The Unreasonable Effectiveness of Optimal Transport in Economics

Alfred Galichon

▶ **To cite this version:**

Alfred Galichon. The Unreasonable Effectiveness of Optimal Transport in Economics. 2021. hal-03936221

HAL Id: hal-03936221

https://sciencespo.hal.science/hal-03936221

Preprint submitted on 12 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

arXiv:2107.04700v1 [econ.GN] 9 Jul 2021

The unreasonable effectiveness of optimal transport in economics

submitted to the proceeding of the 2020 World Congress of the Econometric Society

Alfred Galichon*

July 13, 2021

This paper is dedicated to the memory of Emmanuel Farhi (1978-2020).

1 Introduction

The mathematical theory of optimal transport traces back to Monge in the 18th century, who asked the main questions, for which he provided deep insights but left them unresolved. Regarded as a famous open problem throughout the 19th century, it was revived, and finally solved, with the advent of linear programming and works by Kantorovich, Koopmans, von Neumann, Dantzig and others in the mid 20th century. While the theory

*New York University and Sciences Po, Email: alfred.galichon@nyu.edu. Support from the European Research Council under Grant Agreement No. 866274 "Equiprice" as well as research assistance by Gabriele Buontempo and Giovanni Montanari are gratefully acknowledged.

partly arose out of economic motivations (specifically resource allocation problems), it soon drifted away from economics. A second revival has occured since the 1990s, when insights from convex analysis were introduced by Brenier, Rachev, and Rüschendorf, and from geometry by Gangbo, McCann, Villani and others.

In spite of this, up until recently, optimal transport has still been reported missing from the standard toolbox of quantitative economics. However, it turns out that many basic problems encountered in diverse economic applications in various fields are *optimal transport problems in disguise*. Beyond intellectual curiosity, understanding this connection is useful to make use of the mature set of results of optimal transport to solve the problems, and in particular, to deal with questions of existence, uniqueness, stability, and computation without reinventing the wheel.

This paper is a rapid overview of some of these connections, and some extensions. It is admittedly skewed toward my own work, and borrows much material from my 2016 monograph, *Optimal Transport Methods in Economics*, to which the reader is referred for details. I cover much of the material from an empirical perspective every winter in the January edition of my 'math+econ+code' masterclasses (www.math-econ-code.org). In mathematics, a useful read is Santambrogio's *Optimal Transport for Applied Mathematicians* 2015, or Villani's 2003 introductory lecture notes *Topics in Optimal Transportation*. For computational aspects, Peyré and Cuturi's *Computational Optimal Transport* is a useful complement. Villani's 2009 treatise *Optimal Transport: Old and New* remains the most exhaustive reference on the topic.

2 Optimal transport in a nutshell

2.1 Optimal transport duality

Let us describe the optimal transport problem in the discrete case. Assume a central planner needs to match a population of workers, each of whom is characterized by their type $x \in \mathcal{X}$, with a population of firms, each of whom with type $y \in \mathcal{Y}$. A match between a worker of type x and a firm of type y produces output Φ_{xy}, called *transport surplus*. The set of types $\mathcal{X}$ and $\mathcal{Y}$ are finite, and the total number of workers and firms are identical. We denote by (p_x) and (q_y) the vectors of probability distribution over $\mathcal{X}$ and $\mathcal{Y}$, thus normalizing the total mass of workers and firms to one: $\sum_{x\in\mathcal{X}} p_x = \sum_{y\in\mathcal{Y}} q_y = 1$.

The central planner's problem is to form a matching, and therefore, to decide on the mass π_{xy} of pairs xy to form. In the sequel, we shall call π_{xy} the *optimal transport plan*. This quantity must match everyone, namely satisfy the double set of constraints that all workers of each type x are assigned, $\sum_{y\in\mathcal{Y}} \pi_{xy} = p_x$, and that all firms of type y are assigned, namely $\sum_{x\in\mathcal{X}} \pi_{xy} = q_y$. With these constraints in mind, the workers shall maximize total output, which is $\sum_{x\in\mathcal{X},y\in\mathcal{Y}} \pi_{xy}\Phi_{xy}$ the sum of the pairwise output weighted by the mass of each pair. This yields the problem

$$\max_{\pi\geq 0} \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \pi_{xy}\Phi_{xy} \tag{1}$$
$$s.t. \begin{cases} \sum_{y\in\mathcal{Y}} \pi_{xy} = p_x \ \forall x \in \mathcal{X} \\ \sum_{x\in\mathcal{X}} \pi_{xy} = q_y \ \forall y \in \mathcal{Y} \end{cases}$$

which is clearly a linear programming problem, which we shall call the *primal problem*. Note that the set of π satisfying the constraints is clearly nonempty, as the *random matching* obtained by $\pi_{xy} = p_x q_y$ does satisfy the constraints. However, this matching is not optimal in general.

It is a basic result in linear programming that in this case, the value of the primal problem coincides with the value of the *dual problem*, which is

$$\begin{aligned} &\min_{u,v} \sum_{x \in \mathcal{X}} p_x u_x + \sum_{y \in \mathcal{Y}} q_y v_y \\ &s.t. u_x + v_y \geq \Phi_{xy} \ \forall x \in \mathcal{X}, y \in \mathcal{Y} \end{aligned} \tag{2}$$

where the dual variables u_x and v_y are the Lagrange multipliers respectively associated with the primal constraints $\sum_{y \in \mathcal{Y}} \pi_{xy} = p_x$ and $\sum_{x \in \mathcal{X}} \pi_{xy} = q_y$, while the primal variables $\pi_{xy} \geq 0$ serve as Lagrange multipliers associated with the dual constraints $u_x + v_y \geq \Phi_{xy}$.

Lastly, another important result in linear programming, *complementary slackness*, asserts that $\pi_{xy} > 0 \implies u_x + v_y = \Phi_{xy}$: if a Lagrange multiplier is strictly positive, then the corresponding dual constraint is saturated.

Optimal transport is a far-reaching generalization of the finite-dimensional duality discussed above to the case when $\mathcal{X}$ and $\mathcal{Y}$ are much richer sets; in particular the theory applies to the case when $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional vector spaces, and we will not need more for most of the economic applications we will discuss. In that case, letting P and Q be probability distributions over $\mathcal{X}$ and $\mathcal{Y}$, we shall define $\mathcal{M}(P, Q)$ as the set of joint probability distributions over $\mathcal{X} \times \mathcal{Y}$ with margins P and Q, which is the set of joint

probability distributions π such that if $(X, Y) \sim \pi$, where $\sim$ is understood as "distributed as," then $X \sim P$ and $Y \sim Q$. In this more general setting, the primal problem (1) extends to

$$\max_{\pi \in \mathcal{M}(P,Q)} \int_{\mathcal{X} \times \mathcal{Y}} \Phi(x, y)\, d\pi(x, y) \tag{3}$$

while its dual, extending (2), is

$$\begin{aligned} \min_{u,v} \quad & \int_{\mathcal{X}} u(x)\, dP(x) + \int_{\mathcal{Y}} v(y)\, dQ(y) \\ s.t. \quad & u(x) + v(y) \geq \Phi(x, y) \end{aligned} \tag{4}$$

The **Monge-Kantorovich theorem** provides assumptions under which the former duality results are preserved in more general settings, that is: (i) there exist primal solutions (π_{xy}); (ii) there is no duality gap, that is, the value of the dual problem (4) coincides with the value of the primal (3); and (iii) there exist dual solutions (u_x) and (v_y).

2.2 Some variants

2.2.1 Entropy regularized Optimal Transport

To facilitate computation, consider the previous primal problem with an entropic regularization in the objective function. Take $\sigma > 0$ a parameter

that can be made arbitrarily small. The primal problem

$$\max_{\pi \geq 0} \sum_{x,y} \pi_{xy} \Phi_{xy} - \sigma \sum_{x,y} \pi_{xy} \ln \pi_{xy} \tag{5}$$
$$s.t. \begin{cases} \sum_y \pi_{xy} = p_x \\ \sum_x \pi_{xy} = q_y \end{cases}$$

has dual

$$\min_{u,v} \left\{ \sum_{x \in \mathcal{X}} p_x u_x + \sum_{y \in \mathcal{Y}} q_y v_y + \sigma \sum_{x,y} \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma} \right) - \sigma \right\}, \tag{6}$$

and the optimal π in (5) and the optimal (u, v) in (6) are related by

$$\pi_{xy} = \exp\left(\frac{\Phi_{xy} - u_x - v_y}{\sigma} \right). \tag{7}$$

Again, this problem has extension to the case where $\mathcal{X}$ and $\mathcal{Y}$ are no longer discrete sets: this is the theory of *Bernstein-Schrödinger systems*, surveyed in Léonard (2014). Recently, progresses have been made on the computation of this problem in particular through *coordinate descent*:

Starting at $t = 0$ with an initial estimate of v_y^t :

- Compute (u_x^{t+1}) in order to minimize the objective function in (6), while keeping $v = (v_y^t)$ fixed.
- Compute (v_y^{t+1}) in order to minimize the objective function in (6), while keeping $u = (u_x^{t+1})$ fixed.
- Iterate over t, until the update to the u's and the v's are below toler-

ance.

It is easy to see that both steps are explicit and one full iteration of the algorithm expresses as

$$\begin{cases} u_x^{t+1} = \sigma \log(\frac{1}{p_x} \sum_{y \in \mathcal{Y}} \exp(\frac{\Phi_{xy} - v_y^t}{\sigma})) \\ v_y^{t+1} = \sigma \log(\frac{1}{q_y} \sum_{y \in \mathcal{Y}} \exp(\frac{\Phi_{xy} - u_x^{t+1}}{\sigma})). \end{cases} \tag{8}$$

This is the *iterated proportional fitting algorithm* (IPFP), which has been rediscovered under many names[1]: "matrix scaling", "RAS algorithm", "Sinkhorn-Knopp algorithm", "Kruithof's method", "Furness procedure", "biproportional fitting procedure", "Bregman's procedure". In economics, this algorithm has been proposed at least twice, once by Berry, Levinsohn and Pakes 1995 under the name "contraction mapping algorithm," and once by Guimares-Portugal (2010) in the context of the gravity equation in trade. We will see some of these connections below. This algorithm has been successfully applied to machine learning, see Cuturi (2013) and Peyré and Cuturi (2019). The rates of convergence of this algorithm are by now well understood thanks to the Hilbert projective metric, see Franklin and Lorenz (1989), and more recently, thanks to the theory of Bregman divergences, see Léger (2020).

2.2.2 Variant with unassigned agents

A variant of the problem leaves the agents the possibility of remaining unassigned. When the total mass of workers differs from that of the firms, we still

[1] See Idel (2016) for a historical survey.

denote p_x be the mass of workers of type x (no longer interpreted as a probability) and q_y the mass of firms of type y, and we allow for $\sum_x p_x \neq \sum_y q_y$ to hold. The constraint is now that the total mass of matched workers of type x should be no greater than p_x, and that the total mass of matched firms of type y should not exceed q_y; and the total surplus is still $\sum_{xy} \pi_{xy}\Phi_{xy}$; the primal problem is now

$$\max_{\pi \geq 0} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \pi_{xy}\Phi_{xy} \tag{9}$$
$$s.t. \begin{cases} \sum_{y \in \mathcal{Y}} \pi_{xy} \leq p_x \ \forall x \in \mathcal{X} \\ \sum_{x \in \mathcal{X}} \pi_{xy} \leq q_y \ \forall y \in \mathcal{Y} \end{cases}$$

while the dual problem becomes

$$\min_{u \geq 0, v \geq 0} \sum_{x \in \mathcal{X}} p_x u_x + \sum_{y \in \mathcal{Y}} q_y v_y \tag{10}$$
$$s.t. u_x + v_y \geq \Phi_{xy} \ \forall x \in \mathcal{X}, y \in \mathcal{Y}$$

As we see, these formulations only slightly differ from (1) and (2) respectively: the constraints in the primal switch from an equality to an inequality, while the variables in the dual are now subject to nonnegativity constraints. Because this leaves the possibility of agents to remained unmatched (unemployed in the labor market; singles in the marriage market), these problems are sometimes more relevant for economic modelling. This model is called the Becker-Shapley-Shubik model, after Becker (1973) and Shapley-Shubik (1971).

2.3 Inverse optimal transport problem

Understanding the "direct problem" of optimal transport as determining the optimal transport plan π_{xy} in (1) or (5) based on the transport surplus Φ_{xy}, as described above, we now turn to the "inverse problem" of optimal transport: how to determine the transport surplus Φ_{xy} based on the observation of an optimal transport plan $\hat{\pi}_{xy}$. More specifically, we specify

$$\Phi_{xy}^{\lambda} = \sum_{k} \lambda_k \phi_{xy}^{k}. \tag{11}$$

Setting $\theta = (\lambda_k, u_x, v_y)$ and

$$\pi_{xy}^{\theta} = \exp\left(\Phi_{xy}^{\lambda} - u_x - v_y\right), \tag{12}$$

the inverse optimal transport problem consists of seeking the parameter θ such that π_{xy}^{θ} has the same margins and moments as $\hat{\pi}_{xy}$, that is

$$\sum_{y\in\mathcal{Y}} \pi_{xy}^{\theta} = \sum_{y\in\mathcal{Y}} \hat{\pi}_{xy} =: p_x, \ \sum_{x\in\mathcal{X}} \pi_{xy}^{\theta} = \sum_{x\in\mathcal{X}} \hat{\pi}_{xy} =: q_y, \ \sum_{x,y} \pi_{xy}^{\theta} \phi_{xy}^{k} = \sum_{x,y} \hat{\pi}_{xy} \phi_{xy}^{k}. \tag{13}$$

This question solved by the following convex optimization problem:

Theorem 1 *The unique λ satisfying conditions (13) is unique solution to*

$$\min_{u,v,\lambda} \left\{ \sum_{x\in\mathcal{X}} p_x u_x + \sum_{y\in\mathcal{Y}} q_y v_y + \sum_{x,y} \exp\left(\Phi_{xy}^{\lambda} - u_x - v_y\right) - \sum_{xy} \hat{\pi}_{xy} \Phi_{xy}^{\lambda} \right\} \tag{14}$$

whose dual is

$$\begin{aligned} & \max_{\pi \geq 0} \left\{ -\sum_{xy} \pi_{xy} \ln \pi_{xy} \right\} \\ s.t. \quad & \sum_{y \in \mathcal{Y}} \pi_{xy} = p_x \ \ [u_x] \,, \ \ \sum_{x \in \mathcal{X}} \pi_{xy} = q_y \ \ [v_y] \,, \\ & \sum_{x,y} \pi_{xy} \phi_{xy}^k = \sum_{x,y} \hat{\pi}_{xy} \phi_{xy}^k \ \ [\lambda_k] \,. \end{aligned} \tag{15}$$

The problem of parametric estimation of λ is therefore the problem of a Poisson pseudo-maximum likelihood estimation, similar to the technique employed in trade to estimate the gravity equation (Santos Silva and Tenreyro, 2006). Galichon and Salanié (2021) formulated the initial connection with the Choo-Siow (2006) matching model, in the variant with singles. Dupuy and Galichon (2014) studies a continuous version of this model. Dupuy, Galichon and Sun (2019) add a Lasso-type penalization to estimate λ under sparsity constraint, while Carlier, Dupuy, Galichon and Sun (2021) offer an algorithm called SISTA (Sinkhorn+Iterative Soft Thresholding Algorithm) to compute efficently the regularized problem by alternating coordinate descent steps (Sinkhorn steps) on the u_x's and the v_y's, with a proximal gradient descent step.

3 Optimal transport in economics, finance and statistics

3.1 Family economics

As first understood by Becker (1973) and Shapley-Shubik (1971), the duality in optimal transport can be thought as a powerful welfare theorem, providing the equivalence between optimal matchings (in the sense of the problem of a central planner), and stable matchings (in a sense to be specified). Becker applied this insight in his pioneering analysis of the marriage market, and we now describe his analysis.

Consider the "marriage" problem of heterosexual men and women who need to decide to match. Men are distributed according to a a mass vector (p_x), while women are distributed according to a mass vector (q_y), where the total mass of men and women don't have to coincide. It is assumed that if x and y decide to match, they enjoy a joint utility Φ_{xy}, which they need to split among them. Any agent remaining unmatched gets a reservation utility equal to zero.

A *stable marriage* is a specification of a joint distribution $(\pi_{xy}) \geq 0$ over $\mathcal{X} \times \mathcal{Y}$ as well as payoffs vectors u_x and v_y such that

$$\begin{cases} \sum_{y \in \mathcal{Y}} \pi_{xy} + \pi_{x0} = p_x, \ \sum_{x \in \mathcal{X}} \pi_{xy} + \pi_{0y} = q_y \\ u_x + v_y \geq \Phi_{xy} \\ u_x \geq 0, \ v_y \geq 0 \\ \pi_{xy} > 0 \implies u_x + v_y = \Phi_{xy} \\ \pi_{x0} > 0 \implies u_x = 0, \pi_{0y} > 0 \implies v_y = 0 \end{cases} \tag{16}$$

The first set of conditions implies that all agents either participate in the matching market, or remain unmatched. The next conditions, namely $u_x + v_y \geq \Phi_{xy}$, imply that there is no blocking pair: if $u_x + v_y$ were less than Φ_{xy}, then x and y would have an incentive to quit their existing assignments and form a blocking pair, and each achieve strictly greater utility than u_x and v_y respectively. Similarly, $u_x \geq 0$ and $v_y \geq 0$ indicate that no one can achieve an outcome worse than the reservation utility.

Finally, the last set of conditions expresses that if pairs xy are actually formed, then there must be a way to split the joint surplus Φ_{xy} in such a way that u_x and v_y sum to Φ_{xy}, while if a positive mass of either x or y remain unmatched at equilibrium, the payoff of the corresponding type should be zero.

It is not hard to see that equations (16) are the complementary slackness conditions associated with linear programming problem (9)-(10). Hence:

Theorem 2 (Becker-Shapley-Shubik) *(π, u, v) is a stable marriage in the sense of (16) if and only if π is an optimal solution to (9), and (u, v) is an optimal solution to (10).*

This linear programming formulation is especially attractive for computational purposes, see chapter 3.4 of Galichon (2016).

3.2 Labor economics

In a realistic model of the labor market, not all jobs offering the same wage are as attractive for the workers. Hence, we need to capture the job amenities as the monetary valuations for working certain type of jobs conditional

on being a certain type of worker. Let α_{xy} be the monetary valuation of employer y's amenities for worker x, and let γ_{xy} be the monetary output of worker x working for employer y. As before, we normalize to zero the payoff of unassigned agents.

Let w_{xy} be the wage that x receives if working for y, which is determined at equilibrium. The worker and the firm problems are respectively

$$u_x = \max_{y \in \mathcal{Y}} \{\alpha_{xy} + w_{xy}, 0\} \text{ and } v_y = \max_{x \in \mathcal{X}} \{\gamma_{xy} - w_{xy}, 0\} \tag{17}$$

from which it follows that, defining the total output associated with an xy match as the sum of monetary amenity plus production, namely $\Phi_{xy} = \alpha_{xy} + \gamma_{xy}$, an equilibrium on the labor market should be such that (π, u, v) should be a stable matching in the sense of (16). Once (u, v), which is solution to (10) has been computed, one can compute the vector of equilibrium wages w_{xy} by

$$\gamma_{xy} - v_y \leq w_{xy} \leq u_x - \alpha_{xy}. \tag{18}$$

Note that for pairs xy that are actually formed at equilibirum, $\pi_{xy} > 0$ implies that $u_x + v_y = \Phi_{xy}$, and thus the upper bound $u_x - \alpha_{xy}$ coincides with the lower bound $\gamma_{xy} - v_y$. For other pairs, the upper bound may differ from the lower bound, which is a typical situation in equilibrium, where the price vectors need not be unique outside of the equilibrium path.

3.3 Trade

The structural gravity equations in international trade, introduced by Anderson (2003), with antecedents in Alan Wilson (1969), has been described

as a "workhorse" model in that field (Head and Mayer, 2013). Letting $\mathcal{X}$ be the set of countries, we define $\hat{\pi}_{xy}$ as the observed trade flow from country $x \in \mathcal{X}$ to country $y \in \mathcal{X}$. Letting $p_x = \sum_{y \neq x} \hat{\pi}_{xy}$ be the total volume of country x's exports, and $q_y = \sum_{x \neq y} \hat{\pi}_{xy}$ be the total volume of country y's imports, the gravity model assumes that

$$\pi_{xy}^{\lambda,u,v} = \exp\Big(\Phi_{xy}^{\lambda} - u_x - v_y\Big) \tag{19}$$

where $\Phi_{xy}^{\lambda} = \sum_k \phi_{xy}^k \lambda_k$ and the ϕ_{xy}^k's are various measures of proximity between country x and country y. The exporter and importer fixed effects u_x and v_y are called "multilateral resistances" and are adjusted by fitting the total imports and exports

$$\begin{cases} p_x = \sum_{y \neq x} \exp\big(\Phi_{xy}^{\lambda} - u_x - v_y\big) \\ q_y = \sum_{x \neq y} \exp\big(\Phi_{xy}^{\lambda} - u_x - v_y\big) \end{cases} . \tag{20}$$

As understood by Wilson (1969), $\pi^{\lambda,u,v}$ is the solution to the regularized optimal transport problem (5), while u_x and v_y are solution to its dual (6). Moreover, $\theta = (\lambda, u, v)$ can estimated as an inverse optimal transport problem (14), as suggested by the influencial paper of Santos Silva and Tenreyro (2006), who connect the procedure with a Poisson regression. The link with inverse optimal transport and matching problems is made in Dupuy, Galichon and Sun (2019).

3.4 Hedonic models

Consider a quasilinear hedonic model where each producer $x \in \mathcal{X}$ produces one unit of good and chooses in which quality $z \in \mathcal{Z}$. Each consumer $y \in \mathcal{Y}$ consumes one unit of good, and chooses in which quality $z \in \mathcal{Z}$. The mass of the producers and consumers are respectively distributed according to vectors (p_x) and (q_y). There is a price P_z, determined at equilibrium, for one unit of the good in quality z, and a producer of type x incurs a profit $P_z - C_{xz}$ of producing quality z at that price where C is a cost, while a consumer of type y derives a utility $U_{yz} - P_z$ of consuming utility z at that price. Both producers and consumers can opt out of the market and get profit or utility zero in that case.

In a hedonic equilibrium (Ekeland, Heckman and Nesheim, 2004), demand and supply are formed by the producer's and consumer's problems which are respectively

$$u_x = \max_{z \in \mathcal{Z}} \{P_z - C_{xz}, 0\} \text{ and } v_y = \max_{z \in \mathcal{Z}} \{U_{yz} - P_z, 0\}. \tag{21}$$

Chiappori, McCann and Nesheim (2010) have shown that this problem is actually an optimal transport problem of the type (9) between consumers and producers, with a matching surplus equal to

$$\Phi_{xy} = \max_{z} \{U_{yz} - C_{xz}\} \tag{22}$$

and the indirect utilities u_x and v_y are determined by (10). The intuition for the result is limpid: if x and y decide to exchange a good, they should

pick the good which is cost efficient in the sense that it maximizes their total joint surplus. The price vector P_z will be deduced from u_x and v_y by the set of inequalities

$$\min_{x\in\mathcal{X}}\{u_x + C_{xz}\} \geq P_z \geq \max_{y\in\mathcal{Y}}\{U_{yz} - v_y\} \tag{23}$$

where – similarly to the wage determination in equation (18) – the lower bound and the upper bound will coincide as soon as the quality z is actually traded at equilibrium.

3.5 Discrete choice models

Recently, an intimate connection between optimal transport theory and discrete choice models has been explored, which we now describe. Consider the (additive) discrete choice problem where a consumer i drawn from a population faces a choice between a finite set of alternatives $y \in \mathcal{Y}$. Consumer i's problem is

$$u(\varepsilon_i) = \max_{y\in\mathcal{Y}}\{V_y + \varepsilon_{iy}\} \tag{24}$$

where V_y is the systematic utility that every consumers associate with alternative y, and $(\varepsilon_{iy})_{y\in\mathcal{Y}}$ is drawn from a random vector over $\mathbb{R}^{\mathcal{Y}}$ with distribution $\mathbf{P}$, which is assumed to have a density. The distribution of the random part of the utility ε induces a choice probability, or market share $Q_y(V)$ which is the probability that y is chosen by a consumer i drawn from

the population, formally expresses as[2]

$$Q_y(V) = \Pr\left(y \in \arg\max_{y \in \mathcal{Y}} \{V_y + \varepsilon_{iy}\}\right). \tag{25}$$

The *demand inversion problem*, popularized by Berry (1994) and Berry, Levinsohn and Pakes (1995, hereafter BLP) consists of, given a vector of market shares q_y, how to look for a vector of systematic utility V such that $Q(V) = q$. This problem is a key step in BLP's estimation procedure, which consists of computing V by demand inversion, and then running an instrumental variable regression on V.

Galichon and Salanié (2021) showed that the problem of discrete choice inversion is, in fact, isomorphic to an optimal transport problem.

Theorem 3 (Galichon-Salanié, part 1) *The following statements are equivalent:*

(i) $Q(V) = q$*, that is* V *is the solution to inversion problem of the discrete choice model in (24), and*

(ii) There exist (u, v) *with* $v = -V$ *such that* (u, v) *is solution to the dual optimal transport problem with surplus* $\Phi(\varepsilon, y) := \varepsilon_y$

$$\begin{aligned} &\min_{u,v} \int_{\varepsilon \in \mathbb{R}^{\mathcal{Y}}} u(\varepsilon)\, d\mathbf{P}(\varepsilon) + \sum_{y \in \mathcal{Y}} q_y v_y \\ &s.t.\ u(\varepsilon) + v_y \geq \varepsilon_y \ \forall \varepsilon \in \mathbb{R}^{\mathcal{Y}}, \forall y \in \mathcal{Y}. \end{aligned} \tag{26}$$

This result was extended to the nonsmooth case (where no regularity as-

[2]Note that as ε has a density, the probability of ties is zero, and therefore the arg max has almost surely one element.

sumption is made on the distribution of ε) by Chiong, Galichon and Shum (2016), where a linear programming approch was provided for computational purposes. It has been extended to the continuous choice by Chernozhukov, Galichon, Henry and Pass (2021), and beyond additive random utility models by Bonnet et al. (2021).

A philosophical consequence of theorem 3 is that – at least from a mathematical standpoint – there is no relevant distinction between "one-sided" and "two-sided"models. We think of a discrete choice problem as a situation where conscient creatures called "consumers" choose inanimate objects called "yogurts". However, the equivalence described in theorem 3 shows that this situation is mathematically equivalent to a situation where consumers and yogurts would match, which is itself fully equivalent to a situation where yogurts choose consumers! This is a manifestation of Coase's principle: no matter how the utility is initially distributed, that is, no matter if consumers have preferences for yogurts or if yogurts have preferences for consumers, a Pareto efficient outcome should be reached in any case, and the bargaining process, here the yogurt price adjustment, allows to implement this outcome.

Interestingly, theorem 3 can be extended to mixed logit models, such as BLP's random coefficient logit model. Consider now a variant

$$u\left(\varepsilon_i\right) = \max_{y \in \mathcal{Y}} \left\{ V_y + \varepsilon_{iy} + \sigma \eta_y \right\} \tag{27}$$

where $(\varepsilon_y) \sim \mathbf{P}$ as before, while $\left(\eta_y\right)$ is a vector of i.i.d. random variables with a Gumbel distribution, independent from (ε_y). Let $Q_y^\sigma\left(V\right)$ be the

corresponding market share defined for each entry $y \in \mathcal{Y}$.

Theorem 4 (Galichon-Salanié, part 2) *The following statements are equivalent:*

(i) $Q^{\sigma}(V) = q$*, that is* V *is the solution to inversion problem of the discrete choice model in (27), and*

(ii) There exist (u, v) *with* $v = -V$ *such that* (u, v) *is solution to the dual regularized optimal transport problem with surplus* $\Phi(\varepsilon, y) := \varepsilon_y$

$$\min_{u,v} \int_{\varepsilon \in \mathbb{R}^{\mathcal{Y}}} u(\varepsilon) \, d\mathbf{P}(\varepsilon) + \sum_{y \in \mathcal{Y}} q_y v_y + \sigma \sum_{y \in \mathcal{Y}} \int_{\varepsilon \in \mathbb{R}^{\mathcal{Y}}} \exp\left(\frac{\varepsilon_y - u(\varepsilon) - v_y}{\sigma}\right) d\varepsilon. \tag{28}$$

Note that (28) is the same problem as (6) where the summation on ε has been replaced by a continuous integrals; however, in the sample version, we considering a sample $\varepsilon_1, ..., \varepsilon_N$ from distribution $\mathbf{P}$, and the integrals are replaced by sums.

As shown in Bonnet et al. (2021), the coordinate descent algorithm described in paragraph 2.2.1 coincides with BLP's celebrated "contraction mapping algorithm." This observation led the former authors to propose a demand inversion procedure that extends to the non-additive case.

3.6 Derivative pricing

Consider two stocks, and let X and Y be random variables standing for the value of these stocks at a horizon of time in the future. The fundamental theorem of asset pricing (see Duffie 1992) asserts that if there is a complete market of options with X as an underlying, then there is a distribution $\mathbf{P}$

called *martingale measure* such that the price of an option whose payoff is $u(X)$ shall be $\mathbb{E}_{\mathbf{P}}[u(X)]$. We shall assume that this is the case, and that there is a martingale measure $\mathbf{Q}$ such that the price of any option with payoff $v(Y)$ is $\mathbb{E}_{\mathbf{Q}}[v(Y)]$.

However, we shall not assume that there is a complete market of options on the joint realization of the underlying pair (X, Y), hence we cannot infer a joint martingale measure $\pi(x, y)$ based on the quoted prices. For a trader wishing to introduce a new option on the pair (X, Y), some restrictions must however be considered; in particular, if the option's payoff is of the form $a(X) + b(Y)$, its price must be $\mathbb{E}_{\mathbf{P}}[a(X)] + \mathbb{E}_{\mathbf{Q}}[b(Y)]$, otherwise the trader would face an arbitrage opportunity. But in general, the price of an option with a payoff $\Phi(X, Y)$ that is not additively separable cannot exceed

$$\max_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_{\pi}[\Phi(X, Y)]. \tag{29}$$

The Monge-Kantorovich duality will give us *sharp arbitrage bounds* for the price of this option, and will provide arbitrage strategies, as explained in Galichon, Henry-Labordère and Touzi (2014):

Theorem 5 *An option whose payoff $\Phi(X, Y)$ is priced at V is not subject to an arbitrage opportunity based on the two single-underlying option markets if and only if*

$$\max_{\substack{u,v \\ s.t.\ u(x)+v(y) \leq \Phi(x,y)}} \mathbb{E}_{\mathbf{P}}[u(X)] + \mathbb{E}_{\mathbf{Q}}[v(Y)] \leq V \leq \min_{\substack{u,v \\ s.t.\ u(x)+v(y) \geq \Phi(x,y)}} \mathbb{E}_{\mathbf{P}}[u(X)] + \mathbb{E}_{\mathbf{Q}}[v(Y)] \tag{30}$$

In other words, the price of the option should be bounded above by the

price of the *cheapest overreplicating portfolio*, while it should be bounded below by the price of the *costliest underreplicating portfolio*.

The above discussion has assumed that the pair of underlyings X and Y were the realizations of two assets prices at the same time. However, some derivatives are written on the same underlying asset at two different dates in the future. Assume that X is the value of a stock at a future date, and Y is the stock value at a later date. We then have an additional restriction, which is that in any martingale measure, $\mathbb{E}_\pi [Y|X] = X$ expresses absence of arbitrage. The option bound problem (29) now becomes

$$\begin{aligned} \max_{\pi \in \mathcal{M}(P,Q)} \quad & \mathbb{E}_\pi [\Phi (X, Y)] \\ s.t. \quad & \mathbb{E}_\pi [Y|X] = X \end{aligned} \tag{31}$$

for which the Monge-Kantorovich duality extends and interprets as incorporating dynamic arbitrage strategies; see an exposition from a financial engineering's point of view in Pierre Henry-Labordère (2020)'s insightful book.

3.7 Quantiles

There is an intimate connection between optimal transport and the notion of quantile. Consider the optimal transport problem described in (3) with $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $P = \mathcal{U}([0, 1])$ the uniform distribution on the unit interval, Q a distribution with finite second moments, and $\Phi (x, y) = xy$.

Then, as explained in chapter 4 of Galichon (2016), the solution $(X, Y) \sim \pi$ to problem (3) is a random pair such that $Y = F_Q^{-1}(X)$. Further, the

solution (u, v) to problem (4) is such that $u'(x) = F_Q^{-1}(x)$ and $v'(y) = F_Q(y)$. Hence the primal solution involves the quantile transform, and the dual solutions are simply primitives of the quantile map and the cumulative distribution function.

3.7.1 Multivariate quantiles

This connection led to the definition of a notion of **multivariate quantiles**: when Y is multivariate, say has d dimensions, one can extend the above setting to $X \sim P = \mathcal{U}\left([0,1]^d\right)$ and to $\Phi(x, y) = x^\top y$ and, if (u, v) is a solution to problem (4) in that case, the map $x \to \nabla u(x)$ is defined as the multivariate quantile associated with distribution Q. By Brenier's theorem (Brenier 1987), $\nabla u(X)$ has distribution Q, generalizing the well-known fact in the univariate case that the quantile map associated with a distribution pushes the uniform distribution on the unit interval onto the distribution. This new notion of multivariate quantiles found applications to risk measures (Ekeland, Galichon and Henry, 2012), decision theory (Galichon and Henry, 2012), and multivariate depth (Hallin, Chernozhukov, Galichon and Henry, 2017).

3.7.2 Quantile regression

There is an intimate connection between optimal transport and quantile regression, that is explored in a series of paper by Carlier, Chernozhukov and Galichon (2016, 2017) and Carlier, Chernozhukov, De Bie and Galichon (2020). We follow the latter paper in the present exposition. Quantile regression (see Koenker, 2005) attempts to fit a parametric dependence of

the conditional τ-th quantile of a random variable Y conditional on the value of X, a random vector on $\mathbb{R}^k$, as

$$Q_{Y|X}(\tau|x) = \beta(\tau)^\top x. \tag{32}$$

where $\beta(\tau) \in \mathbb{R}^k$ is the parameter of interest, defined for each value of $\tau \in [0,1]$. Since Koenker and Bassett (1978), this problem has been recognized as a convex optimization problem in the population

$$\min_{\beta(\tau)\in\mathbb{R}^k} \mathbb{E}_{\mathbf{P}}\left[\rho_\tau\left(Y - \beta(\tau)^\top X\right)\right] \tag{33}$$

where the loss function $\rho_\tau(z) = \tau z^+ + (1-\tau)z^-$, and $\mathbf{P}$ denotes the joint distribution of (X,Y). The sample analog of (33) is a linear programming problem, yielding to a simple and computationally efficient estimation of β. The full curve $\tau \to \beta(\tau)$ can be estimated by summation of the objective functions in (33) over $\tau \in [0,1]$, yielding

$$\min_{\beta\in\mathbb{R}^{k\times[0,1]}} \mathbb{E}_{\mathbf{P}}\left[\int_0^1 \rho_\tau\left(Y - \beta(\tau)^\top X\right) d\tau\right]. \tag{34}$$

When specification (32) is correct, the map $\tau \to \beta(\tau)^\top x$ which is picked up is an actual quantile, and therefore nondecreasing. However, if specification (32) is incorrect, there is no guarantee that $\tau \to \beta(\tau)^\top x$ should be monotone. This phenomenon has been widely recognized in the literature on quantile regression and is known as the *quantile crossing problem*. To address the quantile crossing problem, one idea may be to impose directly the monotonicity of $\tau \to \beta(\tau)^\top x$ as an additional constraint in problem (34).

This has been done by Koenker and Ng (2005) but remains computationally challenging and the interpretation of the result is not obvious.

A more indirect approach consists of the following. Rather than imposing the monotonicity of $\tau \to \beta(\tau)^\top x$, one can impose the (weaker) constraint that $\tau \to 1\left\{y \geq \beta(\tau)^\top x\right\}$ should be nonincreasing in τ. Consider the problem

$$\begin{aligned} \min_{\beta \in \mathbb{R}^{k \times [0,1]}} \quad & \mathbb{E}_{\mathbf{P}}\left[\int_0^1 \rho_\tau\left(Y - \beta(\tau)^\top X\right) d\tau\right] \\ s.t. \quad & 1\left\{y \geq \beta(\tau)^\top x\right\} \geq 1\left\{y \geq \beta\left(\tau'\right)^\top x\right\} \ \forall \tau \leq \tau', \forall x \in \mathbb{R}^k, y \in \mathbb{R} \end{aligned} \tag{35}$$

The solution to the previous problem now has a very straightforward interpretation.

Theorem 6 (Carlier-Chernozhukov–De Bie-Galichon) *If the map β is solution to problem (33), then denoting $b(\tau) = \int_0^\tau \beta(t)\, dt$, and letting*

$$\psi(x, y) = \max_{\tau \in [0,1]} \left\{\tau y - x^\top b(\tau)\right\}, \tag{36}$$

the pair (b, ψ) will be solution to the following problem

$$\begin{aligned} & \max_{b,\psi} \mathbb{E}_{\mathbf{P}}[X]^\top \int_0^1 b(\tau)\, d\tau + \mathbb{E}_{\mathbf{P}}[\psi(X, Y)] \\ & s.t.\ x^\top b(\tau) + \psi(x, y) \geq \tau y. \ \blacksquare \end{aligned} \tag{37}$$

Conversely, if (b, ψ) is solution to problem (37) and if b is differentiable, then $\beta(\tau) = b'(\tau)$ is a solution to problem (33).

Theorem 6 sheds new insights on quantile regression. Indeed, an extension of Monge-Kantorovich duality worked out in Carlier, Chernozhukov and Galichon (2016) shows that problem (37) is the dual problem to

$$\begin{aligned} \min_{\pi \in \mathcal{M}(\mathcal{U}([0,1]),\mathbf{P})} \quad & \mathbb{E}_\pi \left[(Y-U)^2 \right] \\ s.t. \quad & \mathbb{E}\left[X|U\right] = \mathbb{E}\left[X\right] \end{aligned} \tag{38}$$

where $\pi \in \mathcal{M}\left(\mathcal{U}\left([0,1]\right), \mathbf{P}\right)$ means that if $(U,X,Y) \sim \pi$, then $U \sim \mathcal{U}\left([0,1]\right)$ and $(X,Y) \sim \mathbf{P}$. If (b,ψ) is a solution to (37) with b differentiable and $(U,X,Y) \sim \pi$ is a solution to (38), then letting $\beta(\tau) = b'(\tau)$, one has the representation

$$Y = X^\top \beta(U) \tag{39}$$

where X is mean-independent from U. Beyond the case when Y is scalar, this formulation allows to get a multivariate extension of quantile regression using the notion of vector quantiles; see Carlier, Chernozhukov, Galichon (2016).

3.8 Partial identification and random sets

Some problems in econometrics specify incomplete restrictions between a model and an observed variable. Assume, following Galichon and Henry (2011), that we observe a random variable $Y \sim Q$ valued in $\mathcal{Y}$, and that the restrictions given by the model specify $Y \in \Gamma_\theta(X)$, and $X \sim P$, where X is a data-generating process valued in $\mathcal{X}$ and $\theta \in \Theta$ is a parameter of the model. Γ_θ is a correspondence from $\mathcal{X}$ to $\mathcal{Y}$, such that $\Gamma_\theta(x)$ is a subset

of $\mathcal{Y}$. The *identified set* is the set Θ_I of $\theta \in \Theta$ such that there is a joint distribution $\pi \in \mathcal{M}(P,Q)$ with $\mathbb{E}_\pi[1\{Y \notin \Gamma_\theta(X)\}] = 0$. Such a problem can be recast as an optimal transport problem

$$V = \min_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi[1\{Y \notin \Gamma_\theta(X)\}] \tag{40}$$

By working on the dual of this problem, one obtains Strassen's theorem (Strassen, 1965)

$$V = \max_B \left\{ Q(B) - P\left(\Gamma^{-1}(B)\right) \right\}, \tag{41}$$

where the maximum extends over the Borel sets B of $\mathcal{Y}$. Therefore $\theta \in \Theta_I$ if and only if $Q(B) \leq P\left(\Gamma^{-1}(B)\right)$ for all B. The sample version of problem (40) allows to use optimal assignment algorithms as efficient computational tools to decide if $\theta \in \Theta_I$, and dual formulation (41) allows to do inference (Galichon and Henry, 2009).

3.9 Generalized linear models

Consider a generalized linear model (GLM) with 2-way fixed effects. The observations are ij; the dependent variable is $\hat{\pi}_{ij}$, while the explanatory variables are $\Phi = \left(\phi_{ij}^k\right)_{ij,k}$ for $k \in \{1, ..., K\}$ and the i and j fixed effects. If l is the link function, which is increasing and continuous, the model is written as

$$\mathbb{E}\left[\hat{\pi}_{ij} | \phi_{ij}^k, i, j\right] = l^{-1}\left(\left(\Phi\beta\right)_{ij} - u_i - v_j\right). \tag{42}$$

Denote $p_i = \sum_j \hat{\pi}_{ij}$ and $q_j = \sum_i \hat{\pi}_{ij}$ the margins of $\hat{\pi}$. Letting L be a primitive of l, and letting $L^*(w) = \max_z \{wz - L(z)\}$ be its convex conjugate,

which is a primitive of l^{-1}, one can show that the GLM model can be fit using

$$\min_{\beta} \left\{ W(\beta) - \sum_{ij} \hat{\pi}_{ij} (\Phi\beta)_{ij} \right\} \tag{43}$$

where

$$\begin{aligned} W(\beta) = \max_{\pi_{ij} \geq 0} \quad & \left\{ \sum_{ij} \pi_{ij} (\Phi\beta)_{ij} - \sum_{ij} L(\pi_{ij}) \right\} \\ s.t. \quad & \sum_{j} \pi_{ij} = p_i, \ \sum_{i} \pi_{ij} = q_j \end{aligned} \tag{44}$$

is a regularized optimal transport problem which can be equivalently expressed by its dual:

$$W(\beta) = \min_{u_i, v_j} \left\{ \sum_{i} p_i u_i + \sum_{j} q_j v_j + \sum_{ij} L^* \left((\Phi\beta)_{ij} - u_i - v_j \right) \right\}. \tag{45}$$

In particular, the log link function $l(z) = \ln z$ yields $l^{-1}(t) = \exp(t)$, and thus $L(z) = z(\ln z - 1)$, and $L^*(t) = \exp(t)$, and $W(\beta)$ is the solution to an entropy regularized optimal transport problem, as described in paragraph 2.2.1.

3.10 Hide-and-seek games

In 1953, von Neumann described the following two-person, zero-sum game. Let (K_{ij}) be a $N \times N$ matrix with positive terms. There are two players, "Hider" and "Seeker". Hider plays first and hides in a cell (i, j). Playing second, Seeker highlights either a row or a column they claims contains Hider. If Seeker's claim is correct, then Hider pays Seeker $K_{ij} > 0$, otherwise

0.

Hider's mixed strategy is described by a vector of probabilities π_{ij} of hiding in cell ij. Once Hider has played, Seeker picks either a column i' or a column j', whichever of these maximizes $\sum_{j'} K_{ij'}\pi_{ij'}$ over i and $\sum_{i'} K_{i'j}\pi_{i'j}$ over j. Let us denote (a, b) the vector of mixed strategies of Seeker, where $a_i \geq 0$ is the probability of highlighting a row i, and $b_j \geq 0$ is the probability of highlighting a column j, and $\sum_{i=1}^{n} a_i + \sum_{j=1}^{n} b_j = 1$. If Hider plays strategy π and Seeker plays strategy (a, b), the expected payoff of Seeker is therefore

$$\sum_{ij} (a_i + b_j) K_{ij}\pi_{ij}$$

and hence the value of this zero-sum game for Seeker is obtained by minimizing the above expression over $x_{ij} \geq 0$, $\sum_{ij} x_{ij} = 1$, and maximizing it over $(a, b) \geq 0$ such that $\sum_i a_i + \sum_j b_j = 1$.

Von Neumann showed that this game is intimately connected with an optimal transport problem. Indeed,

$$V^{-1} = \begin{array}{l} \max_{\pi \geq 0} \sum_{ij} \pi_{ij} K_{ij}^{-1} \\ s.t. \begin{cases} \sum_j \pi_{ij} \leq 1/n \\ \sum_i \pi_{ij} \leq 1/n \end{cases} \end{array} = \begin{array}{l} \min_{u \geq 0, v \geq 0} \sum_i \frac{u_i}{n} + \sum \frac{v_j}{n} \\ s.t.\ u_i + v_j \geq K_{ij}^{-1} \end{array}$$

and the solution $\pi_{ij} \geq 0$ to the primal problem yields Hider's optimal strategy, while setting $a_i = Vu_i/n$ and $b_j = Vv_j/n$ yields Seeker's optimal strategy.

Although von Neumann's paper appeared in 1953, it seems that this important connection between a zero-sum game and a linear programming

problem was known to him decades earlier, in anticipation of Dantzig's general connection between linear programming and zero-sum games, cf. Dantzig (1951). See a historical perspective in Kuhn and Tucker (1958).

4 The mathematics of optimal transport

4.1 Network formulation

As explained in Galichon (2016), chapter 8, the optimal transport problem has the structure of a *min-cost flow* problem. Introduce a network whose set of nodes is $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$ and whose set of arcs is $\mathcal{A} = \mathcal{X} \times \mathcal{Y}$. Such a network is called a *bipartite* one. Define an $\mathcal{A} \times \mathcal{Z}$ matrix ∇ which is such that $\nabla_{xy,z} = 1\{z = y\} - 1\{z = x\}$. Consider the "change of sign trick" where one defined $\tilde{q} = \left(-p^\top, q^\top\right)^\top$ and $\tilde{v} = \left(-u^\top, v^\top\right)^\top$. Define $c_{xy} = -\Phi_{xy}$. The vector $\tilde{q}$ should be interpreted as a vector of quantities, while the vector $\tilde{v}$ should be interpreted as a vector of prices.

Call $C(\tilde{q})$ the value of the optimal transport problem, which rewrites under its primal form as

$$\begin{array}{rll} C(\tilde{q}) = \min\limits_{\pi \geq 0} & & \pi^\top c \\ & s.t. \quad \nabla^\top \pi = & \tilde{q} \end{array} \tag{46}$$

when $q^\top 1_{\mathcal{Z}} = 0$, and $C(\tilde{q}) = +\infty$ otherwise. Equivalently, $C(\tilde{q})$ can be

expressed by its dual value as

$$\begin{aligned} C\left(\tilde{q}\right) = \max_{\tilde{v}} \quad & \tilde{q}^{\top}\tilde{v} \\ s.t. \quad & \nabla\tilde{v} \leq \quad c. \end{aligned}$$

This is an instance of the *min-cost flow problem*, which makes sense more generally on any (not necessarily bipartite) network.

4.2 Equilibrium expression

By convex duality, denoting

$$C^{*}\left(\tilde{v}\right) = \max_{\tilde{q}} \left\{ \tilde{q}^{\top}\tilde{v} - C\left(\tilde{q}\right) \right\}, \tag{47}$$

the convex conjugate of C, it can be seen that $C^{*}\left(\tilde{v}\right) = 0$ if $\nabla\tilde{v} \leq c$ and $+\infty$ otherwise, and one has

$$C\left(\tilde{q}\right) = \max_{\tilde{v}} \left\{ \tilde{q}^{\top}\tilde{v} - C^{*}\left(\tilde{v}\right) \right\}. \tag{48}$$

Further, the set of $\tilde{v} = (-u, v)$ where u and v are solutions to problem (2) are the maximizers of (48).

In the case of the entropy regularized problem (5)-(6), these expressions become respectively

$$\begin{aligned} C_{\sigma}\left(\tilde{q}\right) = \min_{\pi \geq 0} \quad & \pi^{\top}c + \sigma\pi^{\top}\log\pi \\ s.t. \quad & \nabla^{\top}\pi = \quad \tilde{q} \end{aligned} \tag{49}$$

and

$$C_\sigma^*(\tilde{v}) = \sigma 1^\top \exp\left(\frac{c - \nabla\tilde{v}}{\sigma}\right). \tag{50}$$

Keeping in mind the interpretation of $\tilde{q}$ as quantities and $\tilde{v}$ as prices, one should view $C(\tilde{q})$ as a cost function, expression $\tilde{v}^\top\tilde{q} - C_\sigma(\tilde{q})$ as a profit, and $C^*(\tilde{v})$ as an indirect profit function. Hence, expression (47) should be viewed as a profit maximization problem. The optimal transport problem consists of looking for the potentials $\tilde{v}$ that maximize $\tilde{q}^\top\tilde{v} - C^*(\tilde{v})$. By convex duality (see chapter 6 of Galichon, 2016), this is equivalent with

$$\tilde{v} \in \partial C(\tilde{q}), \tag{51}$$

which, still by convex analysis, is equivalent with

$$\tilde{q} \in \partial C^*(\tilde{v}). \tag{52}$$

Therefore, ∂C^* should be interpreted as a *supply correspondence*, while $\partial C = (\partial C^*)^{-1}$ should be interpreted as an *inverse supply correspondence*. The same intepretation extends immediately to the regularized versions of these objects.

4.3 Mathematical structures

The optimal transport problem is blessed with the priviledge to belong to the intersection of two rich theories: convex optimization and gross substitutes. There are, broadly speaking, two structures whithin which the equilibrium

problem

$$\tilde{q} \in Q\left(\tilde{v}\right) \tag{53}$$

is well understood.

- The first one is *convex optimization*: Q is the subdifferential of a convex function. Then the problem is a convex optimization problem, and convex optimization can be put to use to solve problem (53).

- The second case is *gross substitutes*: loosely speaking, q_x cannot increase when $\tilde{v}_y$ increases ($x \neq y$). This setting is needed for coordinate update algorithms such as Jacobi or Gauss-Seidel to converge, see Rheinboldt (1970).

In optimal transport, both structures are met, as we shall now see.

4.3.1 Convex optimization

Recall that the cost function C and the indirect cost function C^* defined above are convex functions, which are dual one to another in the sense of convex analysis. It follows that $\partial C^*\left(\tilde{v}\right)$ and $\partial C\left(\tilde{q}\right)$ are convex sets, and problems (51) and (52) can be solved as convex optimization problems dual to each other, respectively (48) and (47). In the unregularized case, these problems are linear programming problems. In the regularized case, the convexity structure is retained, but the problems are of course no longer linear.

4.3.2 Gross substitutes

When $\sigma > 0$ it is easy to see that the indirect profit function $C_\sigma^*(\tilde{v})$ is submodular. It is not very hard to extend this result to the unregularized case to show that $C^*(\tilde{v})$ is submodular as well. As a result, the corresponding supply function satisfies Kelso and Crawford's (1982) *gross substitutes* property.

A remark is in order here. It may be a surprise that the optimal transport problem has the gross substitutes property, as common sense suggests that workers and firms should be *complements*, and not substitutes. However, keep in mind the "change-of-sign trick" implemented at paragraph 4.1: we defined $\tilde{v} = \left(-u^\top, v^\top\right)^\top$, and therefore we switched the sign of the worker's payoffs (and of their quantities accordingly). This change of sign is the reason why the optimal transport problem, in spite of being a problem with complementarities, reformulates as a problem with gross substitutes. See Sun and Yang (2006).

We can formulate gross substitutes properties of C and C^* in the language of L- and M-convexity, introduced by Murota (1998). Indeed, as the domain of C is the set of $\tilde{q}$ such that $\tilde{q}^\top 1_{\mathcal{Z}} = 0$, and as $C^*(\tilde{v} + \lambda 1_{\mathcal{Z}}) = C^*(\tilde{v})$ for all $\lambda \in \mathbb{R}$, it follows that C is a *M-convex function* and C^* is a *L-convex function*, and the supply bundle $\partial C^*(\tilde{v})$ is a *M-convex set* of $\mathbb{R}^{\mathcal{Z}}$, while $\partial C(\tilde{q})$ is a *L-convex set*, still in the terminology of the same author. In particular, $\partial C(\tilde{q})$ is a *lattice*, while $\partial C^*(\tilde{v})$ is a *base polyhedron*.

4.4 Extensions

As we have seen just above, the optimal transport problem can be formulated (at least under its regularized form) as a set of nonlinear equations $Q(p) = q$, where Q happens to be the subdifferential of a convex function which is also submodular, and hence optimal transport belongs to both convexity *and* gross substitutes families. Some extensions of the optimal transport problem retain both convexity and gross substitutes. This is the case of the min-cost flow problem, for instance, as described in paragraph 4.1.

4.4.1 Problems that retain convexity, but not substitutability

Some problems retain convex optimization but not gross substitutes, such as one-to-many matching problems with transferable utility, see a related discussion in Azevedo and Hatfield (2018). Vector quantile regression, discussed above in paragraph 3.7.2, falls in that category, too.

In ongoing work with Pauline Corblet and Jeremy Fox 2020, we investigate a problem of dynamic matching that retains most of the convexity structure of optimal transport. The problem we study is a two-sided version of Rust (1987)'s model. More specifically, assume that conditional of a worker of type $X = x$ matching with a firm of type y, there is a probability $\mathbb{P}_{x'|xy}$ that the worker will transition to type x' at the next period, and a probability $\mathbb{Q}_{y'|xy}$ that the firm will transition to type y'. In this case, the joint matching surplus should be the sum of the short-term surplus Φ_{xy} and the expected discounted future payoffs of the worker and of the firm, respectively denoted $\beta\mathbb{P}\left[u_X|xy\right]$ and $\beta\mathbb{Q}\left[v_Y|xy\right]$.

In this case, when $\beta = 1$, Corblet et al. (2020) show that both the equilibrium computation and the estimation can be handled by the following saddle-point problem:

$$\max_{n,m} \min_{u,v,\lambda} H(n, m, u, v, \lambda)$$

where one has defined $H(n, m, u, v, \lambda) =$

$$\begin{cases} 2\sum_{xy\in\mathcal{X}\times\mathcal{Y}} \sqrt{n_x m_y} \exp\left(\frac{\sum_k \phi_{xy}^k \lambda_k + \mathbb{P}[u_X|xy] + \mathbb{Q}[v_Y|xy] - u_x - v_y}{2}\right) \\ + \sum_{x\in\mathcal{X}} n_x \exp\left(\sum_k \phi_{x0}^k \lambda_k + \mathbb{E}\left[u_{X'}|X = x\right] - u_x\right) \\ + \sum_{y\in\mathcal{Y}} m_y \exp\left(\sum_k \phi_{0y}^k \lambda_k + \mathbb{E}\left[v_{Y'}|Y = y\right] - v_y\right) \\ - \sum_{x\in\mathcal{X}} n_x - \sum_{y\in\mathcal{Y}} m_y \end{cases}$$

which is convex in (n, m) and concave in (u, v, λ). Corblet et al. (2020) use this formulation to derive an algorithm to estimate the structural parameter λ efficiently. They find that the algorithm extends to the case $\beta < 1$. Dupuy et al. (2020) apply these ideas to family economics and fertility decisions.

4.4.2 Problems that retain substitutability, but not convexity

On the contrary, some problems retain the gross substitutes property, but not the convexity one. This is the case with one-to-one matching models with nontransferable utility, as shown by Adachi (2000), and with one-to-one matching models with imperfectly transferable utility, to handle in particular taxes, salary caps, public goods, etc. See Galichon, Kominers and Weber (2019). Non-additive random utility models and hedonic models beyond quasi-linear utility are also in this case. To handle these challenges, a more

general framework is needed, the *equilibrium flow problem*, which is the subject of current ongoing work by the author with Larry Samuelson and Lucas Vernet (2021). The equilibrium flow problem posits three objects. First a network $(\mathcal{Z}, \mathcal{A})$ is defined as in paragraph 4.1, where $xy \in \mathcal{Z}$ is interpreted as the existence of a trade route from node x to node y, and whose node-incidence matrix is denoted ∇. Second, a vector of outflows $q \in \mathbb{R}^{\mathcal{Z}}$, where q_z is interpreted as the mass that must leave the network at z ($q_z < 0$ means that mass actually appears at z). One assumes that $\sum_{z \in \mathcal{Z}} q_z = 0$, so all the mass that enters the networks must leave it. Finally, a set of *connection functions* $G_{xy} : \mathbb{R} \to \mathbb{R}$ for each $xy \in \mathcal{A}$, which are increasing and whose interpretation is that $G_{xy}(p_y) - p_x$ is the profit of a *carry trade*, consisting of purchasing one unit of the commodity at price p_x at node x, shipping to y, and selling at price p_y at node y.

Given these inputs, the equilibrium flow problem consists of determining a vector of flows $\mu \in \mathbb{R}_+^{\mathcal{Z}}$ and prices $p \in \mathbb{R}^{\mathcal{Z}}$ such that:

(i) *mass balance* holds: the sum of mass that arrives at z minus the sum that leaves is equal to q_z, that is, $\nabla^\top \mu = q$.

(ii) *absence of arbitrage* holds: there cannot be a positive rent associated with the carry trade over any arc, that is, $p_x \geq G_{xy}(p_y)$ for any $xy \in \mathcal{A}$.

(iii) *individual rationality* holds: if the carry trade over arc xy is actually performed, then the associated profit cannot be negative, and thus, $\mu_{xy} > 0$ implies $p_x = G_{xy}(p_y)$.

Galichon, Samuelson and Vernet (2021) show that this framework is general enough to encompass optimal transport problems, min-cost flow problems including shortest path problems, matching models with imperfectly

transferable utility, hedonic models, and supply chain problems.

5 Concluding discussion

To conclude, an attempt should be made to explain the claim to "unreasonable effectiveness" of optimal transport in economics, alluding to a celebrated formula of Wigner (1960). We believe that one of the reasons of the prevalence of optimal transport in economics is that the former strikes a good compromise between what models would like to capture and what they are capable of capturing.

Economics is, in a broad sense, the study of complementarities: capital and labor, worker and firms, supply and demand, buyers and sellers... all exhibit some complementarity which is at the source of economic activity. However, as it is now well understood since the insights of Kelso and Crawford (1982), problems with complementarities are hard to handle, and in particular, hard to compute. Fortunately, due to the bipartite structure, the "change-of-sign trick" described in paragraph 4.1 allowed us to reformulate the problem as a problem with gross substitutes, and therefore, let us enjoy the computational and structural benefits of a problem with gross substitutes. In some sense, the bipartite structure of optimal transport is a meeting point between the complementarity that models would like to capture, and the substitutability structure that they they are able to capture.

To make another analogy with Physics, the situation is similar to the two-body problem in cosmology, which has a tractable formulation and can be fully worked out – while the n-body problem with n larger than two is

notoriously hard. Fortunately, just as in cosmology where many situations can be satisfactorily approximated by a two-body problem, in economics, many phenomenons can be captured using the bipartite approximation. We have surveyed some of these applications in the present paper, but certainly not in an exhaustive way. And optimal transport is a galaxy where there are many more planets, only waiting to be explored.

References

Adachi, H. (2000). On a characterization of stable matchings. *Economics Letters*, *68*(1), 43–49.

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*(1), 170–192.

Azevedo, E. M., & Hatfield, J. W. (2018). Existence of equilibrium in large matching markets with complementarities. *SSRN Electronic Journal*.

Becker, G. S. (1973). A theory of marriage: Part i. *Journal of Political Economy*, *81*(4), 813–846.

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, *25*(2), 242–262.

Bonnet, O., Galichon, A., Hsieh, Y.-W., O'Hara, K., & Shum, M. (2021). Yogurts choose consumers? estimation of random utility models via

two-sided matching [Available at SSRN - Social Science Research Network].

Brenier, Y. (1987). Polar factorization and monotone rearrangement of vector-valued functions. *C. R. Acad. Sci. Paris, Ser. I, Mathematical Analysis*, *305*, 805–808.

Carlier, G., Chernozhukov, V., De Bie, G., & Galichon, A. (2020). Vector quantile regression and optimal transport, from theory to numerics. *Empirical Economics*.

Carlier, G., Chernozhukov, V., & Galichon, A. (2016). Vector quantile regression: An optimal transport approach. *The Annals of Statistics*, *44*(3), 1165–1192.

Carlier, G., Chernozhukov, V., & Galichon, A. (2017). Vector quantile regression beyond the specified case. *Journal of Multivariate Analysis*, *161*, 96–102.

Carlier, G., Dupuy, A., Galichon, A., & Sun, Y. (2021, June 2). *SISTA: Learning optimal transport costs under sparsity constraints* (SSRN Scholarly Paper ID 3855961). Rochester, NY.

Chernozhukov, V., Galichon, A., Hallin, M., & Henry, M. (2017). Monge kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics*, *45*(1), 223–256.

Chernozhukov, V., Galichon, A., Henry, M., & Pass, B. (2021). Identification of hedonic equilibrium and nonseparable simultaneous equations. *Journal of Political Economy*, *129*(3), 842–870.

Chiappori, P.-A., McCann, R. J., & Nesheim, L. P. (2010). Hedonic price equilibria, stable matching, and optimal transport: Equivalence, topology, and uniqueness. *Economic Theory*, *42*(2), 317–354.

Chiong, K. X., Galichon, A., & Shum, M. (2016). Duality in dynamic discrete-choice models. *Quantitative Economics*, *7*(1), 83–115.

Choo, E., & Siow, A. (2006). Who marries whom and why. *Journal of Political Economy*, *114*(1), 175–201.

Ciscato, E., Dupuy, A., Fox, J., Galichon, A., & Weber, S. (2020). Family dynamics: Marriage, fertility and divorce [In progress].

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, *26*.

Dantzig, G. B. (1951). A proof of the equivalence of the programming problem and the game problem. *Activity analysis of production and allocation*, (13), 330–338.

Duffie, D. (1992). *Dynamic asset pricing theory* (Princeton, Ed.).

Dupuy, A., & Galichon, A. (2014). Personality traits and the marriage market. *Journal of Political Economy*, *122*(6), 1271–1319.

Dupuy, A., Galichon, A., & Sun, Y. (2019). Estimating matching affinity matrices under low-rank constraints. *Information and Inference: A Journal of the IMA*, *8*(4), 677–689.

Ekeland, I., Galichon, A., & Henry, M. (2012). Comonotonic measures of multivariate risks. *Mathematical Finance*, *22*(1), 109–132.

Ekeland, I., Heckman, J. J., & Nesheim, L. (2004). Identification and estimation of hedonic models. *Journal of Political Economy*, *112*, S60–S109.

Fox, Galichon, & Corblet. (2020). A dynamic model of two-sided matching [In progress.].

Franklin, J., & Lorenz, J. (1989). On the scaling of multidimensional matrices. *Linear Algebra and its Applications*, *114-115*, 717–735.

Galichon, A., Henry-Labordère, P., & Touzi, N. (2014). A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability*, *24*(1), 312–336.

Galichon, A. (2016). *Optimal transport methods in economics* (Princeton, Ed.).

Galichon, A., & Henry, M. (2009). A test of non-identifying restrictions and confidence regions for partially identified parameters. *Journal of Econometrics*, *152*(2), 186–196.

Galichon, A., & Henry, M. (2011). Set identification in models with multiple equilibria. *The Review of Economic Studies*, *78*(4), 1264–1298.

Galichon, A., & Henry, M. (2012). Dual theory of choice with multivariate risks. *Journal of Economic Theory*, *147*(4), 1501–1516.

Galichon, A., Kominers, S. D., & Weber, S. (2019). Costly concessions: An empirical framework for matching with imperfectly transferable utility. *Journal of Political Economy*, *127*(6), 2875–2925.

Galichon, A., & Salanié, B. (2021). Cupid's invisible hand: Social surplus and identification in matching models. *Review of Economic Studies*, *forthcoming*.

Galichon, A., Samuelson, L., & Vernet, L. (2021). The equilibrium flow problem [In progress].

Guimarães, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, *10*(4), 628–649.

Head, K., & Mayer, T. (2013). Gravity equations: Workhorse, toolkit, and cookbook.

Henry-Labordere, P. (2020). *Model-free hedging: A martingale optimal transport viewpoint - 1st edi*.

Idel, M. (2016). A review of matrix scaling and sinkhorn's normal form for matrices and positive maps.

Kelso, A. S., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica*, *50*(6), 1483–1504.

Koenker, R. (2005). *Quantile regression*.

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Koenker, R., & Ng, P. (2005). Inequality constrained quantile regression. *Sankhyā: The Indian Journal of Statistics (2003-2007)*, *67*(2), 418–440.

Kuhn, H. W., Tucker, A. W. et al. (1958). John von neumann's work in the theory of games and mathematical economics. *Bulletin of the American Mathematical Society*, *64*(3, Part 2), 100–122.

Léger, F. (2020). A gradient descent perspective on sinkhorn. *Applied Mathematics & Optimization*.

Léonard, C. (2014). A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete & Continuous Dynamical Systems*, *34*(4), 1533.

Murota, K. (1998). Discrete convex analysis. *Mathematical Programming*, *83*(1), 313–371.

Ortega, J. M., & Rheinboldt, W. C. (1970). *Iterative solution of nonlinear equations in several variables*.

Peyré, G., & Cuturi, M. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, *11*(5), 355–607.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, 999–1033.

Santambrogio, F. (2015). *Optimal transport for applied mathematicians: Calculus of variations, PDEs, and modeling*. Birkhäuser Basel.

Shapley, L. S., & Shubik, M. (1971). The assignment game i: The core. *International Journal of Game Theory*, *1*(1), 111–130.

Silva, J. M. C. S., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641–658.

Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, *36*(2), 423–439.

Sun, N., & Yang, Z. (2006). Equilibria and indivisibilities: Gross substitutes and complements. *Econometrica*, *74*(5), 1385–1402.

Villani, C. (2003). *Topics in optimal transportation*. American Mathematical Soc.

Villani, C. (2009). *Optimal transport: Old and new*. Springer-Verlag.

Von Neumann, J. (1953). 1. a certain zero-sum two-person game equivalent to the optimal assignment problem. *Contributions to the theory of games (am-28), volume ii* (pp. 5–12).

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. richard courant lecture in mathematical sciences delivered at new york university, may 11, 1959. *Communications on Pure and Applied Mathematics*, *13*(1), 1–14.

Wilson, A. G. (1969). The use of entropy maximising models, in the theory of trip distribution, mode split and route split. *Journal of transport economics and policy*, 108–126.