HAL
open science

A spatial Pareto exchange economy problem

Xavier Bacon, Guillaume Guillaume Carlier, Bruno Nazaret

To cite this version:

Xavier Bacon, Guillaume Guillaume Carlier, Bruno Nazaret. A spatial Pareto exchange economy problem. 2021. hal-03480323v2

HAL Id: hal-03480323

https://hal.science/hal-03480323v2

Preprint submitted on 16 Dec 2021

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.

A spatial Pareto exchange economy problem

X. Bacon * G. Carlier † B. Nazaret ‡

December 16, 2021

Abstract

We use convex duality techniques to study a spatial Pareto problem with transport costs and derive a spatial second welfare theorem. The existence of an integrable equilibrium distribution of quantities is nontrivial and established under general monotonicity assumptions. Our variational approach also enables us to give a numerical algorithm à la Sinkhorn and present simulations for equilibrium prices and quantities in one-dimensional domains and a network of French cities.

Keywords: exchange economy Pareto optimality, optimal transport, convex duality, second welfare theorem, entropic optimal transport, Sinkhorn algorithm.

MS Classification: 49Q22, 91B72.

1 Introduction

The present article proposes a simple model of efficient spatial trade in an exchange economy subject to transportation costs. The notion of efficiency we consider comprises two quite different aspects: efficiency of the transport (import/export) process and efficiency in the Pareto sense of the distribution of goods after import/export has taken place. We take as primitives the spatial (initial) distribution of goods, the transportation cost for these goods and the agent's preferences, given by a (possibly location dependent) utility

*CEREMADE, Université Paris Dauphine, PSL, and SAMM, Université Paris 1 Panthéon Sorbonne, xavier.bacon@dauphine.psl.eu

†CEREMADE, Université Paris Dauphine, PSL, and INRIA-Paris, MOKAPLAN, carlier@ceremade.dauphine.fr

‡SAMM, Université Paris 1 Panthï¿½on-Sorbonne, bruno.nazaret@univ-paris1.fr

function. The unknown is the (final) distribution of goods after trade among the different locations. Our approach is based on the maximization of a global criterion which is a weighted average of the utility of final consumption (Pareto efficiency) net of the minimal total transport cost taking into account mass conservation constraints (optimal transport).

Our model relies to a large extent on optimal transport theory which has been the object of an intensive and fruitful stream of research in the last three decades and has found numerous applications in PDEs, functional inequalities, probability theory, economics and more recently machine learning, see the textbooks of Villani [14] and Santambrogio [13], and for an overview of optimal transport methods in economics and econometrics, see Galichon [6]. Optimal transport aims at explaining *how* to transport mass efficiently (i.e. so as to minimize a total cost) between two given measures with the same total mass. A cornerstone of this theory is the Kantorovich duality which characterizes optimal plans in terms of a pair of dual variables which are naturally interpreted as prices.

In a spatially distributed exchange economy where various goods are available in different quantities according to location, agents located at different locations may have an interest to trade and they will certainly do so if they can all increase their utility for a small transport cost. Possible utility improvements is then the reason *why* some mass is actually transported. To fix ideas, think of two locations A and B, all the resources in water are concentrated at A and all the resources in food are concentrated at B, agents located at B will then import some water from A and export some food to B but the precise amount of water/food transported will depend on: i) the initial endowments of A and B, ii) the preferences of agents located at A and B and iii) on how costly transporting these goods from one location to another is. If we fix the demand for water of B and the demand for food of A, this is a simple optimal transport problem but these demands result from a Pareto (weighted utility maximization) problem between these two locations. A dual version of the problem consists in looking for prices for which markets clear and also reflect the cost of transporting the goods as efficiently as possible.

We propose a simple variational model that combines optimal transport (that is *how* to transport the goods) with Pareto efficiency requirements (that is *why* some goods are actually transported). The variational problem we consider to find Pareto efficient distributions involves separable optimal transport terms and a joint concave utility term. Interestingly, this has a somehow similar structure as steps of the celebrated JKO scheme of Jordan, Kinderlehrer and Otto [7] for Wasserstein gradient flows, see Ambrosio, Gigli

and Savaré [1] for a detailed account of this theory. Convex duality also enables us to derive supporting prices and to prove a sort of spatial second welfare theorem. From a mathematical point, the existence of integrable Pareto optimal final resource distributions is non trivial and we provide a complete proof using quite general monotonicity properties. We also prove existence of continuous prices supporting these Pareto distribution of the goods. The fact that the analysis resorts on convex minimization and duality enables us to use state of the art solvers based on the Sinkhorn algorithm for entropic optimal transport, see Cuturi [4], Benamou et al. [3], Cuturi and Peyré [10], Peyré [9]. We illustrate the practical use of our model by solving numerically one-dimensional examples as well as a graph model representing some cities in France.

The paper is organized as follows. Our variational spatial Pareto model and our assumptions are introduced in section 2. Section 3 is concerned with existence and duality results. Primal-dual optimality conditions are further analysed and a second welfare theorem is derived from the latter in section 4. Finally, section 5 presents an algorithm for the entropic regularization of the problem and numerical simulations for equilibrium prices and quantities in one-dimensional domains and a graph representing a network of French cities.

2 Assumptions and notations

We consider a region denoted by X which we assume to be a compact metric space. We consider agents populating X having initial endowments of N goods, given by nonnegative Borel measures $\mu_1, \ldots, \mu_N \in \mathcal{M}_+(X)^N$. Choosing a reference measure $m \in \mathcal{M}_+(X)$ such that μ_i is absolutely continuous with respect to m for $i = 1, \ldots, N$ (e.g. $m = \sum_{i=1}^N \mu_i$) we denote by α_i the density of μ_i with respect to m i.e.

$$\mu_i = \alpha_i m,\ \alpha_i \in L^1(X, m).$$

For each good i, there is a transport cost $c_i \in C(X \times X)$ which satisfies

$$c_i(x, y) \geq 0,\ c_i(x, x) = 0, \forall (x, y) \in X^2. \tag{2.1}$$

The quantity $c_i(x, y)$ represents the cost of transporting one unit of mass of good i from x to y.

The image (or pushforward) of a Borel measure θ by a Borel map T will be denoted $T_\# \theta$ in the sequel. More precisely, if Y and Z are compact metric spaces, if $\theta \in \mathcal{M}(Y)$ (i.e. θ is a Borel measure on Y) and if $T : Y \to Z$ is

Borel, then $T_\#\theta$ is defined by $T_\#\theta(B) = \theta(T^{-1}(B))$ for every Borel subset B of Z. Equivalently, using test-functions, $T_\#\theta$ is defined by

$$\int_Z \varphi \mathrm{d}T_\#\theta = \int_Y \varphi \circ T \mathrm{d}\theta,\ \forall \varphi \in C(Z).$$

In particular, if $\gamma \in \mathcal{M}_+(X \times X)$, the marginals of γ are $\mathrm{proj}_{1\#}\gamma$ and $\mathrm{proj}_{2\#}\gamma$, where $(\mathrm{proj}_1(x,y), \mathrm{proj}_2(x,y)) = (x,y)$ denote the canonical projections. Given ν_i a Borel measure on X, we set

$$T_{c_i}(\mu_i, \nu_i) := \begin{cases} \min_{\gamma_i \in \Pi(\mu_i,\nu_i)} \int_{X\times X} c_i(x,y)\mathrm{d}\gamma_i(x,y) \text{ if } \nu_i \geq 0 \text{ and } \nu_i(X) = \mu_i(X) \\ +\infty \text{ otherwise} \end{cases}$$

where $\Pi(\mu_i, \nu_i)$ is the set of transport plans between μ_i and ν_i i.e. the set of nonnegative Borel measures on $X \times X$ having μ_i and ν_i as marginals (so that $\gamma_i(A \times B)$ represents the amount of mass of good i initially in A transported to a location in B). In our model, transporting goods is costly but may be worth because it may result in some improvement in terms of the agent's preferences. These preferences are given by a (space-dependent) utility function U: $X \times \mathbb{R}^N_+ \to \mathbb{R} \cup \{-\infty\}$, which satisfies

$$\text{for } m\text{-a.e. } x \in X,\ \beta \in \mathbb{R}^N_+ \mapsto U(x,\beta) \text{ is usc, concave, nondecreasing}^1, \tag{2.2}$$

$$\text{for every } \beta \in \mathbb{R}^N_+,\ x \in X \mapsto U(x,\beta) \text{ is } m\text{-measurable.} \tag{2.3}$$

We shall also assume that there exists $(\overline{\alpha}_1, \ldots, \overline{\alpha}_N) \in L^1(X,m)^N$ such that

$$\overline{\alpha}_i \geq 0,\ \int_X \overline{\alpha}_i \mathrm{d}m = \int_X \alpha_i \mathrm{d}m,\ i = 1, \ldots, N, \tag{2.4}$$

and

$$\int_X U(x, \overline{\alpha}_1(x), \ldots, \overline{\alpha}_N(x))\mathrm{d}m(x) > -\infty. \tag{2.5}$$

Finally, we assume that $(x,\beta) \mapsto U(x,\beta)$ is sublinear with respect to β uniformly in $x \in X$, which means that for every $\delta > 0$ there exists C_δ such that for m-a.e. $x \in X$ one has

$$U(x,\beta) \leq \delta \sum_{i=1}^N \beta_i + C_\delta,\ \forall \beta \in \mathbb{R}^N_+. \tag{2.6}$$

Given $\nu := (\nu_1, \ldots, \nu_N) \in \mathcal{M}_+(X)^N$. Let us define

$$\mathcal{U}(\nu) := \begin{cases} \int_X U(x, \beta_1(x), \ldots, \beta_N(x))\mathrm{d}m(x) \text{ if } \nu_i = \beta_i m, \text{ for } i = 1, \ldots, m \\ -\infty \text{ otherwise} \end{cases}$$

[1] by nondecreasing we mean that $\beta - \beta' \in \mathbb{R}^N_+$ implies $U(x,\beta) \geq U(x,\beta')$.

and

$$\mathcal{T}_\mu(\nu) := \sum_{i=1}^{N} T_{c_i}(\mu_i, \nu_i).$$

We are here interested in allocations which maximize the average utility (Pareto efficiency) taking into account transportation cost of the different goods and therefore consider

$$\sup \left\{ \mathcal{U}(\nu) - \mathcal{T}_\mu(\nu),\ \nu = (\nu_1, \dots \nu_N) \in \mathcal{M}_+(X)^N \right\} \tag{2.7}$$

which can also be expressed as

$$\sup_{\beta \in L^1(X,m)^N} \left\{ \int_X U(x, \beta(x)) \mathrm{d}m(x) - \sum_{i=1}^{N} T_{c_i}(\alpha_i m, \beta_i m) \right\} \tag{2.8}$$

which incorporates the mass conservation constraints

$$\int_X \beta_i(x) \mathrm{d}m(x) = \int_X \alpha_i(x) \mathrm{d}m(x),\ i = 1, \dots, N. \tag{2.9}$$

Note that (2.5) and (2.6) ensure that the value of the maximization problem (2.7) is finite.

3 Existence and duality

3.1 Existence of L^1 optimizers

Notice that the existence of an optimizer in L^1 for (2.7) is not totally straighforward, and this is the object of the next result.

Proposition 3.1. *Assuming (2.1)-(2.2)-(2.3)-(2.4)-(2.5)-(2.6), the maximization problem (2.7) admits at least one solution.*

Proof. Let $\nu^n = \beta^n m$ be a maximizing sequence for (2.7). Since $\beta_i^n \geq 0$ and $\int_X \beta_i^n \mathrm{d}m = \mu_i(X)$, $(\beta^n)_n$ is bounded in $L^1(X, m)^N$. It therefore follows from Komlos' Theorem [8], that there exists a (not relabeled) subsequence such that the Cesaro means

$$\frac{1}{n} \sum_{k=1}^{n} \beta^k$$

converges m-a.e. to some $\beta = (\beta_1, \dots, \beta_N)$. Since both $\mathcal{U}$ and $-\mathcal{T}_\mu$ are concave, the above sequence of Cesaro means is also a maximizing sequence,

hence, slightly abusing notations by calling again β^n this new sequence we may assume that

$$\beta_i^n(x) \to \beta_i(x) \text{ for } m\text{-a.e. } x \in X, \text{ for } i = 1, \dots, N. \tag{3.1}$$

Let $\delta > 0$, the sublinearity assumption (2.6), together with (3.1), the semi-continuity of $U(x, .)$ and Fatou's Lemma yield

$$\begin{aligned}
&\liminf_n \int_X (\delta \sum_{i=1}^N \beta_i^n(x) - U(x, \beta^n(x)))\mathrm{d}m(x) \\
&= \delta \sum_{i=1}^N \mu_i(X) - \limsup_n \int_X U(x, \beta^n(x)))\mathrm{d}m(x) \\
&\geq \int_X (\delta \sum_{i=1}^N \beta_i(x) - U(x, \beta(x))\mathrm{d}m(x)
\end{aligned}$$

by letting $\delta \to 0^+$, we thus get

$$\limsup_n \int_X U(x, \beta^n(x))\mathrm{d}m(x) \leq \int_X U(x, \beta(x))\mathrm{d}m(x). \tag{3.2}$$

Moreover, $\nu^n = \beta^n m$ being bounded in $\mathcal{M}_+(X)^N$, we can assume (again passing to a subsequence if necessary) that it converges weakly-$*$ to some $\nu = (\nu_1, \dots, \nu_N) \in \mathcal{M}_+(X)^N$

$$\nu_i^n := \beta_i^n m \overset{*}{\rightharpoonup} \nu_i, \ i = 1, \dots, N. \tag{3.3}$$

Notice that $\nu_i(X) = \mu_i(X)$, and it is well-known that $T_{c_i}(\mu_i, .)$ is sequentially weakly-$*$ lsc so that

$$\liminf_n \mathcal{T}_\mu(\nu^n) \geq \mathcal{T}_\mu(\nu). \tag{3.4}$$

From (3.2)-(3.4) and the fact that ν^n is a maximizing sequence, we deduce:

$$\sup(2.7) = \int_X U(x, \beta(x))\mathrm{d}m(x) - \mathcal{T}_\mu(\nu). \tag{3.5}$$

Since β and ν may be different and β may violate the mass preservation constraints (2.9), the identity (3.5) is not enough to conclude that (2.7) has a solution. To obtain a solution of (2.7), we shall use some monotonicity arguments relying on (2.1) and (2.2).

Let $f \in C(X)$ with $f \geq 0$. For all $M > 0$, we have,

$$\int_X f(x)\mathrm{d}\nu_i^n(x) \geq \int_X f(x) \min(\beta_i^n(x), M)\mathrm{d}m(x).$$

Letting $n \to +\infty$, using (3.1), Lebesgue's dominated convergence Theorem and (3.3), we deduce

$$\int_X f(x)\mathrm{d}\nu_i(x) \geq \int_X f(x)\min(\beta_i(x), M)\mathrm{d}m(x),$$

so that, letting $M \to \infty$, by monotone convergence, we obtain

$$\nu_i \geq \beta_i m, \quad \text{for } i = 1, \ldots, N. \tag{3.6}$$

Thanks to the Radon-Nikodym Theorem, we can decompose each measure ν_i as

$$\nu_i = \beta_i^a m + \nu_i^s \text{ with } \beta_i^a \in L^1(X, m) \text{ and } \nu_i^s \perp m. \tag{3.7}$$

Since ν_i^s and m are mutually singular there exists a Borel subset of X, which we denote A_i, such that $m(X \setminus A_i) = \nu_i^s(A_i) = 0$ so that for every Borel subset B of X, there holds

$$(\beta_i^a m)(B) = \nu_i(B \cap A_i), \ \nu_i^s(B) = \nu_i(B \setminus A_i). \tag{3.8}$$

As a consequence, with (3.6), we also have

$$\beta_i \leq \beta_i^a, \quad m\text{-a.e. for } i = 1, \ldots, N. \tag{3.9}$$

Let now $\gamma_i \in \Pi(\mu_i, \nu_i)$ be such that $\mathcal{T}_{c_i}(\mu_i, \nu_i) = \int_{X\times X} c_i \mathrm{d}\gamma_i$ and let us decompose γ_i as $\gamma_i = \gamma_i^a + \gamma_i^s$ where γ_i^a and γ_i^s are defined by

$$\int_{X\times X} \varphi \mathrm{d}\gamma_i^a := \int_{X\times A_i} \varphi \mathrm{d}\gamma_i, \ \int_{X\times X} \varphi \mathrm{d}\gamma_i^s := \int_{X\times (X\setminus A_i)} \varphi \mathrm{d}\gamma_i,$$

for every $\varphi \in C(X \times X)$. By construction, we have

$$\mathrm{proj}_{2\#}\gamma_i^a = \beta_i^a m, \ \mathrm{proj}_{2\#}\gamma_i^s = \nu_i^s$$

and one can decompose $\mu_i = \alpha_i m = \mathrm{proj}_{1\#}\gamma_i$ as $\mathrm{proj}_{1\#}\gamma_i^a + \mathrm{proj}_{1\#}\gamma_i^s$. Since these two terms are absolutely continuous with respect to m, we can write them as

$$\alpha_i^a m := \mathrm{proj}_{1\#}\gamma_i^a, \ \alpha_i^s m := \mathrm{proj}_{1\#}\gamma_i^s \text{ (hence } \alpha_i = \alpha_i^a + \alpha_i^s, \ m\text{-a.e)}.$$

Now let us define

$$\widetilde{\gamma}_i := \gamma_i^a + (\mathrm{id}, \mathrm{id})_\# \alpha_i^s m \tag{3.10}$$

and observe that

$$\mathrm{proj}_{1\#}\widetilde{\gamma}_i = \alpha_i^a m + \alpha_i^s m = \alpha_i m = \mu_i$$

and

$$\mathrm{proj}_{2\#}\widetilde{\gamma}_i = \widetilde{\beta}_i m \text{ where } \widetilde{\beta}_i := \beta_i^a + \alpha_i^s$$

so that $\widetilde{\gamma}_i \in \Pi(\mu_i, \widetilde{\beta}_i m)$. Using (2.1) and the definition of $\widetilde{\gamma}_i$ in (3.10), we have

$$\begin{aligned} T_{c_i}(\mu_i, \widetilde{\beta}_i m) &\leq \int_{X\times X} c_i \mathrm{d}\widetilde{\gamma}_i = \int_{X\times X} c_i \mathrm{d}\gamma_i^a \\ &\leq \int_{X\times X} c_i \mathrm{d}\gamma_i = T_{c_i}(\mu_i, \nu_i). \end{aligned}$$

Setting $\widetilde{\beta} = (\widetilde{\beta}_1, \ldots, \widetilde{\beta}_N)$ we thus have

$$\mathcal{T}_\mu(\widetilde{\beta} m) \leq \mathcal{T}_\mu(\nu)$$

and since $\widetilde{\beta}_i \geq \beta_i^a$, from (3.9), we deduce that $\widetilde{\beta} \geq \beta$, hence from the monotonicity part in assumption (2.2), we get

$$\mathcal{U}(\widetilde{\beta} m) \geq \mathcal{U}(\beta m).$$

Together with (3.5), this implies that $\widetilde{\beta}$ solves (2.7).

□

Remark 3.2. The proof above shows that if $\nu \in \mathcal{M}_+(X)^N$ is such that $\nu_i(X) = \mu_i(X)$ for $i = 1, \ldots, N$, and writing the Lebesgue decomposition of ν_i with respect to m as

$$\nu_i = \nu_i^a + \nu_i^s = \beta_i^a m + \nu_i^s \text{ with } \nu_i^s \perp m,$$

there exists $\widetilde{\beta}_i \in L^1(X, m)$ such that

$$\widetilde{\beta}_i \geq \beta_i^a, \ \int_X \widetilde{\beta}_i^a \mathrm{d}m = \mu_i(X) \text{ and } T_{c_i}(\mu_i, \widetilde{\beta}_i m) \leq T_{c_i}(\mu_i, \nu_i).$$

This, together with the monotonicity of U, shows that

$$\mathcal{U}(\beta^a m) - \mathcal{T}_\mu(\nu) \leq \mathcal{U}(\widetilde{\beta} m) - \mathcal{T}_\mu(\widetilde{\beta} m)$$

therefore we have

$$\sup_{\beta \in L^1((X,m))^N} \{\mathcal{U}(\beta m) - \mathcal{T}_\mu(\beta m)\} = \sup_{\nu \in \mathcal{M}_+(X)^N} \{\mathcal{U}(\nu^a) - \mathcal{T}_\mu(\nu)\} \tag{3.11}$$

where in the right-hand side $\nu^a = (\nu_1^a, \ldots, \nu_N^a)$ denotes the absolutely continuous part of ν with respect to m.

3.2 Duality

The Pareto problem (2.7) appears naturally as the dual of a convex minimization problem over continuous functions which we now describe. Let us first introduce for $(x, \varphi) = (x, \varphi_1, \ldots, \varphi_N) \in X \times \mathbb{R}^N$:

$$V(x, \varphi) := \sup_{\beta \in \mathbb{R}^N_+} \{U(x, \beta) - \sum_{i=1}^N \beta_i \varphi_i\}. \tag{3.12}$$

Note that each $V(x, .)$ is convex, lsc, nonincreasing and that $\varphi \in \mathbb{R}^N_+$ whenever $V(x, \varphi) < +\infty$. Let us then consider the convex integral functional

$$\mathcal{V}(\varphi) := \int_X V(x, \varphi_1(x), \ldots, \varphi_N(x)) \mathrm{d}m(x), \ \forall \varphi \in C(X)^N,$$

and observe that, thanks to (2.6), one has

$$\min_{i=1,\ldots,N} \min_{x \in X} \varphi_i(x) = \delta > 0 \Rightarrow \mathcal{V}(\varphi) \leq C_\delta m(X) < +\infty. \tag{3.13}$$

Recall that, for any $i = 1, \ldots, N$ and given $\psi \in C(X)$, the c_i-transform of ψ, denoted by ψ^{c_i}, is defined by

$$\psi^{c_i}(x) := \min_{y \in X} \{c_i(x, y) - \psi(y)\}. \tag{3.14}$$

For $\varphi := (\varphi_1, \ldots, \varphi_N) \in C(X)^N$, let us set

$$K(\varphi) := -\sum_{i=1}^N \int_X \varphi_i^{c_i} \mathrm{d}\mu_i = \sum_{i=1}^N \int_X \max_{y \in X} \{\varphi_i(y) - c_i(x, y)\} \mathrm{d}\mu_i(x)$$

so that K is convex and everywhere continuous w.r.t the uniform norm. Finally, let us consider the convex minimization problem

$$\inf_{\varphi \in C(X, \mathbb{R})^N} \{K(\varphi) + \mathcal{V}(\varphi)\}. \tag{3.15}$$

It is easy to deduce, from (2.5) and the fact that c_i is bounded, that $K + \mathcal{V}$ is bounded from below. Hence, (3.13) together with the continuity of K implies that the value of (3.15) is finite and that the Fenchel-Rockafellar Theorem applies to (3.15). Identifying the dual of $C(X)^N$ with $\mathcal{M}(X)^N$, the Fenchel-Rockafellar dual of (3.15) reads

$$\sup_{\nu \in \mathcal{M}(X)^N} \{-\mathcal{V}^*(-\nu) - K^*(\nu)\}, \tag{3.16}$$

where K^* and $\mathcal{V}^*$ respectively stand for the Legendre transform of K and $\mathcal{V}$ (for the duality between continuous functions and Radon measures), i.e. for every $\nu \in \mathcal{M}(X)^N$:

$$K^*(\nu) := \sup_{\varphi \in C(X)^N} \left\{ \sum_{i=1}^N \int_X \varphi_i \mathrm{d}\nu_i - K(\varphi) \right\},$$
$$\mathcal{V}^*(-\nu) = \sup_{\varphi \in C(X)^N} \left\{ -\sum_{i=1}^N \int_X \varphi_i \mathrm{d}\nu_i - \mathcal{V}(\varphi) \right\}.$$

It turns out that (3.16) is nothing but (2.7), slightly in disguise, as expressed by the next result.

Lemma 3.3. *Let $\nu \in \mathcal{M}(X)^N$, there holds*

$$K^*(\nu) = \mathcal{T}_\mu(\nu),\ \mathcal{V}^*(-\nu) = \begin{cases} -\mathcal{U}(\nu^a) \text{ if } \nu \in \mathcal{M}_+(X)^N, \\ +\infty \text{ otherwise} \end{cases} \tag{3.17}$$

where ν^a is the absolutely continuous part of ν with respect to m. Moreover, we have

$$\inf(3.15) = \max(2.7). \tag{3.18}$$

Proof. The fact that $K^*(\nu) = \mathcal{T}_\mu(\nu)$ is a consequence of the well-known Kantorovich duality formula for optimal transport (see for instance [13]). As for the Legendre transform of $\mathcal{V}$, it follows from the seminal results of Rockafellar [11]-[12] for convex integral functionals. More precisely, thanks to Theorem 4 in [12] and our assumptions on U, we have:

$$\mathcal{V}^*(-\nu) = \inf_{\theta \in L^1(X,m)^N} \Big\{ \int_X V^*(x, -\theta(x)) \mathrm{d}m(x) + \sup_{\varphi \in C(X)^N : \mathcal{V}(\varphi) < +\infty} \sum_{i=1}^N \int_X \varphi_i \mathrm{d}(\theta_i - \nu_i) \Big\}.$$

On the one hand, by the Fenchel-Moreau Theorem and the convexity and lower semicontinuity of $-U(x,.)$ we have

$$V^*(x, -\theta) = \sup_{\varphi \in \mathbb{R}^N} \{-\theta \cdot \varphi - V(x, \varphi)\} = \begin{cases} -U(x, \theta) \text{ if } \theta \in \mathbb{R}^N_+, \\ +\infty \text{ otherwise.} \end{cases}$$

On the other hand, since $\mathcal{V}(\varphi) < +\infty$ whenever each φ_i is strictly positive and $\mathcal{V}(\varphi) < +\infty$ implies that each φ_i is nonnegative, we have

$$\sup_{\varphi \in C(X)^N : \mathcal{V}(\varphi) < +\infty} \sum_{i=1}^N \int_X \varphi_i \mathrm{d}(\theta_i - \nu_i) = \begin{cases} 0 \text{ if } \theta_i \leq \nu_i, \text{ for } i = 1, \dots, N \\ +\infty \text{ otherwise.} \end{cases}$$

As a consequence, using the monotonicity of $U(x,.)$ and the fact that for $\theta_i \in L^1(X,m)$, $0 \le \theta_i \le \nu_i$ is equivalent to $0 \le \theta_i \le \nu_i^a$, we get

$$\begin{aligned}\mathcal{V}^*(-\nu) &= \inf\left\{-\int_X U(x,\theta(x))\mathrm{d}m(x) \ : \ \theta \in L^1(X,m),\ 0 \le \theta_i \le \nu_i,\ i=1,\dots,N\right\} \\ &= \begin{cases} -\mathcal{U}(\nu^a) \text{ if } \nu \in \mathcal{M}_+(X)^N \\ +\infty \text{ otherwise.}\end{cases}\end{aligned}$$

By the Fenchel-Rockafellar Theorem (see [5]), together with the continuity of K and the fact that $\mathcal{V}$ is lsc and $\mathcal{V}(\varphi)$ is finite whenever the components of φ are bounded from below by a positive constant, we get

$$\inf(3.15) = \max_{\nu \in \mathcal{M}_+(X)^N} \{\mathcal{U}(\nu^a) - \mathcal{T}_\mu(\nu)\}$$

from which we deduce (3.18) thanks to proposition 3.1 and (3.11).

□

3.3 Existence for the primal

Let us consider the following qualification condition which is a strengthening of (2.4)-(2.5): there exists $\overline{\alpha} \in L^1(X,m)^N$ such that (2.4) holds and there exists $\varepsilon \in (0,1)$ such that

$$\int_X U(x,(1-\varepsilon)\overline{\alpha}_1(x),\dots,(1-\varepsilon)\overline{\alpha}_N(x))\mathrm{d}m(x) > -\infty. \tag{3.19}$$

We then have an existence result for (3.15):

Proposition 3.4. *Assuming (2.1)-(2.2)-(2.3)-(2.4)-(3.19)-(2.6), the minimization problem (3.15) admits at least one solution.*

Proof. Let $\varphi^n = (\varphi_1^n,\dots,\varphi_N^n)$ be a minimizing sequence for (3.15), let us then define $\widetilde{\varphi}^n := (\widetilde{\varphi}_1^n,\dots,\widetilde{\varphi}_N^n)$ by

$$\widetilde{\varphi}_i^n(y) := \min_{x\in X}\{c_i(x,y) - (\varphi_i^n)^{c_i}(x)\},\ \forall y \in X. \tag{3.20}$$

It is well-known that $(\widetilde{\varphi}_i^n)^{c_i} = (\varphi_i^n)^{c_i}$ and that $\widetilde{\varphi}_i^n \ge \varphi_i^n$ so that $K(\widetilde{\varphi}^n) = K(\varphi^n)$ and (since $V(x,.)$ is nonincreasing) $\mathcal{V}(\widetilde{\varphi}^n) \le \mathcal{V}(\varphi^n)$. Hence $\widetilde{\varphi}^n$ is also a minimizing sequence for (3.15). Now the advantage of working with the minimizing sequence $\widetilde{\varphi}_i^n$ is that it is uniformly equicontinuous (indeed, thanks to (3.20), a modulus of continuity of c_i is also a modulus of modulus of $\widetilde{\varphi}_i^n$

for every n). Since $\mathcal{V}(\widetilde{\varphi}^n)$ is finite, $\widetilde{\varphi}_i^n \geq 0$, let us then set $\delta_i^n := \min_X \widetilde{\varphi}_i^n \geq 0$, and let us prove that δ_i^n is bounded. First, from the equicontinuity, there is a constant $C \geq 0$ such that for every $i \in \{1, \ldots, N\}$ and every n, one has

$$\widetilde{\varphi}_i^n \leq \delta_i^n + C. \tag{3.21}$$

Now fix, M such that

$$M \geq K(\widetilde{\varphi}_n) + \mathcal{V}(\widetilde{\varphi}_n), \ \forall n. \tag{3.22}$$

Since $\widetilde{\varphi}_i^n \geq \delta_i^n$, thanks to (2.1) we have $K(\widetilde{\varphi}_n) \geq \sum_{i=1}^N (\delta_i^n - \min_{X\times X} c_i)\mu_i(X) = \sum_{i=1}^N \delta_i^n \mu_i(X)$. Moreover, by the very definition of V and (3.21)

$$\begin{aligned} V(x, \widetilde{\varphi}^n(x)) &\geq U(x, (1-\varepsilon)\overline{\alpha}(x)) - (1-\varepsilon)\sum_{i=1}^N \overline{\alpha}_i(x)\widetilde{\varphi}_i^n(x) \\ &\geq U(x, (1-\varepsilon)\overline{\alpha}(x)) - (1-\varepsilon)\sum_{i=1}^N \overline{\alpha}_i(x)(\delta_i^n + C). \end{aligned}$$

Integrating and using (2.4)-(3.19), this yields that for some constant C', $\mathcal{V}(\widetilde{\varphi}^n) \geq C' - (1-\varepsilon)\sum_{i=1}^N \delta_i^n \mu_i(X)$. Thanks to (3.22), we therefore get

$$M \geq C' + \varepsilon \sum_{i=1}^N \delta_i^n \mu_i(X)$$

which shows that δ_i^n is bounded so that $\widetilde{\varphi}^n$ is uniformly bounded. Applying Ascoli's Theorem, we may therefore assume that $\widetilde{\varphi}^n$ converges uniformly to some $\widetilde{\varphi}$, so that $K(\widetilde{\varphi}^n)$ converges to $K(\widetilde{\varphi})$, and since for some constant C'', one has $V(x, \widetilde{\varphi}_n(x)) \geq U(x, \overline{\alpha}(x)) - C''$, one deduces from (2.5) and Fatou's lemma that $\liminf_n \mathcal{V}(\widetilde{\varphi}^n) \geq \mathcal{V}(\widetilde{\varphi})$ which finally proves that $\widetilde{\varphi}$ solves (3.15). □

4 Optimality conditions and a second welfare Theorem

4.1 Primal-dual extremality conditions

Now that we know that, under the assumptions of proposition 3.4, there exist solutions both to (2.7) and its primal (3.15), we can easily use (3.18)

to deduce a characterization of these solutions. Indeed, by construction, we have

$$V(x,\varphi) \geq U(x,\beta) - \varphi\cdot\beta = U(x,\beta) - \sum_{i=1}^{N} \varphi_i\beta_i,\ \forall(\varphi,\beta)\in\mathbb{R}_+^N\times\mathbb{R}_+^N \quad (4.1)$$

and for every i, $\beta := (\beta_1,\ldots,\beta_N) \in L^1(X,m)^N$ such that $\beta_i \geq 0$ and $\int_X \beta_i m = \mu_i(X)$ and every $\varphi := (\varphi_1,\ldots,\varphi_N) \in C(X)^N$, we have

$$T_{c_i}(\mu_i,\beta_i m) \geq \int_X \varphi_i^{c_i}\alpha_i m + \int_X \varphi_i\beta_i m. \quad (4.2)$$

Now if φ solves (3.15) and β solves (2.7), from (3.18), we obtain

$$\mathcal{V}(\varphi) - \mathcal{U}(\beta m) + \int_X \varphi\cdot\beta m + \mathcal{T}_\mu(\beta m) - \sum_{i=1}^{N}\int_X \varphi_i^{c_i}\alpha_i m - \int_X \varphi\cdot\beta m = 0$$

so that with (4.1), we get

$$V(x,\varphi(x)) = U(x,\beta(x)) - \varphi(x)\cdot\beta(x) \text{ for } m\text{-a.e. } x. \quad (4.3)$$

Likewise, with (4.2), we get for every i,

$$T_{c_i}(\mu_i,\beta_i m) = \int_X \varphi_i^{c_i}\alpha_i m + \int_X \varphi_i\beta_i m$$

so that

$$\exists\gamma_i \in \Pi(\alpha_i m,\beta_i m) \text{ such that } \varphi_i^{c_i}(x) + \varphi_i(y) = c_i(x,y) \text{ for } \gamma_i\text{-a.e. } (x,y). \quad (4.4)$$

By continuity of φ_i and c_i, this means that, whenever (x,y) is in the support of γ_i, one has

$$\varphi_i(y) - c_i(x,y) \geq \varphi_i(z) - c_i(x,z),\ \forall z\in X. \quad (4.5)$$

Note in particular that (4.3) implies that for m-a.e. x, $\varphi(x)$ is a supergradient of $U(x,.)$ at $\beta(x)$ and then

$$\forall\beta\in\mathbb{R}_+^N,\ \varphi(x)\cdot\beta \leq \varphi(x)\cdot\beta(x) \Rightarrow U(x,\beta) \leq U(x,\beta(x)). \quad (4.6)$$

4.2 A second welfare theorem with transport costs

The optimality conditions (4.4)-(4.6) can easily be interpreted in terms of equilibrium conditions as we shall now explain. Imagine that there is a representative agent at each location x, with utility $U(x,.)$ and an initial endowment $\alpha(x)$ for the goods $1,\dots,N$. Assume also that this agent has a given (monetary) endowment $w(x)$. In an exchange economy, where the agent located at y can trade with agents located at any location x, an equilibrium for the monetary endowment w is a system of prices $y \mapsto p(y)$ together with a final endowment $y \mapsto \beta(y)$ of the goods, satisfying:

- **There is free-mobility of trade**: Given the price, system p, and a location x, agents located at x exporting one unit of good i to location y, get a profit $p_i(y) - c_i(x,y)$, so that their total profit is

$$\pi(x)\cdot\alpha(x) = \sum_{i=1}^{N} \pi_i(x)\alpha_i(x),\ \pi_i(x) = \max_{y\in Y}\{p_i(y) - c_i(x,y)\} = -p_i^{c_i}(x),$$

- **Consumers maximize utility**: Agents located at y maximize their utility $U(y,\beta)$ subject to their budget constraint that expenditure $p(y)\cdot\beta$ is smaller than their total revenue which is $w(y)$ augmented by their export profit $\pi(y)\cdot\alpha(y)$, i.e. $\beta(y)$ should solve

$$\max\{U(y,\beta)\ :\ p(y)\cdot\beta \le \pi(y)\cdot\alpha(y) + w(y)\}$$

- **Markets clear**: which means that for every i, one can find a plan γ_i ($\gamma_i(A\times B)$ represents the quantity of good i transported from a location in A to a location in B), such that $\gamma_i \in \Pi(\alpha_i m, \beta_i m)$ (i.e. demand and supply match) and $\pi_i(x) = p_i(y) - c_i(x,y)$ for every (x,y) in the support of γ_i (which means that this plan is consistent with free-mobility of trade).

We then straighforwardly deduce from the optimality conditions (4.4)-(4.6), a sort of second welfare theorem with transport costs

Theorem 4.1. *Let φ and β solves (3.15) and (2.7) respectively. Define then*

$$p_i := \varphi_i, \pi_i := -p_i^{c_i},\ w := p\cdot\beta - \pi\cdot\alpha$$

then (β, p) is an equilibrium with monetary endowment w.

5 Algorithm and simulations

In this final section, we describe the entropic approximation of (3.15) and its dual (2.7). Since the influential paper of Cuturi [4], entropic optimal transport has become a popular and efficient tool for computational optimal transport thanks to Sinkhorn's celebrated scaling algorithm, and we refer to [10] for a comprehensive exposition. The algorithm detailed in the next paragraph is based on a variant of the Sinkhorn algorithm introduced by Peyré [9] in the context of Wasserstein gradient flows. We finally give numerical results in the one-dimensional case and in the case of a network of cities.

5.1 An entropic approximation algorithm

From now on, we assume that X is finite, take the counting measure m as reference measure on X, and denote by $\alpha_i(x) > 0$ the inital endowment of location $x \in X$ in the good $i \in \{1, \dots, N\}$. Problem (3.15) thus takes the form:

$$\inf_{\varphi \in \mathbb{R}^{X \times N}} \sum_{y \in X} V(y, \varphi(y)) + \sum_{i=1}^{N} \sum_{x \in X} \alpha_i(x) \max_{y \in X} \{\varphi_i(y) - c_i(x, y)\}. \tag{5.1}$$

Given a regularization parameter $\varepsilon > 0$, we consider the smooth approximation of (5.1) where maxima are replaced by soft maxima:

$$\inf_{\varphi \in \mathbb{R}^{X \times N}} \sum_{y \in X} V(y, \varphi(y)) + \varepsilon \sum_{i=1}^{N} \sum_{x \in X} \alpha_i(x) \log \Big(\sum_{y \in X} e^{\frac{\varphi_i(y) - c_i(x,y)}{\varepsilon}} \Big). \tag{5.2}$$

Now, let us observe that (5.2) can be conveniently reformulated by considering the convex function

$$\Phi(\varphi, \psi) := \sum_{y \in X} V(y, \varphi(y)) - \sum_{i=1}^{N} \sum_{x \in X} \alpha_i(x) \psi_i(x) + \varepsilon \sum_{i=1}^{N} \sum_{(x,y) \in X^2} e^{\frac{\psi_i(x) + \varphi_i(y) - c_i(x,y)}{\varepsilon}}$$

for $(\psi, \varphi) \in \mathbb{R}^{X \times N} \times \mathbb{R}^{X \times N}$. Indeed, for fixed φ, the minimizer of $\psi \mapsto \Phi(\varphi, \psi)$ is explicitly given by

$$\psi_i(x) = \varepsilon \log(\alpha_i(x)) - \varepsilon \log \Big(\sum_{y \in X} e^{\frac{\varphi_i(y) - c_i(x,y)}{\varepsilon}} \Big) \tag{5.3}$$

so replacing in Φ, we get

$$\inf_{\psi} \Phi(\varphi, \psi) = C + \sum_{y \in X} V(y, \varphi(y)) + \varepsilon \sum_{i=1}^{N} \sum_{x \in X} \alpha_i(x) \log \Big(\sum_{y \in X} e^{\frac{\varphi_i(y) - c_i(x,y)}{\varepsilon}} \Big)$$

where C is the constant

$$C := \varepsilon \sum_{i=1}^{N} \sum_{x \in X} \alpha_i(x)(1 - \log(\alpha_i(x))).$$

Optimality system in primal and dual form. The optimality condition for (5.2) (which is the minimization of the sum of a convex lsc function with a smooth convex function) writes:

$$-\beta(y) \in \partial V(y, \varphi(y)), \ \forall y \in X \tag{5.4}$$

where $\beta(y) = (\beta_1(y), \ldots, \beta_N(y))$ is given by

$$\beta_i(y) = \sum_{x \in X} \alpha_i(x) \frac{e^{\frac{\varphi_i(y) - c_i(x,y)}{\varepsilon}}}{\sum_{z \in X} e^{\frac{\varphi_i(z) - c_i(x,z)}{\varepsilon}}}. \tag{5.5}$$

Defining

$$\gamma_i(x, y) := \alpha_i(x) \frac{e^{\frac{\varphi_i(y) - c_i(x,y)}{\varepsilon}}}{\sum_{z \in X} e^{\frac{\varphi_i(z) - c_i(x,z)}{\varepsilon}}}$$

we thus have $\gamma_i \in \Pi(\alpha_i, \beta_i)$ and γ_i solves the entropic optimal transport problem

$$T^{\varepsilon}_{c_i}(\alpha_i, \beta_i) := \inf_{\gamma \in \Pi(\alpha_i, \beta_i)} \sum_{(x,y) \in X^2} (c_i(x, y) + \varepsilon \log(\gamma(x, y))\gamma(x, y).$$

It is then easy to check that (5.4)-(5.5) is equivalent to the fact that β solves the dual of (5.2):

$$\sup_{\beta \in \mathbb{R}_+^{X \times N}} \sum_{y \in X} U(y, \beta(y)) - \sum_{i=1}^{N} T^{\varepsilon}_{c_i}(\alpha_i, \beta_i), \tag{5.6}$$

which is the entropic regularization of (2.7) with regularization parameter ε.

Coordinate descent/Sinkhorn. We can find the (regularized) prices φ by solving (5.2) and then the optimal quantities β by using (5.5). As noted above, (5.2) is equivalent to

$$\inf_{(\psi, \varphi) \in \mathbb{R}^{X \times N} \times \mathbb{R}^{X \times N}} \Phi(\varphi, \psi) \tag{5.7}$$

which can be solved by coordinate descent as follows. Starting from (ψ^0, φ^0), we recursively compute $(\psi^{k+1}, \varphi^{k+1})$ by:

$$\psi^{k+1} = \operatorname{argmin}_{\psi \in \mathbb{R}^{X \times N}} \Phi(\varphi^k, \psi)$$

which, using (5.3), gives totally explicit Sinkhorn-like updates:

$$\psi_i^{k+1}(x) = \varepsilon \log(\alpha_i(x)) - \varepsilon \log\Big(\sum_{y\in X} e^{\frac{\varphi_i^k(y)-c_i(x,y)}{\varepsilon}}\Big) \tag{5.8}$$

for $i = 1, \dots, N$ and every $x \in X$. Then, we update the prices by:

$$\varphi_1^{k+1} = \mathrm{argmin}_{\varphi_1\in\mathbb{R}^N}\, \Phi(\varphi_1, \varphi_2^k, \dots \varphi_N^k, \psi^{k+1})$$

which is the same as solving a one-dimensional strictly convex minimization problem for each $y \in X$:

$$\varphi_1^{k+1}(y) = \mathrm{argmin}_t\, V(y, t, \varphi_2^k(y), \dots, \varphi_N^k(y)) + \varepsilon e^{\frac{t}{\varepsilon}} \sum_{x\in X} e^{\frac{\psi_1^{k+1}(x)-c_1(x,y)}{\varepsilon}}. \tag{5.9}$$

This boils down to a strictly monotone equation (if V is differentiable, see the Cobb-Douglas case below) or inclusion (if V is nonsmooth) in t which can be solved efficiently by dichotomy for instance. The next price updates are performed similarly by iteratively solving for, $i = 2, \dots, N-1$ and every $y \in X$:

$$\begin{aligned}\varphi_i^{k+1}(y) = \mathrm{argmin}_t \Big\{ & V(y, \varphi_1^{k+1}(y), \dots, \varphi_{i-1}^{k+1}(y), t, \varphi_{i+1}^k(y), \dots, \varphi_N^k(y)) \\ & + \varepsilon e^{\frac{t}{\varepsilon}} \sum_{x\in X} e^{\frac{\psi_i^{k+1}(x)-c_i(x,y)}{\varepsilon}} \Big\}\end{aligned} \tag{5.10}$$

and for $i = N$

$$\begin{aligned}\varphi_N^{k+1}(y) = \mathrm{argmin}_t \Big\{ & V(y, \varphi_1^{k+1}(y), \dots, \varphi_{N-1}^{k+1}(y), t) \\ & + \varepsilon e^{\frac{t}{\varepsilon}} \sum_{x\in X} e^{\frac{\psi_N^{k+1}(x)-c_N(x,y)}{\varepsilon}} \Big\}.\end{aligned} \tag{5.11}$$

The Cobb-Douglas case. If we further specify the utility to be of Cobb-Douglas form[2] :

$$U(\beta) = \prod_{i=1}^N \beta_i^{a_i},\ \forall \beta \in \mathbb{R}_+^N,\ a_i > 0, a := \sum_{i=1}^N a_i < 1,$$

[2]for simplicity, we take U independent of the location but there is no extra difficulty in having a dependence in y in a multiplicative prefactor or even in the exponents a_i.

then a direct computation gives

$$V(\varphi) = (1-a)\frac{\prod_{i=1}^{N} a_i^{\frac{a_i}{1-a}}}{\prod_{i=1}^{N} \varphi_i^{\frac{a_i}{1-a}}}. \tag{5.12}$$

Hence, each minimization step as (5.10) amounts, for some parameter $A > 0$ and exponent b, to finding the root t of the strictly monotone equation

$$e^t t^b = A$$

which can be solved efficiently with a few steps of a dichotomy method. We can also notice that since V is smooth and locally strongly convex on its domain, the linear convergence of the coordinate scheme described above is guaranteed by the results of Beck and Tetruashvili [2].

5.2 Numerical results

We now present some simulations using the entropic approximation algorithm described in the previous paragraph.

5.2.1 One-dimensional experiments

We first consider the one dimensional case $X = [0, 1]$ (suitably discretized), assume there are twoo goods and that the utility is a Cobb-Douglas function with parameters a_1 and a_2. As for the transport cost, we take a power of the distance:

$$C_{\lambda,p}(x, y) = \lambda \frac{|x-y|^p}{p}, \forall (x, y) \in [0, 1]^2.$$

When λ is large, transport is costly and β_1, β_2 tend to be close to the initial distributions α_1 and α_2. On the contrary, when λ is small, the utility term dominates and β_1 and β_2 tend to be constant.

In our first three simulations, α_1 and α_2 are gaussians of expected values respectively $\mu_1 = 0.25$ and $\mu_2 = 0.75$ and standard deviations $\sigma_1 = \sigma_2 = 0.2$, $(a_1, a_2) = (0.49, 0.49)$ and $\varepsilon = 0.01$. The three simulations correspond to different exponents p for the cost: a concave cost with $p = 0.3$ (Figure 1), the linear cost $p = 1$ (Figure 2) and the quadratic cost $p = 2$ (Figure 3). In the first row, α appears in light grey and β in dark grey, the prices (φ_1, φ_2) are plotted in the second row. The third row represents the convergence of the algorithm: we plot the decay of the functional Φ and the evolution of the error $||\partial V(\varphi(\cdot)) + \beta(\cdot)||_2$ for the first order necessary and sufficient condition (5.4).

The most striking phenomenon shown by these simulations is the emergence of extra modes in the final distributions whereas the initial distributions are single peaked. This phenomenon seems to be related to the concavity of the transport cost. In the concave case of Figure 1, β exhibits three local maxima, the linear case of Figure 2 exhibits two modes and in the quadratic case of case of Figure 3, initial and final distributions roughly have the same unimodal shape.

In order to further explore numerically the impact of the various parameters on the structure of the Pareto optimum, we present in Figure 4 and Figure 5, the case of mixture of three gaussians for α_1 and of two gaussians for α_2 for different values of a_1 and a_2 (in Figure 4, $a_1 + a_2$ is close to 1 and in Figure 5, $a_1 + a_2 = 0.5$).

5.2.2 A network model

Our next application concerns a connected network of French cities. The nodes of this network (see Figure 9) are the nine French cities: Brest, Lille, Lyon, Marseille, Nantes, Paris, Rennes, St-Malo and Strasbourg. The cost between two neighbouring cities is given by the distance as the crow flies [3] and more generally the cost between two cities is given by the shortest path distance between them. To take into account population differences between these cities, we have weighted the Cobb-Douglas utility by a weight $w(x)$ for city x:

City	Brest	Lyon	Marseille	Paris	Strasbourg	Rennes	Lille	Saint-Malo	Nantes
w (thousands)	139	518	868	2,175	285	217	233	46	314

This leads to the weighted utility

$$U(x, \beta_1, \beta_2, \beta_3) = w(x)\beta_1^{a_1}(x)\beta_2^{a_2}(x)\beta_3^{a_3}(x),$$

and the corresponding dual term (see (3.12)):

$$V(x, \varphi) = w(x)^{\frac{1}{1-a}} \frac{\prod_{i=1}^{3} a_i^{\frac{a_i}{1-a}}}{\prod_{i=1}^{3} \varphi_i^{\frac{a_i}{1-a}}}$$

with $a = a_1 + a_2 + a_3 < 1$. In our examples, we have taken $(a_1, a_2, a_3) = (0.4, 0.1, 0.2)$. In Figures 6 to 9, we have represented the initial and final

[3] given by the website https://www.coordonnees-gps.fr/distance.

distributions as well as the optimal transport plans between the cities. Note that the larger ε, the stronger the diffusion effect in the transport plan, which explains why we see more connections in Figure 6 and 7 (large diffusion with $\varepsilon = 0.1$) than in Figure 8 and 9 ($\varepsilon = 0.01$). We also give the prices of the goods in the different cities and analyze the convergence of the algorithm both in terms of the decay of the objective function Φ and in the residuals for the optimality condition.

Acknowledgments: GC is grateful to the Agence Nationale de la Recherche for its support through the projects MAGA (ANR-16-CE40-0014) and MFG (ANR-16-CE40-0015-01). GC also acknowledges the support of the Lagrange Mathematics and Computation Research Center.

References

[1] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] Amir Beck and Luba Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM J. Optim.*, 23(4):2037–2060, 2013.

[3] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):A1111–A1138, 2015.

[4] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, pages 2292–2300, 2013.

[5] Ivar Ekeland and Roger Témam. *Convex analysis and variational problems*, volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, english edition, 1999.

[6] Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, Princeton, NJ, 2016.

[7] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[8] J. Komlós. A generalization of a problem of Steinhaus. *Acta Math. Acad. Sci. Hungar.*, 18:217–229, 1967.

[9] Gabriel Peyré. Entropic approximation of wasserstein gradient flows. *SIAM Journal on Imaging Sciences*, 8(4):2323–2351, 2015.

[10] Gabriel Peyré and Marco Cuturi. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[11] R. T. Rockafellar. Integrals which are convex functionals. *Pacific J. Math.*, 24:525–539, 1968.

[12] R. T. Rockafellar. Integrals which are convex functionals. II. *Pacific J. Math.*, 39:439–469, 1971.

[13] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015. Calculus of variations, PDEs, and modeling.

[14] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

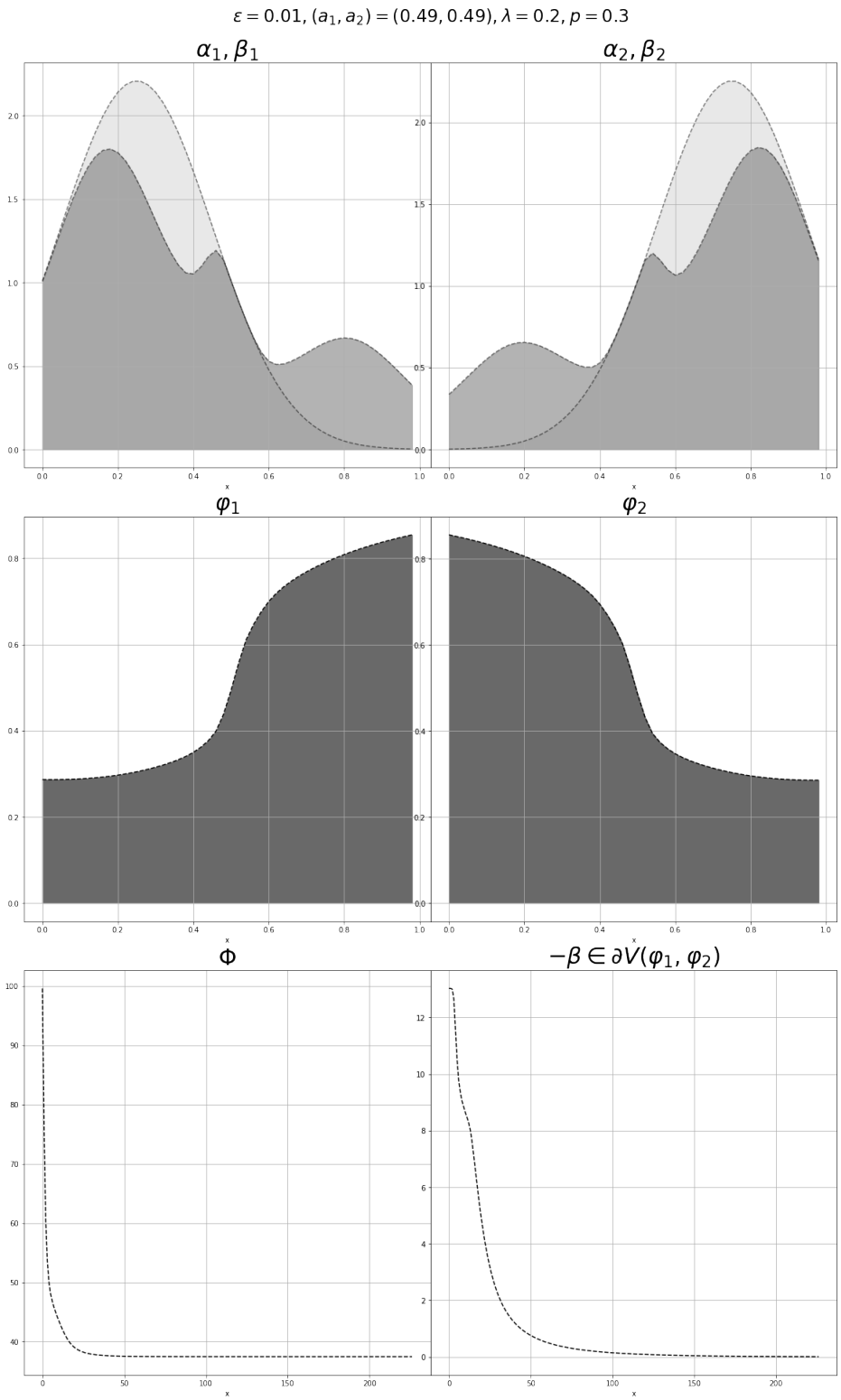


Figure 1: Gaussian distributions with concave cost

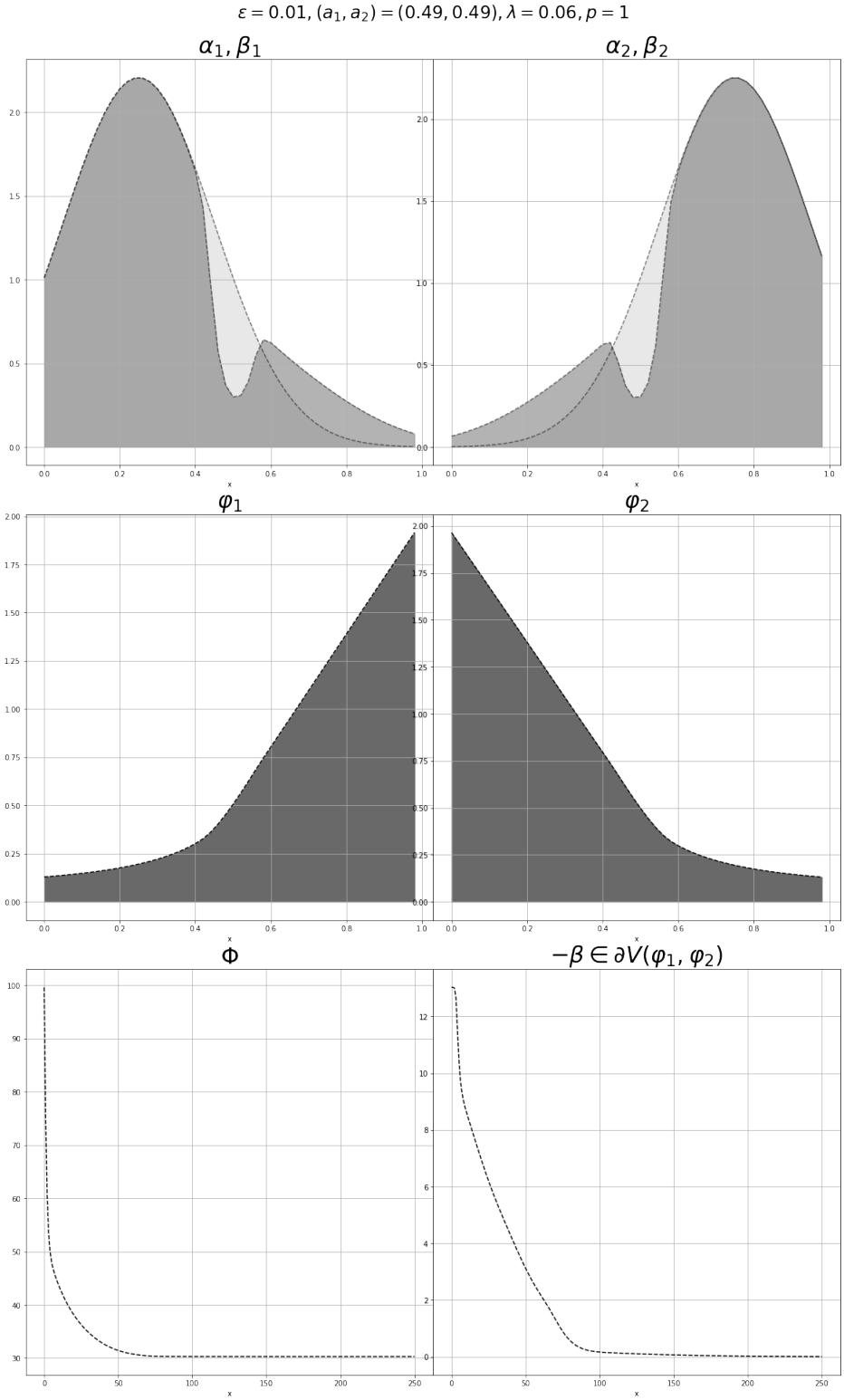


Figure 2: Gaussian distributions with homogeneous cost

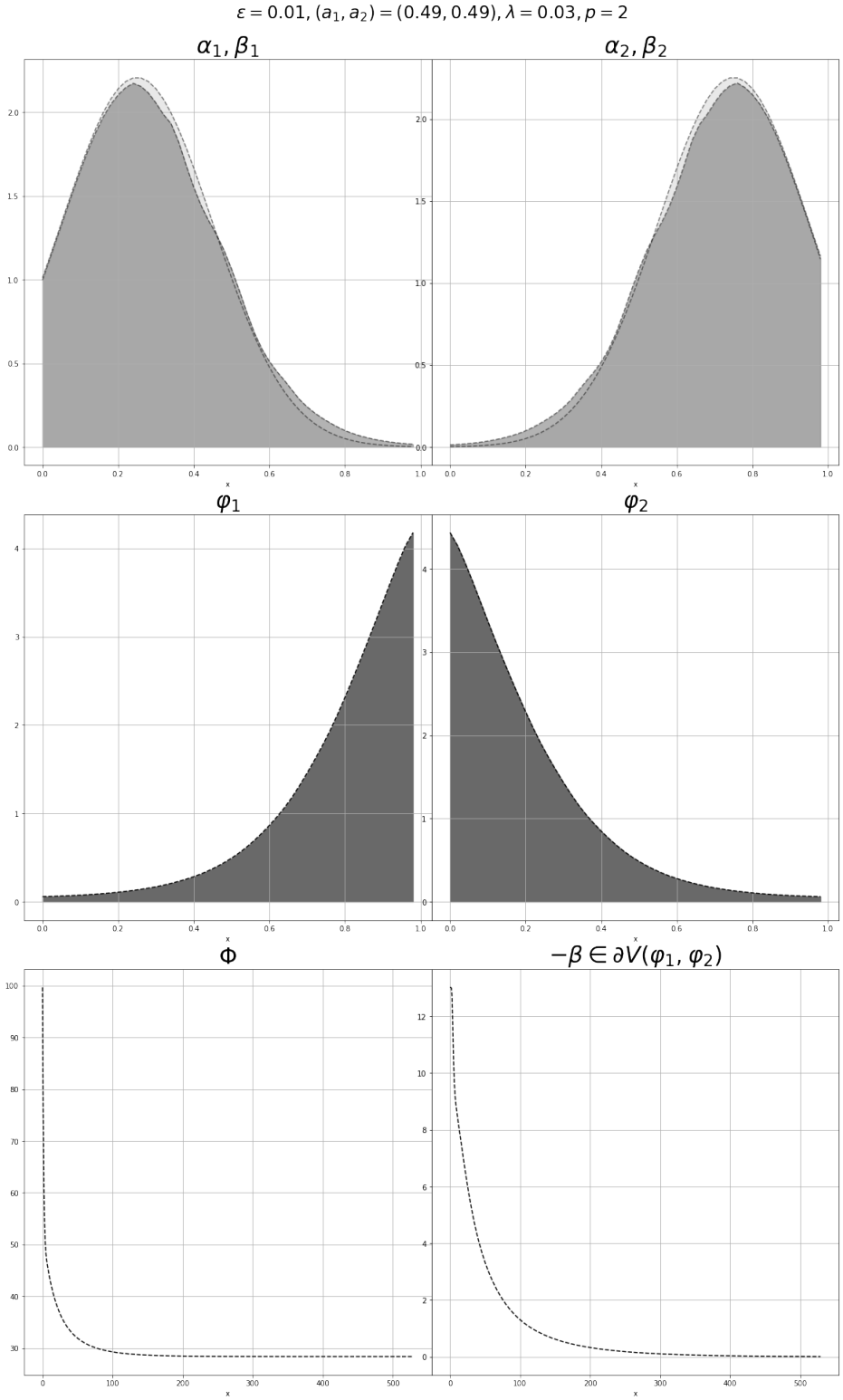


Figure 3: Gaussian distributions with convex cost

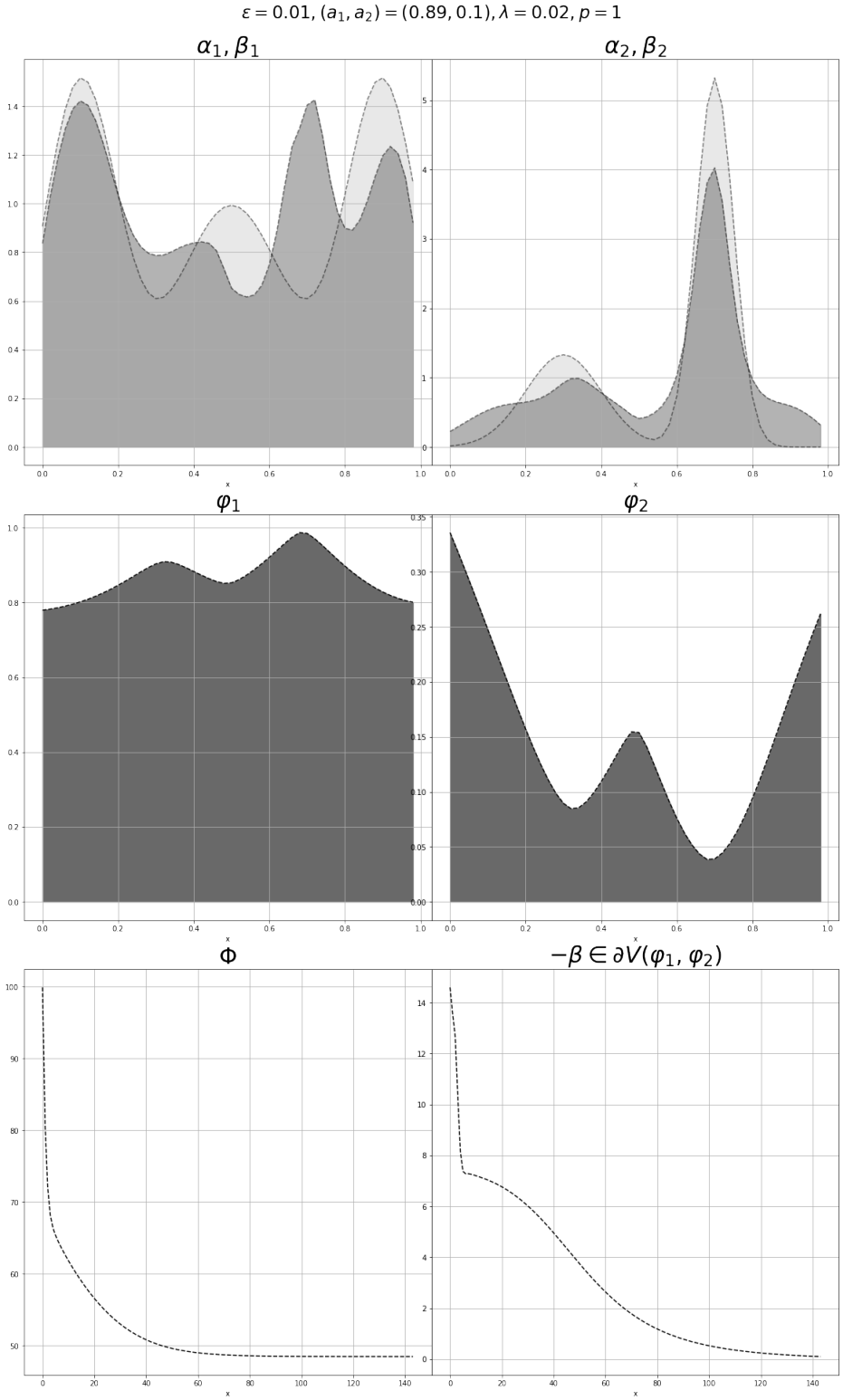


Figure 4: Another example with $a_1 + a_2$ close to 1

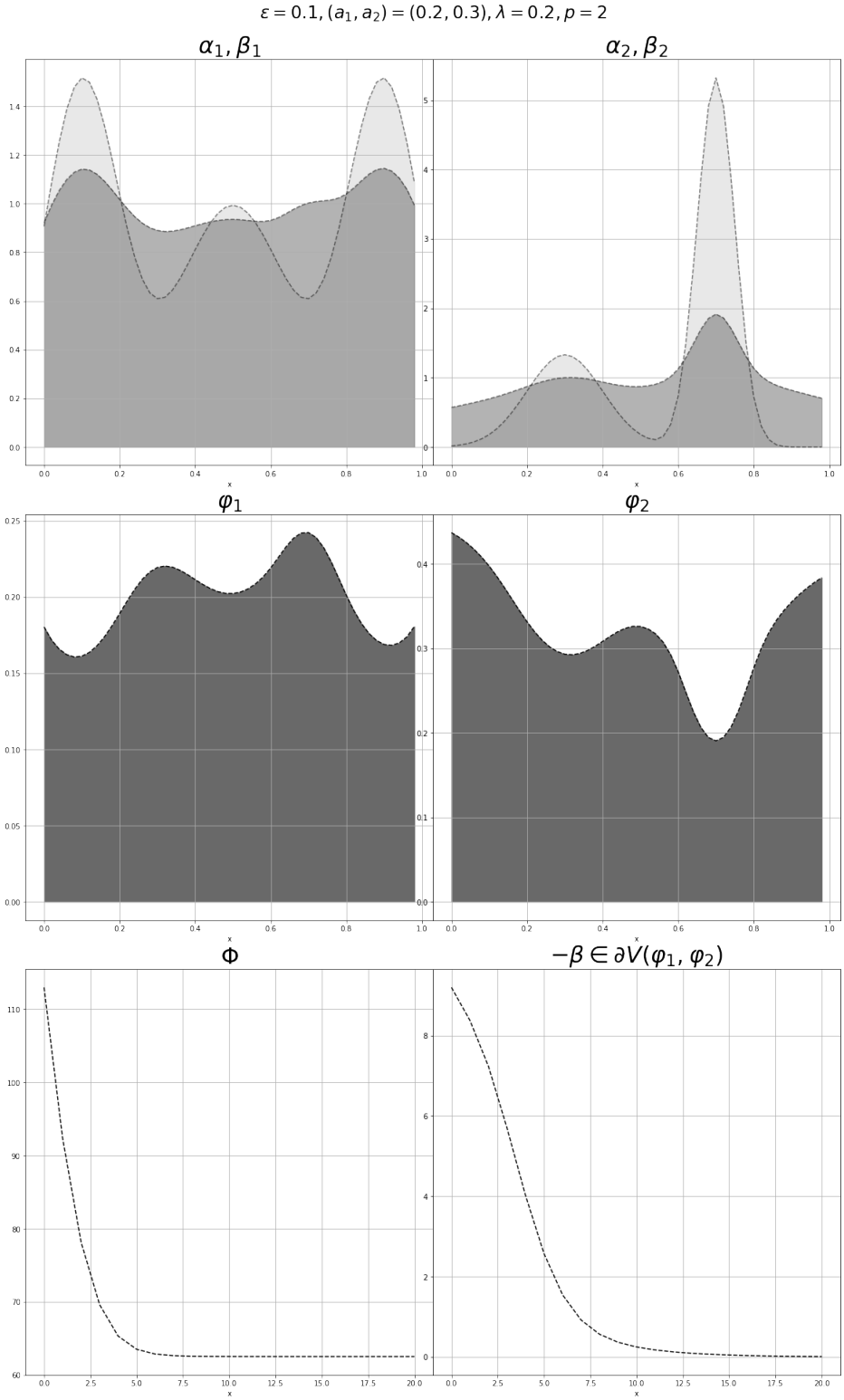


Figure 5: Another example with $a_1 + a_2$ far from 1

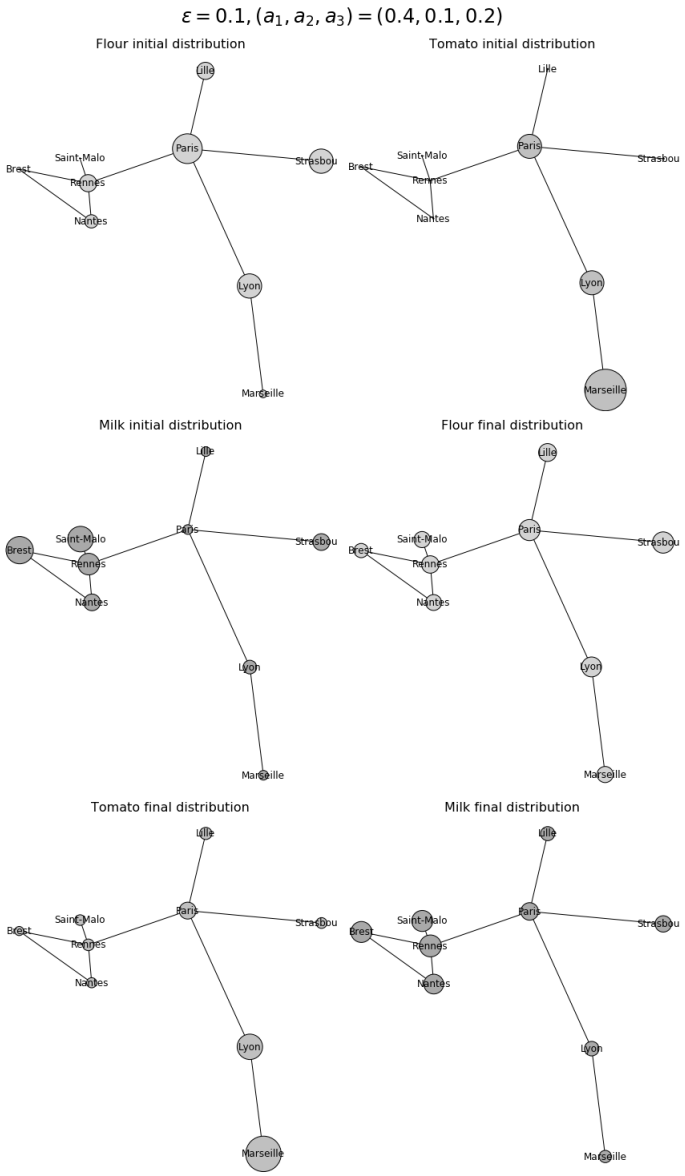
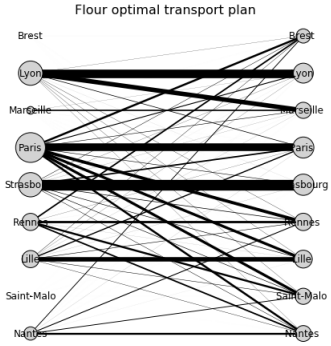
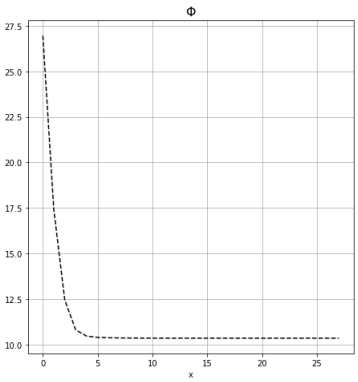
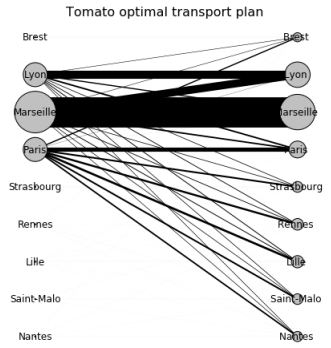


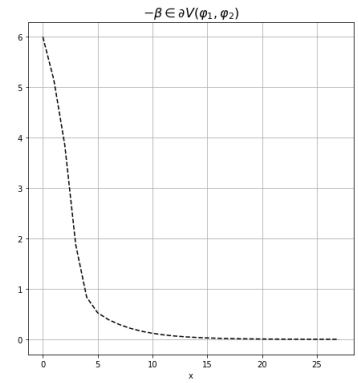
Figure 6: French network with $\varepsilon = 0.1$

	φ_1	φ_2	φ_3	U
BREST	0.4888	0.3004	0.1134	0.42
LYON	0.3476	0.0534	0.3259	1.47
MARSEILLE	0.448	0.0232	0.3829	1.05
PARIS	0.3224	0.1256	0.2331	1.4
STRASBOURG	0.2838	0.2754	0.2393	1.08
RENNES	0.4121	0.2304	0.1299	0.77
LILLE	0.3381	0.1805	0.2708	0.74
SAINT-MALO	0.4359	0.2511	0.1281	0.61
NANTES	0.4296	0.2627	0.1386	0.56









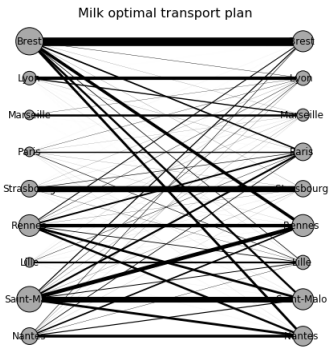


Figure 7: French network with $\varepsilon = 0.1$

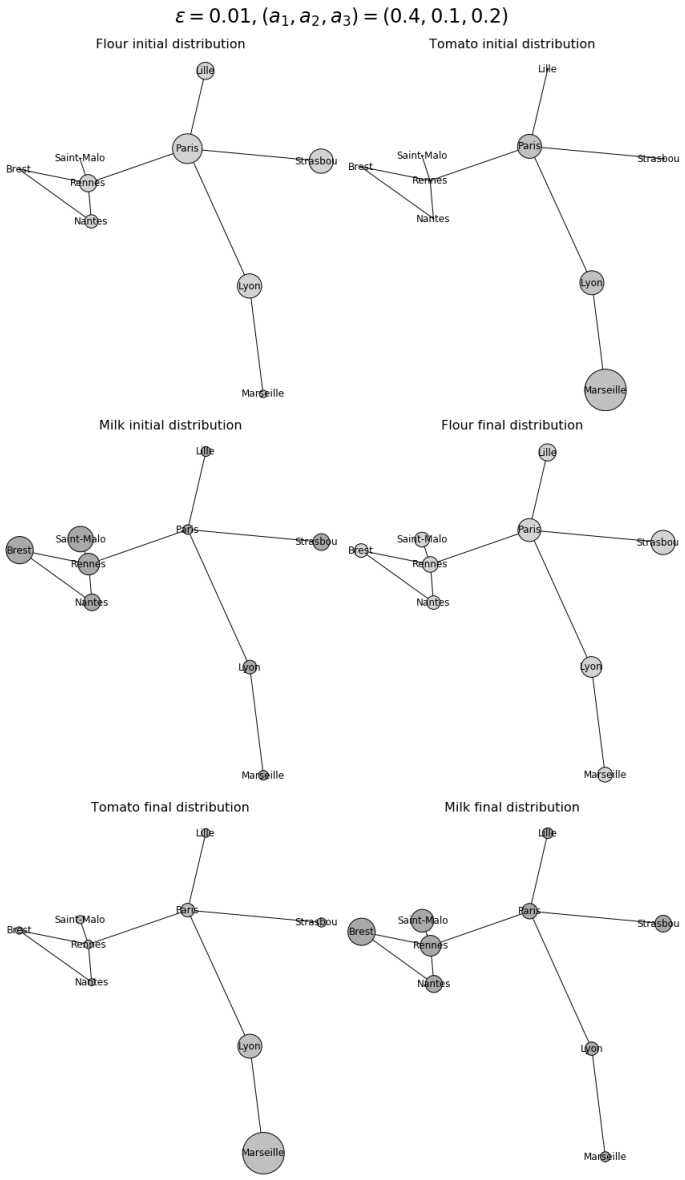
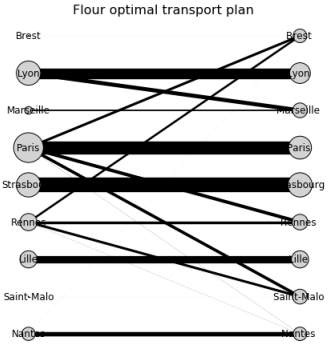
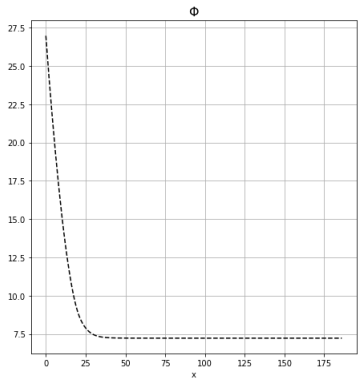
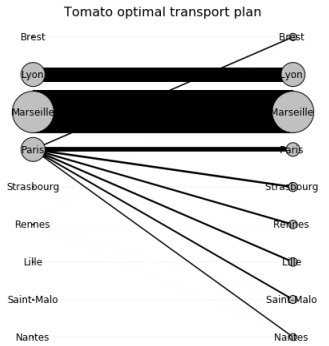


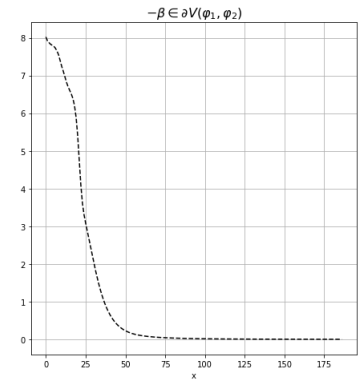
Figure 8: French network with $\varepsilon = 0.01$

	φ_1	φ_2	φ_3	U
BREST	0.5493	0.4755	0.0686	0.36
LYON	0.3171	0.059	0.38	1.53
MARSEILLE	0.4722	0.015	0.5133	0.88
PARIS	0.2657	0.1882	0.2921	1.44
STRASBOURG	0.236	0.42	0.2487	1.34
RENNES	0.432	0.3588	0.1224	0.49
LILLE	0.2939	0.2986	0.3867	0.5
SAINT-MALO	0.47	0.3965	0.1002	0.44
NANTES	0.456	0.4148	0.1443	0.3









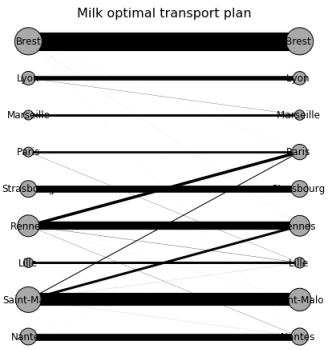


Figure 9: French network with $\varepsilon = 0.01$