HAL
open science

Computation of Cournot-Nash equilibria by entropic regularization

Adrien Blanchet, Guillaume Carlier, Luca Nenna

To cite this version:

Adrien Blanchet, Guillaume Carlier, Luca Nenna. Computation of Cournot-Nash equilibria by entropic regularization. 2016. hal-01363468

HAL Id: hal-01363468

https://inria.hal.science/hal-01363468v1

Preprint submitted on 9 Sep 2016

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.

Computation of Cournot-Nash equilibria by entropic regularization

Adrien Blanchet [*], Guillaume Carlier [†], Luca Nenna [‡]

September 9, 2016

Abstract

We consider a class of games with continuum of players where equilibria can be obtained by the minimization of a certain functional related to optimal transport as emphasized in [7]. We then use the powerful entropic regularization technique to approximate the problem and solve it numerically in various cases. We also consider the extension to some models with several populations of players.

Keywords: Cournot-Nash equilibria, optimal transport, entropic regularization.

AMS Subject Classifications:.

1 Introduction

There is a long tradition in economics and game theory, since the seminal works of Aumann [2], [3], of considering equilibria in games with a continuum of players, each of whom having a negligible influence on the output of the others. In particular, Schmeidler [18] introduced a notion of non-cooperative equilibrium in games with a continuum of agents and, Mas-Colell [15] reformulated Schmeidler's analysis in terms of joint distributions over players' actions and characteristics and emphasized the concept of Cournot-Nash

[*]GREMAQ-TSE, Université de Toulouse, Manufacture des Tabacs 21 allée de Brienne, 31000 Toulouse, FRANCE `Adrien.Blanchet@univ-tlse1.fr`.

[†]Université Paris-Dauphine, PSL Research University, CNRS, CEREMADE, 75016, Paris, France and Inria-Paris, MOKAPLAN `carlier@ceremade.dauphine.fr`.

[‡]Inria-Paris, MOKAPLAN, 2, rue Simone Iff, 75012, Paris, France and CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine `luca.nenna@inria.fr`.

equilibrium distributions. There are many examples where such concepts are relevant such as strategic route use in road traffic or networks, social interactions....

The problem can be described as follows: heterogeneous players each have to choose a strategy (or a probability over strategies, i.e. mixed strategies are allowed) so as to minimize a cost, the latter depending on the choice of the whole population of players only through the distribution of their strategies. In other words, on the one hand, each player, has a negligible influence on the cost. On the other hand, the interactions between players are of mean-field type: it does not matter who plays such and such strategy but rather how many players chose it. There are different mean-field effects, of different nature and which can be either repulsive (i.e. favoring dispersion of strategies) or attractive (favoring concentration of strategies). Congestion (the cost of a strategy is higher if it is frequently played) is a typical example of dispersive effect. In realistic models however, there are also attractive effects: choosing a strategy which is "far" from the strategies played by the other players may be risky or result in some cost.. It should then come as no surprise that, due to such opposite effects, the analysis of equilibria is complex in general. This explains why, in general, one cannot go much further than proving an existence result, as for instance, following the very elegant approach of Mas-Colell [15].

More recently, the first two authors [7] (also see [6]) emphasized the fact that for a separable class of costs, Cournot-Nash equilibria can be obtained by the minimization of a certain functional on the set of measures on the space of strategies. This functional typically involves two terms: an optimal transport cost and a more standard integral functional which may capture both congestion and attractive effects (as in [14]). Interestingly, this kind of minimization problem is very close to the semi-implicit Euler scheme introduced in the seminal work of [13] for Wasserstein gradient flows (for which we refer to [1]).

The variational approach of [7] is somehow more constructive and informative (but less general since it requires a separable cost) than the one relying on fixed-point arguments as in [15] but the optimal transport term cost is delicate to handle. It is indeed costly in general to solve an optimal transport problem and compute an element of the subdifferential of the optimal cost as a function of its marginals. In recent years, however it has been realized that a powerful way to approximate optimal transport is by adding an entropic penalization term. Doing so, the problem becomes projecting for the Kullback-Leibler divergence a given joint measure on the set of measures with fixed marginals, a task that can be achieved very effi-

ciently by alternate projections (see e.g. Bauschke and Lewis [4], Dysktra [10]) as shown by Cuturi [9]. This powerful method is intimately related to Sinkhorn algorithm and the Iterated Proportional Fitting Procedure (IPFP), well-known to statisticians and recently remise au goût du jour by Galichon and Salanié [12] for the estimation of matching models (we also refer to the recent book of Galichon [11] for a broader perspective on optimal transport methods, with or without regularization, in economics and econometrics). Various applications of the IPFP/Sinkhorn algorithm to optimal transport can be found in [5]. In order to take advantage of the power of entropic regularization on Wasserstein gradient flows, Peyré [16] introduced an extension of Dykstra's algorithm which he called Dykstra proximal splitting. It turns out that Peyré's algorithm, recently extended by Chizat et al. [8], is perfectly well-suited to the computation of Cournot-Nash equilibria as we try to explain in the sequel of the paper and illustrate by various numerical examples. We would also like to emphasize that, in the context of Cournot-Nash equilibrium, entropic regularization is also natural from a theoretical point, it amounts to replace exact cost minimization by some Gibbs-like measure or, equivalently to assume that the cost involves some random term.

The paper is organized as follows. In section 2, we recall the concept of Cournot-Nash equilibria, its variational counterpart and the entropic regularization of the latter. In Section 3, we describe the proximal splitting algorithm and the semi-implicit approach. In section 4, we present various numerical results both in dimension one and two and emphasize the influence of the transport cost on the structure of equilibria. Section 5 extends the previous analysis to some models with several populations for which we also present numerical results.

2 Cournot-Nash and entropic Cournot-Nash equilibria

We will restrict ourselves here to the following *finite* Cournot-Nash setting. Not only this will simplify the exposition and enable us to give a simple and self-contained exposition of the variational approach but this will also be consistent with our numerical scheme which anyway considers a finite number of agents' types and a finite number of strategies. We refer to [7] for the analysis of the continuum case. We consider a population of players, each of whom is characterized by a type which takes values in the type set $X := \{x_i\}_{i\in I}$ where I is finite. The frequencies of the players' type in the population is given by a probability $\mu := \{\mu_i\}_{i\in I}$ with $\mu_i \geq 0$ and

$\sum_{i=1}^{N} \mu_i = 1$. Each agent has to choose a strategy y from the strategy set $Y := \{y_j\}_{j\in J}$ with J finite. The unknown of the problem is a matrix $\gamma := \{\gamma_{ij}\}_{i\in I,\, j\in J}$ where γ_{ij} is the probability that a player of type x_i chooses strategy y_j, there is an obvious feasibility constraint on this matrix, obviously it should have nonnegative entries and its first *marginal* should match the given distribution of players μ i.e.:

$$\sum_{j\in J} \gamma_{ij} = \mu_i,\ \forall i \in I. \tag{2.1}$$

The matrix γ induces a probability $\nu = \Lambda_2(\gamma) = \{\nu_j\}_{j\in J}$ on the set of strategies given by its *second marginal*:

$$\nu_j := \sum_{i\in I} \gamma_{ij},\ \forall j \in J. \tag{2.2}$$

Agents of type x_i who play strategy y_j incur a cost that not only depends on x_i and y_j but also on the whole probability $\nu := \{\nu_j\}_{j\in J}$ on the strategy space induced by the behavior of the whole population of players, and we denote this cost by $\Psi_{ij}[\nu]$. An equilibrium is then a probability matrix γ which is feasible and which is consistent with the cost minimizing behavior of players, which is summarized in the next definition:

Definition 2.1. *A Cournot-Nash equilibrium is a matrix $\gamma = \{\gamma_{ij}\}_{i\in I,\, j\in J} \in \mathbb{R}_+^{I\times J}$ which satisfies the feasibility constraint (2.1) and such that, defining the strategy marginal $\nu = \Lambda_2(\gamma)$ by (2.2), one has*

$$\gamma_{ij} > 0 \Rightarrow \Psi_{ij}[\nu] = \min_{k\in J} \Psi_{ik}[\nu].$$

Provided Ψ_{ij} depends continuously on ν, the existence of an equilibrium can easily be proven by Kakutanis' fixed-point theorem, but not much more can be said, at this level of generality. If one further specifies the form of the cost, as we shall do now, following [7], one may obtain equilibria by minimizing a certain cost functional.

2.1 A variational approach to Cournot-Nash equilibria

We now suppose that the cost $\Psi_{ij}[\nu]$ takes the following separable form

$$\Psi_{ij}[\nu] := c_{ij} + f_j(\nu_j) + \sum_{k\in J} \phi_{kj}\nu_k$$

where $c := \{c_{ij}\}_{i\in I,\, j\in J} \in \mathbb{R}^{I\times J}$, each function f_j is nondecreasing and continuous, the matrix $\phi := \{\phi_{kj}\} \in \mathbb{R}^{J\times J}$ is symmetric, i.e. $\phi_{kj} = \phi_{jk}$,

A possible interpretation of this model is the following: the players represent a population of doctors, their type x represent their region of origin and their y strategy represent the location where they chose to dwell, the total cost of x_i-type doctors is the sum of

- a transport cost $c_{ij} = c(x_i, y_j)$,
- a congestion cost $f_j(\nu_j)$: if location y_j is very crowded i.e. if ν_j is large, the doctors settling at y_j will see their benefit decrease,
- an interaction cost with the rest of the population of doctors, one can think that ϕ_{kj} is an increasing function of some distance between y_k and y_j so that $\sum_{k\in J} \phi_{kj}\nu_k$ represents the average distance to the rest of the population.

The variational approach of [7] relies on optimal transport, and we shall give a self-contained and simple presentation in the present discrete setting. Firstly it is useful to introduce the marginal maps:

$$\gamma \in \mathbb{R}^{I\times J} \mapsto \Lambda_1(\gamma) = \alpha \in \mathbb{R}^I,\ \alpha_i := \sum_{j\in J} \gamma_{ij},$$

and

$$\gamma \in \mathbb{R}^{I\times J} \mapsto \Lambda_2(\gamma) = \nu \in \mathbb{R}^J,\ \nu_j := \sum_{i\in I} \gamma_{ij},$$

as well as

$$\mathrm{C}_\mu := \{\gamma = \{\gamma_{ij}\}_{i\in I,\, j\in J} \in \mathbb{R}_+^{I\times J}\ :\ \Lambda_1(\gamma) = \mu\}$$

which is the set of probabilities on $X \times Y$ having μ as first marginal (recall that μ is fixed). For $\nu = \{\nu_j\}_{j\in J} \in \mathbb{R}_+^J$ such that $\sum_{j\in J} \nu_j = 1$, let us also define

$$\mathrm{C}^\nu := \{\gamma = \{\gamma_{ij}\}_{i\in I,\, j\in J} \in \mathbb{R}_+^{I\times J}\ :\ \Lambda_2(\gamma) = \nu\}$$

as the set of probabilities on $X \times Y$ having ν as second marginal. Let us then also define the set of transport plans between μ and ν as

$$\Pi(\mu, \nu) := \mathrm{C}_\mu \cap \mathrm{C}^\nu. \tag{2.3}$$

Given ν a probability on Y, let us define

$$\mathrm{MK}(\nu) := \inf_{\gamma\in\Pi(\mu,\nu)} \Big\{c \cdot \gamma := \sum_{i,j\in I\times J} c_{ij}\gamma_{ij}\Big\} \tag{2.4}$$

that is the value of the optimal transport problem between μ and ν for the cost c. Setting

$$\mathcal{P}(Y) := \{\nu \in \mathbb{R}_+^J\ :\ \sum_{j\in J} \nu_j = 1\}$$

consider the optimization problem

$$\inf_{\nu\in\mathcal{P}(Y)} \mathrm{MK}(\nu) + E(\nu) \tag{2.5}$$

where the energy E is given by

$$E(\nu) := \sum_{j\in J} F_j(\nu_j) + \frac{1}{2}\sum_{k,j\in J\times J} \phi_{kj}\nu_k\nu_j \tag{2.6}$$

and F_j is a primitive of the congestion function f_j:

$$F_j(t) := \int_0^t f_j(s)\mathrm{d}s.$$

We then have

Theorem 2.2. *Let ν solve (2.5) and $\gamma \in \Pi(\mu,\nu)$ be such that $c\cdot\gamma = \mathrm{MK}(\nu)$, then γ is a Cournot-Nash equilibrium. This implies in particular that there exists Cournot-Nash equilibria.*

Proof. We have to prove that whenever $\gamma_{ij} > 0$ one has

$$c_{ij} + f_j(\nu_j) + \sum_{k\in j} \phi_{kj}\nu_k = u_i \tag{2.7}$$

with

$$u_i := \min_{j\in J}\{c_{ij} + f_j(\nu_j) + \sum_{k\in j}\phi_{kj}\nu_k\}.$$

First observe that E is of class C^1 and by construction

$$\frac{\partial E}{\partial \nu_j} = f_j(\nu_j) + \sum_{k\in j}\phi_{kj}\nu_k. \tag{2.8}$$

To treat the transport term, MK, we shall recall the classical Kantorovich duality (see [19], [17]) as follows. Firstly for $v \in \mathbb{R}^J$ let us define

$$\mathrm{K}(v) := -\sum_{i\in I}\min_{j\in J}(c_{ij} - v_j)\mu_i$$

note that K is a convex and Lipschitz function whose conjugate, thanks to Kantorovich duality, can be expressed as

$$K^*(\nu) = \overline{\mathrm{MK}}(\nu) := \begin{cases} \mathrm{MK}(\nu) \text{ if } \nu \in \mathcal{P}(Y), \\ +\infty \text{ otherwise.} \end{cases}$$

Since ν minimizes $\overline{\mathrm{MK}} + E$, one has $0 \in \partial\mathrm{MK}(\nu) + \nabla E(\nu)$, setting $v := -\nabla E(\nu)$, this can be rewritten as $\nu \in \partial\overline{\mathrm{MK}}^*(v) = \partial K(v)$ and since $\overline{\mathrm{MK}}(\nu) = c \cdot \gamma$ this gives

$$\begin{aligned}\mathrm{MK}(\nu) = c \cdot \gamma &= \sum_{j \in \mathcal{J}} v_j \nu_j + \sum_{i \in I} \min_{j \in J} (c_{ij} - v_j)\mu_i \\ &= \sum_{j \in \mathcal{J}} v_j \nu_j + \sum_{i \in I} u_i \mu_i = \sum_{i,j \in I \times J} (u_i + v_j)\gamma_{ij}.\end{aligned}$$

which, since $u_i + v_j \leq c_{ij}$ implies that whenever $\gamma_{ij} > 0$, one has $c_{ij} - v_j = u_i$ which is exactly (2.7). This clearly implies the existence of Cournot-Nash equilibria since $\mathcal{P}(Y)$ is compact and both MK and E are continuous. $\square$

Note that if E is convex then the optimality condition $0 \in \partial\overline{\mathrm{MK}}(\nu) + \nabla E(\nu)$ is necessary and sufficient and there is actually an equivalence between being an equilibrium and being a minimizer in this case.

2.2 Entropic regularization

Solving (2.5) in practice (even if E is convex) might be difficult because of the transport cost term MK for which it is expensive to compute a subgradient. There is however a simple regularization of MK which is much more convenient to handle: the entropic regularization (see [5, 9, 12]). Given a regularization parameter $\varepsilon > 0$, let us define for every $\nu \in \mathcal{P}(Y)$:

$$\mathrm{MK}_\varepsilon(\nu) := \inf_{\gamma \in \Pi(\mu,\nu)} \Big\{ c \cdot \gamma + \varepsilon \sum_{i,j \in I \times J} \gamma_{ij}(\ln(\gamma_{ij}) - 1) \Big\}.$$

We then consider the regularization of (2.5)

$$\inf_{\nu \in \mathcal{P}(Y)} \mathrm{MK}_\varepsilon(\nu) + E(\nu) \tag{2.9}$$

where E is again given by (2.6). Thanks to the entropic regularization term, (2.9) is a smooth minimization problem which consists in minimizing with respect to γ and ν the objective

$$c \cdot \gamma + \varepsilon \sum_{i,j \in I \times J} \gamma_{ij}(\ln(\gamma_{ij}) - 1) + E(\nu)$$

subject to $\gamma_{ij} \geq 0$ (but because of the entropy, these nonnegativity constraints are not binding) and the linear marginal constraints $\gamma \in \Pi(\mu, \nu)$. The first-order optimality conditions give the following Gibbs form for γ_{ij}:

$$\gamma_{ij} = a_i \exp\Big(-\frac{1}{\varepsilon}(c_{ij} + f_j(\nu_j) + \sum_{k \in J} \phi_{kj}\nu_k)\Big) \tag{2.10}$$

for some $a_i > 0$ which has to fulfill the first marginal constraint i.e.

$$a_i = \frac{\mu_i}{\sum_{j\in J} \exp\Big(-\frac{1}{\varepsilon}(c_{ij} + f_j(\nu_j) + \sum_{k\in J} \phi_{kj}\nu_k)\Big)}.$$

Note that these conditions can also be interpreted as a regularized form of a Cournot-Nash equilibrium since they mean that the conditional probabilities on the set of strategies given the players type $\{\frac{\gamma_{ij}}{\mu_i}\}_{j\in J}$ are proportional to $\exp(-\frac{\Psi_{ij}(\nu)}{\varepsilon})$ where $\Psi_{ij}[\nu] = c_{ij}+f_j(\nu_j)+\sum_{k\in J}\phi_{kj}\nu_k$ is the total cost incurred by players x_i when choosing strategy y_j. Another equilibrium interpretation (which is customary in economics and econometrics in the framework of discrete choice models) is to consider that the total cost actually contains a random component that is of the form εX_{ij} where the X_{ij} are i.i.d. logistic random variables (see [11]).

Of course, again when E is convex, since $\overline{\mathrm{MK}}_\varepsilon$ is strictly convex, there is a unique minimizer and the first-order optimality condition for (2.9) is necessary and sufficient so that there is again equivalence between being a minimizer and a (regularized) Cournot-Nash equilibrium.

3 A proximal splitting algorithm

To solve (2.9), we shall use a proximal splitting scheme using the Kullback-Leibler divergence that was recently introduced by Peyré [16] in the context of entropic regularization of Wasserstein gradient flows and extended recently by Chizat et al. [8]. First, let us observe that (2.9) can be rewritten as a special instance of a Bregman proximal problem. To see this, let us first rewrite

$$c\cdot\gamma + \varepsilon \sum_{i,j\in I\times J} \gamma_{ij}(\ln(\gamma_{ij}) - 1) = \varepsilon \sum_{i,j\in I\times J} \gamma_{ij}(\ln\Big(\frac{\gamma_{ij}}{e^{-\frac{c_{ij}}{\varepsilon}}}\Big) - 1)$$

which is the same as $\varepsilon\mathrm{KL}(\gamma|\overline{\gamma})$ where $\overline{\gamma}_{ij} = e^{-\frac{c_{ij}}{\varepsilon}}$ and KL is the Kullback-Leibler divergence

$$\mathrm{KL}(\gamma|\theta) := \sum_{i,j\in I\times J} \gamma_{ij}\Big(\ln\Big(\frac{\gamma_{ij}}{\theta_{ij}}\Big) - 1\Big),\ \gamma\in\mathbb{R}_+^{I\times J},\ \theta\in\mathbb{R}_+^{I\times J}.$$

Note that KL is the Bregman divergence associated to the entropy. Solving (2.9) then amounts to the proximal problem

$$\mathrm{prox}_G^{\mathrm{KL}}(\overline{\gamma}) = \mathrm{argmin}_{\gamma\in\mathbb{R}_+^{I\times J}} \Big\{\mathrm{KL}(\gamma|\overline{\gamma}) + G(\gamma)\Big\} \tag{3.1}$$

with

$$G(\gamma) := \chi_{\{\Lambda_1(\gamma)=\mu\}} + \frac{1}{\varepsilon} E(\Lambda_2(\gamma)).$$

Computing directly $\mathrm{prox}_G^{\mathrm{KL}}(\overline{\gamma})$ may be an involved task, but the idea of Peyré's splitting algorithm is to express G as a sum of more elementary functionals:

$$G := \sum_{l=1}^{L} G_l$$

each of whom being simple in the sense that computing $\mathrm{prox}_{G_l}^{\mathrm{KL}}$ can be done easily (ideally in close form). The algorithm proposed by Peyré generalizes Dykstras' algorithm for KL projections on the intersection of convex sets and can be described as follows. First extend the sequence of functions $G_1, \cdots, G_L$ by periodicity:

$$G_{l+nL} = G_l,\ l = \{1, \cdots, L\},\ n \in \mathbb{N}$$

initialize the algorithm by setting the following values for the $I \times J$ matrices

$$\gamma^{(0)} = \overline{\gamma},\ z^{(0)} = z^{(-1)} = \cdots = z^{(-L+1)} = e,\ e_{ij} = 1,\ (i,j) \in I \times J,$$

and then iteratively define for $n \geq 1$

$$\gamma^{(n)} = \mathrm{prox}_{G_n}^{\mathrm{KL}}\Big(\gamma^{n-1} \odot z^{(n-L)}\Big) \tag{3.2}$$

and

$$z^{(n)} = z^{(n-1)} \odot \Big(\gamma^{(n-1)} \oslash \gamma^{(n)}\Big) \tag{3.3}$$

where $\odot$ and $\oslash$ stand for entry-wise multiplication/division operations:

$$(\gamma \odot \theta)_{ij} = \gamma_{ij}\theta_{ij},\ (\gamma \oslash \theta)_{ij} = \frac{\gamma_{ij}}{\theta_{ij}}.$$

We refer to [16] and [8] for the convergence of this algorithm under suitable assumptions (convexity of the functions G_l and a certain qualification condition), the idea being that at the level of the dual problem, which is smooth, this algorithm amounts to perform an alternate block minimization.

3.1 A class of convex problems

Note that the congestion term $\sum_{j\in J} F_j(\nu_j)$ is convex because f_j is nondecreasing, but the quadratic interaction energy $\nu \mapsto \sum_{j,k\in J\times J} \phi_{kj}\nu_k\nu_j$ is in general not convex. However, using Cauchy-Schwarz inequality, it satisfies

$$\sum_{j,k\in J\times J} \phi_{kj}\nu_k\nu_j \geq -\Big(\sum_{j,k\in J\times J} \phi_{kj}^2\Big)\sum_{j\in J} \nu_j^2$$

so that if F_j is 1-strongly convex:

$$F_j(t) = \frac{1}{2}t^2 + H_j(t)$$

with H_j convex and

$$\sum_{j,k\in J\times J} \phi_{kj}^2 < 1, \tag{3.4}$$

then E is convex as the sum $E = E_2 + E_3$ of the convex quadratic term

$$E_2(\nu) := \frac{1}{2}\sum_{j\in J} \nu_j^2 + \frac{1}{2}\sum_{k,j\in J\times J} \phi_{kj}\nu_k\nu_j$$

and the remaining convex congestion term

$$E_3(\nu) := \sum_{j\in J} H_j(\nu_j).$$

In this setting one can write (2.9) as

$$\inf_{\gamma\in\mathbb{R}_+^{I\times J}} \Big\{\mathrm{KL}(\gamma|\overline{\gamma}) + G_1(\gamma) + G_2(\gamma) + G_3(\gamma)\Big\}$$

where

$$G_1(\gamma) = \chi_{\{\Lambda_1(\gamma)=\mu\}} = \begin{cases} 0 \text{ if } \Lambda_1(\gamma) = \mu \\ +\infty \text{ otherwise} \end{cases}$$

and

$$G_2 = \frac{1}{\varepsilon}E_2 \circ \Lambda_2,\ G_3 = \frac{1}{\varepsilon}E_3 \circ \Lambda_2.$$

To implement the proximal splitting scheme (3.2)-(3.3) in this case, one has to be able to compute the three proximal maps $\mathrm{prox}_{G_l}^{\mathrm{KL}}$ with $l = 1, 2, 3$. The proximal map of G_1 corresponds to the fixed marginal constraint $\Lambda_1(\gamma) = \mu$, it is well-known and it is given in closed form as:

$$\Big(\mathrm{prox}_{G_1}^{\mathrm{KL}}(\theta)\Big)_{ij} = \frac{\mu_i\theta_{ij}}{\sum_{k\in J}\theta_{ik}}.$$

Given $\theta \in \mathbb{R}_+^{I\times J}$, $\gamma := \mathrm{prox}_{G_2}^{\mathrm{KL}}(\theta)$ is of the form

$$\gamma_{ij} = \theta_{ij}\exp\Big(-\frac{\nu_j + \sum_{k\in J}\phi_{kj}\nu_k}{\varepsilon}\Big)$$

where ν denotes the second marginal of γ, so that summing over i, ν is obtained by solving the system:

$$\nu_j = \Big(\sum_{i\in I}\theta_{ij}\Big)\exp\Big(-\frac{\nu_j + \sum_{k\in J}\phi_{kj}\nu_k}{\varepsilon}\Big)$$

which, when (3.4) holds, can be solved in practice in a few Newton's steps. The computation of $\gamma := \mathrm{prox}^{\mathrm{KL}}_{G_3}(\theta)$ is simpler, setting $h_j := H_j'$ the first-order equation first leads to

$$\gamma_{ij} = \theta_{ij}\exp\Big(-\frac{h_j(\nu_j)}{\varepsilon}\Big)$$

and the ν_j's are obtained by solving

$$\nu_j = \Big(\sum_{i\in I}\theta_{ij}\Big)\exp\Big(-\frac{h_j(\nu_j)}{\varepsilon}\Big) \tag{3.5}$$

which is a separable system of monotone equations, which we shall again solve by Newton's method.

3.2 A semi-implicit scheme for more general nonconvex cases

We now go back to the general case where E is not necessarily convex because of the interaction term given by the symmetric matrix ϕ_{kj}. Even though there is no theoretical convergence guarantee (but if the following scheme converges, it converges to an equilibrium), the semi-implicit scheme which we now describe gives good results in practice. The idea is simple and consists in replacing the nonconvex interaction term by its linearization. More precisely, we will approximate our initial problem (2.9):

$$\inf_{\nu\in\mathcal{P}(Y)} \mathrm{MK}_\varepsilon(\nu) + E(\nu) \tag{3.6}$$

where E is the sum of the convex congestion cost and the nonconvex quadratic interaction cost, by a succession of convex problems, starting from $\nu^0 \in \mathcal{P}(Y)$, iteratively solve for $n \geq 1$

$$\nu^{(n+1)} = \mathrm{argmin}_{\nu\in\mathcal{P}(Y)}\, \mathrm{MK}_\varepsilon(\nu) + E^{(n)}(\nu) \tag{3.7}$$

where in $E^{(n)}$ we have linearized the interaction term:

$$E^{(n)}(\nu) = \sum_{j\in J} F_j(\nu_j) + \sum_{j\in J} V_j^{(n)}\nu_j,\ V_j^{(n)} := \sum_{k\in J}\phi_{kj}\nu_k^{(n)}.$$

Of course, we can solve (3.7) by the Dykstra proximal-splitting scheme described in the previous paragraph. More precisely, the linear term can be absorbed by the KL term so that we only have two proximal steps: one corresponding to the (explicit) projection fixed marginal constraint and one corresponding to the congestion cost (corresponding to (3.5) using f_j instead of h_j).

4 Numerical results

We now present some numerical results in dimension $d = 1$ and $d = 2$. As we have pointed out in section 2.2, the strength of the entropic regularization, and consequently of the Dykstra's algorithm, lies in the fact that we can treat optimal transportation problems with any transport cost, in particular both concave and convex cost functions can be considered. Thus, if we consider the cost $c_{ij} = |x_i - y_j|^p$ with $p > 0$ (convex cost if $p > 1$ and concave otherwise), one can analyze how the shape of the unknown marginal ν changes by varying the exponent p. Before showing the results, we want to focus on an other aspect of the entropic regularization, namely diffusion. Indeed, once we add the entropic term to the optimal transport term, then this regularization spreads the support of the plan γ and defines a strongly convex problem with a unique solution. So it is interesting to see how the support of the optimal γ varies by decreasing the parameter ε. Let us consider the standard quadratic cost $c_{ij} = |x_i - y_j|^2$ and the following energy $E(\nu)$

$$E(\nu) = \sum_{j \in J} \nu_j^8 + \frac{1}{2} \sum_{k,j \in J \times J} \phi_{kj} \nu_k \nu_j + \sum |y_j - 9|^4, \tag{4.1}$$

where $\phi_{kj} = 10^{-4}|y_k - y_j|^2$ and the third term is a confinement potential. We notice that there is no need to compute a proximal step for the potential, indeed it can be absorbed by the KL term. We know that in this case the optimal γ (for instance see [7]) is a pure Cournot-Nash equilibrium, which actually means that γ has the form $\gamma_T = (\mathrm{id}, T)_{\#}\mu$ where T is the optimal map. In Figure 1 we plot the support of the optimal γ and its marginal ν for different values of ε. As expected the support of the regularized γ concentrates on the graph of T as ε decreases.

In Section 3.2, we have pointed out that a semi-implicit approach can be applied in order to treat an energy E which is not convex. We want, now, to compare the performances of the implicit and the semi-implicit approach in terms of *CPU time* and *number of iterations* when ε varies. By looking at Figure 2, we notice the number of iterations, as well as the CPU time, of the

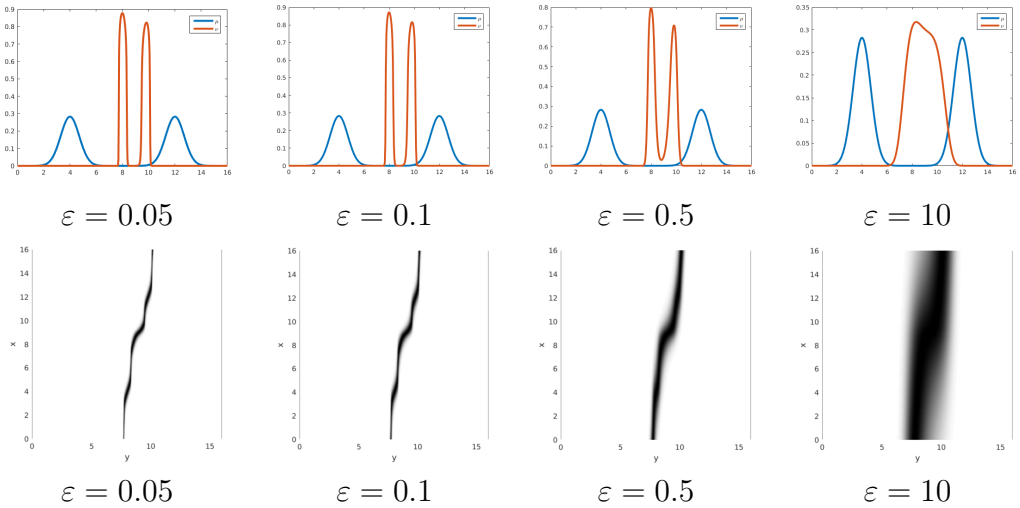


Figure 1: *Top: The initial distribution μ (blue solid line) and the solution ν (red solid line) for $\varepsilon \in \{0.05, 0.1, , 0.5, 10\}$. Bottom: The support of γ for $\varepsilon \in \{0.05, 0.1, , 0.5, 10\}$.*

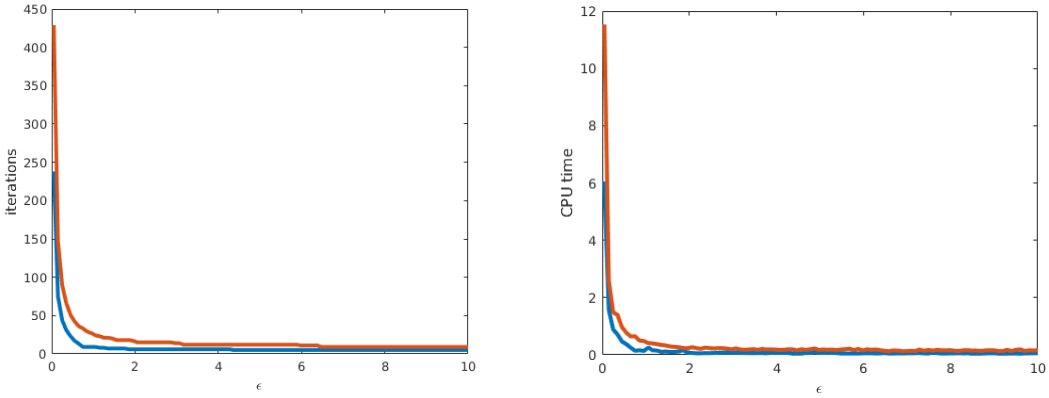
semi-implicit approach are smaller than the ones for the implicit approach. This is quite obvious as in the semi-implicit scheme, the interaction term can be absorbed by the KL term so that one has to compute only two proximal steps instead of three.

4.1 Dimension one

Let us first consider the one-dimensional case. One of the main advantages of the scheme we have proposed is that we can consider any kind of cost function. Thus, take $c_{ij} = |x_i - y_j|^p$ and the energy E given by (4.1), then we want to visualize the optimal ν as $p \in (0, M]$ with M large. For the simulations in Figure 3, we have used a $N = 500$ grid points discretization of $[0, 16]$ and we have treated the interaction term with a semi-implicit approach. Then, we have chosen the smallest ε possible for each cost function tested. As one can notice for $p \leq 1$ the optimal ν has a connected support whereas for $p > 1$, the support of ν is closer to the one of μ. Finally, we obtain an optimal ν which tends to be concentrated near $y = 9$ due to the external potential, except for large p where the optimal transport term becomes *dominant* so that the second marginal ν tends to be *close* to the initial distribution μ.

Let us now consider an energy E given by

$$E(\nu) = \sum_{j \in J} \ln(\nu_j) + \sum_{k,j \in J \times J} \phi_{kj} \nu_k \nu_j + \sum_{j \in J} (y_j - 5)^3 \tag{4.2}$$



iterations CPU time in seconds

Figure 2: *Left: the number of iterations for the semi-implicit (blue) and for the implicit (red). Right: CPU time for the semi-implicit (blue) and for the implicit (red).*

where ϕ_{kj} is a cubic interaction $\phi_{kj} = 10^{-4}|x_i - y_j|^3$. The simulations are presented in Figures 4 and 5 for different initial distribution: a uniform density on $[0, 1]$ and the sum of two translated Gaussians, respectively. For both the numerical experiments we have used $N = 500$ grid points discretization of $[0, 10]$ and treated the interaction term with a semi-implicit approach. One can observe, as in the previous case, that the structure of the optimal ν becomes close to the one of the initial ditribution as p increases.

4.2 Dimension two

For the $2d$ case, we always take $c(x, y) = \|x - y\|^p$, a congestion of the form $F_j(\nu_j) = \nu_j^8$, quadratic interactions $\phi_{kj} = 10^{-4}\|y_k - y_j\|^2$ and a potential $v_j = \|y_j - 3\|^4$. The simulations in Figure 6 are obtained by using a $N \times N$, with $N = 80$, discretization of $[0, 5]^2$ and by treating the interaction term with a semi-implicit approach. As in the $1-$dimensional case, we notice the same effect on the support of ν when we make p vary.

5 Extension to several populations

5.1 A class of two-populations models

We end the paper by briefly explaining how our approach can easily be extended to the case of several populations of players. For the sake of sim-

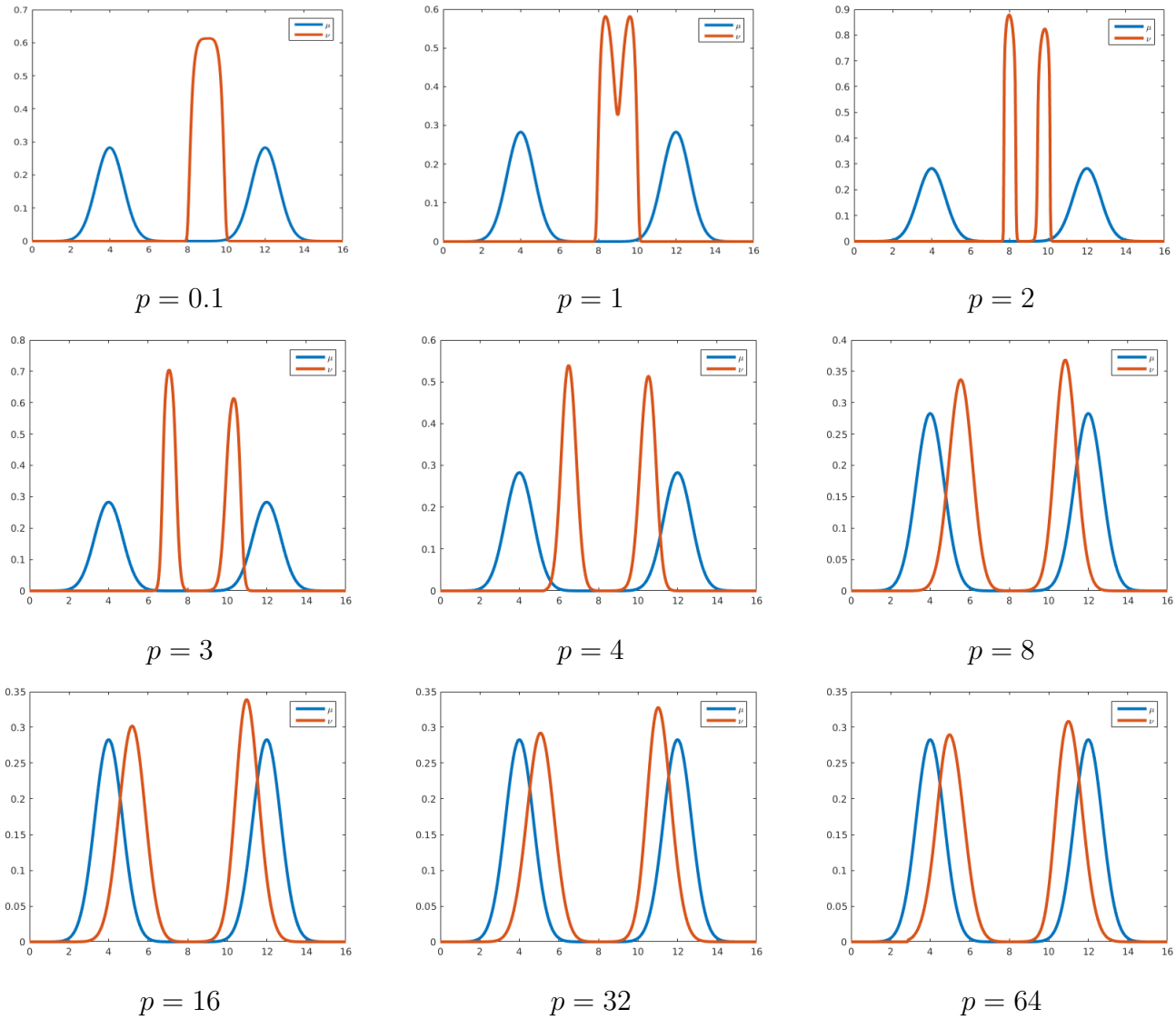


Figure 3: *The initial distribution μ, a sum of two translated Gaussian, (blue solid line) and the solution ν (red solid line) for $p \in \{0.1, 1, 2, 3, 4, 8, 16, 32, 64\}$.*

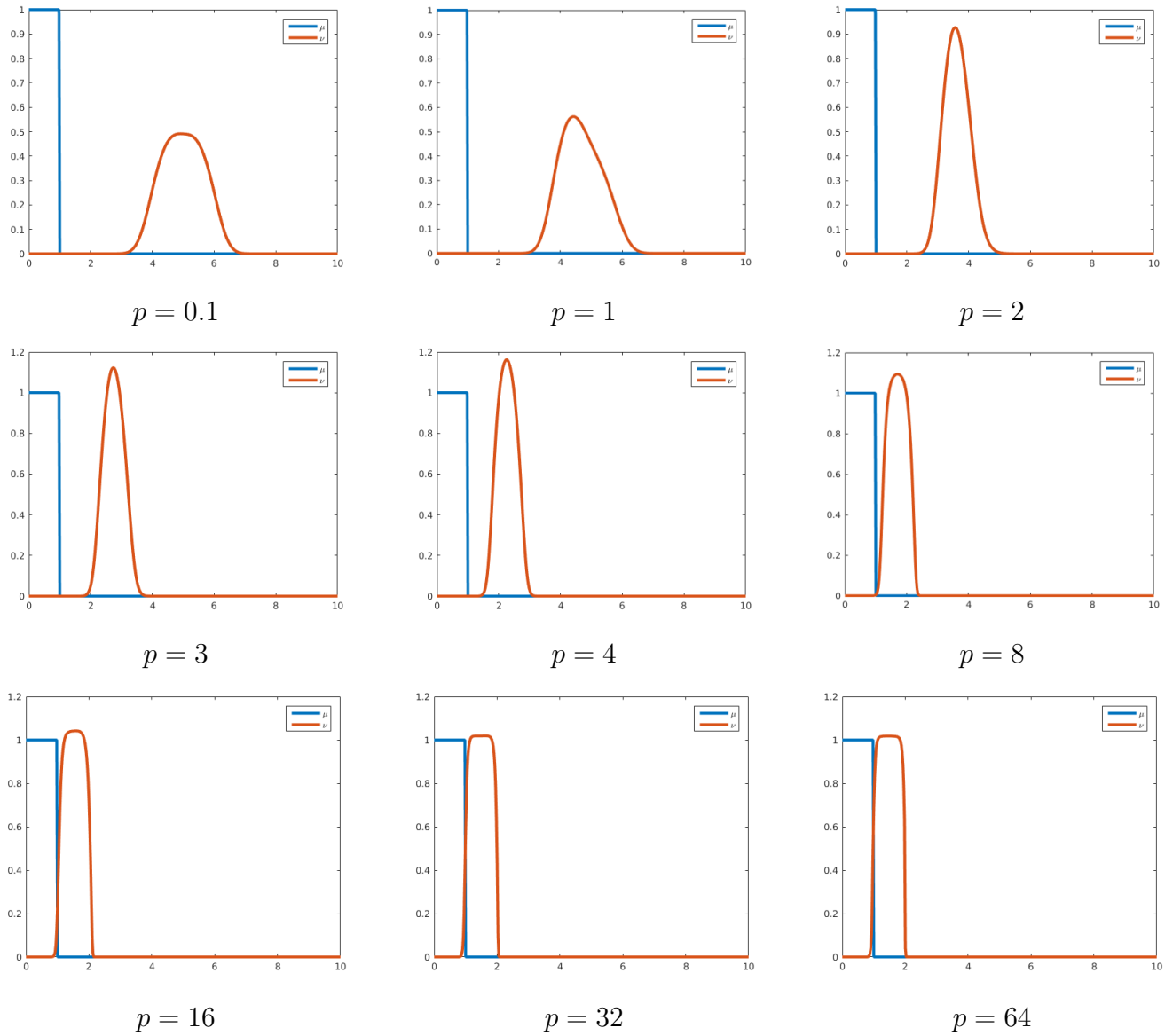


Figure 4: *The initial distribution μ, a uniform density on $[0, 1]$, (blue solid line) and the solution ν (red solid line) for $p \in \{0.1, 1, 2, 3, 4, 8, 16, 32, 64\}$.*

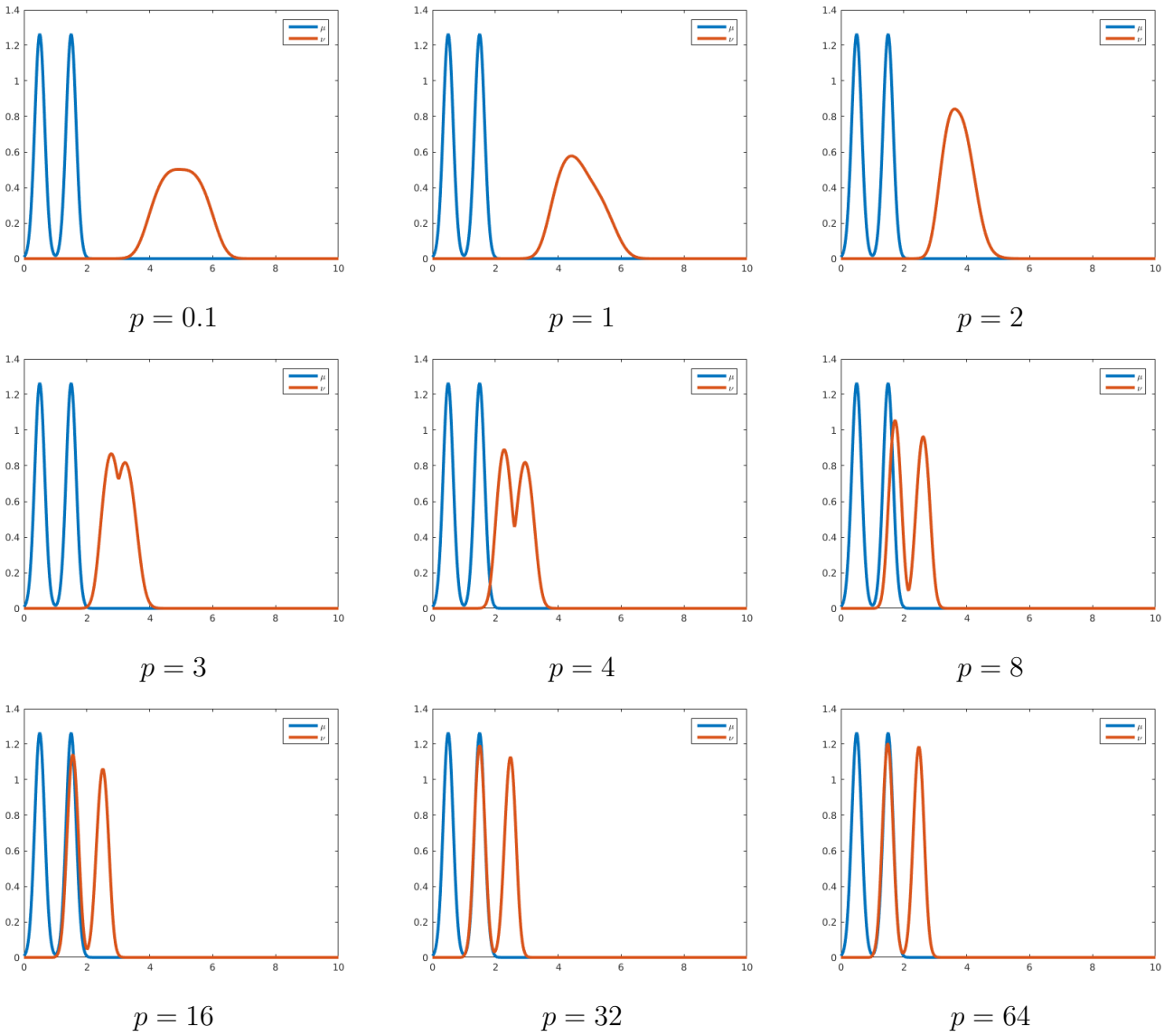
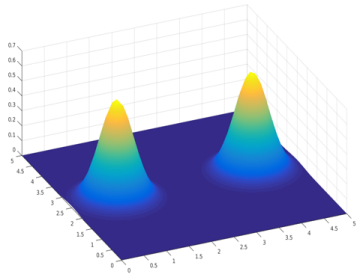
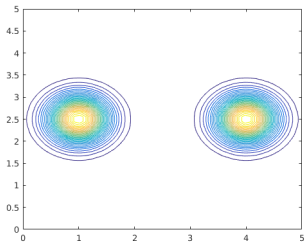


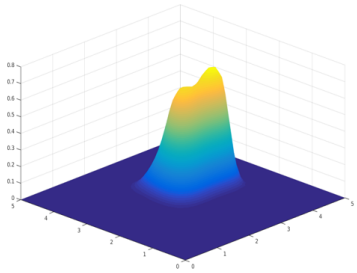
Figure 5: *The initial distribution μ, a sum of two translated Gaussians, (blue solid line) and the solution ν (red solid line) for $p \in \{0.1, 1, 2, 3, 4, 8, 16, 32, 64\}$.*



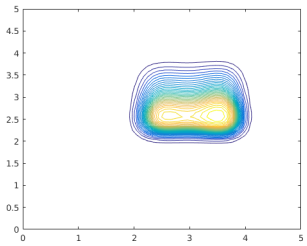
surface plot of μ



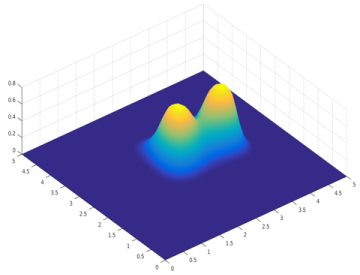
support of μ



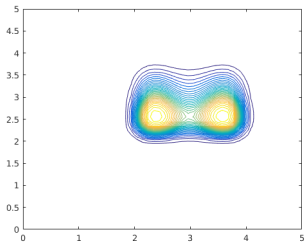
surface plot of ν for $p = 0.5$



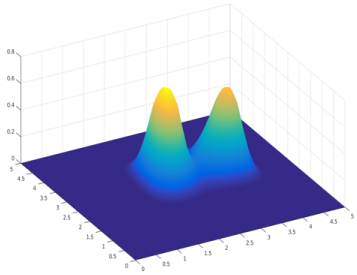
support of ν for $p = 0.5$



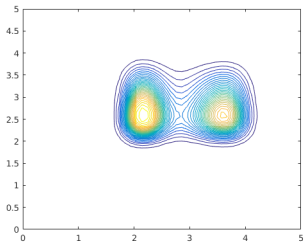
surface plot of ν for $p = 1$



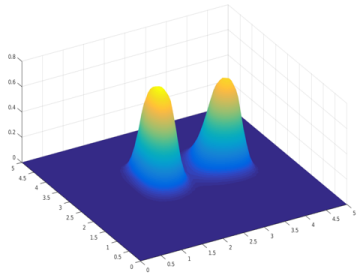
support of ν for $p = 1$



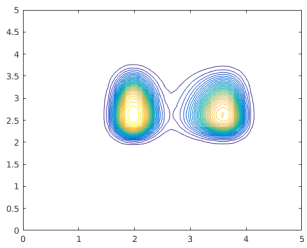
surface plot of ν for $p = 2$



support of ν for $p = 2$



surface plot of ν for $p = 4$



support of ν for $p = 4$

Figure 6: *The initial distribution μ, a sum of two translated Gaussian, and the solution ν for different values of p.*

plicity, we take the two-populations case and assume that these two populations interact through a congestion term. More precisely, we are given two finite type spaces $X_1 = \{x_i^1\}_{i\in I_1}$ and $X_2 = \{x_i^2\}_{i\in I_2}$, a common strategy space $Y = \{y_j\}_{j\in J}$, given distributions of the players types $\mu^1 \in \mathcal{P}(X_1)$, $\mu^2 \in \mathcal{P}(X_2)$, two transport cost matrices $c^1 \in \mathbb{R}^{I_1\times J}$, $c^2 \in \mathbb{R}^{I_2\times J}$, and consider the minimization problem:

$$\inf_{(\nu^1,\nu^2)\in\mathcal{P}(Y)\times\mathcal{P}(Y)} \Big\{ \mathrm{MK}^1_{\varepsilon_1}(\nu^1)+\mathrm{MK}^2_{\varepsilon_2}(\nu^2)+E_1(\nu^1)+E_2(\nu^2)+F(\nu^1+\nu^2) \Big\} \quad (5.1)$$

where for $l = 1, 2$, $\varepsilon_l > 0$ is a regularization (or noise) parameter, $\mathrm{MK}^l_{\varepsilon_l}(\nu^l)$ represents the regularized transport cost:

$$\mathrm{MK}^l_{\varepsilon_l}(\nu^l) := \inf_{\gamma\in\Pi(\mu^l,\nu^l)} \Big\{ c^l \cdot \gamma + \varepsilon_l \sum_{i,j\in I_l\times J} \gamma_{ij}(\ln(\gamma_{ij}) - 1) \Big\},$$

$E_l(\nu^l)$ represents an individual cost for population k, for instance, an interaction cost:

$$E_l(\nu^l) := \sum_{j,k\in J\times J} \phi^l_{kj}\nu^l_j\nu^l_k$$

and F is a total congestion cost

$$F(\nu^1 + \nu^2) := \sum_{j\in J} F_j(\nu^1_j + \nu^2_j)$$

where F_j is convex.

Remark 5.1. The proximal step related to F can be computed as in (3.5) by taking $\nu_j = \nu^1_j + \nu^2_j$.

5.2 Numerical Results

For the two populations case, we take the following energies E_l

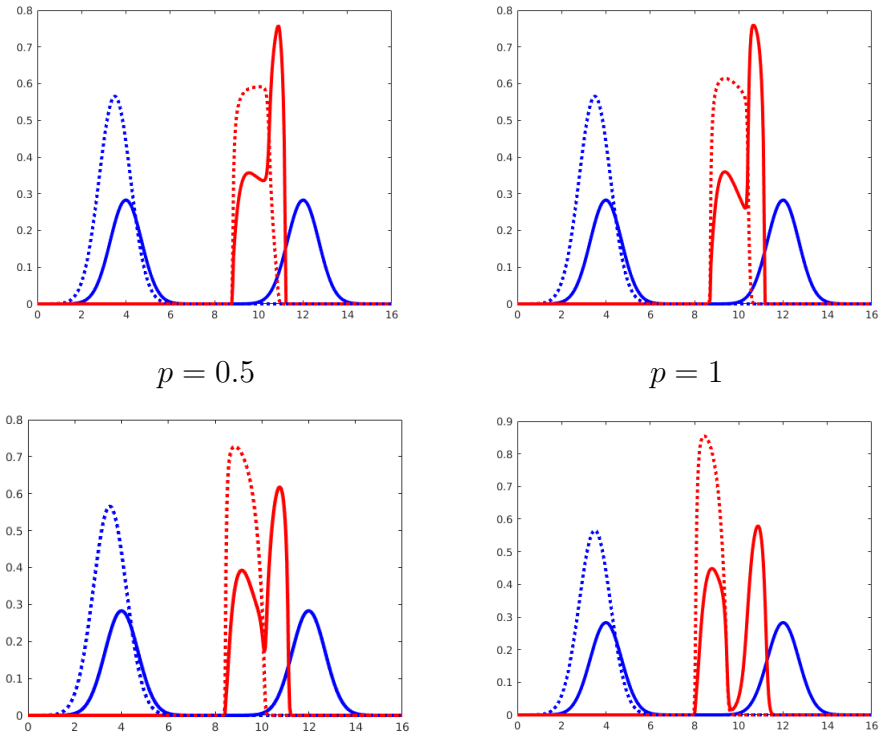
$$E_l(\nu^l) = \sum_{j\in J}(\nu^l_j)^8 + \sum_{k,j\in J\times J} \phi^l_{kj}\nu^l_j\nu^l_k + \sum_{j\in J} |y_j - 10|^4,$$

where $\phi^l_{kj} = 10^{-4}|y_k - y_j|^2$ and the total congestion F_j is given by

$$F_j(\nu^1_j + \nu^2_j) = (\nu^1_j + \nu^2_j)^4.$$

As usual, we consider cost functions of the form $c_{ij} = |x_i - y_j|^p$ and we want to analyze the support of ν^l as p varies. For the simulations in Figure 7 we have used $N = 500$ grid points discretization of $[0, 16]$ and treated the interaction

term with a semi-implicit approach. As we can notice in Figure 7 there is a *competition* between the confinement potential and the total congestion: the two populations tend to concentrated near $y = 10$ by the potential, but the effect of the congestion term makes it costly. This becomes clear if we compare (for instance, the case with $p = 2$) ν^1 with the optimal one in Figure 3; even if the energies are the same, the effect of congestion makes the support of the optimal solutions quite different.



$p = 0.5$ $p = 1$

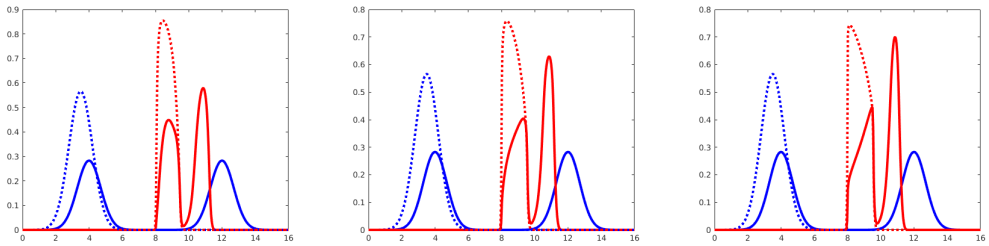
$p = 1.5$ $p = 2$

Figure 7: *The initial distributions μ^1 and μ^2 (blue solid line and blue dotted line) and the solutions ν^1 and ν^2 (red solid line and red dotted line) for different values of p.*

Thus, let us now consider the following case: let E_l be as above and $p = 2$, then we take the total congestion given by

$$F_j(\nu_j^1 + \nu_j^2) = (\nu_j^1 + \nu_j^2)^r$$

and we compute the optimal ν^l for different values of r. In Figure 8 we can see that the congestion term becomes more dominant as r increases so that



$r = 4$ $r = 8$ $r = 32$

Figure 8: *The initial distributions μ^1 and μ^2 (blue solid line and blue dotted line) and the solutions ν^1 and ν^2 (red solid line and red dotted line) for different values of r.*

the two populations try to be as far as possible, despite the effect of the confinement potential which is minimal at $y = 10$.

Acknowledgements: G.C. and L.N. gratefully acknowledge the support from the ANR, through the project ISOTACE (ANR-12- MONU-0013).

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* Lectures in Mathematics. Birkhäuser Boston, 2nd edition, 2004.

[2] R. Aumann. Existence of competitive equilibria in markets with a continuum of traders. *Econometrica*, 32:39–50, 1964.

[3] R. Aumann. Markets with a continuum of traders. *Econometrica*, 34:1–17, 1966.

[4] H. H. Bauschke and A. S. Lewis. Dykstra's algorithm with Bregman projections: a convergence proof. *Optimization*, 48(4):409–427, 2000.

[5] 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.

[6] Adrien Blanchet and Guillaume Carlier. From Nash to Cournot-Nash equilibria via the Monge-Kantorovich problem. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130398, 11, 2014.

[7] Adrien Blanchet and Guillaume Carlier. Optimal transport and Cournot-Nash equilibria. *Math. Oper. Res.*, 41(1):125–145, 2016.

[8] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced transport problems. Technical report, http://arxiv.org/abs/1607.05816, 2016.

[9] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NIPS) 26*, pages 2292–2300, 2013.

[10] R. L. Dykstra. An algorithm for restricted least squares regression. *J. Amer. Stat.*, 78(384):839–842, 1983.

[11] A. Galichon. *Optimal Transport Methods in Economics*. Princeton University Press, 1st edition, 2016.

[12] A. Galichon and B. Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. Technical report, Preprint SSRN-1487307, 2009.

[13] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[14] M. LeBreton and S. Weber. Games of social interactions with local and global externalities. *Econ. Letters*, 111:88–90, 2011.

[15] A. Mas-Colell. On a theorem of Schmeidler. *J. Math. Econ.*, 3:201–206, 1984.

[16] Gabriel Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM J. Imaging Sci.*, 8(4):2323–2351, 2015.

[17] 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.

[18] D. Schmeidler. Equilibrium points of nonatomic games. *J. Stat. Phys.*, 7:295–300, 1973.

[19] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, 2003.