



**HAL**  
open science

## A few models of vector-valued Optimal Transport

Xavier Bacon

► **To cite this version:**

Xavier Bacon. A few models of vector-valued Optimal Transport. Mathematics [math]. ED 386 École doctorale Paris-Centre, 2022. English. NNT: . tel-03913931

**HAL Id: tel-03913931**

**<https://hal.science/tel-03913931v1>**

Submitted on 28 Dec 2022

**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.

UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE

THÈSE DE DOCTORAT

---

Quelques modèles en Transport Optimal de  
mesures à valeurs vectorielles

(A few models of vector-valued Optimal Transport)

---

*Auteur :* Xavier Bacon

*Membres du jury :*

Pierre Cardaliaguet	Examineur
Patrick L. Combettes	Rapporteur
Olivier Guéant	Examineur
Bruno Nazaret	Directeur de thèse
Luca Nenna	Examineur
Filippo Santambrogio	Rapporteur

*Date de la soutenance :* 2 décembre 2022

Laboratoire : Statistiques Analyse et Modélisation Multidisciplinaire (SAMM)

École doctorale : ED 386 (École doctorale de Sciences Mathématiques de Paris Centre)

# Contents

<b>1</b>	<b>Introduction (en français)</b>	<b>5</b>
1.1	Notations . . . . .	5
1.2	Historique de la thèse . . . . .	5
1.3	Transport Optimal entre mesures de probabilité . . . . .	6
1.3.1	Monge et Kantorovich . . . . .	6
1.3.2	Dualité convexe . . . . .	9
1.3.3	Conditions d'optimalité . . . . .	12
1.3.4	Métrique de Wasserstein . . . . .	14
1.3.5	Formulation dynamique . . . . .	14
1.3.6	Transport optimal entropique . . . . .	16
<b>1</b>	<b>Introduction (in english)</b>	<b>21</b>
1.1	Notations . . . . .	21
1.2	History of the thesis . . . . .	21
1.3	Optimal Transport between probability measures . . . . .	22
1.3.1	Monge and Kantorovich . . . . .	22
1.3.2	Convex duality . . . . .	25
1.3.3	Optimality conditions . . . . .	28
1.3.4	Wasserstein metric . . . . .	29
1.3.5	Dynamic formulation . . . . .	30
1.3.6	Entropic Optimal Transport . . . . .	32
<b>2</b>	<b>Multi-species Optimal Transportation</b>	<b>36</b>
2.1	Survol . . . . .	36
2.1	Overview . . . . .	39
2.2	Multi-species Optimal Transportation, X. Bacon, Journal of Optimization Theory and Applications, 2020 . . . . .	43
2.3	Appendix : numerical results . . . . .	82
<b>3</b>	<b>A spatial Pareto exchange Economy Problem</b>	<b>87</b>
3.1	Survol . . . . .	87
3.1	Overview . . . . .	89
3.2	A spatial Pareto exchange economy problem, X. Bacon, G. Carlier, B. Nazaret, HAL, 2021 . . . . .	92

<b>4</b>	<b>Unbalanced Optimal Transport for positive semi-definite matrices</b>	<b>123</b>
4.1	Overview . . . . .	123
4.2	Notations . . . . .	125
4.3	Bures-Wasserstein metric . . . . .	128
4.3.1	Primal, dual and optimality conditions . . . . .	128
4.3.2	Examples of geodesics . . . . .	133
4.4	Dynamic model . . . . .	138
4.5	Appendix : a few words about Riemannian point of view of Bures-Wasserstein metric . . . . .	140

## Remerciements

Mes remerciements vont en premier lieu mon directeur de thèse Bruno Nazaret qui m'a accordé sa confiance pendant ces longues années que j'ai traversées non sans difficultés et qui en plus de ses savoirs a eu une grande capacité d'écoute.

Je remercie ensuite les cinq membres de mon jury en commençant par Pierre L. Combettes et Filippo Santambrogio qui ont accepté d'être les deux rapporteurs de mon manuscrit ainsi que Pierre Cardaliaguet, Olivier Guéant et Luca Nenna qui ont accepté d'être mes examinateurs.

Mes remerciements vont à Guillaume avec qui j'ai pris plaisir à travailler et auprès de qui j'ai énormément appris. Je remercie François-Xavier pour les nombreuses heures que nous avons passées à discuter ainsi que la montagne de notes qu'il a eu la gentillesse de me partager. La dernière partie de ce manuscrit lui est en grande partie redevable.

Je souhaite remercier Lilian qui m'a permis de mettre le pied à l'étrier des mathématiques numériques, ce qui n'était pas une mince affaire me concernant.

Pendant ma thèse j'ai eu l'occasion de travailler à de nombreux endroits. D'abord le SAMM. Ensuite le MOKAPLAN qui m'a permis de rencontrer la communauté francilienne du Transport Optimal et enfin le CEREMADE qui m'a offert la possibilité de terminer ma thèse dans les meilleures conditions. Je pense aussi naturellement aux collègues de Jussieu avec qui j'ai fait mes études. De peur d'en oublier, je ne prendrai pas le risque de me lancer dans une longue énumération de toutes celles et de tous ceux avec qui j'ai passé ces années, alors je dirai juste que je les remercie pour l'accueil qu'ils m'ont fait et pour les moments que nous avons passé ensemble. Je suis toujours agréablement surpris de la gentillesse des membres de la communauté mathématique et plus généralement de la recherche.

Je remercie ma mère pour le soutien infailible qui la caractérise.

Je dédie cette thèse à mon père qui en a vu le début mais pas la fin.

# 1 Introduction (en français)

## 1.1 Notations

- $\mathcal{P}(X)$  l'ensemble des mesures de probabilité sur  $X$ , celui-ci étant toujours métrique, la tribu considérée sera toujours la tribu borélienne.
- $\mathcal{P}_{ac}(X)$  l'ensemble des mesures de probabilité absolument continues par rapport à la mesure de Lebesgue sur  $X$  où  $X$  est un borélien de  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ .
- $\mathcal{P}_p(X)$  pour  $p \in [1, \infty)$  l'ensemble des mesures de probabilité sur  $X$ , admettant un moment d'ordre  $p$ .
- La mesure image  $f\#m$  de  $m$ , mesure sur  $(X, \mathcal{A})$ , par l'application mesurable  $f : X \rightarrow Y$  est la mesure sur  $(Y, \mathcal{B})$  donnée par :

$$\forall B \in \mathcal{B}, f\#m(B) \stackrel{\text{def}}{=} m(f^{-1}(B))$$

- $\mathbb{1}_n = [1, \dots, 1]^T \in \mathbb{R}^n$ ,  $n \in \mathbb{N}$ .
- $\text{supp}(m)$  avec  $m$  une mesure sur  $X$  est le support de  $m$ , défini comme l'intersection de tous les fermés de mesure pleine.
- $\iota$  désigne l'indicatrice convexe d'un ensemble : pour  $E$  un ensemble,  $A$  un sous-ensemble de  $E$  et  $x \in E$ , on pose  $\iota_A(x) = 0$  si  $x \in A$  et  $\iota_A(x) = +\infty$  sinon.

## 1.2 Historique de la thèse

La genèse de ce travail remonte à un mémoire de M2 effectué sous la direction de Bruno Nazaret. Il me proposa suite à celui-ci d'investiguer différentes extensions de la théorie du Transport Optimal qui traite le cas des mesures de probabilité (donc à valeurs dans  $\mathbb{R}_+$  et de sommes 1). Son idée était de proposer des modèles pour des mesures à valeurs dans des ensembles plus généraux :  $\mathbb{R}$  tout entier,  $\mathbb{R}_+^n$ , des espaces de matrices, des espaces de Banach etc.

**Cette thèse est constituée de trois parties indépendantes.**

La première partie présente un travail que j'ai effectué au début de ma thèse. Il s'agit d'un modèle de Transport Optimal entre mesures à valeurs dans  $\mathbb{R}_+^n$  et dont la somme des coordonnées est de masse 1. La synthèse de ce travail fut un papier publié dans le *Journal of Optimization Theory and Applications*. Une annexe (non publiée) présentant une méthode numérique y a été ajoutée. Celle-ci est accompagnée de simulations.

Bruno m'a ensuite rapproché de Guillaume Carlier, professeur à l'Université Paris-Dauphine, avec qui il avait déjà collaboré par le passé. À nous trois nous élaborâmes un modèle mathématique portée vers les Sciences Économiques qui consiste étant donné  $n$  biens distribués initialement dans l'espace à trouver autant de distributions finales qui maximisent la différence d'une utilité totale et d'un coût de transport. Cette collaboration fut pour moi l'occasion d'apprendre le langage Python et de me familiariser avec l'aspect numérique du Transport Optimal. Ce travail a débouché sur un article publié et fait l'objet de la seconde partie de ce manuscrit.

La dernière partie de la thèse est née d'une collaboration étroite avec François-Xavier Vialard, professeur à l'Université Gustave Eiffel, que j'ai rencontré au printemps 2019 à l'INRIA (Institut National de Recherche en Informatique et en Automatique) lors d'une école d'été consacrée au Transport Optimal Numérique où il enseignait. J'appris qu'il s'intéressait à un modèle de Transport Optimal pour les mesures à valeurs dans  $\mathcal{S}_+$  le cône des matrices symétriques et semi-positives. Ce modèle conjugua la récente théorie du Transport Optimal non-équilibré (*unbalanced* en anglais) avec la métrique de Bures-Wasserstein. Il me partagea de très nombreuses notes sur lesquelles nous nous mîmes à travailler de concert. Ce chapitre fera l'objet d'une prépublication. Enfin il est à noter que ces deux dernières années de nombreux articles sur le sujet ont été écrits. Nous revenons en introduction de chapitre sur les contributions apportées par ce manuscrit.

## 1.3 Transport Optimal entre mesures de probabilité

On présente la théorie du Transport Optimal entre mesures de probabilité, il s'agit là de décrire les principaux résultats dont nous aurons besoin dans les différents chapitres de la thèse. Cette section ne comporte aucun résultat inédit et le lecteur désireux d'approfondir le sujet est vivement invité à se reporter aux ouvrages [50], [56], [57] et [47] sur lesquels je me suis appuyé.

### 1.3.1 Monge et Kantorovich

La publication en 1781 par G. Monge de son *Mémoire sur la théorie des déblais et des remblais* [40] inaugure la théorie mathématique de ce que l'on appelle aujourd'hui le Transport Optimal. Il décrit en ces mots cette théorie : « *lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutume de donner le nom de Déblai au volume des terres que l'on doit transporter, & le nom de Remblai à l'espace qu'elles doivent occuper après le transport* ». Cela se traduit en langage mathématique moderne par la donnée de deux mesures de probabilité  $\mu, \nu \in \mathcal{P}(X)$  définies sur un espace métrique  $(X, d)$  muni de sa tribu borélienne :  $\mu$  est la mesure initiale (ou *Déblai*) et  $\nu$  la mesure finale (ou *Remblai*).

L'auteur s'intéresse ensuite au *prix* (comprendre prix minimal) à déboursier pour ef-

fectuer le transport de  $\mu$  jusqu'à  $\nu$  qu'il définit en ces termes : « *le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits des molécules multipliées chacune par l'espace parcouru, [...] & le prix du transport total sera un minimum.* ». Autrement dit, à chaque manière de transporter  $\mu$  vers  $\nu$  est associé un coût total qui se trouvera être la somme des coûts ponctuels qui eux-mêmes seront le produit de la masse déplacée (*i.e.*  $d\mu(x)$ ) par la distance séparant le point de départ  $x$  avec le point d'arrivée choisi  $T(x)$  (*i.e.*  $d(x, T(x))$ ). Ainsi ce qu'il nomme *prix du transport total* correspond au minimum de la fonctionnelle suivante :

$$M(T) \stackrel{\text{def}}{=} \int_X d(x, T(x)) d\mu(x) \quad (\text{Coût selon Monge})$$

parmi toutes les applications mesurables  $T : X \rightarrow X$  où pour assurer que le *déblai*  $\mu$  remplisse bien le *remblai*  $\nu$  on impose la contrainte  $T\#\mu = \nu$ .

On peut d'ores et déjà généraliser ce problème en différenciant l'espace  $X$  où vit  $\mu$  et l'espace  $Y$  où vit  $\nu$  et en remplaçant la distance  $d$  sur  $X$  par un coût général  $c = c(x, y) \in [0, \infty]$  défini sur le produit cartésien  $X \times Y$ . Un exemple important étudié ensuite est celui où  $c(x, y) = h(d(x, y))$  est fonction de la distance. Le problème (dit de Monge) est alors donné par :

**Problème 1** (Problème de Monge). Étant donné  $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$  et un coût  $c : X \times Y \mapsto [0, \infty]$  mesurable, le problème de Monge de valeur  $\mathcal{M}(\mu, \nu)$  est donné par la minimisation suivante :

$$\begin{aligned} \mathcal{M}(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{M(T) : T\#\mu = \nu\} \\ &= \inf \left\{ \int_X c(x, T(x)) d\mu(x) : T\#\mu = \nu \right\}. \end{aligned} \quad (1.1)$$

- Remarque 1.**
1. L'application  $T$  nous décrit le transport de  $\mu$  vers  $\nu$ . Il est important de noter que la totalité de la masse  $d\mu(x)$  se trouvant en  $x$  est transportée vers la destination  $T(x)$  : la formulation de Monge ne permet pas de diviser ponctuellement la masse pour par exemple la transporter vers deux destinations différentes.
  2. Monge parlait de minimum pour définir son prix. On sait au moins depuis les travaux de Weierstraß qu'il faut plutôt parler de borne inférieure. Montrer que ces deux notions sont confondues motive ce qui suit.
  3. La contrainte  $T\#\mu = \nu$  n'est pas linéaire (en  $T$ ). On peut le voir autrement en se plaçant dans un cadre euclidien et régulier ( $T$  lisse et injective par exemple) et



en supposant que  $\mu$  et  $\nu$  sont toutes les deux absolument continues par rapport à la mesure de Lebesgue (de densités respectives notées abusivement  $\mu(x)$  et  $\nu(y)$ ). La formule du changement de variable Jacobien nous assure que  $T\#\mu = \nu$  est équivalent à l'équation aux dérivées partielles :

$$\mu(x) = \nu(T(x))|\det(DT(x))|$$

qui n'est clairement pas linéaire en  $T$ . Comme conséquence, la méthode directe du calcul des variations due justement à Weierstraß échoue à démontrer l'existence d'un minimiseur au problème (1.1). Pour plus de détails, voir [50] (1.1 et Exercice 1).

En 1942, L. Kantorovich propose dans son article *On the Translocation of Masses* [31] une autre approche du problème de Transport Optimal où cette fois-ci le volume  $d\mu(x)$  peut être transporté vers plusieurs localités. Sa modélisation consiste à décrire le transport par une mesure de probabilité  $\gamma \in \mathcal{P}(X \times Y)$  dont les marginales sont données par  $\mu$  et  $\nu$ . Il appelle une telle mesure un *plan de transport* et propose le problème suivant :

**Problème 2** (Problème primal de Kantorovich). Étant donné  $\mu \in \mathcal{P}(X), \nu \in \mathcal{P}(Y)$  et un coût  $c : X \times Y \mapsto [0, \infty]$  mesurable, une mesure de probabilité sur  $X \times Y$ , disons  $\gamma$ , est appelée plan de transport entre  $\mu$  et  $\nu$  si elle vérifie à la fois  $\pi_1\#\gamma = \mu$  et  $\pi_2\#\gamma = \nu$ , où on a noté  $\pi_1(x, y) = x$  et  $\pi_2(x, y) = y$ . L'ensemble de ces plans de transport est noté  $\Pi(\mu, \nu)$ . On interprète  $\gamma(x, y)$  comme le volume acheminé entre  $x$  et  $y$ .

Le problème de Kantorovich de valeur  $\mathcal{K}(\mu, \nu)$  est alors donné par la minimisation suivante :

$$\begin{aligned} \mathcal{K}(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{K(\gamma) : \gamma \in \Pi(\mu, \nu)\} \\ &= \inf \left\{ \iint_{X \times Y} c(x, y) d\gamma(x, y) : \pi_1\#\gamma = \mu, \pi_2\#\gamma = \nu \right\}. \end{aligned} \quad (1.2)$$

**Remarque 2.** 1. Les contraintes sur les marginales  $\pi_1\#\gamma = \mu$  et  $\pi_2\#\gamma = \nu$  admettent chacune une écriture duale en termes de fonctions tests qui est donnée par :

$$\pi_1\#\gamma = \mu \iff \forall \varphi \in L^1(\mu), \int_X \varphi(x) d\mu(x) = \iint_{X \times Y} \varphi(x) d\gamma(x, y) \quad (1.3)$$

$$\pi_2\#\gamma = \nu \iff \forall \psi \in L^1(\nu), \int_Y \psi(y) d\nu(y) = \iint_{X \times Y} \psi(y) d\gamma(x, y) \quad (1.4)$$

et on peut se restreindre aux espaces des fonctions continues et bornées  $C_b(X)$  et  $C_b(Y)$  quand les espaces  $X$  et  $Y$  sont supposés Polonais (*i.e.* séparables et dont les

topologies sont induites par des distances les rendant complets).

2. Le problème primal de Kantorovich se trouve être une relaxation linéaire de celui de Monge : si une fonction  $T : X \rightarrow Y$  vérifie  $T\#\mu = \nu$  alors  $\gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\#\mu \in \Pi(\mu, \nu)$  et les coûts totaux dans (1.1) et (1.2) sont égaux. On en déduit que  $\mathcal{K}(\mu, \nu) \leq \mathcal{M}(\mu, \nu)$ . Nous aborderons les cas d'égalité plus tard.
3. Le problème primal de Kantorovich est un problème de programmation linéaire au sens où le coût total est linéaire et continue et les contraintes sont des polytopes (*i.e.* des intersections en nombre fini de sous-espaces affines).

Une première conséquence du précédent point est que la Méthode directe du Calcul des variations s'applique et permet d'établir l'existence d'un minimiseur au problème (1.2) sous réserve de quelques hypothèses sur les espaces  $X$  et  $Y$  ainsi que sur le coût  $c$ . C'est l'objet du théorème suivant.

**Théorème 3** (Theorem 1.7, [50]). Soient  $X$  et  $Y$  deux espaces Polonais et  $c : X \times Y \rightarrow [0, \infty]$  un coût semi-continu inférieurement. Sous ces conditions, le problème (1.2) admet une solution. Autrement dit, il existe  $\gamma \in \Pi(\mu, \nu)$  tel que

$$K(\gamma) = \mathcal{K}(\mu, \nu).$$

### 1.3.2 Dualité convexe

Le problème de Kantorovich (1.2) étant linéaire, donc convexe, nous pouvons chercher le problème dual qui lui est associé. En plus des références sur le Transport Optimal citées plus haut, on pourra se reporter aux ouvrages [1] et [49] pour une présentation synthétique et complète des techniques de dualité en Analyse Convexe.

Remarquons que le vocable du Transport Optimal appelle « Problème dual » ce que l'Analyse Convexe appelle « Problème primal » et vice versa. Dans ce manuscrit on adoptera les usages du Transport Optimal : les variables du problème primal appartiennent à un espace qui est le dual de l'espace auquel appartiennent les variables du problème dual.

On introduit les multiplicateurs de Lagrange  $\varphi$  et  $\psi$  appartenant respectivement à

$C_b(X, \mathbb{R})$  et  $C_b(Y, \mathbb{R})$ . On *dualise* le problème (1.2) grâce aux équations (1.3) et (1.4) :

$$\begin{aligned}
 \mathcal{K}(\mu, \nu) &= \inf \left\{ \iint_{X \times Y} c(x, y) \, d\gamma(x, y) : \pi_{1\#}\gamma = \mu, \pi_{2\#}\gamma = \nu \right\} \\
 &= \inf_{\gamma \geq 0} \sup_{\varphi, \psi} \left( \iint c \, d\gamma - \int \varphi \, d[\pi_{1\#}\gamma - \mu] + \int \psi \, d[\pi_{2\#}\gamma - \nu] \right) \\
 &\geq \sup_{\varphi, \psi} \left( \int \varphi \, d\mu + \int \psi \, d\nu + \inf_{\gamma \geq 0} \iint [c(x, y) - \varphi(x) - \psi(y)] \, d\gamma(x, y) \right) \\
 &= \sup_{\varphi, \psi} \left( \int \varphi \, d\mu + \int \psi \, d\nu - \iota_{\geq 0}(c - \varphi \oplus \psi) \right) \\
 &= \sup_{\varphi \oplus \psi \leq c} \int_X \varphi \, d\mu + \int_Y \psi \, d\nu
 \end{aligned}$$

où on a noté  $\iota_{\geq 0} = \iota_{\geq 0}(\theta)$  la fonction qui vaut 0 si  $\theta = \theta(x, y)$  est positif en tout  $(x, y)$  et  $+\infty$  sinon. Les calculs précédents nous suggèrent de définir le problème suivant :

**Problème 3** (Problème dual de Kantorovich). Étant donné  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  et un coût  $c : X \times Y \mapsto [0, \infty]$  mesurable, le problème dual de Kantorovich de valeur  $\mathcal{D}(\mu, \nu)$  est donné par la maximisation suivante :

$$\begin{aligned}
 \mathcal{D}(\mu, \nu) &\stackrel{\text{def}}{=} \sup \{ \mathcal{D}(\varphi, \psi) : \varphi \oplus \psi \leq c \} \\
 &= \sup \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \forall (x, y) \in X \times Y, \varphi(x) + \psi(y) \leq c(x, y) \right\},
 \end{aligned} \tag{1.5}$$

où la borne supérieure est prise sur les  $\varphi$  et  $\psi$  continues et bornées.

L'inégalité  $\mathcal{D}(\mu, \nu) \leq \mathcal{K}(\mu, \nu)$  est en réalité souvent une égalité (il n'y a pas de saut de dualité) comme l'affirme le théorème suivant :

**Théorème 4** (Dualité forte, Theorem 1.3, [56]). Sous les mêmes conditions que (1.2) et (1.5), on a l'égalité :

$$\mathcal{K}(\mu, \nu) = \mathcal{D}(\mu, \nu).$$

Si on suppose de plus que le coût  $c$  est uniformément continu et borné (par exemple si  $X$  et  $Y$  sont compacts) alors le problème (1.5) est atteint, autrement dit il existe  $(\varphi, \psi)$  continues et bornées telles que  $\mathcal{D}(\mu, \nu) = \mathcal{D}(\varphi, \psi)$ .

**Remarque 5.** 1. Le problème dual permet de *certifier* un hypothétique minimiseur du problème primal : si on estime que  $\gamma \in \Pi(\mu, \nu)$  est un bon candidat pour être solution dans (1.2) alors il suffit de trouver un couple  $(\varphi, \psi)$  tel que  $\mathcal{D}(\varphi, \psi) = \mathcal{K}(\mu, \nu)$  pour en déduire que  $\gamma$  est effectivement un minimiseur du problème primal. De

plus,  $(\varphi, \psi)$  sera alors un maximiseur du problème dual puisque l'égalité  $K(\gamma) \geq D(\varphi, \psi)$  est toujours vérifiée.

2. On peut donner une interprétation monétaire de  $\varphi$  et  $\psi$  : Alice est la PDG d'une grande multinationale ayant des antennes partout en Europe. Elle souhaite opérer à une redistribution du matériel informatique parmi les différents locaux de ses entreprises. Pour cela elle fait appel à deux sociétés de livraison appartenant respectivement à Bob et Carole qui lui proposent chacune un contrat différent :

(Bob) « Mes camions effectueront le transport selon le plan  $\gamma$  et vous devrez payer le montant  $K(\gamma)$ . **Libre à vous** de me donner le  $\gamma$  que vous préférez. »

(Carole) « Mon entreprise se propose de ne pas facturer le trajet de mes camions mais par contre de vous faire payer un prix  $\varphi(x)$  pour le chargement effectué en  $x$  et d'ajouter à cela un prix  $\psi(y)$  pour le déchargement effectué en  $y$ . Pour vous rassurer je vous garantis que  $\varphi(x) + \psi(y)$  sera toujours inférieur à  $c(x, y)$ . **Libre à moi** de fixer les prix  $\varphi$  et  $\psi$ . »

Si Alice et Carole sont omniscientes, le théorème précédent garantit à Alice que les deux contrats lui coûteront le même montant. Ainsi  $\varphi$  et  $\psi$  peuvent être interprétés comme des prix d'équilibre.

En regardant attentivement la fonctionnelle dans (1.5) on remarque que si  $(\gamma, \varphi, \psi)$  sont optimaux pour leurs problèmes respectifs alors la contrainte  $\varphi(x) + \psi(y) \leq c(x, y)$  doit être saturée en tout point  $(x, y)$  du support de  $\gamma$ . Un procédé simple pour ce faire est, partant de  $\varphi$  et  $\psi$  vérifiant l'inégalité ci-dessus, de remplacer  $\varphi$  par la  $c$ -transformée de  $\psi$  que nous définissons maintenant :

**Définition 6** ( $c$ -transformée et  $\bar{c}$ -transformée). Étant donné une fonction  $f : X \rightarrow \mathbb{R}$ , sa  $c$ -transformée, notée  $f^c$  et définie sur  $Y$  est donnée par :

$$f^c(y) = \inf_{x \in X} c(x, y) - f(x)$$

De la même manière on définit pour une fonction  $g : Y \rightarrow \mathbb{R}$ , sa  $\bar{c}$ -transformée, notée  $f^{\bar{c}}$  et définie sur  $X$  par :

$$g^{\bar{c}}(x) = \inf_{y \in Y} c(x, y) - g(y)$$

Un premier intérêt de cette notion est donc qu'étant donné  $(\varphi, \psi)$  admissible dans le problème (1.5), les couples  $(\varphi, \varphi^{\bar{c}})$  et  $(\varphi^{\bar{c}, c}, \varphi^{\bar{c}})$  sont aussi admissibles et améliorent le

coût dual au sens où

$$D(\varphi, \psi) \leq D(\varphi, \varphi^{\bar{c}}) \leq D(\varphi^{\bar{c},c}, \varphi^{\bar{c}}).$$

Malheureusement on ne peut répéter ce procédé indéfiniment comme l'affirme la proposition suivante :

**Proposition 7** (Proposition 1.34, [50]). Avec les mêmes notations que précédemment, si on suppose de plus que  $c$  est à valeurs réelles, on a les égalités suivantes :

$$f^{c,\bar{c},c} = f^c \text{ et } g^{\bar{c},c,\bar{c}} = g^{\bar{c}}$$

Le problème (1.5) peut par conséquent être écrit comme un problème de maximisation ne dépendant que de  $\varphi$  :

**Problème 4** (Problème dual, bis). Étant donné  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  et un coût  $c : X \times Y \mapsto [0, \infty]$  mesurable, le problème (1.5) admet l'écriture suivante :

$$\mathcal{D}(\mu, \nu) = \sup \left\{ \int \varphi \, d\mu + \int \varphi^c \, d\nu : \varphi \text{ continue et bornée} \right\}. \quad (1.6)$$

où la borne supérieure est prise sur les  $\varphi$  continues et bornées.

**Remarque 8.** La  $c$ -transformée (resp. la  $\bar{c}$ -transformée) étant définies comme la borne inférieure d'une famille de fonctions indexées par  $x$  (resp. par  $y$ ), si cette famille admet un module de continuité commun alors il est en même de la  $c$ -transformée (resp. de la  $\bar{c}$ -transformée). En particulier si  $c$  est supposé Lipschitzien, alors  $f^c$  et  $g^{\bar{c}}$  le seront aussi.

### 1.3.3 Conditions d'optimalité

Désormais on se demande quand est-ce que le problème de Monge (1.1) et celui de Kantorovich (1.2) ont même valeur. Pour répondre à cette question on va chercher des conditions sur  $X, Y, c$  ainsi que  $\mu$  qui nous permettent d'exhiber un plan de transport  $\gamma$  optimal dans (1.2) et qui s'écrit sous la forme  $\gamma = \gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\# \mu$  avec  $T : X \rightarrow Y$  mesurable.

Commençons par nous placer dans les conditions du Théorème 4 et ainsi avoir  $\mathcal{K}(\mu, \nu) = \mathcal{D}(\mu, \nu)$ . On pose  $X = Y$  et on se donne  $\gamma$  et  $\varphi$  optimaux de telle sorte que l'égalité  $K(\gamma) = D(\varphi, \varphi^c)$  implique

$$\begin{aligned} \forall x, y \in X \times X, \varphi(x) + \varphi^c(y) &\leq c(x, y) \\ \forall x, y \in \text{supp}(\gamma), \varphi(x) + \varphi^c(y) &= c(x, y) \end{aligned}$$

où  $\text{supp}(\gamma)$  désigne le support de  $\gamma$  défini comme l'intersection des fermés de mesure pleine. Fixons  $(x_0, y_0) \in \text{supp}(\gamma)$  un couple de points où il y égalité et regardons la fonction  $x \rightarrow \varphi(x) - c(x, y_0)$  qui atteint son minimum en  $x_0$  et est continue dès lors que  $c$  l'est. Si  $x_0$  se trouve à l'intérieur de  $X$  et qu'à la fois  $\varphi$  et  $x \rightarrow c(x, y_0)$  sont différentiables en  $x_0$  alors la condition d'optimalité du premier ordre implique que

$$\nabla\varphi(x_0) = \nabla_x c(x_0, y_0). \quad (1.7)$$

Si on arrive à écrire l'équation précédente comme une équation donnant  $y_0$  en fonction de  $x_0$  ou autrement dit que le point de départ détermine complètement la destination alors on pourra en déduire qu'une solution du problème de Monge est solution du problème de Kantorovich. Une condition suffisante pour cela est de regarder des coûts sous la forme  $c(x, y) = h(x - y)$  avec  $h$  strictement convexe et différentiable (par exemple, si  $X = \mathbb{R}^d$ ,  $h(x) = |x|^p$ ,  $p > 1$ ) puisque ainsi l'équation (1.7) devient

$$\varphi(x_0) = \nabla h(x_0 - y_0)$$

et la différentiabilité de  $h$  associée à sa stricte convexité assurent que  $\nabla h$  est inversible (pour plus de détails, voir [50], 1.3). Après calculs on obtient

$$y_0 = x_0 + (\nabla h)^{-1}(\nabla\varphi(x_0)).$$

En posant  $T(x) = x + (\nabla h)^{-1}(\nabla\varphi(x))$  et  $\gamma = \gamma_T$ , on en déduit que  $\gamma_T$  est optimal pour (1.2) et par conséquent que  $T$  est optimal pour (1.1).

Il reste à trouver un cadre où tous les calculs précédents permettent de définir une application de  $X$  dans  $X$ . Le cadre classique est de prendre  $X = Y = \Omega$  un compact de  $\mathbb{R}^d$  puisque comme  $c$  est localement Lipschitzien par convexité de  $h$ ,  $c$  est alors globalement Lipschitzien par compacité de  $\Omega$  et il en est de même de  $\varphi$  par la Remarque 8. Le théorème de Rademacher nous assure ensuite que  $\varphi$  est différentiable presque partout. Enfin si  $\mu$  est absolument continue par rapport à la mesure de Lebesgue, les points de non différentiabilité de  $\varphi$  seront négligeables pour  $\mu$  et les calculs précédents seront valables presque partout. Pour résumer on a le résultat suivant :

**Théorème 9** ([50], Theorem 1.17). Soient  $\Omega$  un compact de  $\mathbb{R}^d$  et  $\mu, \nu \in \mathcal{P}(\Omega)$ . Si  $c(x, y) = h(x - y)$  où  $h$  est une fonction strictement convexe alors les problèmes (1.2) et (1.6) sont atteints. Si de plus  $\mu$  est absolument continue et  $\mu(\partial\Omega) = 0$  alors on a unicité de la solution dans (1.2) et celle-ci est de la forme  $\gamma = \gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\#\mu$  où  $T$  est donnée par

$$T(x) = x + (\nabla h)^{-1}(\nabla\varphi(x))$$

où on a noté  $\varphi$  une solution de (1.6). De plus  $T$  est optimal dans (1.1).

Terminons en évoquant un théorème spécifique au cas où  $c(x, y) = |x - y|^2/2$  qui est dû à Brenier dans son article [8].

**Théorème 10** (Théorème de Brenier). Étant donné deux mesures  $\mu, \nu \in \mathbb{R}^d$  avec  $\mu$  absolument continue. Il existe une unique application  $T$  vérifiant  $T\#\mu = \nu$  et de plus elle est de la forme  $T = \nabla u$  avec  $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  une fonction convexe.

De plus, en prenant  $c(x, y) = \frac{1}{2}|x - y|^2$  cette application  $T$  est solution de (1.1) et  $\gamma_T$  est solution de (1.2).

### 1.3.4 Métrique de Wasserstein

Si on se donne une distance  $d$  sur un espace Polonais<sup>1</sup>  $X$  et que l'on pose  $c = d$ , il est facile de voir que l'ensemble des mesures de Dirac sur  $X$  muni du coût  $K(\cdot, \cdot)$  est un espace métrique isométrique à  $(X, d)$ . On peut légitimement espérer qu'un tel résultat s'étende aux mesures de probabilité et permette ainsi d'étendre la métrique sur  $X$  à une métrique sur  $\mathcal{P}(X)$ . C'est bien le cas comme le dit le théorème suivant :

**Théorème 11.** Soient  $(X, d)$  un espace Polonais et  $p \in [1, \infty)$ . En considérant le coût  $c(x, y) = d(x, y)^p$ , on définit une distance  $W_p$  sur  $\mathcal{P}_p(X)$  l'ensemble des mesures de probabilité sur  $X$  admettant un moment d'ordre  $p$  en posant

$$W_p(\mu, \nu) = \mathcal{K}^{1/p}(\mu, \nu).$$

Cette distance est appelée distance de Wasserstein d'ordre  $p$ .

Les principales propriétés de cette distance sont réunies dans la proposition qui suit.

**Proposition 12** ([57], Theorems 6.9 and 6.18). [1] Pour  $p \in [1, \infty)$ , la distance  $W_p$  métrise la convergence faible- $*$  des mesures de probabilité.

[2] Si  $(X, d)$  est Polonais, alors  $(\mathcal{P}_p(X), W_p)$  aussi.

Remarquons que la convergence faible- $*$  se lit sur les mesures de Dirac ainsi :  $\delta_x \rightarrow \delta_y$  si et seulement si  $x \rightarrow y$ . Ceci qui n'est pas étonnant puisque compte tenu du commentaire précédent les mesures de Dirac peuvent être vues comme un plongement de  $X$  dans  $\mathcal{P}(X)$ .

### 1.3.5 Formulation dynamique

Arrêtons nous au cas où  $X = Y = \mathbb{R}^d$  et  $c(x, y) = |x - y|^p/p$  avec  $p \in [1, \infty)$ . Dans leur article [3], Benamou et Brenier ont proposé une nouvelle formulation du Transport Optimal basée sur la mécanique des fluides et introduisant une variable de temps. On parle

---

<sup>1</sup>Autrement dit un espace métrique complet et séparable.

alors de Formulation Dynamique du Transport Optimal. D'un point de vue métrique, ils ont établi que sous certaines conditions, la métrique de Wasserstein est géodésique.

Partant au temps  $t = 0$  d'une mesure à densité  $\mu$  sur  $\mathbb{R}^d$ , les auteurs s'intéressent à toutes les cinématiques qui mènent au temps  $t = 1$  jusqu'à  $\nu$  une autre mesure à densité sur  $\mathbb{R}^d$  et ce de telle sorte que la masse totale soit conservée au cours du temps. Mathématiquement, cela revient à s'intéresser à l'ensemble des chemins  $\rho : t \rightarrow \rho_t \in \mathcal{P}_{ac}(\mathbb{R}^d)$  continus (pour la convergence faible-\* des mesures) et reliant  $\mu$  et  $\nu$  (i.e.  $\rho_0 = \mu$  et  $\rho_1 = \nu$ ) soumis au champ de vitesses  $\mathbf{v} = \mathbf{v}_t(x)$  et vérifiant l'équation dite de continuité (ou de conversation de la masse) :

$$\partial_t \rho_t(x) + \operatorname{div}_x [\rho_t(x) \mathbf{v}_t(x)] = 0. \quad (1.8)$$

Le coût total est désormais donné par l'action  $\mathcal{A}_p(\rho, \mathbf{v})$  de cette cinématique (correspondant à la somme des énergies cinétiques quand  $p = 2$ ) définie par

$$\mathcal{A}_p(\rho, \mathbf{v}) = \int_0^1 \int_X \frac{|\mathbf{v}_t(x)|^p}{p} d\rho_t(x) dt.$$

Malheureusement cette fonctionnelle n'est pas convexe en  $(\rho, \mathbf{v})$  et l'astuce pour la rendre convexe est d'effectuer le changement de variable  $\mathbf{v} \leftarrow \mathbf{m} = \rho \mathbf{v}$  pour remplacer la vitesse par la quantité de mouvement (*momentum* en anglais). Le problème est alors donné par :

**Problème 5** (Problème de Benamou-Brenier). Étant donné  $\mu \in \mathcal{P}_{ac}(X), \nu \in \mathcal{P}_{ac}(X)$  et  $p \in [1, \infty)$ , le problème de Benamou-Brenier de valeur  $\mathcal{B}_p(\mu, \nu)$  est donné par la minimisation suivante :

$$\mathcal{B}_p(\mu, \nu) \stackrel{\text{def}}{=} \inf \{ \mathcal{A}(\rho, \mathbf{m}) : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \} \quad (1.9)$$

$$= \inf \left\{ \frac{1}{p} \int_0^1 \int_X \frac{|\mathbf{m}_t(x)|^p}{\rho_t(x)^{p-2}} d\rho_t(x) dt : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \right\} \quad (1.10)$$

Cette formulation étend celle Kantorovich et de Monge en un sens donné par le théorème suivant :

**Théorème 13** ([50], Theorem 5.27). Soient  $X$  un convexe de  $\mathbb{R}^d$  et  $p \in [1, \infty)$ . On se donne  $\mu, \nu \in \mathcal{P}_p(X)$ , alors :

$$W_p(\mu, \nu) = \mathcal{B}_p(\mu, \nu).$$

De plus la borne inférieure est atteinte pour  $\mathcal{B}_p$  dès lors que le problème de Monge associé admet une solution  $T$  et on a alors pour solution  $\rho_t = [(1-t)\operatorname{Id} + tT] \# \mu$ . C'est



donc notamment le cas quand  $\mu$  est absolument continue. Dit autrement,  $(\mathcal{P}_p(X), W_p)$  est un espace géodésique dont les géodésiques sont les solutions du problème (1.9).

Puisque que nous avons affaire un problème de minimisation convexe, on peut *dualiser* ce problème. Faisons le pour  $p = 2$ .

$$\begin{aligned} \mathcal{B}_p(\mu, \nu) &= \inf \{ \mathcal{A}(\rho, \mathbf{m}) : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \} \\ &= \inf_{(\rho, \mathbf{m})} \mathcal{A}(\rho, \mathbf{m}) + \sup_{\phi} \left[ \int_0^1 \int_X ((\partial_t \phi) \rho + \nabla \phi \cdot \mathbf{m}) + \Delta_{\mu}^{\nu}(\phi) \right] \end{aligned}$$

où  $\Delta_{\mu}^{\nu}(\phi) \stackrel{\text{def}}{=} \int_X \phi(1, \cdot) d\nu - \int_X \phi(0, \cdot) d\mu$  correspond aux conditions initiale et finale. On intervertit ensuite la borne inférieure et la borne supérieure tout en remarquant que  $\frac{|\mathbf{m}|^2}{2\rho} - \nabla \phi \cdot \mathbf{m} = \left( \frac{\mathbf{m}}{2\rho} - \nabla \phi \right) \cdot \mathbf{m}$  et donc que minimiser en  $\mathbf{m}$  revient à imposer  $\mathbf{m} = \rho \nabla \phi$  :

$$\begin{aligned} \mathcal{B}_p(\mu, \nu) &\leq \sup_{\phi} \Delta_{\mu}^{\nu}(\phi) + \inf_{\phi \geq 0} \int_0^1 \int_X \left[ -(\partial_t \phi) \rho - \frac{1}{2} |\nabla \phi|^2 \rho \right] \\ &= \sup \left\{ \Delta_{\mu}^{\nu}(\phi) : \partial_t \phi + \frac{1}{2} |\nabla \phi|^2 \leq 0 \right\} \end{aligned}$$

où la contrainte sur  $\phi$  est un cas particulier d'inéquation d'Hamilton-Jacobi. Les conditions d'optimalité primal-dual imposent notamment d'avoir égalité sur le support de la géodésique optimale  $\rho$ .

### 1.3.6 Transport optimal entropique

On présente maintenant une méthode numérique pour déterminer à la fois une approximation du plan de transport optimal et une approximation des potentiels optimaux qui lui sont associés. Connue sous le nom de *Régularisation Entropique du Transport Optimal* (voir l'article de Cuturi [22]). Les quelques notes qui suivent sont entièrement tirées de [47].

On se place dans un cadre discret : étant donné  $n, m \in \mathbb{N}^*$  on se donne deux combinaisons convexes de mesures de Dirac

$$\begin{aligned} \mu &\stackrel{\text{def}}{=} \sum_{i=1}^n \mu_i \delta_{x_i} \\ \nu &\stackrel{\text{def}}{=} \sum_{j=1}^m \nu_j \delta_{y_j} \end{aligned}$$

avec  $\sum_i \mu_i = \sum_j \nu_j = 1$ . Par abus de notation on écrira  $\mu = (\mu_i) \in \mathbb{R}^n$  et  $\nu = (\nu_j) \in \mathbb{R}^m$ . On implémente le coût et les plans de transport dans des matrices rectangulaires  $C = C_{i,j} \stackrel{\text{def}}{=} c(x_i, y_j) \in \mathbb{R}^{n \times m}$  et  $\gamma = \gamma_{ij} \stackrel{\text{def}}{=} \gamma(x_i, y_j) \in \mathbb{R}^{n \times m}$ . Le problème (1.2) se réécrit alors (on a

noté  $\cdot$  le produit matriciel usuel) :

$$\min \{ \langle C | \gamma \rangle : \gamma \cdot \mathbb{1}_n = \mu, \gamma^T \cdot \mathbb{1}_m = \nu \}. \quad (1.11)$$

Partant du problème (1.11), on introduit une famille de nouveaux problèmes paramétrée par  $\varepsilon \in \mathbb{R}_+^*$  et donnée par :

$$\begin{aligned} \mathcal{K}_\varepsilon(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{ K_\varepsilon(\gamma) : \gamma \in \Pi(\mu, \nu) \} \\ &= \inf \{ K(\gamma) - \varepsilon H(\gamma) : \gamma \in \Pi(\mu, \nu) \} \end{aligned} \quad (1.12)$$

où  $H(\gamma) \stackrel{\text{def}}{=} - \sum_{i,j} \gamma_{ij} (\log(\gamma_{ij}) - 1)$  désigne l'entropie (concave) du plan de transport  $\gamma$ . Ces nouveaux problèmes sont  $\varepsilon$ -fortement convexes donc admettent une unique solution. De plus quand  $\varepsilon$  tend vers 0, les solutions de (1.12) convergent vers une solution de (1.11), plus précisément vers celle d'entropie maximale (voir [21] pour plus de détails).

Comme précédemment on introduit les multiplicateurs de Lagrange  $\varphi \in \mathbb{R}^n$  et  $\psi \in \mathbb{R}^m$  :

$$K_\varepsilon(\mu, \nu) = \min_{\gamma} \max_{\varphi, \psi} \langle \gamma | C \rangle - \varepsilon H(\gamma) - \langle \varphi | \gamma \cdot \mathbb{1}_n - \mu \rangle - \langle \psi | \gamma^T \cdot \mathbb{1}_m - \nu \rangle$$

et les conditions d'optimalité nous assurent que  $(\gamma, \varphi, \psi)$  sont optimaux si et seulement si

$$C_{ij} + \varepsilon \log(\gamma_{ij}) - \varphi_i - \psi_j = 0 \quad (1.13)$$

$$\gamma \cdot \mathbb{1}_n = \mu \quad (1.14)$$

$$\gamma^T \cdot \mathbb{1}_m = \nu \quad (1.15)$$

L'équation (1.13) se réécrit :

$$\forall (i, j) \in \llbracket 1, n \rrbracket \times \llbracket 1, m \rrbracket, \gamma_{ij} = u_i K_{ij} v_j$$

où on a posé  $u_i = \exp(\varphi_i/\varepsilon)$ ,  $v_j = \exp(\psi_j/\varepsilon)$  et  $K_{ij} = \exp(-C_{ij}/\varepsilon)$ . Cette dernière équation se réécrit de manière plus compacte :

$$\gamma = \text{diag}(u) K \text{diag}(v) \quad (1.16)$$

et par suite les conditions de marginales (1.14) et (1.15) deviennent (où on a noté  $\odot$  le

produit terme à terme) :

$$u \odot (Kv) = \mu, \quad (1.17)$$

$$v \odot (K^T u) = v. \quad (1.18)$$

Pour résoudre les problèmes (1.17) et (1.18), Cuturi utilise l'algorithme dit de Sinkhorn (voir [52], [25], [32]), qui, partant de  $u_0 = \mathbb{1}_n, v_0 = \mathbb{1}_m$ , consiste à construire de manière itérative  $u^{(l)}$  et  $v^{(l)}$  ( $l \in \mathbb{N}^*$ ) ainsi (où le symbole de fraction désigne une division terme à terme) :

$$u^{(l)} = \frac{\mu}{K \cdot v^{(l-1)}}, v^{(l)} = \frac{v}{K^T \cdot u^{(l)}} \quad (1.19)$$

Ceci revient à demander que  $u^{(l)}$  et  $v^{(l)}$  vérifient alternativement (1.17) et (1.18). On peut voir cette méthode comme une projection alternée sur la contrainte (1.17) puis sur la contrainte (1.18) qui est un cas particulier d'algorithme de Dykstra avec projections de Bregman (voir [2]). La convergence est assurée par le théorème suivant, dont une démonstration se trouve dans [25] ou [46].

**Théorème 14.** La suite définit en (1.19) converge linéairement vers une solution de (1.14) et (1.15).

**Remarque 15** (soft-min). D'un côté les itérations précédentes présentent l'avantage d'être très simples à implémenter (les suites  $u$  et  $v$  sont géométriques) mais d'un point de vue numérique les variables  $u$  ou  $v$  ont tendance à demander un encodage très important (phénomène d'*overflow*). En pratique cela se traduit par  $u$  qui diverge vers  $+\infty$  d'autant plus vite que  $v$  tend vers 0 par valeurs positives. Une manière de pallier ce problème est d'ajouter une étape dite de soft-min à chaque itération. En effet les itérations (1.17) et (1.18) peuvent se réécrire dans les variables  $(\varphi^{(l)}, \psi^{(l)})$  de la manière suivante :

$$\begin{aligned} \varphi^{(l)} &= \varepsilon \log(\mu) - \varepsilon \log \left[ K \exp \left( \frac{\psi^{(l-1)}}{\varepsilon} \right) \right] \\ \psi^{(l)} &= \varepsilon \log(v) - \varepsilon \log \left[ K^T \exp \left( \frac{\varphi^{(l)}}{\varepsilon} \right) \right] \end{aligned}$$

et en se rappelant que  $K = \exp(-C/\varepsilon)$  on peut réécrire ces deux équations à coordon-

---

nées  $i$  et  $j$  fixées :

$$\begin{aligned}\varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) - \varepsilon \log \left[ \sum_j \exp \left( \frac{-C[i, j] + \psi^{(l-1)}[j]}{\varepsilon} \right) \right] \\ \psi^{(l)}[j] &= \varepsilon \log(v[j]) - \varepsilon \log \left[ \sum_i \exp \left( \frac{-C[i, j] + \varphi^{(l)}[i]}{\varepsilon} \right) \right].\end{aligned}$$

En introduisant la notation soft-min, définie pour un vecteur  $\mathbf{z} = (z_i)$  de dimension quelconque :

$$\min_\varepsilon(\mathbf{z}) \stackrel{\text{def}}{=} -\varepsilon \log \left[ \sum_i \exp \left( \frac{-z_i}{\varepsilon} \right) \right],$$

on obtient

$$\begin{aligned}\varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) + \min_\varepsilon(C[i, \cdot] - \psi^{(l-1)}) \\ \psi^{(l)}[j] &= \varepsilon \log(v[j]) + \min_\varepsilon(C[\cdot, j] - \varphi^{(l)}).\end{aligned}$$

Une propriété simple mais fondamentale du soft-min est donnée par :

$$\forall t \in \mathbb{R}, \min_\varepsilon(\mathbf{z}) = t + \min_\varepsilon(\mathbf{z} - t).$$

Cela<sup>2</sup> a pour intérêt de soulager en partie le terme dans l'exponentielle en le transformant en une addition effectuée *a posteriori*, ce qui allège l'*overflow*. Développer ce dernier terme nous donne pour la première équation :

$$\min_\varepsilon(\mathbf{z} - t) = -\varepsilon \log \left[ \sum_i \left( \exp \frac{-C[i, j] + \psi^{(l-1)}[j] + t}{\varepsilon} \right) \right]$$

Et puisque l'on a le choix du  $t$ , autant en prendre un qui rende la somme aussi proche que l'on peut de 0. Le choix de la  $\bar{c}$ -transformée de  $\psi^{(l-1)}$  est alors tout trouvé. Le même raisonnement se tient évidemment aussi pour  $\psi^{(l)}[j]$ . Finalement les nouveaux itérés sont donnés par :

$$\begin{aligned}\varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) + \psi^{(l-1), \bar{c}}[i] + \min_\varepsilon(C[i, \cdot] - \psi^{(l-1)} - \psi^{(l-1), \bar{c}}[i]) \\ \psi^{(l)}[j] &= \varepsilon \log(v[j]) + \varphi^{(l), c}[j] + \min_\varepsilon(C[\cdot, j] - \varphi^{(l)} - \varphi^{(l), c}[j]).\end{aligned}$$

<sup>2</sup>Pour comprendre cette appellation de soft-min, on peut remarquer que  $\min_\varepsilon(\mathbf{z})$  tend vers  $\min_i z_i$  quand  $\varepsilon$  tend vers 0.

Notons que récemment, la méthode de Régularisation Entropique du Transport Optimal a été étendue à d'autres problèmes : Transport Optimal multi-marginales [42], non-équilibré (*unbalanced*) [51] ou flots de gradient [45].

# 1 Introduction (in english)

## 1.1 Notations

- $\mathcal{P}(X)$  denotes the set of probability measures on  $X$ .  $X$  is always endowed with a distance and its Borelian  $\sigma$ -field.
- $\mathcal{P}_{ac}(X)$  denotes the set of probability measures that admits a density with respect to the Lebesgue measure on  $X$ , where  $X$  is a Borelian set of  $\mathbb{R}^d$ ,  $d \in \mathbb{N}$ .
- $\mathcal{P}_p(X)$  with  $p \in [1, \infty)$  et set of probability measures on  $X$ , which admit a moment of order  $p$ .
- $f\#m$  denotes the pushforward measure of  $m$  by  $f$  where  $m$  is a measure on  $(X, \mathcal{A})$  and  $f$  a measurable map  $f : X \rightarrow Y$  and is defined by:

$$\forall B \in \mathcal{B}, f\#m(B) \stackrel{\text{def}}{=} m(f^{-1}(B))$$

- $\mathbb{1}_n = [1, \dots, 1]^T \in \mathbb{R}^n$ ,  $n \in \mathbb{N}$ .
- $\text{supp}(m)$  with  $m$  a measure on  $X$  is the support of  $m$ , defined as the intersection of all closed set, say  $F$ , such that  $m(F) = m(X)$ .
- $\iota$  denotes the convex indicatrix of a set: let  $E$  be a set,  $A$  a subset of  $E$  and  $x \in E$ , we set  $\iota_A(x) = 0$  if  $x \in A$  and  $\iota_A(x) = +\infty$  otherwise.

## 1.2 History of the thesis

The genesis of this work goes back to a master thesis done under the supervision of Bruno Nazaret. He proposed I investigate different extensions of the Optimal Transport theory which deals with the case of probability measures (thus with values in  $\mathbb{R}_+$  and sums 1). His idea was to propose models for measures with values in more general sets:  $\mathbb{R}$  as an integer,  $\mathbb{R}_+^n$ , spaces of matrices, Banach spaces, etc.

### **This thesis consists of three independent parts.**

The first part presents the work I did at the beginning of my thesis. It is a model of Optimal Transport between measures with values in  $\mathbb{R}_+^n$  and whose sum of coordinates is of mass 1. The synthesis of this work was a paper published in the *Journal of Optimization Theory and Applications*. An appendix (unpublished) presenting a numerical method has been added. It is accompanied by simulations.

Bruno then brought me together with Guillaume Carlier, professor at the University of Paris-Dauphine, with whom he had already collaborated in the past. Between the

three of us, we developed a mathematical model for economic sciences which consists, given  $n$  initial distributions of goods, in finding  $n$  final distributions which maximize the difference between a total utility and a transport cost. This collaboration was for me the opportunity to learn the Python language and to get familiar with the numerical aspect of Optimal Transport. This work led to an accepted for publication article and is the subject of the second part of this manuscript.

The last part of the thesis was born from a close collaboration with François-Xavier Vialard, professor at the Gustave Eiffel University, whom I met in the spring of 2019 at INRIA (Institut National de Recherche en Informatique et en Automatique) during a summer school dedicated to Digital Optimal Transport where he was teaching. I learned that he was interested in a model of Optimal Transport for measures with values in  $\mathcal{S}_+$  the cone of symmetric and semi-positive matrices. This model combined the recent theory of *unbalanced* Optimal Transport with the Bures-Wasserstein metric. He shared with me many notes on which we started to work together. A prepublication containing the content of this chapter will be made. Finally, it should be noted that many articles on the subject have been written in the last two years. We come back to the contributions made by this manuscript in the introduction of the chapter.

### 1.3 Optimal Transport between probability measures

The theory of Optimal Transport between probability measures is presented, it is a question of describing the main results that we will need in the different chapters of the thesis. This section does not contain any new results and the reader who wants to go deeper into the subject is strongly invited to refer to the books [50], [56], [57] and [47] on which I relied.

#### 1.3.1 Monge and Kantorovich

The publication in 1781 by G. Monge of his *Mémoire sur la théorie des déblais et des remblais* [40] inaugurates the mathematical theory of what we call today Optimal Transport. He describes this theory in these words: « *lorsqu'on doit transporter des terres d'un lieu dans un autre, on a coutume de donner le nom de Déblai au volume des terres que l'on doit transporter, & le nom de Remblai à l'espace qu'elles doivent occuper après le transport* », which is translated in modern mathematical language by giving two probability measures  $\mu, \nu \in \mathcal{P}(X)$  defined on a metric space  $(X, d)$  with its Borelian  $\sigma$ -algebra:  $\mu$  is the initial measure (or *Déblai*) and  $\nu$  the final measure (or *Remblai*).

The author then looks at the *price* (i.e. the minimum price) to be paid for transport from  $\mu$  to  $\nu$ , which he defines as follows: « *le prix du transport d'une molécule étant, toutes choses d'ailleurs égales, proportionnel à son poids & à l'espace qu'on lui fait parcourir, & par conséquent le prix du transport total devant être proportionnel à la somme des produits*

des molécules multipliées chacune par l'espace parcouru, [...] & le prix du transport total sera un minimum. », i.e. each way of transporting  $\mu$  to  $\nu$  is associated with a total cost which will be the sum of the punctual costs which themselves will be the product of the mass moved (i.e.  $d\mu(x)$ ) by the distance separating the point of departure  $x$  with the chosen point of arrival  $T(x)$  (i.e.  $d(x, T(x))$ ). Thus what he calls *prix du transport total* corresponds to the minimum of the following functional:

$$M(T) \stackrel{\text{def}}{=} \int_X d(x, T(x)) d\mu(x) \quad (\text{Monge cost})$$

subject to all measurable maps  $T : X \rightarrow X$  where to ensure that the *déblai*  $\mu$  fills the *remblai*  $\nu$  the constraint  $T\#\mu = \nu$  is imposed.

We can already generalize this problem by differentiating the space  $X$  where  $\mu$  lives and the space  $Y$  where  $\nu$  lives and by replacing the distance  $d$  on  $X$  by a general cost  $c = c(x, y) \in [0, \infty]$  defined on the Cartesian product  $X \times Y$ . An important example studied afterwards is the one where  $c(x, y) = h(d(x, y))$  is a function of the distance. The problem (called Monge problem) is then given by :

**Problem 1** (Monge Problem). Given  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  and a cost  $c : X \times Y \mapsto [0, \infty]$  measurable, Monge problem with value  $\mathcal{M}(\mu, \nu)$  is given by the following minimization:

$$\begin{aligned} \mathcal{M}(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{M(T) : T\#\mu = \nu\} \\ &= \inf \left\{ \int_X c(x, T(x)) d\mu(x) : T\#\mu = \nu \right\}. \end{aligned} \quad (1.1)$$

**Remark 1.** 1. The map  $T$  describes the transport of  $\mu$  to  $\nu$ . It is important to note that the totality of the mass  $d\mu(x)$  in  $x$  is transported to the destination  $T(x)$  : Monge formulation does not allow the mass to be divided punctually, for example to transport it to two different destinations.

2. Monge spoke of a minimum to define his price. We know at least since the work of Weierstraß that we should rather speak of a infimum. To establish that these two notions coincide motivates the formulation below.

3. The constraint  $T\#\mu = \nu$  is not linear (with respect to  $T$ ). It can be seen in a Euclidean and regular framework ( $T$  smooth and injective for example) by supposing that  $\mu$  and  $\nu$  are both absolutely continuous with respect to the Lebesgue measure (with respective densities noted abusively  $\mu(x)$  and  $\nu(y)$ ). The formula of the Jacobian change of variable asserts that  $T\#\mu = \nu$  is equivalent to the partial



differential equation :

$$\mu(x) = \nu(T(x)) |\det(DT(x))|$$

which is clearly not linear in  $T$ . As a consequence, the direct method of calculus of variations due to Weierstraß fails to prove the existence of a minimizer to the (1.1). For more details, see [50] (1.1 and Exercise 1).

In 1942, L. Kantorovich proposed in his article *On the Translocation of Masses* [31] another approach to the Optimal Transport problem where this time the volume  $d\mu(x)$  can be transported to several localities. His model consists in describing the transport by a probability measure  $\gamma \in \mathcal{P}(X \times Y)$  whose marginals are given by  $\mu$  and  $\nu$ . He calls such a measure a *transport plan* and proposes the following problem:

**Problem 2** (Kantorovich primal problem). Given  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  and a cost  $c : X \times Y \mapsto [0, \infty]$  measurable, a probability measure defined on  $X \times Y$ , say  $\gamma$ , is called transport plan between  $\mu$  and  $\nu$  if it satisfies both  $\pi_1\#\gamma = \mu$  and  $\pi_2\#\gamma = \nu$ , where we have denoted  $\pi_1(x, y) = x$  and  $\pi_2(x, y) = y$ . The set of these transport plans is denoted  $\Pi(\mu, \nu)$ . We interpret  $\gamma(x, y)$  as the volume transported between  $x$  and  $y$ .

Kantorovich problem with value  $\mathcal{K}(\mu, \nu)$  is then given by the following minimization:

$$\begin{aligned} \mathcal{K}(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{K(\gamma) : \gamma \in \Pi(\mu, \nu)\} \\ &= \inf \left\{ \iint_{X \times Y} c(x, y) d\gamma(x, y) : \pi_1\#\gamma = \mu, \pi_2\#\gamma = \nu \right\}. \end{aligned} \quad (1.2)$$

**Remark 2.** 1. The constraints on the marginals  $\pi_1\#\gamma = \mu$  and  $\pi_2\#\gamma = \nu$  each admit a dual writing in terms of test functions which is given by:

$$\pi_1\#\gamma = \mu \iff \forall \varphi \in L^1(\mu), \int_X \varphi(x) d\mu(x) = \iint_{X \times Y} \varphi(x), d\gamma(x, y) \quad (1.3)$$

$$\pi_2\#\gamma = \nu \iff \forall \psi \in L^1(\nu), \int_Y \psi(y) d\nu(y) = \iint_{X \times Y} \psi(y), d\gamma(x, y) \quad (1.4)$$

and we can restrict ourselves to the spaces of continuous and bounded functions  $C_b(X)$  and  $C_b(Y)$  when the spaces  $X$  and  $Y$  are supposed to be Polish (*i.e.* separable and whose topologies are induced by distances making them complete).

2. Kantorovich primal problem turns out to be a linear relaxation of Monge problem: if a function  $T : X \rightarrow Y$  satisfies  $T\#\mu = \nu$  then  $\gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\#\mu \in \Pi(\mu, \nu)$  and the total costs in (1.1) and (1.2) are equal. We deduce from this that  $\mathcal{K}(\mu, \nu) \leq \mathcal{M}(\mu, \nu)$ . We will deal with the cases of equality later.

3. The primal Kantorovich problem is a linear programming problem in the sense that the total cost is linear and continuous and the constraints are polytopes (*i.e.* intersections of a finite number of affine subspaces).

A first consequence of the previous point is that the Direct Method of the Calculus of Variations can straightforward be applied and we establish the existence of a minimizer to the problem (1.2) subject to some assumptions on the spaces  $X$  and  $Y$  as well as on the cost  $c$ . This is the object of the following theorem.

**Theorem 3** (Theorem 1.7, [50]). Let  $X$  and  $Y$  be two Polish spaces and  $c : X \times Y \rightarrow [0, \infty]$  a lower semicontinuous cost. Under these conditions, the (1.2) admits a solution. In other words, there exists  $\gamma \in \Pi(\mu, \nu)$  such that

$$K(\gamma) = \mathcal{K}(\mu, \nu).$$

### 1.3.2 Convex duality

Since the Kantorovich problem is linear, and therefore convex, we can look for the dual problem associated with it. In addition to the references on Optimal Transport cited above, one can refer to the books [1] and [49] for a synthetic and complete presentation of duality techniques in Convex Analysis.

Beware about the Optimal Transport terminology which calls "Dual Problem" what Convex Analysis calls "Primal Problem" and *vice versa*. In this manuscript we adopt the usages of Optimal Transport: the variables of the primal problem belong to a space which is the dual of the space to which the variables of the dual problem belong.

We introduce the Lagrange multipliers  $\varphi$  and  $\psi$  belonging respectively to  $C_b(X, \mathbb{R})$  and  $C_b(Y, \mathbb{R})$ . We *dualize* (1.2) thanks to the equations (1.3) and (1.4):

$$\begin{aligned} \mathcal{K}(\mu, \nu) &= \inf \left\{ \iint_{X \times Y} c(x, y) \, d\gamma(x, y) : \pi_1 \# \gamma = \mu, \pi_2 \# \gamma = \nu \right\} \\ &= \inf_{\gamma \geq 0} \sup_{\varphi, \psi} \iint c \, d\gamma - \int \varphi \, d[\pi_1 \# \gamma - \mu] + \int \psi \, d[\pi_2 \# \gamma - \nu] \\ &\geq \sup_{\varphi, \psi} \left( \int \varphi \, d\mu + \int \psi \, d\nu + \inf_{\gamma \geq 0} \iint [c(x, y) - \varphi(x) - \psi(y)] \, d\gamma(x, y) \right) \\ &= \sup_{\varphi, \psi} \left( \int \varphi \, d\mu + \int \psi \, d\nu - \iota_{\geq 0}(c - \varphi \oplus \psi) \right) \\ &= \sup_{\varphi \oplus \psi \leq c} \int_X \varphi \, d\mu + \int_Y \psi \, d\nu \end{aligned}$$

where  $\iota_{\geq 0} = \iota_{\geq 0}(\theta)$  the function which takes the value 0 if  $\theta = \theta(x, y)$  is positive at

all  $(x, y)$  and  $+\infty$  otherwise. The previous calculations suggest to define the following problem:

**Problem 3** (Kantorovich dual problem). Given  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  and a cost  $c : X \times Y \mapsto [0, \infty]$  measurable, the dual Kantorovich problem with value  $\mathcal{D}(\mu, \nu)$  is given by the following maximization:

$$\begin{aligned} \mathcal{D}(\mu, \nu) &\stackrel{\text{def}}{=} \sup \{D(\varphi, \psi) : \varphi \oplus \psi \leq c\} \\ &= \sup \left\{ \int \varphi \, d\mu + \int \psi \, d\nu : \forall (x, y) \in X \times Y, \varphi(x) + \psi(y) \leq c(x, y) \right\}, \end{aligned} \quad (1.5)$$

where the supremum is taken on the continuous and bounded  $\varphi$  and  $\psi$ .

The inequality  $\mathcal{D}(\mu, \nu) \leq \mathcal{K}(\mu, \nu)$  is in reality often an equality (there is no duality gap) as the following theorem states:

**Theorem 4** (Strong duality, Theorem 1.3, [56]). Under the same conditions as (1.2) and (1.5), we have the equality :

$$\mathcal{K}(\mu, \nu) = \mathcal{D}(\mu, \nu).$$

If we suppose moreover that the cost  $c$  is uniformly continuous and bounded (for example if  $X$  and  $Y$  are compact) then the problem (1.5) is attained, in other words there exists  $(\varphi, \psi)$  continuous and bounded such that  $\mathcal{D}(\mu, \nu) = D(\varphi, \psi)$ .

**Remark 5.** 1. The dual problem allows to *certify* a hypothetical minimizer of the primal problem: if we estimate that  $\gamma$  is a good candidate to be a solution in (1.2) then it is sufficient to find a couple  $(\varphi, \psi)$  such that  $D(\varphi, \psi) = K(\gamma)$  to deduce that  $\gamma$  is indeed a minimizer of the primal problem. Moreover,  $(\varphi, \psi)$  will then be a maximizer of the dual problem since the equality  $K(\gamma) \geq D(\varphi, \psi)$  is always satisfied.

2. We can give a monetary interpretation of  $\varphi$  and  $\psi$  : Alice is the CEO of a large multinational company with branches all over Europe. She wishes to redistribute the computer equipment among the different premises of her companies. To do this, she calls on two delivery companies belonging to Bob and Carole respectively, who each offer her a different contract:

(Bob) "My trucks will carry out the transport according to the  $\gamma$  plan and you will have to pay the  $K(\gamma)$  amount. **You are free** to give me the  $\gamma$  you prefer."

(Carole) "My company proposes not to charge for the journey of my trucks, but to charge you a price  $\varphi(x)$  for loading in  $x$  and to add to that a price  $\psi(y)$  for

unloading in  $y$ . To reassure you, I guarantee that  $\varphi(x) + \psi(y)$  will always be less than  $c(x, y)$ . **Free to me** to set the prices  $\varphi$  and  $\psi$ ."

If Alice and Carol are omniscient, the previous theorem guarantees Alice that both contracts will cost her the same amount. Thus  $\varphi$  and  $\psi$  can be interpreted as equilibrium prices.

Looking carefully at the functional in (1.5) we notice that if  $(\gamma, \varphi, \psi)$  are optimal for their respective problems then the constraint  $\varphi(x) + \psi(y) \leq c(x, y)$  must be saturated at any point  $(x, y)$  of the support of  $\gamma$ . A simple process to do this is, starting from  $\varphi$  and  $\psi$  satisfying the above inequality, to replace  $\varphi$  by the  $c$ -transform of  $\psi$  which we now define:

**Definition 6** ( $c$ -transform and  $\bar{c}$ -transform). Given a function  $f : X \rightarrow \mathbb{R}$ , its  $c$ -transform, denoted  $f^c$  and defined on  $Y$  is given by:

$$f^c(y) = \inf_{x \in X} c(x, y) - f(x)$$

In the same way we define for a function  $g : Y \rightarrow \mathbb{R}$ , its  $\bar{c}$ -transform, noted  $f^{\bar{c}}$  and defined on  $X$  by:

$$g^{\bar{c}}(x) = \inf_{y \in Y} c(x, y) - g(y)$$

A first interest of this notion is thus that given  $(\varphi, \psi)$  admissible in (1.5), the couples  $(\varphi, \varphi^{\bar{c}})$  and  $(\varphi^{\bar{c}, c}, \varphi^{\bar{c}})$  are also admissible and improve the dual cost in the sense

$$D(\varphi, \psi) \leq D(\varphi, \varphi^{\bar{c}}) \leq D(\varphi^{\bar{c}, c}, \varphi^{\bar{c}}).$$

Unfortunately one cannot repeat this process indefinitely as the following proposition states:

**Proposition 7** (Proposition 1.34, [50]). With the same notations as before, if we suppose moreover that  $c$  is real-valued, we have the following equalities:

$$f^{c, \bar{c}, c} = f^c \text{ and } g^{\bar{c}, c, \bar{c}} = g^{\bar{c}}$$

The problem (1.5) can therefore be written as a maximization problem depending only on  $\varphi$ :

**Problem 4** (Dual problem, bis). Given  $\mu \in \mathcal{P}(X)$ ,  $\nu \in \mathcal{P}(Y)$  and a cost  $c : X \times Y \mapsto [0, \infty]$

measurable, the problem (1.5) admits the following writing:

$$\mathcal{D}(\mu, \nu) = \sup \left\{ \int \varphi \, d\mu + \int \varphi^c \, d\nu : \varphi \text{ continuous and bounded} \right\}. \quad (1.6)$$

where the supremum is taken on the continuous and bounded  $\varphi$ .

**Remark 8.** The  $c$ -transform (resp. the  $\bar{c}$ -transform) being defined as the infimum of a family of functions indexed by  $x$  (resp. by  $y$ ), if this family admits a common continuity modulus then it is the same for the  $c$ -transformed (resp. for the  $\bar{c}$ -transformed). In particular if  $c$  is assumed to be Lipschitz, then so it is for  $f^c$  and  $g^{\bar{c}}$ .

### 1.3.3 Optimality conditions

From now on we ask ourselves when the Monge problem and the Kantorovich problem have the same value. To answer this question we will look for conditions on  $X, Y, c$  as well as  $\mu$  which allow us to exhibit an optimal  $\gamma$  transport plan in (1.2) and which is written in the form  $\gamma = \gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\#\mu$  with  $T : X \rightarrow Y$  a measurable map.

Let us start by setting ourselves in the conditions of the Theorem 4 and thus have  $\mathcal{K}(\mu, \nu) = \mathcal{D}(\mu, \nu)$ . We set  $X = Y$  and we give ourselves  $\gamma$  and  $\varphi$  optimal so that the equality  $K(\gamma) = D(\varphi, \varphi^c)$  implies

$$\forall x, y \in X \times X, \varphi(x) + \varphi^c(y) \leq c(x, y) \quad \forall x, y \in \text{supp}(\gamma), \varphi(x) + \varphi^c(y) = c(x, y)$$

where  $\text{supp}(\gamma)$  denotes the support of  $\gamma$  defined as the intersection of closed set with full measure. Let us fix  $(x_0, y_0) \in \text{supp}(\gamma)$  a couple of points where there is equality and let us look at the function  $x \rightarrow \varphi(x) - c(x, y_0)$  which reaches its minimum in  $x_0$  and is continuous as soon as  $c$  is. If  $x_0$  lies inside  $X$  and both  $\varphi$  and  $x \rightarrow c(x, y_0)$  are differentiable in  $x_0$  then the first order optimality condition implies that

$$\nabla \varphi(x_0) = \nabla_x c(x_0, y_0). \quad (1.7)$$

If we manage to write the previous equation as an equation giving  $y_0$  as a function of  $x_0$  or in other words that the starting point completely determines the destination then we can deduce that a solution of the Monge problem is a solution of the Kantorovich problem. A sufficient condition for this is to look at particular costs in the form  $c(x, y) = h(x - y)$  with  $h$  strictly convex and differentiable (for example, if  $X = \mathbb{R}^d$ ,  $h(x) = |x|^p$ ,  $p > 1$ ) since then the equation (1.7) becomes

$$\varphi(x_0) = \nabla h(x_0 - y_0)$$

and the differentiability of  $h$  associated to its strict convexity ensures that  $\nabla h$  is invertible (for more details, see [50], 1.3). After calculations we obtain

$$y_0 = x_0 + (\nabla h)^{-1}(\nabla \varphi(x_0)).$$

By setting  $T(x) = x + (\nabla h)^{-1}(\nabla \varphi(x))$  and  $\gamma = \gamma_T$ , we deduce that  $\gamma_T$  is optimal for (1.2) and consequently that  $T$  is optimal for (1.1).

It remains to find a framework where all the previous calculations allow to define an application of  $X$  in  $X$ . The classical framework is to take  $X = Y = \Omega$  a compact of  $\mathbb{R}^d$  since as  $c$  is locally Lipschitz by convexity of  $h$ ,  $c$  is then globally Lipschitz by compactness of  $\Omega$  and so is  $\varphi$  by the Remark 8. Rademacher theorem then assures us that  $\varphi$  is differentiable almost everywhere. Finally if  $\mu$  is absolutely continuous with respect to the Lebesgue measure, the non-differentiability points of  $\varphi$  will be negligible for  $\mu$  and the previous calculations will be valid almost everywhere. To summarize we have the following result:

**Theorem 9** ([50], Theorem 1.17). Let  $\Omega$  be a compact of  $\mathbb{R}^d$  and  $\mu, \nu \in \mathcal{P}(\Omega)$ . If  $c(x, y) = h(x - y)$  where  $h$  is a strictly convex function then the (1.2) and (1.6) problems are attained. If moreover  $\mu$  is absolutely continuous and  $\mu(\partial\Omega) = 0$  then we have uniqueness of the solution in (1.2) and this one is of the form  $\gamma = \gamma_T \stackrel{\text{def}}{=} (\text{Id}, T)\#\mu$  where  $T$  is given by

$$T(x) = x + (\nabla h)^{-1}(\nabla \varphi(x))$$

where  $\varphi$  is a solution of (1.6). Moreover  $T$  is optimal in (1.1).

Let us finish by mentioning a theorem specific to the case where  $c(x, y) = |x - y|^2/2$  which is due to Brenier in his article [8].

**Theorem 10** (Brenier Theorem). Given two measures  $\mu, \nu \in \mathbb{R}^d$  with  $\mu$  absolutely continuous. There exists a unique map  $T$  satisfying  $T\#\mu = \nu$  and moreover it is of the form  $T = \nabla u$  with  $u : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$  a convex function.

Moreover, by taking  $c(x, y) = \frac{1}{2}|x - y|^2$  this application  $T$  is a solution of (1.1) and  $\gamma_T$  is a solution of (1.2).

### 1.3.4 Wasserstein metric

If we give ourselves  $d$  a distance on a Polish space<sup>1</sup>  $X$  and we set  $c = d$ , it is easy to see that the set of Dirac measures on  $X$  with cost  $K(\cdot, \cdot)$  is a metric space canonically isometric to  $(X, d)$ . One can legitimately hope that such a result extends to probability measures and thus allows to extend the metric on  $X$  to a metric on  $\mathcal{P}(X)$ . This is indeed

---

<sup>1</sup>That is separable and complete.

the case as the following theorem states:

**Theorem 11.** Let  $(X, d)$  be a Polish space and  $p \in [1, \infty)$ . Considering the cost  $c(x, y) = d(x, y)^p$ , we define a distance  $W_p$  on  $\mathcal{P}_p(X)$  the set of probability measures on  $X$  admitting a moment of order  $p$  by setting

$$W_p(\mu, \nu) = \mathcal{K}^{1/p}(\mu, \nu).$$

This distance is called Wasserstein distance of order  $p$ .

The main properties of this distance are gathered in the following proposition.

**Proposition 12** ([57], Theorems 6.9 and 6.18). [1] For  $p \in [1, \infty)$ , the distance  $W_p$  metrizes the weak convergence- $*$  of probability measures.

[2] If  $(X, d)$  is Polish, then so is  $(\mathcal{P}_p(X), W_p)$ .

Note that the weak convergence- $*$  reads on Dirac measures as follows:  $\delta_x \rightarrow \delta_y$  if and only if  $x \rightarrow y$ . This is not surprising since according to the previous comment Dirac measures can be seen as an embedding of  $X$  into  $\mathcal{P}(X)$ .

### 1.3.5 Dynamic formulation

Let us stop at the case where  $X = Y = \mathbb{R}^d$  and  $c(x, y) = |x - y|^p/p$  with  $p \in [1, \infty)$ . In their paper [3], Benamou and Brenier proposed a new formulation of Optimal Transport based on fluid mechanics and introducing a time variable. This is known as the Dynamic Optimal Transport Formulation. From a metric point of view, they established that under certain conditions, the Wasserstein metric is geodesic.

Starting at time  $t = 0$  from a density measure  $\mu$  on  $\mathbb{R}^d$ , the authors are interested in all the paths that lead at time  $t = 1$  to  $\nu$  another density measure on  $\mathbb{R}^d$  in such a way that the total mass is conserved in time. Mathematically, this amounts to being interested in the set of continuous  $\rho : t \rightarrow \rho_t \in \mathcal{P}_{ac}(\mathbb{R}^d)$  paths (for the weak convergence- $*$  of the measures) and linking  $\mu$  and  $\nu$  (i.e.  $\rho_0 = \mu$  and  $\rho_1 = \nu$ ) subjected to the velocity field  $\mathbf{v} = \mathbf{v}_t(x)$  and satisfying the so-called continuity equation (or mass conservation):

$$\partial_t \rho_t(x) + \operatorname{div}_x [\rho_t(x) \mathbf{v}_t(x)] = 0. \quad (1.8)$$

The total cost is now given by the *action*  $\mathcal{A}_p(\rho, \mathbf{v})$  of this kinematics (corresponding to the sum of kinetic energies when  $p = 2$ ) defined by

$$\mathcal{A}_p(\rho, \mathbf{v}) = \int_0^1 \int_X \frac{|\mathbf{v}_t(x)|^p}{p} d\rho_t(x) dt.$$

Unfortunately this functional is not convex with respect to  $(\rho, \mathbf{v})$  and the trick to make

it convex is to make the change of variable  $\mathbf{v} \leftarrow \mathbf{m} = \rho \mathbf{v}$  to replace the velocity by the momentum. The problem is then given by :

**Problem 5** (Benamou-Brenier problem). Given  $\mu \in \mathcal{P}_{ac}(X)$ ,  $\nu \in \mathcal{P}_{ac}(X)$  and  $p \in [1, \infty)$ , the Benamou-Brenier problem of value  $\mathcal{B}_p(\mu, \nu)$  is given by the following minimization:

$$\mathcal{B}_p(\mu, \nu) \stackrel{\text{def}}{=} \inf \{ \mathcal{A}(\rho, \mathbf{m}) : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \} \quad (1.9)$$

$$= \inf \left\{ \frac{1}{p} \int_0^1 \int_X \frac{|\mathbf{m}_t(x)|^p}{\rho_t(x)^{p-2}} d\rho_t(x) dt : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \right\} \quad (1.10)$$

This formulation extends previous one of Kantorovich and Monge in a sense given by the following theorem:

**Theorem 13** ([50], Theorem 5.27). Let  $X$  be a convex of  $\mathbb{R}^d$  and  $p \in [1, \infty)$ . We give  $\mu, \nu \in \mathcal{P}_p(X)$ , then :

$$W_p(\mu, \nu) = \mathcal{B}_p(\mu, \nu).$$

Moreover the infimum is attained for  $\mathcal{B}_p$  as soon as the associated Monge problem admits a solution  $T$  and we then have for solution  $\rho_t = [(1-t)\operatorname{Id} + tT] \# \mu$ . It is thus in particular the case when  $\mu$  is absolutely continuous. In other words,  $(\mathcal{P}_p(X), W_p)$  is a geodesic space whose geodesics are the solutions of the problem (1.9).

Since we are dealing with a convex minimization problem, we can *dualize* this problem. Let us do it for  $p = 2$ .

$$\begin{aligned} \mathcal{B}_2(\mu, \nu) &= \inf \{ \mathcal{A}(\rho, \mathbf{m}) : \partial_t \rho_t(x) + \operatorname{div}_x \mathbf{m}_t(x) = 0 \} \\ &= \inf_{(\rho, \mathbf{m})} \mathcal{A}(\rho, \mathbf{m}) + \sup_{\phi} \left[ \int_0^1 \int_X ((\partial_t \phi) \rho + \nabla \phi \cdot \mathbf{m}) + \Delta_{\mu}^v(\phi) \right] \end{aligned}$$

where  $\Delta_{\mu}^v(\phi) \stackrel{\text{def}}{=} \int_X \phi(1, \cdot) d\nu - \int_X \phi(0, \cdot) d\mu$  corresponds to the initial and final conditions. We then switch the infimum and the supremum while noticing that  $\frac{|\mathbf{m}|^2}{2\rho} - \nabla \phi \cdot \mathbf{m} = \left( \frac{\mathbf{m}}{2\rho} - \nabla \phi \right) \cdot \mathbf{m}$  and thus that minimizing in  $\mathbf{m}$  is equivalent to impose  $\mathbf{m} = \rho \nabla \phi$  :

$$\begin{aligned} \mathcal{B}_2(\mu, \nu) &\leq \sup_{\phi} \Delta_{\mu}^v(\phi) + \inf_{\phi \geq 0} \int_0^1 \int_X \left[ -(\partial_t \phi) \rho - \frac{1}{2} |\nabla \phi|^2 \rho \right] \\ &= \sup \left\{ \Delta_{\mu}^v(\phi) : \partial_t \phi + \frac{1}{2} |\nabla \phi|^2 \leq 0 \right\} \end{aligned}$$

where the constraint on  $\phi$  is a particular case of Hamilton-Jacobi inequation. The primal-dual optimality conditions impose in particular to have equality on the support of the



optimal geodesic  $\rho$ .

### 1.3.6 Entropic Optimal Transport

We now present a numerical method for to determine both an approximation of the optimal transport plan and an approximation of the optimal potentials associated with it. Known as *Entropic Optimal Transport Regularisation* (see Cuturi's article [22]). The few notes that follow are entirely taken from [47].

We place ourselves in a discrete framework: given  $n, m \in \mathbb{N}^*$  we have two convex combinations of Dirac measures

$$\begin{aligned}\mu &\stackrel{\text{def}}{=} \sum_{i=1}^n \mu_i \delta_{x_i} \\ \nu &\stackrel{\text{def}}{=} \sum_{j=1}^m \nu_j \delta_{y_j}\end{aligned}$$

with  $\sum_i \mu_i = \sum_j \nu_j = 1$ . By abuse of notation we will write  $\mu = (\mu_i) \in \mathbb{R}^n$  and  $\nu = (\nu_j) \in \mathbb{R}^m$ . We implement the cost and the transport plans in rectangular matrices  $C = C_{i,j} \stackrel{\text{def}}{=} c(x_i, y_j) \in \mathbb{R}^{n \times m}$  and  $\gamma = \gamma_{ij} \stackrel{\text{def}}{=} \gamma(x_i, y_j) \in \mathbb{R}^{n \times m}$ . The (1.2) is then rewritten (where  $\cdot$  is the usual matricial product) :

$$\min \{ \langle C | \gamma \rangle : \gamma \cdot \mathbb{1}_n = \mu, \gamma^T \cdot \mathbb{1}_m = \nu \}. \quad (1.11)$$

Starting from the (1.11), we introduce a family of new problems parametrized by  $\varepsilon \in \mathbb{R}_+^*$  and given by:

$$\begin{aligned}\mathcal{K}_\varepsilon(\mu, \nu) &\stackrel{\text{def}}{=} \inf \{ K_\varepsilon(\gamma) : \gamma \in \Pi(\mu, \nu) \} \\ &= \inf \{ K(\gamma) - \varepsilon H(\gamma) : \gamma \in \Pi(\mu, \nu) \}\end{aligned} \quad (1.12)$$

where  $H(\gamma) \stackrel{\text{def}}{=} -\sum_{i,j} \gamma_{ij} (\log(\gamma_{ij}) - 1)$  denotes the (concave) entropy of the  $\gamma$  transport plan. These new problems are  $\varepsilon$ -strongly convex and therefore admit a unique solution. Moreover when  $\varepsilon$  tends to 0, the solutions of (1.12) converge towards a solution of (1.11), more precisely towards the one of maximum entropy (see [21] for more details).

As before we introduce the Lagrange multipliers  $\varphi \in \mathbb{R}^n$  and  $\psi \in \mathbb{R}^m$  :

$$K_\varepsilon(\mu, \nu) = \min_{\gamma} \langle \gamma | C \rangle - \varepsilon H(\gamma) - \langle \varphi | \gamma \cdot \mathbb{1}_n - \mu \rangle - \langle \psi | \gamma^T \cdot \mathbb{1}_m - \nu \rangle$$

and the optimality conditions imply that  $(\gamma, \varphi, \psi)$  are optimal if and only if

$$C_{ij} + \varepsilon \log(\gamma_{ij}) - \varphi_i - \psi_j = 0 \quad (1.13)$$

$$\gamma \cdot \mathbb{1}_n = \mu \quad (1.14)$$

$$\gamma^T \cdot \mathbb{1}_m = \nu \quad (1.15)$$

The equation (1.13) is rewritten:

$$\forall (i, j) \in \llbracket 1, n \rrbracket \times \llbracket 1, m \rrbracket, \gamma_{ij} = u_i K_{ij} v_j$$

where  $u_i = \exp(\varphi_i/\varepsilon)$ ,  $v_j = \exp(\psi_j/\varepsilon)$  and  $K_{ij} = \exp(-C_{ij}/\varepsilon)$ . This last equation can be rewritten in a more compact way:

$$\gamma = \text{diag}(u) K \text{diag}(v) \quad (1.16)$$

and consequently the marginal conditions (1.14) and (1.15) become (where  $\odot$  is the element-wise product) :

$$u \odot (Kv) = \mu, \quad (1.17)$$

$$v \odot (K^T u) = \nu. \quad (1.18)$$

To solve the problem (1.17) and (1.18), Cuturi uses the so-called Sinkhorn algorithm (see [52], [25], [32]), which, starting from  $u_0 = \mathbb{1}_n, v_0 = \mathbb{1}_m$ , consists in iteratively constructing  $u^{(l)}$  and  $v^{(l)}$  ( $l \in \mathbb{N}^*$ ) in the following way (where the quotient symbol is the element-wise division):

$$u^{(l)} = \frac{\mu}{K \cdot v^{(l-1)}}, v^{(l)} = \frac{\nu}{K^T \cdot u^{(l)}} \quad (1.19)$$

This amounts to ask that  $u^{(l)}$  and  $v^{(l)}$  satisfy alternately (1.17) and (1.18). This method can be seen as an alternate projection scheme on the constraint (1.17) and then on the constraint (1.18), which is a special case of Dykstra algorithm with Bregman projections (see [2]). The convergence is ensured by the following theorem for which a proof can be found in [25] or [46].

**Theorem 14.** The sequence defined in (1.19) converges linearly to a solution of (1.14) and (1.15).

**Remark 15** (soft-min). On the one hand, the previous iterations have the advantage of being very simple to implement (the sequences  $u$  and  $v$  are geometrical) but from a

numerical point of view the variables  $u$  or  $v$  tend to require a very important encoding (an overflow phenomenon). In practice  $u$  diverging towards  $+\infty$  as fast as  $v$  tends towards 0 by positive values. One way to overcome this problem is to add a so-called soft–min step to each iteration. Indeed, the iterations (1.17) and (1.18) can be rewritten in the variables  $(\varphi^{(l)}, \psi^{(l)})$  as follows:

$$\begin{aligned}\varphi^{(l)} &= \varepsilon \log(\mu) - \varepsilon \log \left[ K \exp \left( \frac{\psi^{(l-1)}}{\varepsilon} \right) \right] \\ \psi^{(l)} &= \varepsilon \log(v) - \varepsilon \log \left[ K^T \exp \left( \frac{\varphi^{(l)}}{\varepsilon} \right) \right]\end{aligned}$$

and remembering that  $K = \exp(-C/\varepsilon)$  we can rewrite these two equations with fixed  $i$  and  $j$  coordinates:

$$\begin{aligned}\varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) - \varepsilon \log \left[ \sum_j \exp \left( \frac{-C[i, j] + \psi^{(l-1)}[j]}{\varepsilon} \right) \right] \\ \psi^{(l)}[j] &= \varepsilon \log(v[j]) - \varepsilon \log \left[ \sum_i \exp \left( \frac{-C[i, j] + \varphi^{(l)}[i]}{\varepsilon} \right) \right].\end{aligned}$$

By introducing the notation soft–min, defined for a vector  $\mathbf{z} = (z_i)$  of any dimension:

$$\min_{\varepsilon}(\mathbf{z}) \stackrel{\text{def}}{=} -\varepsilon \log \left[ \sum_i \exp \left( \frac{-z_i}{\varepsilon} \right) \right],$$

we obtain

$$\begin{aligned}\varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) + \min_{\varepsilon}(C[i, \cdot] - \psi^{(l-1)}) \\ \psi^{(l)}[j] &= \varepsilon \log(v[j]) + \min_{\varepsilon}(C[\cdot, j] - \varphi^{(l)}).\end{aligned}$$

A simple but fundamental property of soft–min is given by :

$$\forall t \in \mathbb{R}, \min_{\varepsilon}(\mathbf{z}) = t + \min_{\varepsilon}(\mathbf{z} - t).$$

It<sup>2</sup> has for interest to relieve partly the term in the exponential by transforming it into an addition carried out *a posteriori*, which lightens the *overflow*. Expanding this last

---

<sup>2</sup>To understand this name of soft–min, we can notice that  $\min_{\varepsilon}(\mathbf{z})$  tends to  $\min_i z_i$  when  $\varepsilon$  tends to 0.

term gives us for the first equation :

$$\min_{\varepsilon}(\mathbf{z} - t) = -\varepsilon \log \left[ \sum_i \exp \left( \frac{-C[i, j] + \psi^{(l-1)}[j] + t}{\varepsilon} \right) \right]$$

And since we can choose  $t$ , we might as well choose one that makes the sum as close as possible to 0. The choice of the  $\bar{c}$ -transform of  $\psi^{(l-1)}$  is then quite obvious. It holds obviously for  $\psi^{(l)}[j]$ . Finally the new iterates are given by :

$$\begin{aligned} \varphi^{(l)}[i] &= \varepsilon \log(\mu[i]) + \psi^{(l-1), \bar{c}}[i] + \min_{\varepsilon}(C[i, \cdot] - \psi^{(l-1)} - \psi^{(l-1), \bar{c}}[i]) \\ \psi^{(l)}[j] &= \varepsilon \log(\nu[j]) + \varphi^{(l), c}[j] + \min_{\varepsilon}(C[\cdot, j] - \varphi^{(l)} - \varphi^{(l), c}[j]). \end{aligned}$$

Note that recently, the method of Entropic Regularisation of Optimal Transport has been extended to other problems: multi-marginal Optimal Transport [42], unbalanced [51] or gradient flows [45].

## 2 Multi-species Optimal Transportation

### 2.1 Survol

Le premier chapitre de ce manuscrit est constitué d'un article publié dans le *Journal of Optimization Theory and Applications* (JOTA) en 2020. Des simulations numériques non présentes dans le papier original ont été ajoutées en fin de chapitre. Celles-ci reposent sur la méthode de relaxation entropique et l'algorithme de Sinkhorn que l'on a présentés en introduction.

On se place dans le cadre classique du Transport Optimal entre mesures de probabilité : soient  $\mu, \nu \in \mathcal{P}(X)$  deux telles mesures que l'on se représente comme étant deux tas de sable distribués sur l'espace  $X$ . Avec deux pincesaux de couleur **bleu** et **rouge**, on colorie certains grains de sable de  $\mu$  en **bleu** et tous les autres en **rouge**. On se retrouve alors avec deux mesures positives,  $\mu_1$  et  $\mu_2$ , dont la somme redonne  $\mu$  :  $\mu = \mu_1 + \mu_2$ . On fait de même avec  $\nu$ , ainsi  $\nu = \nu_1 + \nu_2$ . Plus généralement on peut se donner  $n$  couleurs et *partitionner* nos mesures de départ et d'arrivée en  $n$  sous-mesures. Cela nous donne comme définition :

**Définition 1.** Étant donné  $(X, \mathcal{X})$  un espace mesurable,  $\mathcal{P}^n(X)$  désigne l'ensemble des distributions admissibles de  $n$  espèces défini par

$$\mathcal{P}^n(X) = \left\{ \mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} \in \mathcal{M}_+^n(X) : \sum_{i=1}^n m_i \in \mathcal{P}(X) \right\}.$$

**Remarque 2.** Plutôt que de parler de couleurs, l'article parlait d'espèces :  $\mu$  peut représenter une distribution initiale de prédateurs  $\mu_{predator}$  et de proies  $\mu_{prey}$  de telle sorte que  $\mu = \mu_{predator} + \mu_{prey}$  et  $\nu = \nu_{predator} + \nu_{prey}$  représente alors une distribution finale de tels prédateurs et proies.

Il nous faut définir quels sont les couplages autorisés. On a deux choix :

1. Le sable **bleu** ne peut être transporté que vers du sable **bleu** et le sable **rouge** ne peut être transporté que vers du sable **rouge**. Ceci impose une égalité entre les différentes masses. On se donne alors deux coûts différents et autant de plans de transport.
2. On autorise le sable de couleur  $A \in \{\text{bleu}, \text{rouge}\}$  à être transporté vers du sable de couleur  $B \in \{\text{bleu}, \text{rouge}\}$ . On a alors quatre coûts différents et autant de plans de transport qui sont ceux associés aux transports de  $A$  vers  $A$ ,  $A$  vers  $B$ ,  $B$  vers  $A$

et  $B$  vers  $B$ .

Je me suis dirigé vers le second car j'ai cherché à proposer un modèle qui autorisait les mélanges entre espèces. Aussi, le premier modèle est un cas particulier du second dès lors que l'on prend un coût infini entre  $A$  et  $B$  et  $B$  et  $A$ . Pour reprendre l'exemple précédent avec les couleurs, un couplage entre  $\mu = \mu_1 + \mu_2$  et  $\nu = \nu_1 + \nu_2$  est donc décrit par la concaténation de quatre plans de transport  $\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{22}$  associés à autant de coûts  $c_{11}, c_{12}, c_{21}, c_{22}$  : par exemple  $\gamma_{12}$  décrit quelle partie de  $\mu_1$  est transportée vers  $\nu_2$  et ce pour un coût  $c_{12}$ . Quand on passe à  $n$  couleurs/espèces cela nous donne :

**Définition 3.** Étant donné  $\mu \in \mathcal{P}^n(X)$  et  $\nu \in \mathcal{P}^n(Y)$ ,  $\Pi(\mu, \nu)$  désigne l'ensemble des plans de transport entre  $\mu$  et  $\nu$  défini par

$$\Pi(\mu, \nu) = \left\{ \gamma \in \mathcal{M}_+^{n \times n}(X \times Y) : \forall (i, j) \in \llbracket 1, n \rrbracket^2, \begin{array}{l} \mu_i = \pi_1 \# \left( \sum_{k=1}^n \gamma_{ik} \right) \\ \nu_j = \pi_2 \# \left( \sum_{k=1}^n \gamma_{kj} \right) \end{array} \right\},$$

pour un tel  $\gamma \in \Pi(\mu, \nu)$ ,  $K(\gamma)$ , le coût de transport total est donné par

$$K(\gamma) = \sum_{(i,j) \in \llbracket 1, n \rrbracket^2} \iint_{X \times Y} c_{ij}(x, y) d\gamma_{ij}(x, y)$$

et le problème de minimisation est

$$\mathcal{K}(\mu, \nu) = \inf \{K(\gamma) : \gamma \in \Pi(\mu, \nu)\}.$$

Il s'agit là d'un problème de programmation linéaire dont le problème dual est étudié dans l'article. En particulier on démontre un résultat de dualité forte entre eux.

Une fois le modèle bien posé, je me suis demandé quelles étaient les conditions à imposer aux coûts  $(c_{ij})_{1 \leq i, j \leq n}$  pour obtenir une distance sur l'espace  $\mathcal{P}^n(X)$ . En plus de demander que chaque  $c_{ij} \stackrel{\text{def}}{=} d_{ij}$  soit une distance sur  $X$ , il est nécessaire et suffisant d'imposer une condition que j'ai appelée « Inégalités Triangulaires Mixtes » (*Mixed Triangle Inequalities* en anglais) qui est donnée par :

$$\forall (i, j, k) \in \llbracket 1, n \rrbracket^3, \forall (x, y, z) \in X^3, d_{ik}(x, z) \leq d_{ij}(x, y) + d_{jk}(y, z). \quad (\text{MTI})$$

Celle-ci interdit notamment à un grain de sable **bleu** présent en  $x$  et allant en  $z$  de se changer sur place en **rouge** pour un coût modique et de faire le trajet  $x \rightarrow z$  pour une somme toute aussi faible (prendre  $x = y$  et  $d_{12}(x, x) + d_{21}(x, z) < d_{12}(x, z)$  dans l'inégalité

précédente). Ces différentes considérations m'ont mené à la définition et au théorème suivants.

**Définition 4.** Étant donné  $\mu, \nu \in \mathcal{P}_p^n(X)$ , la distance de transport d'ordre  $p$  entre  $\mu$  et  $\nu$  est définie par

$$W_p(\mu, \nu) = \left( \inf \left\{ \sum_{ij} \iint d_{ij}(x, y)^p d\gamma_{ij}(x, y), \gamma \in \Pi(\mu, \nu) \right\} \right)^{\frac{1}{p}} \\ := (\mathcal{K}(\mu, \nu))^{\frac{1}{p}}.$$

**Théorème 5.** Soit  $X$  un espace Polonais. Soit  $p \in [1, \infty[$ . Étant donné  $(d_{ij})$ ,  $n^2$  fonctions définies sur  $X \times X$  et à valeurs dans  $\mathbb{R}_+$  telles que :

- [1]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2$ ,  $d_{ij}$  est symétrique,
- [2] (MTI) est vérifiée,
- [3]  $\forall i \in \llbracket 1, n \rrbracket$ ,  $\forall x \in X$ ,  $d_{ii}(x, x) = 0$ ,
- [4]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2$ ,  $i \neq j$ ,  $\forall (x, y) \in X \times Y$ ,  $d_{ij}(x, y) \neq 0$ .

Sous ces conditions,  $W_p$  est une distance sur  $\mathcal{P}_p^n(X)$ .

Enfin, j'ai proposé des simulations numériques pour illustrer l'article dont le calcul est basé sur la méthode de relaxation entropique et l'algorithme de Sinkhorn rendus célèbres en Transport Optimal par l'article de Cuturi [22]. Cela fait l'objet de l'Annexe et constitue un travail original qui n'a pas été publié.

**Travaux similaires** Chen et al. proposent plusieurs formulations du Transport Optimal de mesures vectorielles dans [11], [17] et [14]. Celles-ci se basent sur la formulation dynamique de Benamou-Brenier et ils y établissent des résultats de dualité forte ainsi que des simulations numériques.

K.J. Ciosmak a proposé plusieurs modèles de Transport Optimal entre mesures à valeurs dans  $\mathbb{R}^m$ . Son second modèle développé dans [20] a initialement pour but d'étudier les cas d'égalité dans l'inégalité de Hölder matricielle. Pour cela il développe une formulation à la Beckmann (divergence nulle, voir le chapitre 4 de [50] pour plus de détails) du Transport Optimal entre mesures à valeurs vectorielles : étant donné  $\mu = \mu_+ - \mu_- \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$ , il s'intéresse au problème suivant :

$$\inf \{ \|M\|_{\mathcal{M}} : M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{n \times m}), -\operatorname{div} M = \mu \}$$

où  $\|\cdot\|_{\mathcal{M}}$  désigne la norme de variation totale. Comme dans le cas scalaire, ce problème

admet pour dual un problème de maximisation donné par :

$$\sup \left\{ \int \langle F | d\mu \rangle : \|f\|_{C^1} \leq 1 \right\}.$$

Par rapport au modèle proposé dans le chapitre 1, il s'agit ici d'une généralisation du cas particulier du Transport Optimal avec coût  $L^1$ .

Wang et Zhang ont proposé dans [58] un modèle appelé Transport Optimal simultané. On se donne  $\mu \in \mathcal{M}(X, \mathbb{R}_+^d)$  et  $\nu \in \mathcal{M}(Y, \mathbb{R}_+^d)$  telles que  $\mu_i(X) \geq \nu_i(Y)$  pour tout  $i = 1, \dots, d$  (donc non nécessairement de même masse) et ils interprètent  $\nu$  comme une demande et  $\mu$  comme une offre, cette dernière pouvant être supérieure à la demande. Une formulation à la Monge est proposée au moyen d'une fonction  $T : X \rightarrow Y$  vérifiant  $T\#\mu \geq \nu$  et qui s'interprète ainsi : en  $x \in X$  il n'y a qu'un seul moyen de transport et celui-ci transportera avec lui la totalité des ressources (tous les  $d\mu_i(x)$ ,  $i = 1, \dots, d$ ) jusqu'à la destination  $T(x)$ . On demande de plus que la demande  $\nu$  soit satisfaite. Une formulation à la Kantorovich est également proposée. Ici il s'agit de considérer les mesures  $\kappa \in \mathcal{M}(X \times Y, \mathbb{R}_+)$  vérifiant :

$$\int_X \kappa(x, \cdot) d\mu(x) \geq \nu,$$

ce qui a pour conséquence qu'un tel transport répartit équitablement les différentes ressources  $d\mu_i(x)$ ,  $i = 1, \dots, d$  présentes en chaque point.

Les auteurs remarquent qu'il n'y a pas toujours existence de tels transports, indépendamment de leur optimalité. Ils s'attachent alors à trouver des conditions suffisantes pour que le problème soit bien posé et établissent un résultat de dualité forte pour leur modèle.

Notons enfin que Gover et Wolansky dans [28] traitent un modèle proche de celui-ci dans le cas *semi-discret*, c'est-à-dire quand la mesure d'arrivée est discrète.

## 2.1 Overview

The first chapter of this manuscript consists of a paper published in the *Journal of Optimization Theory and Applications* (JOTA) in 2020. Numerical simulations not present in the original paper have been added at the end of the chapter. These are based on the entropy relaxation method and the Sinkhorn algorithm presented in the Introduction of this manuscript.

We place ourselves in the classical framework of Optimal Transport between probability measures: let  $\mu, \nu \in \mathcal{P}(X)$  be two such measures which we represent as being two amounts of sand distributed on the space  $X$ . With two brushes of colour [blue](#) and



red, we colour some grains of sand of  $\mu$  in blue and all the others in red. We then have two positive measures,  $\mu_1$  and  $\mu_2$ , whose the sum gives  $\mu$ :  $\mu = \mu_1 + \mu_2$ . We do the same with  $\nu$ , so  $\nu = \nu_1 + \nu_2$ . More generally, we can give ourselves  $n$  colours and *partitinate* our starting and ending measurements into  $n$  sub-measures. This gives us the following definition:

**Definition 1.** Given  $(X, \mathcal{X})$  a measurable space,  $\mathcal{P}^n(X)$  denotes the set of admissible distributions of  $n$  species defined by

$$\mathcal{P}^n(X) = \left\{ \mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} \in \mathcal{M}_+^n(X) : \sum_{i=1}^n m_i \in \mathcal{P}(X) \right\}.$$

**Remark 2.** Rather than talking about colours, the article talked about species:  $\mu$  can represent an initial distribution of predators  $\mu_{predator}$  and preys  $\mu_{prey}$  such that  $\mu = \mu_{predator} + \mu_{prey}$  and  $\nu = \nu_{predator} + \nu_{prey}$  represents then a final distribution of such predators and preys.

We need to specify which couplings are allowed. We have two choices:

1. Blue sand can only be transported to blue sand and red sand can only be transported to red sand. This imposes an equality between the different masses. We then have two different costs and as many transport plans.
2. The sand of colour  $A \in \{\text{blue}, \text{red}\}$  is authorised to be transported into the sand of colour  $B \in \{\text{blue}, \text{red}\}$ . We then have four different costs and as many transport plans which are those associated with the transports from  $A$  to  $A$ ,  $A$  to  $B$ ,  $B$  to  $A$  and  $B$  to  $B$ .

I went for the second one because I wish I propose a model that allowed mixing between species. Also, the first model is a special case of the second one as soon as we take an infinite cost between  $A$  and  $B$  and  $B$  and  $A$ . To take the previous example with the colours, a coupling between  $\mu = \mu_1 + \mu_2$  and  $\nu = \nu_1 + \nu_2$  is thus described by the concatenation of four transport plans  $\gamma_1, \gamma_{12}, \gamma_{21}, \gamma_{22}$  associated to as many as costs  $c_{11}, c_{12}, c_{21}, c_{22}$ : for example  $\gamma_{12}$  describes which part of 1 is transported to 2 at a cost  $c_{12}$ . Generalizing it to  $n$  colours/species gives us:

**Definition 3.** Given  $\mu \in \mathcal{P}^n(X)$  and  $\nu \in \mathcal{P}^n(Y)$ ,  $\Pi(\mu, \nu)$  denotes the set of transference

plans between  $\mu$  and  $\nu$  defined by

$$\Pi(\mu, \nu) = \left\{ \begin{array}{l} \gamma \in \mathcal{M}_+^{n \times n}(X \times Y) : \forall (i, j) \in \llbracket 1, n \rrbracket^2, \\ \mu_i = \pi_1 \# \left( \sum_{k=1}^n \gamma_{ik} \right) \\ \nu_j = \pi_2 \# \left( \sum_{k=1}^n \gamma_{kj} \right) \end{array} \right\},$$

for a such  $\gamma \in \Pi(\mu, \nu)$ ,  $K(\gamma)$ , the total transportation cost, is given by

$$K(\gamma) = \sum_{(i,j) \in \llbracket 1, n \rrbracket^2} \iint_{X \times Y} c_{ij}(x, y) d\gamma_{ij}(x, y)$$

and the minimization problem is

$$\mathcal{K}(\mu, \nu) = \inf \{K(\gamma) : \gamma \in \Pi(\mu, \nu)\}.$$

This is a linear programming problem whose dual problem is studied in the article. In particular, a strong duality result between them is established. Once the model was well defined, I asked myself on conditions to ask on the costs  $(c_{ij})_{1 \leq i, j \leq n}$  to obtain a distance on the space  $\mathcal{P}^n(X)$ . First we ask to each  $c_{ij} \stackrel{\text{def}}{=} d_{ij}$  to be a distance on  $X$ , and then it is necessary and sufficient to impose a condition that I have called *Mixed Triangle Inequalities* which is given by :

$$\forall (i, j, k) \in \llbracket 1, n \rrbracket^3, \forall (x, y, z) \in X^3, d_{ik}(x, z) \leq d_{ij}(x, y) + d_{jk}(y, z). \quad (\text{MTI})$$

In particular, it is to allowed for a grain of **blue** sand located in  $x$  and going to  $z$  to change into **red** for a low cost and to make the journey  $x \rightarrow z$  for a sum as low (take  $x = y$  and  $d_{12}(x, x) + d_{21}(x, z) < d_{12}(x, z)$  in the previous inequality). These different considerations led me to the following definition and theorem.

**Definition 4.** Given  $\mu, \nu \in \mathcal{P}_p^n(X)$ , the  $p$ -transportation distance between  $\mu$  and  $\nu$  is defined by

$$\begin{aligned} W_p(\mu, \nu) &= \left( \inf \left\{ \sum_{ij} \iint d_{ij}(x, y)^p d\gamma_{ij}(x, y), \gamma \in \Pi(\mu, \nu) \right\} \right)^{\frac{1}{p}} \\ &:= (\mathcal{K}(\mu, \nu))^{\frac{1}{p}}. \end{aligned}$$

**Theorem 5.** Let  $X$  be a Polish space. Let  $p \in [1, \infty[$ . Let  $(d_{ij})$  be  $n^2$  functions on  $X \times X$  valued in  $\mathbb{R}_+$  such that:

- [1]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2, d_{ij}$  is symmetric,
- [2] (MTI) is satisfied,
- [3]  $\forall i \in \llbracket 1, n \rrbracket, \forall x \in X, d_{ii}(x, x) = 0,$
- [4]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j, \forall (x, y) \in X \times Y, d_{ij}(x, y) \neq 0.$

$W_p$  is a distance on  $\mathcal{P}_p^n(X)$ .

Finally, I have proposed numerical simulations to illustrate the paper, whose calculation is based on the entropy relaxation method and the Sinkhorn algorithm become famous in Optimal Transport by Cuturi's paper [22]. This is the subject of the Appendix and constitutes an original work which has not been published.

**Similar works** Chen et al. propose several formulations of Optimal Transport of vector-valued measures in [11], [17] and [14]. These are based on the dynamic formulation of Benamou-Brenier and they establish strong duality results as well as numerical simulations.

K.J. Ciosmak has proposed several models of Optimal Transport between  $\mathbb{R}^m$ -valued measures. His second model developed in [20] initially aims at studying the equality cases in the matrix Hölder inequality. For this he develops a Beckmann formulation (zero divergence, see chapter 4 of [50] for more details) of the Optimal Transport between vector-valued measures: given  $\mu = \mu_+ - \mu_- \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^m)$ , he is interested in the following problem:

$$\inf \{ \|M\|_{\mathcal{M}} : M \in \mathcal{M}(\mathbb{R}^n, \mathbb{R}^{n \times m}), -\operatorname{div} M = \mu \}$$

where  $\|\cdot\|_{\mathcal{M}}$  denotes the total variation norm. As in the scalar case, this problem admits a dual formulation which is given by the following maximization problem:

$$\sup \left\{ \int \langle F | d\mu \rangle : \|f\|_{C^1} \leq 1 \right\}.$$

Compared to the model proposed in chapter 1, this is a generalisation of the particular case of Optimal Transport with cost  $L^1$ .

Wang and Zhang proposed in [58] a model called Simultaneous Optimal Transport. We give  $\mu \in \mathcal{M}(X, \mathbb{R}_+^d)$  and  $\nu \in \mathcal{M}(Y, \mathbb{R}_+^d)$  such that  $\mu_i(X) \geq \nu_i(Y)$  for all  $i = 1, \dots, d$  (thus not sharing necessarily the same mass) and they interpret  $\nu$  as a demand and  $\mu$  as a supply, supply are allowed to be higher than the demand. A formulation *à la Monge* is proposed by means of a function  $T : X \rightarrow Y$  verifying  $T\#\mu \geq \nu$  and which is interpreted like this: in  $x \in X$  there is only one means of transport and this one will transport with it the totality of the resources (all the  $d\mu_i(x), i = 1, \dots, d$ ) to the destination  $T(x)$ . It

is also required that the demand  $\nu$  be satisfied. A formulation *à la Kantorovich* is also proposed. Here it is a question of considering the measures  $\kappa \in \mathcal{M}(X \times Y, \mathbb{R}_+)$  verifying:

$$\int_X \kappa(x, \cdot) d\mu(x) \geq \nu,$$

which has as a consequence that such a transport distributes equitably the different resources  $d\mu_i(x)$ ,  $i = 1, \dots, d$  present at each point.

The authors notice that such transports do not always exist, independently of their optimality. They then try to find sufficient conditions for the problem to be well posed and establish a strong duality result for their model.

Let us finally note that Gover and Wolansky in [28] treat a model close to this one in the *semi-discrete* case, i.e. when the arrival measure is discrete.

## 2.2 Multi-species Optimal Transportation, X. Bacon, *Journal of Optimization Theory and Applications*, 2020

## Multi-Species Optimal Transportation

Xavier Bacon

Received: date / Accepted: date

Communicated by Gabriel Peyré

**Abstract** Given two  $n$ -dimensional measures on Polish spaces, we propose an optimal transportation's formulation, inspired by classical Kantorovitch's formulation in the scalar case. In particular, we established a strong duality result and as a consequence, optimality conditions are investigated. Wasserstein's metrics induced by our formulation are also investigated.

**Keywords** Optimal Transport · Calculus of variations · Wasserstein distance

**Mathematics Subject Classification (2000)** 49J35 · 49K27 · 49K35

---

Xavier Bacon

Université Paris 1 Panthéon-Sorbonne, SAMM

Paris, France

xavier.bacon@etu.univ-paris1.fr

## 1 Introduction

In the present paper, we give a formulation of optimal transportation between vector-valued positive measures. Starting from the classical scalar optimal transportation between two probabilities, we decompose them in  $n$  parts and we ask ourselves what could be a satisfactory formulation of optimal transportation between these two decompositions. We first propose a linear primal problem and then a dual one. Links between these two are established, following the classical theory of convex analysis. The primal formulation is based on the Kantorovitch's point of view and then makes sense in Polish spaces. To achieve the primal formulation, if we think that each part represents a species, we have to declare the transport cost between each species. In other words, we allowed a transport between different coordinates. The complexity of this problem appears when we introduce a new  $\mathbf{c}$ -transformation. This transformation is the main tool to establish a strong duality result, following the scalar case. As often, a metric structure for the total cost can be deduced, these metrics are subject to the last section.

## 2 Preliminaries and Notations

## 2.1 Preliminaries

Starting from the article of G. Monge [1], many mathematical formulations of optimal transportation have been offered (see [2, 3, 4]). In Monge's formulation, given two Polish spaces  $X$  and  $Y$ , if  $\mu$  (resp.  $\nu$ ) is a Borelian probability on  $X$  (resp.  $Y$ ) and if  $c : X \times Y \rightarrow \mathbb{R}$ , then the Monge's formulation consists in a minimization of the total cost among all Borelian maps which push forward  $\mu$  to  $\nu$ , more precisely for  $T$  a Borelian function between  $X$  and  $Y$  and  $m$  a positive measure on  $X$ ,  $T\#\mu$  stands for the push forward measure which is the measure on  $Y$  defined for all measurable set  $B$  by  $T\#\mu(B) := \mu[T^{-1}(B)]$ . Let  $M(\mu, \nu)$  be the set of such maps, Monge transportation problem is then

$$\mathcal{M}(\mu, \nu) := \inf \left\{ \int_X c[x, T(x)] d\mu(x) : T \in M(\mu, \nu) \right\}. \quad (1)$$

In the middle of the 20th century, L. Kantorovitch proposed a relaxation of (1) in [2] by allowing mass splitting. Thinking of  $\mu$  and  $\nu$  as piles of sands, grains located at  $x$  can be sent at different places at the same time. Formally, Kantorovitch's problem consists in minimizing a new total cost among all transference plans  $\gamma \in \Pi(\mu, \nu)$ , where  $\Pi(\mu, \nu)$  is the set of couplings between  $\mu$  and  $\nu$  *i.e.*  $\gamma \in \Pi(\mu, \nu)$  if for all  $A$  Borelian subset of  $X$ ,  $\gamma(A \times Y) = \mu(A)$  and for all  $B$  Borelian subset of  $Y$ ,  $\gamma(X \times B) = \nu(B)$ . Kantorovitch's transportation problem is then

$$\mathcal{K}(\mu, \nu) := \inf \left\{ \iint_{X \times X} c(x, y) d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\} \quad (2)$$

and for reasons that are discussed below, (2) is more accurate to extend the classical theory to vector-valued measures.

When  $c$  is the power of a distance, these two problems induce a metric on the set of probabilities, called here Wasserstein metric (see [5, 6, 7, 8]). In the recent years, generalization of classical optimal transportation have been proposed, such as multimarginal transportation (see [9, 10]) and its application the density functional theory (see [11, 12]), partial optimal transport (see [13, 14]), unbalanced transport and entropy-transport problems (see [15, 16, 17, 18, 19, 20]). Notice also that optimal transportation of signed measures (see [21]), matricial and tensorial measures (see [22, 23, 24]) or vector-valued densities in [25] have already been investigated.

In the present paper, we propose an extension to vector-valued measures. This one is deeply based on Kantorovich's formulation of *scalar* optimal transportation (Section 3). Given two probabilities  $\mu$  and  $\nu$  and two decompositions of them (say)  $\mu = \mu_1 + \dots + \mu_n$  and  $\nu = \nu_1 + \dots + \nu_n$ , more than a transportation between  $\mu$  and  $\nu$ , we are interested in a description of a transportation between these two decompositions. A naive strategy would be to study the  $n$  subproblems of classical optimal transportation between  $\mu_i$  and  $\nu_i$  for  $i \in \llbracket 1, n \rrbracket$ , assuming that for all  $i$ ,  $\mu_i$  and  $\nu_i$  share the same mass. If this new transportation problem leads to a metric, then the topology induced is the product one, due to the independance of each *phasis*. This problem has been explored in [26, 27] and more recently in [28]. To remove the independance of each phasis, we allow transformation similarly as explored in [29]. Introducing  $n^2$  particular costs  $c_{ij}$  and  $n^2$  particular transference plans  $\gamma_{ij}$  which describe the transport of a piece of  $\mu_i$  into a piece of  $\nu_j$ , we consider that the cost



transportation to move  $d\mu_i(x)$  to  $d\nu_j(y)$  is  $c_{ij}(x, y)d\gamma_{ij}(x, y)$ . Compatibility constraints are described by the set  $\mathbf{II}(\boldsymbol{\mu}, \boldsymbol{\nu})$  where we ask that the  $n^2$  transport plans  $\gamma_{ij}$  clear each  $\mu_i$  and fill each  $\nu_j$ . The new minimization problem is given by

$$\inf \left\{ \sum_{(i,j) \in [1,n]^2} \iint_{X \times Y} c_{ij}(x, y) d\gamma_{ij}(x, y) \text{ s.t. } \boldsymbol{\gamma} \in \mathbf{II} \left( \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \begin{bmatrix} \nu_1 \\ \vdots \\ \nu_n \end{bmatrix} \right) \right\}.$$

In Section 3, we give an existence result for this problem as well as various examples. Then, following the shipper's problem interpretation of optimal transportation from L. Caffarelli (presented in [6]), we introduce a dual formulation in Section 4 and prove strong duality theorem. As a consequence of the duality, optimality conditions for primal-dual optimizers are derived. Finally, assuming that costs  $(c_{ij})$  are all the same power of different distances, a metric on vector-valued measures is presented in Section 5.

## 2.2 Notations

In this article, we distinguish vectorial objects from scalar ones by using bold type character like  $\mathbf{II}$  for the first one and non-bold type character like  $II$  for the latter one.

- Given  $X$  a measurable space,  $\mathcal{P}(X)$  stands for the set of probability measures on  $X$  and for all  $n \in \mathbb{N}^*$ ,  $\mathcal{M}^n(X)$  (resp.  $\mathcal{M}^{n \times n}(X)$ ) refers to the set of vectorial measure on  $(X, \mathcal{X})$  valued in  $\mathbb{R}^n$  (resp. in  $\mathbb{R}^{n \times n}$ ) meaning that

each coordinate is a signed measure.  $\mathcal{M}_+^n(X)$  (resp.  $\mathcal{M}_+^{n \times n}(X)$ ) stands for the subset where each coordinate is positive measure. Recall that for  $T$  a measurable function between  $X$  and  $Y$  and  $m$  a positive measure on  $X$ ,  $T\#m$  stands for the push-forward measure which is the image measure of  $m$  through the map  $T$ .

- Given  $X_1 \times \cdots \times X_n$  a product space and  $k \in \llbracket 1, n \rrbracket$ ,  $\pi_k$  denotes the canonical projection on  $X_k$  *i.e.*,

$$\begin{aligned} \pi_k : X_1 \times \cdots \times X_n &\longrightarrow X_k \\ (x_1, \cdots, x_n) &\mapsto x_k \end{aligned}$$

and for  $l \in \llbracket 1, n \rrbracket$  and  $l > k$ ,  $\pi_{k,l}$  denotes the canonical projection on  $X_k \times X_l$  *i.e.*,

$$\begin{aligned} \pi_{k,l} : X_1 \times \cdots \times X_n &\longrightarrow X_k \times X_l \\ (x_1, \cdots, x_n) &\mapsto (x_k, x_l) \end{aligned}$$

- For  $A$  a borelian subset of  $\mathbb{R}$ ,  $\mathcal{L}_A$  stands for the Lebesgue measure on  $A$ . If  $m, M \in \mathcal{M}_+(X)$  satisfy for all  $A \in \mathcal{X}$ ,  $m(A) \leq M(A)$ ,  $m$  is called a submeasure of  $M$  and this property will be written  $m \leq M$ . Note that being a submeasure of  $M$  implies the absolute continuity w.r.t.  $M$ .
- Given  $(X, \mathcal{T})$  a topological space and  $(Y, d)$  a metric space,  $C_b(X, Y)$  refers to the set of bounded continuous functions between  $(X, \mathcal{T})$  and  $(Y, d)$ .
- Given  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\mathbf{E}_{ij}$  refers to the matrix  $n \times n$  whose coordinates are all equal to 0 except  $(i, j)$  which is equal to 1.
- Given a set  $X$  and  $S$  a subset of  $X$ ,  $\iota_S$  denotes the indicatrix function of  $S$  in the convex analysis sense.

- The notation  $\wedge$  will be used to denote the minimum of two reals, and  $\vee$  for the maximum.

### 3 Kantorovitch's Problem

#### 3.1 Presentation

In the remainder of the paper,  $n$  will denote an element of  $\mathbb{N}^*$ .

**Definition 3.1** Given  $(X, \mathcal{X})$  a measurable space,  $\mathcal{P}^n(X)$  denotes the set of admissible distributions of  $n$  species defined by

$$\mathcal{P}^n(X) = \left\{ \mathbf{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix} \in \mathcal{M}_+^n(X) : \sum_{i=1}^n m_i \in \mathcal{P}(X) \right\}.$$

It is straightforward that  $\mathcal{P}^n(X)$  is a non-empty convex subset of  $\mathcal{M}^n(X)$ .

Inspired by Kantorovitch's formulation of optimal transportation, an extension of the notion of transference plan between two scalar measures is now proposed. For a well understanding of the next definition, let us make a short digression and present our model. Given  $(X, \mathcal{X})$  and  $(Y, \mathcal{Y})$  two measurable spaces and  $\boldsymbol{\mu} \in \mathcal{P}^n(X), \boldsymbol{\nu} \in \mathcal{P}^n(Y)$  two distributions of  $n$  species, if we transport each coordinate separately our problem will have no solution at all in most case since the total amount of each species differs. Then, transformations between species are allowed. Given  $(i, j) \in \llbracket 1, n \rrbracket^2$ , the "transportation"

(with "transformation" if  $i \neq j$ ) of a piece of  $\mu_i$  into a piece of  $\nu_j$  is described by a transference plan  $\gamma_{ij} \in \mathcal{M}_+(X \times Y)$ . The constraints on  $\gamma = (\gamma_{ij})_{1 \leq i, j \leq n}$  are given by

$$\begin{aligned} (\text{Clear } \boldsymbol{\mu}) \quad \forall i \in \llbracket 1, n \rrbracket, \forall A \in \mathcal{X}, \mu_i(A) &= \sum_{k=1}^n \gamma_{ik}(A \times Y) \\ (\text{Fill } \boldsymbol{\nu}) \quad \forall j \in \llbracket 1, n \rrbracket, \forall B \in \mathcal{Y}, \nu_j(B) &= \sum_{k=1}^n \gamma_{kj}(X \times B), \end{aligned}$$

or in other words, for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\sum_{k=1}^n \gamma_{ik}$  has  $\mu_i$  as first marginal and  $\sum_{k=1}^n \gamma_{kj}$  has  $\nu_j$  as second marginal. This naturally leads to the following definition.

**Definition 3.2** Given  $\boldsymbol{\mu} \in \mathcal{P}^n(X)$  and  $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ ,  $\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$  denotes the set of transference plans between  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$  defined by

$$\boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \left\{ \gamma \in \mathcal{M}_+^{n \times n}(X \times Y) : \forall (i, j) \in \llbracket 1, n \rrbracket^2, \begin{cases} \mu_i = \pi_1 \# \left( \sum_{k=1}^n \gamma_{ik} \right) \\ \nu_j = \pi_2 \# \left( \sum_{k=1}^n \gamma_{kj} \right) \end{cases} \right\}.$$

**Remark 3.1.** According to the Definition 3.2, every  $\gamma \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$  induces a canonical transference plan (for  $n = 1$ , the two definitions of transference plan are the same) between  $\sum_{i=1}^n \mu_i$  and  $\sum_{j=1}^n \nu_j$  given by  $\sum_{i,j=1}^n \gamma_{ij}$ . However the converse is not true since given  $\gamma \in \boldsymbol{\Pi} \left( \sum_{i=1}^n \mu_i, \sum_{j=1}^n \nu_j \right)$  and  $(x, y) \in X \times Y$ , there is still a choice to make: is the first species sent into the first or the second one or both? And in what proportions? Let us give a short example to clarify this remark. Taking

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathcal{L}_{[-1,0]} \\ \mathcal{L}_{[-1,0]} \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \mathcal{L}_{[0,1]} \\ \mathcal{L}_{[0,1]} \end{bmatrix}$$

and writing  $\tau_1 : x \rightarrow x + 1$ , it is known that  $\gamma = (I, \tau_1) \# (\mu_1 + \mu_2)$  is a transference plan between  $\mathcal{L}_{[-1,0]} = \mu_1 + \mu_2$  and  $\mathcal{L}_{[0,1]} = \nu_1 + \nu_2$ . Given a such  $\gamma$ ,  $\mu_1$  can be sent towards  $\nu_1$ , or towards  $\nu_2$ . A mix is even possible and  $\mu_1$  can be sent towards  $\mathbb{1}_{[0, \frac{1}{2}]} d\nu_1 + \mathbb{1}_{[\frac{1}{2}, 1]} d\nu_2$ . In other words, the following matrix measures are transference plans,

$$\begin{aligned} & \begin{bmatrix} (I, \tau_1) \# \mu_1 & 0 \\ 0 & (I, \tau_1) \# \mu_2 \end{bmatrix}, \\ & \begin{bmatrix} 0 & (I, \tau_1) \# \mu_1 \\ (I, \tau_1) \# \mu_2 & 0 \end{bmatrix}, \\ & \frac{1}{2} \begin{bmatrix} (I, \tau_1) \# \mu_1 & (I, \tau_1) \# \mu_1 \\ (I, \tau_1) \# \mu_2 & (I, \tau_1) \# \mu_2 \end{bmatrix}. \end{aligned}$$

We also introduce matrix-valued cost  $\mathbf{c}$  as a function from  $X \times Y \rightarrow \mathcal{M}_n(\mathbb{R})$ , integrable w.r.t.  $\gamma$  or positive measurable. The associated total cost is given by the following definition.

**Definition 3.3** Given  $\gamma \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$  and a cost matrix  $\mathbf{c}$ ,  $K(\gamma)$  denotes the total transportation cost according to  $\gamma$  defined by

$$K(\gamma) = \sum_{(i,j) \in [1,n]^2} \iint_{X \times Y} c_{ij}(x, y) d\gamma_{ij}(x, y).$$

The Kantorovich's transportation problem between two distributions of  $n$  species  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$  for  $\mathbf{c}$  is given by

$$\inf \{K(\gamma) : \gamma \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})\} =: \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}) \in [-\infty, +\infty] \quad (\text{KP})$$

**Example 3.1.** Note that if  $c_{ij} = c$  for all  $(i, j) \in \llbracket 1, n \rrbracket^2$  then (KP) shares the same value as the *scalar* optimal transportation between  $\sum_{i=1}^n \mu_i$  and  $\sum_{j=1}^n \nu_j$  for the cost  $c$ . The most simple example of non trivial matrix cost is given by the following one: let  $c$  be a *scalar* cost and  $\kappa$  be a real and define the following matrix cost:

$$\forall (x, y) \in X \times Y, \mathbf{c}(x, y) = \begin{bmatrix} c(x, y) & c(x, y) + \kappa \\ c(x, y) + \kappa & c(x, y) \end{bmatrix}.$$

In other words, a constant cost is required for any transformation. See the example 4.1 below for a study of this special cost.

**Example 3.2.** Let  $\mathbf{p}, \mathbf{q} \in \Delta_n := \{\mathbf{x} \in \mathbb{R}_+^n, \sum_{i=1}^n x_i = 1\}$  and  $\mathbf{x}, \mathbf{y} \in X^n$ . Define

$$\boldsymbol{\mu} = \begin{bmatrix} p_1 \delta_{x_1} \\ \vdots \\ p_n \delta_{x_n} \end{bmatrix}, \boldsymbol{\nu} = \begin{bmatrix} q_1 \delta_{y_1} \\ \vdots \\ q_n \delta_{y_n} \end{bmatrix}.$$

Let us first notice that for all  $(i, j)$ , since  $\text{supp}(\sum_{l=1}^n \gamma_{il}) \subseteq \text{supp}(\mu_i) \times Y$  and  $\text{supp}(\sum_{l=1}^n \gamma_{lj}) \subseteq X \times \text{supp}(\nu_j)$  then  $\text{supp}(\gamma_{ij}) \subseteq (x_i, y_j)$  and hence we have  $\gamma_{ij} = t_{ij} \delta_{(x_i, y_j)}$  for some  $t_{ij} \in [0, 1]$ . Constraints on  $\gamma$  give us that for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\sum_{l=1}^n t_{il} = p_i$  and  $\sum_{l=1}^n t_{lj} = q_j$  and finally

$$\mathbf{K}(\boldsymbol{\gamma}) = \sum_{(i,j) \in \llbracket 1, n \rrbracket^2} t_{ij} c_{ij}(x_i, y_j),$$

(KP) becomes in that case

$$\inf \left\{ \sum_{(i,j) \in \llbracket 1, n \rrbracket^2} t_{ij} c_{ij}(x_i, y_j), \mathbf{t} \in \mathcal{M}_n(\mathbb{R}) \text{ s.t. } \sum_{l=1}^n t_{il} = p_i, \sum_{l=1}^n t_{lj} = q_j \right\}$$

which reduces to the discret optimal transportation.

### 3.2 Existence of a Minimizer

Let  $X$  and  $Y$  be two Polish spaces. In this subsection, we prove an existence result for the problem (KP). Arguments used to establish it are the same as in the scalar case (see [6, 7] for instance). Let us first gather the main properties of problem (KP).

**Lemma 3.1** *Given  $\mu \in \mathcal{P}^n(X)$ ,  $\nu \in \mathcal{P}^n(Y)$  and  $c$  a cost matrix, the following assertions are satisfied:*

[1]  $\Pi(\mu, \nu)$  is a non-empty convex subset of  $\mathcal{M}^{n \times n}(X \times Y)$ .

[2]  $\Pi(\mu, \nu)$  is a weakly sequentially compact<sup>1</sup> subset of  $\mathcal{M}^{n \times n}(X \times Y)$ .

[3] If for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $c_{ij}$  is bounded from below, then the restriction of  $K$  to  $\Pi(\mu, \nu)$  is bounded from below.

[4] If for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $c_{ij}$  is l.s.c. and bounded from below then the restriction of  $K$  to  $\Pi(\mu, \nu)$  is weakly l.s.c. with respect to the tight convergence.

#### Proof

[1] Convexity is clear and it is easy to check that  $\left( \frac{\mu_i \otimes \nu_j}{n} \right)_{(i,j) \in \llbracket 1, n \rrbracket^2}$  belongs to  $\Pi(\mu, \nu)$ .

---

<sup>1</sup> w.r.t. the test function space  $C_b(X \times Y, \mathbb{R}^{n \times n})$ .

[2] Let  $(\gamma^k)_{k \in \mathbb{N}} \in \mathbf{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})^{\mathbb{N}}$  and  $(i, j) \in \llbracket 1, n \rrbracket^2$ . We claim that  $(\gamma_{ij}^k)_{k \in \mathbb{N}}$  is tight.

Indeed, let  $\varepsilon \in \mathbb{R}_+^*$  and  $K_X$  (resp.  $K_Y$ ) a compact of  $X$  (resp.  $Y$ ) such that<sup>2</sup>

$$\forall i \in \llbracket 1, n \rrbracket, \mu_i(X \setminus K_X) \leq \varepsilon \text{ (resp. } \forall j \in \llbracket 1, n \rrbracket, \nu_j(Y \setminus K_Y) \leq \varepsilon). \quad (3)$$

Let  $k \in \mathbb{N}$ , the following inequalities are satisfied,

$$\begin{aligned} \gamma_{ij}^k [(X \times Y) \setminus (K_X \times K_Y)] &\leq \gamma_{ij}^k [(X \setminus K_X) \times Y] + \gamma_{ij}^k [X \times (Y \setminus K_Y)] \\ &\leq \sum_{l=1}^n \gamma_{il}^k [(X \setminus K_X) \times Y] + \sum_{\bar{l}=1}^n \gamma_{i\bar{l}}^k [X \times (Y \setminus K_Y)] \\ &= \mu_i(X \setminus K_X) + \nu_j(Y \setminus K_Y) \text{ since } \gamma \in \mathbf{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) \\ &\leq 2\varepsilon. \end{aligned}$$

This proves the claim and thanks to Prokhorov theorem, there exists a non-negative finite measure on  $X \times Y$ ,  $\gamma_{ij}^\infty$  and a subsequence of  $(\gamma_{ij}^k)_{k \in \mathbb{N}}$  (still written  $(\gamma_{ij}^k)_{k \in \mathbb{N}}$ ) such that  $(\gamma_{ij}^k)_{k \in \mathbb{N}}$  tightly converges towards  $\gamma_{ij}^\infty$ . In order to conclude, we only have to check that  $\gamma^\infty \in \mathbf{Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ . Let  $\phi \in C_b(X \times Y, \mathbb{R})$  and notice that for all  $i \in \llbracket 1, n \rrbracket$  and  $k \in \mathbb{N}$ ,

$$\int_X \phi(x) d\mu_i(x) = \sum_{l=1}^n \iint_{X \times Y} \phi(x) d\gamma_{il}^k(x, y) \rightarrow \sum_{l=1}^n \iint_{X \times Y} \phi(x) d\gamma_{il}^\infty(x, y)$$

.

[3] Straightforward.

[4] Let  $(\gamma^k)_{k \in \mathbb{N}} \in \mathbf{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})^{\mathbb{N}}$  and  $\gamma^\infty \in \mathbf{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$  such that  $(\gamma^k)_{k \in \mathbb{N}}$  tightly converges towards  $\gamma^\infty$  in that for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $(\gamma_{ij}^k)_{k \in \mathbb{N}}$  weakly converges

<sup>2</sup> These two compacts exist: all  $\mu_i$  are finite measure on Polish spaces then there exists  $K_X^i$  verifying these inequalities and then we just have to take union of them.



in duality with  $C_b$  towards  $\gamma_{ij}^\infty$ . Then, by lower semi-continuity of  $\langle c_{ij}, \cdot \rangle$  (see [7], Lemma 1.6), for all  $(i, j) \in \llbracket 1, \dots, n \rrbracket^2$ ,

$$\iint_{X \times Y} c_{ij}(x, y) \, d\gamma_{ij}^\infty(x, y) \leq \liminf_{k \rightarrow \infty} \iint_{X \times Y} c_{ij}(x, y) \, d\gamma_{ij}^k(x, y)$$

and since sum of lim inf is less or equal to lim inf of sum, it ends the proof.  $\square$

With these facts in hand, our main result easily follows.

**Theorem 3.1** *Given  $\mathbf{c}$  a cost matrix such that for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $c_{ij}$  is bounded from below and l.s.c., there exists  $\gamma \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$  such that  $K(\gamma) = \mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu})$ .*

### Proof

By the direct method of Calculus of Variations and lemma 3.1, the problem admits a solution.  $\square$

## 4 Duality and Optimality Conditions

### 4.1 Presentation

In this section, we look for a dual formulation of (KP). In order to find it, consider the following situation<sup>3</sup>: mines full of different metals ( $n$  kinds) and refineries ( $n$  kinds) are distributed in space. For each kind of metal corresponds a kind of refinery, for instance a kind refinery for iron, a kind of refinery for

<sup>3</sup> This interpretation is due to L. Caffareli in the scalar case, according to [6].

gold etc. On the one hand we want to minimize the travel cost *i.e.* minimize the associated Kantorovich's problem, on the other hand a character suggests to supervise the travelling operation for us and propose that contract: for each ton of metal  $i$  located in  $x$ , its price will be  $\varphi_i(x)$  to extract it and for each ton of metal  $j$  located in  $y$  its price will be  $\psi_j(y)$  to drop it off. To guarantee our interest, its constraints will be that for all  $(i, j)$  and  $(x, y)$ ,

$$\varphi_i(x) + \psi_j(y) \leq c_{ij}(x, y).$$

All these considerations suggest to give the following definitions.

**Definition 4.1** Given  $\mathbf{c}$  a cost matrix,  $\Delta(\mathbf{c})$  denotes the set of potential couples for the cost  $\mathbf{c}$  defined by

$$\Delta(\mathbf{c}) = \left\{ \begin{array}{l} \varphi_i \in C_b(X) \\ \left[ \begin{array}{c} \varphi \\ \psi \end{array} \right] = \left[ \begin{array}{c} \varphi_1 \cdots \varphi_n \\ \psi_1 \cdots \psi_n \end{array} \right], \forall (i, j) \in \llbracket 1, n \rrbracket^2, \psi_j \in C_b(Y) \\ \varphi_i \oplus \psi_j \leq c_{ij} \end{array} \right\}$$

and, if there is no ambiguity on  $\mathbf{c}$ , we will write  $\Delta$  instead of  $\Delta(\mathbf{c})$ .

**Definition 4.2** Given  $\mu \in \mathcal{P}^n(X), \nu \in \mathcal{P}^n(Y)$ ,  $\mathbf{c}$  a cost matrix and  $\left[ \begin{array}{c} \varphi \\ \psi \end{array} \right] =$

$\left[ \begin{array}{c} \varphi_1 \cdots \varphi_n \\ \psi_1 \cdots \psi_n \end{array} \right] \in \Delta(\mathbf{c})$ ,  $D(\varphi, \psi)$  denotes the dual cost of  $\left[ \begin{array}{c} \varphi \\ \psi \end{array} \right]$  defined by

$$D(\varphi, \psi) = \sum_{i=1}^n \int_X \varphi_i(x) d\mu_i(x) + \sum_{j=1}^n \int_Y \psi_j(y) d\nu_j(y). \quad (4)$$

Finally, the dual transportation problem is given  $\boldsymbol{\mu} \in \mathcal{P}^n(X)$ ,  $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$  and a cost matrix  $\mathbf{c}$ ,

$$\sup \left\{ D(\boldsymbol{\varphi}, \boldsymbol{\psi}) : \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\mathbf{c}) \right\} =: \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu}) \in [-\infty, +\infty] \quad (\text{DP})$$

We establish now a weak duality result.

**Proposition 4.1.** Given  $\boldsymbol{\mu} \in \mathcal{P}^n(X)$ ,  $\boldsymbol{\nu} \in \mathcal{P}^n(Y)$ , a cost matrix  $\mathbf{c}$ ,  $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$

and  $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\mathbf{c})$ , the following inequality is satisfied,

$$D(\boldsymbol{\varphi}, \boldsymbol{\psi}) \leq \mathbf{K}(\boldsymbol{\gamma}).$$

**Proof**

Let  $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$  and  $\begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\mathbf{c})$ . Compute:

$$\begin{aligned} D(\boldsymbol{\varphi}, \boldsymbol{\psi}) &= \sum_{i=1}^n \int_X \varphi_i \, d\mu_i + \sum_{j=1}^n \int_Y \psi_j \, d\nu_j \\ &= \sum_{i=1}^n \iint_{X \times Y} \varphi_i \, d \left( \sum_{j=1}^n \gamma_{ij} \right) + \sum_{j=1}^n \iint_{X \times Y} \psi_j \, d \left( \sum_{i=1}^n \gamma_{ij} \right). \end{aligned}$$

The last equality coming from the fact that  $\boldsymbol{\gamma} \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu})$ . And then,

$$\begin{aligned} D(\boldsymbol{\varphi}, \boldsymbol{\psi}) &\leq \sum_{(i,j) \in [1,n]^2} \iint_{X \times Y} c_{ij} \, d\gamma_{ij} \text{ since } \begin{bmatrix} \boldsymbol{\varphi} \\ \boldsymbol{\psi} \end{bmatrix} \in \Delta(\mathbf{c}) \\ &= \mathbf{K}(\boldsymbol{\gamma}). \end{aligned}$$

This concludes the proof.  $\square$

## 4.2 An Extension of $c$ -Transformation

In order to prove that (DP) is attained, at least in compact case, we propose an extension of the classical  $c$ -transform (see the recall below). First, we make a short digression about modulus of continuity in order to fix conventions.

**Definition 4.3** Given  $(X, d)$  a metric space and  $f : X \rightarrow \mathbb{R}$ , a uniform modulus of continuity for  $f$  according to  $d$  is a function  $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  such that the following conditions are satisfied:

$$[1] \lim_{t \downarrow 0^+} \omega(t) = 0$$

$$[2] \forall (x, x') \in X^2 : |f(x) - f(x')| \leq \omega [d(x, x')].$$

**Remark 4.1.** If  $f$  admits a uniform modulus of continuity  $\omega_f$  and  $g$  admits a uniform modulus of continuity  $\omega_g$  then  $\omega_f + \omega_g$  is a uniform modulus of continuity for  $\min(f, g)$ .

Recall that when  $f$  is a function between  $X$  (resp.  $Y$ ) and  $\mathbb{R} \cup \{-\infty\}$  and  $c$  a cost function, we can define its  $c$ -transform  $f^c$  (resp.  $\bar{c}$ -transform) by:

$$\begin{aligned} f^c : Y &\rightarrow \mathbb{R} \cup \{-\infty, +\infty\} \\ y &\mapsto \inf \{c(x, y) - f(x) : x \in X\} \\ \left( \begin{array}{l} \text{resp.} \\ f^c : X &\rightarrow \mathbb{R} \cup \{-\infty, +\infty\} \\ x &\mapsto \inf \{c(x, y) - f(y) : y \in Y\} \end{array} \right) \end{aligned}$$

We introduce a new transformation and to motivate it just remark than in our case, we have  $2n$  potentials and  $n^2$  inequalities in the dual formulation. A naive idea would be to first substitute  $\varphi_1$  by  $\psi_1^{c_{11}}$  but there is no guarantee

that our new couple of potentials  $\begin{bmatrix} \psi_1^{c_{11}} & \varphi_2 & \cdots & \varphi_n \\ \psi_1 & \cdots & \cdots & \psi_n \end{bmatrix}$  will still be in  $\Delta(\mathbf{c})$ . The following definition answers this problem.

**Definition 4.4** Given  $\mathbf{f} = (f_1, \dots, f_n) : X \rightarrow (\mathbb{R} \cup \{-\infty\})^n$  and a function  $\mathbf{c} = (c_1, \dots, c_n) : X \times Y \rightarrow (\mathbb{R} \cup \{+\infty\})^n$ ,  $\mathbf{f}^{\mathbf{c}}$  (resp.  $\mathbf{f}^{\bar{\mathbf{c}}}$ ) denotes the  $\mathbf{c}$ -transform of  $\mathbf{f}$  (resp.  $\bar{\mathbf{c}}$ -transform of  $\mathbf{f}$ ) defined by

$$\begin{aligned} \forall y \in Y : \mathbf{f}^{\mathbf{c}}(y) &= \min(f_1^{c_1}(y), \dots, f_n^{c_n}(y)) \\ (\text{resp. } \forall x \in X : \mathbf{f}^{\bar{\mathbf{c}}}(x) &= \min(f_1^{\bar{c}_1}(x), \dots, f_n^{\bar{c}_n}(x))) \end{aligned}$$

All benefits of this transformation is contained in the next proposition.

**Proposition 4.2.** Let  $\mathbf{f} = (f_1, \dots, f_n) : X \rightarrow (\mathbb{R} \cup \{-\infty\})^n$  and let  $\mathbf{c} = (c_1, \dots, c_n) : X \times Y \rightarrow (\mathbb{R} \cup \{+\infty\})^n$ , then

[1] Following inequalities are satisfied,

$$\forall i \in \llbracket 1, n \rrbracket, f_i \oplus \mathbf{f}^{\mathbf{c}} \leq c_i \quad (5)$$

$$\forall j \in \llbracket 1, n \rrbracket, \mathbf{f}^{\bar{\mathbf{c}}} \oplus f_j \leq c_j \quad (6)$$

[2] If  $h : Y \rightarrow \mathbb{R} \cup \{-\infty\}$  is such that for all  $i \in \llbracket 1, n \rrbracket, f_i \oplus h \leq c_i$  then  $h \leq \mathbf{f}^{\mathbf{c}}$ . If  $h : X \rightarrow \mathbb{R} \cup \{-\infty\}$  is such that for all  $j \in \llbracket 1, n \rrbracket, h \oplus g_j \leq c_j$  then  $h \leq \mathbf{f}^{\bar{\mathbf{c}}}$ .

**Proof**

[1] Let  $(i, j) \in \llbracket 1, n \rrbracket^2$  and  $(x, y) \in X \times Y$ . Since  $f_i(x) + f_i^{c_i}(y) \leq c_i(x, y)$  and  $\mathbf{f}^{\mathbf{c}} \leq f_i^{c_i}$  the first inequality is deduced and note that the second inequality can be proved following the same way.

[2] If a such function exists, we deduce from  $f_i \oplus h \leq c_i$  that for all  $(x, y)$  belonging to  $X \times Y$ ,  $h(y) \leq c_i(x, y) - f_i(x)$ , then take infimum with respect to  $x$  and arbitrary on  $i$  concludes for the first inequality. The same proof also works for the second inequality.  $\square$

We will show next that this process is a natural way to improve the dual cost while staying in the constraint  $\Delta(\mathbf{c})$ , at least in compact case and continuous costs. Moreover, it provides a common uniform modulus of continuity for all the potentials.

**Lemma 4.1** *Let  $X, Y$  be two compact metric spaces,  $\mathbf{c}$  a continuous cost matrix and  $(\varphi, \psi) \in \Delta(\mathbf{c})$ . There exists  $(\underline{\varphi}, \underline{\psi}) \in \Delta(\mathbf{c})$  such that*

$$[1] \quad D(\varphi, \psi) \leq D(\underline{\varphi}, \underline{\psi}).$$

[2]  $\underline{\varphi}_1, \dots, \underline{\varphi}_n, \underline{\psi}_1, \dots, \underline{\psi}_{n-1}$  and  $\underline{\psi}_n$  admit a common uniform modulus of continuity which depends only on  $\mathbf{c}$ .

**Proof**

First, make the following substitutions:

$$\forall j \in \llbracket 1, n \rrbracket : \psi_j \leftarrow \varphi^{(c_{1j}, \dots, c_{nj})} := \underline{\psi}_j,$$

then, thanks to Proposition 4.2,  $(\varphi, \underline{\psi}) \in \Delta(\mathbf{c})$  and  $D(\varphi, \psi) \leq D(\varphi, \underline{\psi})$ .

Denoting  $\omega_{c_{ij}}$  a uniform modulus of continuity of  $c_{ij}$  for  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\omega_{c_{ij}}$  is also a uniform modulus of continuity of  $\varphi_i^{c_{ij}}$  according to [7] (Box. 1.8).

Thanks to Lemma 4.1, we conclude that  $\omega_{\underline{\psi}_j} = \omega_{c_{1j}} + \dots + \omega_{c_{nj}}$  is a uniform modulus of continuity of  $\underline{\psi}_j$ . Then, make the following substitutions:

$$\forall i \in \llbracket 1, n \rrbracket : \varphi_i \leftarrow \underline{\psi}^{(\overline{c_{i1}, \dots, c_{in}})} := \underline{\varphi}_i$$

and of course the new couple of potentials is still in  $\Delta(c)$  and the dual cost is increased. To conclude, we just have to check that  $\sum_{1 \leq i, j \leq n} \omega_{c_{ij}}$  is a common uniform modulus of continuity for  $(\underline{\varphi}, \underline{\psi})$ , which is clear.  $\square$

**Example 4.1.** Coming back to the example 3.1, let us compute this new  $\mathbf{c}$ -transform to reduce the problem. Fix  $\kappa$  to be strictly positive and assume that  $X = Y$  and  $c$  is symmetric (then,  $\mathbf{c}$ -transform is equivalent to  $\bar{\mathbf{c}}$ -transform).

Constraints of (DP) are given by the following system:

$$\begin{cases} \varphi_1(x) + \psi_1(y) \leq c(x, y) \\ \varphi_1(x) + \psi_2(y) \leq c(x, y) + \kappa \\ \varphi_2(x) + \psi_1(y) \leq c(x, y) + \kappa \\ \varphi_2(x) + \psi_2(y) \leq c(x, y) \end{cases}$$

**First step:** it is easy to check that:

$$\begin{aligned} (f_1, f_2)^{c, c+\kappa} &= [f_1 \wedge (f_2 - \kappa)]^c \\ (f_1, f_2)^{c+\kappa, c} &= [(f_1 - \kappa) \wedge f_2]^c, \end{aligned}$$

then make the following substitutions:

$$\begin{aligned} \psi_1 &\leftarrow (\varphi_1, \varphi_2)^{c, c+\kappa} = [\varphi_1 \wedge (\varphi_2 - \kappa)]^c =: \tilde{\psi}_1 \\ \psi_2 &\leftarrow (\varphi_1, \varphi_2)^{c+\kappa, c} = [(\varphi_1 - \kappa) \wedge \varphi_2]^c =: \tilde{\psi}_2 \end{aligned}$$

**Second step:** following the proof below, we make the following substitutions:

$$\begin{aligned}
\varphi_1 \leftarrow (\tilde{\psi}_1, \tilde{\psi}_2)^{c, c+\kappa} &= \left[ \tilde{\psi}_1 \wedge (\tilde{\psi}_2 - \kappa) \right]^c \\
&= \tilde{\psi}_1^c \vee (\tilde{\psi}_2 - \kappa)^c \text{ since } (\sup_{\alpha} f_{\alpha})^c = \inf_{\alpha} f_{\alpha}^c \\
&= [\varphi_1 \wedge (\varphi_2 - \kappa)]^{cc} \vee [(\varphi_1 - \kappa) \wedge \varphi_2]^{cc} \\
&= [\varphi_1 \wedge (\varphi_2 - \kappa)]^{cc} \vee [(\varphi_1 - 2\kappa) \wedge (\varphi_2 - \kappa)]^{cc} \\
&= [\varphi_1 \wedge (\varphi_2 - \kappa)]^{cc} \text{ since if } f \leq g \text{ then } g^c \leq f^c \\
&= \tilde{\psi}_1^c \\
\varphi_2 \leftarrow (\tilde{\psi}_1, \tilde{\psi}_2)^{c+\kappa, c} &= \tilde{\psi}_2^c \text{ for the same reasons.}
\end{aligned}$$

When  $c = d$  is a distance, according to [7] (Proposition 3.1):

$$\left\{ \begin{array}{l} \tilde{\psi}_1(y) - \tilde{\psi}_1(x) \leq d(x, y) \\ \tilde{\psi}_2(y) - \tilde{\psi}_1(x) \leq d(x, y) + \kappa \\ \tilde{\psi}_1(y) - \tilde{\psi}_2(x) \leq d(x, y) + \kappa \\ \tilde{\psi}_2(y) - \tilde{\psi}_2(x) \leq d(x, y), \end{array} \right.$$

which is equivalent to the following system, thanks to the symmetry of  $d$ :

$$\left\{ \begin{array}{l} |\tilde{\psi}_1(x) - \tilde{\psi}_1(y)| \leq d(x, y) \\ |\tilde{\psi}_1(x) - \tilde{\psi}_2(y)| \leq d(x, y) + \kappa \\ |\tilde{\psi}_2(x) - \tilde{\psi}_2(y)| \leq d(x, y) \end{array} \right.$$

*i.e.*  $(\tilde{\psi}_1, \tilde{\psi}_2)$  are solution to the system below if and only if they are 1-Lipschitz

w.r.t. to  $d$  and satisfy  $\|\tilde{\psi}_1 - \tilde{\psi}_2\|_{\infty} \leq \kappa$ .



### 4.3 Existence of a Maximizer

**Theorem 4.1** *Given  $X$  et  $Y$  two compact metric spaces,  $\mu \in \mathcal{P}^n(X)$ ,  $\nu \in \mathcal{P}^n(Y)$  and  $\mathbf{c}$  a continuous cost matrix, there exists  $(\varphi, \psi) \in \Delta(\mathbf{c})$  such that  $\mathcal{D}(\mu, \nu) = D(\varphi, \psi)$ .*

#### Proof

The constraint set is non-empty since  $\mathbf{c}$  is bounded by below (continuous on compact). Let:

$$\begin{bmatrix} \varphi^k \\ \psi^k \end{bmatrix}_{k \in \mathbb{N}} = \begin{bmatrix} \varphi_1^k \cdots \varphi_n^k \\ \psi_1^k \cdots \psi_n^k \end{bmatrix}_{k \in \mathbb{N}}$$

be a maximizing sequence for (DP). According to Lemma 4.1, we may assume that our  $2n$  sequences share a common uniform modulus of continuity. We now prove that the sequence is uniformly bounded with respect to  $n$ . Indeed, setting for all  $k \in \mathbb{N}$ :

$$m_k := \min \left[ \inf_{x \in X} \varphi_1^k(x), \dots, \inf_{x \in X} \varphi_n^k(x) \right],$$

and since  $m_k$  is finite, we can substitute:

$$\begin{aligned} \forall i \in \llbracket 1, \dots, n \rrbracket : \varphi_i^k &\leftarrow \varphi_i^k - m_k \text{ still written } \varphi_i^k \\ \forall j \in \llbracket 1, \dots, n \rrbracket : \psi_j^k &\leftarrow \psi_j^k + m_k \text{ still written } \psi_j^k, \end{aligned}$$

and these new potentials are still admissible, have the same dual cost and for all  $i \in \llbracket 1, n \rrbracket, k \in \mathbb{N}, \varphi_i^k \geq 0$ . Therefore we have:

$$\forall i \in \llbracket 1, n \rrbracket, k \in \mathbb{N} : \varphi_i^k \leq \omega [\text{diam}(X)],$$

which concludes the case of  $\varphi$ . Next, let us make the new following substitutions:

$$\forall j \in \llbracket 1, n \rrbracket : \psi_j \leftarrow \overline{\varphi^{(c_{1j}, \dots, c_{nj})}} \text{ still written } \psi_j.$$

We have for all  $y \in Y, j \in \llbracket 1, n \rrbracket$  and  $k \in \mathbb{N}$ ,

$$\begin{aligned} & \min(c_{1j}, \dots, c_{nj}) - \omega[\text{diam}(X)] \\ & \leq \min \left[ \inf_{x \in X} c_{1j}(x, y) - \varphi_1^k(x), \dots, \inf_{x \in X} c_{nj}(x, y) - \varphi_n^k(x) \right] := \psi_j^k(y) \\ \text{and } \psi_j^k(y) & := \min \left[ \inf_{x \in X} c_{1j}(x, y) - \varphi_1^k(x), \dots, \inf_{x \in X} c_{nj}(x, y) - \varphi_n^k(x) \right] \\ & \leq \max(c_{1j}, \dots, c_{nj}), \end{aligned}$$

which leads to the conclusion on  $\psi$ . Finally, the Ascoli-Arzelà theorem applied to each sequence provides the existence of a continuous couple

$$\begin{bmatrix} \varphi^\infty \\ \psi^\infty \end{bmatrix} = \begin{bmatrix} \varphi_1^\infty & \dots & \varphi_n^\infty \\ \psi_1^\infty & \dots & \psi_n^\infty \end{bmatrix}$$

which belong to  $\Delta(\mathbf{c})$  thanks to pointwise convergence and  $D(\varphi^\infty, \psi^\infty) = \mathcal{D}(\boldsymbol{\mu}, \boldsymbol{\nu})$  thanks to uniform convergence on finite measure sets.  $\square$

#### 4.4 Strong Duality

We establish a strong duality result. The proof follows the one of strong duality theorem for scalar optimal transportation described in [7], Chapter 1.

**Definition 4.5** Given  $\mu \in \mathcal{P}^n(X)$ ,  $\nu \in \mathcal{P}^n(Y)$  and  $\mathbf{c}$  a cost matrix, we denote by  $H$  the value function of the perturbed dual problem, *i.e.*

$$\forall \varepsilon \in C(X \times Y, \mathbb{R}^{n \times n}), H(\varepsilon) = \sup \left\{ D(\varphi, \psi) : \begin{bmatrix} \varphi \\ \psi \end{bmatrix} \in \Delta(\mathbf{c} - \varepsilon) \right\}$$

**Lemma 4.2** *Let  $X$  and  $Y$  be two metric compact spaces.  $H$  satisfy the following properties:*

[1]  *$H$  is concave.*

[2] *Suppose that  $\mathbf{c}$  is continuous, then  $H$  is u.s.c. with respect to the uniform norm.*

**Proof**

[1] Let  $t \in [0, 1]$ ,  $\varepsilon^0 \in C(X \times Y, \mathbb{R}^{n \times n})$  (resp.  $\varepsilon^1 \in C(X \times Y, \mathbb{R}^{n \times n})$ ) and let  $(\varphi^0, \psi^0)$  (resp.  $(\varphi^1, \psi^1)$ ) be optimal in (DP) associated to  $\mathbf{c} - \varepsilon^0$  (resp. to  $\mathbf{c} - \varepsilon^1$ ). Note that they exist thanks to the existence result below. Define  $\varepsilon_t = (1-t)\varepsilon^0 + t\varepsilon^1$ ,  $\varphi_t = (1-t)\varphi^0 + t\varphi^1$ ,  $\psi_t = (1-t)\psi^0 + t\psi^1$ . Therefore  $(\varphi_t, \psi_t)$  is admissible for the dual problem associated to  $\mathbf{c} - \varepsilon_t$  and then by definition of  $H$  we have

$$H(\varepsilon_t) \geq D(\varphi_t, \psi_t) = (1-t)H(\varepsilon^0) + tH(\varepsilon^1)$$

And the conclusion follows.

[2] Let  $(\varepsilon^k)_{k \in \mathbb{N}} \in C(X \times Y, \mathbb{R}^{n \times n})^{\mathbb{N}}$  and  $\varepsilon^\infty \in C(X \times Y, \mathbb{R}^{n \times n})$  such that for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\varepsilon_{ij}^k \xrightarrow{\|\cdot\|_\infty} \varepsilon_{ij}^\infty$ . Let  $(\varepsilon^k)_{k \in \mathbb{N}}$  be a sequence and let  $(\varepsilon^{k_l})_{l \in \mathbb{N}}$  be a subsequence satisfying for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\limsup_k H(\varepsilon_{ij}^k) = \lim_l H(\varepsilon_{ij}^{k_l})$ . Ascoli-Arzelà theorem ensures that for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $(\varepsilon_{ij}^{k_l})_{l \in \mathbb{N}}$  are equicontinuous

and uniformly bounded with respect to  $l$  therefore chose the corresponding optimal potentials  $(\varphi^{k_l}, \psi^{k_l})$  equicontinuous and uniformly bounded in  $l$ , thanks to the Ascoli-Arzelà theorem again, there is a uniform convergence towards (say)  $(\varphi^\infty, \psi^\infty)$  up to an extraction, therefore thanks to pointwise convergence there is for all  $(i, j) \in \llbracket 1, n \rrbracket^2$ ,  $\varphi_i^\infty \oplus \psi_j^\infty \leq c_{ij} - \varepsilon_{ij}$  and so:

$$H(\varepsilon) \geq D(\varphi^\infty, \psi^\infty) = \lim_l H(\varepsilon^{k_l}) = \limsup_k H(\varepsilon^k).$$

This concludes the proof.  $\square$

Finally, the strong duality theorem follows.

**Theorem 4.2** *Suppose that  $X$  and  $Y$  are both metric compact spaces and that  $c$  is continuous, then for all  $(\mu, \nu) \in \mathcal{P}^n(X) \times \mathcal{P}^n(Y)$ ,  $\mathcal{K}(\mu, \nu) = \mathcal{D}(\mu, \nu)$ .*

**Proof**

Let  $(\mu, \nu) \in \mathcal{P}^n(X) \times \mathcal{P}^n(Y)$ , since  $(-H)$  is convex and l.s.c. and according to the Fenchel-Moreau theorem, we have:

$$\begin{aligned} \mathcal{D}(\mu, \nu) &= H(\mathbf{0}) \\ &= -[-H(\mathbf{0})] \\ &= -[-H]^{**}(\mathbf{0}) \\ &= - \sup_{\gamma \in \mathcal{M}^{n \times n}(X \times Y)} \langle \mathbf{0}, \gamma \rangle - [-H]^*(\gamma) \\ &= \inf_{\gamma \in \mathcal{M}^{n \times n}(X \times Y)} [-H]^*(\gamma). \end{aligned}$$

Next, we compute for all  $\gamma \in \mathcal{M}^{n \times n}(X \times Y)$ ,

$$\begin{aligned} [-H]^*(\gamma) &= \sup_{\varepsilon} \left( \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \varepsilon_{ij} d\gamma_{ij} + \sup_{\varphi \oplus \psi \leq \mathbf{c} - \varepsilon} \sum_{i=1}^n \int_X \varphi_i d\mu_i + \sum_{j=1}^n \int_Y \psi_j d\nu_j \right) \\ &= \sup_{\varepsilon} \left( \sup_{\varphi \oplus \psi \leq \mathbf{c} - \varepsilon} \left[ \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \varepsilon_{ij} d\gamma_{ij} + \sum_{i=1}^n \int_X \varphi_i d\mu_i + \sum_{j=1}^n \int_Y \psi_j d\nu_j \right] \right). \end{aligned}$$

If there exists  $(i_0, j_0)$  such that  $\gamma_{i_0 j_0} \notin \mathcal{M}_+(X \times Y)$  and  $\varepsilon_{i_0 j_0}^0$  such that

$$\iint_{X \times Y} \varepsilon_{i_0 j_0}^0 d\gamma_{i_0 j_0} > 0,$$

then take  $(\varepsilon_{i_0 j_0}^k)_{k \in \mathbb{N}^*}$  such that  $\varepsilon_{i_0 j_0}^k = c_{i_0 j_0} + k\varepsilon_{i_0 j_0}^0$  for  $k \in \mathbb{N}^*$  and take  $\varphi_{i_0} = 0$  and  $\psi_{j_0} = 0$ . Then putting all the other potentials equals at the value 0 and find  $(\varepsilon_{ij})$  such that all the constraints are still satisfied ( $\mathbf{c}$  is bounded), we get  $[-H]^*(\gamma) = +\infty$  if  $\gamma \notin \mathcal{M}_+^{n \times n}(X \times Y)$ . Now, suppose that  $\gamma \in \mathcal{M}_+^{n \times n}(X \times Y)$ , when  $(\varphi, \psi)$  are fixed, we are interested in taking the largest  $\varepsilon_{ij}$  possible for

every  $(i, j) \in \llbracket 1, n \rrbracket^2$ , that is  $\varepsilon_{ij} = c_{ij} - \varphi_i - \psi_j$  and we get

$$\begin{aligned}
[-H]^*(\gamma) &= \sup_{\varphi \oplus \psi \leq C - \varepsilon} \sum_{1 \leq i, j \leq n} \iint_{X \times Y} c_{ij} - \varphi_i - \psi_j \, d\gamma_{ij} \\
&\quad + \sum_{i=1}^n \int_X \varphi_i \, d\mu_i + \sum_{j=1}^n \int_Y \psi_j \, d\nu_j \\
&= \sup_{(\varphi, \psi)} K(\gamma) + \left( \sum_{i=1}^n \int_X \varphi_i \, d\mu_i - \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \varphi_i \, d\gamma_{ij} \right) \\
&\quad + \left( \sum_{j=1}^n \int_Y \psi_j \, d\nu_j - \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \psi_j \, d\gamma_{ij} \right) \\
&= \sup_{(\varphi, \psi)} K(\gamma) + \sum_{i=1}^n \left( \int_X \varphi_i \, d\mu_i - \iint_{X \times Y} \varphi_i \, d \sum_{j=1}^n \gamma_{ij} \right) \\
&\quad + \sum_{j=1}^n \left( \int_Y \psi_j \, d\nu_j - \iint_{X \times Y} \psi_j \, d \sum_{i=1}^n \gamma_{ij} \right) \\
&= \iota_{\Pi(\mu, \nu)}(\gamma) \text{ according to [7], Lemma 1.45.}
\end{aligned}$$

This ends the proof.  $\square$

#### 4.5 Optimality Conditions

In this subsection,  $X$  and  $Y$  are two metric compact spaces. As a direct consequence of Theorem 3.2, we deduce optimality constraints linking (KP) and (DP).

**Proposition 4.3.** Given  $\gamma \in \Pi(\mu, \nu)$  and  $(\varphi, \psi) \in \Delta(c)$ , the following assertions are equivalent:

- [1]  $\gamma$  is optimal in (KP) and  $(\varphi, \psi)$  is optimal in (DP).

$$[2] \quad \forall (i, j), \varphi_i \oplus \psi_j = c_{ij} \quad \gamma_{ij}\text{-a.e.}$$

**Proof**

If [1] is satisfied, according to Theorem 4.2,  $K(\gamma) = D(\varphi, \psi)$ . We then compute  $D(\varphi, \psi)$  as a function of  $\gamma$ .

$$\begin{aligned} D(\varphi, \psi) &:= \sum_{i=1}^n \int_X \varphi_i(x) d\mu_i(x) + \sum_{j=1}^n \int_Y \psi_j(y) d\nu_j(y) \\ &= \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \varphi_i(x) d\gamma_{ij}(x, y) + \sum_{1 \leq i, j \leq n} \iint_{X \times Y} \psi_j(y) d\gamma_{ij}(x, y) \\ &= \sum_{1 \leq i, j \leq n} \iint_{X \times Y} [\varphi_i(x) + \psi_j(y)] d\gamma_{ij}(x, y). \end{aligned}$$

Comparing the latter expression with  $K(\gamma)$  gives

$$\begin{aligned} 0 &= K(\gamma) - D(\varphi, \psi) \\ &= \sum_{ij} \iint_{X \times Y} (c_{ij}(x, y) - [\varphi_i(x) + \psi_j(y)]) d\gamma_{ij}(x, y). \end{aligned}$$

The conclusion follows from the fact that  $(\varphi, \psi) \in \mathbf{\Delta}(c)$ .

Conversely, if [2] is satisfied, clearly  $K(\gamma) = D(\varphi, \psi)$  which implies that both  $\gamma$  and  $(\varphi, \psi)$  are optimal according to Proposition 4.1.  $\square$

The result above is not surprising since any given  $\gamma \in \Pi(\mu, \nu)$  induces  $n^2$  scalar optimal transportation problems between each marginals (say)  $\pi_1 \# \gamma_{ij} := f_{ij} d\mu_i$  and  $\pi_2 \# \gamma_{ij} := g_{ij} d\nu_j$

$$\inf \left\{ \iint_{X \times Y} c_{ij}(x, y) d\gamma_{ij}(x, y) : \gamma_{ij} \in \Pi(f_{ij} d\mu_i, g_{ij} d\nu_j) \right\}, \quad (KP_{ij})$$

and looking at constraints in vectorial Kantorovitch's problem, it is easy to see that  $\gamma$  has to be optimal in every subproblems  $(KP_{ij})$  to be optimal between  $\mu$  and  $\nu$  (if not, take a better one and compare the total cost, which is nothing less than another proof of the result above).

## 5 Induced Metrics

In this section, we take  $X = Y$  a Polish space. We investigate how to extend the well-known Wasserstein distance and answer the question "does the problem (KP) define a distance on the space  $\mathcal{P}^n(X)$ ?"

Let  $(d_{ij})_{(i,j) \in \llbracket 1, n \rrbracket^2}$  be  $n^2$  finite, symmetric and non negative functions on  $X \times X$  satisfying the triangle inequality (we do not assume that they are distances). Then let  $p \in [1, \infty[$ ,  $x_0 \in X$  and define

$$\mathcal{P}_p^n(X) = \left\{ \boldsymbol{\mu} \in \mathcal{P}^n(X), \forall (i, j) \in \llbracket 1, n \rrbracket^2, \int_X d_{ij}^p(x_0, x) + d_{ji}^p(x_0, x) d\mu_i(x) < \infty \right\}$$

and notice that as in the scalar case, this set does not depend on  $x_0$ .

**Definition 5.1** Given  $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathcal{P}_p^n(X)$ , the  $p$ -transportation distance between  $\boldsymbol{\mu}$  and  $\boldsymbol{\nu}$  is defined by

$$\begin{aligned} W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) &= \left( \inf \left\{ \sum_{ij} \iint d_{ij}(x, y)^p d\gamma_{ij}(x, y), \gamma \in \boldsymbol{\Pi}(\boldsymbol{\mu}, \boldsymbol{\nu}) \right\} \right)^{\frac{1}{p}} \\ &:= (\mathcal{K}(\boldsymbol{\mu}, \boldsymbol{\nu}))^{\frac{1}{p}}. \end{aligned}$$

The symmetry of  $W_p$  is clear provided the costs are symmetric themselves. However, the fact that  $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$  implies that  $\boldsymbol{\mu} = \boldsymbol{\nu}$  is never satisfied if all costs are (power of) distances. In place of it, if  $W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) = 0$  then  $\sum_{i=1}^n \mu_i = \sum_{j=1}^n \nu_j$ . In other words,  $W_p$  is pseudodistance in that case. To prevent that, we add new hypothesis on  $(d_{ij})$  described in the next proposition.

**Proposition 5.1.** Let  $(d_{ij})_{(i,j) \in \llbracket 1, n \rrbracket^2}$  be  $n^2$  symmetric finite non negative functions on  $X \times X$  satisfying the triangle inequality. Assume moreover that



for all  $(i, j) \in \llbracket 1, n \rrbracket^2, i \neq j, d_{ij}$  is strictly positive and  $d_{ii}$  is a distance. Then for all  $\mu, \nu \in \mathcal{P}_p^n(X)$ , if  $W_p(\mu, \nu) = 0$  then  $\mu = \nu$ .

**Proof**

We will follow the proof of Theorem 7.3 in [6]. Let  $\mu, \nu \in \mathcal{P}_p^n(X)$  be such that  $W_p(\mu, \nu) = 0$  and let  $\gamma^*$  be optimal in (KP), then

$$0 = \sum_{k=1}^n \iint_{X^2} d_{kk}(x, y)^p d\gamma_{kk}^*(x, y) + \sum_{i \neq j} \iint_{X^2} d_{ij}(x, y)^p d\gamma_{ij}^*(x, y)$$

According to the strict positivity of non diagonal distances, for all  $i \neq j, \gamma_{ij}^* = 0$  and then for all  $k \in \llbracket 1, n \rrbracket, \gamma_{kk}^*$  is a transport plan between  $\mu_k$  and  $\nu_k$  such that  $\iint d_{kk} d\gamma_{kk} = 0$ . As a consequence  $\gamma_{kk}$  is supported on the diagonal and then for all  $\phi : X \rightarrow \mathbb{R}$  bounded continuous we have  $\int \phi(x) d\mu_k(x) = \iint \phi(x) d\gamma_k(x, y) = \iint \phi(y) d\gamma_k(x, y) = \int \phi(y) d\nu_k(y)$ , which implies  $\mu_k = \nu_k$ .  $\square$  However, without any other constraints on  $(d_{ij})$ , the following example shows that the triangle inequality fails.

**Example 5.1.** Let  $X = \mathbb{R}, n = 2$  and set:

$$\mu = \begin{bmatrix} \delta_0 \\ 0 \end{bmatrix}, \nu = \begin{bmatrix} 0 \\ \delta_1 \end{bmatrix}, \lambda = \begin{bmatrix} \delta_2 \\ 0 \end{bmatrix}$$

Then set  $p = 1, d_{11} = |\cdot|$  and  $d_{12} = d_{21} := d_\varepsilon$  the  $\varepsilon$ -discrete distance on  $\mathbb{R}$  (with  $\varepsilon \in \mathbb{R}_+^*$ ) defined by  $d_\varepsilon(x, y) = \varepsilon$  if  $x = y$  and 0 otherwise and an arbitrary distance for  $d_{22}$ . Clearly,

$$W_1(\mu, \lambda) = 2, W_1(\mu, \nu) = W_1(\nu, \lambda) = \varepsilon \quad (7)$$

And these three numbers do not satisfy to triangle inequality as soon as  $\varepsilon$  is smaller enough.

The main problem in the example above is the lack of comparison between all  $(d_{ij})$ . To give a everyday-life example, it could be more expensive to travel between Paris and Berlin using plane than to first travel between Paris and Amsterdam using car and then going to Berlin from Amsterdam using train. To avoid this phenomenon above, we add new constraints on  $(d_{ij})$ :

$$\forall (i, j, k) \in \llbracket 1, n \rrbracket^3, \forall (x, y, z) \in X^3, d_{ik}(x, z) \leq d_{ij}(x, y) + d_{jk}(y, z) \quad (\text{MTI})$$

and from now on we assume that these constraints are satisfied.

**Remark 5.1.** Note that  $(MTI)$  (for Mixed Triangle Inequalities) contain the fact that all costs satisfy triangle inequality (take  $i = j = k$ ) and if one of these inequalities is false for some  $(x_0, y_0, z_0)$  then one can exhibit a counterexample to let the triangle inequality fail on  $W_p$  similar to the (counter)Example 5.1 above.

**Example 5.2.** An easy way to construct objects that satisfy  $(MTI)$  is (and then, we do not work on empty set) given a distance  $d$  on  $X$  and a non negative scalar  $t$  (for transformation),  $d_{ii} = d$  for all  $i$  and  $d_{ij} = d + t$  for all  $(i, j)$  with  $i \neq j$ .

**Proposition 5.2.** Let  $p \in [1, \infty[$  and  $(d_{ij})_{(i,j) \in \llbracket 1, n \rrbracket^2}$  be such that  $(MTI)$  are satisfied. Then  $W_p$  satisfies the triangle inequality.

### Proof

Let  $\gamma^* = (\gamma_{ij}^*)$  (resp.  $\tilde{\gamma}^* = (\tilde{\gamma}_{jk}^*)$ ) be optimal<sup>4</sup> between  $\mu$  and  $\nu$  (resp.  $\nu$  and  $\lambda$ ). Let  $j \in \llbracket 1, n \rrbracket$  and define for all  $i, k \in \llbracket 1, n \rrbracket$  the marginals  $\nu_j^{i, \leftarrow} := \pi_2 \# \gamma_{ij}^*$

<sup>4</sup> they exist according to 3.1, even if it is not necessary here: passing to supremum bound *a posteriori* otherwise.

and  $\nu_j^{k,\rightarrow} := \pi_1 \# \tilde{\gamma}_{jk}^*$ . These marginals are all submeasures of  $\nu_j$  and then, according to Radon-Nikodym theorem, we denote by  $f_j^{i,\leftarrow}$  (resp.  $f_j^{k,\rightarrow}$ ) the density of  $\nu_j^{i,\leftarrow}$  (resp.  $\nu_j^{k,\rightarrow}$ ) w.r.t.  $\nu_j$ . Finally, define for each  $i, j, k \in \llbracket 1, n \rrbracket$  the following transference plans

$$\gamma_{ijk}^* \text{ is defined as the measure with density } (x, y) \rightarrow f_j^{k,\rightarrow}(y) \text{ w.r.t. } \gamma_{ij}^*, \quad (8)$$

$$\tilde{\gamma}_{ijk}^* \text{ is defined as the measure with density } (y, z) \rightarrow f_j^{i,\leftarrow}(y) \text{ w.r.t. } \tilde{\gamma}_{jk}^*, \quad (9)$$

these definitions imply that

$$\forall (i, j) \in \llbracket 1, n \rrbracket^2, \gamma_{ij}^* = \sum_{k=1}^n \gamma_{ijk}^*, \quad (10)$$

$$\forall (j, k) \in \llbracket 1, n \rrbracket^2, \tilde{\gamma}_{jk}^* = \sum_{i=1}^n \tilde{\gamma}_{ijk}^*, \quad (11)$$

$$\forall (i, j, k) \in \llbracket 1, n \rrbracket^3, \pi_2 \# \gamma_{ijk}^* = \pi_1 \# \tilde{\gamma}_{ijk}^*. \quad (12)$$

To obtain the last equality, fix  $B$  a measurable subset of  $Y$  and compute

$$\begin{aligned} \pi_2 \# \gamma_{ijk}^*(B) &= \gamma_{ijk}^*(X \times B) \\ &= \iint_{X \times B} f_j^{k,\rightarrow}(y) d\gamma_{ij}^*(x, y) \text{ by (8)} \\ &= \int_B f_j^{k,\rightarrow}(y) d\nu_j^{i,\leftarrow}(y) \text{ by definition of } \nu_j^{i,\leftarrow} \\ &= \int_B f_j^{k,\rightarrow}(y) f_j^{i,\leftarrow}(y) d\nu_j(y) \text{ by definition of } f_j^{i,\leftarrow} \\ &= \int_B f_j^{i,\leftarrow}(y) d\nu_j^{k,\rightarrow}(y) \text{ by definition of } f_j^{k,\rightarrow} \\ &= \iint_{B \times Z} f_j^{i,\leftarrow}(y) d\tilde{\gamma}_{ij}^*(y, z) \text{ by definition of } \nu_j^{k,\rightarrow} \\ &= \pi_1 \# \tilde{\gamma}_{ijk}^*(B) \text{ by (9)}. \end{aligned}$$

Then, equalities (12) allow us to apply the Gluing Lemma (see [6], Lemma 7.6) and guarantee the existence of  $\Pi_{ijk}$  a measure on  $X \times Y \times Z$  such that

$$\pi_{1,2}\#\Pi_{ijk} = \gamma_{ijk}^* \text{ and } \pi_{2,3}\#\Pi_{ijk} = \tilde{\gamma}_{ijk}.$$

We next define for all  $(i, k) \in \llbracket 1, n \rrbracket^2$ ,  $\Pi_{ik} = \sum_{j=1}^n \Pi_{ijk}$  and compute

$$\begin{aligned} \pi_1\#\sum_{k=1}^n \Pi_{ik} &= \sum_{k=1}^n \sum_{j=1}^n \pi_1\#\Pi_{ijk} \text{ by definition} \\ &= \sum_{k=1}^n \sum_{j=1}^n \pi_1\#\gamma_{ijk}^* \text{ by (13)} \\ &= \sum_{j=1}^n \pi_1\#\left(\sum_{k=1}^n \gamma_{ijk}^*\right) \\ &= \sum_{j=1}^n \pi_1\#\gamma_{ij}^* \text{ by (10)} \\ &= \mu_i. \end{aligned}$$

For identical reasons  $\pi_3\#\sum_{i=1}^n \Pi_{ik} = \lambda_k$  and as a consequence:

$$\gamma := (\pi_{1,3}\#\Pi_{ik})_{1 \leq i, k \leq n} \in \mathbf{\Pi}(\boldsymbol{\mu}, \boldsymbol{\lambda}).$$

Finally, we have:

$$\begin{aligned}
W_p(\boldsymbol{\mu}, \boldsymbol{\lambda}) &\leq \left( \sum_{ik} \iint d_{ik}(x, z)^p d\gamma_{ik}(x, y, z) \right)^{\frac{1}{p}} \\
&= \left( \sum_{ijk} \iiint d_{ik}(x, z)^p d\Pi_{ijk}(x, y, z) \right)^{\frac{1}{p}} \text{ by definition of } \gamma \text{ and } \Pi \\
&\leq \left( \sum_{ijk} \iiint (d_{ij}(x, y) + d_{jk}(y, z))^p d\Pi_{ijk}(x, y, z) \right)^{\frac{1}{p}} \text{ by (MTI)} \\
&\leq \left( \sum_{ijk} \iiint d_{ij}(x, y)^p d\Pi_{ijk}(x, y, z) \right)^{\frac{1}{p}} + \left( \sum_{ijk} \iiint d_{jk}(y, z)^p d\Pi_{ijk}(x, y, z) \right)^{\frac{1}{p}} \\
&= \left( \sum_{ijk} \iint d_{ij}(x, y)^p d\gamma_{ijk}^*(x, z) \right)^{\frac{1}{p}} + \left( \sum_{ijk} \iint d_{jk}(y, z)^p d\tilde{\gamma}_{ijk}^*(x, y, z) \right)^{\frac{1}{p}} \text{ by (13)} \\
&= \left( \sum_{ij} \iint d_{ij}(x, y)^p d\gamma_{ij}^*(x, y) \right)^{\frac{1}{p}} + \left( \sum_{jk} \iint d_{jk}(y, z)^p d\tilde{\gamma}_{jk}^*(y, z) \right)^{\frac{1}{p}} \\
&= W_p(\boldsymbol{\mu}, \boldsymbol{\nu}) + W_p(\boldsymbol{\nu}, \boldsymbol{\lambda}). \quad \square
\end{aligned}$$

**Theorem 5.1** *Let  $X$  be a Polish space. Let  $p \in [1, \infty[$ . Let  $(d_{ij})$  be  $n^2$  functions on  $X \times X$  valued in  $\mathbb{R}_+$  such that:*

- [1]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2$ ,  $d_{ij}$  is symmetric,
- [2] (MTI) is satisfied,
- [3]  $\forall i \in \llbracket 1, n \rrbracket, \forall x \in X, d_{ii}(x, x) = 0$ .
- [4]  $\forall (i, j) \in \llbracket 1, n \rrbracket^2, i \neq j, \forall (x, y) \in X \times Y, d_{ij}(x, y) \neq 0$ .

$W_p$  is a distance on  $\mathcal{P}_p^n(X)$ .

**Example 5.3.** Coming back to the Example 3.2, let us fix all weights

$$\mathbf{p} = \mathbf{q} = \frac{1}{n} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad (13)$$

and given  $x, y \in X^n$ , we define the distance  $w_p(\mathbf{x}, \mathbf{y})$  between them by

$$w_p(\mathbf{x}, \mathbf{y}) = W_p \left( \frac{1}{n} \begin{bmatrix} \delta_{x_1} \\ \vdots \\ \delta_{x_n} \end{bmatrix}, \frac{1}{n} \begin{bmatrix} \delta_{y_1} \\ \vdots \\ \delta_{y_n} \end{bmatrix} \right).$$

According to 3.2,  $w_p$  is given by

$$\begin{aligned} w_p(\mathbf{x}, \mathbf{y})^p &= \inf \left\{ \sum_{(i,j) \in [1,n]^2} t_{ij} d_{ij}(x_i, y_j)^p, \mathbf{t} \in \mathcal{M}_n(\mathbb{R}), \sum_{l=1}^n t_{il} = \frac{1}{n}, \sum_{l=1}^n t_{lj} = \frac{1}{n} \right\} \\ &= \frac{1}{n} \min \left\{ \sum_{(i,j) \in [1,n]^2} t_{ij} d_{ij}(x_i, y_j)^p, \mathbf{t} \text{ bistochastic matrix } n \times n \right\}, \end{aligned}$$

the last equality providing from classical arguments of linear programming. This example show a way to define new distances on a finite product of spaces using  $n^2$  distances.

## 6 Conclusions

The aim of this paper was to present a new point of view in vector-valued optimal transportation. Writing this paper, we discover in [29, 24] that these authors suggest to use the same idea to treat this problem and allowed mixing of species. Their point of view follows a dynamical formulation of optimal transportation, similar to Beckmann's formulation (presented in [7]) which is a mean to transport quantities such that this transport obeys to Kirchhoff's laws. While in our paper, Kantorovitch's point of view of optimal transportation was our approach angle.

Concerning this approach angle, let us make another small digression about Monge's optimal transportation's problem and present it. Given two proba-

bilities  $\mu$  and  $\nu$  we are interested in knowing if the optimal tranference plan between  $\mu$  and  $\nu$  split mass *i.e* if (formaly) the support of  $\gamma^*$  is included in a function's graph, say  $T^*$ . A natural question here is if there exists a similar problem associated to  $(KP)$ ? A possible answer is given  $\gamma \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu})$  we declare that  $\gamma$  has a *Monge's form* if for all  $(i, j)$ ,  $\gamma_{ij}$  is included in a function's graph say  $T_{ij}$ . The main problem here is given  $T_{ij}$ , it is not possible to build the associated  $\gamma_{ij}$ . Indeed, the knowledge of  $T_{ij}$  does not include which parts of  $\mu_i$  is transported into the  $j$ -nth specie and we don't have also the knowledge of the part  $\nu_j$  which receives  $\mu_i$ . In other words we have still to fix  $f_{ij}$  the part of  $\mu_i$  sent into  $\nu_j$  and  $g_{ij}$  the part of  $\nu_j$  where  $\mu_i$  is sent. This remark makes hard to use only entropic relaxations (see [30]) to solve our problem. To understand that, let us remind that entropic relaxations is a mean to solve linear problems in finite dimensions, for instance optimal transportation problems between convex combinations of Dirac measures. In our problem if we know how to connect parts of  $\mu_i$  and parts of  $\nu_j$  to obtain an optimal transport plan  $\gamma$ , in other word if we know the optimal  $f_{ij}^*$  and  $g_{ij}^*$ , we can use entropic relaxation to optimize in these  $n^2$  subproblems. However, obtaining these  $f_{ij}^*$  and  $g_{ij}^*$  is still an open problem.

**Acknowledgements** The author would like to thank B. Nazaret for his advices and his re-reading of the paper and L. Glaudin for his abilities in  $\text{\LaTeX}$ .

## References

1. Monge, G.: Mémoire sur la théorie des déblais et des remblais: Histoire de l'academieroyale des sciences (1781)
2. Kantorovitch, L.: On the translocation of masses. *Management Science* **5**(1), 1–4 (1958)
3. Beckmann, M.: A continuous model of transportation. *Econometrica: Journal of the Econometric Society* pp. 643–660 (1952)
4. Benamou, J.D., Brenier, Y.: A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik* **84**(3), 375–393 (2000)
5. Ambrosio, L., Gigli, N., Savaré, G.: Gradient flows in metric spaces and in the space of probability measures. *Lectures in Mathematics ETH Zürich*, Birkhäuser Verlag, Basel (2005)
6. Villani, C.: *Topics in optimal transportation*. 58. American Mathematical Soc. (2003)
7. Santambrogio, F.: *Optimal transport for applied mathematicians*. Birkäuser, NY (2015)
8. Villani, C.: *Optimal transport: old and new*, vol. 338. Springer Science & Business Media (2008)
9. Carlier, G., Nazaret, B.: Optimal transportation for the determinant. *ESAIM: Control, Optimisation and Calculus of Variations* **14**(4), 678–698 (2008)
10. Kitagawa, J., Pass, B.: The multi-marginal optimal partial transport problem. In: *Forum of Mathematics, Sigma*, vol. 3. Cambridge University Press (2015)
11. Buttazzo, G., De Pascale, L., Gori-Giorgi, P.: Optimal-transport formulation of electronic density-functional theory. *Physical Review A* **85**(6), 062,502 (2012)
12. Cotar, C., Friesecke, G., Pass, B.: Infinite-body optimal transport with coulomb cost. *Calculus of Variations and Partial Differential Equations* **54**(1), 717–742 (2015)
13. Caffarelli, L., McCann, R.J.: Free boundaries in optimal transport and monge-ampere obstacle problems. *Annals of mathematics* **171**(2), 673–730 (2010)
14. Figalli, A.: The optimal partial transport problem. *Archive for rational mechanics and analysis* **195**(2), 533–560 (2010)



15. Liero, M., Mielke, A., Savaré, G.: Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae* **211**(3), 969–1117 (2018)
16. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: Unbalanced optimal transport: geometry and kantorovich formulation (2015)
17. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.X.: An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics* **18**(1), 1–44 (2018)
18. Kondratyev, S., Monsaingeon, L., Vorotnikov, D., et al.: A new optimal transport distance on the space of finite radon measures. *Advances in Differential Equations* **21**(11/12), 1117–1164 (2016)
19. Benamou, J.D., Carlier, G., Cuturi, M., Nenna, L., Peyré, G.: Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing* **37**(2), A1111–A1138 (2015)
20. Benamou, J.D., Carlier, G., Nenna, L.: Generalized incompressible flows, multi-marginal transport and sinkhorn algorithm. *Numerische Mathematik* **142**(1), 33–54 (2019)
21. Ambrosio, L., Mainini, E., Serfaty, S.: Gradient flow of the chapman–rubinstein–schatzman model for signed vortices. In: *Annales de l’IHP Analyse non linéaire*, vol. 28, pp. 217–246 (2011)
22. Carlen, E.A., Maas, J.: An analog of the 2-wasserstein metric in non-commutative probability under which the fermionic fokker–planck equation is gradient flow for the entropy. *Communications in mathematical physics* **331**(3), 887–926 (2014)
23. Chen, Y., Georgiou, T.T., Ning, L., Tannenbaum, A.: Matricial wasserstein-1 distance. *IEEE control systems letters* **1**(1), 14–19 (2017)
24. Ryu, E.K., Chen, Y., Li, W., Osher, S.: Vector and matrix optimal mass transport: theory, algorithm, and applications. *SIAM Journal on Scientific Computing* **40**(5), A3675–A3698 (2018)
25. Zinsl, J., Matthes, D.: Transport distances and geodesic convexity for systems of degenerate diffusion equations. *Calculus of Variations and Partial Differential Equations*

- 54**(4), 3397–3438 (2015)
26. Benamou, J.D., Brenier, Y., Guittet, K.: Numerical resolution of a multiphasic optimal mass transport problem. Ph.D. thesis, INRIA (2000)
27. Benamou, J.D., Brenier, Y., Guittet, K.: Numerical analysis of a multi-phasic mass transport problem. *Contemporary Mathematics* **353**, 1–18 (2004)
28. Lavenant, H.: Time-convexity of the entropy in the multiphasic formulation of the incompressible euler equation. *Calculus of Variations and Partial Differential Equations* **56**(6), 170 (2017)
29. Chen, Y., Georgiou, T.T., Tannenbaum, A.: Vector-valued optimal mass transport. *SIAM Journal on Applied Mathematics* **78**(3), 1682–1696 (2018)
30. Nenna, L.: Numerical methods for multi-marginal optimal transportation. Ph.D. thesis, PSL Research University (2016)

## 2.3 Appendix : numerical results

**Theory.** In this section, an approach to numerically simulate the previous model is proposed, based on entropic relaxation and the Sinkhorn algorithm. We place ourselves in a discrete framework, where the spaces  $X = \{x_1, \dots, x_N\}$  and  $Y = \{y_1, \dots, y_M\}$  are given by two finite collections of points. In the following, we will write without ambiguity

$$\begin{aligned} m[i] &\stackrel{\text{def}}{=} m(x_i) \\ M[i, j] &\stackrel{\text{def}}{=} M(x_i, y_j). \end{aligned}$$

We give ourselves  $n$  source measures and  $m$  target measures (note that in the previous article we restricted ourselves to  $n = m$ ):

$$\begin{aligned} \boldsymbol{\mu} = (\mu_1, \dots, \mu_n) &\in \overbrace{\mathbb{R}^N \times \dots \times \mathbb{R}^N}^{n \text{ times}} \quad (\text{source distributions}) \\ \boldsymbol{\nu} = (\nu_1, \dots, \nu_m) &\in \underbrace{\mathbb{R}^M \times \dots \times \mathbb{R}^M}_{m \text{ times}} \quad (\text{target distributions}) \end{aligned}$$

and also  $n \times m$  costs:

$$\mathbf{C} = (C_{ij})_{1 \leq i \leq n, 1 \leq j \leq m} \in [\mathbb{R}^{N \times M}]^{n \times m}.$$

Given a regularization parameter  $\varepsilon > 0$ , the regularized problem is given by :

$$\mathcal{K}_\varepsilon(\boldsymbol{\mu}, \boldsymbol{\nu}) \stackrel{\text{def}}{=} \min \left\{ K(\boldsymbol{\gamma}) - \sum_{ij} \varepsilon H(\gamma_{ij}), \boldsymbol{\gamma} \in \Pi(\boldsymbol{\mu}, \boldsymbol{\nu}) \right\} \quad (2.1)$$

where  $H$  is the (concave) entropy of transport plan given by

$$H(\boldsymbol{\gamma}) \stackrel{\text{def}}{=} \begin{cases} - \sum_{k,l} \gamma[k, l] (\log \gamma[k, l] - 1) \\ -\infty \text{ otherwise.} \end{cases}$$

Thus, (2.1) is rewritten:

$$\begin{aligned} \min_{\boldsymbol{\gamma}=(\gamma_{ij})} \max_{\boldsymbol{\varphi}=(\varphi_i), \boldsymbol{\psi}=(\psi_j)} \sum_{i,j} \left( \langle C_{ij} | \gamma_{ij} \rangle - \varepsilon H(\gamma_{ij}) \right) - \sum_i \left\langle \left( \sum_j \gamma_{ij} \right) \cdot \mathbb{1}_n - \mu_i \mid \varphi_i \right\rangle \\ - \sum_j \left\langle \left( \sum_i \gamma_{ij}^T \right) \cdot \mathbb{1}_m - \nu_j \mid \psi_j \right\rangle \end{aligned}$$

and differentiating with respect to  $\gamma_{ij}$  we obtain

$$-C_{ij} = \varepsilon \log(\gamma_{ij}) - \varphi_i - \psi_j = 0$$

then

$$\gamma_{ij} = \exp\left(\frac{\varphi_i}{\varepsilon}\right) \exp\left(-\frac{C_{ij}}{\varepsilon}\right) \exp\left(\frac{\psi_j}{\varepsilon}\right), \quad (2.2)$$

where the functions  $\log$  and  $\exp$  act term by term (on a matrix or a vector according to circumstance). Then we differentiate with respect to  $\varphi_i$  and  $\psi_j$  and find marginal constraints, written here

$$\mu_i = \left(\sum_j \gamma_{ij}\right) \cdot \mathbb{1}_n \quad \text{et} \quad \nu_j = \left(\sum_i \gamma_{ij}^T\right) \cdot \mathbb{1}_m \quad (2.3)$$

then setting  $u_i \stackrel{\text{def}}{=} \exp\left(\frac{\varphi_i}{\varepsilon}\right)$ ,  $v_j \stackrel{\text{def}}{=} \exp\left(\frac{\psi_j}{\varepsilon}\right)$  and  $K_{ij} \stackrel{\text{def}}{=} \exp\left(-\frac{C_{ij}}{\varepsilon}\right)$ , equations (2.2) and (2.3) become

$$\begin{aligned} \gamma_{ij} &= \text{diag}(u_i) K_{ij} \text{diag}(v_j), \\ \mu_i &= u_i \odot \left(\sum_j K_{ij} v_j\right), \quad \nu_j = v_j \odot \left(\sum_i K_{ij}^T u_i\right). \end{aligned}$$

The algorithm is then given by, starting at  $u_1^{(0)} = \dots = u_n^{(0)} = \mathbb{1}_n$ ,  $v_1^{(0)} = \dots = v_m^{(0)} = \mathbb{1}_m$ , sequences  $(u_i^{(l)})_{1 \leq i \leq n}^{l \in \mathbb{N}}$  and  $(v_j^{(l)})_{1 \leq j \leq m}^{l \in \mathbb{N}}$  are built by the following scheme

$$i = 1, \dots, n : u_i^{(l+1)} = \frac{\mu_i}{\sum_j K_{ij} v_j^{(l)}} \quad (2.4)$$

$$j = 1, \dots, m : v_j^{(l+1)} = \frac{\nu_j}{\sum_i K_{ij}^T u_i^{(l+1)}}. \quad (2.5)$$

As for the scalar case, at fixed  $i$  and  $j$  the sequences  $(u_i^l)^{l \in \mathbb{N}}$  and  $(v_j^l)^{l \in \mathbb{N}}$  are geometrically growing, which can cause overflow problems if one of the quotients  $C_{ij}/\varepsilon$  is too large. To limit this problem, we use a technique known as soft-min, similar to the scalar case described in Remark 15. To describe it, we return to the variables  $\varphi_i$  and  $\psi_j$  where for

theses variables the algorithm is given by

$$\begin{aligned}\varphi_i &= \varepsilon \log(\mu_i) - \varepsilon \log \left( \sum_j \sum_l \exp \left[ \frac{-C_{ij}(\cdot, l) + \psi_j(l)}{\varepsilon} \right] \right) \\ \psi_j &= \varepsilon \log(\nu_j) - \varepsilon \log \left( \sum_i \sum_k \exp \left[ \frac{-C_{ij}(k, \cdot) + \varphi_i(k)}{\varepsilon} \right] \right).\end{aligned}$$

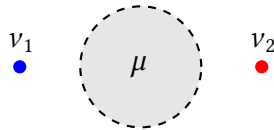
Notice that for example that if  $i$  is fixed, we have the following equality, satisfied for all  $z_i \in \mathbb{R}^N$  :

$$\begin{aligned}-\varepsilon \log \left( \sum_j \sum_l \exp \left[ \frac{-C_{ij}(\cdot, l) + \psi_j(l)}{\varepsilon} \right] \right) \\ = z_i - \varepsilon \log \left( \sum_j \sum_l \exp \left[ \frac{-C_{ij}(\cdot, l) + \psi_j(l) + z_i}{\varepsilon} \right] \right).\end{aligned}$$

To ensure then that  $-C_{ij}(\cdot, l) + \psi_j(l) + z_i$  is as close to zero as possible, we take  $z_i = z_i^* = (\psi_1, \dots, \psi_m)^{C_{i1}, \dots, C_{im}}$ . The same goes for  $\psi_j$ . Thus the algorithm becomes: Starting at  $\varphi_1^{(0)} = \dots = \varphi_n^{(0)} = \mathbb{1}_n$ ,  $\psi_1^{(0)} = \dots = \psi_m^{(0)} = \mathbb{1}_m$ , we build the sequences  $(\varphi_i^{(l)})_{1 \leq i \leq n}^{l \in \mathbb{N}}$  and  $(\psi_j^{(l)})_{1 \leq j \leq m}^{l \in \mathbb{N}}$  like this

$$\begin{aligned}i = 1, \dots, n : \varphi_i &= \varepsilon \log(\mu_i) + z_i^* - \varepsilon \log \left( \sum_j \sum_l \exp \left[ \frac{-C_{ij}(\cdot, l) + \psi_j(l) + z_i^*}{\varepsilon} \right] \right) \\ &\text{where } z_i^* = (\psi_1, \dots, \psi_m)^{C_{i1}, \dots, C_{im}}, \\ j = 1, \dots, m : \psi_j &= \varepsilon \log(\nu_j) + \bar{z}_j^* - \varepsilon \log \left( \sum_i \sum_k \exp \left[ \frac{-C_{ij}(k, \cdot) + \varphi_i(k) + \bar{z}_j^*}{\varepsilon} \right] \right) \\ &\text{where } \bar{z}_j^* = (\varphi_1, \dots, \varphi_n)^{\bar{C}_{1j}, \dots, \bar{C}_{nj}}.\end{aligned}$$

**Simulations.** We now present simulations based on the algorithm above. We place ourselves in the plane  $X = \mathbb{R}^2$  with as source measure a single distribution  $\mu = \mathcal{U}_{B(0,1)}$  uniform measure on the closed unit ball centered in 0 and as target measures two measures  $\nu_1 = \frac{1}{2}\delta_{y^1}$  et  $\nu_2 = \frac{1}{2}\delta_{y^2}$  with  $y^1, y^2 \in X$  :



Costs  $C_1$  (the cost between  $\mu$  and  $\nu_1$ ) and  $C_2$  (cost between  $\mu$  and  $\nu_2$ ) are Hölder

conjugate in the sense:

$$C_1((x_1, x_2), (y_1, y_2)) = \frac{(|x_1 - y_1|^p + |x_2 - y_2|^p)^{1/p}}{p}$$

$$C_2((x_1, x_2), (y_1, y_2)) = \frac{(|x_1 - y_1|^{p'} + |x_2 - y_2|^{p'})^{1/p'}}{p'}$$

where  $\frac{1}{p} + \frac{1}{p'} = 1$  with  $p, p' \in \mathbb{R}^* \cup \{+\infty\}$ .

It should be immediately noticed that  $p$  and  $p'$  do not necessarily belong to  $[1, \infty]$ . In particular, as long as  $p$  (or  $p'$ ) is strictly less than 1, we are not dealing with norms anymore. We have drawn above the level sets of these costs for different values of  $p$ .

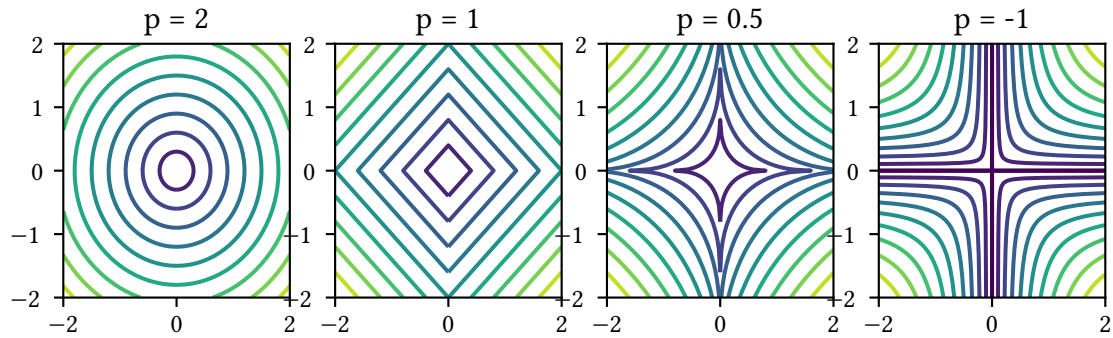
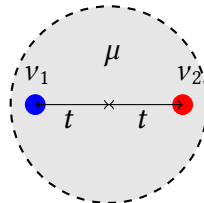


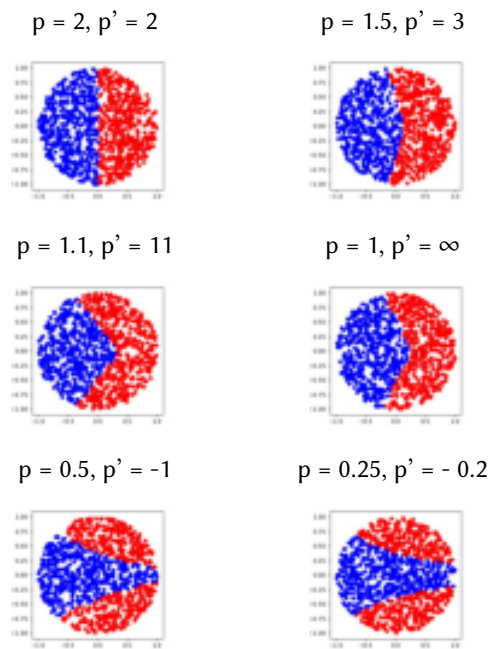
Figure 2.1: Level sets of four costs

We simulated  $\mu$  by a uniform distribution of 1000 Dirac measures of the same mass in the disk centered at 0 and with radius 1. In the following figures, we coloured in blue the part of  $\mu$  which was transported to  $v_1$  and in red the one which is transported to  $v_2$ . The regularisation parameter has been fixed at  $\varepsilon = 10^{-1}$ .

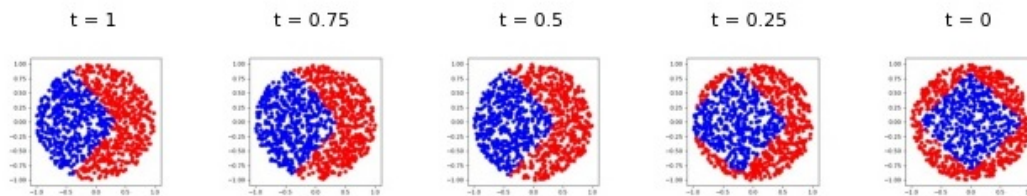
In a second step, we introduce a parameter  $t \in [0, 1]$  so that  $y_1 = (-t, 0)$  and  $y_2 = (t, 0)$ . Then  $t$  moves between 1 and 0. See the following diagram:



We perform the same simulations, for  $p = 1.5$  and  $p' = 3$  and  $\varepsilon = 10^{-1}$ .



**Figure 2.2:** Partitions of  $\mu$  for six  $p$  and  $p'$



**Figure 2.3:** Partitions of  $\mu$  for five  $t \in [0, 1]$  and  $p = 1.5, p' = 3$

## 3 A spatial Pareto exchange Economy Problem

### 3.1 Survol

Le deuxième chapitre de ce manuscrit est constitué d'un article en cours de publication en 2021, écrit en collaboration avec Guillaume Carlier et Bruno Nazaret. Celui-ci propose un modèle spatial d'allocation optimale de ressources où les coûts de transport de ces ressources sont pris en compte.

L'allocation optimale de ressources consiste à répartir entre des agents localisés sur un espace  $X$  une ou plusieurs variables  $v_1, \dots, v_N \in \mathcal{M}_+(X)$  (nos ressources) de quantité totale fixée de manière à maximiser un critère  $\mathcal{U} = \mathcal{U}(v_1, \cdot, v_N)$  appelé *utilité moyenne* :

$$\sup \{ \mathcal{U}(v_1, \dots, v_N) : v_1(X), \dots, v_N(X) \text{ fixés } \},$$

celle-ci étant la somme des utilités particulières  $U(x, \beta_1(x), \dots, \beta_N(x))$  :

$$\mathcal{U}(v_1, \cdot, v_N) = \int_X U(x, \beta_1(x), \dots, \beta_N(x)) dm(x)$$

où  $m$  désigne une mesure de référence sur  $X$  et  $\beta_i$  la densité de  $v_i$  par rapport à  $m$  pour  $i \in \llbracket 1, \dots, N \rrbracket$ .

Supposons  $X$  métrique compact et ajoutons une dimension spatiale au problème en considérant que nos allocations  $v_i$  sont le fruit du transport de ressources initialement réparties dans  $X$  : on se donne  $N$  mesures positives de masses finies  $\mu_1, \dots, \mu_N \in \mathcal{M}_+(X)$  correspondant à autant de distributions de départ. Désormais il faudra payer le déplacement de  $\mu_i$  à  $v_i$ . La théorie du Transport Optimal est toute désignée pour cela et si l'on se donne  $c_1, \dots, c_N$  les coûts de chaque ressource et que l'on note  $\mathcal{T}_{c_i}(\alpha_i m, \beta_i m)$  (où  $\alpha_i$  est la densité de  $\mu_i$  par rapport à  $m$ ) les valeurs des problèmes de Kantorovich (1.2) associés alors le nouveau problème est :

$$\sup \left\{ \int_X U(x, \beta_1(x), \dots, \beta_N(x)) dm(x) - \sum_{i=1}^N \mathcal{T}_{c_i}(\alpha_i m, \beta_i m) \right\} \quad (3.1)$$

où la borne supérieure est prise sur l'ensemble des  $v_1, \dots, v_N$  vérifiant

$$\int_X \alpha_i dm = \int_X \beta_i dm \text{ pour } i \in \llbracket 1, \dots, N \rrbracket.$$

Sous les conditions que  $\beta = (\beta_1, \dots, \beta_N) \in \mathbb{R}_+^N \rightarrow U(x, \beta)$  soient semi-continue



supérieurement, concave et croissante<sup>1</sup> pour  $m$ -presque tout  $x$  et que  $(x, \beta) \rightarrow U(x, \beta)$  soit sous-linéaire par rapport à  $\beta$  uniformément en  $x$  on établit un résultat d'existence dans (3.1). De plus celui-ci étant un problème de maximisation concave, on peut écrire un problème dual de minimisation convexe qui lui est associé. En introduisant les variables duales  $\varphi = (\varphi_1, \dots, \varphi_N)$ , celui-ci est donné par

$$\inf \{ K(\varphi) + \mathcal{V}(\varphi) : \varphi \in C(X, \mathbb{R})^N \}$$

où

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

$$\mathcal{V}(\varphi) = \int_X V(x, \varphi_1(x), \dots, \varphi_N(x)) dm(x)$$

avec

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

Un résultat de dualité forte est établi ce qui nous permet d'interpréter les variables duales optimales comme des prix. Plus précisément on peut donner l'interprétation suivante : les distributions  $\alpha_i$  correspondent à une offre pour différents biens distribués spatialement. En chaque  $x \in X$  se trouve un agent ayant comme dotation (comprendre fortune initiale) la somme  $w(x)$ .

On dit qu'un ensemble de prix  $y \rightarrow p(y)$  et une offre  $y \rightarrow \beta(y)$  forment un équilibre pour la dotation  $w$  si :

- **Libre circulation des échanges** : pour l'agent se trouvant en  $x$ , acheter une unité de  $\alpha_i$  se trouvant en  $y$  coûte  $p_i(y) - c_i(x, y)$ . On demande alors que l'agent se trouvant en  $x$  vende tout son stock et ce au plus offrant, autrement dit que son profit total soit

$$\pi(x) \cdot \alpha(x) = \sum_{i=1}^N \pi_i(x) \alpha_i(x)$$

$$\text{avec } \pi(x) \text{ donné par } \pi_i(x) = \max_{y \in X} p_i(y) - c_i(x, y) = -p_i^{c_i}(x).$$

---

<sup>1</sup>au sens où  $\beta - \beta' \in \mathbb{R}_+^N$  implique  $U(x, \beta) \geq U(x, \beta')$ .

- **Les consommateurs maximisent leur utilité respective** : un agent se trouvant en  $y$  maximise  $U(y, \beta)$  sous la contrainte que ses dépenses  $p(y) \cdot \beta(y)$  soient inférieures à sa dotation initiale  $w(x)$  plus les profits dûs à ses exportations  $\pi(x) \cdot \alpha(x)$ .
- **La demande et l'offre correspondent** : il existe des plans de transport  $\gamma_i \in \Pi(\alpha_i m, \beta_i m)$ ,  $i \in \llbracket 1, \dots, N \rrbracket$  tels que sur le support de ceux-ci les prix réalisent les maxima ci-dessus :  $\pi_i(x) = p_i(y) - c_i(x, y)$ .

Le théorème de dualité forte nous donne alors :

**Théorème 1.** Soient  $\varphi$  et  $\beta$  solutions respectives du problème dual et primal. Posons

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

alors  $(\beta, p)$  est un équilibre pour la dotation  $w$ .

La dernière partie du manuscrit s'attache à proposer des simulations numériques pour ce problème. Pour cela on applique la méthode de relaxation entropique et l'algorithme de Sinkhorn en s'inspirant des travaux de Cuturi [22]. On étudie en particulier des utilités de Cobb-Douglas de la forme

$$U(\beta) = \prod_{i=1}^N \beta_i^{a_i}, \text{ avec } a_i > 0 \text{ et } a \stackrel{\text{def}}{=} \sum_{i=1}^N a_i < 1$$

dont la transformée de Fenchel est donnée par

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

On traite à la fois l'exemple où  $X$  est la discrétisation d'un segment et un exemple plus géométrique où  $X$  est un graphe dont les nœuds sont des villes françaises et les ressources des denrées alimentaires.

## 3.1 Overview

The second chapter of this manuscript consists of an article to be published in 2021, written in collaboration with Guillaume Carlier and Bruno Nazaret. It proposes a spatial model of optimal resource allocation where the transport costs of these resources are taken into account.

The optimal allocation of resources consists in distributing on a space  $X$  (whose elements are called *agents*) one or several variables  $v_1, \dots, v_N \in \mathcal{M}_+(X)$  (our resources)

of fixed total quantity so as to maximize a criterion  $\mathcal{U} = \mathcal{U}(v_1, \cdot, v_N)$  called *average utility* :

$$\sup \{ \mathcal{U}(v_1, \dots, v_N) : v_1(X), \dots, v_N(X) \text{ fixed} \},$$

$\mathcal{U}$  being the sum of the particular utilities  $U(x, \beta_1(x), \dots, \beta_N(x))$  :

$$\mathcal{U}(v_1, \cdot, v_N) = \int_X U(x, \beta_1(x), \dots, \beta_N(x)) dm(x)$$

where  $m$  denotes a reference measure on  $X$  and  $\beta_i$  the density of  $v_i$  with respect to  $m$  for  $i \in \llbracket 1, \dots, N \rrbracket$ .

Let us assume that  $X$  is a compact metric space and let us add a spatial dimension to the problem by considering that our allocations  $v_i$  are the result of the transport of resources initially distributed in  $X$  : we give ourselves  $N$  positive measures of finite masses  $\mu_1, \dots, \mu_N \in \mathcal{M}_+(X)$  corresponding to as many starting distributions. From now on, it will be necessary to pay the displacement from  $\mu_i$  to  $v_i$ . The theory of Optimal Transport is well suited for this and if we give  $c_1, \dots, c_N$  the costs of each resource and if we note  $\mathcal{T}_{c_i}(\alpha_i m, \beta_i m)$  (where  $\alpha_i$  is the density of  $\mu_i$  with respect to  $m$ ) the values of the associated Kantorovich problems (1.2) then the new problem is:

$$\sup \left\{ \int_X U(x, \beta_1(x), \dots, \beta_N(x)) dm(x) - \sum_{i=1}^N \mathcal{T}_{c_i}(\alpha_i m, \beta_i m) \right\} \quad (3.2)$$

where the supremum is taken among  $v_1, \dots, v_N$  satisfying

$$\int_X \alpha_i dm = \int_X \beta_i dm \text{ for } i \in \llbracket 1, \dots, N \rrbracket$$

Under the conditions that  $\beta = (\beta_1, \dots, \beta_N) \in \mathbb{R}_+^N \rightarrow U(x, \beta)$  are upper semi-continuous, concave and nondecreasing<sup>2</sup> for  $m$ -almost all  $x$  as well as  $(x, \beta) \rightarrow U(x, \beta)$  is sublinear with respect to  $\beta$  uniformly in  $x$  we establish an existence result in (3.2). Moreover, this being a concave maximization problem, we can write a dual convex minimization problem associated to it. By introducing the dual variables  $\varphi = (\varphi_1, \dots, \varphi_N)$ , it is given by

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

---

<sup>2</sup>by nondecreasing we mean that  $\beta - \beta' \in \mathbb{R}_+^N$  implies  $U(x, \beta) \geq U(x, \beta')$ .

where

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

$$\mathcal{V}(\varphi) = \int_X V(x, \varphi_1(x), \dots, \varphi_N(x)) dm(x)$$

with

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

A strong duality result is established which allows us to interpret the optimal dual variables as prices. More precisely, we can give the following interpretation: the  $\alpha_i$  distributions correspond to a supply for different spatially distributed goods. In each  $x \in X$  there is an agent whose endowment (understand initial wealth) is the sum  $w(x)$ .

We say that a set of prices  $y \rightarrow p(y)$  and a supply  $y \rightarrow \beta(y)$  form an equilibrium for the endowment  $w$  if:

- **Free-mobility of trade** : for the agent in  $x$ , buying a unit of  $\alpha_i$  in  $y$  costs  $p_i(y) - c_i(x, y)$ . It is then required that the agent in  $x$  sells all his stock to the best purchaser, in other words that his total profit is

$$\pi(x) \cdot \alpha(x) = \sum_{i=1}^N \pi_i(x) \alpha_i(x)$$

$$\text{with } \pi(x) \text{ given by } \pi_i(x) = \max_{y \in X} p_i(y) - c_i(x, y) = -p_i^{c_i}(x)$$

- **Consumers maximise their respective utility**: an agent in  $y$  maximises  $U(y, \beta)$  under the constraint that his spendings  $p(y) \cdot \beta(y)$  is less than his initial endowment  $w(x)$  plus the profits due to his exports  $\pi(x) \cdot \alpha(x)$ .
- **Market clear** : there are transport plans  $\gamma_i \in \Pi(\alpha_i, \beta_i)$ ,  $i \in \llbracket 1, \dots, N \rrbracket$  such that on the support of these the prices attains the maxima above :  $\pi_i(x) = p_i(y) - c_i(x, y)$ .

The strong duality theorem then gives us:

**Theorem 1.** Let  $\varphi$  and  $\beta$  be solutions of the dual and primal problem respectively. Let us put

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

then  $(\beta, p)$  is an equilibrium for the endowment  $w$ .

The last part of the manuscript focuses on proposing numerical simulations for this problem. For this purpose we apply the entropy relaxation method and the Sinkhorn algorithm inspired by the work of Cuturi. In particular, we study Cobb-Douglas utilities of the form

$$U(\beta) = \prod_{i=1}^N \beta_i^{a_i}, \text{ with } a_i > 0 \text{ and } a \stackrel{\text{def}}{=} \sum_{i=1}^N a_i < 1$$

whose Fenchel transform is given by

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

We treat both the example where  $X$  is the discretization of a segment and a more geometrical example where  $X$  is a graph whose nodes are French cities and resources are foodstuffs.

### 3.2 A spatial Pareto exchange economy problem, X. Bacon, G. Carlier, B. Nazaret, HAL, 2021

# A spatial Pareto exchange economy problem

X. Bacon <sup>\*</sup>      G. Carlier <sup>†</sup>      B. Nazaret <sup>‡</sup>

January 3, 2022

## 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

---

<sup>\*</sup>CEREMADE, Université Paris Dauphine, PSL, and SAMM, Université Paris 1 Panthéon Sorbonne, [xavier.bacon@dauphine.psl.eu](mailto:xavier.bacon@dauphine.psl.eu)

<sup>†</sup>CEREMADE, Université Paris Dauphine, PSL, and INRIA-Paris, MOKAPLAN, [carlier@ceremade.dauphine.fr](mailto:carlier@ceremade.dauphine.fr)

<sup>‡</sup>SAMM, Université Paris 1 Panthéon-Sorbonne, [bruno.nazaret@univ-paris1.fr](mailto: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  $\mu_1, \dots, \mu_N \in \mathcal{M}_+(X)^N$ . Choosing a reference measure  $m \in \mathcal{M}_+(X)$  such that  $\mu_i$  is absolutely continuous with respect to  $m$  for  $i = 1, \dots, N$  (e.g.  $m = \sum_{i=1}^N \mu_i$ ) we denote by  $\alpha_i$  the density of  $\mu_i$  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

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

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  $\theta$  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.  $\theta$  is a Borel measure on  $Y$ ) and if  $T : Y \rightarrow Z$  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

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

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

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

where  $\Pi(\mu_i, \nu_i)$  is the set of transport plans between  $\mu_i$  and  $\nu_i$  i.e. the set of nonnegative Borel measures on  $X \times X$  having  $\mu_i$  and  $\nu_i$  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  $U: X \times \mathbb{R}_+^N \rightarrow \mathbb{R} \cup \{-\infty\}$ , which satisfies

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

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

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

$$\bar{\alpha}_i \geq 0, \quad \int_X \bar{\alpha}_i dm = \int_X \alpha_i dm, \quad i = 1, \dots, N, \quad (2.4)$$

and

$$\int_X U(x, \bar{\alpha}_1(x), \dots, \bar{\alpha}_N(x)) dm(x) > -\infty. \quad (2.5)$$

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

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

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

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

---

<sup>1</sup>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

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

which can also be expressed as

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

which incorporates the mass conservation constraints

$$\int_X \beta_i(x) dm(x) = \int_X \alpha_i(x) dm(x), \quad i = 1, \dots, N. \quad (2.9)$$

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 straightforward, 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  $\int_X \beta_i^n dm = \mu_i(X)$ ,  $(\beta^n)_n$  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  $\beta^n$  this new sequence we may assume that

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

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

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

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

$$\limsup_n \int_X U(x, \beta^n(x)) dm(x) \leq \int_X U(x, \beta(x)) dm(x). \quad (3.2)$$

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  $\nu = (\nu_1, \dots, \nu_N) \in \mathcal{M}_+(X)^N$

$$\nu_i^n := \beta_i^n m \xrightarrow{*} \nu_i, \quad i = 1, \dots, N. \quad (3.3)$$

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

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

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

$$\sup(2.7) = \int_X U(x, \beta(x)) dm(x) - \mathcal{T}_\mu(\nu). \quad (3.5)$$

Since  $\beta$  and  $\nu$  may be different and  $\beta$  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,

$$\int_X f(x) d\nu_i^n(x) \geq \int_X f(x) \min(\beta_i^n(x), M) dm(x).$$

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

$$\int_X f(x) d\nu_i(x) \geq \int_X f(x) \min(\beta_i(x), M) dm(x),$$

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

$$\nu_i \geq \beta_i m, \quad \text{for } i = 1, \dots, N. \quad (3.6)$$

Thanks to the Radon-Nikodym Theorem, we can decompose each measure  $\nu_i$  as

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

Since  $\nu_i^s$  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

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

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

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

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

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

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

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

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

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

Now let us define

$$\tilde{\gamma}_i := \gamma_i^a + (\text{id}, \text{id})_{\#} \alpha_i^s m \quad (3.10)$$

and observe that

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

and

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

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

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

Setting  $\tilde{\beta} = (\tilde{\beta}_1, \dots, \tilde{\beta}_N)$  we thus have

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

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

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

Together with (3.5), this implies that  $\tilde{\beta}$  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  $i = 1, \dots, N$ , and writing the Lebesgue decomposition of  $\nu_i$  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  $\tilde{\beta}_i \in L^1(X, m)$  such that

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

This, together with the monotonicity of  $U$ , shows that

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

therefore we have

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

where in the right-hand side  $\nu^a = (\nu_1^a, \dots, \nu_N^a)$  denotes the absolutely continuous part of  $\nu$  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  $(x, \varphi) = (x, \varphi_1, \dots, \varphi_N) \in X \times \mathbb{R}^N$ :

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

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

$$\mathcal{V}(\varphi) := \int_X V(x, \varphi_1(x), \dots, \varphi_N(x)) dm(x), \quad \forall \varphi \in C(X)^N,$$

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

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

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

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

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

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

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

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

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

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

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  $\nu \in \mathcal{M}(X)^N$ :

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

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

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*

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

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

$$\inf(3.15) = \max(2.7). \quad (3.18)$$

*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:

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

On the one hand, by the Fenchel-Moreau Theorem and the convexity and lower semicontinuity of  $-U(x, \cdot)$  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  $\varphi_i$  is strictly positive and  $\mathcal{V}(\varphi) < +\infty$  implies that each  $\varphi_i$  is nonnegative, we have

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

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

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

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  $\varphi$  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  $\bar{\alpha} \in L^1(X, m)^N$  such that (2.4) holds and there exists  $\varepsilon \in (0, 1)$  such that

$$\int_X U(x, (1 - \varepsilon)\bar{\alpha}_1(x), \dots, (1 - \varepsilon)\bar{\alpha}_N(x)) dm(x) > -\infty. \quad (3.19)$$

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  $\tilde{\varphi}^n := (\tilde{\varphi}_1^n, \dots, \tilde{\varphi}_N^n)$  by

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

It is well-known that  $(\tilde{\varphi}_i^n)^{c_i} = (\varphi_i^n)^{c_i}$  and that  $\tilde{\varphi}_i^n \geq \varphi_i^n$  so that  $K(\tilde{\varphi}^n) = K(\varphi^n)$  and (since  $V(x, \cdot)$  is nonincreasing)  $\mathcal{V}(\tilde{\varphi}^n) \leq \mathcal{V}(\varphi^n)$ . Hence  $\tilde{\varphi}^n$  is also a minimizing sequence for (3.15). Now the advantage of working with the minimizing sequence  $\tilde{\varphi}_i^n$  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  $\tilde{\varphi}_i^n$



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

$$\tilde{\varphi}_i^n \leq \delta_i^n + C. \quad (3.21)$$

Now fix,  $M$  such that

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

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

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

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

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

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

## 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

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

and for every  $i$ ,  $\beta := (\beta_1, \dots, \beta_N) \in L^1(X, m)^N$  such that  $\beta_i \geq 0$  and  $\int_X \beta_i m = \mu_i(X)$  and every  $\varphi := (\varphi_1, \dots, \varphi_N) \in C(X)^N$ , 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  $\varphi$  solves (3.15) and  $\beta$  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  $\varphi_i$  and  $c_i$ , this means that, whenever  $(x, y)$  is in the support of  $\gamma_i$ , one has

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

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

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

## 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  $U(x, \cdot)$  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

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

- **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

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

- **Markets clear:** which means that for every  $i$ , one can find a plan  $\gamma_i$  ( $\gamma_i(A \times B)$  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  $\gamma_i$  (which means that this plan is consistent with free-mobility of trade).

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

**Theorem 4.1.** *Let  $\varphi$  and  $\beta$  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  $(\beta, p)$  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 initial endowment of location  $x \in X$  in the good  $i \in \{1, \dots, N\}$ . Problem (3.15) thus takes the form:

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

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

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

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  $\varphi$ , the minimizer of  $\psi \mapsto \Phi(\varphi, \psi)$  is explicitly given by

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

so replacing in  $\Phi$ , we get

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

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:

$$-\beta(y) \in \partial V(y, \varphi(y)), \quad \forall y \in X \quad (5.4)$$

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

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

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  $\gamma_i$  solves the entropic optimal transport problem

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

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

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

which is the entropic regularization of (2.7) with regularization parameter  $\varepsilon$ .

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

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

which can be solved by coordinate descent as follows. Starting from  $(\psi^0, \varphi^0)$ , 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:

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

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

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

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

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

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  $y \in X$ :

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

and for  $i = N$

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

**The Cobb-Douglas case.** If we further specify the utility to be of Cobb-Douglas form<sup>2</sup>:

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

---

<sup>2</sup>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

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

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

$$e^{t^b} = A$$

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 two 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  $\lambda$  is large, transport is costly and  $\beta_1, \beta_2$  tend to be close to the initial distributions  $\alpha_1$  and  $\alpha_2$ . On the contrary, when  $\lambda$  is small, the utility term dominates and  $\beta_1$  and  $\beta_2$  tend to be constant.

In our first three simulations,  $\alpha_1$  and  $\alpha_2$  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,  $\alpha$  appears in light grey and  $\beta$  in dark grey, the prices  $(\varphi_1, \varphi_2)$  are plotted in the second row. The third row represents the convergence of the algorithm: we plot the decay of the functional  $\Phi$  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,  $\beta$  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  $\alpha_1$  and of two gaussians for  $\alpha_2$  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 <sup>3</sup> 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

<sup>3</sup>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  $\varepsilon$ , 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  $\Phi$  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.

$$\varepsilon = 0.01, (a_1, a_2) = (0.49, 0.49), \lambda = 0.2, p = 0.3$$

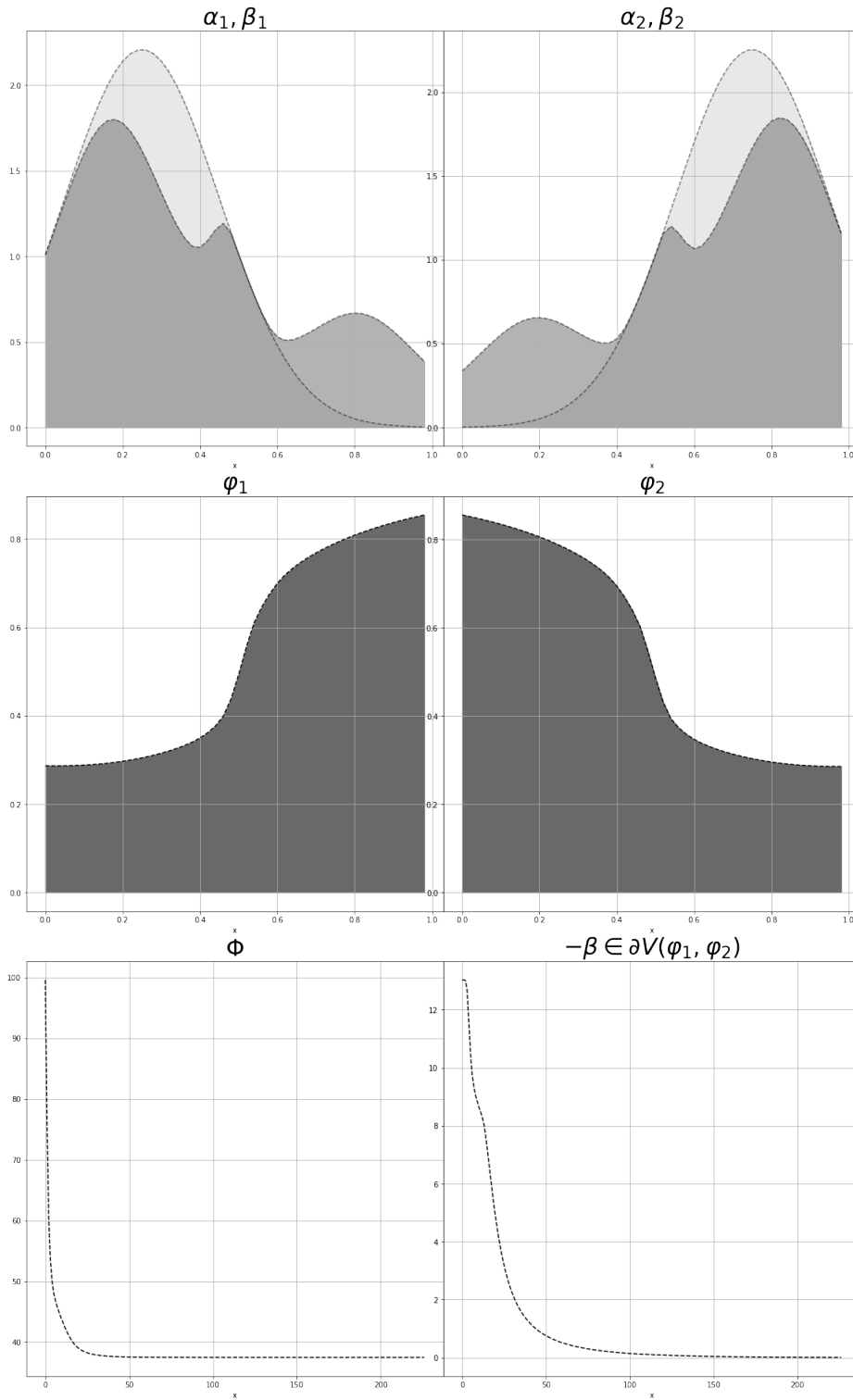


Figure 1: Gaussian distributions with concave cost

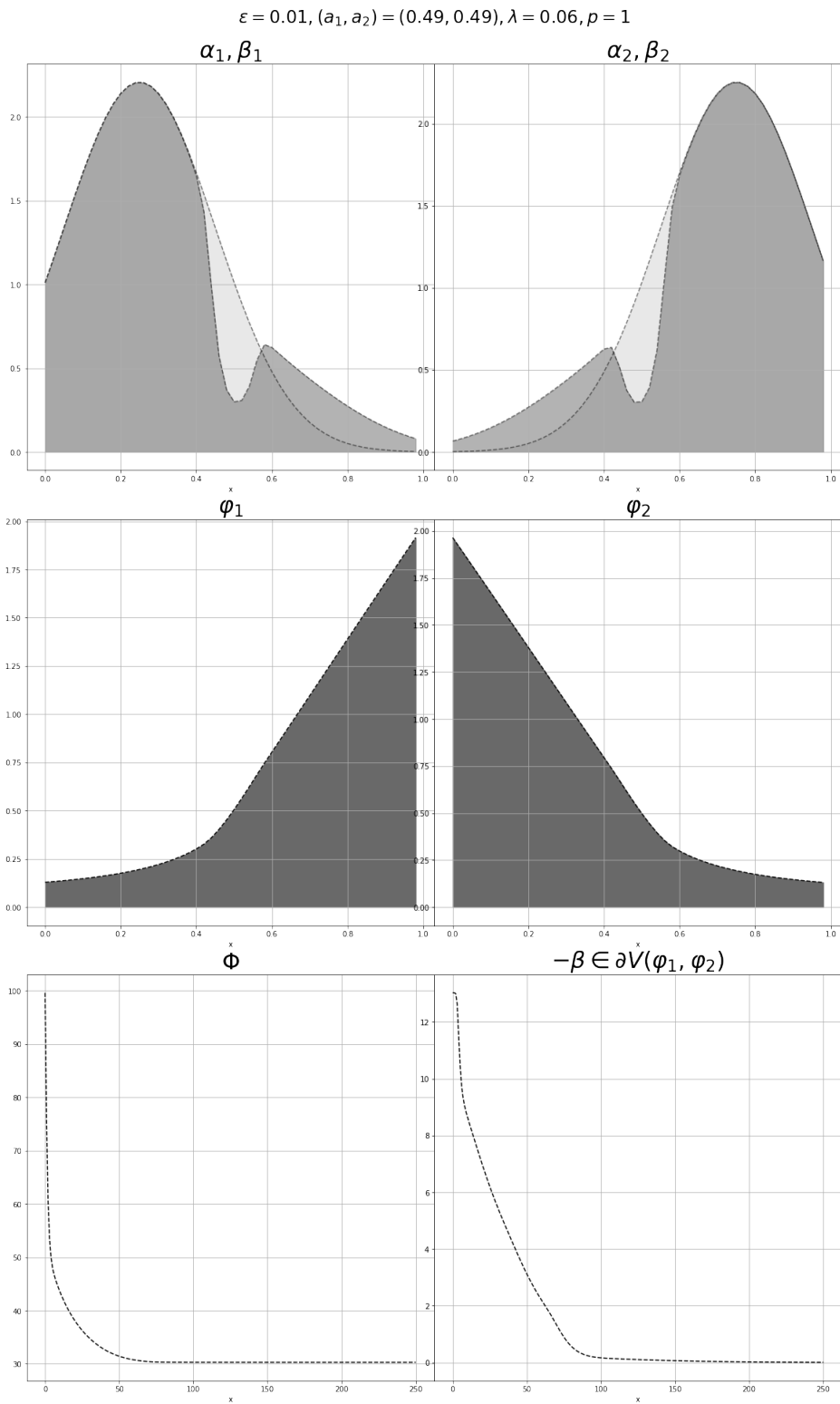


Figure 2: Gaussian distributions with homogeneous cost

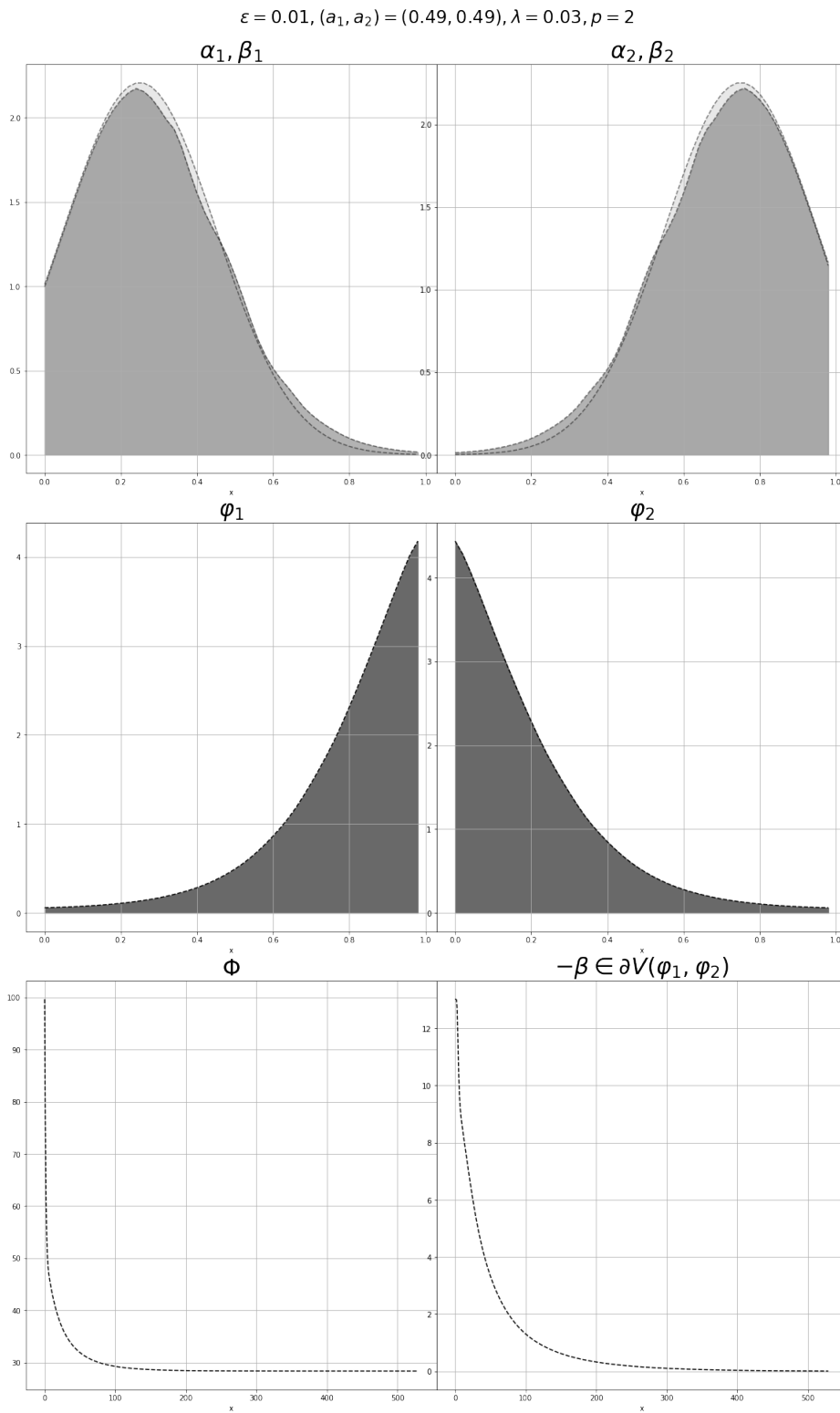


Figure 3: Gaussian distributions with convex cost

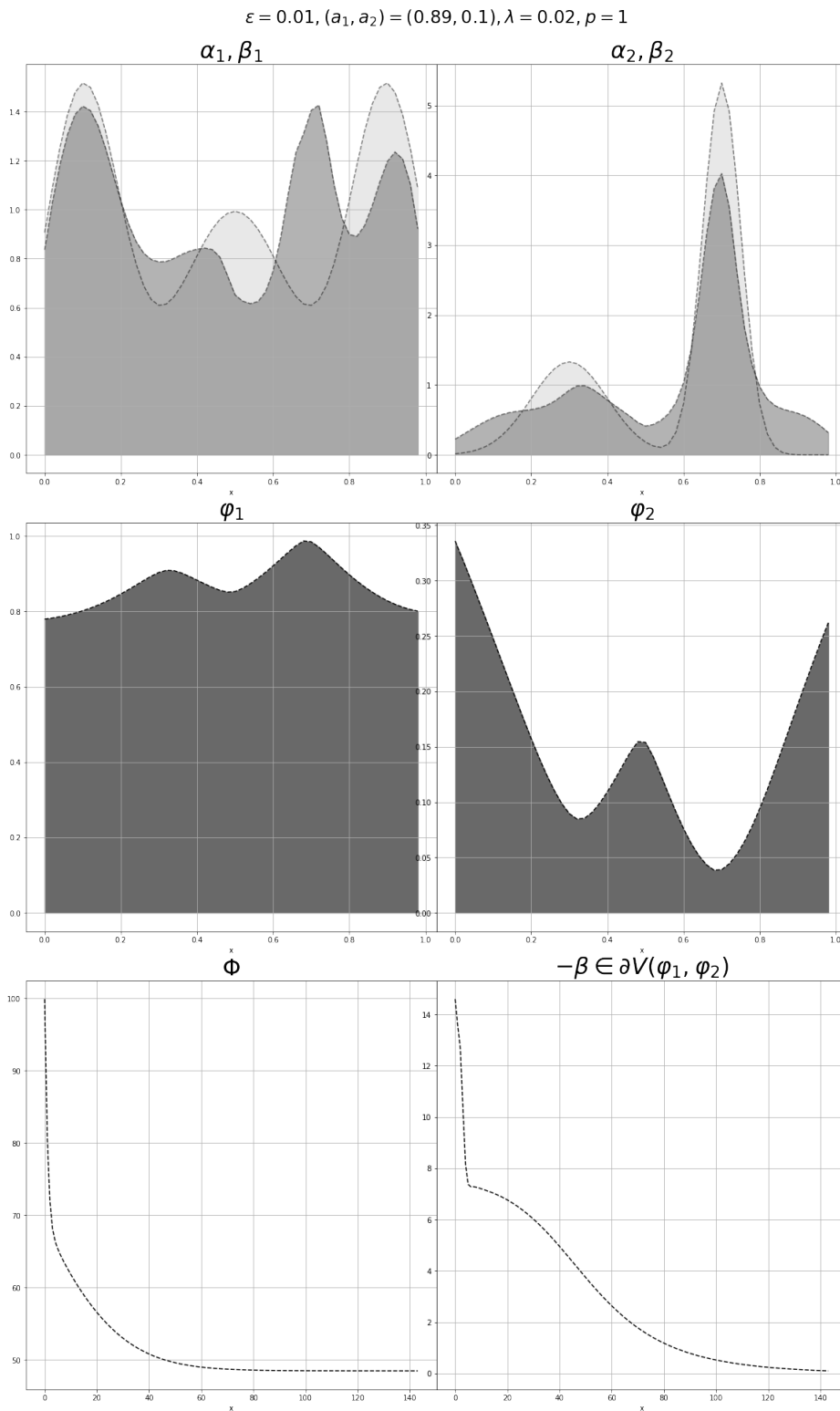


Figure 4: Another example with  $a_1 + a_2$  close to 1

$$\varepsilon = 0.1, (a_1, a_2) = (0.2, 0.3), \lambda = 0.2, p = 2$$

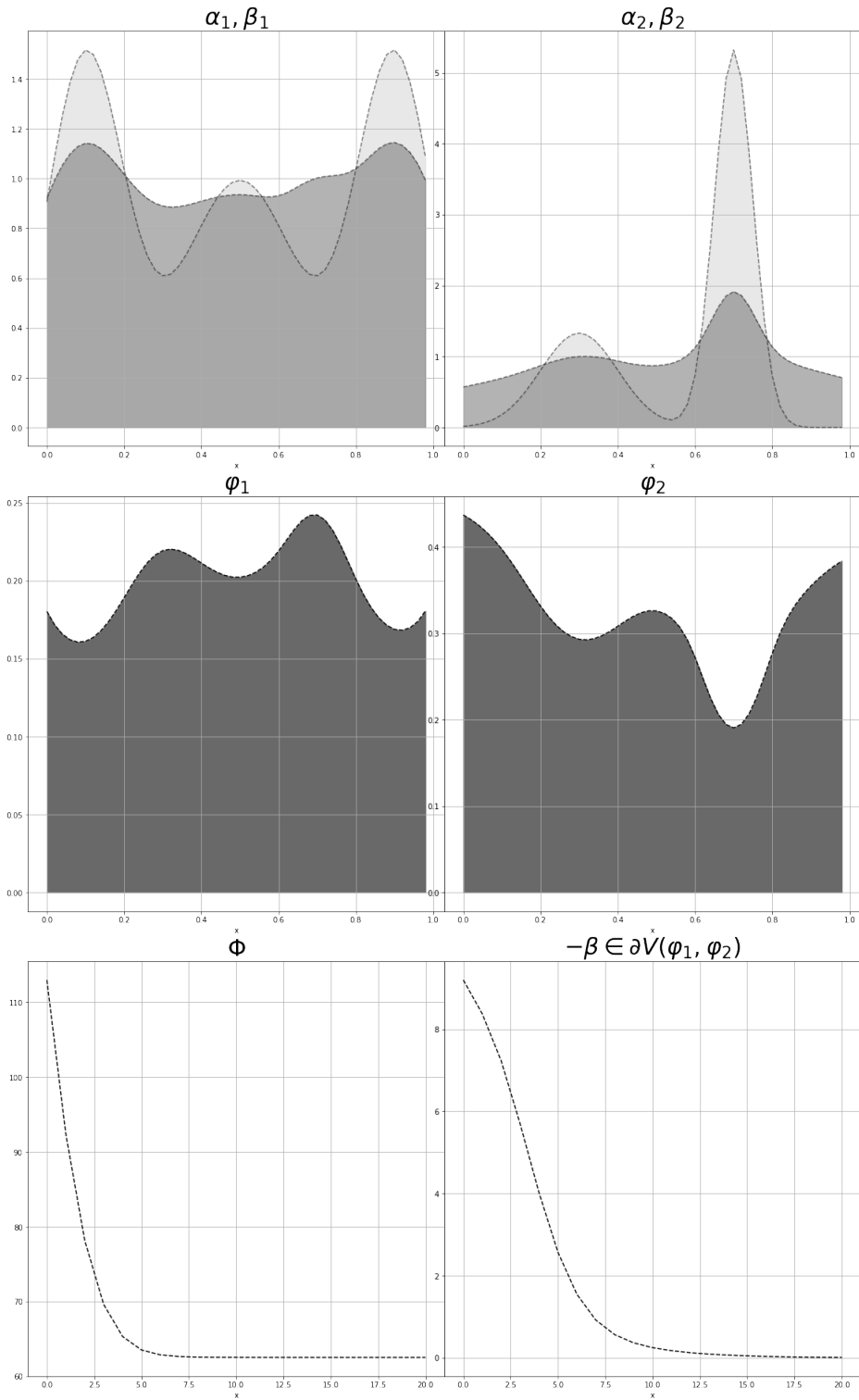


Figure 5: Another example with  $a_1 + a_2$  far from 1

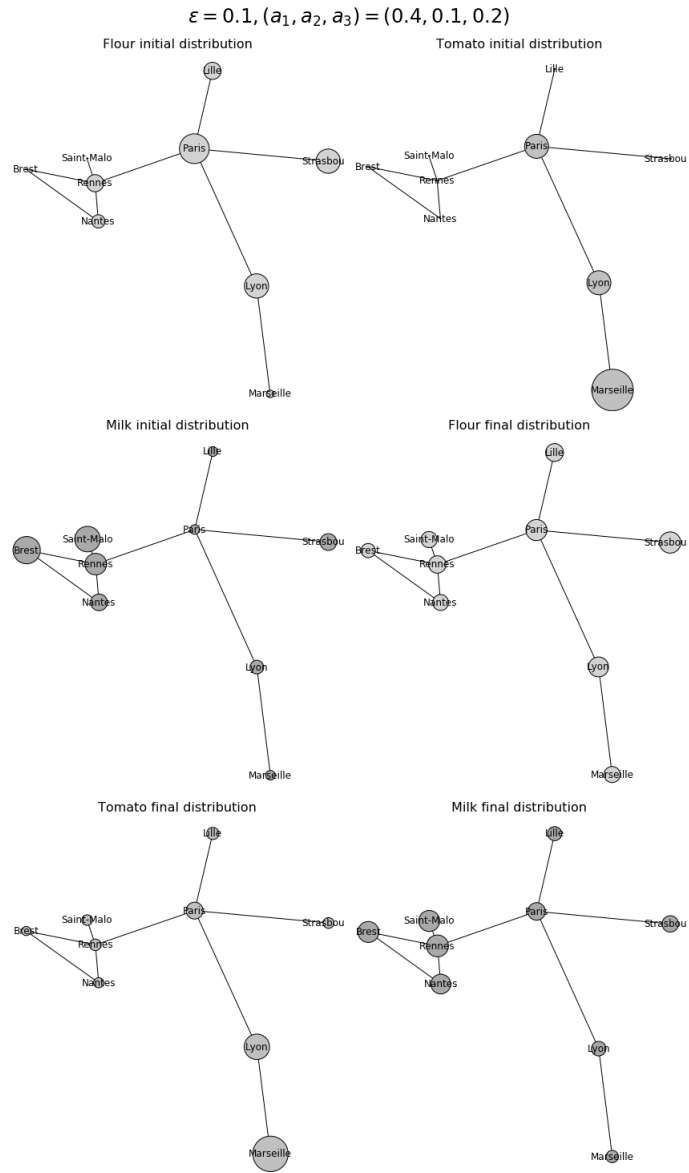


Figure 6: French network with  $\varepsilon = 0.1$



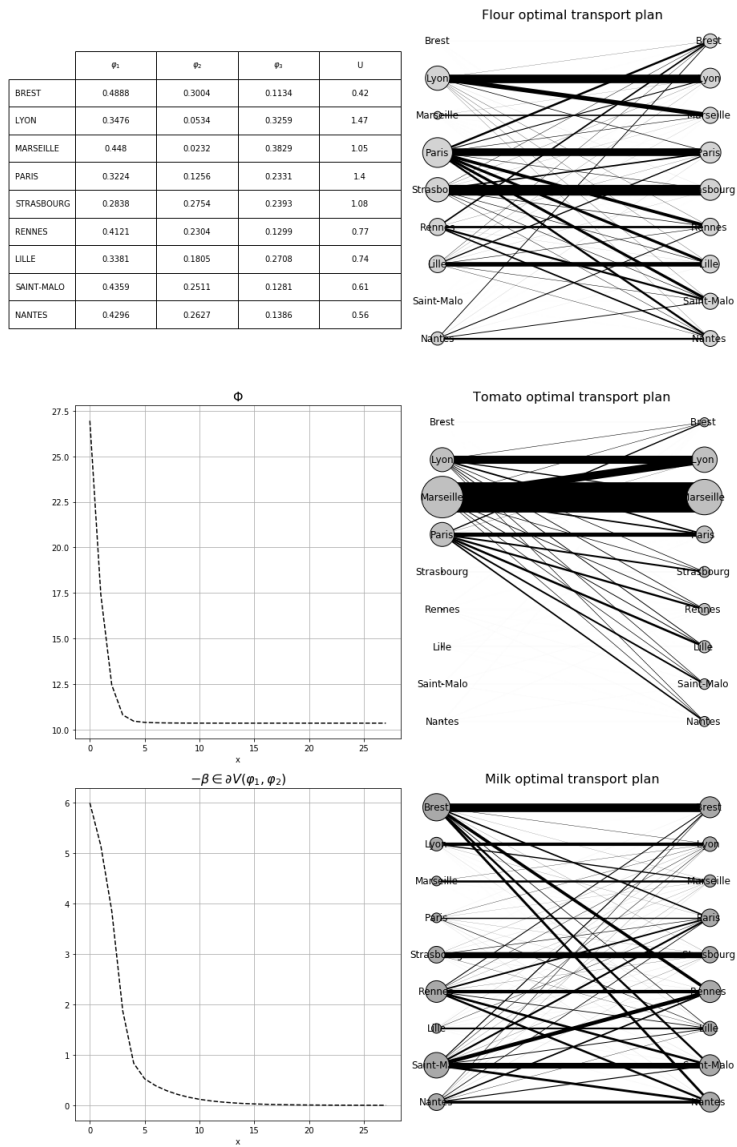


Figure 7: French network with  $\varepsilon = 0.1$

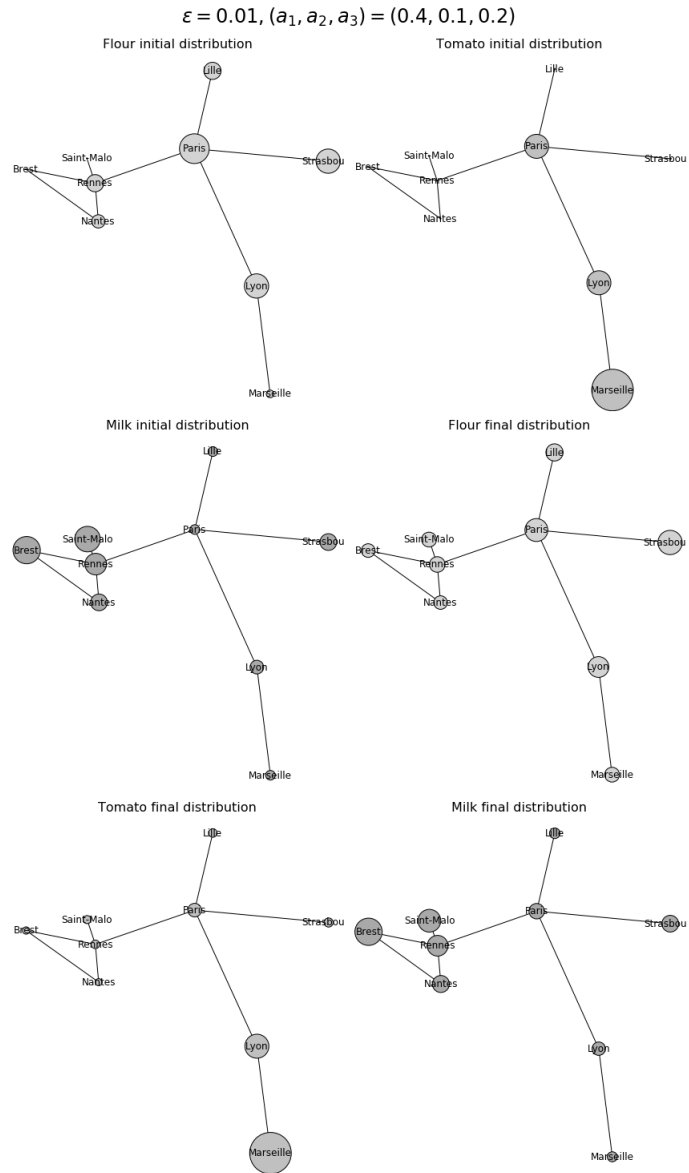


Figure 8: French network with  $\varepsilon = 0.01$

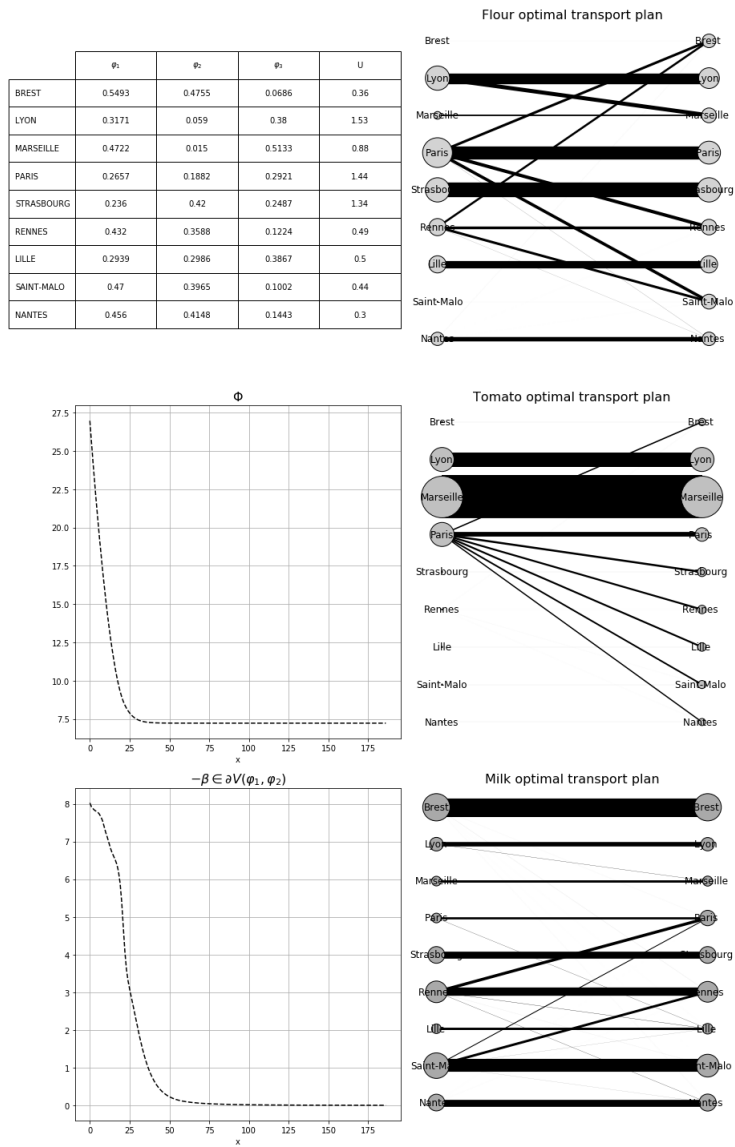


Figure 9: French network with  $\varepsilon = 0.01$

## 4 Unbalanced Optimal Transport for positive semi-definite matrices

### 4.1 Overview

**Unbalanced Optimal Transport.** This chapter is devoted to an extension of the Unbalanced Optimal Transport developed separately by various authors in [18, 19] and [34, 36]. Given  $\rho_0$  and  $\rho_1$  two finite positive measures on  $\Omega$  a domain of  $\mathbb{R}^d$ , the Unbalanced Optimal Transport informally consists of minimizing the functional

$$\frac{1}{2} \int_0^1 \left( \int_{\Omega} \frac{|\omega_t(x)|^2}{\rho} dx + \int_{\Omega} \frac{|\zeta_t(x)|^2}{\rho} dx \right) dt.$$

subject to the so-called *continuity equation with source term* written

$$\partial_t \rho + \operatorname{div}_x \omega = \zeta. \quad (4.1)$$

In this context,  $\rho = \rho_t(x)$  is a time-dependent density such that  $\rho_{t=0} = \rho_0$  and  $\rho_{t=1} = \rho_1$ ,  $\omega = \omega_t(x)$  is interpreted as momentum of  $\rho$  and  $\zeta = \zeta_t(x)$  as a creation/annihilation term.

Both groups of authors have established on their side that the minimization of this action under the constraints (4.1) provides a geodesic distance between positive Radon measurements. The approach of Chizat et al. uses Convex Analysis tools to describe geodesics between atomic measures. On their side, Kondratyev et al. are interested in the topological properties (density, metrisability of weak convergence-\*) of this distance and have developed a differential calculus *à la Otto* [44, 30] in the context of gradient flows.

Afterwards, in [19], Chizat et al. establish a static formulation (*à la Kantorovitch*) dealing with two transport plans  $\gamma_0, \gamma_1 \in \mathcal{M}_+(\Omega \times \Omega)$ : the first describes where the mass coming from  $\rho_0$  goes, while the second where the mass arriving in  $\rho_1$  comes from. In the classical theory of Optimal Transport, these two plans are equal. The authors introduce a cost  $c(x_0, x_1, m_0, m_1)$  which also takes into account the mass deficit (or in excess). The static formulation (called *semi-coupling*) is given by:

$$\inf \left\{ \iint_{\Omega^2} c(x_0, x_1, \gamma_0, \gamma_1) : \pi_1 \# \gamma_0 = \rho_0, \pi_2 \# \gamma_1 = \rho_1 \right\}.$$

The authors finally show that static and dynamic formulations are equivalent. Following these works, the authors Chen, Tannenbaum, Georgiou and Ryu have proposed similar

models: see [15] for an overview. Otherwise see [13], [11], [12], [17] as well as [16] for the integration of the *unbalanced* Optimal Transport to their model.

**Extension to matricial case.** We wish to extend the Unbalanced Optimal Transport to matrix-valued measures. To this end, a first approach is to replace the  $\mathbb{R}_+$ -valued  $\rho$  by a  $\mathcal{S}_{++}$ -valued density  $S = S_t(x)$  where  $\mathcal{S}_{++}$  denotes the set of  $n \times n$  symmetric positive-definite matrices (see [12, 35, 9]). The momentum and the creation/annihilation term are replaced as follows

$$\begin{aligned}\rho &\leftarrow S, \\ \omega &\leftarrow \mathcal{M}, \\ \zeta &\leftarrow \mathcal{R}\end{aligned}$$

in such a way that the equation (4.1) becomes

$$\partial_t S + \operatorname{div}_x \mathcal{M}^{\operatorname{sym}} = \mathcal{R}^{\operatorname{sym}} \quad (4.2)$$

subject to  $S_{t=0} = \mathcal{S}_0$  and  $S_{t=1} = \mathcal{S}_1$  where  $\mathcal{S}_0$  and  $\mathcal{S}_1$  are two prescribed  $\mathcal{S}_{++}$ -valued measures on  $\Omega$ . The new functional is then given by

$$\mathcal{A}(S, \mathcal{M}, \mathcal{R}) = \frac{1}{2} \int_0^1 \left( \int_{\Omega} \operatorname{tr}(\mathcal{M} S^{-1} \mathcal{M}^*) + \operatorname{tr}(\mathcal{R} S^{-1} \mathcal{R}^*) \, dx \right) dt. \quad (4.3)$$

This approach was proposed by Vorotnikov and Brenier in [9] and shortly after by Li and Zu in [35]. The diffusion term  $\mathcal{M}$  in [9] applies to the matrix itself. In the model presented in this chapter, this constraint has been relaxed by introducing a family of  $n^2$  scattering terms applying to each of the coordinates independently (see definitions (4.17) and (4.18) below).

**Relationship with Bures-Wasserstein metric.** Looking at the equation (4.2) and the action (4.3), if  $\mathcal{M} = 0$  that is if there is no diffusion, these equations are reduced to

$$\partial_t S = \mathcal{R}^{\operatorname{sym}}$$

and

$$\mathcal{A}(S, 0, \mathcal{R}) = \frac{1}{2} \int_0^1 \int_{\Omega} \operatorname{tr}(\mathcal{R} S^{-1} \mathcal{R}^*) \, dx \, dt.$$

This happens in particular when  $\Omega$  is reduced to a singleton. In that case  $\mathcal{S}$  and  $\mathcal{R}$  are replaced respectively by matrices  $S$  and  $R$  and the equations become

$$\partial_t S = R^{sym} \quad (4.4)$$

$$A(S, R) = \frac{1}{2} \operatorname{tr}(RS^{-1}R^*) \quad (4.5)$$

under the constraints  $S_{t=0} = S_0$  and  $S_{t=1} = S_1$  where  $S_0$  and  $S_1$  are two prescribed elements of  $\mathcal{S}_{++}$ . The equations (4.4) and (4.5) subject to the starting and ending constraints happen to be exactly the variational formulation of the Bures-Wasserstein metric which has been studied extensively in [53, 6, 55]. While these articles are restricted to the cases of symmetric and positive definite matrices, different authors extend the study to symmetric semidefinite matrices of constant rank (see [38, 37]) and to positive semidefinite (PSD) matrices [54]. This chapter will soon be the subject of a pre-publication article.

**Chapter's organisation** In the first part, we present the Bures-Wasserstein distance, corresponding to our model when there is no spatial motion. The originality in comparison to existing works ([41],[54]) is to propose a convex variational formulation, corresponding to a change of variable in the formulation proposed by [41]. It has to be noted that this convex formulation is a special case of the model developed in [35] when  $\Omega$  is reduced to a singleton. Moreover, a new approach using the study of generalized Riccati equations (*i.e.* with PSD solutions) is proposed, under the assumption of a kernel inclusion.

In the second part, we present a dynamic model of Optimal Transport of PSD-valued measures. We investigate the dual problem and the necessary and sufficient conditions of optimality. Proofs of this part are largely included in the article [35].

An Appendix presents more in-depth the Bures-Wasserstein distance, in particular its relationship with the Riccati equation, the Optimal Transport and its geometric approach.

## 4.2 Notations

### Matrices

- $\mathbb{R}^{n \times n}$  is the space of  $n \times n$  matrices, equipped with the Frobenius product and its induced norm

$$A \cdot B = \operatorname{tr}(A^*B) \text{ and } |A| = \sqrt{A \cdot A}$$

where  $A^*$  is the transpose of  $A$ . Sometimes we will also use the notation  $\langle A | B \rangle = A \cdot B$ .

- $\mathcal{G}$  is the set of invertible matrices.

- $\mathcal{S}$  is the subspace of symmetric  $n \times n$  matrices.
- $\mathcal{S}_+$  is the subspace symmetric positive-semidefinite (PSD)  $n \times n$  matrices.
- $\mathcal{S}_{++}$  is the subspace symmetric positive-definite  $n \times n$  matrices.
- $A^{\text{sym}} \stackrel{\text{def}}{=} \frac{A+A^*}{2}$  will denote the symmetric part of  $A \in \mathbb{R}^{n \times n}$ .
- For  $S \in \mathcal{S}_+$ , there exists a unique matrix  $A \in \mathcal{S}_+$  such that  $A^2 = S$ . It is called the symmetric square root of  $S$  and is denoted  $\sqrt{S}$  or  $S^{1/2}$ .
- $A \preceq B$  for  $A, B \in \mathcal{S}$  means that  $B - A \in \mathcal{S}_+$ .
- $\mathbb{R}^{n \times n \times d}$  is the space of  $n \times n \times d$  3-tensors. An element  $\mathbf{M} \in \mathbb{R}^{n \times n \times d}$  is written  $\mathbf{M} = (M[k])_{1 \leq k \leq d} \in (\mathbb{R}^{n \times n})^d$  and its coordinate  $(i, j, k) \in \llbracket 1, n \rrbracket^2 \times \llbracket 1, d \rrbracket$  is denoted by  $M_{ij}[k]$ . This space is equipped with the sum of Frobenius products and its induced norm

$$\mathbf{M} : \mathbf{N} \stackrel{\text{def}}{=} \sum_{k=1}^d M[k] \cdot N[k] \text{ and } |\mathbf{M}| = \sum_{k=1}^d \sqrt{M[k] \cdot M[k]}.$$

For  $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n \times n \times d}$ , the tensor product is defined by

$$\mathbf{MN} = \sum_{k=1}^d M[k]N[k] \in \mathbb{R}^{n \times n}.$$

For  $\mathbf{M}, \mathbf{N} \in \mathbb{R}^{n \times n \times d}$  and  $A \in \mathbb{R}^{n \times n}$ , the right and left product of  $\mathbf{M}$  with  $A$  is defined by

$$\begin{aligned} \mathbf{AM} &= (AM[k])_{1 \leq k \leq d} \in \mathbb{R}^{n \times n \times d} \\ \mathbf{MA} &= (M[k]A)_{1 \leq k \leq d} \in \mathbb{R}^{n \times n \times d}. \end{aligned}$$

Eventually, we remind the definition of Moore-Penrose pseudoinverse and basic properties of it.

**Definition 1.** A square matrix  $A$  admits a (Moore-Penrose) pseudoinverse  $A^\dagger$  if the four following conditions are satisfied

$$\begin{aligned} AA^\dagger A &= A \\ A^\dagger AA^\dagger &= A^\dagger \\ AA^\dagger &\in \mathcal{S} \\ A^\dagger A &\in \mathcal{S} \end{aligned}$$

**Remark 2.0.** Such a matrix exists and is unique.

1.  $0^\dagger = 0$ .

2. If  $A$  is invertible, then  $A^\dagger = A^{-1}$ . In particular, if  $\lambda \in \mathbb{R}^*$  then  $\lambda^\dagger = 1/\lambda$ .
3. If  $A = \text{diag}(\lambda_1, \dots, \lambda_n)$  then  $A^\dagger = \text{diag}(\lambda_1^\dagger, \dots, \lambda_n^\dagger)$ .
4. More generally, if  $A$  is symmetric and  $A = P \text{diag}(\lambda_1, \dots, \lambda_n) P^*$ , with  $P$  an orthogonal matrix, then  $A^\dagger = P \text{diag}(\lambda_1^\dagger, \dots, \lambda_n^\dagger) P^*$ .

**Proposition 3.** Let  $x \mapsto A(x)$  be a Borelian function. So it is for  $x \mapsto A^\dagger(x)$ .

*Proof.* Lemma 3.1. in [48]. □

### Measures and functions

- $\mathcal{M}(X, E)$  is the set of Borelian measures of  $X$  and *valued* in  $E$ . In general  $X = [0, 1]$ ,  $X = \Omega$  a domain of  $\mathbb{R}^d$  or  $X = [0, 1] \times \Omega$  and  $E = \mathcal{S}$ ,  $E = \mathcal{S}_+$  or  $E = \mathbb{R}^{n \times n}$  or any cartesian product of the previous examples.
- $C^0(X, E)$  is the set of continuous functions defined on  $X$  and valued in  $E$ . As for measures, in general  $X = [0, 1]$ ,  $X = \Omega$  a domain of  $\mathbb{R}^d$  or  $X = [0, 1] \times \Omega$  and  $E = \mathcal{S}$ ,  $E = \mathcal{S}_+$  or  $E = \mathbb{R}^{n \times n}$  or any cartesian product of the previous examples.
- $C^1(X, E)$  is the set of differentiable functions whose derivative is continuous.
- $C_c^1([0, 1] \times \Omega, \mathcal{S})$  is the set of  $C^1$  functions from  $[0, 1] \times \Omega$  to  $\mathcal{S}$  with compact support.

Following [54], we give few geometric vocabulary about curves and geodesics.

### Curves and geodesics

- A curve of  $(X, d)$  is a continuous map  $\gamma : [0, 1] \rightarrow X$ .
- On a metric space  $(X, d)$ , the length of a curve  $\gamma : [0, 1] \rightarrow X$  is defined by the supremum of  $\sum_{k=0}^{l-1} d(\gamma(t_k), \gamma(t_{k+1}))$  over all subdivisions  $(t_k)_{0 \leq k \leq l}$  of  $[0, 1]$ . This supremum is denoted  $L(\gamma)$  and belongs to  $[0, \infty]$ .
- For  $x_0, x_1 \in X$ , the length distance  $d_L$  between  $x_0$  and  $x_1$  is defined by  $d_L(x_0, x_1) = \inf L(\gamma)$  over all curves  $\gamma$  such that  $\gamma(0) = x_0$ ,  $\gamma(1) = x_1$  and  $L(\gamma) \neq \infty$  (a such curve is called *rectifiable*).
- $(X, d)$  is a length space if  $d = d_L$ .
- A curve  $\gamma$  is parameterized at constant speed if there exists a function  $v : [0, 1] \rightarrow \mathbb{R}_+$  such that  $L(\gamma|_{[t_0, t_1]}) = v(t_1 - t_0)$  for all  $t_0, t_1 \in [0, 1]$  and  $t_0 \leq t_1$ .
- A curve  $\gamma$  is minimizing when for all  $t_0, t_1 \in [0, 1]$  and  $t_0 \leq t_1$ ,  $L(\gamma|_{[t_0, t_1]}) = d(\gamma(t_0), \gamma(t_1))$ .
- A geodesic is a minimizing curve of constant speed.
- A length space  $(X, d)$  is a geodesic space if there exists a geodesic between any  $x_0, x_1 \in X$ .



## 4.3 Bures-Wasserstein metric

This section is devoted to the Bures-Wasserstein metric on PSD-matrices of a given size  $n \in \mathbb{N}$ . The particular case of definite matrices have been studied in the past twenty years by several authors [23, 29, 53, 6, 55] extending previous works that focused on matricial metrics [24, 43, 27, 33]. More recently, the extension of Bures-Wasserstein metric on PSD matrices held mathematicians attention in Riemannian geometry and its application to statistics [54] or in quantum mechanics through the Schrödinger Problem on non-commutative space [4, 41] where Hermitian matrices are considered instead of symmetric matrices.

### 4.3.1 Primal, dual and optimality conditions

In relation to [54], we present a variational formulation of Bures-Wasserstein metric on PSD matrices, inspired by [41] we go further by giving a convex formulation for this metric and applying classic Convex Analysis tools to derive optimality conditions.

**Convex formulation of Bures-Wasserstein metric.** Let us start with an explicit definition of the Bures-Wasserstein metric.

**Definition/Proposition 4** (Bures-Wasserstein distance). The Bures-Wasserstein metric on  $\mathcal{S}_+$  is defined by the following formula

$$d_B^2(S_0, S_1) \stackrel{\text{def}}{=} \text{tr } S_0 + \text{tr } S_1 - 2 \text{tr } \sqrt{\sqrt{S_0} S_1 \sqrt{S_0}}. \quad (4.6)$$

Moreover,  $d_B$  is a distance on  $\mathcal{S}_+$ .

*Proof.* Corollary in [24]. □

**Remark 5** (Commutative case). When  $S_0$  and  $S_1$  commutes, this reads explicitly

$$d_B(S_0, S_1) = |\sqrt{S_1} - \sqrt{S_0}|.$$

**Proposition 6** (Topology induced by  $d_B$ ). The topology induced is the topology induced by the Euclidean distance on  $\mathcal{S}_+$  (or by the restriction on  $\mathcal{S}_+$  of any norm defined on  $\mathbb{R}^{n \times n}$ ).

*Proof.* Lemma 3.2 in [54]. □

It appears that  $d_B$  admits a variational formulation given by the following definition [41].

**Proposition 7.**  $(\mathcal{S}_+, d_B)$  is a geodesic space. Moreover, geodesics of  $(\mathcal{S}_+, d_B)$  correspond to minimizers of the following problem (beware of the factor 2)

$$2 d_B^2(S_0, S_1) = \min \frac{1}{2} \int_0^1 \text{tr } U_t^* S_t U_t dt \quad (4.7)$$

subject to  $(S_t, U_t)_{t \in [0,1]}$  such that  $(S, U) \in C^1([0, 1], \mathcal{S}_+ \times \mathcal{S})$  and

$$S_{t=0} = S_0, \quad S_{t=1} = S_1, \quad \dot{S}_t = (S_t U_t)^{\text{sym}}.$$

*Proof.* Proposition 3.4 in [41]. □

However, the variational formulation above is not convex with respect to  $(S, U)$ . It should be put in perspective with the Benamou-Brenier formulation of Optimal Transport described in 1.3.5. The change of variables here is to set  $R = SU \in \mathbb{R}^{n \times n}$  and since  $R = SU$  implies  $\text{im}(R) \subset \text{im}(S)$  this leads us to define the new following action.

**Definition 8.** Let  $S \in C^1([0, 1], \mathcal{S})$  and  $R \in C^1([0, 1], \mathbb{R}^{n \times n})$ . We define the action  $\mathcal{A}$  of the curve  $(S, R)$  by

$$\mathcal{A}(S, R) = \int_0^1 A(S_t, R_t) dt$$

where  $A : \mathcal{S} \times \mathbb{R}^{n \times n} \rightarrow \mathbb{R} \cup \{+\infty\}$  is given by

$$A(S, R) = \begin{cases} \frac{1}{2} \text{tr} R^* S^\dagger R & \text{if } S \in \mathcal{S}_+ \text{ and } \text{im}(R) \subset \text{im}(S) \\ +\infty & \text{otherwise.} \end{cases}$$

Eventually the identity  $S = SS^\dagger S$  leads us to the convex formulation of (4.7).

**Problem 6.** Let  $S_0, S_1 \in \mathcal{S}_+$ , the convex relaxation of (4.7) is given by

$$(\mathcal{P}) = \inf \mathcal{A}(S, R) \tag{4.8}$$

subject to  $(S, R) \in C^1([0, 1], \mathcal{S}_+ \times \mathbb{R}^{n \times n})$  and

$$S_{t=0} = S_0, \quad S_{t=1} = S_1, \quad \dot{S}_t = R_t^{\text{sym}}.$$

A minimizer of (4.8) will be called a Bures-Wasserstein geodesic between  $S_0$  and  $S_1$  or just geodesic if there is no ambiguity.

**Dual formulation and optimality conditions** The convex formulation above allows us to look for a dual formulation and derive optimality conditions. Notice that the previous problem (4.8) can be written in the a more convenient way

$$(\mathcal{P}) = \inf \mathcal{A}(S, R) + \iota_{C(S_0, S_1)}(L(S, R))$$

where  $L : C^1([0, 1], \mathcal{S} \times \mathbb{R}^{n \times n}) \rightarrow C^0([0, 1], \mathcal{S})$  is given by

$$L(S, R) = -\partial_t S + R^{\text{sym}}$$

and  $C(S_0, S_1)$  is the closed convex set given by

$$C(S_0, S_1) = \left\{ \Sigma \in C^0([0, 1], \mathcal{S}) : \int_0^1 Q_t \cdot \Sigma_t dt = \Delta_{S_0}^{S_1}(Q), \forall Q \in C^1([0, 1], \mathcal{S}) \right\}$$

where

$$\Delta_{S_0}^{S_1}(Q) = Q_1 \cdot S_1 - Q_0 \cdot S_0.$$

In order to find a dual formulation, Fenchel-Rockafellar theorem formally gives us the equality

$$\begin{aligned} (\mathcal{P}) &= \inf \mathcal{A}(S, R) + \iota_{C(S_0, S_1)}(L(S, R)) \\ &= \sup \iota_{C(S_0, S_1)}^*(Q) - \mathcal{A}^*(L^*(Q)) \\ &\stackrel{\text{def}}{=} (\mathcal{D}) \end{aligned}$$

and motivates the next proposition.

**Proposition 9** (Fenchel conjugates). According to notations just above, the following equalities are satisfied for all  $(S, R) \in C^1([0, 1], \mathcal{S}_+ \times \mathbb{R}^{n \times n})$

$$\begin{aligned} \iota_{\mathcal{O}}^*(S, R) &= \mathcal{A}(S, R), \\ (-\Delta_{S_0}^{S_1})^*(-L^*(S, R)) &= \iota_{C(S_0, S_1)}(S, R). \end{aligned}$$

*Proof.* We establish the first identity. The proof is divided in three parts.

- Let  $S \in \mathcal{S} \setminus \mathcal{S}_+$ . It exists  $v \in \mathbb{R}^n$  such that<sup>1</sup>  $\langle v | Sv \rangle = -|v|^2 < 0$  and set

$$\bar{S}_\lambda = -\lambda v v^* \text{ for } \lambda \in \mathbb{R}_+ \text{ and } \bar{R}_0 = 0,$$

so for all  $R$  it follows

$$\begin{aligned} \iota_{\mathcal{O}}^*(S, R) &= \sup_{(\bar{S}, \bar{R})} \langle (\bar{S}, \bar{R}) | (S, R) \rangle - \iota_{\mathcal{O}}(\bar{S}, \bar{R}) \\ &\geq \langle (\bar{S}_\lambda, \bar{R}_0) | (S, R) \rangle - \iota_{\mathcal{O}}((\bar{S}_\lambda, \bar{R}_0)) \\ &= \langle S_\lambda | S \rangle \\ &= -\lambda \langle v v^* | S \rangle \\ &= -\lambda \langle v | Sv \rangle \xrightarrow{\lambda \rightarrow +\infty} +\infty, \end{aligned}$$

then

$$\iota_{\mathcal{O}}^*(S, R) = +\infty.$$

- Let  $S \in \mathcal{S}, R \in \mathbb{R}^{n \times n}$  be such that  $\text{im}(R) \not\subset \text{im}(S)$ . Denoting by  $\Pi_{\ker(S)}$  the orthogonal

---

<sup>1</sup>For instance, set  $v$  a renormalized eigenvector associated to a non-positive eigenvalue of  $S$ .

projector onto the kernel of  $S$ , let us set

$$\bar{R} = \Pi_{\ker(S)}R \text{ and } \bar{S} = -\frac{1}{2}\bar{R}\bar{R}^*.$$

Clearly,  $(\bar{S}, \bar{R}) \in \mathcal{O}$  and  $S\bar{R} = 0$ , then

$$\begin{aligned} i_{\mathcal{O}}^*(S, R) &\geq \langle S | \bar{S} \rangle + \langle R | \bar{R} \rangle \\ &= \langle R | \bar{R} \rangle \\ &> 0 \end{aligned}$$

where the last inequality is a consequence of the assumption  $\text{im}(R) \not\subset \text{im}(S)$ . Then  $i_{\mathcal{O}}^*(S, R) = +\infty$ .

- Let  $S \in \mathcal{S}_+$ ,  $R \in \mathbb{R}^{n \times n}$  such that  $\text{im } R \subset \text{im } S$ . The very definition of  $i_{\mathcal{O}}^*$  gives us

$$\begin{aligned} i_{\mathcal{O}}^*(S, R) &= \sup_{(\bar{S}, \bar{R}) \in \mathcal{O}} \langle (S, R) | (\bar{S}, \bar{R}) \rangle \\ &= \sup_{(\bar{S}, \bar{R}) \in \mathcal{O}} \left\langle S \mid \bar{S} + \frac{1}{2}\bar{R}\bar{R}^* \right\rangle + \langle R | \bar{R} \rangle - \frac{1}{2} \underbrace{\langle S | \bar{R}\bar{R}^* \rangle}_{=\|S^{1/2}\bar{R}\|^2}, \end{aligned}$$

then the identity  $R = \sqrt{\bar{S}^\dagger} \bar{S}^{-1/2} R$  (which is a consequence of the range inclusion) enables us to complete the square

$$\langle R | \bar{R} \rangle - \frac{1}{2}\|S^{1/2}\bar{R}\|^2 = -\frac{1}{2}\|S^{1/2}\bar{R} - \sqrt{S^\dagger}R\|^2 + \frac{1}{2}\|\sqrt{S^\dagger}R\|^2$$

and obtain

$$i_{\mathcal{O}}^*(S, R) = \sup_{(S, R) \in \mathcal{O}} \left\langle S \mid \bar{S} + \frac{1}{2}\bar{R}\bar{R}^* \right\rangle - \frac{1}{2}\|S^{1/2}\bar{R} - \sqrt{S^\dagger}R\|^2 + \frac{1}{2}\|\sqrt{S^\dagger}R\|^2.$$

Notice that the first term is negative due to the constraints on  $\bar{S}$  and  $\bar{R}$ , as well as the second term. Then setting  $\bar{R} = S^\dagger R$  and  $\bar{S} = -1/2\bar{R}\bar{R}^*$  gives us

$$i_{\mathcal{O}}^*(S, R) = \frac{1}{2}\|\sqrt{S^\dagger}R\|^2.$$

We establish the second identity. Let  $S \in \mathcal{S}$ ,  $R \in \mathbb{R}^{n \times n}$ . By the very definition of the

Fenchel conjugate we have

$$\begin{aligned} (-\Delta_{S_0}^{S_1})^* (-L^*(S, R)) &= \sup_{Q \in C^1([0,1], S)} \langle Q \mid -L^*(S, R) \rangle + \Delta_{S_0}^{S_1}(Q) \\ &= \sup_{Q \in C^1([0,1], S)} \langle L(Q) \mid -(S, R) \rangle + \Delta_{S_0}^{S_1}(Q). \end{aligned}$$

For an arbitrary  $Q \in C^1([0, 1], S)$ , the previous term can be written

$$\begin{aligned} \langle L(Q) \mid -(S, R) \rangle + \Delta_{S_0}^{S_1}(Q) &= \text{tr} \left( - \int_0^1 \dot{Q}_t S_t + Q_t R_t \, dt + Q_1 S_1 - Q_0 S_0 \right) \\ &= \text{tr} \left( \int_0^1 Q_t [\dot{S}_t - R_t^{\text{sym}}] \, dt + Q_1 [S_1 - S_{t=1}] - Q_0 [S_0 - S_{t=0}] \right), \end{aligned}$$

where the last equality is a consequence of an integration by parts. By linearity, it is clear that a necessary and sufficient condition for  $(-\Delta_{S_0}^{S_1})^* (-L^*(S, R)) = 0$  is to have  $S_0 = S_{t=0}, S_1 = S_{t=1}$  and

$$\text{tr} \left( \int_0^1 Q_t [\dot{S}_t - R_t^{\text{sym}}] \, dt \right) = 0, \forall Q \in C^1([0, 1], S)$$

or equivalently, by continuity and symmetry of  $\dot{S}_t - R_t^{\text{sym}}$ ,

$$\dot{S}_t - R_t^{\text{sym}} = 0, \forall t \in [0, 1]$$

and the result follows.  $\square$

Calculations of the two previous Fenchel conjugates lead us to the dual formulation of (4.8). Moreover there is no duality gap between these two problems, according to the next theorem.

**Definition/Theorem 10** (Dual formulation and no duality gap). Let  $S_0, S_1 \in \mathcal{S}_+$ , a dual problem for (4.8) is given by

$$\begin{aligned} (\mathcal{D}) &\stackrel{\text{def}}{=} \sup_Q \Delta_{S_0}^{S_1}(Q) \\ &= \sup_Q Q_1 \cdot S_1 - Q_0 \cdot S_0 \end{aligned} \tag{4.9}$$

subject to  $Q \in C^1([0, 1], S)$  and  $\dot{Q}_t + \frac{1}{2}Q^2 \preceq 0$ . Moreover the following no duality gap equality holds

$$(\mathcal{P}) = (\mathcal{D})$$

and the problem  $(\mathcal{P})$  is attained.

*Proof.* This result is a consequence of the Fenchel-Rockafellar theorem (see for example

[49], [10],[1]). Fenchel-Rockafellar theorem asserts that

$$\sup_Q \Delta_{S_0}^{S_1}(Q) - \iota_{\mathcal{O}}(L(Q)) = \inf_{S,R} (-\Delta_{S_0}^{S_1})^*(-L^*(S, R)) + \iota_{\mathcal{O}}^*(S, R)$$

or according to Proposition 9,

$$\sup_Q \Delta_{S_0}^{S_1}(Q) - \iota_{\mathcal{O}}(L(Q)) = \inf_{S,R} \mathcal{A}(S, R) + \iota_{\mathcal{L}(S_0, S_1)}(S, R).$$

However the following hypothesis are satisfied:

- $L$  is linear and bounded (according to supremum norms on  $C^1$  and  $C^0$ ).
- $\Delta_{S_0}^{S_1}$  is convex and l.s.c. as bounded linear functional.
- Taking  $Q = 0$ ,  $\Delta_{S_0}^{S_1}$  is continuous at 0 as bounded linear functional. Moreover  $L(0) = (0, 0)$  and  $\iota_{\mathcal{O}}(0, 0) = 0 < +\infty$ ,

then constraints qualifications are satisfied and Fenchel-Rockafellar theorem can be applied.  $\square$

As a straightforward consequence of the previous theorem, we derive the optimality conditons.

**Proposition 11** (Optimality conditions). Let  $(S, R) \in C^1([0, 1], \mathcal{S} \times \mathbb{R}^{n \times n})$  be minimizer of (4.8). The supremum of (4.9) is attained at  $Q \in C^1([0, 1], \mathcal{S})$  if and only there exists  $S \in C^1([0, 1], \mathcal{S}_+)$  such that

$$\begin{cases} S_{t=0} = S_0, S_{t=1} = S_1 \\ \dot{S} = (SQ)^{\text{sym}} \\ \dot{Q} + \frac{1}{2}Q^2 \leq 0 \\ S \cdot (\dot{Q} + \frac{1}{2}Q^2) = 0 \end{cases} \quad (4.10)$$

**Remark 12.** Since  $R = SQ$  for optimal  $(S, Q)$ , in these coordinates  $d_B$  is given by

$$d_B^2(S_0, S_1) = \frac{1}{4} \int_0^1 \text{tr } Q_t S_t Q_t \, dt.$$

### 4.3.2 Examples of geodesics

**Rank one matrices.** We begin with explicit geodesics between rank one PSD matrices.

**Proposition 13** (Rank one). Let  $s_0, s_1 \in \mathbb{R}^n$  and set  $S_0 = s_0 s_0^*$ ,  $S_1 = s_1 s_1^*$  be two PSD rank 1 matrices. The Bures-Wasserstein geodesic between  $S_0$  and  $S_1$  is given for all  $t \in [0, 1]$  by

$$S_t \stackrel{\text{def}}{=} s_t s_t^* \quad \text{where} \quad s_t \stackrel{\text{def}}{=} (1-t)s_0 + ts_1. \quad (4.11)$$

*Proof.* The result is clear if  $s_0 = s_1$ . Let us assume that  $s_0 \neq s_1$ . The potential  $Q = Q_t$  given for all  $t \in [0, 1]$  by

$$Q_t = \frac{2}{f_t}(s_1 - s_0)(s_1 - s_0)^*$$

where  $f_t \stackrel{\text{def}}{=} \langle s_t | s_1 - s_0 \rangle$  (notice that  $f_t \neq 0$ ), satisfies the optimality conditions (4.10) since on one hand

$$\dot{Q}_t = -\frac{2|s_1 - s_0|^2}{f_t^2}(s_1 - s_0)(s_1 - s_0)^*,$$

then

$$\dot{Q}_t + \frac{1}{2}Q_t^2 = 0$$

and on the other hand, direct calculations give us

$$\begin{aligned} \dot{S}_t &= \dot{s}_t s_t^* + s_t \dot{s}_t^* \\ &= (s_1 - s_0) \dot{s}_t^* + s_t (s_1 - s_0)^*. \end{aligned}$$

Finally we obtain

$$\dot{S}_t = (S_t Q_t)^{\text{sym}}.$$

The first and last equations of (4.10) are clear and then  $(S, Q)$  is optimal and  $S$  is a geodesic between  $S_0$  and  $S_1$  for the Bures-Wasserstein metric.  $\square$

**Remark 14.** As a consequence, for  $S_0 = s_0 s_0^*$  and  $S_1 = s_1 s_1^*$  two PSD rank 1 matrices and  $(S, Q) = (S_t, Q_t)$  their associated optimal path and potential, we clearly have

$$\frac{1}{4}Q_t S_t Q_t = (s_1 - s_0)(s_1 - s_0)^*$$

and the distance can be easily compute as follows

$$\begin{aligned} d_B^2(S_0, S_1) &= \frac{1}{4} \int_0^1 \text{tr} Q_t^* S_t Q_t \, dt \\ &= \text{tr}(s_1 - s_0)(s_1 - s_0)^* \\ &= |s_1 - s_0|^2. \end{aligned}$$

In other words, PSD semi-positive rank 1 matrices endowed with  $d_B$  are isometric to  $(\mathbb{R}_+^n, |\cdot|)$ .

We extend the previous result to fixed ranked matrices assuming a kernel inclusion hypothesis. For this aim we have to make a small digression about the Riccati equation.

---

**Riccati equation.** The formula (4.11) can be understood in a different way using presentation of Bures-Wasserstein metric as push-forward distance of the Euclidean metric on  $\mathbb{R}^{n \times n}$ : in the non-singular case, when  $S_0$  and  $S_1$  are both invertible, Takatsu proposed in [53] a geometric point of view of  $d_B$  (see also [55] and [29]). Considering the flat Frobenius metric  $G(Z, W) = Z \cdot W$  on  $\mathcal{G}$ , the map  $\Pi(A) = AA^*$  is a Riemannian submersion (more details are given in the Appendix) of  $\mathcal{G}$  into  $\mathcal{S}_{++}$  endowed with metric tensor  $g$  defined by

$$g_S(X, Y) = \text{tr}(\mathcal{L}_S^{-1}(X) \cdot S \cdot \mathcal{L}_S^{-1}(Y)).$$

where  $\mathcal{L}_S(Z) = ZS + SZ^*$ . As a consequence, geodesics are mapped to geodesics and we deduce an explicit formula for  $S$  the geodesic between  $S_0$  and  $S_1$ ,

$$S_t = (A + tZ) \cdot (A + tZ)^* \quad (4.12)$$

where  $A = S_0^{1/2}$ ,  $Z = [\Lambda - \text{Id}] S_0^{1/2}$  (notice that  $(A + Z)^2 = S_1$ ) and  $\Lambda$  is the unique PSD matrix solution of the Riccati equation

$$\Lambda S_0 \Lambda^* = S_1. \quad (4.13)$$

and is explicitly given by

$$\Lambda = S_1^{1/2} (S_1^{1/2} S_0 S_1^{1/2})^{-1/2} S_1^{1/2}. \quad (4.14)$$

The formula (4.12) is nothing but the previous one written in (4.11). Eventually, these two formulas can be seen as the pushforward of the geodesic in  $\mathbb{R}_+^n$  through the squared map

$$\mathbb{R}_+^n \ni s \rightarrow s s^* \in \mathcal{S}_+.$$

However, tools used in non-singular case can not easily be adapted to  $\mathcal{S}$  due to the equation (4.14). Yet we have the following recent result.

**Theorem 15** ([26]). Let  $S_0$  and  $S_1$  be positive semidefinite matrices satisfying the kernel inclusion  $\ker S_0 \subset \ker S_1$ . Then

$$\Lambda = S_1^{1/2} [(S_1^{1/2})^\dagger S_0^\dagger (S_0^{1/2})^\dagger]^{1/2} S_1^{1/2}$$

is a solution of the following Riccati equation

$$\Lambda S_0 \Lambda = S_1.$$

Moreover, the uniqueness of its solution is ensured under the additional assumption  $\ker S_0 \subset \ker \Lambda$ .

**Fixed ranked matrices.** The following proposition implies in particular the existence of several geodesics (one for each solution of the Riccati equation) between two matrices



sharing the same rank and admitting the previous kernel inclusion.

**Proposition 16** (Pythagorean theorem). Let  $S_0, S_1 \in \mathcal{S}_+$  with  $\text{rank}(S_0) = \text{rank}(S_1) = d$  and  $\ker(S_0) \subset \ker(S_1)$ . We write them as a sum of symmetric rank 1 matrices

$$\begin{aligned} S_0 &= \sum_{i=1}^d S_0^i = \sum_{i=1}^d \lambda_i s_i s_i^* \\ S_1 &= \sum_{i=1}^d S_1^i \end{aligned} \quad (4.15)$$

where  $(s_i)$  is an orthonormal family of  $\mathbb{R}^n$ , i.e.  $\langle s_i | s_j \rangle = \delta_{ij}$  and  $(\lambda_i)$  the eigenvalues of  $S_0$ . The following equality is satisfied

$$d_B^2(S_0, S_1) = \sum_{i=1}^d d_B^2(S_0^i, S_1^i).$$

*Proof.* The kernel condition ensures that there exists (at least)  $\Lambda \in \mathcal{S}$  (according to Theorem 15) such that the following Riccati equation is satisfied

$$S_1 = \Lambda S_0 \Lambda \quad (4.16)$$

or equivalently  $S_1 = (\Lambda \sqrt{S_0}) (\Lambda \sqrt{S_0})^*$ . According to (4.15) the equation above becomes

$$S_1 = \sum_{i=1}^d \left( \sqrt{\lambda_i} \Lambda s_i \right) \left( \sqrt{\lambda_i} \Lambda s_i \right)^*.$$

Then, as a consequence of the previous formula for rank 1 matrices we obtain

$$\sum_{i=1}^d d_B^2(S_0^i, S_1^i) = \sum_{i=1}^d \text{tr} S_0^i + \text{tr} S_1^i - 2\lambda_i \langle s_i | \Lambda s_i \rangle.$$

The scalar product above can be rewritten in a more convenient way since

$$\sqrt{S_0} = \sum_{i=1}^d \sqrt{\lambda_i} s_i s_i^*$$

and then we obtain

$$\left( \sqrt{S_0} \Lambda \right) \left( \sqrt{S_0} \Lambda \right)^* = \sum_{1 \leq i, j \leq d} \sqrt{\lambda_i} \sqrt{\lambda_j} \langle s_i | \Lambda s_j \rangle s_i s_j^*.$$

The orthogonality identity  $\text{tr}(s_i s_j^*) = \delta_{ij}$  gives us

$$\text{tr} \left( \sqrt{S_0} \Lambda \right) \left( \sqrt{S_0} \Lambda^* \right)^* = \sum_{i=1}^d \lambda_i \langle s_i | \Lambda s_i \rangle,$$

which implies

$$\begin{aligned} \sum_{i=1}^d d_B^2(S_0^i, S_1^i) &= \text{tr} S_0 + \text{tr} S_1 - 2 \text{tr} \left( \sqrt{S_0} \Lambda \right) \left( \sqrt{S_0} \Lambda^* \right)^* \\ &= \text{tr} S_0 + \text{tr} S_1 - 2 \text{tr} \sqrt{S_0} \Lambda \sqrt{S_0}. \end{aligned}$$

Notice that

$$\begin{aligned} (\sqrt{S_0} \Lambda \sqrt{S_0})^2 &= \sqrt{S_0} \Lambda S_0 \Lambda \sqrt{S_0} \\ &= \sqrt{S_0} S_1 \sqrt{S_0} \text{ by (4.16),} \end{aligned}$$

then

$$\sqrt{S_0} \Lambda \sqrt{S_0} = \sqrt{\sqrt{S_0} S_1 \sqrt{S_0}}$$

and finally

$$\begin{aligned} \sum_{i=1}^d d_B^2(S_0^i, S_1^i) &= \text{tr}(S_0) + \text{tr}(S_1) - 2 \text{tr} \sqrt{\sqrt{S_0} S_1 \sqrt{S_0}} \\ &= d_B^2(S_0, S_1) \text{ by definition.} \end{aligned}$$

□

**A few words about the general case** In their recent article [54], the two authors use Differential Geometry techniques to give a complete description of Bures-Wasserstein geodesics. We have introduced Convex Analysis tools to achieve the same goal however our method fails when we do not make the kernel inclusion assumption of the Theorem (15).

To be more precise, the authors established that for any  $S_0, S_1 \in \mathcal{S}_+$ , setting  $X_0, X_1 \in \mathbb{R}^{n \times n}$  such that  $S_0 = X_0 X_0^*$ ,  $S_1 = X_1 X_1^*$  and  $O$  be an orthogonal matrix such that  $X_0^* X_1 O^* \in \mathcal{S}_+$  and finally denoting by  $s$  the following curve of  $\mathbb{R}^{n \times n}$

$$\begin{aligned} s &: [0, 1] \rightarrow \mathbb{R}^{n \times n} \\ t &\mapsto (1-t)X_0 + tX_1 O^*, \end{aligned}$$

then  $S = S_t = s_t s_t^*$  is a geodesic between  $S_0$  and  $S_1$ .

One wish I would have had would have been to find an optimal potential  $Q$  associated to each of its geodesics and then re-demonstrate the optimality of these curves.

## 4.4 Dynamic model

Let  $\Omega$  be a bounded domain on  $\mathbb{R}^d$  and let  $\mathcal{S} = \mathcal{S}_t(x)$ ,  $\mathcal{M} = \mathcal{M}_t(x)$  and  $\mathcal{R} = \mathcal{R}_t(x)$  be three measures respectively valued in  $\mathcal{S}_+$ ,  $\mathbb{R}^{n \times n \times d}$  and  $\mathbb{R}^{n \times n}$ . The continuity equation with source term is formally written

$$\partial_t \mathcal{S} + \operatorname{div}_x \mathcal{M}^{\operatorname{sym}} = \mathcal{R}^{\operatorname{sym}} \quad (4.17)$$

where the divergence operator applies on each coordinate of  $\mathcal{A} = \mathcal{M}^{\operatorname{sym}}$ , i.e.

$$\operatorname{div}_x \mathcal{A} = \left( \operatorname{div}_x \mathcal{A}_{ij} \right)_{ij} = \left( \sum_{k=1}^d \partial_k \mathcal{A}_{ij}[k] \right)_{ij} \in \mathbb{R}^{n \times n}. \quad (4.18)$$

At the same time we define the gradient operator by duality with the operator  $\operatorname{div}_x$  thanks to equality

$$\langle \operatorname{div}_x \mathbf{B} \mid \mathcal{Q} \rangle = -\langle \mathbf{B} \mid \operatorname{grad}_x \mathcal{Q} \rangle. \quad (4.19)$$

where  $\mathbf{B}$  is a smooth function of  $\Omega$  valued in  $\mathbb{R}^{n \times n \times d}$  and  $\mathcal{Q}$  a smooth function of  $\Omega$  valued in  $\mathbb{R}^{n \times n}$ . Notice that the definition of  $\operatorname{grad}_x$  given below encodes no flux boundary conditions.

**Definition 17** (Continuity equation with source term). Let  $\mathcal{S}_0, \mathcal{S}_1 \in \mathcal{M}(\Omega, \mathcal{S}_+)$ . A triplet  $(\mathcal{S}, \mathcal{M}, \mathcal{R}) \in \mathcal{M}([0, 1] \times \Omega, \mathcal{S}_+ \times \mathbb{R}^{n \times n \times d} \times \mathbb{R}^{n \times n})$  is a weak solution of the equation (4.17) with initial condition  $\mathcal{S}_0$  and final condition  $\mathcal{S}_1$  if the following equation is satisfied for all test functions  $\mathcal{Q} \in C_c^1([0, 1] \times \bar{\Omega}, \mathcal{S})$ :

$$\begin{aligned} & \int_0^1 \int_{\Omega} \partial_t \mathcal{Q}_t(x) \cdot d\mathcal{S}_t(x) dt + \int_0^1 \int_{\Omega} \operatorname{grad}_x \mathcal{Q}_t(x) : d\mathcal{M}_t(x) dt \\ & + \int_0^1 \int_{\Omega} \mathcal{Q}_t(x) \cdot d\mathcal{R}_t(x) dt = \int_{\Omega} \mathcal{Q}(1, x) \cdot d\mathcal{S}_1(x) - \int_{\Omega} \mathcal{Q}(0, x) \cdot d\mathcal{S}_0(x) \end{aligned} \quad (4.20)$$

and  $\mathcal{C}(\mathcal{S}_0, \mathcal{S}_1)$  will design the set of these triplets.

On  $\mathcal{S} \times \mathbb{R}^{n \times n \times d}$  we define the action  $\bar{A}$  as the sum of previous actions on  $\mathcal{S} \times \mathbb{R}^{n \times n}$ , that is

$$\bar{A}(\mathcal{S}, \mathbf{M}) = \sum_{k=1}^d A(\mathcal{S}, M[k]).$$

By Proposition (9),  $A = A(\mathcal{S}, R)$  and  $\bar{A} = \bar{A}(\mathcal{S}, \mathbf{M})$  are both 1-homogeneous with respect to  $(\mathcal{S}, R)$  and  $(\mathcal{S}, \mathbf{M})$  respectively and then  $\tilde{A} = \tilde{A}(\mathcal{S}, \mathbf{M}, R) \stackrel{\text{def}}{=} \bar{A}(\mathcal{S}, \mathbf{M}) + A(\mathcal{S}, R)$  is also 1-homogeneous. As a consequence we can define a nonnegative measure  $\mathcal{A} = \mathcal{A}(\mathcal{S}, \mathbf{M}, \mathcal{R})$  on  $[0, 1] \times \Omega$  by setting for  $B$  a borelian of  $[0, 1] \times \Omega$

$$\mathcal{A}(\mathcal{S}, \mathbf{M}, \mathcal{R})(B) = \int_B \tilde{A} \left( \frac{d(\mathcal{S}, \mathbf{M}, \mathcal{R})}{d\lambda} \right) d\lambda$$

where  $\lambda$  is any scalar measure on  $[0, 1] \times \Omega$  such that  $(|\mathcal{S}|, |\mathcal{M}|, |\mathcal{R}|) \ll \lambda$ . This definition does not depend on  $\lambda$ . For more details, see [35]. By notation abuse, we will write  $\mathcal{A}(\mathcal{S}, \mathcal{M}, \mathcal{R})$  instead of  $\mathcal{A}(\mathcal{S}, \mathcal{M}, \mathcal{R})([0, 1] \times \Omega)$ .

**Problem 7** (Primal dynamic problem). Given  $\mathcal{S}_0, \mathcal{S}_1 \in \mathcal{M}(\Omega, \mathcal{S}_+)$ , we consider the problem

$$(\mathcal{P}) = \inf \{ \mathcal{A}(\mathcal{S}, \mathcal{M}, \mathcal{R}) : (\mathcal{S}, \mathcal{M}, \mathcal{R}) \in \mathcal{C}(\mathcal{S}_0, \mathcal{S}_1) \}. \quad (4.21)$$

Problem (4.21) above can be written in the following more convenient way

$$(\mathcal{P}) = \inf \mathcal{A}(\mathcal{S}, \mathcal{M}, \mathcal{R}) + \iota_{\mathcal{C}(\mathcal{S}_0, \mathcal{S}_1)}(\mathcal{L}(\mathcal{S}, \mathcal{M}, \mathcal{R}))$$

where  $\mathcal{L} : C^1([0, 1], \mathcal{S} \times \mathbb{R}^{n \times n \times d} \times \mathbb{R}^{n \times n}) \rightarrow C^0([0, 1] \times \Omega, \mathcal{S})$  is given by

$$\mathcal{L}(\mathcal{S}, \mathcal{M}, \mathcal{R}) = -\partial_t \mathcal{S} - \operatorname{div}_x \mathcal{M}^{\operatorname{sym}} + \mathcal{R}^{\operatorname{sym}}$$

and  $\tilde{\mathcal{C}}$  is the closed convex given by

$$\tilde{\mathcal{C}}(\mathcal{S}_0, \mathcal{S}_1) = \left\{ \begin{array}{l} \Sigma \in C^0([0, 1] \times \Omega, \mathcal{S}) : \int_0^1 \mathcal{Q}_t \cdot \Sigma_t(x) dx dt = \int_{\Omega} \Delta_{\mathcal{S}_0}^{\mathcal{S}_1}(\mathcal{Q}(x)) dx, \\ \forall \mathcal{Q} \in C^1([0, 1] \times \Omega, \mathcal{S}) \end{array} \right\}.$$

Fenchel-Rockafellar theorem formally gives us again the equality

$$\begin{aligned} (\mathcal{P}) &= \inf \mathcal{A}(\mathcal{S}, \mathcal{M}, \mathcal{R}) + \iota_{\tilde{\mathcal{C}}(\mathcal{S}_0, \mathcal{S}_1)}(\mathcal{L}(\mathcal{S}, \mathcal{M}, \mathcal{R})) \\ &= \sup \iota_{\tilde{\mathcal{C}}(\mathcal{S}_0, \mathcal{S}_1)}^*(\mathcal{Q}) - \mathcal{A}^*(\mathcal{L}^*(\mathcal{Q})) \\ &\stackrel{\text{def}}{=} (\mathcal{D}) \end{aligned}$$

and then motivates again the next proposition.

**Proposition 18** ([35]).  $\mathcal{A}$  is convex, l.s.c., 1-homogeneous and its Fenchel conjugate is given by

$$\mathcal{A}^*(\overline{\mathcal{S}}, \overline{\mathcal{M}}, \overline{\mathcal{R}}) = \iota_{\mathcal{O}}(\overline{\mathcal{S}}, \overline{\mathcal{M}}, \overline{\mathcal{R}})$$

where

$$\mathcal{O} \stackrel{\text{def}}{=} \left\{ (\overline{\mathcal{S}}, \overline{\mathcal{M}}, \overline{\mathcal{R}}) \in \mathcal{M}([0, 1] \times \Omega, \mathcal{S} \times \mathbb{R}^{n \times n \times d} \times \mathbb{R}^{n \times n}) : \overline{\mathcal{S}} + \frac{1}{2} \overline{\mathcal{M}} \cdot \overline{\mathcal{M}}^* + \frac{1}{2} \overline{\mathcal{R}} \cdot \overline{\mathcal{R}}^* \leq 0 \right\}.$$

And the Fenchel conjugates of  $\iota_{\mathcal{C}(\mathcal{S}_0, \mathcal{S}_1)}$  is given by

$$\iota_{\mathcal{C}(\mathcal{S}_0, \mathcal{S}_1)}^*(\mathcal{Q}) = \int_{\Omega} \Delta_{\mathcal{S}_0}^{\mathcal{S}_1}(\mathcal{Q}(\cdot, x)) dx$$

Finally, a Fenchel-Rockafellar theorem is established and optimality conditions follow.

**Definition/Theorem 19** (Dynamic dual formulation and no duality gap). Let  $\mathcal{S}_0, \mathcal{S}_1 \in \mathcal{M}(\mathcal{S}_+)$ , a dual problem for (4.21) is given by

$$(\mathcal{D}) \stackrel{\text{def}}{=} \sup_{\mathcal{Q}} \int_{\Omega} \Delta_{\mathcal{S}_0}^{\mathcal{S}_1}(\mathcal{Q})(x) \, dx \quad (4.22)$$

subject to  $\mathcal{Q} \in C^1([0, 1] \times \Omega, \mathcal{S})$  and  $\partial_t \mathcal{Q} + \frac{1}{2} (\mathbf{grad}_x \mathcal{Q})^2 + \frac{1}{2} \mathcal{Q}^2 \leq 0$ . Moreover the following equality holds

$$(\mathcal{P}) = (\mathcal{D})$$

And problem  $(\mathcal{P})$  is attained.

*Proof.* The proof is an easy induction of the proof present in [35], Proposition 3.1. for  $n = k = m$  and  $\Lambda_1 = \Lambda_2 = \text{Id}$ .  $\square$

As a consequence of the previous theorem, we derive the optimality conditions.

**Proposition 20** (Optimality conditions). Let  $(\mathcal{S}, \mathcal{M}, \mathcal{R}) \in \mathcal{M}([0, 1] \times \Omega, \mathcal{S}_+ \times \mathbb{R}^{n \times n \times d} \times \mathbb{R}^{n \times n})$  be minimizer of  $(\mathcal{P})$ . The supremum of  $(\mathcal{D})$  is attained at  $\mathcal{Q} \in C^1([0, 1] \times \Omega, \mathcal{S})$  if and only if there exists  $\mathcal{S} \in \mathcal{M}([0, 1] \times \Omega, \mathcal{S}_+)$  such that

$$\begin{cases} (\mathcal{S}, \mathcal{S} \cdot \mathbf{grad}_x \mathcal{Q}, \mathcal{S} \cdot \mathcal{Q}) \in \mathcal{C}(\mathcal{S}_0, \mathcal{S}_1) \\ \partial_t \mathcal{Q} + (\mathbf{grad}_x \mathcal{Q})^2 + \mathcal{Q}^2 \leq 0 \\ \mathcal{S} \cdot (\partial_t \mathcal{Q} + (\mathbf{grad}_x \mathcal{Q})^2 + \mathcal{Q}^2) = 0. \end{cases}$$

## 4.5 Appendix : a few words about Riemannian point of view of Bures-Wasserstein metric

We give here a small presentation of the Riemannian point of view of Bures-Wasserstein metric when  $S_0$  and  $S_1$  are both invertible. This case has been studied by several authors and what follows is totally based on [39], [53] and [6] and [5].

**Optimal transport of Gaussians and Riccati equation.** In [39], McCann proposes to connect two probability measures  $\mu_0$  and  $\mu_1$  by a path  $\mu_t \in \mathcal{P}(\mathbb{R}^n)$ , the so-called McCann interpolant, defined thanks to the (convex) Brenier map  $\psi$  solution of  $\mu_1 = \nabla \psi \# \mu_0$ . The interpolant is then given by

$$\forall t \in [0, 1], \mu_t = ((1-t)\text{Id} + t\nabla \psi) \# \mu_0.$$

In Example 1.7. of the same paper, the case of Gaussian measures is discussed, more precisely, the author proves that this space, endowed with the square Wasserstein distance  $W_2$ , is geodesic and he gives a description of these geodesics. We recall that a measure on  $\mathbb{R}^n$  is called a (*centred*) *Gaussian measure* if there exists  $S \in S_{++}$  (its *covariance*) such

that its density  $f_S$  with respect to the Lebesgue measure on  $\mathbb{R}^n$  is given by

$$\forall x \in \mathbb{R}^n, f_S(x) = \frac{1}{(2\pi)^{n/2} (\det S)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}x^* S^{-1}x\right).$$

Let  $\mu_0$  (with covariance  $S_0$ ) and  $\mu_1$  (with covariance  $S_1$ ) be two Gaussian measures. The idea is to look for the Brenier map among the linear maps  $x \mapsto \Lambda x$  : this map is the gradient of  $x \mapsto \frac{1}{2}x^* \Lambda x$  and a necessary condition for this map to be convex is  $\Lambda$  to be positive semi-definite. The covariance of the second term is given by the following proposition, which is a consequence of the abstract change-of-variable variables theorem (see, e.g., [7]).

**Proposition 21** (Covariance of a linear pushforward Gaussian). Let  $\mu_0$  be a centered Gaussian measure with covariance  $S_0$  and  $\Lambda$  an endomorphism of  $\mathbb{R}^n$  (its canonical matrix representation will be denoted  $\Lambda$  also). The pushforward measure  $\mu_1 = \Lambda\#\mu_0$  is a centered Gaussian measure with covariance  $S_1 = \Lambda S_0 \Lambda^T$ .

As a consequence,  $\Lambda$  must satisfied the equation

$$\Lambda S_0 \Lambda^* = S_1. \quad (4.23)$$

Among all the solutions of this equation, only one is PSD, namely

$$\Lambda \stackrel{\text{def}}{=} S_1^{1/2} (S_1^{1/2} S_0 S_1^{1/2})^{-1/2} S_1^{1/2}.$$

The uniqueness in Brenier theorem allows us to conclude. Moreover, at any time  $t$  the measure remains a Gaussian measure. To summarize:

**Theorem 22** ([39]). The set of all Gaussian measures endowed with the  $L^2$ -Wasserstein metric is geodesic. And the geodesic between  $\mu_0$  and  $\mu_1$  is given by

$$\begin{aligned} \forall t \in [0, 1], \mu_t &= ((1-t)\text{Id} + t\Lambda)\#\mu_0, \\ \text{where } \Lambda &= S_1^{1/2} (S_1^{1/2} S_0 S_1^{1/2})^{-1/2} S_1^{1/2}. \end{aligned} \quad (4.24)$$

Formula (4.24) has already been highlighted in [24], [43], [27] and [33]. This interpolation between measures gives us naturally an interpolation between matrices of  $\mathcal{S}_{++}$ .

**Remark 23.** The equation (4.23) is known as the *Riccati equation* in the theory of control. Investigating symmetric solutions leads us to consider the equation

$$X S_0 X = S_1 \quad (\text{with } S_0 \in \mathcal{S}_{++}, S_1 \in \mathcal{S}_+). \quad (4.25)$$

It turns out that symmetric positive solution of (4.25) is unique since  $X S_0 X = Y S_0 Y$  implies  $(X S_0^{1/2})(X S_0^{1/2})^* = (Y S_0^{1/2})(Y S_0^{1/2})^*$  and then  $X$  and  $Y$  are equal modulo an orthogonal matrix.

Multiplying the equation (4.25) by  $S_0^{1/2}$  on both sides, then taking its square root and

finally multiplying the equation by  $S_0^{-1/2}$  on both sides gives us

$$S_0^{-1/2}(S_0^{1/2}XS_0X S_0^{1/2})^{1/2}S_0^{-1/2} = S_0^{-1/2}(S_0^{1/2}S_1S_0^{1/2})^{1/2}S_0^{-1/2}.$$

Taking  $X = S_0^{-1/2}(S_0^{1/2}S_1S_0^{1/2})^{1/2}S_0^{-1/2}$  suits well. Clearly, making the change of variable  $S_0 \leftarrow S_1^{-1}$  and  $S_1 \leftarrow S_0^{-1}$  does not change the solution and then by uniqueness, it turns out that

$$\begin{aligned} \Lambda &= S_1^{1/2} (S_1^{1/2}S_0S_1^{1/2})^{-1/2} S_1^{1/2} \\ &= S_0^{-1/2}(S_0^{1/2}S_1S_0^{1/2})^{1/2}S_0^{-1/2}. \end{aligned} \tag{4.26}$$

**Geometric mean of matrices.** The formula (4.24) seems quite obscure at first sight. Another way to introduce it is to see  $\Lambda = S_1^{1/2} (S_1^{1/2}S_0S_1^{1/2})^{-1/2} S_1^{1/2}$  as a *geometric mean* between  $S_0^{-1}$  and  $S_1$ . To understand this designation of *mean*, we describe a construction, following [5]. A convenient property asked to a *mean* in  $\mathbb{R}$  is to be 1-homogeneous, that is

$$\forall a, b \in \mathbb{R}_+, \forall x \in \mathbb{R}_+^*, M(xa, xb) = xM(a, b).$$

In the context of matrices, it seems natural to ask both a right and a left 1-homogeneity, that is, for all  $A, B > 0$  and  $X$  invertible,

$$\begin{aligned} M(XA, XB) &= XM(A, B) \text{ and} \\ M(AX, BX) &= M(A, B)X. \end{aligned}$$

An easy consequence of these two equalities is the *congruence invariance* of the mean

$$M(X^*AX, X^*BX) = X^*M(A, B)X. \tag{4.27}$$

Notice that  $(X, A) \rightarrow X^*AX \stackrel{\text{def}}{=} \Gamma_X(A)$  defines a group action on matrices which preserves the set of strictly positive matrices. With this notation, equalities (4.27) becomes

$$M(\Gamma_X(A), \Gamma_X(B)) = \Gamma_X(M(A, B)).$$

Now, for fixed  $A, B \in S_{++}$ , we will derive a formula for  $M(A, B)$  that satisfies (4.27). Notice first that

$$\begin{cases} \Gamma_{A^{-1/2}}(A) = \text{Id}, \\ \Gamma_{A^{-1/2}}(B) = A^{-1/2}BA^{-1/2}. \end{cases}$$

The matrix  $A^{-1/2}BA^{-1/2}$  is symmetric and positive. Let  $(U, D)$  be an orthogonal and a diagonal matrices such that  $A^{-1/2}BA^{-1/2} = U^*DU = \Gamma_U(D)$ , then:

$$\begin{cases} \Gamma_{A^{-1/2}U}(A) = \text{Id}, \\ \Gamma_{A^{-1/2}U}(B) = U^*A^{-1/2}BA^{-1/2}U = D, \end{cases}$$

and by setting  $X = A^{-1/2}U$ , (4.27) becomes

$$M(\text{Id}, D) = U^*A^{-1/2}M(A, B)A^{-1/2}U.$$

Since  $I$  and  $D$  commutes, it seems reasonable to set  $M(I, D) = D^{1/2}$ . But  $D^{1/2} = U^*(A^{-1/2}BA^{-1/2})^{1/2}U$ , a formula for  $M(A, B)$  derives:

$$M(A, B) = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}.$$

Eventually, by direct calculations, it is clear that  $M(A, B)$  is solution of  $XA^{-1}X = B$  and then, taking back notations  $S_0$  and  $S_1$ , it follows that  $M(S_0^{-1}, S_1) = \Lambda$  (according to (4.26)).

**Remark 24.** Note that the uniqueness of solutions in Riccati equation and inversibility of  $M(A, B)$  implies symmetry of  $M$ , that is,  $M(A, B) = M(B, A)$ .

**A geometric point of view.** We have just seen that the geodesic Wasserstein distance on probabilities induces a geodesic distance on Gaussian probabilities and so a geodesic distance on  $\mathcal{S}_{++}$ . Another point of view is possible (see [53]):  $\mathcal{S}_{++}$  can be viewed as a Riemannian manifold for a certain metric tensor  $g$ , and the induced distance  $d_g$  is actually  $W_2$ . See also [6] and [55].

**Remark 25.** Formula (4.6) is known as the *Bures distance* in quantum information theory. Writing it for diagonal matrices (say)  $D_1 = \text{diag}(\lambda_1, \dots, \lambda_n)$  and  $D_2 = \text{diag}(\mu_1, \dots, \mu_n)$ , we get:

$$d_B^2(D_1, D_2) = \sum_{i=1}^n \left( \sqrt{\lambda_i} - \sqrt{\mu_i} \right)^2$$

which is also the Hellinger distance between  $(\lambda_1, \dots, \lambda_n)$  and  $(\mu_1, \dots, \mu_n)$ , known as *Rao-Fisher* metric in information theory. This distance can be seen as the pushforward of the euclidian distance on  $\mathbb{R}^n$  by the pointwise square function  $f$ :

$$\begin{aligned} f : \quad \mathbb{R}^n &\quad \rightarrow \quad \mathbb{R}^n \\ x = (x_1, \dots, x_n) &\quad \mapsto \quad (x_1^2, \dots, x_n^2) \end{aligned}$$

See for instance [24] where  $d(S_0, S_1)^2 = \text{tr} S_0 + \text{tr} S_1 - 2 \text{tr}(\sqrt{S_0 S_1})$  for more information. In some way, the following construction is a generalization of this pushforward distance in a non commutative case.

Inspired by the previous remark, Takatsu proposed in [53] a geometric point of view of the  $L^2$ -Wasserstein metric (see also [55] and [29]). It consists of a construction of



a Riemannian metric on  $\mathcal{S}_{++}$ , which comes from the flat Frobenius metric  $G(Z, W) = \text{tr}(ZW^T)$  on  $\mathcal{G}$ . We now describe it.

Given  $A \in \mathcal{G}$ , the tangent space of  $\mathcal{G}$  at  $A$  is identified with  $\mathbb{R}^{n \times n}$ . The scalar product  $G$  defined by

$$G(Z, W) = \text{tr}(Z^*W)$$

is a (flat) Riemannian metric on  $\mathcal{G}$ . The map  $\Pi : \mathcal{G} \rightarrow \mathcal{S}_{++}$  defined by  $\Pi(A) = AA^*$  is smooth, surjective and its differential map is given by

$$d\Pi_A(Z) = ZA^* + AZ^*. \quad (4.28)$$

For  $S \in \mathcal{S}_{++}$  and  $A \in \mathcal{G}$  such that  $\Pi(A) = S$ , the tangent space  $T_A(\mathcal{G})$  (identified with  $\mathbb{R}^{n \times n}$ ) satisfied the following orthogonal decomposition:

$$T_A(\mathcal{G}) = \text{Ver}(\Pi, A) \oplus \text{Hor}(\Pi, A),$$

where  $\text{Ver}(\Pi, A)$  is the collection of tangent vectors to  $\Pi^{-1}(S)$  (called *vertical* vectors) and  $\text{Hor}(\Pi, A)$  is the collection of normal vectors to  $\Pi^{-1}(S)$  (called *horizontal* vectors). Easy calculations gives us:

$$\begin{aligned} \text{Ver}(\Pi, A) &= \{Z \in \mathbb{R}^{n \times n} : ZA^* \text{ is anti-symmetric} \} \\ \text{Hor}(\Pi, A) &= \{Z \in \mathbb{R}^{n \times n} : ZA^{-1} \text{ is symmetric} \}. \end{aligned}$$

In the following, for  $Z \in T_A(\mathcal{G})$ ,  $Z_H$  will denote the orthogonal projection of  $Z$  onto  $\text{Hor}(\Pi, A)$  (which depends of  $A$ ). To go further, we recall a definition and a theorem of Differential Geometry:

**Definition 26** (Riemannian submersion, see [55]). A submersion  $\Pi : (\mathcal{M}, g) \rightarrow (\mathcal{N}, h)$  is called Riemannian if for all  $p \in \mathcal{M}$ ,  $d\Pi_p|_{\text{Hor}(\Pi, p)}$  is a vector space isometry, that is, for all  $z, w \in \text{Hor}(\Pi, p)$ , the following holds:

$$g_p(z, w) = h_{\Pi(p)}(d\Pi_p(z), d\Pi_p(w)). \quad (4.29)$$

**Theorem 27.** Let  $\Pi : (\mathcal{M}, g) \rightarrow (\mathcal{N}, h)$  be a Riemannian submersion. If  $\gamma$  is a geodesic in  $(\mathcal{M}, g)$  such that  $\dot{\gamma}(0) \in \mathcal{H}(\pi, p)$ , then  $\pi \circ \gamma$  is a geodesic in  $(\mathcal{N}, h)$  and these two geodesics have the same length.

According to Theorem 27 above, to guarentee that  $\Pi$  is a Riemannian submersion, we define  $g$  the Riemannian metric on  $\mathcal{S}_{++}$  by the following formula,

$$g_{\Pi(A)}(d\Pi_A(Z), d\Pi_A(W)) = G(Z_H, W_H) = \text{tr}(Z_H W_H^*) \quad (4.30)$$

where  $A \in \mathcal{G}$  and  $Z, W \in \mathbb{R}^{n \times n}$ .

To go further, we rewrite this metric in a more common way by noting that we have

according to (4.28):

$$g_{\Pi(A)}(ZA^* + AZ^*, WA^* + AW^*) = \text{tr}(Z_H W_H^*).$$

Now note that  $Z \rightarrow ZA$  is a bijection from  $\mathbb{R}^{n \times n}$  to itself and then the change of variables  $Z \leftarrow ZA$  and  $W \leftarrow WA$  gives us

$$g_{\Pi(A)}(ZAA^* + AA^*Z^*, WAA^* + AA^*W^*) = \text{tr}([AZ]_H [AW]_H^*)$$

or by setting  $S = \Pi(A) = AA^* \in \mathcal{S}_{++}$ :

$$g_S(ZS + SZ^*, WS + SW^*) = \text{tr}([AZ]_H [AW]_H^*).$$

The expression  $ZS + SZ^*$  deserves our attention and is the object of the following lemma.

**Lemma 28.** Given  $S \in \mathcal{S}_{++}$ , the map  $\mathcal{L} : \mathbb{R}^{n \times n} \rightarrow \mathcal{S}$  defined by  $\mathcal{L}(Z) = ZS + SZ^*$  is linear, surjective and realizes a bijection between  $\mathcal{S}$  and itself.

*Proof.* Linearity is clear. Bijectivity comes from the fact that a solution  $Z$  of the equation

$$S^{-1}ZS = -Z^*$$

has only 0 as eigenvalue. Diagonalizability of  $Z$  concludes.  $\square$

Now, if we set  $X \stackrel{\text{def}}{=} ZS + SZ^*$  and  $Y \stackrel{\text{def}}{=} WS + SW^*$  and we note that  $\mathcal{L}_S^{-1}(X) \cdot A \in \text{H}(\Pi, A)$  since  $\mathcal{L}_S^{-1}(X) \in \mathcal{S}$ , we obtain the expression

$$g_S(X, Y) = \text{tr}(\mathcal{L}_S^{-1}(X) \cdot S \cdot \mathcal{L}_S^{-1}(Y)). \quad (4.31)$$

Finally, the metric induces by  $g$  is given by the following proposition.

**Proposition 29** ([53]). The metric  $g$  given in (4.31) induces the  $L^2$ -Wasserstein distance, that is  $d_g = W_2$ .

*Proof.* Let  $S_0, S_1 \in \mathcal{S}_{++}$ . We recall that

$$\Lambda = M(S_0^{-1}, S_1) = S_1^{1/2} (S_1^{1/2} S_0 S_1^{1/2})^{-1/2} S_1^{1/2}$$

and set

$$\begin{aligned} A &\stackrel{\text{def}}{=} S_0^{1/2} \\ Z &\stackrel{\text{def}}{=} [\Lambda - \text{Id}] S_0^{1/2}, \end{aligned}$$

so  $Z \in \text{Hor}(\Pi, A)$  since  $Z^* A^{-1}$  is symmetric. Moreover  $\Pi(A) = S_0$  and  $\Pi(A + Z) = S_1$ . The curve  $(Z(t))_{t \in [0,1]}$  defined by

$$Z(t) = A + tZ$$

is the geodesic between  $A$  and  $A+Z$  since  $(\mathcal{G}, G)$  is flat. According to previous discussion, since  $\dot{Z}(0) = Z$  is horizontal,  $\Pi$  is a Riemannian submersion and then  $\gamma \stackrel{\text{def}}{=} \Pi \circ Z$  is a geodesic between  $S_0$  and  $S_1$ . Finally, we obtain the following equalities:

$$\begin{aligned}
 W_2(S_0, S_1)^2 &= \text{tr } S_0 + \text{tr } S_1 - 2 \text{tr} \left( S_1^{1/2} S_0 S_1^{1/2} \right)^{1/2} \\
 &= \text{tr}(Z \cdot Z^*) \\
 &= G(Z, Z) \text{ by definition of } G \\
 &= g_{\Pi(A)}(d\Pi_A(Z), d\Pi_A(Z)) \text{ by (4.30)} \\
 &= d_g \left( \dot{Z}(0), \dot{Z}(0) \right)^2 \\
 &= d_g(S_0, S_1)^2 \text{ since } Z \text{ is a geodesic .}
 \end{aligned}$$

The proposition is established. □

**Remark 30.** As a consequence, the geodesic between  $S_0$  and  $S_1$  is given by:

$$(A + tZ) \cdot (A + tZ)^* = [(1 - t)\text{Id} + t\Lambda] S_0 [(1 - t)\text{Id} + t\Lambda]^* . \quad (4.32)$$

## Bibliography

- [1] Heinz H Bauschke, Patrick L Combettes, et al. *Convex analysis and monotone operator theory in Hilbert spaces*, volume 408. Springer, 2011.
- [2] Heinz H Bauschke and Adrian S Lewis. Dykstras algorithm with bregman projections: A convergence proof. *Optimization*, 48(4):409–427, 2000.
- [3] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [4] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017.
- [5] Rajendra Bhatia. *Positive definite matrices*. Princeton university press, 2009.
- [6] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2019.
- [7] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [8] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.
- [9] Yann Brenier and Dmitry Vorotnikov. On optimal transport of matrix-valued measures. *SIAM Journal on Mathematical Analysis*, 52(3):2849–2873, 2020.
- [10] Haim Brezis and Haim Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.
- [11] Yongxin Chen, Giovanni Conforti, and Tryphon T Georgiou. Measure-valued spline curves: An optimal transport viewpoint. *SIAM Journal on Mathematical Analysis*, 50(6):5947–5968, 2018.
- [12] Yongxin Chen, Wilfrid Gangbo, Tryphon T Georgiou, and Allen Tannenbaum. On the matrix Monge–Kantorovich problem. *European Journal of Applied Mathematics*, 31(4):574–600, 2020.
- [13] Yongxin Chen, Tryphon T Georgiou, Lipeng Ning, and Allen Tannenbaum. Matricial Wasserstein-1 distance. *IEEE control systems letters*, 1(1):14–19, 2017.
- [14] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Vector-valued optimal mass transport. *SIAM Journal on Applied Mathematics*, 78(3):1682–1696, 2018.
- [15] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Wasserstein geometry of quantum states and optimal transport of matrix-valued measures. In *Emerging Applications of Control and Systems Theory*, pages 139–150. Springer, 2018.
- [16] Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum. Interpolation of matrices and matrix-valued densities: The unbalanced case. *European Journal of Applied Mathematics*, 30(3):458–480, 2019.

- [17] Yongxin Chen, Eldad Haber, Kaoru Yamamoto, Tryphon T Georgiou, and Allen Tannenbaum. An efficient algorithm for matrix-valued and vector-valued optimal mass transport. *Journal of Scientific Computing*, 77(1):79–100, 2018.
- [18] Lenaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.
- [19] Lenaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.
- [20] Krzysztof J Ciosmak. Matrix Holder’s inequality and divergence formulation of optimal transport of vector measures. *SIAM Journal on Mathematical Analysis*, 53(6):6932–6958, 2021.
- [21] Roberto Cominetti and J San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1):169–187, 1994.
- [22] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- [23] J Dittmann. Explicit formulae for the Bures metric. *Journal of Physics A: Mathematical and General*, 32(14):2663, 1999.
- [24] DC Dowson and BV Landau. The fréchet distance between multivariate normal distributions. *Journal of multivariate analysis*, 12.
- [25] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and its applications*, 114:717–735, 1989.
- [26] Masatoshi Fujii. Riccati equation for positive semidefinite matrices (recent developments of operator theory by Banach space technique and related topics). *Mathematical Analysis Institute*, 2073:105–113, 2018.
- [27] Clark R Givens and Rae Michael Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2):231–240, 1984.
- [28] Shlomi Gover and Gershon Wolansky. On semi-discrete sub-partitions of vector-valued measures. *arXiv preprint arXiv:2007.03982*, 2020.
- [29] Anna Jenčová. Geodesic distances on density matrices. *Journal of Mathematical Physics*, 45(5):1787–1794, 2004.
- [30] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
- [31] Leonid V Kantorovich. On the translocation of masses. *Journal of mathematical sciences*, 133(4):1381–1382, 2006.
- [32] Philip A Knight. The Sinkhorn–Knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275, 2008.
- [33] Martin Knott and Cyril S Smith. On the optimal mapping of distributions. *Journal of Optimization Theory and Applications*, 43(1):39–49, 1984.

- 
- [34] Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Advances in Differential Equations*, 21(11/12):1117–1164, 2016.
- [35] Bowen Li and Jun Zou. On a general matrix valued unbalanced optimal transport and its fully discretization: dynamic formulation and convergence framework. *arXiv preprint arXiv:2011.05845*, 2020.
- [36] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.
- [37] Estelle Massart and P-A Absil. Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices. *SIAM Journal on Matrix Analysis and Applications*, 41(1):171–198, 2020.
- [38] Estelle Massart, Julien M Hendrickx, and P-A Absil. Curvature of the manifold of fixed-rank positive-semidefinite matrices endowed with the Bures–Wasserstein metric. In *International Conference on Geometric Science of Information*, pages 739–748. Springer, 2019.
- [39] Robert J McCann. A convexity principle for interacting gases. *Advances in mathematics*, 128(1):153–179, 1997.
- [40] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences de Paris*, 1781.
- [41] Léonard Monsaingeon and Dmitry Vorotnikov. The Schrödinger problem on the non-commutative Fisher-Rao space. *Calculus of Variations and Partial Differential Equations*, 60(1):1–48, 2021.
- [42] Luca Nenna. *Numerical methods for multi-marginal optimal transportation*. PhD thesis, Université Paris sciences et lettres, 2016.
- [43] Ingram Olkin and Friedrich Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.
- [44] Felix Otto. The geometry of dissipative evolution equations: the porous medium equation. 2001.
- [45] Gabriel Peyré. Entropic approximation of Wasserstein gradient flows. *SIAM Journal on Imaging Sciences*, 8(4):2323–2351, 2015.
- [46] 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.
- [47] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.
- [48] James B Robertson and Milton Rosenberg. The decomposition of matrix-valued measures. *Michigan Mathematical Journal*, 15(3):353–368, 1968.
- [49] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.
-

- [50] 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.
- [51] Thibault Séjourné, Jean Feydy, François-Xavier Vialard, Alain Trounev, and Gabriel Peyré. Sinkhorn divergences for unbalanced optimal transport. *arXiv preprint arXiv:1910.12958*, 2019.
- [52] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.
- [53] Asuka Takatsu. On Wasserstein geometry of Gaussian measures. In *Probabilistic approach to geometry*, pages 463–472. Mathematical Society of Japan, 2010.
- [54] Yann Thanwerdas and Xavier Pennec. Bures-Wasserstein minimizing geodesics between covariance matrices of different ranks. *arXiv preprint arXiv:2204.09928*, 2022.
- [55] Jesse van Oostrum. Bures-Wasserstein geometry. *arXiv preprint arXiv:2001.08056*, 2020.
- [56] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
- [57] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.
- [58] Ruodu Wang and Zhenyuan Zhang. Simultaneous optimal transport. *arXiv preprint arXiv:2201.03483*, 2022.