second term on the right, which is also feasible for the first term so that  $\hat{\phi}(y) - \hat{\phi}(z) \leq c(y, w_0) - c(z, w_0)$ . To bound this further, we use the triangle inequality of Lemma 9 twice to write  $c(y, w_0) - c(z, w_0) \leq c(y, z)$  as well as  $c(y, z) \leq c(y, x) + c(x, z)$ . Using the triangle inequalities, we thus write  $\hat{\phi}(y) - \hat{\phi}(z) \leq c(y, x) + c(x, z)$  which is what we wanted to show since  $y$  is arbitrary.

4. If worker and job  $x, z \in J$ , use (A.28) to write

$$\phi(x) + \psi(z) = \max_{y \in I \cup J} (\hat{\phi}(y) - c(y, x)) - \max_{y \in I \cup J} (\hat{\phi}(y) - c(y, z)).$$

To bound the right-hand side, let  $y_0$  denote the value that attains the supremum in the first term, that is also feasible for the second term. Hence, the right-hand side is bounded above by  $-c(y_0, x) + c(y_0, z)$ . By the triangle inequality of Lemma 9 it follows that  $-c(y_0, x) + c(y_0, z) \leq c(x, z)$  and hence we have  $\phi(x) + \psi(z) \leq c(x, z)$ .

The second part of the proof shows that the equality holds everywhere on  $I \cup J$  with respect to the optimal assignment  $\pi$ . We distinguish two cases:



1. The worker is perfectly matched to their job, or  $(x, z) \in \{(x, x) : x \in \mathbb{R}\}$ . Since the dual functions are defined as  $\psi(x) = -\phi(x)$  we have  $\phi(x) + \psi(x) = 0$ . As a result,  $\phi(x) + \psi(x) = 0 = c(x, x)$ , as the cost of mismatch is zero.
2. The worker is mismatched in their job, or  $(x, z) \in \Gamma_\pi$ , implying worker  $x \in I$  and job  $z \in J$ . Using definition (A.28),  $\psi(z) = -\phi(z)$ ,

$$\phi(x) + \psi(z) = \hat{\phi}(x) - \max_{y \in I \cup J} (\tilde{\phi}(y) - c(y, z)).$$

By Lemma 10,  $\hat{\phi}(x) + \tilde{\psi}(z) \leq c(x, z)$  for all  $(x, z)$ . In particular, for a given job  $z$ ,  $\hat{\phi}(x) + \tilde{\psi}(z) \leq c(x, z)$  for all  $x$ , and  $\min_{x \in I \cup J} (c(x, z) - \hat{\phi}(x)) \geq \tilde{\psi}(z)$ , or, equivalently,  $-\max_{x \in I \cup J} (\hat{\phi}(x) - c(x, z)) \geq \tilde{\psi}(z)$ . Combining this inequality with the previous expression, we obtain the inequality

$$\phi(x) + \psi(z) \geq \hat{\phi}(x) + \tilde{\psi}(z) = c(x, z)$$

where the final equality follows by Lemma 10. Since we have shown the opposite inequality above in the first case of the first part to this proof, we obtain that  $\phi(x) + \psi(z) = c(x, z)$ .

In the third part of the proof we further establish that the dual inequality  $\phi(x) + \psi(z) \leq c(x, z)$  holds when  $x \in K$  or  $z \in K$ . There are three cases.

1.  $x \notin K$ ,  $z \in K$ . For any  $x' \notin K$ , we have  $\phi(x) + \psi(x') = \phi(x) - \phi(x') \leq c(x, x')$  when  $x \notin K$  by the first part of this proof. Following the triangle inequality of Lemma 9,  $\phi(x) - \phi(x') \leq c(x, x') \leq c(x, z) + c(z, x')$ , giving  $\phi(x) - c(x, z) \leq c(z, x') + \phi(x') = c(z, x') - \psi(x')$ . Taking infimum over  $x' \in I \cup J$  gives  $\phi(x) - c(x, z) \leq \phi(z) = -\psi(z)$  using the definition of  $\phi$ .
2.  $x \in K$ ,  $z \notin K$ . For any  $x' \in I \cup J$ , by the definition of the wage function, we have that  $\phi(x) = \min_{x' \in I \cup J} c(x, x') - \psi(x')$ , such that  $\phi(x) - \phi(x') \leq c(x, x') \leq c(x, z) + c(z, x')$ , where the final step follows by the triangle inequality of Lemma 9. Alternatively, we write  $\phi(x) - c(x, z) \leq c(z, x') - \psi(x')$ . Taking infimum in  $x' \in I \cup J$  gives  $\phi(x) + \psi(z) \leq c(x, z)$  using the definition of  $\phi$ .
3.  $x, z \in K$ . We want to establish  $\phi(x) + \psi(z) \leq c(x, z)$ . Using the definitions of the dual potentials in (A.29),

$$\phi(x) + \psi(z) = \min_{x' \in I \cup J} (c(x, x') - \psi(x')) + \max_{x' \in I \cup J} (-c(z, x') + \psi(x')).$$

Suppose the maximum in the second term is attained by the worker value  $x_0$ , and also evaluate the first term at  $x_0$  where it may not attain the minimum, implying  $\phi(x) + \psi(z) \leq c(x, x_0) - c(z, x_0)$ . By the triangle inequality  $c(x, x_0) - c(z, x_0) \leq c(x, z)$  and hence it indeed follows that  $\phi(x) + \psi(z) \leq c(x, z)$ .

By observing that  $\phi = -\psi$  on the set  $K$ , the equality  $\phi(x) + \psi(z) = 0 = c(x, z)$  holds when  $x, z \in K$  and  $(x, z) \in \Gamma_\pi$ . This completes the proof in view of Lemma 1.  $\square$

# Composite Sorting

## Technical Appendix

Job Boerma, Aleh Tsyvinski, Ruodu Wang, and Zhenyuan Zhang

March 2023

## B Additional Results

In this appendix we present additional technical results.

### B.1 General Production Function

In this section, we present two generalizations of the model.

#### B.1.1 Concave Distance Function

Using Legendre transformations, we show that the indirect output function is generally a strictly concave function in mismatch given strictly convex cost functions. To be specific, consider the same cost function  $\Psi$  as above, and use  $d := |x - y|$  to denote the distance to obtain:

$$\mathcal{C}(d) = \min_{\gamma \geq 0} \gamma d + \Psi(\gamma) \tag{A.32}$$

This problem has a unique solution characterized by  $d = -\Psi'(\gamma(d))$ . From the envelope condition, we obtain  $\mathcal{C}'(d) = \gamma > 0$ , showing that the cost function is strictly increasing in the distance.

To characterize the second derivative, we write the cost minimization problem as a maximization problem of the form:

$$\hat{\mathcal{C}}(d) = \max_{\gamma \geq 0} -\gamma d - \Psi(\gamma)$$

where  $\hat{\mathcal{C}} = -\mathcal{C}$  which shows  $\hat{\mathcal{C}}$  is the Legendre transformation of the strictly convex function  $\Psi$ . Since the Legendre transformation of a strictly convex function is also strictly convex, the indirect cost function  $\mathcal{C}$  is a strictly concave function of the distance. As a result, choosing an assignment to maximize:

$$y(x, z) = z + x - \mathcal{C}(|x - y|)$$

where  $\mathcal{C}$  is now our concave cost distance function.

### B.1.2 Asymmetric Distance Function

Next, we incorporate differential distance functions for both  $x - z > 0$  and  $x - z < 0$ . This is a trivial extension, let  $\bar{\Psi}$  denote the cost function for  $x - z > 0$  and  $\underline{\Psi}$  denote the cost function for  $x - z < 0$ . In this case, the cost minimization problem is:

$$\bar{C}(d) = \min_{\gamma \geq 0} \gamma d + \bar{\Psi}(\gamma)$$

when  $d > 0$ . By the same arguments on the Legendre transformation, this gives rise to a strictly concave function of the distance  $\bar{C}(d)$ . Analogously, when  $d < 0$ , we generically obtain a distinct strictly concave function of the distance  $\underline{C}(d)$ . As a result, we choose an assignment to maximize:

$$y(x, z) = z + x - \bar{C}(\{x - y\}_+) - \underline{C}(\{y - x\}_+).$$

## B.2 Convex Cost of Mismatch

We show that the optimal assignment features positive sorting when the mismatch cost function is strictly convex in the distance between worker skills and job difficulties. We establish this by showing that the cost function is submodular when the mismatch cost function is strictly convex. In turn, the production function is strictly supermodular, and hence the optimal sorting is positive following [Becker \(1973\)](#).

**Lemma 11.** *Convex Cost of Mismatch.* Suppose  $c(x, z) = h(z - x)$  with  $h : \mathbb{R} \rightarrow \mathbb{R}$  strictly convex. Then the mismatch cost function  $c$  is strictly submodular.

*Proof.* To establish that the cost of mismatch is submodular, we need to establish that

$$c(x_1, z_1) + c(x_2, z_2) < c(x_1, z_2) + c(x_2, z_1),$$

for any  $x_1 < x_2$  and  $z_1 < z_2$ .

It follows from the strict convexity of the cost function that  $h(z_1 - x_1) \leq c_1 h(z_1 - x_2) + c_2 h(z_2 - x_1)$  where the weighting coefficients are  $c_1 = \frac{(z_2 - x_1) - (z_1 - x_1)}{(z_2 - x_1) - (z_1 - x_2)}$  and  $c_2 = 1 - c_1 = \frac{(z_1 - x_1) - (z_1 - x_2)}{(z_2 - x_1) - (z_1 - x_2)}$ .<sup>43</sup> Analogously, it follows that  $h(z_2 - x_2) \leq c_2 h(z_1 - x_2) + c_1 h(z_2 - x_1)$ .<sup>44</sup> We use these two inequalities to show that

$$c(x_1, z_2) + c(x_2, z_1) = h(z_2 - x_1) + h(z_1 - x_2) > h(z_1 - x_1) + h(z_2 - x_2) = c(x_1, z_1) + c(x_2, z_2),$$

where the inequality follows by addition of the two previous inequalities.  $\square$

<sup>43</sup>We observe that  $c_1(z_1 - x_2) + (1 - c_1)(z_2 - x_1) = z_2 - x_1 + c_1((z_1 - x_2) - (z_2 - x_1)) = z_2 - x_1 - ((z_2 - x_1) - (z_1 - x_1)) = (z_1 - x_1)$  and further note that the coefficient  $c_1$  is between 0 and 1 since the denominator is positive the numerator is positive but smaller than the denominator.

<sup>44</sup>For completeness, we observe that  $c_2(z_1 - x_2) + (1 - c_2)(z_2 - x_1) = z_2 - x_1 + c_2((z_1 - x_2) - (z_2 - x_1)) = z_2 - x_1 - ((z_1 - x_1) - (z_1 - x_2)) = (z_2 - x_2)$ .

### B.3 Efficiency Properties of Algorithm

Recall our algorithm for computing dual potentials from Appendix A.7. The total runtime is dominated by the runtime for solving the system of inequalities (A.3), which we recall as follows. Suppose that  $(x_1, z_1), \dots, (x_p, z_p)$  are ordered subpairs of pair  $(x_0, z_0)$  in the optimal assignment  $\pi$ . Define  $c_{ij} := c(x_i, z_j)$ . Then the system of inequalities, where for all  $1 \leq n < m \leq p$ :

$$\max(c_{00} + c_{nn} - c_{0n} - c_{m0}, c_{nn} - c_{mn}) \leq \sum_{k=n+1}^m \beta_k \leq \min(c_{0m} + c_{n0} - c_{00} - c_{mm}, c_{nm} - c_{mm}) \quad (\text{A.33})$$

(a) admits a solution  $(\beta_2, \dots, \beta_p)$ .

Since the system of inequalities (A.33) can be solved via linear programming in  $p$  steps where  $p \leq n$ , the worst-case runtime for our algorithm is  $O(n^4)$  (when  $p = n$ ). However, it becomes much more efficient when more layers of arcs are introduced. This is because compared with the trivial linear programming, our algorithm solves the problem in the order from bottom arcs to top, while at each step the values of  $\phi$  in the hidden arcs need not be computed again, but only adjusted with constant factors. Typically, the number  $N$  will not be as large as  $n$ . The following proposition provides a general upper bound of the number  $N$ , which is a consequence of the absence of intersecting pairs. Define the number of crossings of  $F - G$  at  $x \in \mathbb{R}$  as

$$C_{F-G}(x) := \sum_{1 \leq k \leq n} \mathbb{1}_{\{(F-G)(x_k)=x\}} + \sum_{1 \leq k \leq n} \mathbb{1}_{\{(F-G)(z_k)=x\}}.$$

In other words, every time a flat part of  $F - G$  takes value  $x$ , we count that as an  $x$ -crossing. Note that this is nonzero only if  $nx \in \mathbb{Z}$  and  $|x| \leq 1$ .

**Proposition 6.** Suppose that  $(x_1, z_1), \dots, (x_p, z_p)$  are the ordered subpairs of the pair  $(x_0, z_0)$  in the optimal matching  $\pi$ . Then there exists  $x \in \mathbb{Z}/n$  such that the measure of underqualification  $F - G$  crosses the level  $x$  for  $N$  times, i.e.,  $C_{F-G}(x) \geq N$ .

*Proof.* Consider numbers  $t_i \in (\max(x_i, z_i), \min(x_{i+1}, z_{i+1}))$ ,  $1 \leq i < p$ . By the property of no intersecting pairs,  $F(t_i) - G(t_i)$  is a constant in  $i$ . On the other hand,  $F - G$  cannot be constant on the interval  $[t_i, t_{i+1}]$ . Thus the claim follows.  $\square$

In fact, the number of crosses of  $F - G$  at a certain level is typically much smaller if we consider empirical processes. By empirical we mean that  $X_1, \dots, X_n$  are random samples drawn independently from distribution  $F$  on  $\mathbb{R}$ , and the workers are uniformly distributed on  $\{X_1, \dots, X_n\}$ , and similarly for the jobs. In this case, we may further refine the bound for the runtime of our algorithm, as is remarked below.

**Remark 1.** Suppose that  $F, G$  are independent empirical measures of the uniform distribution on  $[0, 1]$ . Then the runtime of the algorithm is  $O(n^{2.5}(\log \log n)^{3/2})$  almost surely.

In order to prove Remark 1, consider the (random) empirical cumulative densities  $F_n, G_n$ , drawn from two independent sequences  $\{X_i\}_{1 \leq i \leq n}$  and  $\{Z_i\}_{1 \leq i \leq n}$  uniformly in  $[0, 1]$ , i.e.,

$$F_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k \leq t\}} \quad \text{and} \quad G_n(t) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{Z_k \leq t\}}.$$

It is well known that the scaled measure of underqualification  $\sqrt{n}(F_n - G_n)$  can be well approximated by a Brownian bridge, where we recall that a (standard) Brownian bridge  $B = \{B(t)\}_{t \in [0,1]}$  is a centered Gaussian process with covariance  $\mathbb{E}[B(s)B(t)] = \min(s, t) - st$ . We denote the local time of a standard Brownian bridge  $B$  on  $[0, 1]$  at  $x \in \mathbb{R}$  by  $L_B(x)$ . By definition, the local time process  $\{L_B(x)\}_{x \in \mathbb{R}}$  is such that for any bounded Borel function  $f$ ,

$$\int_0^1 f(B(t)) dt = \int_{\mathbb{R}} f(x) L_B(x) dx.$$

The following Lemma is a special case of Theorem 5 of [Khoshnevisan \(1992\)](#).

**Lemma 12** (Theorem 5 of [Khoshnevisan \(1992\)](#)). There exists a suitable probability space carrying  $F_n, G_n$ , and a sequence of Brownian bridges  $\{B_n\}$ , such that

$$\lim_{n \rightarrow \infty} \max_{k \in \mathbb{Z}} \left| n^{-1/2} C_{F_n - G_n} \left( \frac{k}{n} \right) - \sqrt{2} L_{B_n} \left( \frac{k\sqrt{2}}{\sqrt{n}} \right) \right| = O(n^{-0.24}) \quad \text{a.s.}$$

We also have the following Lemma on fluctuations of the local time for Brownian bridges. This is taken from Lemma 3.2 of [Bass and Khoshnevisan \(1995\)](#) applied with  $n_k = k$  and  $\varepsilon_n = \sqrt{2/n}$ .

**Lemma 13.** Let  $\{B_n\}$  be any sequence of Brownian bridges. It holds that

$$\sup_{|x-y| < \sqrt{\frac{2}{n}}} |L_{B_n}(x) - L_{B_n}(y)| = O(n^{-0.24}) \quad \text{a.s.}$$

With a Borel-Cantelli argument in [Csörgő, Shi, and Yor \(1999\)](#) applied to the sequence of Brownian bridges  $\{B_n\}$  (with the tail estimates supplied by Theorem 5.1 therein), the following lemma can be similarly established as Theorem 1.4 of [Csörgő, Shi, and Yor \(1999\)](#).

**Lemma 14.** Let  $\{B_n\}$  be any sequence of Brownian bridges. There is a constant  $C > 0$  such that

$$\mathbb{P} \left[ \int_{\mathbb{R}} L_{B_n}(x)^4 dx > y \right] \leq \exp \left( -\frac{y^{2/3}}{C} \right).$$

Moreover,

$$\int_{\mathbb{R}} L_{B_n}(x)^4 dx = O \left( (\log \log n)^{3/2} \right) \quad \text{a.s.}$$

*Proof.* The first claim is Theorem 5.1 of Csörgő, Shi, and Yor (1999) applied with  $p = 4$ . The second claim can be proved in a similar way as (3.7a) of Bass and Khoshnevisan (1995).  $\square$

*Proof of Remark 1.* Recall that solving (A.33) requires  $N^4$  steps. In view of Proposition 6, the runtime of our algorithm has the upper bound

$$\sum_{|k| \leq n} C_{F-G} \left( \frac{k}{n} \right)^4.$$

Using Lemma 12, we get that almost surely,

$$C_{F-G} \left( \frac{k}{n} \right) = \sqrt{2n} L_{B_n} \left( \frac{\sqrt{2}k}{\sqrt{n}} \right) + O(n^{0.26}).$$

Therefore, using the elementary inequality  $(A + B)^4 \leq 16(A^4 + B^4)$  we have almost surely,

$$\begin{aligned} \sum_{|k| \leq n} C_{F-G} \left( \frac{k}{n} \right)^4 &\leq 64 \sum_{|k| \leq n} \left( n^2 L_{B_n} \left( \frac{\sqrt{2}k}{\sqrt{n}} \right)^4 + O(n^{1.04}) \right) \\ &\leq O(n^{2.04}) + 1024n^{2.5} \left( \int_{-\sqrt{n}}^{\sqrt{n}+1/\sqrt{n}} L_{B_n}(x)^4 dx + \sum_{|k| \leq n} \frac{1}{\sqrt{n}} O(n^{-0.96}) \right) \\ &\leq O(n^{2.04}) + 1024n^{2.5} \int_{\mathbb{R}} L_{B_n}(x)^4 dx, \end{aligned}$$

where we applied Lemma 13 in the second inequality. Applying now Lemma 14 concludes the proof.  $\square$