université
PARIS-SACLAY

# Sur quelques généralisations du problème du transport optimal

**Habilitation à diriger des recherches**
**de l'Université Paris-Saclay**

**présentée et soutenue à Orsay, le 06/03/2024, par**


## Luca NENNA


### Composition du jury

| | |
|---|---|
| **Yann BRENIER**<br>DR, CNRS et Université Paris-Saclay | Rapporteur |
| **Young-Heon KIM**<br>Professeur des universités, UBC | Rapporteur |
| **Gabriel PEYRÉ**<br>DR, CNRS et ENS Paris | Rapporteur |
| **Julie DELON**<br>Professeure des universités, Université Paris-Cité | Examinatrice |
| **Jean-Marie MIREBEAU**<br>DR, ENS Paris-Saclay | Examinateur |
| **Simona ROTA-NODARI**<br>Professeure des universités, Université Côte d'Azur | Examinatrice |

Habilitation à diriger des recherches

*per aspera ad astra*

# On pains au chocolat, electrons, entropy and the theory (and numerics) to rule them all.

# Remerciements

Le temps passe vite, même trop vite, et l'écriture de ce manuscrit pour l'Habilitation à Diriger les Recherches a été l'occasion pour moi de mesurer une fois de plus l'immense chance que j'ai de faire un travail qui me passionne. Je souhaiterais profiter de cette occasion pour adresser mes plus profonds remerciements aux différentes personnes qui m'ont permis d'avoir cette chance.

Tout d'abord, je voudrais remercier très chaleureusement Jean-David Benamou et Guillaume Carlier pour avoir accepté, il y a longtemps désormais d'encadrer ma thèse de doctorat. Je pense que si j'avais fait un choix différent il y a presque 11 ans, je ne serais pas si heureux (et fier) du travail accompli et plein d'énergie pour tout le travail qui m'attend. Encore merci pour tout ce que vous m'avez donné.

I am deeply thankful to Gabriel Peyré and Young-Heon Kim for accepting to be referees of this thesis. I am glad that they appreciated this work and I have found their comments very inspiring.

Je tiens également à remercier Yann Brenier (merci pour les discussions sur les mathématiques, la politique et les trains...ça a été et ça sera toujours un vrai plaisir!!), Julie Delon, Jean-Marie Mirebeau et Simona Rota-Nodari qui m'ont fait l'honneur d'être les membres de mon jury.

Un moka-merci à l'équipe MOKAPLAN (Vincent, JD, Antonin, Irène, Paul, Flavien, Derya et tous les moka-thésards et aussi les moka-anciens) qui m'a accueilli pendant ma thèse et encore aujourd'hui me donne abri au moins une fois par semaine dans ses locaux sur Paris! Vous devez être fiers de cette communauté que vous avez construit et que vous continuez à construire. Toujours côté Inria, je voudrais remercier l'équipe MATHERIALS pour m'avoir accueilli en délégation ces deux dernières années (délégation qui m'a permis de rédiger ce manuscrit). Bien sûr je peux pas oublier le 4éme étage du bâtiment A

J'ai rencontré beaucoup trop de personnes passionnantes ces dernières années (une pensée spéciale pour Rodrigue Lelotte et Mathias Dus avec son alter ego J.-C. ); parmi elles merci encore à mes collaborateur.trice.s : Brendan Pass, Mathieu Lewin, Simone Di Marino, Paul Pegon, Virginie Ehrlacher, Luigi De Pascale, Quentin Mérigot, Léonard Monsaigeon, Jean-Baptiste Casteras, Hamza Ennaji, Augusto Gerolin...merci à vous tou.te.s car à chaque discussion, à chaque pause café ou match de tennis vous m'apprenez avec patience (et il faut beaucoup de patience avec moi) énormément de choses sur la vie et sur les maths. Je sais qu'il n'est pas toujours facile de travailler avec moi, mais j'espère que vous avez apprécié nos résultats et le temps que nous avons passé (et que nous passerons encore) ensemble.
Merci à Maxime, Paul, Quentin, Thomas et Léonard pour leur relectures, pas seulement de ce document, (oui Thomas j'ai copié tes remerciements ;) mais c'est factuel), pour GOTA, pour

ParMA, les matchs de tennis et les discussions de maths autour d'une bière fraîche sous le ciel bleu de Lisbonne.

Merci au Laboratoire de Mathématiques d'Orsay, et surtout à l'équipe ANEDP, pour m'avoir fait confiance et recruté il y a quelques années comme McF. J'espère que ce manuscrit puisse témoigner que la confiance n'était pas mal placée. Merci à Jean-Marie, Maya et Cyril pour avoir partagé et partager encore le bureau 2F1 avec la vue (plus ou moins) sur la bucolique vallée de l'Yvette. Et surtout merci à Thomas Galloüet, Quentin Mérigot, Yann Brenier, Bertrand Maury, Bruno Lévy, Hugo Leclerc et Sylvain Faure avec qui on commence cette nouvelle aventure avec l'équipe ParMA.

Merci à vous tou.te.s avec qui je partage ma vie à Paris (Ci, Simon, Max, Ricky, JLO, Dario, Franchie, Vituzzo, Meggie, André, Pierre, Damian, Paul, Samra...et les plus petits Lupo et VL...et encore les anciens parisiens Vale, Paolo, Anna et Davide...et j'ai peut-être oublié queluq'un), mes vacances, mes séjours à Lisbonne ou en Italie, à mes amis de toujours (Laura, Claudio, Monica, Daniele, Alby, Paolo, Carlotta, Eleonora×2), à mes parents, à ma famille et à Antonietta et Roger qui m'ont accueilli il y a presque 11 ans et m'ont permis de faire de Paris ma maison. À vous tou.te.s, je veux dédier ce manuscrit même si je sais que c'est rien par rapport au support, à l'amitié et à l'amour que vous me donnez chaque jour.

# Contents

# Notations

In all this manuscript, $d \in \mathbb{N}^*$ denotes the dimension of the ambient space $\mathbb{R}^d$ and $m \in \mathbb{N}$ is an integer such that $m \geq 2$.

| | |
|---|---|
| $X$ | a subset of $\mathbb{R}^d$ |
| $X_i$ | a subset of $\mathbb{R}^d$ for any index $i \in \{1, \dots, m\}$ |
| $\boldsymbol{X}$ | product $X_1 \times \dots \times X_m$ whenever $(X_i)_{1 \leq i \leq m}$ is a family of $m$ subsets of $\mathbb{R}^d$; |
| $\boldsymbol{X}_{-i}$ | product $\prod_{1 \leq j \leq m, j \neq i} X_j$ |
| $x_i, x, \boldsymbol{x}$ | a point in $X_i$, in some $X$, and in $\boldsymbol{X}$ respectively; |
| $x_q$ | $(x_i)_{i \in q}$ if $q \subseteq \{1, \dots, m\}$ and $\boldsymbol{x} \in \boldsymbol{X}$; |
| $e_i$ | $i$-th coordinate map $e_i : \boldsymbol{x} = (x_1, \dots, x_m) \mapsto x_i$; |
| $A \Subset B$ | $A \subseteq K \subseteq B$ for some compact set $K$; |
| $\langle \cdot, \cdot \rangle$ | Euclidean scalar product; |
| $\lvert \cdot \rvert$ | Euclidean norm on $\mathbb{R}^d$; |
| $\lVert \cdot \rVert$ | norm on $\mathbb{R}^d \times \dots \times \mathbb{R}^d$ defined by $\lVert \boldsymbol{x} \rVert = \max_{1 \leq i \leq m} \lvert x_i \rvert$ if $\boldsymbol{x} = (x_1, \dots, x_m)$; |
| $B_r(\boldsymbol{x})$ | open ball of radius $r$ centered at $\boldsymbol{x} \in (\mathbb{R}^d)^m$ for the above norm; |
| $\mathcal{C}^{0,1}_{\mathrm{loc}}(X)$ | space of real-valued locally Lipschitz functions on $X$ which is a sub-manifold of $\mathbb{R}^d$ or $(\mathbb{R}^d)^m$; |
| $[f]_{\mathcal{C}^{0,1}(X)}$ | Lipschitz constant of $f : X \to \mathbb{R}$ where $X$ is a subset of $\mathbb{R}^d$ or $(\mathbb{R}^d)^m$ for the above norms; |
| $\mathcal{C}^{1,1}_{\mathrm{loc}}(X)$ | space of differentiable real-valued functions on $X$, a sub-manifold of $\mathbb{R}^d$ or $(\mathbb{R}^d)^m$, with locally Lipschitz differential; |
| $\mathcal{P}(X)$ | space of probability measures on a metric space $X$; |
| $\Pi(\mu_1, \dots, \mu_m)$ | space of probability measures on $\boldsymbol{X}$ having $\mu_1, \dots, \mu_m$ as marginals; |
| $\Pi(\rho)$ | space of probability measures on $\boldsymbol{X}$ having all marginals equals to $\rho$; |
| $\lVert \cdot \rVert_{L^p(\mu)}$ | $L^p$ norm induced by a measure $\mu$, where $p \in [1, +\infty]$; |
| $\operatorname{spt} \mu$ | support of the measure $\mu$; |
| $\mathcal{H}^s_X$ | $s$-dimensional Hausdorff measure on the metric space $X$ endowed with the Borel $\sigma$-algebra (the subscript $X$ will often be dropped); |
| $M_d(\mathbb{R})$ | space of real matrices of size $d \times d$, endowed with the Frobenius norm induced by the scalar product $A \cdot B \coloneqq \mathrm{Tr}\,(A^T B)$, for $A, B \in M_d(\mathbb{R})$; |
| $S_d(\mathbb{R})$ | subspace of real symmetric matrices of size $d \times d$; |
| $\Delta_P$ | simplex of P-uples $t = (t_p)_{p \in P}$ such that $t_p \geq 0$ for all $p \in P$ and $\sum_{p \in P} t_p = 1$; |
| $\mathcal{H}^m_1$ | is the set of admissible electronic wavefunctions for a system of $m$ electrons with finite kinetic energy, that is $\mathcal{H}^m_1 := \bigwedge_{i=1}^m H^1(\mathbb{R}^3)$. In the same fashion we define $\mathcal{H}^m_0 := \bigwedge_{i=1}^m L^2(\mathbb{R}^3)$ and $\mathcal{H}^m_0 := \bigwedge_{i=1}^m H^2(\mathbb{R}^3)$; |
| $\mathfrak{S}^+_1(\mathcal{H}^m_0)$ | denotes the set of trace-class self-adjoint non-negative operators on $\mathcal{H}^m_0$. |

# Introduction (en français)

Le transport optimal (OT) est un domaine de recherche très dynamique et il y a tant d'exemples et d'applications qu'il est difficile de choisir l'un d'entre eux pour introduire ce sujet fascinant; nous mentionnons les manuels "old and new" comme [AGS04; Vil09; Vil03; San15; Gal16; PC19]. Nous avons décidé d'introduire le problème du transport optimal classique en utilisant un exemple donné par Villani dans [Vil03].

Situons notre contexte à Paris et prenons en considération l'ensemble des boulangeries qui produisent des pains au chocolat/chocolatines (le lecteur peut bien sûr choisir la viennoiserie qui lui plaît le plus). Ces derniers doivent être livrés chaque matin aux cafés dans lesquels les clients pourront les goûter. La production et la consommation de pains au chocolat sont décrites par $\mu_1$ et $\mu_2$, respectivement. Nous supposons que la quantité total de la production et de la consommation est la même. Le problème du transport optimal consiste alors à transporter la quantité de pains au chocolat/chocolatines produite par la boulangerie $x_1 \in X_1$ ($X_1 \subset \mathbb{R}^{d_1}$ est l'ensemble des boulangeries caractérisées par $d_1$ propriétés e.g., emplacement précis, taille, etc.) à un café $x_2 \in X_2$ ($X_2 \subset \mathbb{R}^{d_2}$ est l'ensemble des cafés caractérisés par $d_2$ propriétés) de sorte que le coût de transport $c(x_1, x_2)$, par exemple la distance entre la boulangerie et le café, soit minimal. En transport optimal classique, nous avons $d_1 = d_2$ , mais dans ce qui suit, nous allons essayer d'aborder le problème lorsque le nombre de caractéristiques (alias les dimensions) diffère. Le problème introduit par Monge dans [Mon81] consiste à chercher un transport $x_2 = T(x_1)$ qui nous indique le café $x_2$ où l'intégralité des pains au chocolat produits par $x_1$ sera délivrés. Toutefois, ce problème est difficile à traiter et on nécessite une relaxation (connue sous le nom de problème de **Monge-Kantorovich**) : on permet, par exemple, qu'une partie des pains au chocolat produite en $x_1$ soit envoyée au café $x_2$ et l'autre à $x_2'$. Dans ce cas, nous cherchons un couplage optimal $\gamma(x_1, x_2)$ qui nous indique comment la masse en $x_1$ est répartie sur chaque $x_2 \in X_2$. Le problème peut alors être formulé comme suit : étant donné deux mesures de probabilité $\mu_1 \in \mathcal{P}(X_1)$ et $\mu_2 \in \mathcal{P}(X_2)$ et une fonction de coût continue $c : X_1 \times X_2 \to \mathbb{R}$ le problème de *Monge-Kantorovich* est définie comme suit

$$\boxed{\mathsf{OT}^c(\mu_1, \mu_2) := \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{X_1 \times X_2} c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2),} \tag{1}$$

où $\Pi(\mu_1, \mu_2)$ désigne l'ensemble des mesures de probabilité sur $X_1 \times X_2$ ayant $\mu_1$ et $\mu_2$ comme marginales, c'est-à-dire qu'étant donnée la projection canonique $e_i : X_1 \times X_2 \to X_i$, nous avons $e_{i,\sharp}\gamma = \mu_i$ pour $i = 1, 2$.

Si le couplage optimal $\gamma(x_1, x_2)$ consiste à assigner un unique café à chaque boulangerie $x_1$, alors on dit que le couplage est *déterministe* et qu'un transport optimal $T$ existe. Lorsque le coût $c(x_1, x_2)$ est la distance au carré, l'existence d'un couplage déterministe a été prouvée par Yann Brenier dans [Bre91] ; de plus, dans ce cas, (1) est une distance au carré sur l'ensemble

des mesures de probabilité.
De nombreuses questions se posent à ce stade : que se passe-t-il si le nombre de caractéristiques des boulangeries et des cafés est différent (e.g., $d_1 \neq d_2$)? Et si l'une des distributions est inconnue et que nous souhaitons l'obtenir en minimisant le $\mathsf{OT}^c$ plus une autre fonctionnelle ? De plus, nous pourrions ajouter une troisième catégorie, les hôtels, de sorte que nous devons maintenant minimiser le coût d'envoi des pains au chocolat/chocolatines à la fois aux cafés et aux hôtels à Paris. Ce problème peut-il encore être interprété comme un problème d'OT ? Ce manuscrit (qui résume en partie mes recherches après la thèse) est, en fait, consacré à répondre à ces questions et à quelques autres concernant les modèles de transport optimal apparaissant en physique, en particulier en physique quantique.
Plus en détail, ce manuscrit est divisé en trois chapitres organisés, selon la thématique, comme suit :

- Le **chapitre 1** est consacré (1) au transport optimal multi-marginales et à une application aux mesures de risque, (2) aux taux de convergence pour le transport optimal multi-marginales entropique et, enfin, (3) à une caractérisation du transport optimal multi-marginales entropique (discret) par une EDO. Mes co-auteurs sur ces sujets sont H. Ennaji, Q. Mérigot, B. Pass et P. Pegon et les résultats des articles suivants [Enn+22; NP23c; NP23a] y sont résumés ;

- Dans le **chapitre 2** nous décrivons une généralisation du transport optimal (1) au cas où nous laissons le nombre de marginales varier et (2) au cadre quantique dans lequel on propose une approximation parcimonieuse de la fonctionnelle de Lieb. Mes co-auteurs sur ces sujets sont S. Di Marino, M. Lewin et V. Ehrlacher et nous nous concentrons sur les articles suivants [DLN22; EN23]

- Dans le **chapitre 3** nous considérons le transport optimal lorsque les dimensions des marginales (à savoir le nombre de caractéristiques dans l'exemple ci-dessus) sont différentes et certains problèmes variationnels connexes découlant de la théorie des jeux (e.g., l'équilibre de Cournot-Nash). De plus, nous proposons une classe de métriques sur l'ensemble des mesures de probabilité, impliquant des mesures singulières, qui nous aident à prouver la convergence d'une méthode numérique pour résoudre les problèmes variationnels mentionnés ci-dessus. Mon co-auteur sur ces sujets est B. Pass et nous considérons les résultats obtenus dans [NP20; NP23b].

Des perspectives sont presentées à la fin de chaque chapitre. Nous allons maintenant présenter plus précisément le contenu des différents chapitres.

## Transport optimal multi-marginales et régularisation entropique

Au début de cette introduction, nous avons présenté le transport optimal classique dans le cas où nous devons comparer deux populations, à savoir les boulangeries et les cafés. Cependant, il est très naturel de traiter plusieurs populations (boulangeries, cafés, hôtels, supermarchés, etc.), de sorte qu'une manière simple de les comparer est de définir le problème de transport optimal multi-marginales suivant : étant donné $m$ distributions $\mu_i \in \mathcal{P}(X_i)$, avec $X_i \subset \mathbb{R}^{d_i}$ (pour simplifier, nous pouvons prendre $d_i = d$ pour tout $i = 1, \ldots, m$), et une fonction de coût continue

$c : \boldsymbol{X} \to \mathbb{R}$, le problème d'optimisation se lit comme suit

$$\boxed{\mathsf{MOT}^c(\mu_1, \ldots, \mu_m) := \inf_{\gamma \in \Pi(\mu_1,\ldots,\mu_m)} \int_{\boldsymbol{X}} c(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m),} \tag{2}$$

où $\Pi(\mu_1, \ldots, \mu_m)$ désigne l'ensemble des mesures de probabilité sur $\boldsymbol{X} := \prod_{i=1}^m X_i$ ayant $\mu_i$ comme marginales. Il s'avère que ce type de problème se pose dans de nombreux domaines d'application, y compris l'économie, [CE10], les mathématiques financières [BHP13; DS14a; DS14b; Enn+22], les statistiques [BK18; CCG16], le traitement d'images [Rab+11], la tomographie [Abr+17], l'apprentissage automatique [Haa+21; TJK22], la dynamique des fluides [Bre89] et la physique et la chimie quantique, dans le cadre de la théorie de la fonctionnelle de la densité [BDG12; CFK13]. La structure des solutions du problème de transport optimal multi-marginales est une question notoirement délicate, et n'est pas encore bien comprise, malgré les efforts considérables déployés par de nombreux chercheurs [GŚ98; Car03; CN08; Hei02; Pas11b; Pas12; KP14b; KP15; CS16; PV21; CDD15; PV22; MP17] ; voir aussi [Pas15] et [DGN17].
La classe des problèmes de transport optimal multi-marginales unidimensionnels qui se posent dans le contexte de l'estimation du risque présente un intérêt particulier. Dans ce cadre, pour des raisons de sécurité, il est nécessaire d'avoir une estimation pessimiste du risque. De tels problèmes se posent, par exemple, pour évaluer le risque d'une installation industrielle proche d'une rivière et protégée par une digue (voir [IL15]). En particulier, on sait comment évaluer un risque (tel que le niveau d'eau dans la rivière, comparé à la hauteur de la digue) en fonction de quelques variables (par exemple, la largeur de la rivière, le débit annuel maximal, etc. Ce que l'on ne sait pas, en revanche, c'est le **couplage** entre ces variables qui permet d'obtenir le scénario le plus défavorable. En gros, nous voulons résoudre le problème de maximisation suivant

$$\boxed{\max_{\gamma \in \Pi(\mu_1,\ldots,\mu_m)} \mathcal{R}_\alpha(b_\# \gamma),} \tag{3}$$

où $\mathcal{R}_\alpha(\mu) := \int_0^1 F_\mu^{-1}(t)\alpha(t)\mathrm{d}t$ est la mesure spectrale du risque pour une fonction non négative et non décroissante $\alpha : [0, 1] \to \mathbb{R}_+$ avec $\int_0^1 \alpha(t)\mathrm{d}t = 1$, $F_\mu^{-1}$ est la pseudo-inverse de la fonction de répartition de $\mu$ et $b$ est la fonction output (e.g., elle représente la hauteur simulée d'une rivière à risque d'inondation). Dans [Enn+22], nous parvenons **(1)** à montrer que pour tout problème de mesures de risque (3) peut être reformulé comme un problème $\mathsf{MOT}^c$ avec $m + 1$ marginales ; **(2)** à prouver que la solution peut être explicitement caractérisée en utilisant un raffinement de la théorie existante du transport multi-marginales sur un espace ambiant unidimensionnel ; **(3)** à étudier le cas dans lequel les variables sont en dimension supérieure ; **(4)** à étendre au cadre dans lequel la fonction output est multivariée.

Une façon naturelle de résoudre numériquement le problème multi-marginales présenté ci-dessus est de recourir à la régularisation entropique, ce qui nous amène au deuxième type de généralisation du transport optimal présenté dans ce manuscrit : **le transport optimal multi-marginales entropique**. Étant données comme avant $m$ mesures de probabilité $\mu_i$ et une fonction coût $c$, le problème de minimisation est défini comme suit

$$\boxed{\mathsf{MOT}^c_\varepsilon(\mu_1, \cdots, \mu_m) := \inf_{\gamma \in \Pi(\mu_1,\ldots,\mu_m)} \int_{\boldsymbol{X}} c \mathrm{d}\gamma + \varepsilon \mathrm{Ent}(\gamma | \otimes_{i=1}^m \mu_i),} \tag{4}$$

où $\mathrm{Ent}(\cdot| \otimes_{i=1}^m \mu_i)$ est l'entropie relative de Boltzmann-Shannon (ou la divergence de Kullback-Leibler) par rapport à la mesure produit $\otimes_{i=1}^m \mu_i$. Grâce aux propriétés algébriques de l'entropie la méthode des projections alternées sur les $m$ contraintes de marginal correspond, dans ce cas, au célèbre algorithme de Sinkhorn [Sin67] qui est aujourd'hui largement utilisé pour résoudre numériquement les problèmes de transport optimal et les problèmes variationnels connexes (voir par exemple [PC19]). Puisque $\mathsf{MOT}^c_\varepsilon$ peut être considéré comme une perturbation de $\mathsf{MOT}^c$, il est naturel d'en étudier le comportement lorsque $\varepsilon$ tend vers 0. Cela nous a motivés, voir [NP23c], à étudier le taux de convergence du coût entropique $\mathsf{MOT}^c_\varepsilon$ vers $\mathsf{MOT}^c_0$ sous certaines hypothèses sur les fonctions coût et les marginales. Les résultats que nous avons obtenus avec P. Pegon, et détaillés dans la section 1.4, peuvent être résumés comme suit : **(1)** Nous établissons deux **bornes supérieures**, l'une valable pour des coûts localement Lipschitz et une autre plus fine valable pour des coûts localement semi-concaves. Pour cette dernière, nous améliorons la borne supérieure (obtenue pour les coûts localement Lipschitz) d'un facteur de 1/2, obtenant l'inégalité suivante pour une certaine $C^\star \in \mathbb{R}_+$ :

$$\boxed{\mathsf{MOT}^c_\varepsilon \leq \mathsf{MOT}^c_0 + \frac{1}{2}\left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)\varepsilon\log(1/\varepsilon) + C^\star\varepsilon.} \tag{5}$$

**(2)** Pour la **borne inférieure**, en fonction d'un $\kappa$ dépendant d'une condition de signature sur les dérivées secondes mixtes du coût, nous avons

$$\boxed{\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 + \frac{\kappa}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon.} \tag{6}$$

pour une certaine $C_* \in \mathbb{R}_+$. À remarquer que quand $\kappa = \left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)$ on obtient un développement à l'ordre 1 du coût entropique. Ces résultats nous ont également permis d'étendre les résultats de [CPT23; EN22] aux fonctions de coût dégénérées dans le cas des deux marginales.

La dernière section de ce chapitre 1.5 traite d'une caractérisation du problème de transport optimal multi-marginales entropique discret pour le cas d'une fonction coût *pairwise*, c'est-à-dire $c(x_1,\dots,x_m) = \sum_{i<j} w(x_i,x_j)$ avec $w$ symétrique, via une équation différentielle ordinaire. Des exemples de fonctions coût de ce type sont le coût Gangbo-Święch où $w(x_i,x_j) = |x_i - x_j|^2$, qui est lié au problème du barycentre de Wasserstein [AC11], et le coût de Coulomb $w(x_i,x_j) = 1/|x_i - x_j|$, qui apparaît dans le cadre de la théorie fonctionnelle de la densité [CFK13; BDG12]. Supposons maintenant que $X_i = X$, pour tout $i$, sont des ensembles finis et que les marginales sont des sommes de diracs. Dans le cas où les marginales sont toutes égales $\mu_i = \rho = \sum_{x\in X}\rho_x\delta_x$ (e.g., en chimie quantique car elles représentent des électrons indiscernables), on peut montrer que le problème discret dual de (4) prend la forme suivante (pour simplifier, nous conservons la notation du cas continu, mais les intégrales doivent être comprises comme des sommes) :

$$\inf_{\varphi:X\to\mathbb{R}}\left\{\tilde{\Phi}(\varphi,\eta)\right\}, \tag{7}$$

où $\eta > 0$ est un paramètre qui permet d'ajouter/enlever certaines interactions entre les marginales,

$$\tilde{\Phi}(\varphi,\eta) := -(m-1)\int_X \varphi \mathrm{d}\rho + \varepsilon\int_X \log\left(\int_{X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi(x_i) - c_\eta(\boldsymbol{x})}{\varepsilon}\right)\mathrm{d}\otimes^{m-1}\rho\right)\mathrm{d}\rho,$$

et

$$c_\eta(x_1,\dots,x_m) := \eta\sum_{i=2}^m\sum_{j=i+1}^m w(x_i,x_j) + \sum_{i=2}^m w(x_1,x_i).$$

Notez que lorsque $\eta = 1$, nous retrouvons le coût pair-wise défini ci-dessus et nous devons minimiser un seul potentiel car nous exploitons l'indiscernabilité des marginales. Ensuite, en différenciant la condition d'optimalité de (20) par rapport à $\eta$, nous obtenons le problème de Cauchy suivant

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\eta),\eta)]^{-1}\dfrac{\partial}{\partial\eta}\nabla_\varphi\tilde{\Phi}(\varphi(\eta),\eta), \\ \varphi(0) = \varphi_w, \end{cases}} \tag{8}$$

où la valeur initiale $\varphi(0)$ de $\varphi$ lorsque $\eta = 0$ coïncide avec le potentiel optimal $\varphi_w$ pour le problème de transport optimal à deux marginales avec le coût $w$. Après avoir prouvé que l'EDO dans (21) est bien posée, on peut l'utiliser pour caractériser le problème multi-marginales entropique original à $\eta = 1$. En exploitant ensuite la régularité du membre de droite, nous pouvons résoudre numériquement (21) en utilisant des méthodes d'ordre élevé, montrant ainsi que c'est (parfois) une approche préférable (en termes de nombre d'itérations pour converger et de temps CPU) à celle de Sinkhorn. Nous concluons le chapitre en utilisant cette approche EDO également pour le cas où la fonction coût est associée à la discrétisation temporelle du principe de Brenier pour les équations d'Euler incompressibles [Bre89].

## Transport optimal en physique mathématique

Nous continuons à explorer certaines généralisations du transport optimal, en particulier le transport multi-marginales, en nous concentrant sur l'application du transport optimal dans la théorie de la fonctionnelle de densité (DFT). L'une des quantités clés de la DFT est celle que l'on appelle la fonctionnelle de Levy-Lieb, qui est l'énergie minimale (cinétique plus interaction électron-électron) d'un système quantique ayant $\rho$ comme densité électronique prescrite. En d'autres termes, nous devons résoudre le problème de minimisation suivant

$$\boxed{F_{LL}[\rho] := \inf_{\substack{\Psi\in\mathcal{H}^m_1 \\ \rho_\Psi=\rho}} \frac{1}{2}\int_{\mathbb{R}^{3m}} |\nabla\Psi|^2 + \int_{\mathbb{R}^{3m}} V|\Psi|^2,} \tag{9}$$

où $\mathcal{H}^m_1$ est l'ensemble des fonctions d'onde antisymétriques $\Psi$ pour un système de $m$ électrons avec une énergie cinétique finie, la fonction $V$ est le potentiel d'interaction coulombienne électron-électron, et $\rho_\Psi$ est la densité électronique associée à la fonction d'onde $\Psi$, c'est-à-dire la fonction réelle définie sur $\mathbb{R}^3$ comme suit

$$\forall x \in \mathbb{R}^3, \quad \rho_\Psi(x) := m\int_{(\mathbb{R}^3)^{m-1}} |\Psi(x, x_2, \ldots, x_m)|^2 \,\mathrm{d}x_2 \ldots \mathrm{d}x_m.$$

Il se trouve que $F_{LL}$ n'est pas convexe, il est donc pratique de considérer une convexification proposée par Lieb [Lie83], où la minimisation est effectuée sur l'ensemble des états mixtes plutôt que sur l'ensemble des états purs comme dans (9): nn gros, les fonctions d'onde sont maintenant remplacées par des opérateurs et l'énergie dans (9) est exprimée en termes de traces. Plus précisément, nous considérons le problème de minimisation sur l'ensemble $\mathfrak{S}^+_1$ des opérateurs auto-adjoints non négatifs de trace 1 ayant $\rho$ comme marge, c'est-à-dire

$$\boxed{F_L[\rho] := \inf_{\substack{\Gamma\in\mathfrak{S}^+_1 \\ \rho_\Gamma=\rho}} \mathrm{Tr}\,(H_m\Gamma),} \tag{10}$$

où $H_m := -\frac{1}{2}\Delta + V$. Remarquez également que l'opérateur optimal $\Gamma$ peut être diagonalisé, c'est-à-dire qu'il existe $(\Psi_i)_{i\in\mathbb{N}^*}$ et $(\alpha_i)_{i\in\mathbb{N}^*}$ tels que $\Gamma = \sum_i \alpha_i \Psi_i$. On considère maintenant un paramètre semi-classique $\varepsilon = \hbar^2$, en faisant un rescaling de la densité $\rho$, on peut étudier la limite lorsque $\varepsilon \to 0$ à la fois de (22) et de (23). Les fonctionnelles ci-dessus convergent (voir [Lew18] pour la fonctionnelle de Lieb et [CFK13; CFK18; BD17] pour celle de Lévy-Lieb) vers le problème de transport optimal multi-marginales avec un coût de Coulomb

$$\inf_{\gamma\in\Pi(\rho)} \int_{\mathbb{R}^{3m}} \sum_{i<j} \frac{1}{|x_i - x_j|} \mathrm{d}\gamma \tag{11}$$

où maintenant les marginales sont toutes égales à la densité électronique donnée $\rho$. Le $\mathsf{MOT}^c$ avec un coût de Coulomb est donc une manière rigoureuse d'approcher le terme de répulsion électron-électron dans la fonctionnelle de Levy-Lieb/Lieb. Pour cette raison, (11) joue un rôle central dans la DFT computationnelle, et trouver des moyens de la calculer (ou de l'approcher) de manière efficace est crucial. Ce chapitre est en quelque sorte consacré à trouver des généralisations de MOT pour approcher (11) d'une manière plus facile à traiter (numériquement).

La première partie de ce chapitre 2.2 est consacrée à une généralisation du $\mathsf{MOT}^c$ qui approxime la répulsion électron-électron dans la fonctionnelle de Levy-Lieb en donnant des informations sur l'importance des fluctuations des corrélations dans le système. En particulier, nous permettons maintenant au nombre nombre de marginales $m$ de varier. Ce modèle est inspiré de la physique statistique, où il est généralement connu sous le nom de **grand-canonique** [Rue99]. Le problème du Transport Optimal Grand Canonique (GC-OT) peut être formulé sous la forme suivante :

$$\boxed{\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) := \inf\left\{\sum_{n=0}^{\infty}\int_{X^n} c_n \,\mathrm{d}\gamma_n \ :\ \sum_{n=0}^{\infty}\gamma_n(X^n) = 1, \quad \sum_{n=1}^{\infty} n\,\gamma_n(\cdot, X^{n-1}) = \rho\right\}} \tag{12}$$

où chaque $\gamma_n$ est une mesure symétrique sur $X^n$, $c_n$ est le coût symétrique pour le problème $n$-marginal (par exemple, le potentiel de Coulomb) et $\gamma_n(\cdot, X^{n-1})$ est la première marge de $\gamma_n$. Remarquez le facteur $n$ multipliant la marge dans la contrainte impliquant $\rho$, ce qui tient compte du fait qu'il y a $n$ marges égales de ce genre puisque toutes les $\gamma_n$ sont symétriques. La famille $\gamma = (\gamma_n)_{n\geq 0}$ forme une probabilité qui décrit le comportement de certains agents dont le nombre est inconnu ou peut varier. Dans cette interprétation, $\gamma_n(X^n)$ est la probabilité qu'il y ait $n$ agents et $\gamma_0 \in [0,1]$ est celle qu'il n'y ait aucun agent du tout. Dans le problème GC-OT (12), seule la quantité *moyenne* $\rho$ est fixée et des fluctuations du nombre d'agents sont autorisées. Résoudre le problème de minimisation $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ nécessite en particulier de déterminer la meilleure façon de répartir le nombre d'agents à travers les mesures $\gamma_n$, afin de reproduire la moyenne donnée $\rho$, en fonction des coûts correspondants $\boldsymbol{c} = (c_n)n \geq 0$. Il est à noter que, dans le cas où nous considérons le potentiel de Coulomb comme $c_n$, résoudre ce problème est censé fournir des informations sur les corrélations dans le système, en fonction de la densité $\rho$, qui pourraient ensuite être utilisées en DFT computationnelle. À remarquer que la contrainte de symétrie sur les mesures $\gamma_n$ ne nous empêche pas de traiter des systèmes avec différents types d'agents. Revenons à l'exemple des boulangeries/cafés : si nous voulons transporter une certaine quantité de pains au chocolat/chocolatines, alors $X$ sera un ensemble fini contenant les propriétés (emplacement précis, taille, etc.) à la fois des boulangeries et des cafés. La symétrie signifie simplement que les pains au chocolat/chocolatines sont tous identiques et que nous ne voulons pas distinguer lequel est envoyé où. La principale différence avec l'approche standard est que maintenant nous pouvons définir un problème de transport sans connaître exactement ni la quantité de pains au chocolat/chocolatines produits ni la demande. Nous discutons plusieurs propriétés mathématiques du problème GT-OT (12), dans le cadre de la théorie du transport optimal. **(1)** Nous formulons le problème et montrons l'existence d'un minimiseur $\gamma = (\gamma_n)_{n\geq 0}$, sous des hypothèses sur les coûts. **(2)** Nous déduisons certaines propriétés sur le *support* de $\gamma$, c'est-à-dire sur le nombre de $\gamma_n$ qui sont non nuls (nous discutons en particulier le cas du potentiel de Coulomb). **(3)** Nous étudions également le problème tronqué où toutes les $\gamma_n$ sont supposées s'annuler après un certain $N_{\max}$, et la convergence vers le problème réel lorsque $N_{\max} \to \infty$. Cela est utile pour effectuer des calculs numériques. **(4)** Nous nous concentrons sur la théorie de la dualité et l'existence du potentiel dual. **(5)** Nous étudions le cas de la dimension $d = 1$, confirmant ainsi certaines conjectures sur la forme du plan optimal formulées dans [MSG13]. **(6)** Enfin, nous étudions la régularisation entropique du problème GC-OT.

Dans la dernière partie du chapitre 2.3, nous nous concentrons sur la fonctionnelle de Lieb (23), qui peut être considérée en quelque sorte comme une régularisation quantique du MOT avec un coût de Coulomb : dans la DFT classique (voir par exemple [Eva79; JLM23]), l'énergie libre (en gros, l'équivalent de la fonctionnelle de Lieb pour des particules classiques) d'un système de $m$ particules avec une densité donnée $\rho$ est exactement définie comme un problème de transport multi-marginal entropique $\mathsf{MOT}_\varepsilon^c$, comme (4), avec $c$ étant le potentiel de Coulomb. Cependant, l'avantage principal est que la fonctionnelle de Lieb est en réalité linéaire en $\Gamma$, ce qui la rend similaire à un problème standard de MOT. À partir de ces observations, nous visons à utiliser certains résultats récents sur l'approximation par moments du MOT. Plus précisément, l'approche d'abord considérée dans [Alf+21] consiste à introduire une approximation, lorsque le nombre de moments $M$ tend vers l'infini, des problèmes de transport multi-marginaux où les contraintes de marginales sont remplacées par des contraintes de moment associées à $M$ "fonctions de moment" qui sont des fonctions à valeurs réelles définies sur $\mathbb{R}^3$.

La solution de ce problème de transport optimal avec contrainte des moments est toujours une mesure de probabilité définie sur $\mathbb{R}^{3m}$, mais elle a également une structure parcimonieuse au sens où elle peut être écrite sous la forme d'une mesure discrète concentrée sur un certain nombre de points appartenant à $\mathbb{R}^{3m}$. Ce nombre de points évolue au plus de manière linéaire avec le nombre de contraintes de moments. Dans notre cadre, cela se traduit par la démonstration que les solutions de l'approximation de (10) par contraintes de moments peuvent être écrites sous la forme $\Gamma = \sum_{k=1}^{K} \alpha_k |\Psi_k\rangle\langle\Psi_k|$, où $K \in \mathbb{N}^*$ évolue au plus de manière linéaire avec le nombre de contraintes de moments et $\Psi_k$ sont des fonctions d'onde. En particulier, nous sommes en mesure de prouver **(1)** l'existence de minimiseurs pour le problème approché ainsi que la structure parcimonieuse. **(2)** Nous montrons la convergence de l'approximation vers la fonctionnelle de Lieb exacte. De plus, nous prouvons également le taux de convergence de l'approximation de l'énergie de l'état fondamental associée vers la valeur exacte. **(3)** Enfin, nous présentons quelques résultats sur la formulation duale du problème approché, en remarquant en particulier que le problème dual peut être reformulé comme un problème de programmation semi-définie positive.

## Transport optimal entre dimensions différentes

Le dernier chapitre de ce manuscrit traite du cas où le nombre de propriétés des boulangeries et des cafés diffère, c'est-à-dire $d_1 \neq d_2$, ce qui conduit à une autre généralisation du transport optimal, bien étudiée dans [CMP17; MP20], connue sous le nom de **transport optimal entre dimensions différentes**. Nous nous concentrons notamment sur l'étude de certains problèmes variationnels impliquant ce type de terme de transport optimal :

$$\boxed{\inf_{\mu_1,\mu_2} \mathcal{J}(\mu_1, \mu_1) := \mathsf{OT}^c(\mu_1, \mu_2) + \mathcal{F}(\mu_2) + \mathcal{G}(\mu_1).} \tag{13}$$

On s'intéresse à la caractérisation des minimiseurs de $(\mu_1, \mu_2) \mapsto \mathcal{J}(\mu_1, \mu_2)$, ainsi que les minimiseurs des sous-problèmes obtenus lorsque soit $\mu_1$ soit $\mu_2$ est fixé : $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$ et $\mu_1 \mapsto \mathcal{J}(\mu_1, \mu_2)$. Des problèmes de ces formes, pour divers choix des fonctionnelles $\mathcal{F}$ et $\mathcal{G}$, se posent dans une grande variété d'applications, notamment : les flots de gradient sur l'espace de Wasserstein (où la minimisation $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$ représente une étape dans un flot de gradient discret en temps, JKO [JKO98]). Ou encore l'interpolation par déplacement (lorsque $\mathcal{F}$ est la distance de Wasserstein par rapport à une deuxième mesure de probabilité), les équilibres de Cournot-Nash en théorie des jeux, la planification urbaine, etc. Pour clarifier la raison de la prise en compte d'un terme OT de dimensions inégales, considérons le sous-problème où la distribution $\mu_1$, disons la distribution des boulangeries, est fixée et où une personne très riche, ayant le désir incontrôlable d'ouvrir de nombreux cafés le long de la ligne de métro la plus fréquentée de Paris, veut trouver la distribution des cafés qui minimise $\mu_2 \mapsto \mathsf{OT}^c(\mu_1, \mu_2) + \mathcal{F}(\mu_2)$ : le coût de transport des pains au chocolat/chocolatines plus une fonctionnelle qui, par exemple, évite une concentration trop élevée de cafés au même endroit. Maintenant, en modélisant le fait que $\mu_2$ est concentré sur une courbe (la ligne de métro), on assume que l'espace $X_2$ a une dimension plus basse que $X_1$, par exemple $d_2 = 1$.
La première section du chapitre 3.1 est entièrement consacrée à l'étude des minimiseurs de tels problèmes variationnels. Remarquez que le problème OT entre dimensions différentes est traitable (c'est-à-dire que nous pouvons construire une solution presque explicite) lorsqu'une certaines condition, dite de **nestedness** (emboîtement), concernant le coût $c$ et les marginales $\mu_1$ et $\mu_2$ est satisfaite. Malheureusement, dans le contexte actuel, seulement le coût et **une** des marginales (et **aucune** dans le cas de doubles minimisations) sont connus. Nous prouvons ici que, sous diverses hypothèses sur $c$, $\mu_1$ et $X_2$, $(c, \mu_1, \mu_2)$ est nested chaque fois que $\mu_2$ minimise $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$, pour différentes choix de la fonctionnelle $\mathcal{F}$ ; des résultats analogues sont également établis pour certaines formes spécifiques de $\mathcal{G}$ pour les minimisations dans l'espace de dimension supérieure, $\mu_1 \mapsto \mathcal{J}(\mu_1, \mu_2)$, et pour les doubles minimisations $(\mu_1, \mu_2) \mapsto \mathcal{J}(\mu_1, \mu_2)$. Nous montrons ensuite que cette garantie a priori de nestedness rend le problème de caractérisation ou d'identification des minimiseurs beaucoup plus traitable ; dans différents contextes, en fonction de la forme précise de $\mathcal{F}$, nous établissons que les solutions peuvent être caractérisées par des équations différentielles (locales), peuvent être calculées numériquement par un schéma itératif convergent, ou peuvent être obtenues sous une forme presque explicite.

La deuxième partie de ce chapitre 3.2 traite d'une classe de métriques sur l'espace des mesures de probabilité impliquant des mesures singulières. Plus précisément, nous développons la théorie d'une métrique, que nous appelons la métrique de Wasserstein avec base $\nu$ et que nous notons $\mathcal{W}_\nu$, sur l'ensemble des mesures de probabilité $\mathcal{P}(X)$ sur un domaine $X \subseteq \mathbb{R}^d$. Cette métrique est basée sur un léger affinement de la notion de *géodésiques généralisées*, voir [AGS04], par rapport à une mesure de base $\nu$ et est pertinente en particulier dans le cas où $\nu$ est singulière par rapport à la mesure de Lebesgue $d$-dimensionnelle ; elle est également liée au concept de transport optimal linéarisé [Wan+13]. La métrique $\mathcal{W}_\nu$ est définie en termes d'un problème variationnel impliquant le transport optimal vers $\nu$ ; nous la caractérisons également en termes d'intégrations de la métrique de Wasserstein classique (transport optimal avec le coût quadratique) entre les probabilités conditionnelles lorsque les mesures sont désintégrées par rapport au

transport optimal vers $\nu$, et à travers la limite de certains problèmes de transport optimal multi-marginales. En revenant au problème d'appariement entre les boulangeries et les cafés. L'idée du $\mathcal{W}_\nu$ consiste à apparier les deux populations en tenant compte des informations fournies par la mesure de base $\nu$. Cela est particulièrement clair lorsqu'on considère la caractérisation en termes d'intégrations de la métrique de Wasserstein classique entre les probabilités conditionnelles : considérez $\nu$ comme une mesure de probabilité unidimensionnelle telle que, par exemple, pour un quartier/arrondissement $y$ de Paris (elle est unidimensionnelle car elle pourrait simplement être un ensemble d'étiquettes), $\nu(y)$ vous indique à quel point il est important d'apparier les boulangeries et les cafés à proximité. Ensuite, l'appariement via la métrique $\mathcal{W}_\nu$ fonctionne comme suit : (1) identifiez les boulangeries et les cafés proches d'un certain $y$ (la désintégration par rapport au transport optimal vers $\nu$), (2) appariez-les en utilisant la distance de Wasserstein classique.
À mesure que nous faisons varier la mesure de base $\nu$, la métrique $\mathcal{W}_\nu$ interpole entre la métrique de Wasserstein quadratique habituelle (obtenue lorsque $\nu$ est une mesure de Dirac) et une métrique associée aux uniques géodésiques généralisées définies lorsque $\nu$ est suffisamment régulière (par exemple, absolument continue par rapport à la mesure de Lebesgue). Lorsque $\nu$ se concentre sur une sous-variété de dimension inférieure de $\mathbb{R}^d$, nous démontrons que le problème variationnel dans la définition de la métrique $\mathcal{W}_\nu$ a une solution unique et nous établissons la convexité géodésique de la classe habituelle de fonctionnelles. Nous introduisons également une classe de métriques définies par rapport à une mesure de base fixée $\mu$ de dimension supérieure, sur l'ensemble des mesures qui sont absolument continues par rapport à une deuxième mesure de base fixée $\sigma$. À l'aide de cette métrique, nous prouvons la convergence d'un schéma itératif pour résoudre un problème variationnel étudié dans la première section du chapitre.

Permettez-moi de conclure cette introduction en mentionnant quelques travaux qui ne sont pas résumés dans ce manuscrit : [Ben+18] (étude des jeux à champ moyen variationnels du second ordre et lien avec la minimisation d'une entropie relative), [DGN23] (analyse des minimiseurs de la fonctionnelle de Levy-Lieb dans le cas bosonique) et [NP22] (méthodes numériques pour résoudre des problèmes variationnels impliquant le transport optimal entre dimensions différentes).

# Introduction (in English)

Optimal Transportation is a very dynamic research field and there are so many examples and applications that it is actually difficult to choose one in order to introduce this fascinating subject; we refer the reader to some "old and new" textbooks such as [AGS04; Vil09; Vil03; San15; Gal16]. Here, we have decided to introduce the classical optimal transport problem by using an example given by Villani in [Vil03].

Imagine that we are in Paris and consider the bakeries, producing pains au chocolat/chocolatines (the reader is free to choose his favorite French pastry), which should be shipped each morning to the local cafés where customers will enjoy eating them. The production and the consumption of pains au chocolat/chocolatines are described by a distribution $\mu_1$ and $\mu_2$, respectively; we assume that the total amount of the production and the consumption is the same. Then, the Optimal Transportation problem consists in transporting the amount of pains au chocolat/chocolatines produced by the bakery $x_1 \in X_1$ ($X_1 \subset \mathbb{R}^{d_1}$ is the set of bakeries differentiated by $d_1$ properties e.g., precise location, size, etc) to a café $x_2 \in X_2$ ($X_2 \subset \mathbb{R}^{d_2}$ is the set of cafés differentiated by $d_2$ properties) such that the transport cost $c(x_1, x_2)$, for instance the distance between the bakery and the café, is minimized. In classical optimal transport we have that $d_1 = d_2$ , but in the following we will try to tackle the problem when the number of chracteristcs (a.k.a. the dimensions) differs. The problem introduced by Monge in [Mon81] conists in searching for a transport map $x_2 = T(x_1)$ which tells us the café $x_2$ where *all* the amount of pains au chocolat/chocolatines produced by $x_1$ is sent. However this problem is difficult to treat and a relaxation (known as the **Monge-Kantorovich** problem) is needed: we allow to split the "mass" at $x_1$ such that, as an example, a fraction is sent to $x_2$ and the other one to $x'_2$. So, in this second case we are looking for an optimal coupling $\gamma(x_1, x_2)$ which tells us how mass at $x_1$ is distributed to all $x_2 \in X_2$. The problem can be then formulated as follows: given two probability measures $\mu_1 \in \mathcal{P}(X_1)$ and $\mu_2 \in \mathcal{P}(X_2)$ and a continuous cost function $c : X_1 \times X_2 \to \mathbb{R}$ the *Monge-Kantorovich* problem reads as

$$\boxed{\mathsf{OT}^c(\mu_1, \mu_2) := \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{X_1 \times X_2} c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2),} \tag{14}$$

where $\Pi(\mu_1, \mu_2)$ denotes the set of probability measures on $X_1 \times X_2$ having $\mu_1$ and $\mu_2$ as marginals, that is, given the canonical projection $e_i : X_1 \times X_2 \to X_i$, we have $e_{i,\sharp}\gamma = \mu_i$ for $i = 1, 2$. In the case in which the "cheapest" coupling $\gamma(x_1, x_2)$ is to assign a unique café to each bakery $x_1$, then we say that the optimal coupling is *deterministic* and an optimal map exists. When cost $c(x_1, x_2)$ is the squared distance, the existence of such a deterministic coupling was proved by Yann Brenier in his seminal work [Bre91]; moreover in this case (14) is a squared distance over the set of probability measures.
Many questions arise at this point: what happens if the number of characteristics of bakeries and cafés is different (e.g., $d_1 \neq d_2$)? and if one of the distributions is unknown and we wish to

obtain it via minimizing the $\mathsf{OT}^c$ plus another functional? Moreover, we can also add a third category, e.g., the hotels, so that now we have to minimize the cost of sending pains au chocolat/chocolatines to both the cafés and the hotel in Paris. Can this problem still be interpreted as an OT problem? This manuscript (which partially summarizes my research after the Ph.D.) is, actually, devoted to answering these questions and a few other ones dealing with optimal transport models arising in Physics, especially in Quantum Physics.
More precisely, this manuscript is divided in three chapters, thematically, organized as follows:

- **Chapter 1** is devoted to (1) multi-marginal optimal transport and an application to risk measures, (2) convergence rates for entropic multi-marginal optimal transport and, finally, (3) an ODE characterization of (discrete) entropic multi-marginal optimal transport. My co-authors on these topics are H. Ennaji, Q. Mérigot, B. Pass and P. Pegon and the results of the following papers [Enn+22; NP23c; NP23a] are summarized;

- In **Chapter 2** we describe a generalization of optimal transport to (1) the case in which we let the number of marginals vary and (2) to the quantum framework deriving a sparse approximation of the Lieb Functional. My co-authors on these topics are S. Di Marino, M. Lewin and V. Ehrlacher and we focus on the following papers [DLN22; EN23]

- In **Chapter 3** we consider the optimal transport when the dimensions of the marginals (namely the number of characteristics in the example above) are different and some related variational problems arising in Game Theory (e.g., the Cournot-Nash equilibrium). Moreover, we propose a class of metrics on the set of probability measures, involving singular measures, which help us to prove convergence of a numerical method to solve the above mentioned variational problems. My co-author on these topics is B. Pass and we consider the results obtained in [NP20; NP23b].

At the end of each chapter we present some perspectives. We now introduce more precisely the content of the different chapters.

## Multi-Marginal Optimal Transport and Entropic Regularization

At the beginning of the introduction we presented the standard optimal transport in the case in which we have to match/compare two populations, namely the bakeries and the cafés. However, it is very natural to deal with many populations (e.g., bakeries, cafés, hotels, supermarkets, etc) so that a simple way to compare them is to define the following **multi-marginal optimal transport problem**: given $m$ distributions $\mu_i \in \mathcal{P}(X_i)$, with $X_i \subset \mathbb{R}^{d_i}$ (for simplicity we can take $d_i = d$ for all $i = 1, \dots, m$), and a continuous cost function $c : \boldsymbol{X} \to \mathbb{R}$ the optimization problem reads as

$$\boxed{\mathsf{MOT}^c(\mu_1, \dots, \mu_m) := \inf_{\gamma \in \Pi(\mu_1, \dots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \dots, x_m) \mathrm{d}\gamma(x_1, \dots, x_m),} \tag{15}$$

where $\Pi(\mu_1, \dots, \mu_m)$ denotes the set of probability measures on $\boldsymbol{X} := \prod_{i=1}^m X_i$ having $\mu_i$ as marginals. It turns out that this kind of problem arises in many areas of applications including economics [CE10], financial mathematics [BHP13; DS14a; DS14b; Enn+22], statistics [BK18; CCG16], image processing [Rab+11], tomography [Abr+17], machine learning [Haa+21; TJK22], fluid dynamics [Bre89] and quantum physics and chemistry, in the framework of density functional theory [BDG12; CFK13]. The structure of solutions to the multi-marginal optimal transport problem is a notoriously delicate issue, and is still not well understood, despite substantial efforts by many researchers [GŚ98; Car03; CN08; Hei02; Pas11b; Pas12; KP14b; KP15; CS16; PV21; CDD15; PV22; MP17]; see also the surveys [Pas15] and [DGN17].

Of particular interest is the class of one dimensional multi-marginal optimal transport problems arising in the context of **risk estimation**. In this framework, for safety reasons, one needs to have a pessimistic estimation of the risk. Such problems arise for evaluating the risk of an industrial facility close to a river and protected by a dyke (see [IL15]). In particular, one knows how to evaluate a risk (such as the level of water in the river, compared to the height of the dyke) depending on a few variables (e.g. river width, maximal annual flowrate, etc.) whose probability distribution is known. What is unknown however, is the **coupling** between these variables yielding the worst-case scenario. Roughly speaking we want to solve the following maximization problem

$$\boxed{\max_{\gamma\in\Pi(\mu_1,\dots,\mu_m)} \mathcal{R}_\alpha(b_\#\gamma),} \tag{16}$$

where $\mathcal{R}_\alpha(\mu) := \int_0^1 F_\mu^{-1}(t)\alpha(t)\mathrm{d}t$ is the spectral risk measure for a non-negative and non-decreasing function $\alpha : [0,1] \to \mathbb{R}_+$ with $\int_0^1 \alpha(t)\mathrm{d}t = 1$, $F_\mu^{-1}$ is the pseudo-inverse function of $\mu$ and $b$ is the output function (e.g., it represents the simulated height of a river at risk of flooding). In [Enn+22] we **(1)** show that for any risk measures problem (16) can be reformulated as a $\mathsf{MOT}^c$ problem with $m+1$ marginals; **(2)** prove that the solution can be explicitly characterized by using a careful refinement of the existing theory of multi-marginal transport over one dimensional ambient space; **(3)** study the case in which the underline variables are in higher dimension; **(4)** extend to the framework in which the output function is multivariate.

A natural way to solve numerically the multi-marginal problem presented above is by means of the well known entropic regularization which brings us to the second kind of generalization of optimal transport presented in this manuscript: **the entropic multi-marginal optimal transport problem**. Given as before $m$ probability measures $\mu_i$ and a cost function $c$ the minimization problem is defined as

$$\boxed{\mathsf{MOT}_\varepsilon^c(\mu_1,\cdots,\mu_m) := \inf_{\gamma\in\Pi(\mu_1,\dots,\mu_m)} \int_{\boldsymbol{X}} c\mathrm{d}\gamma + \varepsilon \mathrm{Ent}(\gamma| \otimes_{i=1}^m \mu_i),} \tag{17}$$

where $\mathrm{Ent}(\cdot|\otimes_{i=1}^m \mu_i)$ is the Boltzmann-Shannon relative entropy (or Kullback-Leibler divergence) w.r.t. the product measure $\otimes_{i=1}^m \mu_i$. Notice that when it comes to solve the regularized problem by using the alternating Kullback-Leibler projections on the $m$ marginals constraints, by the algebraic properties of the entropy such an algorithm corresponds to the celebrated Sinkhorn's algorithm [Sin67] which is nowadays widely used to numerically solve optimal transport and related variational problems (see for instance [PC19]). Now, since $\mathsf{MOT}_\varepsilon^c$ can be seen as perturbation of $\mathsf{MOT}^c$ it is natural to study the behaviour as $\varepsilon$ vanishes. This motivated us, see [NP23c], to investigate the rate of convergence of the entropic cost $\mathsf{MOT}_\varepsilon^c$ to $\mathsf{MOT}_0^c$ under some mild assumptions on the cost functions and the marginals. The results we obtained together with P. Pegon, and detailed in Section 1.4, can be summarized as follows:

**(1)** we establish two **upper bounds**, one valid for locally Lipschitz costs and a finer one valid for locally semi-concave costs. For the latter, we improve the upper bound by a 1/2 factor, obtaining the following inequality for some $C^* \in \mathbb{R}_+$

$$\boxed{\mathsf{MOT}^c_\varepsilon \leq \mathsf{MOT}^c_0 + \frac{1}{2}\left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)\varepsilon\log(1/\varepsilon) + C^*\varepsilon.} \tag{18}$$

**(2)** For the **lower bound**, given a $\kappa$ depending on a signature condition on the second mixed derivatives of the cost, we have

$$\boxed{\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 + \frac{\kappa}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon.} \tag{19}$$

for some $C_* \in \mathbb{R}_+$.
These also helped us to extend the results in [CPT23; EN22] to degenerate cost function in the two marginals case. Notice that when $\kappa = \left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)$ we obtain a first order development of the entropic cost.
The last section of this chapter 1.5 deals with a characterisation of the discrete entropic multi-marginal optimal transport problem for the case of pair-wise cost function, that is $c(x_1,\dots,x_m) = \sum_{i<j} w(x_i,x_j)$ with $w$ is symmetric, via an ordinary differential equation. Typical examples of pair-wise cost functions are the Gangbo-Święch cost where $w(x_i,x_j) = |x_i - x_j|^2$, which is related to the Wasserstein barycenter problem [AC11], and the Coulomb cost $w(x_i,x_j) = 1/|x_i - x_j|$, which appears in the framework of Density Functional theory [CFK13; BDG12]. Assume now that $X_i = X$, for all $i$, are finite sets and the marginals are sum of diracs, one can show that if the marginals are all equal $\mu_i = \rho = \sum_{x\in X}\rho_x\delta_x$ (this is the case of the Coulomb cost since they represent electrons which are indistinguishable) then the discrete problem dual to (17) takes the following simple form (for simplicity we keep the continuous case notation but integrals must be intended as sums)

$$\inf_{\varphi:X\to\mathbb{R}}\left\{\tilde{\Phi}(\varphi,\eta)\right\}, \tag{20}$$

where

$$\tilde{\Phi}(\varphi,\eta) := -(m-1)\int_X \varphi \mathrm{d}\rho + \varepsilon\int_X \log\left(\int_{X^{m-1}}\exp\left(\frac{\sum_{i=2}^m \varphi(x_i) - c_\eta(\boldsymbol{x})}{\varepsilon}\right)\mathrm{d}\otimes^{m-1}\rho\right)\mathrm{d}\rho,$$

$$c_\eta(x_1,\dots,x_m) := \eta\sum_{i=2}^m\sum_{j=i+1}^m w(x_i,x_j) + \sum_{i=2}^m w(x_1,x_i),$$

and $\eta > 0$ is a parameter which allows to interpolate between the original and a special cost function for which the problem's complexity scales linearly in $m$. Notice that when $\eta = 1$ we retrieve the pair-wise cost defined above and we have to miniminize just over one potential since we exploit the undistinguishability of the marginals. Now, differentiating the optimality condition of (20) with respect to $\eta$ we obtain the following Cauchy problem:

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\eta),\eta)]^{-1}\dfrac{\partial}{\partial\eta}\nabla_\varphi\tilde{\Phi}(\varphi(\eta),\eta), \\ \varphi(0) = \varphi_w, \end{cases}} \tag{21}$$

where, the initial value $\varphi(0)$ of $\varphi$ when $\eta = 0$ coincides with the optimal potential $\varphi_w$ for the two marginal optimal transport problem with cost $w$. After proving the well-posedness of ODE in (21) one can use it in order to characterize the original entropic multi-marginal problem obtained for $\eta = 1$. By exploiting then the regularity of the rhs, we can solve numerically (21) by using some high order methods showing that it is (sometimes) a preferable approach (in terms of number of iterations to converge and CPU times) to Sinkhorn's. We conclude the chapter by using this ODE approach also for the case in which the cost function associated to the time discretization of Brenier's principle for incompressible Euler equations [Bre89].

## Optimal Transport in mathematical physics

We study some other generalizations of optimal transport, and in particular multi-marginal transport, by focusing on the application of OT in Density Functional Theory (DFT). One of the key quantities in DFT is the so-called Levy-Lieb functional which is the least energy (kinetic plus electron-electron interaction) of a quantum system having the prescribed electron density $\rho$. Namely we have to solve the following minimization problem

$$\boxed{F_{LL}[\rho] := \inf_{\substack{\Psi \in \mathcal{H}^m_1 \\ \rho_\Psi = \rho}} \frac{1}{2}\int_{\mathbb{R}^{3m}} |\nabla \Psi|^2 + \int_{\mathbb{R}^{3m}} V|\Psi|^2,} \tag{22}$$

where $\mathcal{H}^m_1$ is the set of antysimmetric wavefunctions $\Psi$ for a system of $m$ electrons with finite kinetic energy, the function $V$ is the electron-electron Coulomb interaction potential, $\rho_\Psi$ is the electronic density associated to the wavefunction $\Psi$, namely the real-valued function defined over $\mathbb{R}^3$ as follows:

$$\forall x \in \mathbb{R}^3, \quad \rho_\Psi(x) := m \int_{(\mathbb{R}^3)^{m-1}} |\Psi(x, x_2, \dots, x_m)|^2 \, \mathrm{d}x_2 \dots \mathrm{d}x_m.$$

It happens that $F_{LL}$ is not convex so it is convenient to look at a convexification proposed by Lieb [Lie83] where the minimization is performed over the set of mixed states instead of the set of pure ones as in (22): roughly speaking the wavefunctions are now replaced by operators and the energy in (22) is expressed in terms of traces. More precisely, we consider the minimization problem over the set $\mathfrak{S}^+_1$ of trace-class self-adjoint non-negative operators having $\rho$ as marginal, that is

$$\boxed{F_L[\rho] := \inf_{\substack{\Gamma \in \mathfrak{S}^+_1 \\ \rho_\Gamma = \rho}} \mathrm{Tr}\,(H_m \Gamma),} \tag{23}$$

where $H_m := -\frac{1}{2}\Delta + V$. Notice also that the optimal $\Gamma$ can be diagonalized, that is there exists $(\Psi_i)_{i\in\mathbb{N}^*}$ and $(\alpha_i)_{i\in\mathbb{N}^*}$ such that $\Gamma = \sum_i \alpha_i \Psi_i$. Consider now an effective semi-classical parameter $\varepsilon = \hbar^2$, by scaling the density $\rho$, and look at the limit for $\varepsilon \to 0$ of both (22) and (23). The above functionals converge (see [Lew18] for the Lieb functional and [CFK13; CFK18; BD17] for the Lévy-Lieb one) to the multi-marginal optimal transport problem with Coulomb cost

$$\inf_{\gamma \in \Pi(\rho)} \int_{\mathbb{R}^{3m}} \sum_{i<j} \frac{1}{|x_i - x_j|} \mathrm{d}\gamma \tag{24}$$

where now the marginals are all equal to the given electron density $\rho$. The $\mathsf{MOT}^c$ with Coulomb cost is then a rigorous way to approximate the electron-electron repulsion term in Levy-Lieb/Lieb

functional. For this reason, (24) plays a central role in computational DFT and so finding ways to efficiently compute (or approximate) is crucial. This chapter is in some sense devoted to find generalizations of MOT to approximate (24) in a more (numerical) tractable way.
The first part of this chapter 2.2 is devoted to a generalization of $\mathsf{MOT}^c$ which approximates the electron-electron repulsion in Levy-Lieb functional by giving some information on the amount of fluctuations of correlations in the system. In particular we now allow the number of marginals $m$ to vary. This model takes its root in statistical physics, where it usually goes under the name **grand-canonical** [Rue99]. In short, the Grand-Canonical Optimal Transport (GC-OT) problem can be formulated in the Kantorovich form as follows

$$\boxed{\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) := \inf\left\{\sum_{n=0}^{\infty}\int_{X^n} c_n \,\mathrm{d}\gamma_n \ : \ \sum_{n=0}^{\infty}\gamma_n(X^n) = 1, \quad \sum_{n=1}^{\infty} n\,\gamma_n(\cdot, X^{n-1}) = \rho\right\}} \tag{25}$$

where each $\gamma_n$ is a symmetric measure on $X^n$, $c_n$ is the symmetric cost for the $n$-marginal problem (for instance the Coulomb potential) and $\gamma_n(\cdot, X^{n-1})$ is the first marginal of $\gamma_n$. Notice the factor $n$ multiplying the marginal in the constraint involving $\rho$, which accounts for the fact that there are $n$ equal such marginals since all the $\gamma_n$ are symmetric. The family $\gamma = (\gamma_n)_{n\geq 0}$ forms a probability which describes the behavior of some agents whose number is unknown or can vary. In this interpretation $\gamma_n(X^n)$ is the probability that there are $n$ agents and $\gamma_0 \in [0,1]$ is the one that there is no agent at all. In the GC-OT problem (25) only the *average* quantity $\rho$ is fixed and fluctuations of the number of agents are allowed. Solving the minimization problem $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ requires in particular to determine the best way to distribute the number of agents through the measures $\gamma_n$, in order to reproduce the given average $\rho$, depending on the corresponding costs $\boldsymbol{c} = (c_n)_{n\geq 0}$. It is worth noticing that, in the case in which we consider the Coulomb potential as $c_n$, solving this problem is believed to give some information on the amount of correlations in the system, depending on the density $\rho$, which could then be used in the approximation of the more precise quantum problem. Let us mention that the symmetric constraint on the measures $\gamma_n$ does not prevent us from treating systems with different kinds of agents. Going back to the bakeries/cafés example, if we want to transport a certain amount of pains au chocolat/chocolatines then $X$ will be a finite set containing the properties (precise location, size, etc) of both bakeries and cafés. The symmetry just means that the pains au chocolat/chocolatines are all the same and we do not want to distinguish which one is sent where. The main difference with the standard approach is that now we can define an OT problem without knowing exactly either the quantity of pains au chocolat/chocolatines produced or the demand. We discuss several mathematical properties of the GT-OT problem (25), within the framework of optimal transport theory. **(1)** We formulate the problem and show the existence of a minimizer $\gamma = (\gamma_n)_{n\geq 0}$, under appropriate assumptions on the costs. **(2)** We derive some properties on the *support* of $\gamma$, that is, on how many of the $\gamma_n$'s are non zero (we particularly discuss the Coulomb case). **(3)** We also study the truncated problem where all the $\gamma_n$ are assumed to vanish after some $N_{\max}$ and the convergence to the true problem when $N_{\max} \to \infty$. This is useful for numerical purposes. **(4)** We focus on duality theory and the existence of the dual potential. **(5)** We study the 1D case, confirming thereby some conjectures on the shape of the optimal plan made in [MSG13]. **(6)** Finally, we study the entropic regularization of the GC-OT problem.
In the last part of the chapter 2.3 we focus on the Lieb functional (23) which can be considered in some sense a quantum regularization of MOT with Coulomb cost: in classical DFT (see for instance [Eva79; JLM23]) the free energy (roughly speaking the counterpart of the Lieb

functional for classical particles) of a system of $m$ particles with given density $\rho$ is exactly defined as an entropic multi-marginal transport problem $\mathsf{MOT}^c_\varepsilon$, such as (17), with $c$ being the Coulomb potential. However, the main advantage is that the Lieb functional is actually linear in $\Gamma$ which, at the same time, makes it similar to a standard MOT problem. Starting from these observations we aim at using some recent results about the moment approximation of MOT. More precisely, the approach first considered in [Alf+21] consists in introducing an approximation, as the number of moments $M$ tends to infinity, of the exact multi-marginal transport problems where the marginal constraints are replaced by moment constraints associated to a $M$ "moment functions" which are real-valued functions defined on $\mathbb{R}^3$. The solution of this moment-constrained optimal transport problem is still a probability measure defined on $\mathbb{R}^{3m}$ but is also a sparse object in the sense that it can be written as a discrete measure charging a number of points belonging to $\mathbb{R}^{3m}$ which scales at most linearly with the number of moment constraints. Back to our setting, this translates into proving that the solutions of moment constrained approximations of (23) can be written under the form $\Gamma = \sum_{k=1}^{K} \alpha_k |\Psi_k\rangle\langle\Psi_k|$, where $K \in \mathbb{N}^*$ scales at most linearly with the number of moment constraints and $\Psi_k$ are some wave-functions. In particular we are able to prove **(1)** existence of minimizers for the approximate problem as well as the sparse structure. **(2)** We show convergence of the moment-constrained approximation towards the exact Lieb functional. Moreover, we also prove some rates of convergence of the associated approximation of the ground state energy to the exact one. **(3)** We finally present some results about the dual formulation of the moment-constrained problem noticing in particular that the dual problem can be re-written as a semi-definite positive programming problem.

## Unequal dimensional Optimal Transport

The final chapter of this manuscript deals with the case in which the number of properties of bakeries and cafés differs, that is $d_1 \neq d_2$ which yields to another generalization of optimal transport, well studied in [CMP17; MP20], know as **unequal dimensional optimal transport**. We especially focus on studying some variational problems involving this kind of optimal transport term, that is

$$\boxed{\inf_{\mu_1,\mu_2} \mathcal{J}(\mu_1,\mu_1) := \mathsf{OT}^c(\mu_1,\mu_2) + \mathcal{F}(\mu_2) + \mathcal{G}(\mu_1).} \tag{26}$$

We are interested in characterizing the minimizers of $(\mu_1,\mu_2) \mapsto \mathcal{J}(\mu_1,\mu_2)$, as well as the minimizers of the subproblems obtained when either $\mu_1$ or $\mu_2$ is fixed: $\mu_2 \mapsto \mathcal{J}(\mu_1,\mu_2)$ and $\mu_1 \mapsto \mathcal{J}(\mu_1,\mu_2)$. Problems of these general forms, for various choices of the functionals $\mathcal{F}$ and $\mathcal{G}$, arise in a wide variety of applications, including: gradient flows on Wasserstein space (where the minimization $\mu_2 \mapsto \mathcal{J}(\mu_1,\mu_2)$ represents one step in a discrete in time gradient flow, the celebrated JKO scheme [JKO98]), displacement interpolation (when $\mathcal{F}$ is the Wasserstein distance to a second probability measure), Cournot-Nash equilibria in game theory, city planning problems, etc. To clarify the reason for considering an unequal dimensional OT term, consider the subproblem where the population $\mu_1$, say the distribution of bakeries, is fixed and one, say a very rich man/woman with the uncontrollable desire of opening many cafés along the busiest metro line in Paris, wants to find the distribution of cafés which minimizes $\mu_2 \mapsto \mathsf{OT}^c(\mu_1,\mu_2) + \mathcal{F}(\mu_2)$: the cost of transporting the pains au chocolat/chocolatines plus a functional which, for example, avoids an excessive concentration of cafés at the same place. Now modeling the fact that $\mu_2$ is concentrated on a curve (the metro line) amounts to saying that the space $X_2$ is lower dimensional than $X_1$, for example $d_2 = 1$.

The first section of the chapter 3.1 is exactly devoted to studying the minimizers of such variational problems. Notice that when a *joint* condition, the **nestedness**, on the cost $c$ and marginals $\mu_1$ and $\mu_2$ is satisfied, the unequal dimensional OT problem is tractable (namely we can build an almost explicit solution). Unfortunately, in the present context, only the cost and **one** of the marginals (**neither** in the case of double minimizations) is prescribed. Here we prove that, under various conditions on $c$, $\mu_1$ and $X_2$, $(c, \mu_1, \mu_2)$ is nested whenever $\mu_2$ minimizes $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$, for a variety of different choices of the functional $\mathcal{F}$; analogous results for certain specific forms of $\mathcal{G}$ are also established for minimizations on the higher dimensional space, $\mu_1 \mapsto \mathcal{J}(\mu_1, \mu_2)$, and for double minimizations $(\mu_1, \mu_2) \mapsto \mathcal{J}(\mu_1, \mu_2)$. We go on to demonstrate that this a priori guarantee of nestedness makes the problem of characterizing or identifying the minimizers much more tractable; in different contexts, depending on the precise form of $\mathcal{F}$, we establish that solutions can be characterized by (local) differential equations, can be computed numerically by a convergent iterative scheme, or can be derived in almost closed form.
The second part of this chapter 3.2 deals with a class of metrics on the space of probability measures involving singular measures. More precisely, we develop the theory of a metric, which we call the $\nu$-based Wasserstein metric and denote by $\mathcal{W}_\nu$, on the set of probability measures $\mathcal{P}(X)$ on a domain $X \subseteq \mathbb{R}^d$. This metric is based on a slight refinement of the well-known notion of *generalized geodesics*, see [AGS04], with respect to a base measure $\nu$ and is relevant in particular for the case when $\nu$ is singular with respect to $d$-dimensional Lebesgue measure; it is also closely related to the concept of linearized optimal transport [Wan+13]. The $\nu$-based Wasserstein metric is defined in terms of an iterated variational problem involving optimal transport to $\nu$; we also characterize it in terms of integration of classical Wasserstein metric ($\mathsf{OT}^c$ with the quadratic cost) between the conditional probabilities when measures are disintegrated with respect to optimal transport to $\nu$, and through limits of certain multi-marginal optimal transport problems. Going back to the matching problem between bakeries and cafés, the idea underlying $\mathcal{W}_\nu$ consists in matching the two populations by taking into account the information given by the base measure $\nu$. This is particularly clear when considering the characterization in terms of integration of classical Wasserstein metric between the conditional probabilities: consider $\nu$ as a one dimensional probability measure such that, for example, for a neighborhood/arrondissement $y$ of Paris (it is one dimensional because it could be just a set of labels), $\nu(y)$ tells you how important is matching bakeries and cafés near to it. Then the matching via the $\mathcal{W}_\nu$ metrics works then as follows: (1) identify the bakeries and cafés close to some $y$ (the disintegration with respect to optimal transport to $\nu$), (2) match them using the classic Wasserstein distance.
As we vary the base measure $\nu$, the $\nu$-based Wasserstein metric interpolates between the usual quadratic Wasserstein metric (obtained when $\nu$ is a Dirac mass) and a metric associated with the uniquely defined generalized geodesics obtained when $\nu$ is sufficiently regular (eg, absolutely continuous with respect to Lebesgue). When $\nu$ concentrates on a lower dimensional submanifold of $\mathbb{R}^d$, we prove that the variational problem in the definition of the $\nu$-based Wasserstein metric has a unique solution. We also establish geodesic convexity of the usual class of functionals. We also introduce a class of metrics defined relative to a fixed higher dimensional base measure $\mu$ , on the set of measures which are absolutely continuous with respect to a second fixed based measure $\sigma$. By means of this metric we prove convergence of an iterative scheme to solve a variational problem studied in the first section of the chapter.

Let me end this introduction by mentioning some works which are not summarized in this manuscript: [Ben+18] (study of second order variational mean-field games and equivalence with the minimization of a relative entropy), [DGN23] (analysis of the minimizers of the Levy-Lieb functional in the bosonic case) and [NP22] (numerical methods in order to solve variational problems involving unequal dimensional optimal transport).

**List of papers:**


- J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. "Iterative Bregman Projections for Regularized Transportation Problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138. ISSN: 1064-8275. DOI: `10.1137/141000439`
- J.-D. Benamou, G. Carlier, and L. Nenna. "A Numerical Method to Solve Multi-Marginal Optimal Transport Problems with Coulomb Cost". In: *Splitting Methods in Communication, Imaging, Science, and Engineering.* Springer International Publishing, 2016, pp. 577–601
- S. Di Marino, A. Gerolin, and L. Nenna. "Optimal Transportation Theory with Repulsive Costs". In: *Topological Optimization and Optimal Transport.* Radon Series on Computational and Applied Mathematics. https://www.degruyter.com/view/books/9783110430417/978311043041 010/9783110430417-010.xml: De Gruyter, 2017
- M. Seidl, S. Di Marino, A. Gerolin, L. Nenna, K. J. H. Giesbertz, and P. Gori-Giorgi. "The Strictly-Correlated Electron Functional for Spherically Symmetric Systems Revisited". In: *Physical Review A: Atomic, Molecular, and Optical Physics* (2017)
- A. Blanchet, G. Carlier, and L. Nenna. "Computation of Cournot–Nash Equilibria by Entropic Regularization". In: *Vietnam Journal of Mathematics* (2017). ISSN: 2305-2228. DOI: `10.1007/s10013-017-0255-x`
- J.-D. Benamou, G. Carlier, and L. Nenna. "Generalized Incompressible Flows, Multi-Marginal Transport and Sinkhorn Algorithm". In: *Numerische Mathematik* 142.1 (2019), pp. 33–54. ISSN: 0945-3245. DOI: `10.1007/s00211-018-0995-x`
- J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna. "An Entropy Minimization Approach to Second-Order Variational Mean-Field Games". 2018
- L. Nenna and B. Pass. "Variational Problems Involving Unequal Dimensional Optimal Transport". In: *Journal de Mathématiques Pures et Appliquées* 139 (2020), pp. 83–108
- S. Di Marino, M. Lewin, and L. Nenna. *Grand-Canonical Optimal Transport.* 2022. URL: `http://arxiv.org/abs/2201.06859`. preprint
- H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass. "Robust Risk Management via Multi-Marginal Optimal Transport". In: *to appear on Journal of Optimization Theory and Applications* (2022). DOI: `10.48550/ARXIV.2211.07694`
- L. Nenna and B. Pass. "A Note on Cournot-Nash Equilibria and Optimal Transport between Unequal Dimensions". In: *Optimal Transport Statistics for Economics and Related Topics.* Vol. 483. Studies in Systems, Decision and Control. arXiv, 2022. DOI: `10.48550/arXiv.2209.14888`

- S. Di Marino, A. Gerolin, and L. Nenna. “Universal Diagonal Estimates for Minimizers of the Levy-Lieb Functional”. In: *Letters in Mathematical Physics* (2023). DOI: 10.48550/arXiv.2303.00496
- L. Nenna and B. Pass. “Transport Type Metrics on the Space of Probability Measures Involving Singular Base Measures”. In: *Applied Mathematics & Optimization* 87.2 (2023), p. 28
- L. Nenna and B. Pass. *An ODE Characterisation of Multi-Marginal Optimal Transport with Pairwise Cost Functions*. 2023. URL: http://arxiv.org/abs/2212.12492. preprint
- L. Nenna and P. Pegon. “Convergence Rate of Entropy-Regularized Multi-Marginal Optimal Transport Costs”. In: *to appear on Canadian Journal of Mathematics* arXiv:2307.03023 (2023). DOI: 10.48550/arXiv.2307.03023. preprint
- V. Ehrlacher and L. Nenna. *A Sparse Approximation of the Lieb Functional with Moment Constraints*. 2023. URL: http://arxiv.org/abs/2306.00806. preprint

**List of papers in preparation:**

- J.B. Casteras, L. Monsaingeon, L. Nenna, *Large Deviation Principle and Gamma-convergence for the Sticky-Schödinger problem.*
- V. Ehrlacher, L. Nenna, *Reduced-order modeling for parametrized optimal transport problems.*
- L. De Pascale, L. Nenna, *Variational multi-populations mean field games.*
- S. Di Marino, A. Gerolin, L. Nenna, M. Seidl , P. Gori-Giorgi, *The strictly-correlated electron functional for spherically symmetric systems revisited II: SGS CONJECTURE.*

# Chapter 1

# Multi-Marginal Optimal Transport and Entropic Regularization

This chapter summarizes the main results on Multi-Marginal Optimal Transport with a particular focus on the risk measure applications and the entropic counterpart.

All the results are contained in the following papers

Section 1.2 H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass. "Robust Risk Management via Multi-Marginal Optimal Transport". In: *to appear on Journal of Optimization Theory and Applications* (2022). DOI: 10.48550/ARXIV.2211.07694

Section 1.4 L. Nenna and P. Pegon. "Convergence Rate of Entropy-Regularized Multi-Marginal Optimal Transport Costs". In: *to appear on Canadian Journal of Mathematics* arXiv:2307.03023 (2023). DOI: 10.48550/arXiv.2307.03023. preprint

Section 1.5 L. Nenna and B. Pass. *An ODE Characterisation of Multi-Marginal Optimal Transport with Pairwise Cost Functions.* 2023. URL: http://arxiv.org/abs/2212.12492. preprint

## 1.1 Introduction to Multi-Marginal Optimal Transport

In this section we briefly introduce a generalization of standard Optimal Transport problem to the case in which many marginals are involved instead of two.

Given $m$ compactly supported probability measures $\mu_i$ on sub-manifolds $X_i$ of dimension $d_i$ in $\mathbb{R}^d$ for $i \in \{1, \ldots, m\}$ and a continuous cost function $c : X_1 \times X_2 \times \ldots \times X_m \to \mathbb{R}_+$, the multi-marginal optimal transport problem consists in solving the following optimization problem

$$\mathsf{MOT}^c(\mu_1, \ldots, \mu_m) := \inf_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \int_{\boldsymbol{X}} c(x_1, \ldots, x_m) \mathrm{d}\gamma \tag{MOT}$$

where $\boldsymbol{X} := X_1 \times X_2 \times \ldots \times X_m$ and $\Pi(\mu_1, \ldots, \mu_m)$ denotes the set of all probability measures $\gamma$ having $\mu_i$ as $i$-th marginal, i.e. $(e_i)_\sharp \gamma = \mu_i$ where $e_i : (x_1, \ldots, x_m) \mapsto x_i$, for every $i \in \{1, \ldots, m\}$. The formulation above is also known as the *Kantorovich problem* and it amounts to a linear minimization problem over a convex, weakly compact set; it is then not difficult to prove the existence of a solution by the direct method of calculus of variations. Much of the attention in the optimal transport community is rather focused on uniqueness and the structure of the minimizers. In particular one is mainly interested in determining if the solution is concentrated on the graph of a function $(T_2, \ldots, T_m)$ over the first marginal, where $(T_i)_\sharp \mu_1 = \mu_i$ for $i \in \{1, \ldots, m\}$, in which case this function induces a solution *à la Monge*, that is $\gamma = (\mathrm{id}, T_2, \ldots, T_m)_\sharp \mu_1$.

In the two marginals setting, the theory is fairly well understood and it is well-known that under mild conditions on the cost function (e.g. twist condition) and marginals (e.g. being absolutely continuous with respect to Lebesgue), the solution to (MOT) is unique and is concentrated on the graph of a function ; we refer the reader to [San15] to have a glimpse of it. The extension to the multi-marginal case is still not well understood, but it has attracted recently a lot of attention due to a diverse variety of applications.

In particular in his seminal works [Pas11b; Pas12] Pass established some conditions, more restrictive than in the two marginals case, to ensure the existence of a solution concentrated on a graph. In this chapter we rely, for the high dimensional case, on the following (local) result in [Pas12] giving an upper bound on the dimension of the support of the solution to (MOT). Let $P$ be the set of partitions of $\{1, \ldots, m\}$ into two non-empty disjoint subsets: $p = \{p_-, p_+\} \in P$ if $p_- \bigcup p_+ = \{1, \ldots, m\}$, $p_- \bigcap p_+ = \emptyset$ and $p_-, p_+ \neq \emptyset$. Then for each $p \in P$ we denote by $g_p$ the bilinear form on the tangent bundle $T\boldsymbol{X}$

$$g_p \coloneqq D^2_{p_- p_+} c + D^2_{p_+ p_-} c \quad \text{where} \quad D^2_{pq} c \coloneqq \sum_{i \in p, j \in q} D^2_{x_i x_j} c$$

for every $p, q \subseteq \{1, \ldots, m\}$, and $D^2_{x_i x_j} c \coloneqq \sum_{\alpha_i, \alpha_j} \frac{\partial^2 c}{x_i^{\alpha_i} x_k^{\alpha_k}} \mathrm{d}x_i^{\alpha_i} \otimes \mathrm{d}x_j^{\alpha_j}$, defined for every $i, j$ on the whole tangent bundle $T\mathbf{X}$. Define

$$G_c \coloneqq \left\{ \sum_{p \in P} t_p g_p \mid (t_p)_{p \in P} \in \Delta_P \right\} \tag{1.1}$$

to be the convex hull generated by the $g_p$, then it is easy to verify that each $g \in G_c$ is symmetric and therefore its signature, denoted by $(d^+(g), d^-(g), d^0(g))$, is well defined. Then, the following result from [Pas12] gives a control on the dimension of the support of the optimizer(s) in terms of these signatures.

**Theorem 1.1.1** (Part of [Pas12, Theorem 2.3])**.** *Let $\gamma$ a solution to* (MOT) *and suppose that the signature of some $g \in G_c$ at a point $\boldsymbol{x} \in \boldsymbol{X}$ is $(d^+, d^-, d^0)$, that is the number of positive, negative and zero eigenvalues. Then, there exists a neighbourhood $N_{\boldsymbol{x}}$ of $\boldsymbol{x}$ such that $N_{\boldsymbol{x}} \bigcap \operatorname{spt} \gamma$ is contained in a Lipschitz sub-manifold of $\boldsymbol{X}$ with dimension no greater than $\sum_{i=1}^m d_i - d^+$.*

**Remark 1.1.2.** *For the following it is important to notice that by standard linear algebra arguments we have for each $g \in G_c$ that $d^+(g) \leq \sum_{i=1}^m d_i - \max_i d_i$. This implies that the smallest bound on the dimension of* $\operatorname{spt} \gamma$ *which Theorem 1.1.1 can provide is $\max_i d_i$.*

**Remark 1.1.3** (Two marginals case)**.** *When $m = 2$, the only $g \in G_c$ coincides precisely with the pseudo-metric introduced by Kim and McCann in [KM10]. Assuming for simplicity that $d_1 = d_2 = d$, they noted that $g$ has signature $(d, d, 0)$ whenever $c$ is non-degenerate so Theorem 1.1.1 generalizes their result since it applies even when non-degeneracy fails providing new information in the two marginals case: the signature of $g$ is $(r, r, 2d - 2r)$ where $r$ is the rank of $D^2_{x_1 x_2} c$. Notice that this will help us in Section 1.4 to generalize the results established in [CPT23; EN22], concerning the rate of convergence of entropic optimal transport, to the case of a degenerate cost function.*

It is well known that under some mild assumptions the Kantorovich problem (MOT) is dual to the following

$$\sup\left\{\sum_{i=1}^{m}\int_{X_i}\varphi_i(x_i)\mathrm{d}\mu_i \mid \varphi_i\in\mathcal{C}_b(X_i),\ \sum_{i=1}^{m}\varphi_i(x_i)\leq c(x_1,\dots,x_m)\right\}. \tag{MD}$$

Besides, it admits solutions $(\varphi_i)_{1\leq i\leq m}$, called *Kantorovich potentials*, when $c$ is continuous and all the $X_i$'s are compact, and these solutions may be assumed *c-conjugate*, in the sense that for every $i\in\{1,\dots,m\}$

$$\forall x\in X_i,\quad \varphi_i(x)=\inf_{(x_j)_{j\neq i}\in\boldsymbol{X}_{-i}} c(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_m)-\sum_{1\leq j\leq m, j\neq i}\varphi_j(x_j). \tag{1.2}$$

**One dimensional case** The most tractable of Multi-Marginal Optimal Transport problems occurs when the underlying spaces $X_i$ are one dimensional, and the mixed second derivatives of $c$ interact in a certain way.

**Definition 1.1.4** (Compatibility). *Suppose each $X_i\subset\mathbb{R}$ is a bounded real interval. Assume that $c\in\mathcal{C}^2(X_1\times\dots\times X_m)$. We say that $s$ is strictly compatible if for each three distinct indices $i,j,k\in\{1,1,\dots,m\}$ and each $(x_1,\dots,x_m)\in X_1\times\ldots\times X_m$ we have*

$$\frac{\partial^2 c}{\partial x_i\partial x_j}\left[\frac{\partial^2 c}{\partial x_k\partial x_j}\right]\frac{\partial^2 c}{\partial x_k\partial x_i}(x_1,\dots,x_m)<0. \tag{1.3}$$

*We will say that $c$ is weakly compatible (or simply compatible) if for each $i\neq j$, we have either $\frac{\partial^2 c}{\partial x_i\partial x_j}\geq 0$ throughout $X_1\times\ldots\times X_m$ or $\frac{\partial^2 c}{\partial x_i\partial x_j}\leq 0$ throughout $X_1\times\ldots\times X_m$, and*

$$\frac{\partial^2 c}{\partial x_i\partial x_j}\left[\frac{\partial^2 c}{\partial x_k\partial x_j}\right]\frac{\partial^2 c}{\partial x_k\partial x_i}(x_1,\dots,x_m)\leq 0, \tag{1.4}$$

*for each distinct $i,j,k\in\{1,\dots,m\}$ and each $(x_1,\dots,x_m)\in X_1\times\ldots\times X_m$.*

Note that if $c$ is strictly compatible, (1.3) implies that each $\frac{\partial^2 c}{\partial x_i\partial x_j}\neq 0$ throughout the connected domain $X_1\times\ldots\times X_m$; continuity then yields that each $\frac{\partial^2 c}{\partial x_i\partial x_j}\neq 0$ is either always positive or always negative. As (1.3) clearly implies (1.4), then strict compatibility implies compatibility.

The fact that the mixed partials $\frac{\partial^2 c}{\partial x_i\partial x_j}$ do not change signs under the compatibility condition allows us to partition the set $\{1,2,\dots,m\}=P_+\cup P_-$ of indices into disjoint subsets $P_+$ and $P_-$ such that $1\in P_-$ and for each $i\neq j$, $\frac{\partial^2 c}{\partial x_i\partial x_j}\leq 0$ throughout $X_1\times\dots\times X_m$ if either both $i$ and $j$ are in $P_-$ or if both are in $P_+$, and $\frac{\partial^2 c}{\partial x_i\partial x_j}\geq 0$ throughout $X_1\times\ldots\times X_m$ otherwise. The inequalities are strict if strict compatibility holds .

**Definition 1.1.5** (Pseudo-inverse). *Given a probability measure $\mu$ on the real line $\mathbb{R}$, the generalized or pseudo-inverse of the cumulative function $F_\mu$ is defined as*

$$F_\mu^{-1}(t)=\inf\{x\in\mathbb{R}\mid F_\mu(x)\geq t\}.$$

**Definition 1.1.6.** *For a compatible $c$, we define the $c$ -* comonotone coupling *by:*

$$\gamma = (G_1, \ldots, G_m)_{\#}\mathcal{L}_{[0,1]}, \tag{1.5}$$

*where $G_1 = F^{-1}_{\mu_1}$ and for each $i = 2, \ldots, m$*

$$G_i(t) = \begin{cases} F^{-1}_{\mu_i}(t) & \text{if } i \in P_-, \\ F^{-1}_{\mu_i}(1-t) & \text{if } i \in P_+. \end{cases} \tag{1.6}$$

The following result can be found in [Car03] for strictly submodular functions (all the second mixed derivatives are striclty negative) as well as in [Pas11a] for compatible functions (where its proof appears together with the formulation of the compatibility conditions and the observation that it is equivalent to strict submodularity after changing coordinates).

**Theorem 1.1.7** ([Car03; Pas11a])**.** *Suppose that $c$ is strictly compatible. Then the $c$-comonotone coupling* (1.5) *is the unique optimizer in* (MOT).

Though the following result, we have obtained in [Enn+22], does not seem to be available in the literature, it is a straightforward consequence of Theorem 1.1.7 and the well known stability of optimal transport with respect to perturbations of the cost function. It asserts that the optimality of the measure constructed in Theorem 1.1.7 still holds if the strong compatibility assumption is relaxed to weak compatibility, although the uniqueness assertion may not.

**Proposition 1.1.8** ([Enn+22])**.** *Suppose that $c$ is weakly compatible. Then the $c$-comonotone coupling* (1.5) *is optimal in* (MOT).

## 1.2 From Risk measures to multi-marginal optimal transport

**Remark 1.2.1** (Notations)**.** *Throughout all this section we adapt the language and the notations to the case in which optimal transport is applied to financial mathematics; that is instead of defining problem* (MOT) *as a minimization problem, we consider it as a maximization one and the cost function is now denoted as a surplus function $b$, namely*

$$\sup_{\gamma \in \Pi(\mu_1, \ldots, \mu_m)} \int_{\boldsymbol{X}} b(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m). \tag{1.7}$$

*Notice that the two formulations are clearly equivalent by taking $b = -c$. All the theorems stated in the previous section still hold and will be used in the following.*

In a variety of problems in operations research, a variable of interest $b = b(x_1, x_2, \ldots, x_m)$ depends on several underlying random variables, whose individual distributions are known (or can be estimated) but whose joint distribution is not. A natural example arises when the variables represent parameters in a physical system whose individual distributions can be estimated empirically or through modeling (or a combination of both) but whose dependence structure cannot; an example of this flavour, originating in [IL15], in which the output $b$ represents the simulated height of a river at risk of flooding, and the underlying variables include various design parameters and climate dependent factors can be found below. Methods used in risk management to evaluate the resulting aggregate level of risk naturally depend on the distribution of the output variable $b$, and therefore, in turn, on the joint distribution of the $x_i$. A natural problem is therefore to determine bounds, or worst-case scenarios, on these risk measures; that is, to maximize a given risk measure over all possible joint distributions of the $x_i$ with known marginal distributions.

Problems such as this have received extensive attention within the financial risk management community; see, for example, the monographs [Rüs13] and [MFE15] and the references within them. In the simplest of these problems, the underlying variables $x_i$ are typically real valued and the output variable $b$ is often assumed to have a particular structure (in many cases, it is a weighted sum of the $x_i$, reflecting the value of a portfolio built out of underlying assets with values $x_i$, or a function of this weighted sum). In these cases, explicit solutions for the maximizing couplings of the $x_i$ can sometimes be obtained (see, for example, [EWW15; EP06; PR13; Puc13; WW11]).

For more general output functions $b$, and possibly multi-dimensional underlying variables $x_i$, much less is known about the dependence structure of the maximizing joint distributions. Motivated by applications arising in the analysis of counter-party credit risk, a recent paper of Ghossoub-Hall-Saunders [GHS20] studies this more general setting systematically, allowing the underlying variables to take values in very general spaces and the output $b$ to take a very general form. They observe that the resulting problem is a generalization of the optimal transport problem; in fact, in the simplest case, when the spectral function is identically equal to one and the number $m$ of underlying variables is 2, the problem is exactly a classical optimal transport problem with surplus function $b = -c$. Although the analysis in [GHS20] focused on the $m = 2$ case, they note that their results can be extended to the $m \geq 3$ setting, in which case maximizing spectral risks becomes a generalization of the multi-marginal optimal transport problem.

In a joint work with H. Ennaji, Q. Mérigot and B. Pass we obtain the following results which will be more detailed below:

- We show that for *any* spectral risk measure, the maximization can in fact be formulated as a traditional multi-marginal optimal transport problem with $m+1$ marginals: the given marginals distributions of the $x_i$ as well as another distribution arising from the particular form of the spectral function (see Theorem 1.2.7 below). Although this formulation slightly increases the underlying dimension, it makes the maximization problem linear and much more tractable – indeed, the results and techniques, both theoretical and computational, in the substantial literature on multi-marginal optimal transport become applicable. Moreover, we derive an interesting equivalence between partial optimal transport and multi-marginal optimal transport which, up to our knowledge, has not been observed before in the literature, see Remark 1.2.4 and 1.2.8.

- We derive an explicit characterization of solutions for very general spectral risk measures and a substantial class of output functions $b$, through a careful refinement of the existing theory of multi-marginal optimal transport on one dimensional ambient spaces (see Theorem 1.2.10 and the examples following it). For $b$ falling outside this class, explicit solutions are likely generally unattainable; however, our formulation of the problem can potentially facilitate the use of a very broad range of computational methods for optimal transport to approximate solutions numerically (see Section 8.1 in [Ben21] and the references therein).

- For underlying variables in higher dimensional spaces, explicit solutions are generally not possible; however, there are known conditions on the cost/surplus function under which the solution to the multi-marginal optimal transport problem concentrates on a graph over the first variable, see [KP14a]. These conditions were not developed with our application in mind; in particular, most previous work has focused on the case when each of the variables lives in set of the same dimension, whereas in our setting, the extra variable $x_0$ arising from the spectral function is one dimensional while the underlying variables $x_1, ...., x_m$ are higher dimensional. When $m = 2$, we provide a condition on $b$ under which the solution is of Monge type (Theorem 1.2.13).

- We also extend these ideas to the setting where the output function $b$ is multi-variate valued, using the maximal correlation risk measures developed in [Rüs06; EGH12]. We are again able to identify conditions under which the solution concentrates on a graph (Propositions 1.2.15 and 1.2.16); once more, these results do not seem readily attainable without using our optimal transport reformulation.

### 1.2.1 Equivalence between maximizing spectral risk measures and multi-marginal optimal transport

Let the real line describe a certain level of risk (e.g., the level of radiation in the nuclear power plant) and $\mu$ be a probability measure on $\mathbb{R}$ which can be interpreted as the distribution of risk. We will consider the following form of quantifier of the risk associated with $\mu$.

**Definition 1.2.2** (spectral risk measure). *A functional* $\mathcal{R} : \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ *is called a* spectral risk measure *if it takes the form* $\mathcal{R} = \mathcal{R}_\alpha$*, where*

$$\mathcal{R}_\alpha(\mu) := \int_0^1 F_\mu^{-1}(t)\alpha(t)\mathrm{d}t. \tag{1.8}$$

*for a non-negative and nondecreasing function* $\alpha : [0, 1] \to \mathbb{R}_+$ *with* $\int_0^1 \alpha(t)\mathrm{d}t = 1$.

**Remark 1.2.3.** *Note that by taking* $\alpha$ *to be defined and real valued on* $[0, 1]$*, we are tacitly assuming that it is bounded, since* $\alpha(0) \le \alpha(t) \le \alpha(1)$ *by monotonicity; the same assumption was made in [GHS20]. This is mostly for technical convenience; we expect that most of our results can be extended to the more general case* $\alpha : [0, 1) \to \mathbb{R}_+$*, where, as above,* $\alpha$ *is monotone and* $\int_0^1 \alpha(t)\mathrm{d}t = 1$*, but possibly* $\lim_{t\to 1} \alpha(t) = \infty$*, under suitable hypothesis (for instance, decay conditions on* $\alpha_\# \mathcal{L}_{[0,1]}$*).*

For a given $\alpha$, we will often refer to the spectral risk measure $\mathcal{R}_\alpha$ as the *$\alpha$-risk*, and $\alpha$ as the the *spectral function*.

It is well know in literature (see, for example, [Rüs06] and [EGH12], in which this fact is the basis for a multi-variate extension) the following variational characterization of spectral risk measures, that is given a probability measure $\mu$ with $\int_\mathbb{R} x\mathrm{d}\mu(x) > -\infty$,

$$\mathcal{R}_\alpha(\mu) = \max_{\pi \in \Pi(\alpha_\sharp \mathcal{L}_{[0,1]}, \mu)} \int_{\mathbb{R}\times\mathbb{R}} xy\mathrm{d}\pi(x, y), \tag{1.9}$$

where we have denoted $\Pi(\alpha_\sharp \mathcal{L}_{[0,1]}, \mu)$ the space of probability measures on $\mathbb{R}^2$ with marginals $\alpha_\sharp \mathcal{L}_{[0,1]}$ and $\mu$. Notice that thanks to this characterization the spectral risk measure is indeed a standard optimal transport problem with cost function $c(x, y) = -xy$. One can easily prove this by exploiting the fact that in one dimension the optimal transport plan is obtained by using the pseudo inverse of the cumulative function, as (1.5). Return now to the case in which there are $m$ parameters which enter into the estimation of risk through the output function $b$, and we wish to evaluate the worst case scenario for the $\alpha$-risk $\mathcal{R}_\alpha(b_\# \gamma)$ of $b_\# \gamma$ among couplings $\gamma \in \Pi(\mu_1, \mu_2, \ldots, \mu_m)$ of the marginal distributions $\mu_i$. That is, we want to maximize:

$$\max_{\gamma \in \Pi(\mu_1,\mu_2,\ldots,\mu_m)} \mathcal{R}_\alpha(b_\# \gamma). \tag{1.10}$$

**Remark 1.2.4** (Maximizing Expected Shortfall is optimal partial transport)**.** *A special case of particular importance in risk management applications occurs when* $\alpha = \alpha_{m_0} := \frac{1}{m_0}\mathbf{1}_{[1-m_0,1]}$*, in which case* (1.8) *is also know as the* Expected Shortfall*, or* Conditional Value at Risk*. In this setting, the maximization problem* (1.10) *is actually equivalent to a well known variant of the optimal transport problem; indeed, it can be reformulated into*

$$\boxed{\max_{\gamma \in \Pi_{m_0}(\mu_1,\ldots,\mu_m)} \frac{1}{m_0} \int b(x_1, \ldots, x_m) \mathrm{d}\gamma(x_1, \ldots, x_m)}$$

*where* $\Pi_{m_0}(\mu_1, \ldots, \mu_m)$ *denotes the set of non-negative measures* $\gamma$ *on* $X_1 \times \cdots \times X_m$ *with total mass* $m_0$*, such that its* $i-$*marginal* $\gamma_i$ *is dominated by* $\mu_i$ *for each* $i$*; that is* $\int_{X_i} \varphi \mathrm{d}\gamma_i \leq \int_{X_i} \varphi \mathrm{d}\mu_i$ *for all non-negative test functions* $\varphi \in \mathcal{C}^0(X_i)$*. This is known as the* optimal partial transport problem *when* $m = 2$ *[CM10; Fig10], and the* multi-marginal optimal partial transport problem *[KP14c] when* $m > 2$*. As we will show below, the problem is in fact equivalent to an ordinary multi-marginal optimal transport problem with an additional marginal; see Theorem 1.2.7 and Remark 1.2.8 below.*

Our first main contribution is to show that the spectral risk maximization problem (1.10) is equivalent to the multi-marginal optimal transport problem (1.7) with $X_0 = [\alpha(0), \alpha(1)] \subseteq \mathbb{R}$, $\mu_0 = \alpha_\sharp \mathcal{L}_{[0,1]}$, the other $X_i$ and $\mu_i$ representing the domains and distributions of the underlying variables, respectively, and

$$s(x_0, x_1, \ldots, x_m) = x_0 b(x_1, \ldots, x_m). \tag{1.11}$$

The intuition behind this is simple: if the variables $(x_0, ..., x_m)$ are optimally coupled for $s$, $x_0$ and the variable $b$ must be coupled in an increasing way, this is captured precisely by Lemma 1.2.5. This monotonicity leads exactly to the spectral risk measure evaluated on the distribution of $b$.

**Lemma 1.2.5** ([Enn+22])**.** *Suppose that* $\pi \in \Pi(\mu_0, \mu_1, \ldots, \mu_m)$ *and let* $\gamma$ *be the* $(1, \ldots, m)$ *-marginal of* $\pi$*, that is the projection of* $\pi$ *on* $X_1 \times \ldots \times X_m$*. Then, for* $s$ *given by* (1.11)*,*

$$\int_{X_0 \times \boldsymbol{X}} s(x_0, x_1, \ldots, x_m) \mathrm{d}\pi(x_0, x_1, \ldots, x_m) \leq \mathcal{R}_\alpha(b_\# \gamma)$$

*Furthermore, we have equality if and only if the support of*

$$\tau_\pi = \Big((x_0, x_1, x_2, \ldots, x_d) \mapsto (x_0, b(x_1, x_2, \ldots, x_d))\Big)_\# \pi \in \mathcal{P}(\mathbb{R}^2)$$

*is monotone increasing.*

**Lemma 1.2.6** ([Enn+22])**.** *Given any measure* $\gamma \in \Pi(\mu_1, \ldots, \mu_m)$*, there exists a* $\pi \in \Pi(\mu_0, \mu_1, \ldots, \mu_m)$ *whose* $(1, \ldots, m)$*-marginal is* $\gamma$*, such that* $\Big((x_0, x_1, \ldots, x_m) \mapsto (x_0, b(x_1, x_2, \ldots, x_m)\Big)_\# \pi$ *has monotone increasing support.*

The proof of Lemma 1.2.5 relies on the variational characterization of the spectral risk (1.9) where as for Lemma 1.2.6 one uses comonotonic transport plans and measure disintegration. We are now ready to establish the equivalence between the maximal $\alpha-$risk problem (1.10) and the multi-marginal optimal transport problem (MOT).

**Theorem 1.2.7** ([Enn+22])**.** *A probability measure $\pi$ in $\Pi(\mu_0, \mu_1, \dots, \mu_m)$ is optimal for* (1.7) *with cost function* (1.11) *if and only if its $(1, \dots, d)$-marginal is optimal in* (1.10)*, and*

$$\tau_\pi = \Big((x_0, x_1, x_2, \dots, x_d) \mapsto (x_0, b(x_1, x_2, \dots, x_d))\Big)_{\#}\pi$$

*has monotone increasing support.*

**Remark 1.2.8.** *Returning to the Expected Shortfall Case, $\alpha = \alpha_{m_0} = \frac{1}{m_0}\mathbf{1}_{[1-m_0,1]}$, in view of Remark 1.2.4, Theorem 1.2.7 shows that the optimal partial transport problem for the surplus $b(x_1, \dots, x_m)$ can be transformed into a multi-marginal transport problem for the surplus $s = x_0 b$ and additional marginal $\mu_0 = \alpha_\sharp \mathcal{L}_{[0,1]}$. To the best of our knowledge, this equivalence between these two well studied mathematical problems has not been observed before in the optimal transport literature. In addition, this perspective, together with results in [CM10; Fig10; KP14c] allows us to immediately identify conditions under which the active (that is, the part that couples to $\alpha > 0$) part of the optimal $\gamma$ in* (1.10) *is uniquely determined and concentrates on a graph over $x_1$. Furthermore, algorithms to compute the solution are readily available [Ben+15; IN18].*

### 1.2.2 Solutions for one-dimensional assets and compatible payouts

We now turn our attention to the structure of maximizers in (1.10) when the underlying variables are one dimensional. Assume that each $x_i \in \mathbb{R}$ and each $\mu_i$ is supported on an interval, $X_i = [\underline{x}_i, \overline{x}_i]$, for all $i = 1, \dots, m$. We note that in our setting, the first marginal, $\mu_0$ is supported always on the interval $[\underline{x}_0, \overline{x}_0] := [\alpha(0), \alpha(1)]$ with $\alpha(0) \geq 0$.

**Lemma 1.2.9** ([Enn+22])**.** *Suppose that $b$ is compatible and monotone increasing in each $x_i \in P_+$ and monotone decreasing for each $x_i \in P_-$. Then the $s$-comonotone coupling* (1.5) *is optimal for* (1.7) *with surplus function given by* (1.11)*. The maximal value is*

$$\int_0^1 \alpha(t) b(G_1(t), G_2(t), \dots, G_m(t))\mathrm{d}t. \tag{1.12}$$

*Furthermore, if in addition the monotonicity of $b$ with respect to each argument is strict, the solution is unique on the support of $\alpha$; that is, any other solution $\overline{\pi}$ coincides with $\pi = (G_\alpha, G_1, \dots, G_m)_{\#}\mathcal{L}_{[0,1]}$ on $(0, \alpha(1)) \times [\underline{x}_1, \overline{x}_1] \times \dots \times [\underline{x}_m, \overline{x}_m]$.*

Clearly full uniqueness of the optimal $\pi$ cannot hold if $\alpha = 0$ on a set of positive measure, or, equivalently, $\mu_0(\{0\}) > 0$; in this case, we can rearrange the part of the $(1, ..., d)-$ marginal $\gamma$ of $\pi$ that couples to $\alpha = 0$ in any way without affecting the value of $\int s \mathrm{d}\pi$. The preceding Lemma identifies conditions under which this is the only source of non-uniqueness, so that the optimal $\pi$ is uniquely determined on $(0, \alpha(1)) \times [\underline{x}_1, \overline{x}_1] \times \dots \times [\underline{x}_m, \overline{x}_m]$.

Note also that we are able to obtain uniqueness off the set where $\alpha = 0$ despite the fact that the surplus function may not be strongly compatible. This is done by exploiting an idea developed in [PV22] (see in particular Proposition 4.1 there); namely that one only needs strongly compatible interactions between certain key pairs of variables (rather than all of them).

Thanks to Theorem 1.2.7, the preceding result then easily yields the following characterization of solutions to (1.10).

**Theorem 1.2.10** ([Enn+22])**.** *Assume that $b$ is compatible, monotone increasing in each $x_i \in P_+$ and monotone decreasing in each $x_i \in P_-$. Then $(G_1, \ldots, G_m)_\# \mathcal{L}_{[0,1]}$ maximizes* (1.10) *and the maximal value is given by* (1.12)*. If in addition the monotonicity is strict, then letting $a = \mu_0\{0\}$ any other maximizer $\gamma$ must couple the regions defined by $G_i([0, a])$, and must coincide with the $b$-comonotone coupling $(G_1, \ldots, G_m)_\# \mathcal{L}_{[0,1]}$ on $G_1([a, 1]) \times G_2([a, 1]) \times \ldots \times G_m([a, 1])$.*

**Example 1.2.11.** *If $b(x_1, \ldots, x_m) = -\sum_{1 \le i < j \le m}(x_i - x_j)^2$ and if $\alpha \equiv 1$, problem* (1.10) *is equivalent to the computation of Wasserstein barycenters [AC11], while with $\alpha = \alpha_{m_0}$ we get the partial Wasserstein barycenter problem [KP14c]. In both cases, since $b$ is compatible the answer can be calculated explicitly by Theorem 1.2.10.*

**Example 1.2.12** (Sensitivity analysis and maximal river flow)**.** *We briefly describe a recurring example from [IL15], used throughout that paper to illustrate issues in sensitivity analysis. In that setting, one wants to understand the influence of the dependence structure (among other factors) between several contributing inputs on an output behavior. We consider a simple model which involves the height of river at risk of flooding and compares it to the height of a dyke which protects industrial facilities. The maximal annual overflow $S$ of a river is modelled by*

$$S = Z_\nu + \left( \frac{Q}{BK_s \sqrt{\frac{Z_m - Z_\nu}{L}}} \right)^{0.6} - H_d - C_b,$$

*where* $\left( \frac{Q}{BK_s \sqrt{\frac{Z_m - Z_\nu}{L}}} \right)^{0.6}$ *is the maximal annual height of the river and the variables $Q$, $K_s$, $Z_\nu$, $Z_m$, $H_d$, $C_b$, $L$, $B$ are physical quantities whose values are modelled as random variables due to their variation in time and space, measurement inaccuracies, or uncertainty of their true values. Their individual distributions are modelled in [IL15] (see Table 1 on p.4), and are all absolutely continuous with respect to Lebesgue measure. The $\alpha$-risk* (1.8)*, where the $x_i$ are the variables $Q$, $K_s$, $Z_\nu$, $Z_m$, $H_d$, $C_b$, $L$, $B$ and $b = S$ is the overflow, quantifies the risk of overflow. In [IL15], the variables were assumed to be independent, although other dependence structures are certainly possible;* (1.10) *asks what is the maximal risk over all possible dependence structures. Notice then that the surplus function $b$ is compatible and satisfies the strict monotonicity with respect to each variable required in Theorem 1.2.10. The* $(1, ..., m)$ *marginal $\gamma$ of the $s$-comonotone solution $\pi$ defined in* (1.5) *is optimal, and the unique optimizer on the support of $\alpha$, $\{\alpha > 0\}$. In this case, the $G_i$ corresponding to $Z_\nu, Q$ and $L$ are monotone increasing, while the $G_i$ corresponding to the other variables are decreasing.They can be retrieved explicitly by computing the quantile functions associated to each marginals as described above.*

### 1.2.3 Higher dimensional assets

As explained at the beginning of the chapter, there is an analogous theory of multi-marginal optimal transport when the underlying variables lie in more general spaces. Although in these cases it is generally not possible to derive explicit solutions as we did above, it is possible to prove analogous structural properties of optimal couplings, namely, that solutions are of Monge type (that is, concentrated on graphs over $x_1$), for certain surplus functions $s$ (see [GŚ98] for an early result in this direction, for a particular $s$, and [KP14a] and [Pas11b] for general, sufficient conditions on $s$).

We present one such result below, when each $X_i \subseteq \mathbb{R}^d$, to illustrate how the theory can be adapted to the present setting. For simplicity, we restrict our attention to the $m = 2$ case.

The conditions we impose on $b$ are related to the local differential condition found in [Pas11b]; however, the proof in [Pas11b] was restricted to the case when each marginal was supported on a space of the same dimension, whereas here our first marginal $\mu_0$, corresponding to $x_0 = \alpha$, has one dimensional support while the other marginals are supported in $\mathbb{R}^d$. Our proof is inspired by the argument in [Pas11b], but modified to fit the present setting here. Similar results can be established for larger $m$, using similar arguments. However, the conditions imposed on $b$ become more stringent for larger $m$, and more complicated to state. In what follows, we will assume that $b$ is $\mathcal{C}^2$; $D_{x_i}b(x_1,x_2) = (\frac{\partial b}{\partial x_i^1}, \frac{\partial b}{\partial x_i^2}, ..., \frac{\partial b}{\partial x_i^d})$ represents the gradient of the function $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ with respect to the variable $x_i \in \mathbb{R}^d$, for $i = 1, 2$. Similarly, $D^2_{x_1x_2}b(x_1,x_2) = \big(\frac{\partial^2 b}{\partial x_1^r \partial x_2^l}\big)_{rl}$ is the $d$ by $d$ matrix of mixed second order derivatives (each entry is a derivative with respect to one coordinate from each of $x_1$ and $x_2$).

**Theorem 1.2.13** (Non-linear Brenier Theorem, [Enn+22])**.** *Suppose that $m = 2$, that the domains $X_1, X_2 \subseteq \mathbb{R}^d$ are compact and that $\mu_1$ is absolutely continuous with respect to Lebesgue measure. Assume that $x_2 \mapsto D_{x_1}b(x_1,x_2)$ is injective for each fixed $x_1 \in X_1$, and that for each $(x_1,x_2) \in X_1 \times X_2$, $\det(D^2_{x_1x_2}b(x_1,x_2)) \neq 0$ and*

$$\boxed{\big\langle D_{x_2}b(x_1,x_2), [D^2_{x_1x_2}b(x_1,x_2)]^{-1} D_{x_1}b(x_1,x_2)\big\rangle > 0.} \tag{1.13}$$

*Then the part of the solution to* (1.10) *away from $\alpha = 0$ concentrates on the graph of a function over $x_1$. Furthermore, if $|\{\alpha = 0\}| = 0$, the solution is unique.*

**Remark 1.2.14.** *The injectivity of $x_2 \mapsto D_{x_1}b(x_1,x_2)$ is known as the* twist condition *in the optimal transport literature; it is well known that, together with the absolute continuity of $\mu_1$, it guarantees the Monge structure of the solution to the two marginal optimal transport problem with surplus $b$ (see, for example, [San15]). The invertibility of $D^2_{x_1x_2}b(x_1,x_2)$ is frequently referred to as* non-degeneracy*, and can be seen as a linearized version of the twist. Here we require additional hypotheses as we are dealing with the more sophisticated 3 marginal problem with surplus $s(\alpha,x_1,x_2) = \alpha b(x_1,x_2)$, which is equivalent to the spectral risk maximization* (1.10) *by Theorem 1.2.7.*

### 1.2.4 Multidimensional measures of risk

We end this section by showing a final result obtained in [Enn+22] concerning the framework proposed in [Rüs06] in which risk is measured in a multi-dimensional way: instead of a single output variable, we now have several, depending on the same underlying random variables; this corresponds to a vector valued output function $b : X_1 \times \ldots \times X_m \to \mathbb{R}^d$. A natural form of multivariate risk measures on the distribution $b_\# \gamma$ of output variables is then the maximal correlation measure from [Rüs06], which is defined by

$$\mathcal{R}_\nu(b_\#\gamma) = \max_{\eta \in \Pi(b_\#\gamma, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle z, y \rangle \, \mathrm{d}\eta$$

for some probability measure $\nu \in \mathcal{P}(\mathbb{R}^d)$. We note that it was proven in [EGH12] that any strongly coherent multi-variate risk measure takes this form. We consider the problem of maximizing $\mathcal{R}_\nu(b_\#\gamma)$ over all $\gamma \in \Pi(\mu_1, \ldots, \mu_m)$, where the $\mu_i$ as before represent the distributions of the underlying variables. Exactly as in Theorem 1.2.7, one can show that this problem is equivalent to the multi-marginal problem

$$\max_{\pi \in \Pi(\nu, \mu_1, \ldots, \mu_d)} \int \langle b(x_1, \ldots, x_m), y \rangle \, \mathrm{d}\pi. \tag{1.14}$$

This problem is more challenging than the case of a scalar valued $b$; nonetheless, we are able to obtain the following results in particular cases.

**Proposition 1.2.15** ([Enn+22]). *Suppose that the underlying variables $x_i$ are one dimensional. Assume that $\nu$ is concentrated on a smooth curve, that is $\nu = f_\# \mathcal{L}_{[0,1]}$ with $f : [0,1] \to \mathbb{R}^d$, where each component of $f$ is positive and monotone increasing, and each component $b_j$ of $b = (b_1, b_2, \dots, b_n)$ has positive second mixed derivatives and is monotone increasing in each $x_i$. Then $\Big((t, x_1, \dots, x_m) \mapsto (f(t), x_1, \dots, x_m)\Big)_\# \pi$ is optimal in (1.14), where $\pi = (\mathrm{id}, F_{\mu_1}^{-1}, F_{\mu_2}^{-1}, \dots, F_{\mu_m}^{-1})_\# \mathcal{L}_{[0,1]}$ is the comonotone coupling of $\mu_0, \mu_1, \dots, \mu_m$, where $\mu_0 = \mathcal{L}_{[0,1]}$.*

Notice that the positivity of mixed second derivatives of each components $b_j$ implies that they are compatibles so that one can again exploit the fact that in one dimension the optimal plan can be obtained by using the pseudo-inverse of the cumulative functions. In particular it turns out that the cost function $\langle b(x_1, \dots, x_m), f(t) \rangle = \sum_{j=1}^d b_j(x_1, \dots, x_m) f_j(t)$ is compatible, and so Proposition 1.1.8 implies the desired result.

For more general, diffuse $\nu$, we are able to prove that the solution is of Monge form and unique, provided that $b$ is invertible.

**Proposition 1.2.16** ([Enn+22]). *Assume that $\nu$ is absolutely continuous with respect to $d$ dimensional Lebesgue measure and $b : X_1 \times \ldots \times X_m \to \mathbb{R}^d$ is invertible. Then there exists a unique solution to (1.14). Furthermore, it concentrates on a graph over $y$.*

Note that the invertibility assumption on $b$ implies that the sum of the dimensions of the $X_i$ must be less than or equal to the dimension $d$ of $\nu$

## 1.3 Entropic Multi-Marginal Optimal Transport

We consider a $m-$uple of probability measures $\mu_i$ compactly supported on sub-manifolds $X_i \subseteq \mathbb{R}^d$of dimension $d_i$ and a cost function $c : X_1 \times \ldots \times X_m \to \mathbb{R}_+$. The Entropic Multi-Marginal Optimal Transport problem is defined as :

$$\mathsf{MOT}_\varepsilon^c := \inf \left\{ \int_{X_1 \times \ldots \times X_m} c \mathrm{d}\gamma + \varepsilon \mathrm{Ent}(\gamma | \otimes_{i=1}^m \mu_i) \mid \gamma \in \Pi(\mu_1, \dots, \mu_m) \right\}, \tag{$\mathrm{MOT}_\varepsilon$}$$

where $\mathrm{Ent}(\cdot | \otimes_{i=1}^m \mu_i)$ is the Boltzmann-Shannon relative entropy (or Kullback-Leibler divergence) w.r.t. the product measure $\otimes_{i=1}^m \mu_i$, defined for general probability measures $p, q$ as

$$\mathrm{Ent}(p \,|\, q) = \begin{cases} \displaystyle\int_{\mathbb{R}^d} \rho \log(\rho) \, \mathrm{d}q & \text{if } p = \rho q, \\ +\infty & \text{otherwise.} \end{cases}$$

Notice that the multi-marginal optimal transport problem we have seen before corresponds to the case where $\varepsilon = 0$.

Let us highlight that Entropic optimal transport (EOT) has found applications and proved to be an efficient way to approximate Optimal Transport (OT) problems, especially from a computational viewpoint. Indeed, when it comes to solving EOT by alternating Kullback-Leibler projections on the two marginal constraints, by the algebraic properties of the entropy such iterative projections correspond to the celebrated Sinkhorn's algorithm [Sin64], applied in this framework in the pioneering works [Cut13; Ben+15]. The simplicity and the good convergence guarantees (see [FL89; MG20]) of this method compared to the algorithms used for the OT problems, then determined the success of EOT for applications in machine learning, statistics, image processing, language processing and other areas (see the monograph [PC19] and references therein). As for multi-marginal optimal transport one can derive the dual problem of ($\text{MOT}_\varepsilon$) which reads as

$$\mathsf{MOT}^c_\varepsilon = \varepsilon + \sup\left\{\sum_{i=1}^m \int_{X_i} \varphi_i(x_i)\mathrm{d}\mu_i - \varepsilon \int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^m \varphi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}} \mathrm{d} \otimes_{i=1}^m \mu_i \mid \varphi_i \in \mathcal{C}_b(X_i)\right\}, \quad (\text{MD}_\varepsilon)$$

which is invariant by $(\varphi_1, \dots, \varphi_m) \mapsto (\varphi_1 + \lambda_1, \dots, \varphi_m + \lambda_m)$ where $(\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$ and $\sum_{i=1}^m \lambda_i = 0$, see [Léo14; NW22; MG20] for some recent presentations. It admits an equivalent *log-sum-exp* form:

$$\mathsf{MOT}^c_\varepsilon = \sup\left\{\sum_{i=1}^m \int_{X_i} \varphi_i(x_i)\mathrm{d}\mu_i - \varepsilon \log\left(\int_{\boldsymbol{X}} e^{\frac{\sum_{i=1}^m \varphi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}} \mathrm{d} \otimes_{i=1}^m \mu_i\right) \mid \varphi_i \in \mathcal{C}_b(X_i)\right\}, \quad (\text{MD}'_\varepsilon)$$

which is invariant by the same transformations without assuming $\sum_{i=1}^m \lambda_i = 0$.

From ($\text{MOT}_\varepsilon$) and ($\text{MD}_\varepsilon$) we recover, as $\varepsilon \to 0$, the unregularized multi-marginal optimal transport (MOT) and its dual (MD) we have introduced above. The link between multi-marginal optimal transport and its entropic regularization is very strong and a consequence of the $\Gamma-$convergence of ($\text{MOT}_\varepsilon$) towards (MOT) (one can adapt the proof in [Car+17] or see [BCN19; GKR20] for $\Gamma-$convergence in some specific cases) is that

$$\lim_{\varepsilon \to 0} \mathsf{MOT}^c_\varepsilon = \mathsf{MOT}^c_0.$$

By the direct method in the calculus of variations and strict convexity of the entropy, one can show that ($\text{MOT}_\varepsilon$) admits a unique solution $\gamma_\varepsilon$, called *optimal entropic plan*. Moreover, there exist $m$ real-valued Borel functions $\varphi_i^\varepsilon$ such that

$$\gamma_\varepsilon = \exp\left(\frac{\oplus_{i=1}^m \varphi_i^\varepsilon - c}{\varepsilon}\right) \otimes_{i=1}^m \mu_i, \tag{1.15}$$

where $\oplus_{i=1}^m \varphi_i^\varepsilon := (x_1, \dots, x_m) \mapsto \sum_{i=1}^m \varphi_i^\varepsilon(x_i)$, and in particular we have that

$$\mathsf{MOT}^c_\varepsilon = \sum_{i=1}^m \int_{X_i} \varphi_i^\varepsilon \,\mathrm{d}\mu_i \tag{1.16}$$

and these functions have continuous representatives and are uniquely determined up a.e. to additive constants. The reader is referred to the analysis of [MG20], to [Nen16] for the extension to the multi-marginal setting, and to [BL92; BLN94; Csi75; FG97; RT98] for earlier references on the two marginals framework.

The functions $\varphi_i^\varepsilon$ in (1.15) are called *Schrödinger potentials*, the terminology being motivated by the fact that they solve the dual problem ($\text{MD}_\varepsilon$) and are as such the (unique up to invariance) solutions to the so-called *Schrödinger system*: for all $i \in \{1, \dots, m\}$,

$$\varphi_i(x_i) = -\varepsilon \log \int_{\boldsymbol{X}_{-i}} e^{\frac{\oplus_{1\le j\le m, j\neq i} \varphi_j^\varepsilon - c(\boldsymbol{x})}{\varepsilon}} \mathrm{d} \oplus_{1\le j\le m, j\neq i} \mu_j \quad \text{for } \mu_i\text{-a.e. } x_i, \tag{1.17}$$

where $\boldsymbol{X}_{-i} = \prod_{1\le j\le m, j\neq i}^m X_j$. Note that (1.17) is a *softmin* version of the multi-marginal $c$-conjugacy relation for Kantorovich potentials.

## 1.4 Rate of convergences for entropic multi-marginal optimal transport

Since $\mathsf{MOT}^c_\varepsilon$ can be seen as a perturbation of $\mathsf{MOT}^c_0$, it is natural to study the behaviour as $\varepsilon$ vanishes. In this section we summarize the joint work with P. Pegon [NP23c] where we were mainly interested in investigating the rate of convergence of the entropic cost $\mathsf{MOT}^c_\varepsilon$ to $\mathsf{MOT}^c_0$ under some mild assumptions on the cost functions and marginals. In particular we have extended the techniques introduced in [CPT23] to the multi-marginal case which will also let us generalize the bounds in [CPT23] to the case of degenerate cost functions. For the two marginals and non-degenerate case we also refer the reader to a very recent (and elegant) paper [MS23] where the authors push a little further the analysis of the convergence rate by disentangling the roles of $\int c\mathrm{d}\gamma$ and the relative entropy in the total cost and deriving convergence rate for both these terms. Notice that concerning the convergence rate of the entropic multi-marginal optimal transport an upper bound has been already established in [EN22], which depends on the number of marginals and the quantization dimension of the optimal solutions to ($\mathrm{MOT}_\varepsilon$) with $\varepsilon = 0$.

Our main findings can be summarized as follows:

- we establish two **upper bounds**, one valid for locally Lipschitz costs and a finer one valid for locally semi-concave costs. For the latter, we improve the upper bound by a 1/2 factor, obtaining the following inequality for some $C^* \in \mathbb{R}_+$

$$\boxed{\mathsf{MOT}^c_\varepsilon \leq \mathsf{MOT}^c_0 + \frac{1}{2}\left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)\varepsilon\log(1/\varepsilon) + C^*\varepsilon.} \tag{1.18}$$

  We stress that this upper bound is smaller or equal than the one provided in [EN22, Theorem 3.8], which is of the form $\frac{1}{2}(m-1)D\varepsilon\log(1/\varepsilon) + O(\varepsilon)$ where $D$ is a quantization dimension of the support of an optimal transport plan. Thus $D$ must be greater or equal than the maximum dimension of the support of the marginals, and of course $\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i \leq (m-1)\max_{1\leq i\leq m} d_i$. The inequality may be strict for example in the two marginals case with unequal dimension.

- For the **lower bound**, from the dual formulation of ($\mathrm{MOT}_\varepsilon$) we have

$$\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 - \varepsilon\log\int_{\boldsymbol{X}} e^{-\frac{E(x_1,\dots,x_m)}{\varepsilon}}\mathrm{d}\otimes_{i=1}^m \mu_i(x_i),$$

  where $E(x_1,\dots,x_m) = c(x_1,\dots,x_m) - \oplus_{i=1}^m\varphi_i(x_i)$ is the duality gap and $(\varphi_1,\dots,\varphi_m)$ are Kantorovich potentials for the un-regularized problem ($\mathrm{MOT}_\varepsilon$) with $\varepsilon = 0$. By proving that $E$ detaches quadratically from the set $\{E = 0\}$ allows us to estimate the previous integral so that we get the following: given a $\kappa$ depending on a signature condition (see ($\mathrm{PS}(\kappa)$)) on the second mixed derivatives of the cost, the lower bound can be summerized as follows

$$\boxed{\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 + \frac{\kappa}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon.} \tag{1.19}$$

  for some $C_* \in \mathbb{R}_+$.

### 1.4.1 Upper bound

In this section we detail a little more our main results concerning the upper bounds and we give some elements of proof. We first consider the case of locally Lipschitz costs for which we obtain the following theorem.

**Theorem 1.4.1** ([NP23c]). *Assume that for $i \in \{1, \dots, m\}$, $\mu_i \in \mathcal{P}(X_i)$ is a compactly supported measure on a Lipschitz sub-manifold $X_i$ of dimension $d_i$ and $c \in \mathcal{C}^{0,1}_{\text{loc}}(\boldsymbol{X})$, then*

$$\boxed{\mathsf{MOT}^c_\varepsilon \leq \mathsf{MOT}^c_0 + \left(\sum_{i=1}^m d_i - \max_{j\in\{1,\dots,m\}} d_j\right)\varepsilon \log(1/\varepsilon) + O(\varepsilon).} \tag{1.20}$$

The proofs rely, as in [CPT23], on a multi-marginal variant of the block approximation introduced in [Car+17]. In particular, given the optimal solution $\gamma_0$ to the un-regularized problem one can build a competitor $\gamma_\delta$ for the entropic problem by using the block approximation. Then we have

$$\mathsf{MOT}^c_\varepsilon - \mathsf{MOT}^c_0 \leq \int c \mathrm{d}(\gamma_\delta - \gamma_0) + \varepsilon \mathrm{Ent}(\gamma_\delta | \otimes_{i=1}^m \mu_i), \tag{1.21}$$

where the first term can be controlled by using the fact that the cost function is Lipschitz and the second term is smaller equal than $\sum_{i=}^{m-1} H_\delta(\mu_i)$ where $H_\delta(\mu_i)$ is Renyi's[1] dimension of the measure $\mu_i$. Now, by exploiting the fact we deal with a relative entropy and the block approximation, we can get rid of the Renyi's dimension of one measure which can be the one concentrated on the higher dimensional sub-manifold (in this case $\mu_m$ in order to simplify notation). Thus, since the $\mu_i$ are concentrated on submanifold of dimension $d_i$ we have that $H_\delta(\mu_i) \leq d_i \log(1/\delta) + C$, with $C > 0$, and by choosing $\delta = \varepsilon$ we obtain the desired result. Notice that in this case the bound will depend only on the dimension of the support of the marginals **and not on the optimal transport plan**. Besides, notice that by taking $m = 2$ and $d_1 = d_2 = d$, one easily retrieves [CPT23, Proposition 3.1]. More involved, it is the case in which we consider a locally semi-concave cost functions. In particular we can use an integral variant of Alexandrov Theorem which is proved in [CPT23].

**Lemma 1.4.2** ([CPT23, Lemma 3.6]). *Let $f : \Omega \to \mathbb{R}$ be a $\lambda$-concave function defined on a convex open set $\Omega \subseteq \mathbb{R}^d$, for some $\lambda \geq 0$. There exists a constant $C \geq 0$ depending only on $d$ such that:*

$$\int_\Omega \sup_{y\in B_r(x)\cap\Omega} |f(y) - (f(x) + \langle \nabla f(x), (y - x)\rangle)| \, \mathrm{d}x \leq Cr^2 \mathcal{H}^{d-1}(\partial\Omega)([f]_{\mathcal{C}^{0,1}(\Omega)} + \lambda \operatorname{diam}(\Omega)), \tag{1.22}$$

*where $[f]_{\mathcal{C}^{0,1}(\Omega)}$ denotes the Lipschitz constant of $f$.*

[1] **Rényi dimension** (following [You82]): If $\mu$ is a probability measure over a metric space $X$, we set for every $\delta > 0$,

$$H_\delta(\mu) = \inf\left\{\sum_{n\in\mathbb{N}} \mu(A_n)\log(1/\mu(A_n)) \mid \forall n, \operatorname{diam}(A_n) \leq \delta, \text{ and } X = \bigsqcup_{n\in\mathbb{N}} A_n\right\},$$

where the infimum is taken over countable partitions $(A_n)_{n\in\mathbb{N}}$ of $X$ by Borel subsets of diameter less than $\delta$, and we define the *lower and upper entropy dimension of* $\mu$ respectively by:

$$\underline{\dim}_R(\mu) := \liminf_{\delta\to 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}, \quad \overline{\dim}_R(\mu) := \limsup_{\delta\to 0^+} \frac{H_\delta(\mu)}{\log(1/\delta)}.$$

However, since we are considering (and exploiting) probability measures concentrated on smooth sub-manifolds of dimension $d_i$, we have to carefully define local semi-concavity in order to apply the previous Lemma:

**Definition 1.4.3.** *A function $f : X \to \mathbb{R}$ defined on a $\mathcal{C}^2$ sub-manifold $X \subseteq \mathbb{R}^N$ of dimension $d$ is* locally semiconcave *if for every $x \in X$ there exists a local chart (i.e. a $\mathcal{C}^2$ diffeomorphism) $\psi : U \to \Omega$ where $U \subseteq X$ is an open neighborhood of $x$ and $\Omega$ is an open convex subset of $\mathbb{R}^d$, such that $f \circ \psi^{-1}$ is $\lambda$-concave for some $\lambda \in \mathbb{R}$, meaning $f \circ \psi^{-1} - \lambda \frac{|\cdot|^2}{2}$ is concave on $\Omega$.*

We obtain then the following theorem

**Theorem 1.4.4.** *Let $c$ be locally semiconcave and assume that for every $i \in \{1, \ldots, m\}$, $X_i \subseteq \mathbb{R}^N$ is a $\mathcal{C}^2$ sub-manifold of dimension $d_i$ and $\mu_i \in L^\infty(\mathcal{H}^{d_i}_{X_i})$ is a probability measure compactly supported in $X_i$. Then there exists constants $\varepsilon_0, C^* \geq 0$ such that for $\varepsilon \in (0, \varepsilon_0]$*

$$\boxed{\mathsf{MOT}^c_\varepsilon \leq \mathsf{MOT}^c_0 + \frac{1}{2}\left(\sum_{i=1}^m d_i - \max_{1\leq i\leq m} d_i\right)\varepsilon \log(1/\varepsilon) + C^*\varepsilon.} \tag{1.23}$$

The proof consists in better estimating the term $\int c\mathrm{d}(\gamma_\delta - \gamma_0)$ in (1.21). Indeed we have that

$$\int c\mathrm{d}(\gamma_\delta - \gamma_0) = \int E \mathrm{d}\gamma_\delta,$$

where $E = c - \oplus_{i=1}^m \varphi_i$ with $\varphi_i$ being the optimal potentials for the un-regularized problem. By exploiting the fact that $c \in \mathcal{C}^2$, the local semiconcavity of the potentials (inherited by the cost function), the fact that $\mu_i \in L^\infty$ and Lemma 1.4.2 we can improve the upper bound of a factor 1/2.

### 1.4.2 Lower bound

We are going to establish a a lower bound in the same form as the fine upper bound of 1.4.4, the dimensional constant being this time related to the signature of some bilinear forms by using Theorem 1.1.1 and following ideas from the proof of it proposed in [Pas12]. Let us consider the following signature condition:

$$\text{for every } \boldsymbol{x} \in \boldsymbol{X}, \quad d^+_c(\boldsymbol{x}) \geq \kappa \quad \text{where} \quad d^\pm_c(\boldsymbol{x}) \coloneqq \max\left\{d^\pm(g)(\boldsymbol{x}) \mid g \in G_c\right\}. \tag{PS($\kappa$)}$$

Then we can state the main result concerning the lower bound.

**Theorem 1.4.5.** *Let $c \in \mathcal{C}^2(\boldsymbol{X})$ and assume that for every $i \in \{1, \ldots, m\}$, $X_i \subseteq \mathbb{R}^d$ is a $\mathcal{C}^2$ sub-manifold of dimension $d_i$ and $\mu_i \in L^\infty(\mathcal{H}^{d_i}_{X_i})$ be a probability measure compactly supported in $X_i$. If* (PS($\kappa$)) *is satisfied, then there exists a constant $C_* \in [0, \infty)$ such that for every $\varepsilon > 0$,*

$$\boxed{\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 + \frac{\kappa}{2}\varepsilon \log(1/\varepsilon) - C_*\varepsilon.} \tag{1.24}$$

As already mentioned at the beginning of this section the proof of this result relies on the dual formulation of (MOT$_\varepsilon$)

$$\mathsf{MOT}^c_\varepsilon \geq \mathsf{MOT}^c_0 - \varepsilon \log \int_{\boldsymbol{X}} e^{-\frac{E(x_1,\ldots,x_m)}{\varepsilon}} \mathrm{d} \otimes_{i=1}^m \mu_i(x_i),$$

where $E(x_1, \dots, x_m) = c(x_1, \dots, x_m) - \oplus_{i=1}^m \varphi_i(x_i)$ is the duality gap and $(\varphi_1, \dots, \varphi_m)$ are Kantorovich potentials for the un-regularized problem ($\mathrm{MOT}_\varepsilon$) with $\varepsilon = 0$. By using the singular values decomposition of a well-chosen bilinear form obtained as an average of mixed second derivatives of the cost and the signature condition introduced above we are able to prove that $E$ detaches quadratically from the set $\{E = 0\}$ that is

$$\frac{E(\boldsymbol{x}') + E(\boldsymbol{x})}{2} \geq \left|u^+(\boldsymbol{x}') - u^+(\boldsymbol{x})\right|^2 - \left|u^-(\boldsymbol{x}') - u^-(\boldsymbol{x})\right|^2 - \eta(r)\left|u(\boldsymbol{x}') - u(\boldsymbol{x})\right|^2 \tag{1.25}$$

for $\eta(r) \geq 0$, $\eta \to 0$, and where $u = (u^0, u^+, u^-)$ are the local coordinate associate to the singular values decomposition of a suitable bilinear form in $G_c$, as defined in (1.1). This allows us to estimate the previous integral, in a similar way as in [CPT23], namely for some constant $C > 0$ and for every $\varepsilon > 0$, we get

$$\int e^{-E/\varepsilon} \mathrm{d} \otimes_{i=1}^m \mu_i \leq C\varepsilon^{\kappa/2}.$$

**Remark 1.4.6.** *Theorem 1.4.5 improves the result in [EN22] where only an upper bound depending on the quantization dimension of the solution to the un-regularized problem is provided. Moreover, this slightly more flexible use of Minty's trick compared to [CPT23] allows us to obtain a lower bound also for degenerate cost functions in the two marginals setting.*

### 1.4.3 Matching bound

We present here some examples where the upper and lower bounds coincide giving us a developement of the entropic cost.

**Example 1.4.7 (Two marginals case).** *In previous works [CPT23; EN22] concerning the rate of convergence for the two marginals problem, it was assumed that the cost function must satisfy a non degeneracy condition, that is $D^2_{x_1x_2}c$ must be of full rank. A direct consequence of our analysis is that we can provide a lower bound (the upper bound does not depend on such a condition) for costs for which the non-degeneracy condition fails. Let $r$ be the rank of $D^2_{x_1x_2}c$ at the point where the non-degeneracy condition fails, then the signature of $\overline{g}$ at this point is given by $(r, r, 2d-2r)$ meaning that locally the support of the optimal $\gamma_0$ is at most $2d-r$ dimensional. Thus, the bounds become*

$$\boxed{\frac{r}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon \leq \mathsf{OT}^c_\varepsilon - \mathsf{OT}^c_0 \leq \frac{d}{2}\varepsilon\log(1/\varepsilon) + C^*\varepsilon,}$$

*for some constants $C_*, C^* > 0$. Notice that if $D^2_{x_1,x_2}c$ has full rank then $r = d$ and we retrieve the matching bound results of [CPT23; EN22].*

**Example 1.4.8 (Two marginals case and unequal dimension).** *Consider now the two marginals case but unequal dimensional (the unequal dimensional OT will be discussed more precisely in chapter 3), that is for example $d_1 > d_2$. Then, if $D^2_{x_1,x_2}c$ has full rank, that is $r = d_2$, we obtain a matching bound depending only on the lower dimensional marginal*

$$\boxed{\frac{d_2}{2}\varepsilon\log(1/\varepsilon) - C_*\varepsilon \leq \mathsf{OT}^c_\varepsilon - \mathsf{OT}^c_0 \leq \frac{d_2}{2}\varepsilon\log(1/\varepsilon) + C^*\varepsilon,}$$

*for some constants $C_*, C^* > 0$. If $\mu_1$ is absolutely continuous with respect to $\mathcal{H}^{d_1}$ on some smooth sub-manifold of dimension $d_1$, then any optimal transport plan would be concentrated on a set of Hausdorff dimension no less than $d_1$, and thus the upper bound given in [EN22, Theorem 3.8] would be $\frac{d_1}{2}\varepsilon\log(1/\varepsilon) + O(\varepsilon)$, which is strictly worse than our estimate.*

**Example 1.4.9 (Gangbo-Święch cost and Wasserstein barycenter).** *Suppose that* $c(x_1,\dots,x_m)=\sum_{i<j}|x_i-x_j|^2$*, known as the Gangbo-Święch cost [GŚ98]. Notice that the cost is equivalent to* $c(x_1,\dots,x_m)=h(\sum_{i<j}x_i)$ *where* $h$ *is* $\mathcal{C}^2$ *and* $D^2h<0$*,then the signature of* $\overline{g}$ *is* $((m-1)d,d,0)$ *and we have a matching bound*

$$\boxed{\frac{1}{2}\Big((m-1)d\Big)\varepsilon\log(1/\varepsilon)-C_*\varepsilon\leq \mathsf{MOT}_\varepsilon^c-\mathsf{MOT}_0^c\leq\frac{1}{2}\Big((m-1)d\Big)\varepsilon\log(1/\varepsilon)+C^*\varepsilon.}$$

*Notice now that considering the* $\mathsf{MOT}_0^c$ *problem with a cost* $c(x_1,\dots,x_m)=\sum_i|x_i-T(x_1,\dots,x_m)|^2$*, where* $T(x_1,\dots,x_m)=\sum_{i=1}^m\lambda_i x_i$ *is the Euclidean barycenter , is equivalent to the* $\mathsf{MOT}_0^c$ *with the Gangbo- Święch cost and the matching bound above still holds. Moreover, the multi-marginal problem with this particular cost has been shown [AC11] to be equivalent to the Wasserstein barycenter, that is* $T_\sharp\gamma_0=\nu$ *is the barycenter of* $\mu_1,\dots,\mu_m$*.*

## 1.5 An ODE characterisation of discrete entropic MMOT

In many cases of interest, the cost function $c(x_1,....x_m)=\sum_{i,j=1,i\neq j}^m w(x_i,x_j)$ is given by a sum of two marginal cost functions; when $w(x_i,x_j)=|x_i-x_j|^2$, for instance the multi-marginal problem is equivalent to the well known Wasserstein barycenter problem (see Proposition 4.2 in [AC11]), while the Coulomb cost $w(x_i,x_j)=\frac{1}{|x_i-x_j|}$ plays a central role in the quantum chemistry applications pioneered in [CFK13] and [BDG12].

In a joint work with B. Pass [NP23a] for such pairwise interaction costs, we develop a continuation method which, by introducing a suitable one parameter family of cost functions, establishes a link between the original multi-marginal problem and a simpler one whose complexity scales linearly in the number of marginals. For discrete marginals, we show that, after the addition of an entropic regularization term, the solution of the original multi-marginal problem can be recovered by solving an ordinary differential equation (ODE) whose initial condition is the solution to the simpler problem.

This method is actually inspired by the one introduced in [CGS09] to compute the Monge solution of the two marginal problem, starting from the Knothe-Rosenblatt rearrangement; note, however, note that since we apply this strategy to a regularized problem, our resulting ODE enjoys better regularity than the one in [CGS09], which, in turn, makes it amenable to higher order numerical schemes. The above mentioned differential equation will be derived by differentiating the optimality conditions of the dual problem; in particular by penalizing the constraints with the soft-max function we will obtain a well defined ODE for which existence and uniqueness of a solution can be established.

### 1.5.1 Discrete Entropic Multi-Marginal Optimal Transport problem

From now on we take finite sets $X_i=\{x_i^1,\dots,x_i^N\}$ and discrete measures $\mu_i=\sum_{j=1}^N\mu_i^j\delta_{x_i^j}$. For sake of simplicity we will keep the continuous in space notation, but integral must be understood as finite sums. Consider then the discrete entropic multi-marginal optimal transport problem defined as

$$\mathsf{MOT}_\varepsilon^{c,M}(\mu_1,\dots,\mu_m):=\inf\left\{\int_{\boldsymbol{X}}c\mathrm{d}\gamma+\varepsilon\mathrm{Ent}(\gamma|\otimes_{i=1}^m\mu_i)\mid\gamma\in\Pi(\mu_1,\dots,\mu_m)\right\},\qquad(\mathrm{MOT}_\varepsilon^M)$$

as well as the associated dual problem (written in a slightly different way with respect to the previous section)

$$\sup\left\{\sum_{i=1}^{m}\int_{X_i}\varphi_i \mathrm{d}\mu_i - \varepsilon\int_{\boldsymbol{X}}\exp\left(\frac{\sum_i \varphi_i - c(\boldsymbol{x})}{\varepsilon}\right)\mathrm{d}\otimes_{i=1}^{m}\mu_i\right\}, \tag{1.26}$$

where now the $\varphi_i \in \mathbb{R}^{X_i}$ are functions over finite sets. We note in particular that (1.26) is an *unconstrained* finite dimensional concave maximization problem. Solutions may be computed using a multi-marginal version of the Sinkhorn algorithm [Ben+15; Cut13; PC19; Gal16; Chi+16], and one can then recover the optimal $\gamma$ in ($\mathrm{MOT}_\varepsilon^M$) from the solutions $\varphi_1, \dots, \varphi_m$ to (1.26) via the well known formula:

$$\gamma_{\boldsymbol{x}} = \exp\left(\frac{\sum_{i=1}^m \varphi_i(x_i) - c(\boldsymbol{x})}{\varepsilon}\right)\mu^1_{x^1}\mu^2_{x_2}...\mu^m_{x_m}.$$

Here, we are especially interested in cost functions $c(x_1, \dots, x_m)$ involving pair-wise interactions, that is

$$c(x_1, \dots, x_m) = \sum_{i<j}^{m} w(x_i, x_j).$$

Such costs are ubiquitous in applications: for example, for systems of interacting classical particles in [CFK13; BDG12], $c$ is a pair-wise cost, with $w(x-y) = \dfrac{1}{|x-y|}$, known as the Coulomb cost. The case where $w(x, y) = |x-y|^2$ is the quadratic distance is well known to be equivalent to the Wasserstein barycenter problem (see Proposition 4.2 in [AC11]), which has a wide variety of applications in statistics, machine learning and image processing, among other areas. Pairwise costs with pair-dependent interactions, that is, costs of the form $c(x_1, \dots, x_m) = \sum_{i<j}^m w_{ij}(x_i, x_j)$ where the $w_{ij}$ are not necessarily the same for each choice of $i$ and $j$, also arise in the time discretization of Brenier's relaxation [Bre89] of Arnold's variational formulation of incompressible Euler equation, and in inference problems for probabilistic graphical models [Haa+21]. Our results adapt immediately to such costs. Let us now consider costs $c_\eta$ of the form

$$c_\eta(x_1, \dots, x_m) := \eta\sum_{i=2}^{m}\sum_{j=i+1}^{m} w(x_i, x_j) + \sum_{i=2}^{m} w(x_1, x_i). \tag{1.27}$$

It is clear that when $\eta = 1$ we retrieve a pair-wise cost as defined above whereas in the limit $\eta \to 0$ we obtain a cost involving only the interactions between $x_1$ and the other $x_i$ individually. Later on, we will develop an ordinary differential equation that governs the evolution with $\eta$ of the solutions to the regularized dual problem (1.26). In particular Proposition 1 in [NP23a] asserts that the initial condition for that equation (that is, the solutions when $\eta = 0$) can be recovered by solving each of the individual two marginal problems between $\mu_1$ and $\mu_i$.

### 1.5.2 An ODE characterisation of multi-marginal optimal transport

We now turn our attention to developing an ODE for the Kantorovich potentials, woorking with the regularized discrete problem ($\mathrm{MOT}_\varepsilon^M$) and its dual (1.26) with pairwise cost (1.27), we make the following, standing assumptions throughout this section:

1. (Equal marginals) All the marginals are equal $\mu_i = \rho = \sum_{x\in X}\rho_x\delta_x$, where $X$ is a finite subset.

2. (Symmetric cost) The two body cost $w$ is symmetric $w(x,y) = w(x,y)$.

3. (Finite cost) The two body cost function $w : X \times X \to \mathbb{R}$ is everywhere real-valued.

Notice now that although the cost (1.27) at $\eta = 1$ is symmetric in the variables $x_1, x_2, \dots, x_m$, the one at $\eta < 1$ is not. It is, however, symmetric in the variables $x_2, ..., x_m$; this means that the optimal $\varphi_i$ in (1.26) satisfy $\varphi_i = \varphi_j = \varphi$ for $i, j \geq 2$ and so, setting $\varphi_1 = \psi$, we can rewrite (1.26) as

$$\inf_{\varphi,\psi \in \mathbb{R}^X} \{\Phi(\varphi, \psi, \eta)\}, \tag{1.28}$$

where

$$\Phi(\varphi, \psi, \eta) := -(m-1)\int_X \varphi \mathrm{d}\rho - \int_X \psi \mathrm{d}\rho + \varepsilon \int_{X^m} \exp\left(\frac{\sum_{i=2}^m \varphi(x_i) + \psi(x_1) - c_\eta(\boldsymbol{x})}{\varepsilon}\right) \mathrm{d}\otimes^m \rho.$$

Since the functional $\Phi(\varphi, \psi, \eta)$ is convex on the set $\{\varphi, \psi : X \to \mathbb{R}\} \approx \mathbb{R}^{2|X|}$, as the sum of a linear and an exponential function, optimal solutions $(\varphi^*, \psi^*)$ can be characterized by the first order optimality conditions $\nabla_\varphi \Phi = \nabla_\psi \Phi = 0$ which allow to rewrite the optimal potential $\psi^*$ as a function of $\psi^*$. By exploiting this problem (1.28) can be re-written in a more compact form

$$\inf_{\varphi : X \to \mathbb{R}} \left\{\tilde{\Phi}(\varphi, \eta)\right\}, \tag{1.29}$$

where

$$\tilde{\Phi}(\varphi, \eta) := -(m-1)\int_X \varphi \mathrm{d}\rho + \varepsilon \int_X \log\left(\int_{X^{m-1}} \exp\left(\frac{\sum_{i=2}^m \varphi(x_i) - c_\eta(\boldsymbol{x})}{\varepsilon}\right) \mathrm{d}\otimes^{m-1} \rho\right) \mathrm{d}\rho.$$

It is well known that the solution to (1.26) is unique up to the addition of constants $\varphi_i \mapsto \varphi_i + C_i$ adding to 0, $\sum_{i=1}^m C_i = 0$; thus, solutions to (1.29) are unique up to the addition of a single constant, $\varphi \mapsto \varphi + C$. We therefore impose the normalization

$$\varphi(x_0) = 0 \tag{1.30}$$

for all $\eta \in [0,1]$ and a fixed $x_0 \in X$. The problem (1.29), restricted to $\varphi$'s satisfying (1.30) then has a unique solution; the function $\tilde{\Phi}(\cdot, \eta)$ is strictly convex when restricted to this set, and the solution $\varphi^*_\eta$ can be characterized by the optimality condition $\nabla_\varphi \tilde{\Phi}(\varphi^*, \eta) = 0$ Our numerical method consists then in solving an ODE for the evolution of $\varphi(\eta) = \varphi^*_\eta$ (we consider now a curve of optimal potential $\varphi^*_\eta$) obtained by differentiating the above mentioned optimality condition with respect to $\eta$. It turns out that $\varphi$ can be fully characterized as the solution to the following Cauchy problem:

$$\boxed{\begin{cases} \dfrac{\mathrm{d}\varphi}{\mathrm{d}\eta}(\eta) = -[D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi(\eta), \eta)]^{-1} \dfrac{\partial}{\partial \eta} \nabla_\varphi \tilde{\Phi}(\varphi(\eta), \eta), \\ \varphi(0) = \varphi_w, \end{cases}} \tag{1.31}$$

where, the initial value $\varphi(0)$ of $\varphi$ when $\eta = 0$ coincides with the optimal potential $\varphi_w$ for the two marginal optimal transport problem with cost $w$.

The following well-posedness theorem then holds.

**Theorem 1.5.1.** *Let $\varphi(\eta)$ be the solution to* (1.29) *for all $\eta \in [0,1]$. Then $\eta \mapsto \varphi(\eta)$ is $\mathcal{C}^1$ and is the unique solution to the Cauchy problem* (1.31).

The main step of the proof consists in showing that the pure second derivatives with respect to $\varphi$ as well as the mixed second derivatives with respect to $\varphi$ and $\varepsilon$ exist and are Lipschitz, and the Hessian with respect to $\varphi$ is invertible. In particular thanks to the boundedness of the cost function, namely $||c_\eta||_\infty \leq M$, allows to prove that the potentials (the techniques we used were largely by inspired by Carlier's beautiful work on the linear convergence of multi-marginal Sinkhorn [Car22]) are also bounded and one can restrict the study of the well-posedness of the ODE on the set

$$U := \{\varphi \in \mathbb{R}^X \mid \varphi_{x_0} = 0, \ \|\varphi\|_\infty \leq 4M\}. \tag{1.32}$$

On this set the functional $\tilde{\Phi}$ is now strongly concave and this ensures that the Hessian is invertible.

### 1.5.3 The algorithm

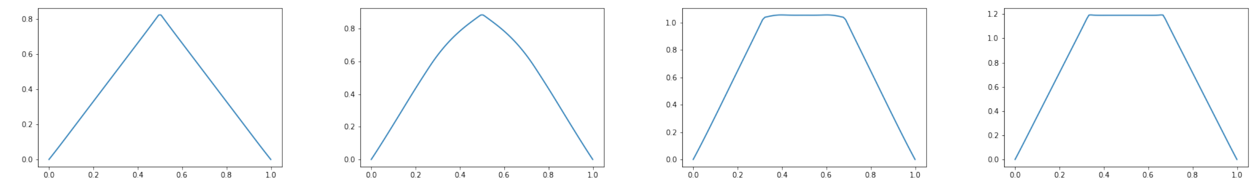
The algorithm consists in discretizing (1.31) by an explicit Euler scheme (notice that one could also use some high order method for the ODEs). Let $h$ be the step size and set $\varphi(0) = \varphi_w$ the solution of a 2 marginal problem with cost $w$, then the $\varphi$ can be defined inductively as detailed in 1.

**Algorithm 1** Algorithm to compute the $\varphi$ via explicit Euler method

**Require:** $\varphi(0) = \varphi_w$
**while** $||\varphi^{(k+1)} - \varphi^{(k)}|| <$tol **do**
  $D^{(k)} := D^2_{\varphi,\varphi}\tilde{\Phi}(\varphi^{(k)}, kh)$
  $b^{(k)} := -\dfrac{\partial}{\partial \eta}\nabla_\varphi \tilde{\Phi}(\varphi^{(k)}, kh)$
  Solve $D^{(k)}z = b^{(k)}$
  $\varphi^{(k+1)} = \varphi^{(k)} + hz$
**end while**

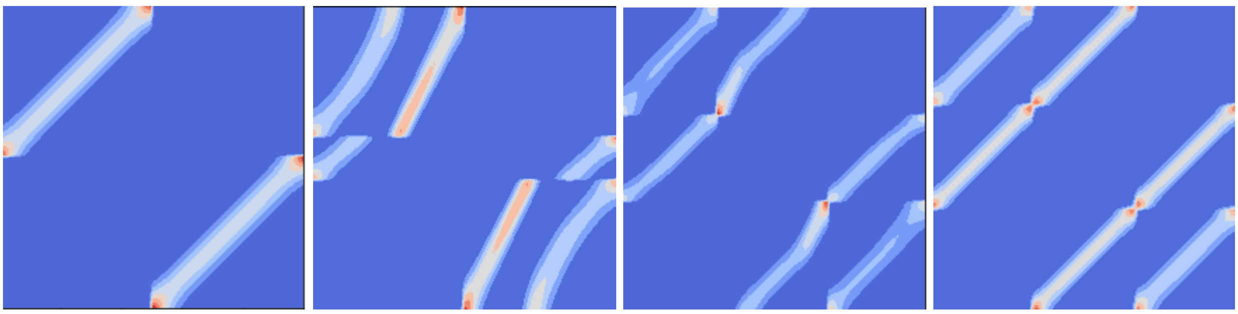
**Remark 1.5.2** (Complexity in the marginals)**.** *We highlight that this ODE approach does not overpass the NP hardness (see [AB21]) of some multi-marginal problems, as the one with the Coulomb cost, and it still suffers the exponential complexity in the number of marginals. However, since the ODE is smooth, one can try apply an high order method (we will do a careful analysis of it later in the section) which can converge quickly in the number of iterations and have a computational time competitive with respect to Sinkhorn. We also note that, if the symmetry assumptions are dropped, as we have brifley discussed above, one needs to solve a system of $(N-1)m$ ODEs. The matrix $D^{(k)}$ in Algorithm 1 is then be $(N-1)m \times (N-1)m$ instead of $m \times m$, increasing the leading term in the complexity by the relatively manageable multiplicative factor $(N-1)^2$. Numerical results for one such case are presented below.*

In Figures 1.2 and 1.1 we show the numerical result obtained with $\varepsilon = 0.006$, $h = 1/1000$, $m = 3$, the uniform measure on $[0,1]$ uniformily discretized with 100 gridpoints and the pairwise interaction $w(x,y) = -\log(0.1 + |x-y|)$. Notice that since we have developed our continuation method by the entropic regularization of optimal transport, we can easily reconstruct the optimal (regularized) plan at each time $k$ by using the potential $\varphi^{(k)}$. Moreover, it is interesting to notice that the optimal plan at each time step of the ODE stays deterministic (taking into account the blurring effect of the entropic regularization); that is, it is concentrated on a low dimensional structure.



$\eta = 0$ $\eta = 0.25$ $\eta = 0.75$ $\eta = 1$

Figure 1.1: (Log cost) potential $\varphi(\eta)$.



$\eta = 0$ $\eta = 0.25$ $\eta = 0.75$ $\eta = 1$

Figure 1.2: (Log cost) support of the coupling $\gamma^{\eta}_{1,2}$.

**Remark 1.5.3** (Intermediate step of the ODE approach)**.** *It is interesting to highlight that each ith-step of the ODE approach actually returns the optimal solution to the multi-marginal problem with the cost* $c_{kh}$*. This, in particular, implies that one can also use this approach in order to retrieve the solutions of many multi-marginal problems with different costs, by choosing a suitable* $c_\eta$*, instead of solving these problems individually using Sinkhorn.*

We are finally interested in comparing the ODE approach and the Sinkhorn algorithm in terms of performance. In order to do this we consider the optimal solution of the regularized dual problem obtained with a gradient descent with backtracking and take it as the reference solution to compute the relative error $\frac{\|\varphi - \varphi_{ref}\|_\infty}{\|\varphi_{ref}\|_\infty}$. Concerning the ODE, since we have already remarked above that it is smooth, we consider different high order methods such as 3rd, 5th and 8th order Runge-Kutta methods. By looking at Table 1.1, it is clear that all the methods achieve almost the same relative error, but the number of iterations to reach it as well as the CPU time in seconds slightly differs. In particular we remark that a 3rd order RK is faster than a Sinkhorn in terms of computational time and iterations but less precise. The other RK methods achieve comparable precision to Sinkhorn with less iterations but the computational cost at the step of the ODE becomes now quite onerous implying a significant increase (especially 8th RK) in terms of CPU time.

| | 3rd RK | 5th RK | 8th RK | Sinkhorn |
|---|---|---|---|---|
| relative error | $1.47 \times 10^{-5}$ | $7.8 \times 10^{-6}$ | $7.62 \times 10^{-6}$ | $5.46 \times 10^{-6}$ |
| iterations | 87 | 87 | 87 | 820 |
| CPU time (sec) | 72.39 | 158.9 | 385.1 | 102.8 |

Table 1.1: Comparison between the ODE approach and Sinkhorn for the uniform density and 400 gridpoints

**The non-symmetric ODE: Euler case** In a series of papers [Bre89; Bre93; Bre99] Brenier proposed a relaxation of the incompressible Euler equation with constrained initial and final data interpreted as a geodesic on the group of measure preserving diffeomorphisms. Brenier's relaxed formulation consists in finding a probability measure over absolutely continuous paths which minimizes the average kinetic energy. In this framework the incompressibility is encoded by an additional constraint that at each time $t$, the distribution of position need be uniform. If we consider a uniform discretization of $[0, T]$ (where $T$ is the final time) with $m$ steps in time, we recover a multi-marginal formulation of the Brenier principle with the specific cost function

$$c(x_1, \dots, x_m) = \frac{m^2}{2T^2} \sum_{i=1}^{m-1} |x_{i+1} - x_i|^2$$

representing the discretized, in time, kinetic energy ($|\cdot|$ denotes the standard euclidean norm). The coupling $\gamma$ is the probability to find a *generalized particle* on the discrete path $x_1, \dots, x_m$. Because the fluid is incompressible, the $i$th marginal $\mu_i$ of $\gamma$ is the uniform on the $d-$dimensional cube $[0, 1]^d$. For each $i \in \{1, \dots, m\}$ the transition probability from time 1 to time $i$ is given by the coupling

$$\gamma_{1,i} = \Big((x_1, \dots, x_m) \mapsto (x_1, x_i)\Big)_{\#} \gamma,$$

which represents the probability of finding a *generalized particle* initially at $x_1$ to be at $x_i$ at time $i$. In order to impose the initial and final constraint we include by penalization, that is by adding a term to the cost function which now reads as

$$c(x_1, \dots, x_m) = \frac{m^2}{2T^2} \sum_{i=1}^{m-1} |x_{i+1} - x_i|^2 + \beta |F(x_1) - x_m|^2,$$

where $\beta > 0$ is a penalization parameter in order to enforce the initial-final constraint. $F(x_1)$ represents the prescribed final position of the particle initially at position $x_1$, and the coupling $\gamma$ can be interpreted as a (generalization of) a discrete time geodesic between the identity mapping and $F$ on the space of measure preserving maps. For an efficient implementation of Sinkhorn in this case we refer the reader to [BCN19; Nen16; Ben+15]. If we consider now the ODE setting, we have now to deal with a non symmetric case (the cost is not symmetric in the marginals anymore) and so to solve a system, still well posed, of ODEs. In particular we consider the following $c_\eta$ cost

$$c_\eta(x_1, \dots, x_m) = \frac{m^2}{2T^2} |x_2 - x_1|^2 + \eta \left( \frac{m^2}{2T^2} \sum_{i=2}^{m-1} |x_{i+1} - x_i|^2 \right) + \beta |F(x_1) - x_m|^2.$$

For the numerical simulations we took 100 gridpoints discretization of $[0, 1]$, $m = 9$ time marginals, $\beta = 20$ and $\varepsilon = 0.002$. In this case, we solved the ODE system by using a $5th$

order Runge-Kutta method and $h = 1/100$. In Figures 1.3 and 1.4 we plot the transition couplings for the initial-final for two different initial-final configurations: $F(x) = 1 - x$ and $F(x) = (x + 1/2) \bmod 1$.

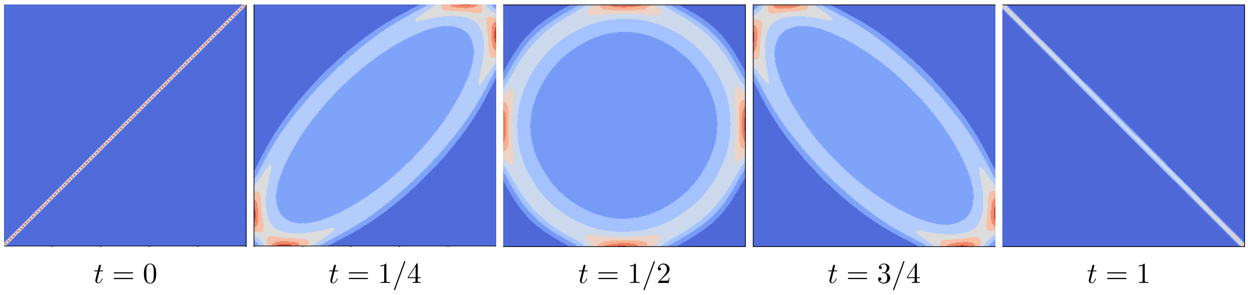



Figure 1.3: Transition plan $\gamma_{1,i}$ for $F(x) = 1 - x$ of Brenier relaxed formulation of incompressible Euler equations.

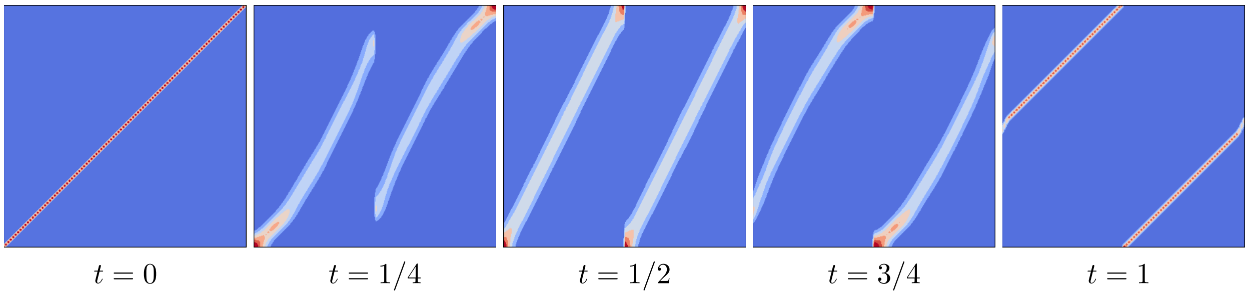



Figure 1.4: Transition plan $\gamma_{1,i}$ for $F(x) = (x + 1/2) \bmod 1$ of Brenier relaxed formulation of incompressible Euler equations.

## 1.6 Perspectives

Let us mention in this section some research perspectives:

- **Particle discretization of $\mathsf{MOT}^c$:** we want study the behaviour of a simple **particle discretization** of the problem, where the solution is a sum of a finite number of Dirac masses whose positions are free to move, that is $\gamma = \sum_i^N \delta_{X_i}$ where $X_i \in \mathbb{R}^{dm}$. The constraints could be enforced via penalization, using optimal transport data attachment terms. The problem we want to solve then takes the following form

$$\boxed{\min_{X^1,\cdots,X^m} \sum_{i=1}^{N} c(X_i^1, \cdots, X_i^m) + \lambda\Big(\sum_{j=1}^{m} \mathsf{OT}^{d^2}(\mu_j, e_{j\sharp}\gamma)\Big),} \tag{1.33}$$

where we take the quadratic cost $d^2(x, y) = |x - y|^2$ in the penalization terms and $X_i^j$ is the $j-$th component of the vector $X_i$. This discretization makes the discrete problem non-convex, but one could hope to study the convergence of global minimizers of $\mathsf{MOT}^c$

depending on the dimension of the support of the solution, on the size of the discretization and the strength of the penalization. Establishing convergence results for this discretization in the spirit of $\Gamma$-convergence should then be within hand, using similar arguments as [Sar22], and will extend the result in [BGV19] for the case of Wasserstein splines. Notice that a coarser discretization, thanks to a better understanding of the structure of the solution, will make the numerical method competitive with respect to the actual ones: in the entropic case the dimension of the support of the solution is naturally high dimensional since the regularization effect, meaning that a finer discretization is needed. A recent result of [MSS21] makes us optimistic that despite the non-convexity of this particle discretization, it would be possible to prove (at least in some cases) that a simple gradient descent algorithm is sufficient to reach a low energy configuration. This is a difficult question, that would be of prime interest to many communities, from optimal transport to machine-learning. Moreover, we also plan to study the discretization in the case in which the OT penalization terms are replaced by the entropic counterpart $\mathsf{OT}^c_\varepsilon(\mu_1, \mu_2)$. Finally we would like to implement this algorithm and experiment it numerically on some applications as the risk estimation problem presented in [IL15]. A second step will consist in testing (and comparing performances) other numerical methods such as the **Sinkhorn** algorithm based on the entropic regularization of Optimal Transport [BCN19; Ben+15; Ben+18] or the ones arising from the approximation of **optimal transport problems with marginal moments constraints** [Alf+21; ACE22]. In particular we are also interested in studying a recent approach based on genetic column generation method proposed in [FP22; FSV22]: convergence for multi-marginal problem is still an open question as well as a possible link with Frank-Wolfe method.

- **Entropic optimal transport and Mean Field Games:** Establish convergence rates for optimizers of the $\varepsilon$-regularized primal problem $\mathsf{OT}^c_\varepsilon$ and its dual as $\varepsilon \to 0$. While convergence (up to subsequence) is easily obtained for compactly supported measures by simple compactness arguments, obtaining quantitative rates is substantially more difficult. The fully discrete case is covered by the general results of [CM94; Wee18] and the semi-discrete case by [ANS22; Del22]. We propose to focus on the continuous case, which is not so well understood, and establish sharp convergence rates for the primal problem (e.g. in terms of Wasserstein distance) and for the dual problem, with assumptions on the cost ($\mathcal{C}^2$ regularity, twist condition in particular) and on the marginals but as few assumptions as possible on the unregularized optimizers (in contrast to what is usually done in the literature). Study the asymptotics as $\varepsilon \to 0$ in more general situations including the multi-marginal case and non-twisted or degenerate costs. A first direction is to generalize the convergence rates of the entropic cost established in [CPT23] to the multi-marginal case with possibly degenerate costs. A second direction, as another development from [CPT23], is to tackle the question of the convergence of entropic plans.

  Furthermore, consider the dual problem to $\mathsf{OT}^c_\varepsilon$, that is

$$\sup_{\varphi_1,\varphi_2} \int \varphi_1(x_1)\mathrm{d}\mu_1 + \int \varphi_2(x_2)\mathrm{d}\mu_2 - \varepsilon \int e^{\frac{\varphi_1(x_1)+\varphi_2(x_2)-c(x_1,x_2)}{\varepsilon}}\mathrm{d}\mu_1\mathrm{d}\mu_2, \tag{1.34}$$

  by optimality condition one can derive the Sinkhorn's algorithm

$$\varphi_i^{n+1}(x_i) = -\log\left(\int e^{\frac{\varphi_j(x_j)-c(x_1,x_2)}{\varepsilon}}\mathrm{d}\mu_j\right), \text{ for } i = 1,2 \text{ and } i \neq j. \tag{1.35}$$

Linear convergence for the multi-marginal case has been recently established in [Car22] however the same result for a generalized Sinkhorn in order to solve variational second order Mean Field Games (see [Ben+18] for more details) is still an open question. In this case we would like to establish it by starting with simple congestion functionals and then relax the assumptions in order to treat more general case. Moreover, we plan to extend the variational formulation to the case of a **multi-population** systems in order to study the dynamic of several populations. We also want to tackle the convergence for the variational formulation of second order mean field games [Ben+18] when the noise vanishes.

Finally, for the entropic transport we have developed an ODE approach for the case of pair-wise cost function, we would like to extend it to a more general cost functions. Notice that in this case one can easily take $c_\eta(x_1,\dots,x_m) = \eta c(x_1,\dots,x_m)$ and having the trivial null solutions as initial condition for the Cauchy problem.

# Chapter 2

# Generalized Optimal Transport in Mathematical Physics

This chapter summarizes the main results obtained for some applications of Optimal Transport in Mathematical Physics. In particular we will focus on (1) a Statistical Mechanics extension of multi-marginal transport, namely the Grand Canonical OT, which allows to deal with the case of a varying number of marginals; (2) an approximation of Lieb's functional which can be seen as a quantum regularization of optimal transport.

All the results are contained in the following papers

Section 2.2 S. Di Marino, M. Lewin, and L. Nenna. *Grand-Canonical Optimal Transport.* 2022. URL: `http://arxiv.org/abs/2201.06859`. preprint

Section 2.3 V. Ehrlacher and L. Nenna. *A Sparse Approximation of the Lieb Functional with Moment Constraints.* 2023. URL: `http://arxiv.org/abs/2306.00806`. preprint

## 2.1 Optimal Transport in Density Functional Theory and beyond

Density Functional Theory (see the seminal paper by Hohenberg and Kohn [HK64]) attempts to describe all the relevant information about a many-body quantum system at ground state in terms of the one electron density $\rho$. Following Levy and Lieb's approach [Lev79; Lie83], the ground state energy can be rephrased as the following variational principle involving only the electron density

$$\mathcal{E}_0[v] = \inf_{\substack{\rho\in\mathcal{A}^m \\ \int_{\mathbb{R}^3} v(x)\, d\rho<+\infty}} \left\{ F_{LL}[\rho] + \int_{\mathbb{R}^3} v(x)\,\mathrm{d}\rho \right\},$$

where $\mathcal{A}^m = \{\rho \in L^1(\mathbb{R}^3) \ : \ \rho \geq 0, \sqrt{\rho} \in H^1, \rho(\mathbb{R}^3) = m\}$ is the set of admissible densities, $v$ is an external potential and the Levy-Lieb functional $F_{LL}$ is defined as

$$\boxed{F_{LL}[\rho] := \inf_{\substack{\Psi\in\mathcal{H}^m_1 \\ \rho_\Psi=\rho}} \frac{1}{2}\int_{\mathbb{R}^{3m}} |\nabla\Psi|^2 + \int_{\mathbb{R}^{3m}} V|\Psi|^2,} \tag{2.1}$$

where

(i) $\mathcal{H}_1^m := \bigwedge_{i=1}^m H^1(\mathbb{R}^3)$ is the set of admissible electronic wavefunctions for a system of $m$ electrons with finite kinetic energy, that is the set of antisymmetric functions of $H^1(\mathbb{R}^{3m})$;

(ii) for any $\Psi \in \mathcal{H}_1^m$ and $x \in \mathbb{R}^3$, $\rho_\Psi$ is the electronic density associated to the wavefunction $\Psi$, namely the real-valued function defined over $\mathbb{R}^3$ as follows:

$$\forall x \in \mathbb{R}^3, \quad \rho_\Psi(x) := m \int_{(\mathbb{R}^3)^{m-1}} |\Psi(x, x_2, \dots, x_m)|^2 \, \mathrm{d}x_2 \dots \mathrm{d}x_m.$$

(iii) the function $V : (\mathbb{R}^3)^m \to \mathbb{R}_+ \cup \{+\infty\}$ is the electron-electron Coulomb interaction potential

$$V(x_1, \dots, x_m) := \sum_{i<j} \frac{1}{|x_i - x_j|}.$$

The Levy-Lieb functional is indeed the lowest possible (kinetic plus interaction) energy of a quantum system having the prescribed density $\rho$. This universal functional is the central object of Density Functional Theory, since knowing it would allow one to compute the ground state energy of a system with any external potential $v$. For a complete review on it we refer the reader to [LLS23]. However, it turns out that $F_{LL}$ is not convex, it is therefore convenient to look at a convexification proposed by Lieb [Lie83] where the minimization is performed over the set of mixed states instead of the set of pure ones as in (2.1). More precisely, we consider here the alternative minimization problem

$$\boxed{F_L[\rho] := \inf_{\substack{\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m) \\ \rho_\Gamma = \rho}} \mathrm{Tr}\,(H_m \Gamma),} \tag{2.2}$$

where $H_m := -\frac{1}{2}\Delta + V$ is a self-adjoint operator on $\mathcal{H}_0^m := \bigwedge_{i=1}^m L^2(\mathbb{R}^3)$ with domain $D(H_m) = \mathcal{H}_2^m := \bigwedge_{i=1}^m H^2(\mathbb{R}^3)$, $\mathfrak{S}_1^+(\mathcal{H}_0^m)$ denotes the set of trace-class self-adjoint non-negative operators on $\mathcal{H}_0^m$. For all $\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m)$, there exists an orthonormal basis $(\Psi_i)_{i \in \mathbb{N}^*}$ of $\mathcal{H}_0^m$ and a non-increasing sequence $(\alpha_i)_{i \in \mathbb{N}^*}$ of non-negative numbers such that

$$\Gamma = \sum_{i=1}^{+\infty} \alpha_i |\Psi_i\rangle\langle\Psi_i|, \tag{2.3}$$

using so-called bra-ket notation. Then, the associated electronic density $\rho_\Gamma$ is defined as follows: for all $x \in \mathbb{R}^3$,

$$\rho_\Gamma(x) := m \sum_{i=1}^{+\infty} \alpha_i \int_{(\mathbb{R}^3)^{m-1}} |\Psi_i(x, x_2, \dots, x_m)|^2 \, \mathrm{d}x_2 \dots \mathrm{d}x_m = \sum_{i=1}^{+\infty} \alpha_i \rho_{\Psi_i}(x).$$

We know that there exist positive constants $\varepsilon, D > 0$ such that $H_m + D \geq \varepsilon(-\Delta + \mathrm{Id})$ (in the sense of self- adjoint operators on $\mathcal{H}_0^m$). We also denote by $\mathfrak{S}_{1,1}(\mathcal{H}_0^m)$ the set of self-adjoint operators $\Gamma$ on $\mathcal{H}_0^m$ with finite kinetic energy, i.e. such that $\mathrm{Tr}\left(|H_m + D|^{1/2} \Gamma |H_m + D|^{1/2}\right) < +\infty$.

**Remark 2.1.1.** *It can then be easily checked that,* $\Gamma \in \mathfrak{S}_{1,1}(\mathcal{H}_0^m)$ *if and only if* $\Gamma \in \mathfrak{S}_1(\mathcal{H}_0^m)$ *and* $\mathrm{Tr}(H_m \Gamma) < +\infty$. *Then, if* $\Gamma$ *admits an eigendecomposition of the form (2.3), necessarily* $\Psi_i \in \mathcal{H}_1^m$ *as soon as* $\alpha_i > 0$.

It is well-known, see for instance [Lie83], that the infimum in (2.1) and (2.2) is attained.

**Remark 2.1.2** (Convexification). *It is worth highlighting that $F_L$ is indeed the convexification of $F_{LL}$ in the sense that*

$$F_L[\rho] = \inf_{\substack{\forall i\geq 1,\ \alpha_i\geq 0,\ \rho_i\in\mathcal{A}^m \\ \sum_{i=1}^{+\infty}\alpha_i=1 \\ \sum_{i=1}^{+\infty}\alpha_i\rho_i=\rho}} \sum_{i=1}^{+\infty}\alpha_i F_{LL}[\rho_i]$$

It is useful noticing that $F_L$ admits a dual problem.

**Theorem 2.1.3** ([Lie83]). *Duality holds in the sense that*

$$F_L[\rho] = \sup_{\substack{v\in L^{3/2}(\mathbb{R}^3)+L^\infty(\mathbb{R}^3) \\ H_m^v\geq 0}} \left\{ \int_{\mathbb{R}^3} v(x)\rho(x)\mathrm{d}x \right\}, \tag{2.4}$$

*where*

$$H_m^v = H_m - \sum_{i=1}^m v(x_i).$$

The constraint in (2.4) has to be understood in the sense of self-adjoint operators, namely for all $\Psi\in\mathcal{H}_1^m$, $\langle\Psi|H_m^v|\Psi\rangle\geq 0$.

**Remark 2.1.4.** *Notice that* (2.4) *is very similar to the dual problem introduced defined in* (MD).

It is possible to introduce an effective semi-classical parameter $\varepsilon=\hbar^2$ by scaling the density $\rho$. Namely, for $\rho_\varepsilon=\varepsilon^3\rho(\varepsilon x)$ we have

$$\begin{aligned} \frac{F_L[\rho_\varepsilon]}{\varepsilon} &= \inf_{\substack{\Gamma\in\mathfrak{S}_1^+(\mathcal{H}_0^m) \\ \rho_\Gamma=\rho}} \mathrm{Tr}\left(-\frac{\varepsilon}{2}\Delta+V\right)\Gamma, \\ \frac{F_{LL}[\rho_\varepsilon]}{\varepsilon} &= \inf_{\substack{\Psi\in\mathcal{H}_1^m \\ \rho_\Psi=\rho}} \frac{\varepsilon}{2}\int_{\mathbb{R}^{3m}}|\nabla\Psi|^2+\int_{\mathbb{R}^{3m}}V|\Psi|^2 \end{aligned} \tag{2.5}$$

In the limit $\varepsilon\to 0$, the above functionals converge (see [Lew18] for Lieb functional and [CFK13; CFK18; BD17] for the Lévy-Lieb ones) to the multi-marginal optimal transport problem with Coulomb cost

$$\boxed{\inf_{\gamma\in\Pi(\rho)}\int_{\mathbb{R}^{3m}}\sum_{i<j}\frac{1}{|x_i-x_j|}\mathrm{d}\gamma(x_1,\dots,x_m)} \tag{2.6}$$

where now the marginals are all equal to the given electron density $\rho$.

**Remark 2.1.5.** *Notice that now the mass of $\rho$ is not equal to* 1 *anymore even if by a slight abuse of notation we will still define the OT problems on the set of probability measures. However the fact that the mass is equal to the number of particles/electrons $m$ of the system does not affect the minimizers but only the value of the energy.*

In Theoretical Chemistry literature the 0−limit regime is also know as the Strictly Correlated Electrons (SCE) state and it has been firstly introduce in [SGS07] and it is used in order to approximate the electron-electron repulsion term in Lévy-Lieb/Lieb functionals. The literature about the SCE approximation is growing considerably. Recent developments include results on the existence and non-existence of Monge-type solutions (e.g.,[CDD15; CS16; CFK13; BDG12; BDK21; CD15]), structural properties of Kantorovich potentials (e.g., [CDS19; BCD18]), and design of efficient computational algorithms (e.g., [BCN16; Nen16; KY19; Alf+21; Kho+20; ACE22; FP22; FSV22; FP23]) .

In this chapter we mainly focus on two different generalization of multi-marginal optimal transport:

- **Grand Canonical Optimal Transport** This model takes its roots in Statistical Physics, where it usually goes under the name *Grand-Canonical* [Rue99]. It has recently been used for Coulomb and Riesz costs where it naturally occurs in the large-$m$ limit of the multi-marginal problem [LLS19; LLS18; LLS23; CP19]. A truncated version has been studied in [Bou+21a; Bou+21b]. Its entropic regularization is well known in the literature [CCL84; CC84; JKT23] and plays a central role in the density functional theory of inhomogeneous classical fluids at positive temperature [Eva79]. Here we will define both Grand Canonical OT and entropic grand canonical OT, study the existence and characterization of solutions as well as develop the duality theory.
- **Quantum Optimal Transport** We provide a sparse approximation of the Lieb functional (2.2) by using some techniques originally developed to solve multi-marginal optimal transport [Alf+21]. In particular we can look at this functional as a (linear) regularization of multi-marginal OT in the quantum framework.

## 2.2 Grand Canonical Optimal Transport

The Grand-Canonical Optimal Transport (GC-OT) problem can be formulated in the Kantorovich form as follows

$$\boxed{\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) := \inf_{\gamma\in\Pi_{\mathrm{GC}}} \sum_{n=0}^{\infty} \int_{X^n} c_n \,\mathrm{d}\gamma_n,} \tag{2.7}$$

where

$$\Pi_{\mathrm{GC}}(\rho) := \Big\{ \gamma = (\gamma_n)_{n\geq 0} \;:\; \gamma_0 \in [0,1], \quad \gamma_n \in \mathcal{M}_{\mathrm{sym}}(X^n) \\ \gamma_0 + \sum_{n\geq 1} \gamma_n(X^n) = 1, \quad \rho_\gamma = \rho \Big\}, \tag{2.8}$$

where $\mathcal{M}_{\mathrm{sym}}(X^n)$ denotes the set of symmetric positive measures on $X^n$, $c_n$ is the symmetric cost for the $n$-marginal problem and $\rho_{\gamma_n}$ is the first marginal of $\gamma_n$. Notice the factor $n$ multiplying the marginal in the constraint involving $\rho$, which accounts for the fact that there are $n$ equal such marginals since all the $\gamma_n$ are symmetric. The family $\gamma = (\gamma_n)_{n\geq 0}$ forms a probability which describes the behavior of some agents whose number is unknown or can vary. In this interpretation $\gamma_n(X^n)$ is the probability that there are $n$ agents and $\gamma_0 \in [0,1]$ is the one that there is no agent at all. In the GC-OT problem (2.7) only the *average* quantity $\rho$ is fixed and fluctuations of the number of agents are allowed. Solving the minimization problem $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ requires in particular to determine the best way to distribute the number of agents through the measures $\gamma_n$, in order to reproduce the given average $\rho$, depending on the corresponding costs $\boldsymbol{c} = (c_n)_{n\geq 0}$.

In order to motivate the problem, let us first explain why it is natural to let the number of marginals vary, even if we start with a system which has a well defined number of agents

$m$. Consider a symmetric Borel probability measure $\mathscr{P}$ over $X^m$, where $X \subset \mathbb{R}^n$ is any given Borel and $m \geq 2$ is typically a large number. In applications, $\mathscr{P}$ describes $d$ properties of $m$ agents, taking their values in $X$ (e.g. their space location along the $d$ coordinates in $\mathbb{R}^d$), with $\mathscr{P}(A_1 \times \cdots \times A_m)$ being the probability that agent 1 is in $A_1$, agent 2 is in $A_2$, etc. The symmetry of $\mathscr{P}$ simply means that our agents are all identical and indistinguishable from one another. When $m$ is very large it seems natural to allow it to vary a little, for instance to account for the fact that it can probably not be known exactly. However, grand-canonical states also occur naturally at fixed $m$ when we look at *subsystems*. Let us explain this now. Let us fix a subset $A \subset X$. Imagine that we would like to ignore what is happening outside of $A$ and only concentrate on what the agents in $A$ are doing. Of course, although there are exactly $m$ agents in total, the number of agents in the subset $A$ can vary from 0 to $m$. A simple counting argument leads us to introducing the measures $\gamma_n$ on $A^n$ defined by

$$\begin{cases} \gamma_0 = \mathscr{P}\left((X \setminus A)^m\right) & \text{for } n = 0, \\ \gamma_n(B_n) = \binom{m}{n} \mathscr{P}\Big(B_n \times (X \setminus A)^{m-n}\Big) & \text{for } 1 \leq n \leq m-1, \\ \gamma_m(B_m) = \mathscr{P}\left(B_m\right) & \text{for } n = m, \\ 0 & \text{for } n \geq m+1, \end{cases} \tag{2.9}$$

for any $B_n \subset A^n$. In other words, the conditional measure $\gamma_n$ describes what $n$ agents are doing in $A$, in the situation that there are exactly $n$ agents in $A$ and $m - n$ outside. The combinatorial factor in the definition of $\gamma_n$ is because in the expression it was assumed, by symmetry, that the $n$ first are in $A$ whereas the $m - n$ last ones are outside. Then $\gamma_n(A^n)$ is the probability that there are exactly $n$ agents in $A$ and thus

$$\gamma_0 + \sum_{n=1}^{m} \gamma_n(A^n) = 1.$$

On the other hand, the average number of agents in $A$ is given by

$$\sum_{n=1}^{m} n\, \gamma_n(A^n) \in [0, m].$$

The corresponding density of agents in $A$ is the positive measure defined by

$$\rho_\gamma(B) := \gamma_1(B) + \sum_{n=2}^{m} n\, \gamma_n(B \times A^{n-1})$$

for every $B \subset A$. It is the sum of the marginals with the multiplication factor $n$ seen before. A calculation shows that

$$\rho_\gamma(B) := m\mathscr{P}(B \times X^{m-1}),$$

that is, $\rho_\gamma$ is just the restriction of $m$ times the first marginal of $\mathscr{P}$ to the set $A$. Our conclusion is that although the total number of agents $m$ can be fixed, it is never fixed as soon as we look at a subsystem, which is always represented by a grand-canonical probability $\gamma$. Here we have $\gamma_n \equiv 0$ for $n \geq m + 1$ because we start with $m$ agents in total. But if we let $m$ be arbitrarily large with $A$ fixed, then $\gamma$ can in principle have infinitely many non-trivial $\gamma_n$'s, as in (2.7).

Note that there is definitely some loss of information when we look at this $\gamma$ instead of the big probability $\mathscr{P}$ since everything which is happening outside of $A$ has been averaged over. Even if we would cover $X$ with several $A$'s and look at the corresponding localized probabilities, we would in general not be able to reconstruct $\mathscr{P}$, since the correlation between the different domains is discarded. We believe that approximating a large multi-marginal problem by a collection of smaller local grand-canonical ones is a strategy which might be helpful in practice.

### 2.2.1 The theory of Grand Canonical Optimal Transport

**Existence**

We start to state some results concerning the existence of minimizers for (2.7) as well as the relation with a classical multi-marginal optimal transport problem. Although one can allow any possible cost $c_n$ for a mathematical exercise, in practice the $c_n$ are often related with one another. A large part of our results is stated by using costs describing pairwise interactions between agents/particles, that is all the $c_n$ are expressed in terms of two-agent cost $c_2$ only.

**Definition 2.2.1** (Pairwise costs). *A* pairwise grand-canonical cost *takes the form*

$$c_0 = c_1 = 0, \qquad c_n(x_1, ..., x_n) = \sum_{1\le j<k\le n} c_2(x_j, x_k) \quad \text{for } n \ge 3. \tag{2.10}$$

The well-posedness of the grand-canonical problem (2.7) requires specific assumptions on $c_n$ to avoid a collapse due to the possibility of having infinitely many agents in the system. The first assumption we need is the following one

**Definition 2.2.2** (Stability). *Let $X \subset \mathbb{R}^d$. We say that the family of symmetric costs $\boldsymbol{c} = (c_n)_{n\ge 0}$ is* stable *whenever there exist two constants $A, B \ge 0$ such that $c_n \ge -A - Bn$ on $X^n$ for all $n \ge 0$.*

Stability implies that $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) > -\infty$ and so our problem is a well defined minimization problem, which is manifestly convex in $\rho$. Although stability is a good condition for $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ to be well defined for all finite measures $\rho$, it is not sufficient to obtain the existence of optimizers. Consider for instance costs which are positive but not large enough for $n \gg 1$, *e.g.*, $c_0 = 0$ and $c_n > 0$ with $||c_n||_{L^\infty(X^n)} = o(n)$. Now using the following trial state

$$\gamma_0 = 1 - \frac{\rho(X)}{m}, \quad \gamma_m = \frac{\rho(X)}{m} \otimes^m \left(\frac{\rho}{\rho(X)}\right), \quad \gamma_n \equiv 0 \quad \forall n \notin \{0, m\} \tag{2.11}$$

we have that $0 < \mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) \le \rho(X)||c_m||_{L^\infty(X^m)}/m \to 0$. Hence $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) = 0$ is never attained for $\rho \ne 0$. This example shows that the cost $c_n$ should be at least as large as $n$ for some configurations to hope to have minimizers. Notice that for a pairwise cost as in the definition above with $c_2(0,0) > 0$ then we have $c_n(0, \dots, 0) = \frac{n(n-1)}{2} c_2(0,0)$ which blows up like $n^2$. The following remedy is the adaptation of another classical concept in statistical mechanics [Rue99].

**Definition 2.2.3** (Super-stability). *We say that the family of costs $\boldsymbol{c} = (c_n)_{n\ge 0}$ is* super-stable *if it is stable and if for any compact set $K \subset \mathbb{R}^d$, there exists $\varepsilon_K > 0$ and $n_K \in \mathbb{N}$ such that*

$$c_n(x_1, \dots, x_n) \ge -\frac{n}{\varepsilon_K} + \varepsilon_K \left(\sum_{j=1}^n \mathbb{1}_{X\cap K}(x_j)\right)^2 \quad \text{on } X^n \text{ for all } n \ge n_K. \tag{2.12}$$

Note that such a cost blows up quadratically in the number of agents in any fixed domain $K$. Then the following existence theorem holds

**Theorem 2.2.4** (Existence of optimizers [DLN22])**.** *Let $X \subset \mathbb{R}^d$ be any Borel set. Let $\boldsymbol{c} = (c_n)_{n\geq 0}$ be a superstable family of lower semi-continuous costs. Then any finite $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ admits a minimizer $\gamma^*$. Moreover $\rho \mapsto \mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ is convex and lower semi-continuous for the tight convergence of measures.*

The proof relies on the standard direct method of calculus of variations. However let us remark that the super-stability plays a central role in showing that the candidate $\gamma^\star$ to be the optimizer has the right average marginal: it is clear that $\rho_{\gamma^*} \leq \rho$ and we also have that for any compact $K$

$$\rho_{\gamma^*}(K) \geq \lim_{k\to+\infty}\left(\rho(K) - \sum_{n\geq M}\rho_{\gamma_n^k}\right),$$

where $\gamma^k$ is the minimizing (sub)sequence. Hence thanks to superstability we get that the tail of the sum in the above equation behaves as $1/\sqrt{M}$. By letting $M \to \infty$ we can conclude that $\rho_{\gamma^*}(K) \geq \rho(K)$ obtaining existence for $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$.

If we decompose $\rho$ into any possible convex combination of densities with integer masses, then as an immediate consequence of the existence theorem we have that the $\mathsf{GCOT}^{\boldsymbol{c}}$ is the the convex hull of multi-marginal optimal transport problem that is

$$\boxed{\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) = \min_{\substack{\rho=\sum_{n\geq 1}\alpha_n\rho_n \\ \rho_n(X)=n \\ \sum_{n\geq 0}\alpha_n=1}} \sum_{n\geq 0}\alpha_n\,\mathsf{MOT}_0^c(\rho_n),} \tag{2.13}$$

where $\mathsf{MOT}_0^c(\rho_n)$ is the standard $n-$marginal problem with cost $c_n$. Under an extra slightly technical hypothesis on the family of cost

$$\lim_{\substack{\min_k |y_k|\to\infty \\ \min_{k\neq\ell}|y_k-y_\ell|\to\infty}} c_{n+m}(x_1,...,x_n,y_1,...,y_m) = c_n(x_1,...,x_n), \tag{2.14}$$

one can also prove that $\mathsf{GCOT}^{\boldsymbol{c}}$ is also the weak closure of $\mathsf{MOT}_0^c$.

**Theorem 2.2.5** (Weak lower semi-continuous envelope [DLN22])**.** *Let $\rho$ be any finite measure so that $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) < +\infty$. Then there exists a sequence $\rho^k$ such that $N_k := \rho^k(\mathbb{R}^d) \in \mathbb{N}$,*

$$\rho^k \rightharpoonup \rho \quad \textit{locally and} \qquad \lim_{k\to\infty}\mathsf{MOT}_0^c(\rho^k) = \mathsf{GCOT}_0^{\boldsymbol{c}}(\rho).$$

The fact that we only have local convergence is due to the possible unboundedness of the number $N_k$ of agents in $\rho^k$. This is because we need infinitely many agents to be able to reproduce a $\gamma = (\gamma_n)_{n\geq 0}$ with $\gamma_n \neq 0$ for arbitrarily large $n$. If we know that there exists a minimizer which satisfies $\gamma_n \equiv 0$ for $n$ large, then the local convergence can be replaced by weak convergence in duality with $\mathcal{C}_0(\mathbb{R}^d)$.

### Truncation of the support

When $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ has a minimizer $\gamma$, a natural question is to ask how many agents are necessary to minimize the given grand-canonical cost $\boldsymbol{c}$, that is, how many of the $\gamma_n$'s are non zero. Let us firstly introduce the following definition of support for a grand canonical probability measure.

**Definition 2.2.6** (Support)**.** *We call*

$$\mathrm{Supp}(\gamma) = \{n \geq 0 \ : \ \gamma_n \neq 0\}$$

*the* support *in $n$ of a grand-canonical probability $\gamma = (\gamma_n)_{n\geq 0}$ and we say that $\gamma$ has a* compact support *whenever* $\mathrm{Supp}(\gamma)$ *is bounded.*

Notice now that in numerical simulations we have to truncate the support of the $\gamma$ in order to make $\mathsf{GCOT}^{\boldsymbol{c}}$ computable that is

$$\mathsf{GCOT}^{\boldsymbol{c},\leq N}(\rho) := \inf_{\substack{\gamma\in\Pi_{\mathrm{GC}}(\rho)\\ \mathrm{Supp}(\gamma)\subset[0,N]}} \sum_{n\geq 0}\int_{X^n} c_n \mathrm{d}\gamma_n. \tag{2.15}$$

Thus, knowing the size of the support of exact minimizers plays a central role in order to suppress or diminish the approximation due to the truncation. Good quantitative estimates are then useful and they will often depend on $\rho$.

**Remark 2.2.7** (Truncated Grand Canonical)**.** *Notice that $\mathsf{GCOT}^{\boldsymbol{c},\leq N}(\rho)$ has been deeply studied in a beautiful paper [Bou+21b] by Bouchitté and coauthors where it was obtained as the weak closure of $\mathsf{MOT}_0^c$ in a similar spirit has in our work. Moreover, in [DLN22][Thm 3.1] we can provide that $\mathsf{GCOT}^{\boldsymbol{c},\leq N}(\rho)$ convergences toward $\mathsf{GCOT}^{\boldsymbol{c}}(\rho)$ when $N \to +\infty$.*

Notice that in our work [DLN22] we discuss several possible conditions on $c_2 > 0$ which imply that the support of minimizers is always compact. There are some overlap between these conditions but we have not found a general simple theory which covers everything. To be more precise, we give a quantitative estimate on the support of minimizers in the following cases

(i) when $X$ is a bounded domain,

(ii) when $1/c_2$ satisfies a triangle-type inequality, which is for instance the case of Riesz interactions $c_2(x, y) = |x - y|^{-s}$ with $s > 0$ in $\mathbb{R}^d$,

(iii) when $c_2$ is asymptotically doubling.

In the Coulomb case $c_2(x, y) = |x-y|^{-1}$ we will show that the length of the support is controlled by $\sqrt{\rho(X)}$ but it can in principle grow with $\rho(X)$. Let us summarise the results concerning the support of minimisers:

- [DLN22][Thm 3.2]: Consider the case where the cost is pairwise and such that $0 < m \leq c_2 \leq M \leq \infty$ and the domain is bounded. Then $\mathsf{GCOT}^{\boldsymbol{c}}(\rho)$ is finite for any non-negative measure $\rho$ and when $\rho(X) > 1$ a minimizer $\gamma^*$ satisfies

  $$\mathrm{Supp}(\gamma^*) \subset \left[\frac{m}{M}\lfloor\rho(X)\rfloor \ , \ 1 + \frac{M}{m}(\lceil\rho(X)\rceil - 1)\right].$$

  When $\rho(x) \leq 1$ then $\mathsf{GCOT}^{\boldsymbol{c}}(\rho) = 0$ and the unique minimizers is of the form $\gamma^* = (1 - \rho(X), \rho, 0, \ldots)$.

- [DLN22][Thm 3.3]: Consider as before a bounded domain and a pairwise cost with $c_2$ lower semi-continuous and such that $1/c_2$ satisfies the triangle-type inequality

  $$\frac{1}{c_2(x,y)} \leq Z\left(\frac{1}{c_2(x,z)} + \frac{1}{c_2(z,y)}\right), \qquad \forall x, y, z \in X,$$

for a some positive constant $Z$ then for any non-negative measures $\rho$ such that $\mathsf{GCOT}^{\mathbf{c}}(\rho) < +\infty$ and $\rho(X) > 1$ a minimizer has a compact support

$$\mathrm{Supp}(\gamma^*) \subset \left[\frac{\lfloor\rho(X)\rfloor + 1}{2Z+1}\,,\,(2Z+1)(\lceil\rho(X)\rceil - 1)\right].$$

The above results is indeed inspired by a celebrated paper by Lieb [Lie84] on the maximum number of electrons that a molecule can bind. Moreover, notice that our theorem can pe applied to the 3D Coulomb potential or more generally to Riesz costs where $c_2(x,y) = |x-y|^{-s}$, $s > 0$ in any dimension. In these cases we have $Z = \max(1, 2^{s-1})$.

- [DLN22][Thm 3.4]: We also consider the case of an aymptotically doubling cost, namely a condition to quantify how fast $c_2$ can converge to zero. For a cost like this there exists a positive constant $C$ such that for some large enough $R$ we have $m(2r) \geq CM(r)$ for all $r \geq R$, where

$$m(r) = \inf\{c_2(x,y) \,:\, |x-y| \leq r\}, \quad M(r) = \sup\{c_2(x,y) \,:\, |x-y| \geq r\}.$$

A typical example is the Riesz class of costs where $C = 2^{-s}$. In this case, if we assume that $\rho(X) > 1$ and it is not too concentrate, that is $\kappa := \sup_{x\in X}\rho(B_r(x)) < 1$ (for instance $\rho$ has no atom). The support of the minimiser is characterized as follows

$$\mathrm{Supp}(\gamma^*) \subset \left[0\,,\,1 + \rho(X)\max\left(\frac{2}{C^2}, \frac{M(r)}{(1-\kappa)m(2R_0)}\right)\right] \tag{2.16}$$

where $R_0$ is the smallest radius such that $\rho(X \setminus B_{R_0}) \leq 1/2$ and $R_0 \geq R$. Notice that the Coulomb cost is indeed asympotically doublig but the support estimate (2.16) is actually less precise then in our previous result.

- **The 3D Coulomb cost:** the case of the Coulomb cost deserves a special attention since it plays a central role in Density Functional Theory and the approximation of Lévy-Lieb/Lieb functionals. First, notice that the bound of theorem [DLN22][3.3] gives

$$\mathrm{Supp}(\gamma^*) \subset \left[\frac{N_0+1}{3}, 3N_0 - 3\right], \quad \rho(X) = N_0 \in \mathbb{N}\setminus 1.$$

A natural question is then to ask whether the length of th support is really of the order $N_0$ or smaller. We remind that the *ionization conjecture* is a celebrated problem in quantum mathematical physics [Sim00; LS10] which states that a nucleus with charge $Z$ can bind at most $m \leq Z + M$ electrons for some $M$. A natural similar conjecture in our case would be that the support of any minimizers for $\mathsf{GCOT}^{\mathbf{c}}(\rho)$ is included in $[\rho(\mathbb{R}^d) - M, \rho(\mathbb{R}^d) + M]$ for a universal constant $M$. This happens to be **true** in dimension $d = 1$ with $M = 1$, but **wrong** in dimension $d \geq 2$. We can however prove that the length of the support is much smaller that the average $\rho(\mathbb{R}^d)$, at most of the order of $\sqrt{\rho(\mathbb{R}^d)}$ in any dimension. In particular [DLN22][Thm 3.5] state that for $X = \mathbb{R}^d$ and $c_2(x,y) = |x-y|^{-1}$ for any non-negative measures such that the grand canonical cost is finite and $\rho(\mathbb{R}^d) > 1$ any minimizers satisfies

$$\mathrm{supp}(\gamma^*) \subset \left[\lfloor\rho(X)\rfloor - \frac{1}{2}\sqrt{8\lfloor\rho(X)\rfloor + 9} + \frac{3}{2}\,,\,\lceil\rho(X)\rceil + \frac{1}{2}\sqrt{8\lceil\rho(X)\rceil - 7} - \frac{1}{2}\right]. \tag{2.17}$$

Moreover, when $\rho(\mathbb{R}^d) = 2$ then $\mathrm{Supp}(\gamma^*) = \{2\}$. Let us remark that this result has been obtained by exploiting a method introduced in [FNV18; FKN16] and an extension of classical *c-monotonicity* to the grand canonical case, see [DLN22][Lemma 3.1]. We end the discussion on the high dimensional case with an example which has support of the length of the order $\rho(X)^\alpha$ with $\alpha \sim 0.38$. Let us consider 6 points $x_1, ..., x_6$ in the plane $\mathbb{R}^2$ and the uniform measure $\rho = \frac{1}{2}\sum_{i=1}^6 \delta_{x_i}$ such that $\mathsf{GCOT}^{\boldsymbol{c}}(\rho)$ admits

$$\gamma_2^* = \frac{1}{2}\delta_{x_1} \otimes_s \delta_{x_2}, \qquad \gamma_4^* = \frac{1}{2}\delta_{x_3} \otimes_s \delta_{x_4} \otimes_s \delta_{x_5} \otimes_s \delta_{x_6} \tag{2.18}$$

as unique minimizer (in [DLN22][section 3.4.2] we explain in detail how this is possible). Then, form these points it is possible to construct inductively a sequence $(y_j^{(k)})_{j=1}^{6^k} \subset \mathbb{R}^d$ so that,

$$\rho^{(k)} = \frac{1}{2}\sum_{j=1}^{6^k} \delta_{y_j^{(k)}}, \qquad \rho^{(k)}(\mathbb{R}^d) = \frac{6^k}{2},$$

the grand-canonical problem $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho^{(k)})$ admits a unique minimizer $\gamma^{(k)}$, which satisfies

$$\mathrm{Supp}(\gamma^{(k)}) = \left\{\frac{6^k - 2^k}{2}, \frac{6^k + 2^k}{2}\right\}.$$

Notice that this example disproves the *ionization-kind conjecture* for high dimensional Coulomb cost.

- **The 1D problem:** as in the classical case of (multi-marginal) optimal transport the one dimensional case happens to be more tractable by exploiting the structure of $\mathbb{R}$. In this case we are able to extend the results of [CDD15] to the grand canonical framework, confirming in particular the shape of the optimal plans considered [MSG13]. We prove, see [DLN22][Thm 5.1], that for $m < \rho(\mathbb{R}) < m + 1$ we have that $\mathrm{Supp}(\gamma) = \{m, m + 1\}$, whereas for $\rho(\mathbb{R}) \in \mathbb{N}$ the grand-canonical optimal solution is actually the canonical one. In addition, we show that the points are *strictly correlated* on the support of the optimal plan, that is, is given by a Monge state. When the mass of $\rho$ is not an integer, only one point is removed in a some region of the support. The running hypothesis is that $\boldsymbol{c} = (c_n)_{n\geq 0}$ is a pairwise cost with $c_2(x, y)$ satisfying

$$c_2(x, y) = w(|x - y|),\ w : \mathbb{R} \to [0, +\infty] \text{ convex and decreasing.} \tag{2.19}$$

Let us just clarify a little bit what a Monge state means in this case. For any atomless $\rho \in \mathcal{M}(\mathbb{R})$ let us consider the unique $m \in \mathbb{N}$ and $\eta \in [0, 1)$ such that $\rho(\mathbb{R}) = m + \eta$. We then split $\mathbb{R}$ into consecutive intervals where $\rho$ has the alternating masses $\eta$ and $1 - \eta$: $x_0 = -\infty \leq x_1 \leq x_2 \leq \ldots \leq x_{2m} \leq x_{2m+1} = +\infty$ where

$$\rho\big((x_{2i}, x_{2i+1})\big) = \eta, \qquad \rho\big((x_{2i+1}, x_{2i+2})\big) = 1 - \eta, \qquad \forall i = 0, \ldots, m - 1.$$

Of course, we automatically obtain $\rho((x_{2n}, x_{2n+1})) = \eta$ too. We now define $T : (x_0, x_{2m-1}) \to (x_2, x_{2m+1})$ as the $\rho$-a.e. unique increasing function such that $\rho((x, T(x))) = 1$. We have $T((x_i, x_{i+1})) \subseteq (x_{i+2}, x_{i+3})$. Iteratively we then define $T_1(x) = T(x)$ and $T_{i+1}(x) = T(T_i(x))$. Next we define $\gamma^*$ by

$$\gamma_i^* = \begin{cases} 0 & \text{for } i \notin \{m, m+1\}, \\ \mathrm{Sym.}\big[(\mathrm{id}, T_1, T_2, \ldots, T_{n-1})_\sharp \rho|_{(x_1, x_2)}\big] & \text{for } i = m, \\ \mathrm{Sym.}\big[(\mathrm{id}, T_1, T_2, \ldots, T_n)_\sharp \rho|_{(x_0, x_1)}\big] & \text{for } i = m+1. \end{cases} \tag{2.20}$$

where Sym. means that we are symmetrizing the plan. We have $\gamma^* \in \Pi_{GC}(\rho)$. If $\rho(\mathbb{R}) \in \mathbb{N}$, then $\gamma^*$ is exactly the canonical optimal plan considered in [CDD15]. Note that $\gamma^*$ does not depend on the cost, it is only a function of the density $\rho$.

### Duality

We conclude this section by studying the dual problem to (2.7). We know that the variables dual to the density $\rho$ are one-agent costs of the form $\mathbf{\Phi} = (\Phi_n)_{n\geq 0}$ with $\Phi_0 = 0$ and $\Phi_n = \sum_{j=1}^n \varphi(x_j)$ for a given $\varphi \in \mathcal{C}_b^0(\mathbb{R}^d)$. We should however not forget the other constraint that $\gamma$ forms a probability, $\gamma_0 + \sum_{n\geq 1} \gamma_n(X^n) = 1$, which requires the introduction of an additional Lagrange multiplier $\beta$. This constraint is independent of the density constraint, on the contrary to the usual multi-marginal problem. This leads us to the following *dual problem*

$$\mathsf{D}(\rho) := \sup \Bigg\{ \beta + \int_X \varphi(x)\,\mathrm{d}\rho(x) \ : \ \beta \leq c_0, \quad \varphi \in \mathcal{C}_b(\mathbb{R}^d), \\ \beta + \sum_{j=1}^n \varphi(x_j) \leq c_n(x_1, ..., x_n), \quad \forall n \geq 1 \Bigg\}. \tag{2.21}$$

If we take any $\beta$ and $\varphi \in \mathcal{C}_b(\mathbb{R}^d)$ satisfying the above constraints and any $\gamma \in \Pi_{\mathrm{GC}}(\rho)$, then we have

$$\langle \gamma, \boldsymbol{c} \rangle := c_0\gamma_0 + \sum_{n\geq 1} \int_{X^n} c_n \,\mathrm{d}\gamma_n \geq \beta + \int_X \varphi(x)\,\mathrm{d}\rho(x)$$

which proves that $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) \geq \mathsf{D}(\rho)$ in all situations. We would like to have equality. It is possible to rewrite the dual problem in a slightly different manner. Let us introduce the grand-canonical ground state energy in the potential $\varphi$

$$E(\varphi) := \inf_{\substack{n\geq 0 \\ x_1,...,x_n \in X^n}} \left( c_n(x_1, ..., x_n) - \sum_{j=1}^n \varphi(x_j) \right) = \inf_\gamma \langle \gamma, \boldsymbol{c} - \Phi \rangle.$$

The last infimum taken over all grand-canonical probabilities $\gamma$ so that $\langle \gamma, \boldsymbol{c} \rangle < \infty$ and $\rho_\gamma(X) < \infty$ (without any other constraint on $\rho_\gamma$). Like in Density Functional Theory [Lie83; LLS23], we rewrite the infimum over $\gamma$ as an infimum over $\rho$ and then an infimum over all $\gamma$ having this density $\rho$:

$$E(\varphi) := \inf_\rho \left\{ \mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) - \int_X \varphi(x)\,\mathrm{d}\rho(x) \right\}$$

with the infimum taken over all finite non-negative measures. We see that $E$ is nothing but the Legendre-Fenchel transform of $\mathsf{GCOT}_0^{\boldsymbol{c}}$. On the other hand, in (2.21) the largest possible $\beta$ at fixed $\varphi$ is indeed equal to $E(\varphi)$, and therefore we can rewrite (2.21) in the form

$$\mathsf{D}(\rho) = \sup_{\varphi \in \mathcal{C}_b(\mathbb{R}^d)} \left\{ \int_X \varphi(x)\,\mathrm{d}\rho(x) + E(\varphi) \right\}. \tag{2.22}$$

Thus, $\mathsf{D}$ is the Legendre-Fenchel transform of $E$. From Theorem 2.2.4 and the Fenchel duality theorem for convex lower semi-continuous functions [Sim11], we conclude the following, which is an extension of a well known result in the multi-marginal case [Kel].

**Theorem 2.2.8** (Duality)**.** *Let $X \subset \mathbb{R}^d$ be any Borel set. Let $\boldsymbol{c} = (c_n)_{n\geq 0}$ be a superstable family of lower semi-continuous costs. Then we have $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) = \mathsf{D}(\rho)$.*

We end up by discussing the more complicated question of the existence of an optimal pair $(\beta, \varphi)$ for the dual problem. As a first step we have to relax a bit the notion of dual potentials and assume that $\varphi$ is in $L^\infty(\mathrm{d}\rho)$ instead of $\mathcal{C}_b$:

$$\widetilde{\mathsf{D}}(\rho) := \sup\Big\{\beta + \int_X \varphi(x)\,\mathrm{d}\rho(x) \ : \ \beta \le c_0, \quad \varphi \in L^\infty(\mathrm{d}\rho), \\ \beta + \sum_{j=1}^n \varphi(x_j) \le c_n(x_1, ..., x_n) \quad \otimes^n \rho\text{–a.e.} \quad \forall n \ge 1\Big\}. \tag{2.23}$$

The following is a rather simple result which provides the existence of the dual pair $(\beta, \varphi)$.

**Theorem 2.2.9** (Existence of a dual potential [DLN22])**.** *Let $X$ be any open set in $\mathbb{R}^d$ and $\rho$ be any finite measure on $X$. Let $\boldsymbol{c} = (c_n)_{n\ge 0}$ be a pariwise superstable family of lower semi-continuous costs such that $c_0 = c_1 = 0$, $c_2 > 0$ and for $n \ge 2$, $\{c_n < \infty\}$ is an open subset of $X$ on which $c_n$ is continuous. Then we have $\mathsf{D}(\rho) = \widetilde{\mathsf{D}}(\rho) = \mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$. If $\mathsf{GCOT}_0^{\boldsymbol{c}}((1+\varepsilon)\rho) < +\infty$ for some $\varepsilon > 0$. Then there exists $(\beta^*, \varphi^*) \in \mathbb{R} \times L^\infty(X, \mathrm{d}\rho)$ which is optimal for $\widetilde{\mathsf{D}}(\rho)$ in* (2.23)*. For any optimizer $\gamma^* = (\gamma_n^*)_{n\ge 0}$ for $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$, we have*

$$\beta^* + \sum_{j=1}^n \varphi^*(x_j) = c_n(x_1, ..., x_n) \quad \gamma_n^*\text{–a.e., for all } n \ge 0. \tag{2.24}$$

**Remark 2.2.10.** *Notice that the original statement of the previous theorem (see [DLN22][Thm 4.2]) we have considered a generic family of costs so that we have to make some extra assumptions*

- $c_1 \in L^\infty(X, \mathrm{d}\rho)$*;*
- **Monotonicity** $c_{n+1}(x_1, ..., x_{n+1}) \ge c_n(x_1, ..., x_n) - A$ *for some $A \in \mathbb{R}$, all $n \ge 0$ and all $x_1, ..., x_{n+1} \in X$.*

*The $\mathsf{GCOT}_0^{\boldsymbol{c}}((1+\varepsilon)\rho) < +\infty$ assumption (strongly inspired by [CCL84]) means that, in some sense, $\rho$ must be in the interior of the convex set of densities for which $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho) < +\infty$.*

**Remark 2.2.11.** *We finally remark that in the special case of Coulomb cost (even if [DLN22][Thm 4.3] holds for a more general class of grand canonical costs) we are also able to prove that the optimal potential $\varphi^*$ for* (2.23) *is Lipschitz with constant depending only on the marginal $\rho$. in the case in which* $\mathrm{Supp}(\rho) = \mathbb{R}^d$ *then $\varphi^* \in \mathcal{C}_b(\mathbb{R}^d)$ and it is also optimal for $\mathsf{D}(\rho)$.*

### 2.2.2 Entropic Grand Canonical Optimal Transport

As we did for the standard multi-marginal optimal transport problem we discuss the entropic regularization of $\mathsf{GCOT}_0^{\boldsymbol{c}}(\rho)$ which, from a statistical mechanics point of view, amounts to adding temperature in the system. In the Physics literature, the problem of finding the external potential for which the associated grand-canonical equilibrium classical (Gibbs) state at a temperature $\varepsilon > 0$ has the given density $\rho$, plays an important role in the density functional theory of classical inhomogeneous systems [Eva79]. Notice that the canonical model is just the entropic regularization of the multi-marginal problem $\mathsf{MOT}_0^{c}(\rho)$, see for instance Section 1.3.

**Remark 2.2.12.** *On the mathematical side, our work relies on the fundamental works of Chayes, Chayes and Lieb [CCL84; CC84] in 1984 who, to our knowledge, were the first to prove the existence of the dual potential $v$ for such systems with entropy, both in the $N$-marginal and grand-canonical cases.*

In order to have a correct behavior it is necessary to place the right 'Boltzmann $n!$' factor in the definition of the entropy, an issue which does not occur for the multi-marginal problem at fixed $n$. The entropy of a grand-canonical probability $\gamma = (\gamma_n)_{n\geq 0}$ is defined by [Rue99; CCL84; CC84]

$$\mathcal{S}(\gamma) := -\sum_{n\geq 0}\int_{X^n}\gamma_n \log\big(n!\gamma_n\big), \tag{2.25}$$

with the condition that each $\gamma_n$ is absolutely continuous with respect to the Lebesgue measure on $X^n$. Otherwise, $\mathcal{S}(\gamma)$ is taken equal to $-\infty$. Note the minus sign in the definition (the usual convention in statistical mechanics). Notice the $n!$ which is due to the fact that our system describes independent agents/particles and so we should in principle not work on $X^n$ but, rather, in the simplex of volume $1/n!$ obtained by moding out the permutations. consider now the following grand canonical relative entropy

$$\mathrm{Ent}(\gamma|\pi_\rho) = \sum_{n\geq 0}\int_{X^n}\gamma_n \log\left(\frac{\gamma_n n! e^{\rho(X)}}{\rho^{\otimes n}}\right),$$

where $\pi_\rho$ is the *Poisson grand-canonical probability* $\pi_\rho = (\pi_{\rho,n})_{n\geq 0} \in \Pi_{\mathrm{GC}}(\rho)$ given by

$$\pi_{\rho,0} = e^{-\rho(X)}, \qquad \pi_{\rho,n} = e^{-\rho(X)}\frac{\otimes^n\rho}{n!}.$$

For any finite non-negative measure $\rho$ on $X$, the grand-canonical entropic optimal transport problem reads

$$\boxed{\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) := \inf_{\gamma\in\Pi_{\mathrm{GC}}(\rho)}\Big\{\langle\gamma,\boldsymbol{c}\rangle + \varepsilon\mathrm{Ent}(\gamma|\pi_\rho)\Big\}.} \tag{2.26}$$

We will assume that the cost $\boldsymbol{c}$ is stable, $c_n \geq -A - Bn$, so that $\langle\gamma,\boldsymbol{c}\rangle \geq -A - B\rho(X)$. The infimum is thus finite for one $\varepsilon > 0$ if and only if there exists a $\gamma \in \Pi_{\mathrm{GC}}(\rho)$ such that $\langle\gamma,\boldsymbol{c}\rangle$ and $\mathrm{Ent}(\gamma|\pi_\rho)$ are simultaneously finite. In this case $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho)$ is actually finite for all $\varepsilon > 0$. Otherwise, we have $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) = +\infty$ for all $\varepsilon > 0$. For instance we can simply assume that the cost $\langle\pi_\rho,\boldsymbol{c}\rangle$ is finite for the Poisson state. Note also that $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho)$ is non-decreasing and concave in $\varepsilon$. The following existence result is a consequence of the known properties of the relative entropy Ent.

**Theorem 2.2.13** (Existence for the positive temperature problem)**.** *Let $X \subset \mathbb{R}^d$ be any Borel set. Let $\boldsymbol{c} = (c_n)_{n\geq 0}$ be a stable family of lower semi-continuous costs. Let $\rho$ be a finite measure such that $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) < \infty$ for one (hence all) $\varepsilon > 0$.*

*(i)* $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho)$ *admits a* unique minimizer $\gamma^{(\varepsilon)}$ *for all $\varepsilon > 0$.*

*(ii) In the limit $\varepsilon \to 0^+$ we have*

$$\lim_{\varepsilon\to 0^+}\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) = \inf_{\substack{\gamma\in\Pi_{\mathrm{GC}}(\rho)\\ \mathrm{Ent}(\gamma|\pi_\rho)<\infty}}\langle\gamma,\boldsymbol{c}\rangle. \tag{2.27}$$

*If $\boldsymbol{c} = (c_n)_{n\geq 0}$ is superstable and the right side is equal to $\mathsf{GCOT}^{\boldsymbol{c}}_0(\rho)$, then $\gamma^{(\varepsilon)}$ converges, up to subsequences, to a minimizer $\gamma^* \in \Pi_{\mathrm{GC}}(\rho)$ for $\mathsf{GCOT}^{\boldsymbol{c}}_0(\rho)$, in the sense that $\gamma^{(\varepsilon)}_n \rightharpoonup \gamma^*_n$ tightly for all $n \geq 0$.*

*(iii) If $\langle \pi_\rho, \boldsymbol{c}\rangle < \infty$, then in the limit $\varepsilon \to \infty$ we have*

$$\lim_{\varepsilon\to\infty} \mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) = \langle \pi_\rho, \boldsymbol{c}\rangle,$$

$$\lim_{\varepsilon\to\infty} \varepsilon \operatorname{Ent}(\gamma^{(\varepsilon)}|\pi_\rho) = \lim_{\varepsilon\to\infty} \sum_{n\geq 0} \int_{X^n} |\gamma_n^{(\varepsilon)} - \pi_{\rho,n}| = 0.$$

We emphasize that for the existence of $\gamma^{(\varepsilon)}$ the cost $\boldsymbol{c} = (c_n)_{n\geq 0}$ is only assumed to be *stable* and not *superstable* as we required at $\varepsilon = 0$ in Theorem 2.2.4. This is because the entropy gives us an additional control on the large-$n$ behavior.

We end the section by discussing the dual problem. Consider a measurable potential $\psi : X \to \mathbb{R}$ such that $\int_X e^{\frac{\psi(x)}{\varepsilon}}\,\mathrm{d}\rho(x) < \infty$. With $\Psi_n = \sum_{j=1}^n \psi(x_j)$ we define

$$\boxed{D_{\varepsilon,\rho}(\psi) := \inf_\gamma \left\{\langle \gamma, \boldsymbol{c} - \Psi\rangle + \varepsilon \operatorname{Ent}(\gamma|\pi_\rho)\right\} = -\varepsilon \log Z_{\varepsilon,\rho}(\psi)} \tag{2.28}$$

with the partition function

$$Z_{\varepsilon,\rho}(\psi) := e^{-\frac{c_0}{\varepsilon} - \rho(X)} + \sum_{n\geq 1} \frac{e^{-\rho(X)}}{n!} \int_{X^n} \exp\left(\frac{-c_n(x_1, ..., x_n) + \sum_{j=1}^n \psi(x_j)}{\varepsilon}\right) \mathrm{d} \otimes^n \rho. \tag{2.29}$$

Recalling that $c_n \geq -A - Bn$ we find

$$Z_{\varepsilon,\rho}(\psi) \leq \exp\left(\frac{A}{\varepsilon} - \rho(X) + e^{\frac{B}{\varepsilon}} \int_X e^{\frac{\psi}{\varepsilon}} \mathrm{d}\rho\right) \tag{2.30}$$

and since $Z_{\varepsilon,\rho}(\psi) \geq e^{-c_0/\varepsilon - \rho(X)}$, we obtain

$$-A + \varepsilon\rho(X) - \varepsilon e^{\frac{B}{\varepsilon}} \int_X e^{\frac{\psi}{\varepsilon}} \mathrm{d}\rho \leq D_{\varepsilon,\rho}(\psi) \leq c_0 + \varepsilon\rho(X). \tag{2.31}$$

The formula $D_{\varepsilon,\rho}(\psi) = -\varepsilon \log Z_{\varepsilon,\rho}(\psi)$ is well known. The corresponding unique minimizer is the Gibbs state $\gamma_\psi$ given by

$$\gamma_{\psi,0} = \frac{e^{-\frac{c_0}{\varepsilon} - \rho(X)}}{Z_{\varepsilon,\rho}(\psi)}, \qquad \gamma_{\psi,n} = \frac{e^{\frac{-c_n + \sum_{j=1}^n \psi(x_j)}{\varepsilon} - \rho(X)} \rho^{\otimes n}}{Z_{\varepsilon,\rho}(\psi) n!}. \tag{2.32}$$

To prove this claim, just notice that

$$\langle \gamma, \boldsymbol{c}\rangle + \varepsilon \operatorname{Ent}(\gamma|\pi_\rho) - \langle \gamma^*, \boldsymbol{c}\rangle - \varepsilon \operatorname{Ent}(\gamma^*|\pi_\rho) = \varepsilon \operatorname{Ent}(\gamma|\gamma^*)$$

is positive and vanishes only at $\gamma^*$. Next we discuss the existence of a dual potential. The following is a small adaptation of results proved in [CCL84; CC84].

**Theorem 2.2.14** (Duality at $\varepsilon > 0$ [CCL84; CC84])**.** *Let $X \subset \mathbb{R}^d$ be any Borel set. Let $\boldsymbol{c} = (c_n)_{n\geq 0}$ be a stable family of lower semi-continuous costs with $c_1 \in L^\infty(X, \mathrm{d}\rho)$. Let $\rho$ be a finite measure such that $\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) < \infty$ for one (hence all) $\varepsilon > 0$. Then we have*

$$\mathsf{GCOT}^{\boldsymbol{c}}_\varepsilon(\rho) = \sup_{\int_X e^{\psi/\varepsilon} \mathrm{d}\rho < \infty} \left\{\int_X \psi\, \mathrm{d}\rho + D_{\varepsilon,\rho}(\psi)\right\}. \tag{2.33}$$

*If there exists $\varepsilon > 0$ such that for one (hence all) $\varepsilon > 0$*

$$\mathsf{GCOT}_\varepsilon^{\mathbf{c}}\big((1+\varepsilon)\rho\big) < \infty, \tag{2.34}$$

*then there exists a unique potential $\psi$ realizing the supremum in (2.33) and it is such that $\psi \in L^1(X, \mathrm{d}\rho)$. The corresponding unique minimizer $\gamma^{(\varepsilon)}$ from Theorem 2.2.13 is equal to the Gibbs state $\gamma_\psi$ in (2.32).*

## 2.3 Quantum Optimal Transport and Lieb Functional

In this section we focus on a moment constraints approximation of the Lieb Functional. Recall that Lieb functional takes the following form

$$\boxed{F_L[\rho] := \inf_{\substack{\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m) \\ \rho_\Gamma = \rho}} \mathrm{Tr}\,(H_m \Gamma),} \tag{2.35}$$

where $H_m := -\frac{1}{2}\Delta + V$ is a self-adjoint operator on $\mathcal{H}_0^m$ with domain $D(H_m) = \mathcal{H}_2^m := \bigwedge_{i=1}^m H^2(\mathbb{R}^3)$, $\mathfrak{S}_1^+(\mathcal{H}_0^m)$ denotes the set of trace-class self-adjoint non-negative operators on $\mathcal{H}_0^m$. Actually, problem (2.35) is a particular instance of *quantum optimal transport* problem. We refer the reader to [GMP16; CGP21] for references on earlier works on closely related types of problems. The aim of the joint work with V. Ehrlacher [EN23] is to prove that solutions of approximations of problems (2.35) where the partial trace constraint is relaxed by a finite number of moment constraints enjoy similar sparsity properties than solutions of moment constrained multi-marginal symmetric classical optimal transport problems, such as those which were established in [Alf+21]. More precisely:

- we prove by using a generalization of Tchakaloff's theorem, that the solutions of moment constrained approximations of (2.35) can be written under the form $\Gamma = \sum_{k=1}^K \alpha_k |\Psi_k\rangle\langle\Psi_k|$, where $K \in \mathbb{N}^*$ scales at most linearly with the number of moment constraints, and where for all $1 \le k \le K$, $\alpha_k \in [0,1]$, $\Psi_k \in \mathcal{H}_1^m$ and $|\Psi_k\rangle\langle\Psi_k|$ is the orthogonal projector of $\mathcal{H}_0^m$ onto the vectorial space spanned by $\Psi_k$ (using bra-ket notation);

- under appropriate assumptions on the set of moment functions we show that the value of the moment constraints approximation functional converges to the value of the exact Lieb functional as the number of moments go to infinity;

- we study the mathematical properties of the associated dual problem;

- we obtain some rates of convergence on the associated approximation of the ground state energy.

Let us finally mention here that particular moment-constrained approximations of the Lieb functional have already been considered in [Gar22] for the construction of Kohn-Sham potentials. The novel results brought by our contribution in comparison to this previous work is (i) the extension of existence and convergence results to more general moment constraints that the one considered in [Gar22]; (ii) the results on the sparsity structure of associated minimizers; (iii) convergence rate of the approximate ground state energy; and (iv) study of the mathematical properties of the associated dual problem.

**Remark 2.3.1** (Classical DFT)**.** *Classical DFT is typically used at interfaces between liquid-gas, liquid-liquid (in fluid mixtures), crystal-liquid and crystal-gas phases at bulk coexistence. In this framework the free energy for $m$ particles takes the following form*

$$\boxed{\mathcal{F}_\varepsilon(\rho) = \inf_{\gamma \in \Pi(\rho)} \int_{\mathbb{R}^{dm}} \sum_{i<j} w(x_i - x_j)\mathrm{d}\gamma + \varepsilon \int_{\mathbb{R}^{dm}} \gamma \log(\gamma),}$$

*which is exactly the entropic multi-marginal optimal transport problem for a cost function given by the interaction potential $\sum_{i<j} w(x_i - x_j)$.*

### 2.3.1 Moment approximation of Lieb functional: the primal problem

Recall that $\mathcal{A}^m = \{\rho \in L^1(\mathbb{R}^3) \ : \ \rho \geq 0, \sqrt{\rho} \in H^1, \rho(\mathbb{R}^3) = m\}$, we fix an electronic density $\rho \in \mathcal{A}^m$. We have $\mathcal{F} := L^{3/2}(\mathbb{R}^3) + L^\infty(\mathbb{R}^3) \subset L^1_\rho(\mathbb{R}^3)$ and for any $f \in \mathcal{F}$, we denote by

$$\|f\|_{\mathcal{F}} := \inf_{\substack{f_{3/2} \in L^{3/2}(\mathbb{R}^3),\ f_\infty \in L^\infty(\mathbb{R}^3), \\ f_{3/2} + f_\infty = f}} \|f_{3/2}\|_{L^{3/2}(\mathbb{R}^3)} + \|f_\infty\|_{L^\infty(\mathbb{R}^3)}.$$

Let $M \in \mathbb{N}^*$, given a collection of $M$ functions $\Phi := (\varphi_1, \ldots, \varphi_M) \in \mathcal{F}^M$, the main idea of the moment-constrained approximation consists in replacing the density constraint in (2.35) with the $M$ scalar moment constraints associated to the functions $\varphi_1, \ldots, \varphi_M$, that is

$$\int_{\mathbb{R}^3} \varphi_n \rho_\Gamma = \int_{\mathbb{R}^3} \varphi_n \rho, \quad \forall n = 1, \ldots, M. \tag{2.36}$$

Notice that (2.36) is equivalent to

$$\int_{\mathbb{R}^3} \varphi \rho_\Gamma = \int_{\mathbb{R}^3} \varphi \rho, \quad \forall \varphi \in \mathrm{Span}\{\Phi\}. \tag{2.37}$$

We denote by $\mathfrak{S}_1^+(\mathcal{H}_0^m, \Phi, \rho)$ the set of $\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m)$ satisfying constraints (2.36) (or equivalently (2.37)). For any Hilbert space $\mathcal{H}$, we denote by $\Sigma_{\mathcal{H}}$ the Borel $\sigma$-algebra of $\mathcal{H}$. and we have the following version of Tchakaloff's theorem.

**Proposition 2.3.2** (Tchakaloff)**.** *Let $\mu$ be a Borelian measure on a Hilbert space $\mathcal{H}$ concentrated on a Borel set $\mathcal{A} \in \Sigma_{\mathcal{H}}$. Let $J_0 \in \mathbb{N}^*$ and $\Lambda : \mathcal{H} \to \mathbb{R}^{J_0}$ a Borel measurable map. Assume that the first moments of $\Lambda_\sharp \mu$ exists, that is*

$$\int_{\mathbb{R}^{J_0}} \|x\| \mathrm{d}\Lambda_\sharp \mu(x) = \int_{\mathcal{H}} \|\Lambda(\Psi)\| \mathrm{d}\mu(\Psi) < +\infty,$$

*where $\|\cdot\|$ denotes the Euclidean norm of $\mathbb{R}^{J_0}$. Then there exists an integer $1 \leq K \leq J_0$, elements $\Psi_1, \cdots, \Psi_K \in \mathcal{A}$ and weights $\omega_1, \cdots, \omega_K > 0$ such that*

$$\forall j = 1, \cdots, J_0, \ \int_{\mathcal{H}} \Lambda_j(\Psi) \mathrm{d}\mu(\Psi) = \sum_{k=1}^K \omega_k \Lambda_j(\Psi_k) = \int_{\mathcal{H}} \Lambda_i(\Psi)\, \mathrm{d}\mu_d(\Psi),$$

*where $\Lambda_j$ is the $j-th$ component of $\Lambda$, and $\mu_d = \displaystyle\sum_{k=1}^K \omega_k \delta_{\Psi_k}$.*

The main idea of the proof of the sparsity result announced above is to define a measure associated to an operator $\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m)$. Assume that the operator $\Gamma$ can be written as

$$\Gamma = \sum_{i=1}^{+\infty} \alpha_i |\Psi_i\rangle\langle\Psi_i| \tag{2.38}$$

for some sequence $(\Psi_i)_{i\in\mathbb{N}^*}$ of normalized functions of $\mathcal{H}_0^m$ and non-negative real numbers $(\alpha_i)_{i\in\mathbb{N}^*}$ such that $\sum_{i\in\mathbb{N}^*} \alpha_i = m$. Then we can define a Borelian measure $\mu_\Gamma : \Sigma_{\mathcal{H}_0^m} \to \mathbb{R}_+$ associated to the decomposition (2.38) of the operator $\Gamma$ as

$$\mu_\Gamma = \sum_{i=1}^{+\infty} \alpha_i \delta_{\Psi_i}.$$

Naturally, there is no unique such measure $\mu_\Gamma$ associated with an operator $\Gamma$ since it heavily depends on the decomposition (2.38). We denote by $\mathbb{1}$ the function defined over $\mathbb{R}^3$ which is identically equal to 1. Then our main result concerning existence for the approximate problem reads as follows.

**Theorem 2.3.3** (Existence for the primal, [EN23]). *Let $\rho \in \mathcal{A}^m$, $M \in \mathbb{N}^*$ and $\Phi := (\varphi_1, \dots, \varphi_M) \in \mathcal{F}^M$ such that $\mathbb{1} \in \mathrm{Span}\{\Phi\}$. Let us assume in addition that*

*(A$\theta$) there exists a non-negative non-decreasing continuous function $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\theta(r) \underset{r\to+\infty}{\longrightarrow} +\infty$ and $C_\rho := \int_{\mathbb{R}^3} \theta(|x|)\rho(x)\,\mathrm{d}x < +\infty$.*

*For all $C > 0$, let us introduce the Moment-Constrained Approximation of the Lieb functional (MCAL)*

$$\boxed{F_{L,\theta}^{\Phi,C}[\rho] := \inf_{\substack{\Gamma\in\mathfrak{S}_1^+(\mathcal{H}_0^m,\Phi,\rho)\\ \mathrm{Tr}\,(\Theta\Gamma)\le C}} \mathrm{Tr}\,(H_m\Gamma),} \tag{MCAL}$$

*where $\Theta(x_1, \dots, x_m) := \frac{1}{m}\sum_{i=1}^m \theta(|x_i|)$ for all $x_1, \dots, x_m \in \mathbb{R}^3$. Then, for all $C \ge C_\rho$, $F_{L,\theta}^{\Phi,C}[\rho]$ is finite and a minimum. Moreover, for all $C \ge C_\rho$, there exists a minimizer $\Gamma_{\mathrm{opt},\theta}^{\Phi,C}$ to* (MCAL) *such that $\Gamma_{\mathrm{opt},\theta}^{\Phi,C} = \sum_{k=1}^K \omega_k |\Psi_k\rangle\langle\Psi_k|$, for some $1 \le K \le M+1$, with $\omega_k \ge 0$ and $\Psi_k \in \mathcal{H}_1^m$ for all $1 \le k \le K$.*

Notice that assumption (A$\theta$) is needed in order to obtain tightness and then use a similar argument as in [Lie83]. This is needed because we are considering operators defined on the space $\mathcal{H}_0^m = \bigwedge_{i=1}^m L^2(\mathbb{R}^3)$. Such a technical assumption is not needed in the case when one considers operators acting on functions acting on a finite domain with Dirichlet boundary conditions.

### 2.3.2 Convergence to the exact Lieb functional

Let us denote here by $\mathcal{D}(\mathbb{R}^3)$ the set of $\mathcal{C}^\infty$ real-valued functions defined on $\mathbb{R}^3$ with compact support. More precisely, let $\rho \in \mathcal{A}^m$ such that there exists a function $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying assumption (A$\theta$). Let $C_\rho := \int_{\mathbb{R}^3} \theta(|x|)\rho(x)\,\mathrm{d}x$ and let $C > C_\rho$.

For all $n \in \mathbb{N}^*$, let $M_n \in \mathbb{N}^*$ and $\Phi^n := (\varphi_k^n)_{1\le k\le M_n} \subset \mathcal{F}$ be a sequence of functions belonging to $\mathcal{F}$ and which satisfies $\mathbb{1} \in \mathrm{Span}\{\Phi^n\}$ for all $n \in \mathbb{N}^*$ together with the following density conditions:

(A$\Phi$) for all $f \in \mathcal{D}(\mathbb{R}^3)$,

$$\inf_{g_n \in \mathrm{Span}\{\Phi^n\}} \|f - g_n\|_{\mathcal{F}} \underset{n\to+\infty}{\longrightarrow} 0.$$

**Remark 2.3.4.** *One example of sequence $(\Phi_n)_{n\in\mathbb{N}^*}$ satisfying (A$\Phi$) is the following: for all $n \in \mathbb{N}^*$, let $\Omega_n := (-n,n)^3$ and let $\mathcal{T}_n := \{T_1^n, \dots, T_{N_n}\}$ (with $N_n := \#\mathcal{T}_n$) be a regular conforming triangular mesh of $\Omega_n$, the elements of which have a maximal diameter size $h_n$ such that $h_n \le \frac{1}{n}$. Let $M_n := \#\mathcal{T}_n + 1 = N_n + 1$. Denoting by $\varphi_k^n := \mathbb{1}|_{T_k^n}$ for $1 \le m \le M_n - 1$ and by $\varphi_{M_n}^n := \mathbb{1}|_{\Omega_n^c}$ and by $\Phi^n = (\varphi_k^n)_{1\le k\le M_n}$ for all $n \in \mathbb{N}^*$, one can easily check that the sequence $(\Phi^n)_{n\in\mathbb{N}^*}$ satisfies (A$\Phi$).*

We then have the following convergence result, which may be seen as an extension of [Gar22][Theorem 3.2] to more general set of moment functions, up to the additional tightness assumption (A$\theta$).

**Theorem 2.3.5** (Convergence,[EN23])**.** *Let $\rho \in \mathcal{A}^m$ such that there exists a function $\theta : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying assumption (A$\theta$). Let $C_\rho := \int_{\mathbb{R}^3} \theta(|x|)\rho(x)\,dx$ and $C \ge C_\rho$. For all $n \in \mathbb{N}^*$, let $M_n \in \mathbb{N}^*$ and $\Phi^n := (\varphi_k^n)_{1\le k\le M_n} \subset \mathcal{F}$ such that assumption (A$\Phi$) holds. We assume in addition that there exists $n_0 \in \mathbb{N}^*$ such that $\mathbb{1} \in \mathrm{Span}\{\Phi^n\}$ for all $n \ge n_0$. Then, for all $n \ge n_0$, there exists at least one sparse minimizer to (MCAL) with $\Phi = \Phi^n$ in the sense of Theorem 2.3.3. Besides, it holds that*

$$\boxed{\lim_{n\to+\infty} F_{L,\theta}^{\Phi^n,C}[\rho] = F_L[\rho].} \tag{2.39}$$

*Moreover, from any sequence $(\Gamma_n)_{n\ge n_0}$ such that $\Gamma_n$ is a minimizer for* (MCAL) *with $\Phi = \Phi^n$, one can extract a subsequence which strongly converges in $\mathfrak{S}_{1,1}(\mathcal{H}_0^m)$ to $\Gamma_\infty$, where $\Gamma_\infty$ is a minimizer of* (2.2).

Like for the existence theorem 2.3.3, we can state a similar result with less technical assumptions in the case when we consider operators acting on functions defined on a bounded subdomain $X \subset \mathbb{R}^3$ with Dirichlet boundary conditions.

### 2.3.3 Convergence rate of the ground state energy in the bounded domain case

We restrict ourselves to the case of a bounded subdomain $X \subset \mathbb{R}^3$. Let $M \in \mathbb{N}^*$, $\Phi := (\varphi_k)_{1\le k\le M} \subset \mathcal{F}(X)$ be a set of moment functions. For all $v \in \mathcal{F}(X)$, let us introduce the ground state energy associated to the potential $v$:

$$E[v] := \inf_{\Psi\in\mathcal{H}_1^m(X)} \langle\Psi|H_{m,X}^v|\Psi\rangle = \inf_{\Gamma\in\mathfrak{S}_1^+(\mathcal{H}_0^m(X))} \mathrm{Tr}\,(H_{m,X}^v\Gamma),$$

where

$$H_{m,X}^v := H_{m,X} - \sum_{i=1}^m v(x_i).$$

Rewriting the minimization over $\Gamma$ as an external minimization over $\rho \in \mathcal{A}_m(X)$ and then as an internal one over all $\Gamma$ such that $\mathrm{Tr}\,\Gamma = \rho$, it can easily be checked that

$$\boxed{E[v] = \inf_{\rho\in\mathcal{A}_m(X)} \left\{ F_{L,X}[\rho] - \int_X v\,\mathrm{d}\rho \right\},} \tag{2.40}$$

where $F_{L,X}$ is the Lieb functional defined on a bounded domain. Let us also define by

$$\boxed{E^{\Phi}[v] := \inf_{\rho \in \mathcal{A}_m(X)} \left\{ F^{\Phi}_{L,X}[\rho] - \int_X v \, \mathrm{d}\rho \right\},} \tag{2.41}$$

where $F^{\Phi}_{L,X}$ is the functional (MCAL) defined on a bounded domain so that we can drop off the assumption ($A\theta$). Similarly, let us point out that, if $v \in \mathrm{Span}\{\Phi\}$, rewriting the minimization over $\Gamma$ as an external minimization over $\rho \in \mathcal{A}_m(X)$ and then as an internal one over all $\Gamma \in \mathfrak{S}_1^+(\mathcal{H}_0^m(X), \Phi, \rho)$, it holds that

$$E[v] = E^{\Phi}[v], \quad \forall v \in \mathrm{Span}\{\Phi\}.$$

We have then the following approximation result.

**Proposition 2.3.6** ([EN23]). *Let us assume that $v \in L^\infty(X)$ and that $\Phi = (\varphi_k)_{1 \le k \le M} \subset L^\infty(X)$. Then, it holds that*

$$\boxed{|E[v] - E^{\Phi}[v]| \le 2m \min_{w \in \mathrm{Span}\{\Phi\}} \|v - w\|_{L^\infty(X)}.} \tag{2.42}$$

Proposition 2.3.6 then enables to quantify the rate of convergence of $|E[v] - E^{\Phi^n}[v]|$ as $n$ goes to infinity for some particular sequences of moment functions $(\Phi^n)_{n \in \mathbb{N}}$ provided that $v$ is regular enough. As an illustration, we analyze here the rate of convergence of a numerical method inspired from the external dual charge approach recently proposed in [Lel22].

**Corollary 2.3.7** ([EN23]). *Let $l \ge 0$ and $X$ be a bounded regular subdomain of $\mathbb{R}^3$. Let $\mu \in H^{l+1}(X)$ be an external density of charge and define $v \in H_0^1(X) \cap H^{l+3}(X)$ as the unique solution to*

$$\begin{cases} -\Delta v = \mu & \text{in } X, \\ v = 0 & \text{on } \partial X. \end{cases}$$

*Let $(\mathcal{T}_h)_{h>0}$ be a sequence of triangular regular meshes of $X$ such that*

$$h := \max_{K \in \mathcal{T}_h} \mathrm{diam}(K).$$

*Let $k \in \mathbb{N}$ and $P_h^k \subset L^\infty(X)$ be the subspace of continuous $\mathbb{P}_k$ finite element functions associated to the mesh $\mathcal{T}_h$. We denote by $V_{h,k}$ the subspace of $H_0^1(X) \cap H^2(X)$ containing all functions $v_{h,k} \in H_0^1(X) \cap H^2(X)$ solution to*

$$\begin{cases} -\Delta v_{h,k} = \mu_{h,k} & \text{in } X, \\ v_{h,k} = 0 & \text{on } \partial X, \end{cases}$$

*for some $\mu_{h,k} \in P_h^k$. Let $\Phi_{h,k}$ be a basis of $V_{h,k}$. Then, assuming that $l \le k$, there exists a constant $C > 0$ such that for all $h > 0$,*

$$\boxed{|E[v] - E^{\Phi_{h,k}}[v]| \le C m h^{l+1} \|v\|_{H^{l+3}(X)}.}$$

**Remark 2.3.8.** *Denoting by $M_{h,k}$ the dimension of $V_{h,k}$, it holds that $M_{h,k} = \mathcal{O}\left(\frac{k}{h^3}\right)$. As a consequence, the above result implies that the rate of convergence of $E^{\Phi_{h,k}}[v]$ to $E[v]$ decays like $\mathcal{O}\left(\frac{m}{M_{h,k}^{(l+1)/3}}\right)$ where $M_{h,k}$ is the number of moment constraints in the MCAL approximation.*

### 2.3.4 The dual problem

We end this section by discussing duality for the case in which we restrict our-self to a bounded domain $X$ as above. We know that the dual variable associated to the density $\rho \in \mathcal{I}_N(X)$ is a one-body interaction potential of the form $W^v(x_1, \dots, x_m) := \sum_{i=1}^m v(x_i)$ for a given $v \in \mathcal{F}(X)$. We then consider the following natural dual problem

$$\boxed{\mathcal{D}^{\Phi}_{L,X}[\rho] = \sup_{\substack{v \in \mathrm{Span}\{\Phi\}, \\ \forall \Psi \in \mathcal{H}^m_1(X),\ \langle \Psi | H^v_{m,X} | \Psi \rangle \geq 0}} \int_X v \mathrm{d}\rho,} \tag{2.43}$$

where

$$H^v_{m,X} := H_{m,X} - \sum_{i=1}^m v(x_i) = H_{m,X} - W^v.$$

If we take any $v := \sum_{k=1}^M \alpha_k \varphi_k \in \mathrm{Span}\{\Phi\}$ satisfying the above constraints and any $\Gamma \in \mathfrak{S}^+_1(\mathcal{H}^m_0(X), \Phi, \rho)$ then we have

$$\begin{aligned} \mathrm{Tr}\,(H_{m,X}\Gamma) \geq \mathrm{Tr}\,(W^v\Gamma) = \int_X v \mathrm{d}\rho_\Gamma &= \int_X \left( \sum_{k=1}^M \alpha_k \varphi_k \right) \mathrm{d}\rho_\Gamma \\ &\geq \int_X \left( \sum_{k=1}^M \alpha_k \varphi_k \right) \mathrm{d}\rho = \int_X v \mathrm{d}\rho \end{aligned}$$

which proves that $F^{\Phi}_{L,X}[\rho] \geq \mathcal{D}^{\Phi,C}_{L,X}[\rho]$. We would like to prove that this inequality is actually an equality. Let us introduce the ground state energy associated to the potential $v$:

$$E[v] = \inf_{\Psi \in \mathcal{H}^m_1(X)} \langle \Psi | H^v_m | \Psi \rangle = \inf_{\Gamma \in \mathfrak{S}^+_1(\mathcal{H}^m_0(X))} \mathrm{Tr}\,(H^v_m \Gamma).$$

We rewrite now the minimization over $\Gamma$ as an external minimization over $\rho \in \mathcal{A}^m(X)$ and then as an internal one over all $\Gamma$ in $\mathfrak{S}^+_1(\mathcal{H}^m_0(X), \Phi, \rho)$ (we are considering the ground state for a potential $v \in \mathrm{Span}\{\Phi\}$):

$$E[v] = \inf_{\rho \in \mathcal{A}^m(X)} \left\{ F^{\Phi}_{L,X}[\rho] - \int_X v \mathrm{d}\rho \right\}.$$

Notice that $E$ is nothing but the Legendre-Fenchel transform of $F^{\Phi}_{L,X}[\rho]$. On the other hand, we rewrite (2.43) in the form

$$\mathcal{D}^{\Phi}_{L,X}[\rho] = \sup_{v \in \mathrm{Span}\{\Phi\}} \left\{ \int_X v \mathrm{d}\rho - E[v] \right\}. \tag{2.44}$$

Thus, $\mathcal{D}^{\Phi}_{L,X}[\rho]$ is the Legendre transform of $E$. From lower semi-continuity of $F^{\Phi}_{L,X}$ and Fenchel duality theorem for convex lower semi-continuous functions we conclude the following

**Theorem 2.3.9** (Strong duality, [EN23])**.** *Under the assumptions of Theorem 2.3.3, we have* $F^{\Phi}_{L,X}[\rho] = \mathcal{D}^{\Phi}_{L,X}[\rho]$.

We now have the following result which, taking into account the sparsity result of Theorem 2.3.3, gives a more convenient formulation of $\mathcal{D}^{\Phi}_{L,X}[\rho]$.

**Theorem 2.3.10** (Existence for the dual, [EN23]). *Under the assumptions of Theorem 2.3.3, there exists at least one maximizer to (2.43), and it holds that*

$$\begin{aligned}\mathcal{D}^{\Phi}_{L,X}[\rho] &= \max_{\substack{v\in\mathrm{Span}\{\Phi\},\\ \forall\Psi\in\mathcal{H}^m_1(X),\quad \langle\Psi|H^v_{m,X}|\Psi\rangle\geq 0}} \int_X v\mathrm{d}\rho \\ &= \max_{\substack{v\in\mathrm{Span}\{\Phi\},\\ \forall\Psi\in\mathrm{Span}\{\Psi_1,\dots,\Psi_K\},\quad \langle\Psi|H^v_{m,X}|\Psi\rangle\geq 0}} \int_X v\mathrm{d}\rho,\end{aligned}$$

*where*

$$\Gamma^{\Phi}_{\mathrm{opt},X} = \sum_{k=1}^{K} \omega_k |\Psi_k\rangle\langle\Psi_k|$$

*for some* $1 \leq K \leq M+1$*, with* $\omega_k > 0$ *and* $\Psi_k \in \mathcal{H}^m_1(X)$ *for all* $1 \leq k \leq K$ *is a minimizer of* $F^{\Phi}_{L,X}$.

The proof of existence result for the dual problem exploits the fact that the problem can be re-written as semi-definite positive program, this is clear by looking at the constraint $\forall\Psi \in \mathrm{Span}\{\Psi_1,\dots,\Psi_K\}$, $\langle\Psi|H^v_{m,X}|\Psi\rangle \geq 0$, and then one can use classical existence results for this kind of problem.

## 2.4 Perspectives

Let us mention in this section some research perspectives:

- **Numerical methods for repulsive MMOT and Grand canonical OT:** As Explained throughout all this chapter the MMOT problem gives a rigorous lower bound of the electron-electron repulsion energy. However numerical simulations to compute the solution to MMOT cannot be afforded only for a very large number of electrons/marginals [BCN16; FP22; Alf+21] as need in computational chemistry. This implies that algorithms for Coulombic optimal transport are still unsuitable to be plugged in self-consistent field methods of computational chemistry. The idea consists in developing efficient numerical methods based on (1) the particle discretization proposed in the previous chapter (2) or on reduced order model approximation and to adopt them, as sub-routine, in order to compute the ground state of molecules and atoms. Moreover, we plan to generalize these numerical methods to the case of grand canonical optimal transport. Notice that the Grand Canonical OT allows to decompose a molecule in sub-domain (with a small number of electrons/marginals in order to **bypass the curse of dimensionality**) and compute the electronic configuration in each of them. We are interested in understanding better the computational error when one tries to approximate the total electron-electron repulsion with a sum of grand canonical optimal transport problems computed on each sub-molecules. Furthermore, it would be interesting establishing a way to decompose a molecule as already initiated in [Can+04; Gal+05].

- **Moments approximation of Lieb functional:** This sparsity structure we have proven in [EN23] leads, naturally to propose a numerical scheme in order to approximate the solution of the moments Lieb functional. The main idea is to develop an iterative scheme which shares some common features with the GenCol algorithm [FP22; FSV22], in the sense that, at each iteration, the "support" of the minimizer is adapted using the solution of an associated dual problem.

# Chapter 3

# Unequal dimensional Optimal Transport

This chapter is devoted to the case in which we consider an Optimal Transport problem between two probability measures on space having different dimension. In particular we are going to summarize the results contained in the following papers:

Section 3.1 L. Nenna and B. Pass. "Variational Problems Involving Unequal Dimensional Optimal Transport". In: *Journal de Mathématiques Pures et Appliquées* 139 (2020), pp. 83–108

Section 3.2 L. Nenna and B. Pass. "Transport Type Metrics on the Space of Probability Measures Involving Singular Base Measures". In: *Applied Mathematics & Optimization* 87.2 (2023), p. 28

Before detailing the above papers let us briefly recall some basics facts of Optimal Transport involving unequal dimensional measures.

Given a bounded and continuous cost function $c : \overline{X}_1 \times \overline{X}_2 \to \mathbb{R}$ and two measures $\mu_1 \in \mathcal{P}(\overline{X}_1)$ and $\mu_2 \in \mathcal{P}(\overline{X}_2)$ defined on the closure of open bounded domains $X_1 \subseteq \mathbb{R}^{d_1}$ and $X_2 \subseteq \mathbb{R}^{d_2}$, then the optimal transport problem

$$\boxed{\mathsf{OT}^c(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_1,\mu_2)} \int_{X_1 \times X_2} c(x_1, x_2) \mathrm{d}\gamma(x_1, x_2)} \tag{3.1}$$

always admits at least one solution. We will assume throughout this paper that $c \in C^2(\overline{X}_1 \times \overline{X}_2)$ satisfies the *twist* condition, which asserts that for each $x_1 \in X_1$

$$x_2 \mapsto D_{x_1} c(x_1, x_2) \text{ is injective on } X_2, \tag{3.2}$$

as well as the *non-degeneracy* condition, asserting that the $d_1 \times d_2$ matrix $D^2_{xy} c(x_1, x_2)$ of mixed second order partial derivatives has full rank for each $(x_1, x_2) \in \overline{X}_1 \times \overline{X}_2$ (as $d_1 > d_2$, this means $D^2_{x_1 x_2} c(x_1, x_2)$ has rank $d_2$).

Before discussing the case $d_1 > d_2$, we remind that (3.1) admits a useful dual formulation

$$\mathsf{OT}^c(\mu_1, \mu_2) := \sup_{(u,v) \in L^1(\mathrm{d}\mu_1) \times L^1(\mathrm{d}\mu_2)} \int_{X_1} u(x_1) \mathrm{d}\mu_1(x_1) + \int_{X_2} v(x_2) \mathrm{d}\mu_2(x_2). \tag{3.3}$$

Under mild conditions, there is a unique solution $(u, v)$ to the dual problem, up to the addition $(u, v) \mapsto (u + C, v - C)$ of constants adding to 0, known as the Kantorovich potentials, and these potentials are $c$-concave; that is, they satisfy

$$u(x_1) = v^c(x_1) := \min_{x_2 \in X_2} [c(x_1, x_2) - v(x_2)], \ v(x_2) = u^c(x_2) := \min_{x_1 \in X} [c(x_1, x_2) - u(x_1)],$$

This solution is used to define generalized nestedness. The following definition is slightly adapted from [MP20].

**Definition 3.0.1.** *When $d_1 > d_2$, we will say that the model $(c, \mu_1, \mu_2)$ satisfies the generalized nestedness condition if for $\mu_2$ almost every $x_2$ the potential $v$ is differentiable and we have*

$$\begin{aligned} \partial^c v(x_2) &:= \{x_1 : u(x_1) + v(x_2) = c(x_1, x_2)\} \\ &= X_=(x_2, Dv(x_2)) := \{x_1 : Dv(x_2) = D_y c(x_1, x_2)\}. \end{aligned} \tag{3.4}$$

The containment $\partial^c v(x_2) \subseteq X_=(x_2, Dv(x_2))$ holds automatically throughout the domain of $Dv(x_2)$; it is therefore the opposite containment that distinguishes nested from non-nested models. The origin of the term nestedness lies in the one dimensional target setting, in which case the condition is equivalent to nestedness of certain *super-level* sets of $x_1 \mapsto \frac{\partial c}{\partial x_2}(x_1, x_2)$ [CMP17]. This is discussed in more detail below (see Proposition 3.0.2).

If both $\mu_2$ and $\mu_1$ are absolutely continuous, the potential $v$ satisfies the Monge-Ampere type equation almost everywhere [MP20][1]:

$$\overline{\mu}_2(x_2) = \int_{\partial^c v(x_2)} \frac{\det(D^2_{x_2 x_2} c(x_1, x_2) - D^2 v(x_2))}{\sqrt{|\det(D^2_{x_2 x_1} c D^2_{x_1 x_2} c)(x_1, x_2)|}} \overline{\mu}_1(x_1) \mathrm{d}\mathcal{H}^{d_1 - d_2}(x_1), \tag{3.5}$$

where $\overline{\mu}_1(x_1) := \frac{d\mu_1}{dx_1}(x_1)$ and $\overline{\mu}_2(x_2) := \frac{d\mu_2}{dx_2}(x_2)$ are the densities of $\mu_1$ and $\mu_2$. In general, this is a *non-local* differential equation for $v(x_2)$, since the domain of integration $\partial^c v(x_2)$ is defined using the values of $v$ and $u = v^c$ throughout $X_2$; however, when the model satisfies the generalized nestedness condition (namely $\partial^c v(x_2) = X_=(x_2, Dv(x_2))$), it reduces to the local equation [MP20]:

$$\overline{\mu}_2(x_2) = \int_{X_=(x_2, Dv(x_2))} \frac{\det(D^2_{x_2 x_2} c(x_1, x_2) - D^2 v(x_2))}{\sqrt{|\det(D^2_{yx} c D^2_{xy} c)(x_1, x_2)|}} \overline{\mu}_1(x_1) \mathrm{d}\mathcal{H}^{d_1 - d_2}(x_1). \tag{3.6}$$

**Multi-to one-dimensional optimal transport** We consider now the optimal transport problem in the case in which $d_1 > d_2 = 1$ (for more details we refer the reader to [CMP17]). In this case, generalized nestedness follows from a relatively simple condition, related to the following heuristic attempt to construct solutions to (3.1). Let us define the level and super-level sets of $D_y c$ as follows

$$X_=(x_2, k) := \{x_1 \in X_1 \ : \ \frac{\partial c}{\partial x_2}(x_1, x_2) = k\},$$

$$X_\geq(x_2, k) := \{x_1 \in X_1 \ : \ \frac{\partial c}{\partial x_2}(x_1, x_2) \geq k\},$$

[1] Note that, here and below, our notation differs somewhat from [CMP17] and [MP20], since we have adopted the convention of minimizing, rather than maximizing, in (3.1).

as well as the strict variant $X_>(x_2,k) := X_\geq(x_2,k) \setminus X_=(x_2,k)$. In order to build an optimal transport map $T$, we take the unique level set splitting the mass proportionately with $x_2$; that is, defining $k(x_2)$ such that

$$\mu_1(X_\geq(x_2,k(x_2))) = \mu_2((-\infty,x_2]), \tag{3.7}$$

then we set $x_2 = T(x_1)$ for all $x_1$ which belong to $X_=(x_2,k(x_2))$. Notice that if there exists $x_1 \in X_1$ such that $x_1 \in X_=(x_2,k(x_2)) \cap X_=(x_2',k(x_2'))$ then the map $T$ is not well-defined. The absence of such a degenerate case is equivalent to the super-level sets being nested; this is the definition of nestedness from [CMP17], which implies the more general Definition 3.0.1 when $d_2 = 1$, as the following result from [MP20] affirms.

**Proposition 3.0.2** (Nestedness for one dimensional targets)**.** *The model $(c,\mu_1,\mu_2)$ satisfies the generalized nestedness condition if*

$$\forall x_2, x_2' \text{ with } x_2' > x_2,\ \mu_2([x_2,x_2']) > 0 \Longrightarrow X_\geq(x_2,k(x_2)) \subseteq X_>(x_2',k(x_2')) \tag{3.8}$$

We will say $(c,\mu_1,\mu_2)$ is *nested* if (3.8) is satisfied.

If the model $(c,\mu_1,\mu_2)$ is nested then [CMP17][Theorem 4] assures that $\gamma_T = (\mathrm{id},T)_\sharp\mu_1$, where the map $T$ is built as above, is the unique minimizer of (3.1) in $\Pi(\mu_1,\mu_2)$. Moreover, the optimal potential $v(x_2)$ is given by $v(x_2) = \int_{-\infty}^{x_2} k(t)dt$, and so (3.6) becomes

$$\overline{\mu}_2(x_2) = \int_{X_=(x_2,k(x_2))} \frac{D^2_{x_2x_2}c(x_1,x_2) - k'(x_2)}{|D^2_{x_1x_2}c(x_1,x_2)|}\overline{\mu}_1(x_1)\mathrm{d}\mathcal{H}^{d_1-1}(x_1). \tag{3.9}$$

## 3.1 Some variational problems involving Unequal dimensional Optimal Transport

This Section is devoted to the study of functionals of the form

$$\boxed{\mathcal{J}(\mu_1,\mu_1) = \mathsf{OT}^c(\mu_1,\mu_2) + \mathcal{F}(\mu_2) + \mathcal{G}(\mu_1)} \tag{3.10}$$

depending on probability measures $\mu_1$ and $\mu_2$. We are interested in characterizing the minimizers of $(\mu_1,\mu_2) \mapsto \mathcal{J}(\mu_1,\mu_2)$, as well as the minimizers of the subproblems obtained when either $\mu_1$ or $\mu_2$ is fixed: $\mu_2 \mapsto \mathcal{J}(\mu_1,\mu_2)$ and $\mu_1 \mapsto \mathcal{J}(\mu_1,\mu_2)$. Problems of these general forms, for various choices of the functionals $\mathcal{F}$ and $\mathcal{G}$, arise in a wide variety of applications, including: gradient flows on Wasserstein space (where the minimization $\mu_2 \mapsto \mathcal{J}(\mu_1,\mu_2)$ represents one step in a discrete gradient flow), displacement interpolation (when $\mathcal{F}$ is the Wasserstein distance to a second probability measure), Cournot-Nash equilibria in game theory, city planning problems, hedonic pricing in economics [BC14a; BC16; BC14b; BS05; AC11; CE10], and have consequently received a fair bit of attention in the literature. Most of the analytical progress so far, however, has been restricted to the case where the dimensions of the spaces $X_1$ and $X_2$ coincide, $d_1 = d_2$. In a wide variety of applications, particularly in economics and game theory, however, these dimensions may differ: in the Cournot-Nash problem, for example, $X_1$ parameterizes a space of agents, (differentiated by $d_1$ characteristics of an agent $x$) while $X_2$ represents a space of strategies. The dimensions of these spaces reflect the number of characteristics used to differentiate among agents, and the number of parameters involved in the choice of strategy, respectively, and need not be the same in general.

Minimizers of (3.10) when $d_1 = d_2$ have been studied extensively; for the subproblems where one marginal is fixed, under mild conditions on the functionals $\mathcal{F}$ and $\mathcal{G}$, existence and uniqueness of minimizers has been established, and, depending on the precise forms of $\mathcal{F}$ and $\mathcal{G}$, various regularity results and bounds on solutions exist. Our purpose is to study the minimizers of $\mathcal{J}$ when $d_1 > d_2$. As mentioned above when the target $X_2$ is unidimensional ($d_1 > d_2 = 1$), the *nestedness* condition allows one to construct almost closed form solutions from the cost function $c$ and marginals $\mu_1$ and $\mu_2$. For higher dimensional targets, ($d_1 > d_2 > 1$), an analogous condition ensures that a certain, generally non-local, partial differential equation on the lower dimensional space $X_2$ characterizing the solution is, in fact, local and degenerate elliptic [MP20]. Notice that the nestedness conditions are *joint* conditions on the cost $c$ and marginals $\mu_1$ and $\mu_2$, whereas in the present context, only the cost and **one** of the marginals (**neither** in the case of double minimizations) is prescribed; the other marginal is part of the solution to the problem. In [NP20] we prove here that, under various conditions on $c$, $\mu_1$ and $X_2$, $(c, \mu_1, \mu_2)$ is nested whenever $\mu_2$ minimizes $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$, for a variety of different choices of the functional $\mathcal{F}$; analogous results for certain specific forms of $\mathcal{G}$ are also established for minimizations on the higher dimensional space, $\mu_1 \mapsto \mathcal{J}(\mu_1, \mu_2)$, and for double minimizations $(\mu_1, \mu_2) \mapsto \mathcal{J}(\mu_1, \mu_2)$. We go on to demonstrate that this a priori guarantee of nestedness makes the problem of characterizing or identifying the minimizers much more tractable; in different contexts, depending on the precise form of $\mathcal{F}$, we establish that solutions can be characterized by (local) differential equations, can be computed numerically by a convergent iterative scheme, or can be derived in almost closed form. The main results we establish in [NP20] are the following:

- **Characterization of nestedness:** Theorem 3.1.1 characterizes nestedness for a given model $(c, \mu_1, \mu_2)$ when the target is one dimensional ($d_2 = 1$). Its consequences include Corollaries 3.1.2 and 3.1.4, which give sufficient conditions for nestedness in terms of either a lower bound on $\mu_2$, which depends on $c$ and $\mu_1$, or an upper bound on $\mu_1$, depending on $c$ and $\mu_2$.

- **Nestedness for the congestion case:** Theorem 3.1.7 ensures nestedness of the model when the lower dimensional measure $\mu_2$ minimizes $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$, $\mathcal{F}$ is a congestion term and the target is one dimensional;

- **Nestedness for the interaction case:** Theorem 3.1.10 provides sufficient conditions for generalized nestedness when the lower dimensional measure $\mu_2$ minimizes $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$ (with the target dimension $n$ not necessarily one) and the functional $\mathcal{F}$ is composed of interaction and potential terms;

- **Best reply scheme:** Theorem 3.1.13 generalizes a previous result in [BC14b], providing convergence of the best reply scheme to compute minimizers of $\mu_2 \mapsto \mathcal{J}(\mu_1, \mu_2)$ when $\mathcal{F}$ is made of interaction and potential terms;

- **The high dimensional congestion case** Theorem 3.1.9 proves nestedness when the higher dimensional measure $\mu_1$ minimizes $\mu_1 \mapsto \mathcal{J}(\mu_1, \mu_2)$, $\mathcal{G}$ is a congestion functional and the target is one dimensional ($n = 1$);

- **The double minimization:** Theorem 3.1.16 provides conditions guaranteeing nestedness for minimizers $(\mu_1, \mu_2)$ of the double minimization problem $(\mu_1, \mu_2) \mapsto \mathcal{J}(\mu_1, \mu_2)$, when $\mathcal{F}$ is composed of interaction and potential terms, $\mathcal{G}$ is of congestion type, and the target is one dimensional;

- **The hedonic pricing and Wasserstein barycenter:** Theorem 3.1.19 considers the case in which the functional $\mathcal{F}(\nu) := \mathsf{OT}^{c_2}(\mu_2, \nu)$ reflects the cost of optimal transport to a second fixed measure, as is the case in the hedonic pricing problem in economics, so that, after relabeling $\mu_1$ and $c$ from (3.10) as $\mu_1$ and $c_1$, the goal is to minimize $\nu \mapsto \mathsf{OT}^{c_1}(\mu_1, \nu) + \mathsf{OT}^{c_2}(\mu_2, \nu)$. When the target $\nu$ is one dimensional, the theorem implies that under a condition we call *hedonic nestedness* one can find solutions $\nu$ to this problem almost explicitly. Among other consequences, this condition ensures nestedness of $(c_i, \mu_i, \nu)$ for both $i = 1$ and 2.

### 3.1.1 A sufficient condition for nestedness

Fix $x_2 < x_2'$ (where $x_2, x_2' \in Y$), $k_0 \in D_{x_2}c(X, x_2)$ and set

$$k_{max}(x_2, x_2', k_0) = \sup\{k : X_{\geq}(x_2, k_0) \subseteq X_{\geq}(x_2', k)\}.$$

We then define the *minimal mass difference*, $D^{min}_{\mu_1}$, as follows:

$$D^{min}_{\mu_1}(x_2, x_2', k_0) = \mu_1(X_{\geq}(x_2', k_{max}(x_2, x_2', k_0)) \setminus X_{\geq}(x_2, k_0)).$$

The minimal mass difference represents the smallest amount of mass that can lie between $x_2$ and $x_2'$, and still have the corresponding level curves $X_=(x_2, k_0)$ and $X_=(x_2', k_1)$ not intersect. In the following we assume that $\mathrm{d}\mu_1(x) = \overline{\mu}_1(x_1)\mathrm{d}x_1$ and $\mathrm{d}\mu_2(x_2) = \overline{\mu}_2(x_2)\mathrm{d}x_2$ are absolutely continuous with respect to the Lebesgue measure.

**Theorem 3.1.1** ([NP20])**.** *Assume that $\mu_1$ and $\mu_2$ are absolutely continuous with respect to Lebesgue measure. If $D^{min}_{\mu_1}(x_2, x_2', k(x_2)) < \mu_2([x_2, x_2'])$ for all $x_2 < x_2'$ where $k$ is defined by (3.7), then $(c, \mu_1, \mu_2)$ is nested. Conversely, if $(c, \mu_1, \mu_2)$ is nested, we must have $D^{min}_{\mu_1}(x_2, x_2', k(x_2)) \leq \mu_2([x_2, x_2'])$ for all $x_2' > x_2$.*

The proof, essentially, follows from the definition of nestedness (3.8). The following sufficient condition is more convenient when one has bounds on the density $\overline{\mu}_2(x_2)$.

**Corollary 3.1.2** ([NP20])**.** *If for each $x_2 \in X_2$, we have*

$$\boxed{\sup_{x_2' \in X_2, x_2 \leq y \leq x_2'} \left[ \frac{D^{min}_{\mu_1}(x_2, x_2', k(x_2))}{x_2' - x_2} - \overline{\mu}_2(y) \right] < 0,}$$

*where $\overline{\mu}_2 = \frac{\mathrm{d}\mu_2(x_2)}{\mathrm{d}x_2}$, then $(c, \mu_1, \mu_2)$ is nested.*

As a consequence, if the quantity $\frac{D^{min}_{\mu_1}(x_2, x_2', k_0)}{x_2' - x_2}$ is bounded above for all $x_2 < x_2'$ and $k_0 \in D_y c(X, x_2)$, then a corresponding lower bound on $\overline{\mu}_2$ will ensure nestedness. We illustrate this with an example.

**Example 3.1.3.** *Letting $\mu_1$ be uniform measure on the quarter disk so that $d\mu_1(x) = \frac{4}{\pi}dx$, $X_2 = (0, \overline{x}_2)$, with $\overline{x}_2 \leq \frac{\pi}{2}$, parametrize a segment of the unit circle, and $c(x_1, x_2) = -(x_1^1 \cos(x_2) + x_1^2 \sin(x_2))$ the bilinear cost, we note that the level curves $X_=(x_2, k)$ are line segments parallel to the line segment joining $(0, 0)$ with $(\cos(x_2), \sin(x_2))$.*

*Now, for fixed $x_2 < x_2', k_0$, the lines $X_=(x_2, k_0)$ and $X_=(x_2', k_{max}(x_2, x_2', k_0))$ intersect on the $x_2$ axis, if $k_0 < 0$ and the $x_1$ axis if $k_0 > 0$. In either case, the region $X_{\geq}(x_2', k_{max}(x_2, x_2', k_0)) \setminus$*

*$X_{\geq}(x_2, k_0)$ is the part of the wedge of angle $x_2' - x_2$ between the two lines which intersects $X_1$; this wedge is smaller than the corresponding wedge $X_{\geq}(x_2', 0) \setminus X_{\geq}(x_2, 0)$, for which the intersection point is at the origin. Therefore,*

$$D^{min}_{\mu_1}(x_2, x_2', k_0) \leq \mu_1(X_{\geq}(x_2', 0) \setminus X_{\geq}(x_2, 0)) = (x_2' - x_2)\frac{2}{\pi}$$

*It therefore follows from Corollary 3.1.2 that the model $(c, \mu_1, \mu_2)$ is nested for any $\mathrm{d}\mu_2 = \overline{\mu}_2(x_2)\mathrm{d}x_2$ such that*

$$\overline{\mu}_2(x_2) > \frac{2}{\pi}. \tag{3.11}$$

When considering problems where the measure $\mu_2$ on $\mathbb{R}$ is fixed but the high dimensional marginal $\mu_1$ is not, the following reformulation is sometimes convenient; it implies that an appropriate upper bound on $\overline{\mu}_1$ yields nestedness.

For $k_0 \in D_{x_2}c(X_1, x_2)$, set

$$S^{min}_{\mu_1}(x_2, x_2', k_0) := X_{\geq}(x_2', k_{max}(x_2, x_2', k_0)) \setminus X_{\geq}(x_2, k_0).$$

We then have:

**Corollary 3.1.4** ([NP20]). *Suppose that for all $x_2 \in X_2$*

$$\boxed{\sup_{x_2' \in X_2, x_2 \leq y \leq x_2'} [||\overline{\mu}_1||_{L^\infty(S^{min}_{\mu_1}(x_2, x_2', k(x_2)))} \frac{D^{\min}_{vol}(x_2, x_2', k(x_2))}{x_2' - x_2} - \overline{\mu}_2(x_2)] < 0.}$$

*Then $(c, \mu_1, \mu_2)$ is nested.*

We now turn our focus to minimizing functionals of the form (3.10), and the subproblems obtained when one of the measures is fixed:

$$\boxed{\min_{\mu_2 \in \mathcal{P}(\overline{X}_2)} \mathsf{OT}^c(\mu_1, \mu_2) + \mathcal{F}(\mu_2),} \tag{3.12}$$

and

$$\boxed{\min_{\mu_1 \in \mathcal{P}(\overline{X}_1)} \mathsf{OT}^c(\mu_1, \mu_2) + \mathcal{G}(\mu_1).} \tag{3.13}$$

### 3.1.2 The congestion case

#### Bounds on the density

Let us consider (3.12) where $\mathcal{F}$ is of *congestion*, or internal energy, form: $\mathcal{F}(\mu_2) := \int_Y f(\overline{\mu}_2(x_2))\mathrm{d}x_2$ with $f : [0, \infty) \to \mathbb{R}$ continuously differentiable on $(0, \infty)$, strictly convex with superlinear growth at infinity, satisfying,

$$\lim_{\overline{\mu}_2 \to 0^+} f'(\overline{\mu}_2) = -\infty.$$

A prototypical example is the entropy, $f(\overline{\mu}_2) = \overline{\mu}_2 \ln(\overline{\mu}_2)$. This is a popular type of functional in a variety of settings; in the Cournot-Nash case, it reflects agents' desires to choose strategies that are not too close to each other. Notice that the assumptions on $\mathcal{F}$ guarantees the existence of a minimizer of (3.12).

In the proposition below, we let

$$M_c := \sup_{(x_1,x_2,x_2')\in(\overline{X}_1\times\overline{X}_2\times\overline{X}_2)} \frac{|c(x_1,x_2)-c(x_1,x_2')|}{|(x_1,x_2)-(x_1,x_2')|},$$

where $|\cdot|$ denotes the Euclidean norm, be a global Lipschitz constant for $x_2 \mapsto c(x_1,x_2)$ and for a bounded real valued function $v : X_2 \to \mathbb{R}$, denote by $K_v : (0,\infty) \to (-\infty,\infty)$ the inverse of the monotone increasing function $z \mapsto \int_Y (f')^{-1}(z - v(x_2))\mathrm{d}x_2 : (-\infty,\infty) \to (0,\infty)$.

For simplicity, we assume below that $0 \in \overline{X}_2$. The argument below is inspired by [San15, Section 7.4.1].

**Proposition 3.1.5** ([NP20])**.** *The minimizing $\mu_2$ in* (3.12) *is absolutely continuous with respect to Lebesgue, and its density $\overline{\mu}_2$ satisfies*

$$(f')^{-1}(K_{-M_c|x_2|}(1) - M_c|x_2|) \leq \overline{\mu}_2(x_2) \leq (f')^{-1}(K_{M_c|x_2|}(1) + M_c|x_2|)$$

We remark that the bounds on the density $\overline{\mu}_2$ we have established above do not depend on the dimensions of $X_1$ and $X_2$. The lower bound is most relevant in (3.12) (since it is the lower bound on $\overline{\mu}_2$ that implies nestedness in the multi-to one-dimensional optimal transport problem, via Corollary 3.1.2); however, since the dimensions play no role, the result also applies to minimizers $\mu_1$ in problem (3.13), in which case the upper bound can be used to prove nestedness via Corollary 3.1.4 (again with $d_2 = 1$).

Before developing these applications, we illustrate how the result above can be used to find an explicit bound in an example.

**Example 3.1.6.** *Recall the quarter disk to arc problem from Example 3.1.3: $\mu_1$ is uniform on the quarter disk, so that $\mathrm{d}\mu_1(x_1) = \frac{4}{\pi}\mathrm{d}x_1$, $Y = (0,\overline{x}_2)$, and the cost $c(x_1,x_2) = -\langle x_1, (\cos x_2, \sin x_2)\rangle$. We take $\mathcal{F}(\mu_2) = \int_X \overline{\mu}_2 \ln(\overline{\mu}_2)\mathrm{d}x_2$ so that up to a non-vital constant $f'(\lambda) = \ln(\lambda)$. We get that $K_v$ is the inverse of $z \mapsto \int_0^{\overline{x}_2} \exp(z - v(x_2))\mathrm{d}x_2 = \exp(z)\int_0^{\overline{x}_2}\exp(-v(x_2))\mathrm{d}x_2$; that is, $K_v(1) = \ln\left([\int_0^{\overline{x}_2}\exp(-v(x_2))\mathrm{d}x_2]^{-1}\right)$. Noting that $M_c = 1$, we have, for any minimizer $\mu_2$ of* (3.12)*,*

$$\overline{\mu}_2(x_2) \geq \exp\left(\ln\left([\int_0^{\overline{x}_2}\exp(x_2)\mathrm{d}x_2]^{-1}\right) - x_2\right) = \frac{e^{-x_2}}{e^{\overline{x}_2}-1}. \tag{3.14}$$

### The one dimensional target

We now turn our attention to proving that minimizers in (3.12) are nested. We begin by considering one dimensional targets and congestion (or internal energy) forms for $\mathcal{F}$. We consider (3.12), when the dimension of $X_2$ is $d_2 = 1$, and $\mathcal{F}[\mu_2] = \int_{X_2} f(\overline{\mu}_2)\mathrm{d}x_2$, and $f$ satisfies the conditions above.

Combining Corollary 3.1.2 with the lower bound on the target density from Proposition 3.1.5 yields the following.

**Theorem 3.1.7** ([NP20])**.** *Suppose that $\mu_2$ minimizes* (3.12) *over $\mathcal{P}(\overline{X}_2)$, where $X_2 = (0,\overline{x}_2)$ and $X_1 \subseteq \mathbb{R}^{d_1}$. Then $(c,\mu_1,\mu_2)$ is nested provided*

$$\boxed{\sup_{x_2'\in X_2, x_2\leq y\leq x_2'} \frac{D^{min}_{\mu_1}(x_2,x_2',k(x_2))}{x_2'-x_2} - (f')^{-1}(K_{-M_c y}(1) - M_c y) < 0}$$

*for all $x_2 \in X_2$.*

In particular, we note the following consequence for our example matching the quarter circle to an arc with the bilinear cost.

**Corollary 3.1.8** ([NP20]). *Suppose that $\mu_2$ minimizes (3.12) over $\mathcal{P}(\overline{X}_2)$, where $\mu_1$ is uniform measure on the quarter disk, $X_2 = (0, \overline{x}_2)$, $F[\mu_2] = \int_{X_2} \overline{\mu}_2 \ln(\overline{\mu}_2) \mathrm{d}x_2$ is the entropy and $c(x_1, x_2) = -\langle x_1, (\cos x_2, \sin x_2)\rangle$ . Then the model $(c, \mu_1, \mu_2)$ is nested provided $\overline{x}_2 \leq \ln\left(\frac{1+\sqrt{(1+2\pi)}}{2}\right) \approxeq 0.61$.*

Once we have determined a priori that the solution must be nested, we can characterize it by a differential equation on the lower dimensional space (ie, an ordinary differential equation). Note that, for a nested model, setting $k(x_2) = v'(x_2)$, one has, by (3.9),

$$\overline{\mu}_2(x_2) = \int_{X_=(x_2,k(x_2))} \frac{D^2_{x_2x_2}c(x_1,x_2) - k'(x_2)}{|D^2_{xy}c(x_1,x_2)|} \overline{\mu}_1(x_1) \mathrm{d}\mathcal{H}^{d_1-1} := G(x_2, k(x_2), k'(x_2))$$

and so differentiating the first order condition $v(x_2) + f'(\overline{\mu}_2(x_2)) = C$ , we get the following second order differential equation for $k$:

$$\boxed{k(x_2) + f''(G(x_2, k(x_2), k'(x_2))) \frac{\mathrm{d}}{\mathrm{d}x_2}[G(x_2, k(x_2), k'(x_2))] = 0.} \tag{3.15}$$

In addition, we can derive boundary conditions for (3.15): since we know the solution is nested, we have $\lim_{x_2 \to 0^+} \mu_1(X_\geq(x_2, k(x_2))) = \lim_{x_2 \to 0^+} \mu_2(0, x_2) = 0$, so that $k(0) := \lim_{x_2 \to 0^+} k(x_2)$ exists, we have $\mu_1(X_\geq(0, k(0))) = 0$ and $X_=(0, k(0))$ is tangent to $\partial X$. This, and a similar argument as $x_2$ approaches $\overline{x}_2$ suggests that we impose the boundary conditions:

$$k(0) = \max_{x_1 \in \overline{X}_1} \frac{\partial c}{\partial x_2}(x_1, 0), \; k(\overline{x}_2) = \min_{x_1 \in \overline{X}_1} \frac{\partial c}{\partial x_2}(x_1, \overline{x}_2). \tag{3.16}$$

Under the conditions of Corollary 3.1.2, any minimizer of (3.12) gives a solution to the above boundary value problem. Conversely, it is not hard to see that given a solution $k(x_2)$ to (3.15) with boundary conditions (3.16), then $\overline{\mu}_2(x_2) = G(x_2, k(x_2), k'(x_2))$ is a minimizer provided that $v(x_2) = \int_0^{x_2} k(y)\mathrm{d}y$ is $c$-concave.

### The high dimensional source

We now consider problems where the high dimensional measure $\mu_1$ is allowed to vary, restricting to a fixed one dimensional target setting, $d_2 = 1$. Consider fixing $\mu_2$ and minimizing $\mu_1 \mapsto \mathsf{OT}^c(\mu_1, \mu_2) + \mathcal{G}(\mu_1)$, where $\mathcal{G}(\mu_1) = \int_{X_1} g(\overline{\mu}_1(x_1))\mathrm{d}x_1$ is a congestion type functional, with $g$ satisfying the conditions on $f$ we have established in the above sub-section the domains $X_1 \subseteq \mathbb{R}^{d_1}$ and $X_2 \subseteq \mathbb{R}$ is one dimensional. Combining Corollary 3.1.4 and Proposition 3.1.5, we immediately obtain the following.

**Theorem 3.1.9** ([NP20]). *Assume that for all $x_2 < y < x_2'$, and $x_1 \in X_\geq(x_2', k_{max}(x_2, x_2', k(x_2))) \setminus X_\geq(x_2, k(x_2))$ we have*

$$\boxed{(g')^{-1}(K_{M_c|x_1|}(1) + M_c|x_1|) < \frac{\overline{\mu}_2(y)(x_2' - x_2)}{D^{min}_{vol}(x_2, x_2', k(x_2))}.}$$

*where $M_c$ is a Lipschitz constant for $x_2 \mapsto c(x_1, x_2)$. Then the model $(c, \mu_1, \mu_2)$ is nested for any minimizer $\mu_1$.*

Note that the equality $\overline{\mu}_1(x_1) = (g')^{-1}(C - u(x_1))$, and the general fact that the potential $u(x_1)$ is Lipschitz implies that the optimal marginal $\overline{\mu}_1(x_1)$ is Lipschitz as well. This allows one to use Theorem 7.1 in [CMP17] to obtain interior $\mathcal{C}^{2,1}$ estimates on $v = u^c$. It is not clear to us whether this can be bootstrapped to obtain higher regularity.

### 3.1.3 The interaction case and fixed point characterization

We now go back to the minimization over the low dimensional marginal and consider the case where $\mathcal{F}$ consists of interaction and potential terms; that is, suppose that $\mu_2$ minimizes the following functional on $\mathcal{P}(\overline{X}_2)$:

$$\mu_2 \mapsto \mathsf{OT}^c(\mu_1, \mu_2) + \int_{X_2} V(x_2) \mathrm{d}\mu_2(x_2) + \frac{1}{2} \int_{X_2} \int_{X_2} W(x_2, x_2') \mathrm{d}\mu_2(x_2') \mathrm{d}\mu_2(x_2) \tag{3.17}$$

where $\mu_1$ is a given probability measure on a set $X_1 \subseteq \mathbb{R}^{d_1}$, $X_2 \subseteq \mathbb{R}^{d_2}$ with $d_1 > d_2$ and the interaction term $W(x_2, x_2') = W(x_2', x_2)$ is symmetric. We will denote by $F_{V,W}[\mu_2]$ the first variation of $\mathcal{F}$; that is,

$$F_{V,W}[\mu_2](x_2) := V(x_2) + \int_{X_2} W(x_2, x_2') \mathrm{d}\mu_2(x_2').$$

In this case, we do not generally expect lower bounds on the density $\overline{\mu}_2$, and so the results from the *Bounds on density* Section tell us little about the structure of solutions. However, under certain conditions, we will be able to use the optimality conditions directly to infer generalized nestedness, as we will see below.

Assume throughout this section convexity of $X_2$ and uniform convexity of $y \mapsto c(x, y) + V(y) + W(z, y)$ throughout $X_1 \times X_2 \times X_2$; that is, for all $x, y, z$, we have

$$D^2_{x_2 x_2} c(x_1, x_2) + D^2 V(x_2) + D^2_{x_2 x_2} W(x_2', x_2) \geq C > 0. \tag{3.18}$$

Also assume that for each $x_1 \in X_1$, $x_2' \in X_2$ and $x_2 \in \partial X_2$ that

$$\big\langle D_{x_2} c(x_1, x_2) + DV(x_2) + D_{x_2} W(x_2, x_2'), \mathbf{n}_{X_2}(x_2) \big\rangle \geq 0, \tag{3.19}$$

where $\mathbf{n}_{X_2}$ is the outward unit normal to $X_2$.

**Theorem 3.1.10** ([NP20])**.** *Under the uniform convexity* (3.18) *and outward gradient* (3.19) *assumptions,* $(c, \mu_1, \mu_2)$ *satisfies the generalized nestedness condition for any minimizer* $\mu_2$ *of* (3.17)*. Furthermore, if* $c$*,* $V$ *and* $W$ *are* $\mathcal{C}^k$ *smooth (for any integer* $k \geq 2$*), then the optimal map between* $\mu_1$ *and* $\mu_2$ *is* $\mathcal{C}^{k-1}$*.*

Generalized nestedness of the solution and [MP20] now combine to imply the following result:

**Corollary 3.1.11** ([NP20])**.** *Assume that* $c$ *is twisted and non-degenerate, and adopt the assumptions of Theorem 3.1.10. Then the minimizer* $\mu_2$ *is absolutely continuous and its density satisfies the following integral Monge-Ampere type equation almost everywhere.*

$$\boxed{\overline{\mu}_2(x_2) = \int_{X=(x_2, Dv(x_2))} \frac{\det[D^2_{x_2 x_2} c - D^2 v(x_2)]}{\sqrt{\det[D^2_{x_2 x_1} c D^2_{x_1 x_2} c]}} \overline{\mu}_1(x_1) \mathrm{d}\mathcal{H}^{d_1 - d_2}(x_1).} \tag{3.20}$$

**Remark 3.1.12.** *The integro Monge-Ampere operator appearing in* (3.20) *has regularity controlled by a variety of quantities depending on $c$, $X_1$, $X_2$ and $\mu_1$ (see Theorem 11 in [MP20]). Since the potential $v(x_2) = -V(x_2) - \int_{X_2} W(x_2, x_2')\mathrm{d}\mu_2(x_2')$ is as smooth as $V$ and $W$ on the support of $\mu_2$,* (3.6) *then yields regularity estimates on $\overline{\mu}_2(x_2)$.*

Finally, we note that on the support of $\mu_2$, we can eliminate $v$ from (3.6) to obtain the following partial differential equation for $\overline{\mu}_2(x_2)$:

$$\overline{\mu}_2(x_2) = G(x_2, -DF_{V,W}[\mu_2](x_2), -D^2(F_{V,W}[\mu_2])(x_2)) \tag{3.21}$$

where $F_{V,W}[\mu_2](x_2) = V(x_2) + \int_{X_2} W(x_2, x_2')\mathrm{d}\mu_2(x_2')$ is linear in $\mu_2$ and

$$G(x_2, p, Q) = \int_{X=(x_2,p)} \frac{\det[D^2_{x_2x_2}c(x_1, x_2) - Q]}{\sqrt{\det[D^2_{x_2x_1}c(x_1, x_2)D^2_{x_1x_2}c(x_1, x_2)]}}$$

is the integro Monge-Ampere operator from [MP20].

Two complications, absent in the congestion case, arise here: first, the operator $F_{V,W}$ depends *non-locally* on $\mu_2$, and so the PDE (3.21) is non-local, even though the model satisfies the generalized nestedness condition, which eliminates potential non-locality arising from the integro Monge-Ampere operator $G$ as in [MP20]. Second, we do not know the support of $\mu_2$ in advance, only that it is a connected subset of $X_2$; we therefore cannot impose boundary conditions. These issues are not artefacts of the unequal dimensional setting; they arise in equal dimensional problems as well. Since they seem to make solving the problem via the PDE approach challenging, they serve as good motivation for an iteration scheme, adapted from Blanchet-Carlier [BC14b] and developed below.

### Fixed point characterization

Noting that by differentiating with respect to $x_2$ the optimality condition for (3.17) we obtain

$$D_{x_2}c(x_1, x_2) + DF_{V,W}[\mu_2](x_2) = 0, \tag{3.22}$$

we denote by $B_{\mu_2} : X_1 \to X_2$ the map such that

$$D_{x_2}c(x_1, B_{\mu_2}(x_1)) + DF_{V,W}[\mu_2](B_{\mu_2}(x_1)) = 0, \tag{3.23}$$

which is well defined under conditions (3.19) and (3.18). Then, the scheme introduced in [BC14b] consists in iterating the application defined as

$$\mathcal{B}(\mu_2) := (B_{\mu_2})_\sharp\mu_1. \tag{3.24}$$

The following Theorem establishes the existence of a unique fixed point $\mu_2^\star$ of (3.24) which is a solution to (3.17).

**Theorem 3.1.13.** *(The best reply iteration scheme-unequal dimensional case)[[NP20]] Let $\mu_1 \in \mathcal{P}(X_1)$ and the application $\mathcal{B} : \mathcal{P}(X_2) \to \mathcal{P}(X_2)$ defined in* (3.24)*. Assume that the transport cost $c(x_1, x_2)$ is uniformly convex in $x_2$, that is $D^2_{x_2x_2}c \geq \eta\,\mathrm{id}$ with $\eta > 0$, $D^2_{x_1x_2}c$ has maximal rank and $F_{V,W}[\mu_2]$ satisfies the following hypothesis*

$$D^2F_{V,W}[\mu_2] \geq \lambda\,\mathrm{id} \text{ in } X_2, \lambda > 0; \tag{3.25}$$

$$\mathcal{H}^{d_1-d_2}(B^{-1}_{\mu_2}(x_2)) \leq M\ \forall x_2 \in X_2, M \in \mathbb{R}; \tag{3.26}$$

$$JB_{\mu_2} \geq k > 0 \text{ in } X_1; \tag{3.27}$$

$$\int_{X_2} |DF_{V,W}[\nu_1] - DF_{V,W}[\nu_0]|\mathrm{d}x_2 \leq C\mathcal{W}_1(\nu_1, \nu_0) \tag{3.28}$$

*where $JB_{\mu_2}$ is the $d_2-$dimensional Jacobian of $B_{\mu_2}$ Moreover, let $\mu_1 \in \mathcal{P}(X_1)$ absolutely continuous with respect to Lebesgue and such that $||\mu_1||_\infty MC < k(\eta + \lambda)$ .*
*Then $\mathcal{B}$ is a contraction of $(\mathcal{P}(X_2), \mathcal{W}_1)$, where we denote by $\mathcal{W}_1$ the 1-Wasserstein distance (namely the Optimal Transport problem with the Monge cost) and the unique fixed point $\mu_2^\star$ is solution to* (3.17).

**Remark 3.1.14.** *One can get rid of hypothesis eq.* (3.27) *by noticing that the Jacobian of $B_{\mu_2}$ depends on other quantities: $D^2F_{V,W}[\mu_2](x_2)$, $D^2_{x_2x_2}c(x_1,x_2)$ and $D^2_{x_1x_2}c(x_1,x_2)$.*

**Remark 3.1.15** (The equal dimensional case)**.** *When $d_1 = d_2$, the above proposition is an extension of [BC14b, Theorem 5.1] to the case in which a general cost function is involved.*
*We highlight the map $B_{\mu_2}(x_1)$ is actually not explicit or simple to compute. However in the special case in which $c(x_1,x_2) = h(x_1 - x_2)$, with $h$ is strictly convex, the map $B_{\mu_2}(x)$ takes the form*

$$B_{\mu_2}(x_1) = (\mathrm{id} + Dh^{-1}(-DF_{V,W}[\mu_2]))^{-1}(x_1).$$

*So far we have assumed that the cost is double twisted, but we can avoid this assumption and notice that the co-area formula still holds. In this case we have that*

$$\mathcal{B}(\mu_2)(x_2) = \int_{B^{-1}_{\mu_2}(x_2)} \frac{\mu_1(x_1)}{JB_{\mu_2}} \mathrm{d}\mathcal{H}^0(x_1),$$

*and, since $\mathcal{H}^0$ is simply the counting measure (under the non-degeneracy condition), this implies that* (3.26) *can be interpreted as a bound on the number of points in the pre-image of $B_{\mu_2}$.*

### 3.1.4 The double minimization

Consider now the problem where neither measure is fixed, and where $\mathcal{G}(\mu_1) = \int_{X_1} g(\overline{\mu}_1(x_1))\mathrm{d}x_1$ and $F_{V,W}$ have the forms in subsections 3.1.2 and 3.1.3, respectively. That is, consider the minimization problem

$$\boxed{\inf_{(\mu_1,\mu_2)\in\mathcal{P}(\overline{X}_1)\times\mathcal{P}(\overline{X}_2)} \mathsf{OT}^c(\mu_1,\mu_2) + \mathcal{G}(\mu_1) + \int_{X_2} F_{V,W}[\mu_2]\mathrm{d}\mu_2(x_2).} \tag{3.29}$$

The results established above can be used to prove the following.

**Theorem 3.1.16** ([NP20])**.** *Adopt the assumptions on $X_1, X_2, c, V$ and $W$ from the previous section, and assume that $g$ satisfies the conditions in subsection 3.1.2. Then, whenever $(\mu_1,\mu_2)$ minimizes* (3.29),

1. *$(c, \mu_1, \mu_2)$ is nested.*
2. *$\mu_1$ is absolutely continuous with an everywhere positive density.*
3. *The optimal map $T$ is two degrees less smooth than $c, V$ and $W$, while the Kantorovich potential $u(x_1)$ and density $\overline{\mu}_1(x_1)$ are one degree less smooth than $c, V$ and $W$.*
4. *$\mu_2$ is absolutely continuous.*

### 3.1.5 The hedonic pricing problem

In this section, we study the hedonic pricing problem found in [Eke05] and [CMN10]; economically, this problem involves matching distributions $\mu_1$ and $\mu_2$ of buyers and sellers on spaces $X_1 \subseteq \mathbb{R}^{d_1}$ and $X_2 \subseteq \mathbb{R}^{d_2}$ (both assumed bounded and open), with $m_1, m_2 \geq 1$, according to their preferences for goods in a space $Y$ (which we will assume is one dimensional). Mathematically, this amounts to taking $\mathcal{F}(\nu)$ to be the optimal transport distance to another fixed measure in (3.12) [Eke05][CMN10]. We therefore seek to minimize:

$$\boxed{\min_{\nu \in \mathcal{P}(\overline{Y})} \mathsf{OT}^{c_1}(\mu_1, \nu) + \mathsf{OT}^{c_2}(\mu_2, \nu)}, \tag{3.30}$$

where the $\mu_i \in \mathcal{P}(X_i)$ are absolutely continuous probability measures on the $X_i$ and $Y \subseteq \mathbb{R}$. Each $\mathsf{OT}^{c_i}$ represents the optimal transport distance (3.1) between $\mu_i$ and $\nu$ with respect to a $\mathcal{C}^2$, non-degenerate cost function $c_i(x_i, y)$. We attempt to construct a solution by adapting the construction for the straight optimal transport problem in [CMP17] as follows:

Fix $y$. For each $M \in [0, 1]$, choose the unique $k_i = k_i(y, M)$ such that $\mu_i(X^i_{\geq}(y, k_i)) = M$, where

$$X^i_{\geq}(y, k_i) := \{x_i \in X_i : D_y c_i(x_i, y) \geq k_i\};$$

we adopt similar notation for the level sets $X^i_{=}(y, k_i)$. Now consider the function $M \mapsto k_1(y, M) + k_2(y, M)$. The map is continuous and strictly decreasing.

**Lemma 3.1.17** ([NP20]). *Assume $y$ is in the interior of $Y$ and $y \in \operatorname{argmin}(c_1(x_1, y) + c_2(x_2, y))$ for some $(x_1, x_2) \in X_1 \times X_2$. Then the mapping $M \mapsto k_1(y, M) + k_2(y, M)$ has a unique* 0.

We can then state the following definition of hedonic nestedness.

**Definition 3.1.18** (Hedonic nestedness). *Denote the zero from the preceding Lemma by $M(y)$. We say the problem* (3.30) *is* hedonically nested *if*

$$X^i_{\geq}(y, k_i(y, M(y))) \subseteq X^i_{>}(\overline{y}, k_i(\overline{y}, M(\overline{y}))) \tag{3.31}$$

*for $i = 1, 2$, whenever $y, \overline{y} \in Y$ with $y < \overline{y}$.*

Note that this is equivalent to $M$ being the cumulative distribution function of a probability measure $\nu$ *and* $(c_i, \mu_i, \nu)$ being nested for $i = 1$ and 2.

**Theorem 3.1.19** ([NP20]). *The problem is hedonically nested if and only if $M(y)$ is the cumulative distribution function of some probability measure $\nu$ and $(c_i, \mu_i, \nu)$ is nested for $i = 1, 2$. In this case, $\nu$ is optimal in* (3.30).

**Remark 3.1.20** (Wasserstein barycenter). *Note that a similar construction will hold for the matching for teams problem from [CE10], where one minimizes $\nu \mapsto \sum_{i=1}^{N} \mathsf{OT}^{c_i}(\mu_i, \nu)$ over probability measures on $Y \subseteq \mathbb{R}$. Notice that in the case in which $c_i(x, y) = \lambda_i |x - y|^2$ with $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$, this problem can be read as an unequal dimensional version of the Wasserstein barycenters problem introduced in [AC11].*

We end this section by noting that the nestedness of either one of the $(c_i, \mu_i, \nu)$ (implied, for instance, by hedonic nesting) implies that the solution $\nu$ vanishes at the boundary; in economic terms, this means that neither the lowest nor highest quality goods are exchanged in equilibrium.

**Corollary 3.1.21** ([NP20]). *Suppose $Y = (\underline{y}, \overline{y})$ is an interval. Assume $(c_i, \mu_i, \nu)$ is nested for the optimal $\nu$, for either $i = 1$ or $2$, and that the density $\overline{\mu}_i$ is bounded. Set $\underline{k} = \max_{x \in \overline{X}} D_y c_i(x, \underline{y})$ and $\overline{k} = \min_{x \in \overline{X}} D_y c_i(x, \overline{y})$.*

*If $\lim_{k \to \underline{k}^-} \mathcal{H}^{d_i - 1}(X^i_=(\underline{y}, k)) = 0$, then the optimal density is zero at $\underline{y}$, $\overline{\nu}(\underline{y}) = 0$.*

*Similarly, if $\lim_{k \to \overline{k}^+} \mathcal{H}^{d_i - 1}(X^i_=(\overline{y}, k)) = 0$, then the optimal density is zero at $\overline{y}$, $\overline{\nu}(\overline{y}) = 0$.*

The condition $\lim_{k \to \underline{k}^+} \mathcal{H}^{d_i - 1}(X^i_=(\underline{y}, k)) = 0$ heuristically means that the first level set of $x \mapsto D_y c_i$ that intersects $\overline{X}_i$ does so in a lower dimensional way. Since this level curve is tangent to $\partial X_i$, this is generically true. Note that when $c_i(x, y) = \langle x, \alpha(y) \rangle$ for some curve $\alpha : \mathbb{R} \to \mathbb{R}^{d_i}$, the level sets are hyperplanes and so strict convexity of $X_i$, or in fact the slightly weaker condition that $\partial X_i$ has no $d_i - 1$ dimensional facets introduced in [FKM11], suffices.

## 3.2 A class of metrics involving singular measures

We now turn our attention to the results obtained in [NP23b] where we exploit the unequal dimensional optimal transport in order to define a new class of metrics involving singular measures. We will also see that these metrics help us to obtain new information about the variational problems we have studied above. Recall that given two probability measures $\mu_0$ and $\mu_1$ (we change a little bit the labelling of the measures with respect to the previous section for sake of simplicity in terms of notations) on a convex, bounded domain $X \subseteq \mathbb{R}^d$, the**Wasserstein distance** between them is defined as the infimal value in the Monge-Kantorovich optimal transport problem; that is,

$$\boxed{\mathsf{OT}^2(\mu_0, \mu_1) := \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{X \times X} |x_0 - x_1|^2 \mathrm{d}\pi(x_0, x_1)} \tag{3.32}$$

Among the many important properties of the Wasserstein distance (reviewed in [San15; Vil09; Vil03] for example) is the fact that it is a metric on the set $\mathcal{P}(X)$ of probability measures on $X$. In turn, the geodesics induced by this metric, known as *displacement interpolants* and introduced in [McC97], play key roles in many problems, both theoretical and applied. Variants of displacement interpolants, known as *generalized geodesics* (introduced in [AGS04], see Definition 3.2.1 below), are a natural and important tool in the analysis of problems involving a fixed base measure $\nu \in \mathcal{P}(X)$ as the ones treated in the previous section.

**Definition 3.2.1** (Generalized geodesics). *A generalized geodesic with base measure $\nu$ from $\mu_0$ to $\mu_1$ is a curve $\mu_t$ in $\mathcal{P}(X)$ of the form $\mu_t = ((1-t)e_0 + te_1)_\# \gamma$ for some $\gamma \in \mathcal{P}(X \times X \times X)$ where $\gamma_{yx_i} \in \Pi_{\text{opt}}(\nu, \mu_i)$ for $i = 0, 1$ where $\Pi_{\text{opt}}(\nu, \mu_i)$ is the set of optimal couplings between $\nu$ and $\mu_i$ with respect to optimal transport for the quadratic cost function.*

It is natural in certain variational problems to consider *singular* base measures; for instance in game theory problems derived from spatial economics, $\nu$ may represent a population of players, parametrized by their location $y \in \mathbb{R}^2$ [BC16]; it is often the case that the population is essentially concentrated along a one dimensional subset, such as a major highway or railroad. Our goal here is to develop an appropriate framework for these: we study a metric on an appropriate subset of $\mathcal{P}(X)$ for each choice of base measure $\nu$, which we call the *$\nu$-based Wasserstein metric* (see Definition 3.2.2); essentially, this metric arises from minimizing the average squared distance among all couplings of $\mu_0$ and $\mu_1$ corresponding to generalized geodesics. Geodesics with respect to the $\nu$-based Wasserstein metric will always be generalized geodesics with respect to $\nu$, but, for different structures of $\nu$, very different geometry is induced. We pay special attention to the cases when $\nu$ concentrates on a lower dimensional submanifold of $\mathbb{R}^d$, in which case the problem has a natural interplay with the unequal dimensional optimal transport problem explored in [CMP17; MP20; NP20]. In the particular case when $\nu$ concentrates on a line segment, we show that our metric coincides with the layerwise-Wasserstein metric (see (3.34)) introduced in [KPS20] to analyze anisotropic data such as plants' root shapes. We establish the following results:

- We obtain three equivalent characterizations of the $\nu$-based Wasserstein metric, roughly speaking:
  1. as an optimal transport problem restricted to couplings which are correlated along $\nu$ (we take this as the definition);
  2. by optimally coupling conditional probabilities of $\mu_0$ and $\mu_1$ after disintegrating with respect to optimal transport to $\nu$;
  3. as limits of multi-marginal optimal transport between $\mu_0$, $\mu_1$ and $\nu$.
- We establish uniqueness of the corresponding geodesics (although uniqueness of generalized geodesics for regular base measures was established in [AGS04], they are generally not unique when the base measure is singular).
- We study geodesic convexity of several functionals which play a key role in optimal transport research; many of these functionals were originally introduced by McCann [McC97], who established their displacement convexity. Using the standard terminology convexity along $\mathcal{W}_\nu$ geodesics of potential energies, interaction energies and the Wasserstein distance to $\nu$ follow immediately from known results, whereas convexity along $\mathcal{W}_\nu$ geodesics of the internal energy, under certain conditions, requires a new proof (convexity of the internal energy is known to hold along some, but not all, generalized geodesics [AGS04]). We note that this applies in particular to the layerwise-Wasserstein distance, yielding a far reaching improvement to [KPS20][Corollary 4.2]. We also show that when $\nu$ concentrates on a lower dimensional submanifold, the set of measures $\mu$ for which the model $(|x-y|^2, \mu, \nu)$ satisfies a strengthening of the generalized nested condition ( see Definition 3.0.1) is geodesically convex, see [NP23b][Proposition 33].
- We also introduce a class of metrics, said $\mu$-based Wasserstein metric, relevant in the case when the measure $\mu$ on $X \subset \mathbb{R}^d$ is fixed, and one would like to interpolate between measures on a fixed, lower dimensional submanifold $Y$. In particular we compare and interpolate between Kantorovich potentials on the $Y$ side in order to compare and interpolate between measures on $Y$.
- We, finally, show how the ideas introduced [NP23b] can be applied to prove convergence of computational methods to find equilibria in some game theoretic models. In particular we identify conditions under which equilibria are fixed points of a contractive mapping, implying uniqueness of the equilibrium (although this is easily deduced by other methods as well) and, perhaps more importantly, that it can be computed by iterating the mapping. This iteration had already been introduced as a method of computation by Blanchet-Carlier when $d = 1$, but without a proof of convergence, and in higher dimensions, with a proof of convergence but for simpler interaction terms [BC14b]. We prove that the relevant mapping is a contraction with respect to a variant of the $\mu$-based Wasserstein metric.

### 3.2.1 The $\nu-$based Wasserstein metric

We now define our metric with base point $\nu$ as follows.

**Definition 3.2.2** ($\nu-$based Wasserstein metric)**.** *Let $\nu \in \mathcal{P}(X)$. For $\mu_0, \mu_1 \in \mathcal{P}(X)$, we define the $\nu$-based Wasserstein metric as*

$$\boxed{\mathcal{W}_\nu(\mu_1, \mu_0) := \sqrt{\inf_{\gamma \in \Gamma} \int_{X \times X \times X} |x_0 - x_1|^2 \mathrm{d}\gamma(y, x_0, x_1)},} \tag{3.33}$$

*where $\Gamma := \{\gamma \in \mathcal{P}(X \times X \times X) \mid \gamma_{yx_i} \in \Pi_{\text{opt}}(\mu_i, \nu), i = 0, 1\}$.*

**Remark 3.2.3.** *This definition is closely related to the concept of linear optimal transport, introduced in [Wan+13]. The difference is that in linear optimal transport, a* fixed *optimal transport $\pi_{\mu_i} \in \Pi_{\text{opt}}(\nu, \mu_i)$ is selected for each $\mu_i$ (see equation (3) in [Wan+13]), whereas in the definition of $\mathcal{W}_\nu$ one minimizes over the entire set $\Pi_{\text{opt}}(\nu, \mu_i)$. For $\mu_i \in \mathcal{P}^u_\nu(X)$, the two concepts clearly coincide, and it is on this set that $\mathcal{W}_\nu$ yields a metric (see Lemma 3.2.4 below). Outside of this set, $\mathcal{W}_\nu$ still yields a semi-metric, whereas linear optimal transport might be better described as defining a metric on the selected $\pi_{\mu_i} \in \Pi_{\text{opt}}(\nu, \mu_i)$, since it is dependent on these choices.*

**Lemma 3.2.4** ([NP23b])**.** *$\mathcal{W}_\nu$ is a semi-metric on $\mathcal{P}(X)$. It is a metric on $\mathcal{P}^u_\nu(X)$.*

The proof of the following Lemma is very similar to the proof that the classical Wasserstein metric is in fact a metric (see for example, [San15][Proposition 5.1]).

**Example 3.2.5.** *We recall the main example used to introduce generalized geodesics in Section 9.2 of [AGS04]. Let $\nu$ be absolutely continuous with respect to d-dimensional Lebesgue measure on $X$. Then by Brenier's theorem [Bre91] there exist unique optimal couplings of the form $(\mathrm{id}, \tilde{T}_i)_\# \nu$ in $\Pi_{\text{opt}}(\nu, \mu_i)$, and therefore the only measure $\gamma$ with $\gamma_{yx_i} \in \Pi_{\text{opt}}(\nu, \mu_i)$ for $i = 0, 1$ is $\gamma = (Id, \tilde{T}_0, \tilde{T}_1)_\# \nu$. We then have*

$$\mathcal{W}^2_\nu(\mu_0, \mu_1) = \int_X |\tilde{T}_0(y) - \tilde{T}_1(y)|^2 \mathrm{d}\nu(y)$$

*so that the metric space $(\mathcal{P}(X), \mathcal{W}_\nu)$ is isometric to a subset of the Hilbert space $L^2(\nu)$. Geodesics for this metric take the form $t \mapsto \Big(tT_1 + (1-t)T_0\Big)_\# \nu$; these are the standard generalized geodesics found in, for example, Definition 7.31 of [San15].*

**Example 3.2.6.** *At the other extreme, suppose $\nu = \delta_y$ is a Dirac mass. Then for any coupling $\pi \in \Pi(\mu_0, \mu_1)$, the measure $\gamma = \delta_y \otimes \pi$ has $\gamma_{yx_i} = \delta \otimes \mu_i \in \Pi_{\text{opt}}(\nu, \mu_i)$. Since*

$$\int_{X \times X \times X} |x_0 - x_1|^2 \mathrm{d}\gamma(y, x_0, x_1) = \int_{X \times X} |x_0 - x_1|^2 \mathrm{d}\pi(x_0, x_1)$$

*we have*

$$\mathcal{W}^2_\nu(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{X \times X} |x_0 - x_1|^2 \mathrm{d}\pi(x_0, x_1)$$

*which is exactly the standard quadratic Wasserstein metric.*

We are especially interested in the cases in between these extremes, when $\nu$ is singular with respect to Lebesgue measure but not a Dirac mass. One of the simplest such cases is the following example.

**Example 3.2.7** ([NP23b][Proposition 16). *Suppose that $\nu$ concentrates on a line segment and is absolutely continuous with respect to one dimensional Hausdorff measure. It then turns out that the $\nu$-based Wasserstein metric coincides with the layerwise-Wasserstein metric introduced in [KPS20] and defined as follows*

$$\mathrm{d}_{LW}^2(\mu_0,\mu_1) := \mathsf{OT}^2(\mu_0^V,\mu_1^V) + \int_0^1 \mathsf{OT}^2(\tilde{\mu}_0^l,\tilde{\mu}_1^l)\mathrm{d}l \tag{3.34}$$

*where the $\mu_i^V = \Big(x = (x^1,x^2,...,x^d) \mapsto x^1\Big)_{\#}\mu_i$ are the vertical marginals of the $\mu_i$, $\tilde{\mu}_i$ are rescaled versions of the $\mu_i$, defined by $\tilde{\mu}_i = (F_{\mu_i},\mathrm{id})_{\#}\mu_i$ where $F_{\mu_i}$ is the cumulative distribution function of $\mu_i^V$, with the identity mapping* id *applied on the last $d-1$ coordinates (so that $(F_{\mu_i},\mathrm{id})(x^1,x^2,...,x^d) = (F_{\mu_i}(x^1),x^2,...,x^d)$) and $\tilde{\mu}_i = \mu_i^l \otimes dl$ is disintegrated with respect to its uniform vertical marginal.*

As we show below, $\mathcal{W}_\nu$ is also related to the following multi-marginal optimal transport problem. Fix $\varepsilon > 0$ and set

$$\boxed{\mathsf{MOT}^{\varepsilon c}(\nu,\mu_0,\mu_1) := \inf_{\gamma\in\Pi(\nu,\mu_0,\mu_1)} \int_{X\times X\times X} [\varepsilon|x_0-x_1|^2 + |x_0-y|^2 + |x_1-y|^2]\mathrm{d}\gamma(y,x_0,x_1).} \tag{3.35}$$

The following result establishes two different characterizations of the $\nu$-based Wasserstein metric.

**Theorem 3.2.8** ([NP23b]). *The following holds*

1. $$\boxed{\mathcal{W}_\nu^2(\mu_1,\mu_0) = \inf_{\pi_i\in\Pi_{\mathrm{opt}}(\mu_i,\nu),i=0,1} \int_X \mathsf{OT}^2(\mu_0^y,\mu_1^y)\mathrm{d}\nu(y),}$$

   *where $\mu_i^y$ is the conditional probability given $y$ of the optimal coupling $\pi_i = \nu(y)\otimes\mu_i^y(x) \in \Pi_{\mathrm{opt}}(\nu,\mu_i)$ between $\nu$ and $\mu_i$.*

2. *Furthermore, any weak limit point $\overline{\gamma}$ as $\varepsilon \to 0$ of minimizers $\gamma_\varepsilon$ of the multi-marginal problem* (3.35) *is an optimal coupling between $\mu_0$ and $\mu_1$ for the problem* (3.33) *defining $\mathcal{W}_\nu$.*

we can also establish the following $\Gamma$-convergence result relating $MM_\nu^\varepsilon$ and $W_\nu$.

**Proposition 3.2.9** ([NP23b]). *The functional on $\mathcal{P}(X)\times\mathcal{P}(X)$ defined by*

$$\boxed{(\mu_0,\mu_1) \mapsto \mathcal{F}_\nu^\varepsilon(\mu_0,\mu_1) := \frac{1}{\varepsilon}\mathsf{MOT}^{\varepsilon c}(\nu,\mu_0,\mu_1) - \frac{1}{\varepsilon}\mathsf{OT}^2(\nu,\mu_0) - \frac{1}{\varepsilon}\mathsf{OT}^2(\nu,\mu_0)}$$

$\Gamma$*-converges to $(\mu_0,\mu_1) \mapsto \mathcal{W}_\nu^2(\mu_0,\mu_1)$ as $\varepsilon \to 0$ with respect to the product on $\mathcal{P}(X)\times\mathcal{P}(X)$ of the weak topology on $\mathcal{P}(X)$ with itself*

We end this section on the properties of the $\nu-$based Wasserstein metric by the following corollary which characterizes the optimal solution in the case $d=2$ through the Knote-Rosenblatt rearrangement.

**Corollary 3.2.10** ([NP23b]). *Let $m=2$. Suppose $\nu$ is concentrated on the line segment $\{(t,0) : t\in\mathbb{R}\}$ and is absolutely continuous with respect to 1-dimensional Hausdorff measure. Then the optimal rearrangement $\gamma$ in* (3.33) *satisfies $\gamma_{x_0x_1} = (\mathrm{id},G)_{\#}\mu_0$, where $G$ is the Knothe-Rosenblatt rearrangement.*

### 3.2.2 Structure of geodesics for the $\nu-$based Wasserstein metric

We note that geodesics for the metric $W_\nu$ are generalized geodesics with base $\nu$, according to the general definition in [AGS04]. When $\nu$ is sufficiently regular, generalized geodesics are *uniquely* determined by $\mu_0$ and $\mu_1$. In this case, $\Pi_{\text{opt}}(\nu, \mu_i)$ consists of a single measure concentrated on the graph of a function $\tilde{T}_i : X \mapsto X$ pushing $\nu$ forward to $\mu_i$; $\gamma = (\text{id}, \tilde{T}_0, \tilde{T}_1)_\# \nu$ is then the unique measure satisfying $\gamma_{yx_i} \in \Pi_{\text{opt}}(\nu, \mu_i)$, and therefore the unique coupling minimizing the integral in the definition of $\mathcal{W}_\nu(\mu_0, \mu_1)$. If $\nu$ is singular, this reasoning does not apply; existence of minimizers in (3.33) follows immediately from standard continuity-compactness results, but they may be non-unique in general. Consider now the case when $\nu$ concentrates on some lower dimensional submanifold and take as cost function $c(x, y) = -\langle x, f(y)\rangle$ where $f \in \mathcal{C}^2(Y)$ is injective and non-degenerate, that is $D^2 f(y)$ has full rank everywhere.

In this setting, if the $\mu_i$ are absolutely continuous with respect to Lebesgue, we obtain a uniqueness result as a consequence of Theorem 3.2.8 and Lemma 3.2.13. We also characterize geodesics with respect to our metric is this setting.

**Theorem 3.2.11** ([NP23b]). *Assume that both $\mu_i$ are absolutely continuous with respect to Lebesgue measure and that $\nu$ is supported on a smooth $n$-dimensional submanifold of $X$, and is absolutely continuous with respect to $n$-dimensional Hausdorff measure. Then the optimal coupling $\gamma$ in* (3.33) *is unique and its two-fold marginal $\gamma_{x_0x_1}$ is induced by a mapping $G : X \to X$ pushing $\mu_0$ forward to $\mu_1$. Letting $T_i$ denote the optimal maps from $\mu_i$ to $\nu$, the restriction of $G$ to $T_i^{-1}(y)$ is the optimal mapping pushing $\mu_0^y$ forward to $\mu_1^y$ for a.e. $y$, where, as in Theorem 3.2.8, $\mu_i(x) = \mu_i^y(x) \otimes \nu(y)$. This mapping takes the form $\nabla\varphi^y : X_1(y, Dv_0(y)) \to X_1(y, Dv_1(y))$, where $\varphi^y : X_1(y, Dv_0(y)) \to \mathbb{R}$ is a convex function defined on the $(d-n)-$ dimensional affine subset $X_1(y, Dv_0(y))$.*

We now turn to the study of minimizing $\mathcal{W}_\nu$ geodesics.

**Proposition 3.2.12** ([NP23b]). *Suppose that $\pi_0, \pi_1 \in \Pi_{\text{opt}}(\nu)$. Then $\pi_t = ((1-t)e_0 + te_1, e_2)_\# \gamma$ is a minimizing geodesic for the $\nu$-based Wasserstein metric on couplings defined by*

$$\mathcal{W}_\nu(\mu_0, \mu_1) = \inf_{\pi_i \in \Pi_{\text{opt}}(\nu, \mu_i), i=0,1} \tilde{\mathcal{W}}_\nu(\pi_0, \pi_1),$$

*where*

$$\tilde{\mathcal{W}}_\nu(\pi_0, \pi_1) = \sqrt{\inf_{\gamma \in \mathcal{P}(X\times X\times X) | \gamma_{yx_i} = \pi_i, i=0,1} \int_{X\times X\times X} |x_0 - x_1|^2 \mathrm{d}\gamma(y, x_0, x_1)} \tag{3.36}$$

*$\gamma$ is optimal in* (3.36)*, where $e_0(x_0, x_1, y) = x_0$, $e_1(x_0, x_1, y) = x_1$, and $e_2(x_0, x_1, y) = y$. Suppose that $\mu_0, \mu_1 \in \mathcal{P}_\nu^u$. Then the generalized geodesic $\mu_t = ((1-t)e_0 + te_1)_\# \gamma$ is a minimizing geodesic for the $\nu$ based Wasserstein metric for each optimal $\gamma$ in* (3.33)*, provided that $\mu_t \in \mathcal{P}_\nu^u$ for all $t \in [0, 1]$.*

We end this section on the structure of geodesic by giving a uniqueness result. Indeed, under much stronger conditions on the marginals $\mu_0$ and $\mu_1$ and the reference measure $\nu$, we are able to use Theorem 3.2.11 to show that the geodesic for the $\mathcal{W}_\nu$ metric between $\mu_0$ and $\mu_1$ is unique.

**Lemma 3.2.13** (Structure of geodesics, [NP23b]). *Let $\mu_0$, $\mu_1 \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be absolutely continuous with respect to the $d-$ dimensional and $n-$dimensional Lebesgue measure respectively. Then the geodesic $\mu_t$ between $\mu_0$ and $\mu_1$ is uniquely defined. The Kantorovich potential for optimal transport between $\mu_t$ and $\nu$ is $v_t = tv_1 + (1-t)v_0$, where $v_i : Y \to \mathbb{R}$ is the Kantorovich potential* *between $\mu_i$ and $\nu$, for $i = 0, 1$. For almost every $y$, the conditional probability $\mu_t^y$, concentrated on $X_1(y, Dv_t(y))$, is the unique displacement interpolant between $\mu_0^y$ and $\mu_1^y$. It is absolutely continuous with respect to $(d-n)$-dimensional Hausdorff measure; denoting its density by $\overline{\mu}_t^y$, and the density of $\mu_t$ with respect to $\mathcal{L}^d$ by $\overline{\mu}_t$, we have,*

$$\boxed{\overline{\mu}_t^y(x_t) = \frac{\overline{\mu}_t(x_t)}{JT_t^y(x_t)\overline{\nu}(y)}}$$

*where*

$$JT_t^y(x_t) := \frac{\sqrt{\det(Df^T Df(y))}}{\det[-D^2 v_t(y) - x_t \cdot D^2 f(y)]}$$

*is the Jacobian of the optimal map $T_t$ from $\mu_t$ to $\nu$, evaluated at $x_t \in T_t^{-1}(y) \subseteq X_1(y, Dv_t(y))$ and $f$ is defined at the beginning of this section. The Brenier map between $\mu_0^y$ and $\mu_t^y$ is then given by $\nabla\varphi_t^y = (1-t)I + t\nabla\varphi^y$, where $\nabla\varphi^y : T_0^{-1}(y) \to T_1^{-1}(y)$ is the Brenier map between $\mu_0^y$ and $\mu_1^y$.*

### 3.2.3 Geodesic convexity of functionals

We now turn our attention to certain convexity properties of $\mathcal{W}_\nu$ geodesics and we start by studying the convexity of various functionals along the geodesics. We consider the functionals

$$\mu \mapsto \mathsf{OT}^2(\mu, \nu) \tag{3.37}$$

$$\mu \mapsto \int_X V(x)\mathrm{d}\mu(x) \tag{3.38}$$

$$\mu \mapsto \int_{X^2} W(x-z)\mathrm{d}\mu(x)\mathrm{d}\mu(z) \tag{3.39}$$

and

$$\mu \mapsto \begin{cases} \int U(\overline{\mu}(x))\mathrm{d}x & \text{if } \mathrm{d}\mu = \overline{\mu}(x)\mathrm{d}x \text{ is a.c. wrt Lebesgue measure on } X \\ +\infty & \text{otherwise.} \end{cases} \tag{3.40}$$

We are able show that these functionals are geodesically convex with respect to the metric $\mathcal{W}_\nu$. Explicitly, geodesic convexity means that $\mathcal{F}(\mu_t)$ is a convex function of $t \in [0,1]$, where $\mathcal{F} : \mathcal{P}(X) \to \mathbb{R}$ is any of (3.39), (3.40), (3.38) or (3.37). We will restrict our attention to $\mathcal{W}_\nu$ geodesics $\mu_t$ of the form in Proposition 3.2.12; that is, $\mu_t = (te_0 + (1-t)e_1)_\#\gamma$, where $\gamma$ is optimal in (3.33), under the assumption that each $\mu_t \in \mathcal{P}_\nu^u(X)$, since in more general situations the existence of a minimizing $\mathcal{W}_\nu$ geodesic in the metric space $\mathcal{P}_\nu^u$ joining $\mu_0$ and $\mu_1$ is not clear. We are particularly interested in the case when the endpoints $\mu_0$ and $\mu_1$ and reference measure $\nu$ satisfy the assumptions in Theorem 3.2.11, under which the existence of a unique geodesic follows from Lemma 3.2.13. Since the geodesics we consider for $\mathcal{W}_\nu$ are always generalized geodesics, the following result follows immediately from Lemma 9.2.1 and Propositions 9.3.2 and 9.3.5 [AGS04].

**Proposition 3.2.14** ([NP23b])**.** *The following functionals are convex along all minimizing $\mathcal{W}_\nu$ geodesics of the form in Proposition 3.2.12, under the corresponding conditions:*

1. (3.37) *is geodesically 1-convex for any $\nu$.*
2. (3.38) *is geodesically convex if $V$ is convex. It is geodesically strictly convex if $V$ is strictly convex.*

*3.* (3.39) *is geodesically convex if W is convex. It is geodesically strictly convex along the subset of measures with the same barycenter if W is strictly convex.*

The geodesic convexity of the internal energy is somewhat more involved; in particular unlike the other three forms, it does not hold for all generalized geodesics. Our proof, under additional restrictions on $\nu$, uses the structure of the transport map $G$ in a crucial way.

**Lemma 3.2.15** ([NP23b]). *Assume that the reference measure $\nu$ is absolutely continuous with respect to n-dimensional Hausdorff measure on a smooth n-dimensional submanifold $Y$ and that $\mu_0$ and $\mu_1$ are absolutely continuous with respect to Lebesgue measure on $X$. The internal energy at the interpolant is given by*

$$\int_X U(\overline{\mu}_t(x))\mathrm{d}x = \int_Y \frac{1}{\overline{\nu}(y)} \int_{T_0^{-1}(y)} U\Big(\frac{\overline{\mu}_0^y(x_0)\overline{\nu}(y) JT_t^y(\nabla\varphi_t^y(x_0))}{\det(D^2\varphi_t^y(x_0))}\Big)\frac{1}{JT_t^y(\nabla\varphi_t^y(x_0))}\det(D^2\varphi_t^y(x_0))\mathrm{d}\mathcal{H}^{d-n}(x_0)\mathrm{d}y \tag{3.41}$$

From the previous Lemma it follows the geodesic convexity of the internal energy.

**Corollary 3.2.16** ([NP23b]). *Under the assumption in the preceding Lemma, the internal energy* (3.40) *is geodesically convex assuming $U(0) = 0$ and $r \mapsto r^m U(r^{-m})$ is convex and non-increasing.*

Note that the condition that $r \mapsto r^m U(r^{-m})$ is convex and non-increasing is quite standard and goes back to the seminal paper by McCann [McC97].

### 3.2.4 The $\mu-$based Wasserstein metric

We consider now the case in which the base measure is actually the one in higher dimension. We can define a class of metrics on a lower dimensional space. More precisely, these metrics use the Kantorovich potentials $v(y)$ on the variable $y$ arising from optimal transport to a fixed reference measure $\mu(x)$ to compare measures on the $y$ variable. Notice that on the contrary the $\mathcal{W}_\nu$ uses in part the Kantorovich potential $v(y)$ for optimal transport from a reference measure $\nu(y)$ on $Y$ to measures on $X$ in order to compare those free measures. Fix reference measures $\mu \in \mathcal{P}(X)$ and $\sigma \in \mathcal{P}(Y)$ on bounded domains $X \subseteq \mathbb{R}^d$ and $Y \subseteq \mathbb{R}^n$, with $n \leq d$, absolutely continuous with respect to $m$ and $n$ dimensional Lebesgue measure, respectively. Let $c : \overline{X} \times \overline{Y} \to \mathbb{R}$ be a cost function which is $\mathcal{C}^1$ up to the boundary and satisfies the *twist condition.*

We let $\mathcal{P}_{ac,\sigma}(Y)$ be the set of probability measures on $Y$ which are absolutely continuous with respect to $\sigma$. For measures $\nu_0, \nu_1 \in \mathcal{P}_{ac,\sigma}(Y)$, we define

$$\boxed{\mathcal{W}_{\mu,\sigma,c,p}(\nu_0,\nu_1) := ||Dv_0 - Dv_1||_{L^p(\sigma)}} \tag{3.42}$$

for some $p \in [1, \infty]$ where $v_i$ is the $c$-concave Kantorovich potential, that is, $c$-conjugate solutions to the dual problem of (3.1) corresponding to the optimal transport problem between $\mu$ and $\nu_i$ with cost function $c$.

**Proposition 3.2.17** ([NP23b]). *$\mathcal{W}_{\mu,\sigma,c,p}$ is a metric on $\mathcal{P}_{ac,\sigma}(Y)$.*

To help explain the motivation behind this metric, it is useful to first consider the $m = n$ case with $c(x, y) = |x - y|^2$. The metric $\mathcal{W}_{\mu,\sigma,c,2}$ is then similar to the $\nu$-based Wasserstein metric. Recall from Example 3.2.5 that in this case $\mathcal{W}_\nu(\mu_0, \mu_1)$ between two measures in $\mathcal{P}(X)$ coincides with the $L^2$ metric on the optimal maps from $\nu$ to the $\mu_i$; on the other hand, $\mathcal{W}_{\mu,\sigma,c,2}(\nu_0, \nu_1)$ for two measures in $\mathcal{P}(Y)$ coincides with the $L^2$ metric on the optimal maps from the $\nu_i$ to $\mu$. That is, $\mathcal{W}_{\mu,\sigma,c,2}(\nu_0, \nu_1)$ compares gradients of potential on the opposite, rather than same, side of the problem from the reference measure. The case where $d > n$ is an attempt to generalize this idea in the same way that the $\nu$-based Wasserstein metric generalizes the $L^2$ metric in Example 3.2.5. The metric $W^*_{\mu,\sigma,c,p}$ arises naturally in certain variational problems among measures on $Y$ involving optimal transport to the fixed measure $\mu$ on $X$ like the one treated in Section 3.1.2, we will expand this viewpoint in the section below.

### 3.2.5 Fixed point characterization of solutions to variational problems

We consider a fixed measure $\mu$ on the bounded domain $X \subseteq \mathbb{R}^m$, absolutely continuous with respect to Lebesgue measure with density $\overline{\mu}$, bounded above and below, and a fixed domain $Y := (\underline{y}, \overline{y}) \subset \mathbb{R}$. We are interested in the minimization problem

$$\boxed{\inf_{\nu \in \mathcal{P}(Y)} \mathsf{OT}^c(\mu, \nu) + \int \overline{\nu}(y) \log(\overline{\nu}(y)) \mathrm{d}y.} \tag{3.43}$$

We recall that this problem belongs to a class arising in certain game theory problems, introduced in a series of papers by Blanchet-Carlier [BCN17; BC16; BC14b], in which the fixed measure $\mu$ represents a distribution of players and $Y$ a space of strategies; minimizing measures $\nu$ in (3.10) represent equilibrium distributions of strategies. In general, the dimensions of $d$ and $n$ of the spaces of players and strategies represent the number of charateristics used to distinguish between them. The case $d > n$ (including the case $n = 1$ treated here) are of particular interest, since it is often the case that **players are highly homogeneous**, whereas **the strategies available to them are less varied**. When $m = 1$ (so that both players and strategy spaces are one dimensional), Blanchet-Carlier introduced a characterization of minimizers as fixed points of a certain mapping. They then iterated that mapping as a way to compute solutions [BC14b]; however, they did not prove convergence of this scheme. Here we introduce a similar fixed point characterization for $d \geq 1$. Furthermore, we prove for any $m$, under certain assumptions, that the mapping is a contraction with respect to a strategically chosen metric, implying convergence of the iterative scheme; even for $d = 1$, this is a new result. The key to the proof is choosing a suitable metric; our choice is a slight variant of the metric $\mathcal{W}_{\mu,\mathcal{L},c,1}$ introduced in Section 3.2.4. Choose $\underline{d} = \min_{x \in \overline{X}, y \in \overline{Y}} \frac{\partial c}{\partial y}(x, y), \overline{d} = \max_{x \in \overline{X}, y \in \overline{Y}} \frac{\partial c}{\partial y}(x, y)$. Let $k \in L^1\Big((\underline{y}, \overline{y}); [\underline{d}, \overline{d}]\Big)$, the set of $L^1$ functions on $(\underline{y}, \overline{y})$ taking values in $[\underline{d}, \overline{d}]$, and set $v_k(y) := \int_{\underline{y}}^y k(s) ds$. By using optimality condition of (3.43) we can define the following mapping $L^1\Big((\underline{y}, \overline{y}); [\underline{d}, \overline{d}]\Big) \to \mathcal{P}^{ac}((\underline{y}, \overline{y}))$,

$$k \mapsto \overline{\nu}_k(y) := \frac{e^{-v_k(y)}}{\int_{\underline{y}}^{\overline{y}} e^{-v_k(s)} \mathrm{d}s}. \tag{3.44}$$

This is a probability density on $(\underline{y}, \overline{y})$ (we denote the corresponding measure by $\nu_k(y)$); moreover note that for a given $\nu \in \mathcal{P}^{ac}((\underline{y}, \overline{y}))$ the mass splitting property

$$\mu(X^c_{\geq}(y, k)) = \nu((\underline{y}, y)),$$

selects a unique $k = k(y) \in [\underline{d}, \overline{d}]$, for each $y \in (\underline{y}, \overline{y})$. $k(y)$ is then bounded and so clearly in $L^1\Big((\underline{y},\overline{y});[\underline{d},\overline{d}]\Big)$. We can then think of this process as defining a mapping $\mathcal{P}^{ac}((\underline{y},\overline{y})) \to L^1\Big((\underline{y},\overline{y});[\underline{d},\overline{d}]\Big)$, $\nu \mapsto k(\cdot)$. Composing this mapping with (3.44) then yields a $F : L^1\Big((\underline{y},\overline{y});[\underline{d},\overline{d}]\Big) \to L^1\Big((\underline{y},\overline{y});[\underline{d},\overline{d}]\Big)$. Explicitly, defining $\tilde{k} := F[k]$, we have

$$\nu_k(\underline{y}, y) := \mu(X_{\geq}(y, \tilde{k}(y))). \tag{3.45}$$

This approach let us define the following algorithm 2 to compute solution to (3.43). We can then

**Algorithm 2** Algorithm to compute the optimal $\nu$

**Require:** $k^{(0)}$
1: **while** $||k^{(l+1)} - k^{(l)}|| <$tol **do**
2:

$$\overline{\nu}^{(l)}(y) := \frac{e^{-v^{(l)}(y)}}{\int_{\underline{y}}^{\overline{y}} e^{-v^{(l)}(s)} \mathrm{d}s}.$$

3: Compute $k^{(l)}$ by using the mass splitting $\mu(X^c_{\geq}(y,k)) = \nu^{(l)}((\underline{y}, y))$.
4: **end while**

characterize the minimizer of (3.43) as a fixed point of this mapping $F$.

**Lemma 3.2.18** ([NP23b])**.** *Let $\nu$ be a minimizer of* (3.10) *such that $(c, \mu, \nu)$ is nested. Then $k(y) = v'(y)$ is a fixed point of $F$, where $v$ is the corresponding Kantorovich potential.*

We can prove that, under certain conditions, $F$ is a contraction. The Banach fixed point theorem will then imply that $F$ has a unique fixed point. Under conditions ensuring nestedness of the minimizer, Lemma 3.2.18, will then ensure that that fixed point is the minimizer of (3.43). We need the following conditions on $X$, $c$ and $\mu$ to ensure $F$ is a contraction:

- $A := \max |D_x c_y(x, y)| < \infty$,
- $B := \min \overline{\mu}(x) > 0$,
- There is a $C > 0$ such that, for all $y \in Y$, $x \in X$, using $c_y(x, y)$ as a shorthand for $\frac{\partial c}{\partial y}(x, y)$,

$$\mathcal{H}^{d-1}(X^c_1(y, c_y(x, y))) \geq C \min\Big\{\mu(X^c_{\geq}(y, c_y(x, y))), \mu(X^c_{\leq}(y, c_y(x, y)))\Big\}. \tag{3.46}$$

Let us discuss briefly the third condition above. For any $x, y$, setting $p = c_y(x, y)$, the level set $X^c_1(y, p)$ is an $d - 1$ dimensional submanifold, dividing the region $X$ into the sub and super level sets $X^c_{\geq}(y, p)$ and $X^c_{\leq}(y, p)$. The condition implies that if $X^c_1(y, p)$ is small, then at least one of $\mu(X^c_{\geq}(y, p))$ or $\mu(X^c_{\leq}(y, p))$ must also be (quantitatively) small. In addition, we set

$$H(y) = \max\left\{\frac{|e^{(\overline{d}-2\underline{d})(y-\underline{y})} - 1|}{|1 - e^{-\overline{d}(y-\underline{y})}|}, \frac{|e^{(\overline{d}-2\underline{d})(\overline{y}-\underline{y})} - e^{(\overline{d}-2\underline{d})(y-\underline{y}}|}{|(e^{-\overline{d}(y-\underline{y})} - e^{-\overline{d}(\overline{y}-\underline{y})}|}\right\}.$$

**Theorem 3.2.19** ([NP23b]). *Assume the three conditions above, and that*

$$\frac{2A\overline{d}^3(1-e^{-\underline{d}(\overline{y}-\underline{y})})^2}{BC\underline{d}^2[1-e^{-\overline{d}(\overline{y}-\underline{y})}]^2|\overline{d}-2\underline{d}|}\int_{\underline{y}}^{\overline{y}}H(y)\mathrm{d}y<1.$$

*Then $F$ is a contraction.*

**Remark 3.2.20.** *The condition appearing in the statement of the theorem looks complicated; however, all quantities except for $B$ and $C$ can be computed using the cost function alone ($B$ involves the reference marginal $\mu$ while $C$ involves both the cost and $\mu$). Since the factor in front of the integral and the function $H(y)$ are bounded, the limit as $\overline{y}-\underline{y}\to 0$ of the left hand side above is $0$; therefore, the condition will always hold for sufficiently small intervals. We illustrate these points in a simple example below, where we compute each of these quantities explicitly.*

**Remark 3.2.21** (Relationship to $\mu-$based metrics)**.** *We show that $F$ is a contraction with respect to the $L^1$ metric. Noting that the mapping $\nu\mapsto k$, where $k$ is defined by the mass splitting condition $\nu(\underline{y},y)=\mu(X_{\geq}(y,k(y)))$ is a bijection on the set of non-atomic measures $\nu\in\mathcal{P}(Y)$ (and in particular on the set $\mathcal{P}^{ac}(Y)$ of measures which are absolutely continuous with respect to Lebesgue measure $\mathcal{L}$), we can consider the $L^1$ metric to be a metric on $\mathcal{P}^{ac}(Y)$. If $(c,\mu,\nu)$ is nested, then $k(y)=v'(y)$ where $v$ is the Kantorovich dual potential. In this case, this metric is exactly the metric $\mathcal{W}_{\mu,\mathcal{L},c,1}$ from Section 3.2.4.*

Now combining Lemma 3.2.18, condition 3.1.2 ensuring nestedness of the solution, and the Banach Fixed Point Theorem, we get the convergence of the iterative algorithm to compute the solution we discussed above.

**Corollary 3.2.22** ([NP23b])**.** *Under the assumptions in Theorem 3.2.19 and 3.1.2, for any $k\in L^1\Big((\underline{y},\overline{y});[\underline{d},\overline{d}]\Big)$, the sequence $k^{(l)}$ generated by algorithm 2 converges to a function $k_{fixed}$ whose anti-derivative is the Kantorovich potential corresponding to the unique minimizer $\nu$ of* (3.43).

Let us conclude this section by showing some numerical results obtained by using the algorithm 2. In all the simulations we have taken the uniform density on the arc $(0,\frac{\pi}{6})$ as initial density $\nu^{(0)}$, $\mu$ is always the uniform on quarter disk and $f(\nu)$ is the entropy as above. The cost is given by $c(x,y)=|x_1-\cos(y)|^2+|x_2-\sin(y)|^2$ we know that the model is nested provided $\overline{y}\leq 0.61$ (which is exactly the case we consider for the numerical tests). Moreover, we have also considered an additional potential term (the convergence proof can be straightforward adapted to this simple variant), which will favour a concentration in a certain area of $(0,\frac{\pi}{6})$, so that the functional $\mathcal{F}$ has the form

$$\mathcal{F}(\nu)=\int_Y f(\nu)\mathrm{d}y+\int V(y)\mathrm{d}\nu(y),$$

where $V(y)=10|y-0.1|^2$. In Figure 3.1 we show the final density we have obtained (on the left) and the intersection between the level-set $X_{\geq}(y,k^{\star}(y))$ for the final solution and the support of the fix measure $\mu$. Notice that, as expected, the level set $X_{\geq}(y,k^{\star}(y))$ do not cross each other meaning that the model is nested.

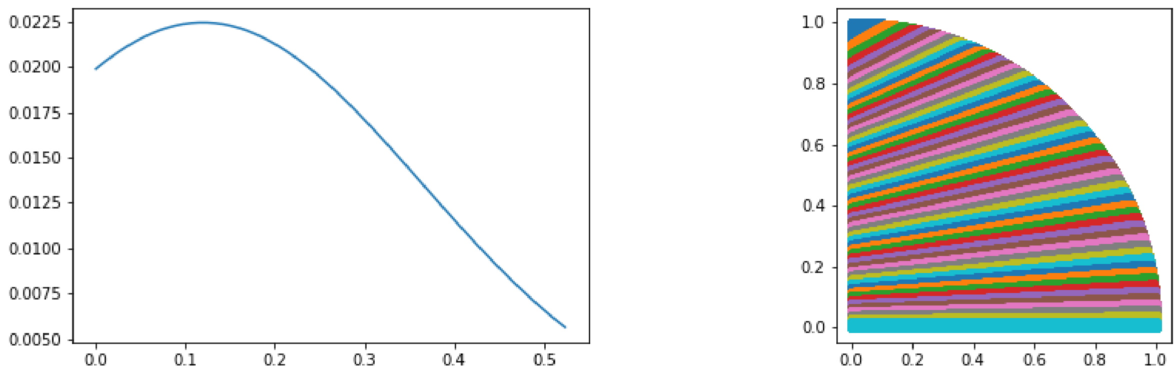



Figure 3.1: (Left) Final density $\nu^\star$. (Right) Intersection between the level-set $X_{\geq}(y, k^\star(y))$ for the final solution and the support of the fix measure $\mu$. These figures were firstly presented in a joint work with B. Pass [NP22].

## 3.3 Perspectives

The research perspectives I would like to address in the future can be developed along the two main axis of this chapter (1) the unequal dimensional optimal transport and (2) the class of metrics involving singular measures.

- Concerning the former an interesting variational problem where optimal transport arises naturally is the principal agent's problem (see for instance [FKM11; Car01] as well as the seminal work by Rochet and Choné [RC98]) where the population of agents $\mu$ is given and the monopolist must choose the best strategies in order to maximise the utility. However, as in the Cournot-Nash case explained above, there is no need for the space of strategies to have the same dimension as the ones of the agents, actually it is natural to assume that it is one dimensional. It turns out then that an unequal dimensional optimal transport problems arises. As in the joint work with B. Pass it is natural to study if the model (which can suffers a lack of convexity) is nested in order to make the solutions more tractable and numerically computable.

- As for the class of metrics I would like to further develop the possible applications: as we have highlighted above the $\nu-$based metric is useful to compare/interpolate anisotropic measures (as it happens for the plants' root [KPS20]) since the geodesic preserves this property of the measures. So it is natural to define a barycenter of anisotropic measures and study its property as well as the possible numerical methods to compute it. Finally, notice that this metric is in some sense a non-linear (and more difficult to compute numerically) version of the sliced Wasserstein metric. As for sliced Wasserstein ([Bon13]) one can study its relation with the standard Wasserstein metric and, up to make it numerically tractable, study the associated gradient flows: does the generated sequence converges to the solution of a specific PDE (in the same flavour as done in [JKO98])? Can be applied to understand better some machine learning problems as for the sliced counterpart [Bon+22] ?

# Bibliography


[Abr+17] I. Abraham, R. Abraham, M. Bergounioux, and G. Carlier. "Tomographic Reconstruction from a Few Views: A Multi-Marginal Optimal Transport Approach". In: *Applied Mathematics & Optimization* 75.1 (2017), pp. 55–73.

[AC11] M. Agueh and G. Carlier. "Barycenters in the Wasserstein Space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924. ISSN: 0036-1410, 1095-7154. DOI: 10.1137/100805741.

[ACE22] A. Alfonsi, R. Coyaud, and V. Ehrlacher. "Constrained Overdamped Langevin Dynamics for Symmetric Multimarginal Optimal Transportation". In: *Mathematical Models and Methods in Applied Sciences* 32.03 (2022), pp. 403–455. ISSN: 0218-2025. DOI: 10.1142/S0218202522500105.

[Alf+21] A. Alfonsi, R. Coyaud, V. Ehrlacher, and D. Lombardi. "Approximation of Optimal Transport Problems with Marginal Moments Constraints". In: *Mathematics of Computation* 90.328 (2021), pp. 689–737. ISSN: 0025-5718, 1088-6842. DOI: 10.1090/mcom/3568.

[AB21] J. M. Altschuler and E. Boix-Adserà. "Hardness Results for Multimarginal Optimal Transport Problems". In: *Discrete Optimization* 42 (2021), p. 100669. ISSN: 1572-5286. DOI: 10.1016/j.disopt.2021.100669.

[ANS22] J. M. Altschuler, J. Niles-Weed, and A. J. Stromme. "Asymptotics for Semidiscrete Entropic Optimal Transport". In: *SIAM Journal on Mathematical Analysis* 54.2 (2022), pp. 1718–1741. ISSN: 0036-1410. DOI: 10.1137/21M1440165.

[AGS04] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures.* 2nd ed. Lectures in Mathematics. Birkhäuser Boston, 2004.

[BHP13] M. Beiglböck, P. Henry-Labordere, and F. Penkner. "Model-Independent Bounds for Option Prices—a Mass Transport Approach". In: *Finance and Stochastics* 17.3 (2013), pp. 477–501.

[Ben+18] J.-D. Benamou, G. Carlier, S. Di Marino, and L. Nenna. "An Entropy Minimization Approach to Second-Order Variational Mean-Field Games". 2018.

[BCN16] J.-D. Benamou, G. Carlier, and L. Nenna. "A Numerical Method to Solve Multi-Marginal Optimal Transport Problems with Coulomb Cost". In: *Splitting Methods in Communication, Imaging, Science, and Engineering.* Springer International Publishing, 2016, pp. 577–601.

[Ben21] J.-D. Benamou. "Optimal Transportation, Modelling and Numerical Simulation". In: *Acta Numerica* 30 (2021), pp. 249–325. ISSN: 0962-4929. DOI: 10.1017/S0962492921000040.

[Ben+15] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. "Iterative Bregman Projections for Regularized Transportation Problems". In: *SIAM Journal on Scientific Computing* 37.2 (2015), A1111–A1138. ISSN: 1064-8275. DOI: 10.1137/141000439.

[BCN19] J.-D. Benamou, G. Carlier, and L. Nenna. "Generalized Incompressible Flows, Multi-Marginal Transport and Sinkhorn Algorithm". In: *Numerische Mathematik* 142.1 (2019), pp. 33–54. ISSN: 0945-3245. DOI: 10.1007/s00211-018-0995-x.

[BGV19] J.-D. Benamou, T. O. Gallouët, and F.-X. Vialard. "Second-Order Models for Optimal Transport and Cubic Splines on the Wasserstein Space". In: *Foundations of Computational Mathematics* 19.5 (2019), pp. 1113–1143. ISSN: 1615-3383. DOI: 10.1007/s10208-019-09425-z.

[BK18] J. Bigot and T. Klein. "Characterization of Barycenters in the Wasserstein Space by Averaging Optimal Transport Maps". In: *ESAIM: Probability and Statistics* 22 (2018), pp. 35–57.

[BD17] U. Bindini and L. De Pascale. "Optimal Transport with Coulomb Cost and the Semiclassical Limit of Density Functional Theory". 2017.

[BDK21] U. Bindini, L. De Pascale, and A. Kausamo. *On Deterministic Solutions for Multi-Marginal Optimal Transport with Coulomb Cost.* 2021. URL: http://arxiv.org/abs/2011.05063. preprint.

[BC14a] A. Blanchet and G. Carlier. "From Nash to Cournot-Nash Equilibria via the Monge-Kantorovich Problem". In: *Philosophical Transactions of the Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences* 372.2028 (2014), pp. 20130398, 11. ISSN: 1364-503X. DOI: 10.1098/rsta.2013.0398.

[BC16] A. Blanchet and G. Carlier. "Optimal Transport and Cournot-Nash Equilibria". In: *Mathematics of Operations Research* 41.1 (2016), pp. 125–145. ISSN: 0364-765X. DOI: 10.1287/moor.2015.0719.

[BC14b] A. Blanchet and G. Carlier. "Remarks on Existence and Uniqueness of Cournot–Nash Equilibria in the Non-Potential Case". In: *Mathematics and Financial Economics* 8.4 (2014), pp. 417–433.

[BCN17] A. Blanchet, G. Carlier, and L. Nenna. "Computation of Cournot–Nash Equilibria by Entropic Regularization". In: *Vietnam Journal of Mathematics* (2017). ISSN: 2305-2228. DOI: 10.1007/s10013-017-0255-x.

[Bon+22] C. Bonet, N. Courty, F. Septier, and L. Drumetz. *Efficient Gradient Flows in Sliced-Wasserstein Space.* 2022. URL: http://arxiv.org/abs/2110.10972. preprint.

[Bon13] N. Bonnotte. "Unidimensional and Evolution Methods for Optimal Transportation". PhD thesis. Université Paris Sud - Paris XI ; Scuola normale superiore (Pise, Italie), 2013.

[BL92] J. M. Borwein and A. S. Lewis. "Decomposition of Multivariate Functions". In: *Canadian Journal of Mathematics* 44.3 (1992), pp. 463–482. ISSN: 0008-414X, 1496-4279. DOI: 10.4153/CJM-1992-030-9.

[BLN94] J. M. Borwein, A. S. Lewis, and R. D. Nussbaum. "Entropy Minimization, DAD Problems, and Doubly Stochastic Kernels". In: *Journal of Functional Analysis* 123.2 (1994), pp. 264–307. ISSN: 0022-1236. DOI: 10.1006/jfan.1994.1089.

[Bou+21a] G. Bouchitté, G. Buttazzo, T. Champion, and L. De Pascale. "Dissociating Limit in Density Functional Theory with Coulomb Optimal Transport Cost". In: *ANNALI SCUOLA NORMALE SUPERIORE - CLASSE DI SCIENZE* (2021), pp. 1429–1471. ISSN: 2036-2145, 0391-173X. DOI: 10.2422/2036-2145.201903_014.

[Bou+21b] G. Bouchitté, G. Buttazzo, T. Champion, and L. De Pascale. "Relaxed Multi-Marginal Costs and Quantization Effects". In: *Annales de l'Institut Henri Poincaré C* 38.1 (2021), pp. 61–90. ISSN: 0294-1449. DOI: 10.1016/j.anihpc.2020.06.004.

[Bre93] Y. Brenier. "The Dual Least Action Problem for an Ideal, Incompressible Fluid". In: *Archive for Rational Mechanics and Analysis* 122.4 (1993), pp. 323–351. ISSN: 1432-0673. DOI: 10.1007/BF00375139.

[Bre99] Y. Brenier. "Minimal Geodesics on Groups of Volume-Preserving Maps and Generalized Solutions of the Euler Equations". In: *Communications on Pure and Applied Mathematics* 52.4 (1999), pp. 411–452. ISSN: 1097-0312. DOI: 10.1002/(SICI)1097-0312(199904)52:4<411::AID-CPA1>3.0.CO;2-3.

[Bre91] Y. Brenier. "Polar Factorization and Monotone Rearrangement of Vector-Valued Functions". In: *Communications on Pure and Applied Mathematics* 44.4 (1991), pp. 375–417. ISSN: 1097-0312. DOI: 10.1002/cpa.3160440402.

[Bre89] Y. Brenier. "The Least Action Principle and the Related Concept of Generalized Flows for Incompressible Perfect Fluids". In: *Journal of the American Mathematical Society* 2.2 (1989), pp. 225–255.

[BCD18] G. Buttazzo, T. Champion, and L. De Pascale. "Continuity and Estimates for Multimarginal Optimal Transportation Problems with Singular Costs". In: *Applied Mathematics & Optimization* 78.1 (2018), pp. 185–200. ISSN: 1432-0606. DOI: 10.1007/s00245-017-9403-7.

[BDG12] G. Buttazzo, L. De Pascale, and P. Gori-Giorgi. "Optimal-Transport Formulation of Electronic Density-Functional Theory". In: *Physical Review A* 85.6 (2012), p. 062502.

[BS05] G. Buttazzo and F. Santambrogio. "A Model for the Optimal Planning of an Urban Area". In: *SIAM journal on mathematical analysis* 37.2 (2005), pp. 514–530.

[CM10] L. A. Caffarelli and R. J. McCann. "Free Boundaries in Optimal Transport and Monge-Ampère Obstacle Problems". In: *Annals of Mathematics. Second Series* 171.2 (2010), pp. 673–730. ISSN: 0003-486X. DOI: 10.4007/annals.2010.171.673.

[CGP21] E. Caglioti, F. Golse, and T. Paul. *Towards Optimal Transport for Quantum Densities.* 2021. URL: http://arxiv.org/abs/2101.03256. preprint.

[Can+04] E. Cancès, R. Keriven, F. Lodier, and A. Savin. "How Electrons Guard the Space: Shape Optimization with Probability Distribution Criteria". In: *Theoretical Chemistry Accounts* 111.2 (2004), pp. 373–380. ISSN: 1432-2234. DOI: 10.1007/s00214-003-0509-4.

[Car03] G. Carlier. "On a Class of Multidimensional Optimal Transportation Problems". In: *Journal of Convex Analysis* 10.2 (2003), pp. 517–529. ISSN: 0944-6532.

[CGS09] G. Carlier, A. Galichon, and F. Santambrogio. "From Knothe's Transport to Brenier's Map and a Continuation Method for Optimal Transport". In: *Siam Journal On Mathematical Analysis* 41.6 (2009), pp. 2554–2576. ISSN: 0036-1410. DOI: 10.1137/080740647.

[Car01] G. Carlier. "A General Existence Result for the Principal-Agent Problem with Adverse Selection". In: *Journal of Mathematical Economics* 35.1 (2001), pp. 129–150. ISSN: 0304-4068. DOI: 10.1016/S0304-4068(00)00057-4.

[Car22] G. Carlier. "On the Linear Convergence of the Multimarginal Sinkhorn Algorithm". In: *SIAM Journal on Optimization* 32.2 (2022), pp. 786–794. ISSN: 1052-6234. DOI: 10.1137/21M1410634.

[CCG16] G. Carlier, V. Chernozhukov, and A. Galichon. "Vector Quantile Regression: An Optimal Transport Approach". In: *The Annals of Statistics* 44.3 (2016), pp. 1165–1192.

[Car+17] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. "Convergence of Entropic Schemes for Optimal Transport and Gradient Flows". In: *SIAM Journal on Mathematical Analysis* 49.2 (2017), pp. 1385–1418. ISSN: 0036-1410. DOI: 10.1137/15M1050264.

[CE10] G. Carlier and I. Ekeland. "Matching for Teams". In: *Economic theory* 42.2 (2010), pp. 397–418.

[CN08] G. Carlier and B. Nazaret. "Optimal Transportation for the Determinant". In: *ESAIM. Control, Optimisation and Calculus of Variations* 14.4 (2008), pp. 678–698. ISSN: 1292-8119. DOI: 10.1051/cocv:2008006.

[CPT23] G. Carlier, P. Pegon, and L. Tamanini. "Convergence Rate of General Entropic Optimal Transport Costs". In: *Calculus of Variations and Partial Differential Equations* 62.4 (2023), p. 116. ISSN: 1432-0835. DOI: 10.1007/s00526-023-02455-0.

[CC84] J. T. Chayes and L. Chayes. "On the Validity of the Inverse Conjecture in Classical Density Functional Theory". In: *Journal of Statistical Physics* 36.3 (1984), pp. 471–488. ISSN: 1572-9613. DOI: 10.1007/BF01010992.

[CCL84] J. T. Chayes, L. Chayes, and E. H. Lieb. "The Inverse Problem in Classical Statistical Mechanics". In: *Communications in Mathematical Physics* 93.1 (1984), pp. 57–121. ISSN: 1432-0916. DOI: 10.1007/BF01218639.

[CMN10] P.-A. Chiappori, R. McCann, and L. Nesheim. "Hedonic Price Equilibria, Stable Matching and Optimal Transport; Equivalence, Topology and Uniqueness." In: *Econom. Theory.* 42.2 (2010), pp. 317–354.

[CMP17] P.-A. Chiappori, R. J. McCann, and B. Pass. "Multi- to One-Dimensional Optimal Transport". In: *Communications on Pure and Applied Mathematics* 70.12 (2017), pp. 2405–2444. ISSN: 0010-3640. DOI: 10.1002/cpa.21707.

[Chi+16] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. *Scaling Algorithms for Unbalanced Transport Problems*. http://arxiv.org/abs/1607.05816, 2016.

[CDD15] M. Colombo, L. De Pascale, and S. Di Marino. "Multimarginal Optimal Transport Maps for One-Dimensional Repulsive Costs". In: *Canadian Journal of Mathematics* 67.2 (2015), pp. 350–368. ISSN: 0008-414X. DOI: 10.4153/CJM-2014-011-x.

[CD15] M. Colombo and S. Di Marino. "Equality between Monge and Kantorovich Multimarginal Problems with Coulomb Cost". In: *Annali di Matematica Pura ed Applicata (1923 -)* 194.2 (2015), pp. 307–320. ISSN: 1618-1891. DOI: 10.1007/s10231-013-0376-0.

[CDS19] M. Colombo, S. Di Marino, and F. Stra. "Continuity of Multimarginal Optimal Transport with Repulsive Cost". In: *SIAM Journal on Mathematical Analysis* 51.4 (2019), pp. 2903–2926. ISSN: 0036-1410. DOI: `10.1137/19M123943X`.

[CS16] M. Colombo and F. Stra. "Counterexamples in Multimarginal Optimal Transport with Coulomb Cost and Spherically Symmetric Data". In: *Mathematical Models & Methods in Applied Sciences* 26.6 (2016), pp. 1025–1049. ISSN: 0218-2025. DOI: `10.1142/S021820251650024X`.

[CM94] R. Cominetti and J. S. Martín. "Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming". In: *Mathematical Programming* 67.1 (1994), pp. 169–187. ISSN: 1436-4646. DOI: `10.1007/BF01582220`.

[CFK13] C. Cotar, G. Friesecke, and C. Klüppelberg. "Density Functional Theory and Optimal Transportation with Coulomb Cost". In: *Communications on Pure and Applied Mathematics* 66.4 (2013), pp. 548–599. ISSN: 1097-0312. DOI: `10.1002/cpa.21437`.

[CFK18] C. Cotar, G. Friesecke, and C. Klüppelberg. "Smoothing of Transport Plans with Fixed Marginals and Rigorous Semiclassical Limit of the Hohenberg-Kohn Functional". In: *Archive for Rational Mechanics and Analysis* 228.3 (2018), pp. 891–922. ISSN: 0003-9527. DOI: `10.1007/s00205-017-1208-y`.

[CP19] C. Cotar and M. Petrache. "Next-Order Asymptotic Expansion for N-marginal Optimal Transport with Coulomb and Riesz Costs". In: *Advances in Mathematics* 344 (2019), pp. 137–233. ISSN: 0001-8708. DOI: `10.1016/j.aim.2018.12.008`.

[Csi75] I. Csiszar. "*I*-Divergence Geometry of Probability Distributions and Minimization Problems". In: *The Annals of Probability* 3.1 (1975), pp. 146–158. ISSN: 0091-1798, 2168-894X. DOI: `10.1214/aop/1176996454`.

[Cut13] M. Cuturi. "Sinkhorn Distances: Lightspeed Computation of Optimal Transport". In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc., 2013.

[Del22] A. Delalande. "Nearly Tight Convergence Bounds for Semi-discrete Entropic Optimal Transport". In: *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*. International Conference on Artificial Intelligence and Statistics. PMLR, 2022, pp. 1619–1642.

[DGN17] S. Di Marino, A. Gerolin, and L. Nenna. "Optimal Transportation Theory with Repulsive Costs". In: *Topological Optimization and Optimal Transport*. Radon Series on Computational and Applied Mathematics. https://www.degruyter.com/view/books/978311043041 010/9783110430417-010.xml: De Gruyter, 2017.

[DGN23] S. Di Marino, A. Gerolin, and L. Nenna. "Universal Diagonal Estimates for Minimizers of the Levy-Lieb Functional". In: *Letters in Mathematical Physics* (2023). DOI: `10.48550/arXiv.2303.00496`.

[DLN22] S. Di Marino, M. Lewin, and L. Nenna. *Grand-Canonical Optimal Transport*. 2022. URL: `http://arxiv.org/abs/2201.06859`. preprint.

[DS14a] Y. Dolinsky and H. M. Soner. "Martingale Optimal Transport and Robust Hedging in Continuous Time". In: *Probability Theory and Related Fields* 160.1 (2014), pp. 391–427.

[DS14b] Y. Dolinsky and H. M. Soner. “Robust Hedging with Proportional Transaction Costs”. In: *Finance and Stochastics* 18.2 (2014), pp. 327–347.

[EN22] S. Eckstein and M. Nutz. *Convergence Rates for Regularized Optimal Transport via Quantization*. 2022. URL: http://arxiv.org/abs/2208.14391.

[EN23] V. Ehrlacher and L. Nenna. *A Sparse Approximation of the Lieb Functional with Moment Constraints*. 2023. URL: http://arxiv.org/abs/2306.00806. preprint.

[Eke05] I. Ekeland. “An Optimal Matching Problem”. In: *ESAIM Control. Optim. Calc. Var.* 11 (2005), pp. 57–71.

[EGH12] I. Ekeland, A. Galichon, and M. Henry. “Comonotonic Measures of Multivariate Risks”. In: *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* 22.1 (2012), pp. 109–132.

[EP06] P. Embrechts and G. Puccetti. “Bounds for Functions of Multivariate Risks”. In: *Journal of Multivariate Analysis* 97.2 (2006), pp. 526–547. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2005.04.001.

[EWW15] P. Embrechts, B. Wang, and R. Wang. “Aggregation-Robustness and Model Uncertainty of Regulatory Risk Measures”. In: *Finance and Stochastics* 19.4 (2015), pp. 763–790. ISSN: 0949-2984. DOI: 10.1007/s00780-015-0273-z.

[Enn+22] H. Ennaji, Q. Mérigot, L. Nenna, and B. Pass. “Robust Risk Management via Multi-Marginal Optimal Transport”. In: *to appear on Journal of Optimization Theory and Applications* (2022). DOI: 10.48550/ARXIV.2211.07694.

[Eva79] R. Evans. “The Nature of the Liquid-Vapour Interface and Other Topics in the Statistical Mechanics of Non-Uniform, Classical Fluids”. In: *Advances in Physics* 28.2 (1979), pp. 143–200. ISSN: 0001-8732. DOI: 10.1080/00018737900101365.

[Fig10] A. Figalli. “The Optimal Partial Transport Problem”. In: *Archive for Rational Mechanics and Analysis* 195.2 (2010), pp. 533–560. ISSN: 0003-9527. DOI: 10.1007/s00205-008-0212-7.

[FKM11] A. Figalli, Y.-H. Kim, and R. J. McCann. “When Is Multidimensional Screening a Convex Program?” In: *Journal of Economic Theory* 146.2 (2011), pp. 454–478. ISSN: 0022-0531. DOI: 10.1016/j.jet.2010.11.006.

[FG97] H. Föllmer and N. Gantert. “Entropy Minimization and Schrödinger Processes in Infinite Dimensions”. In: *The Annals of Probability* 25.2 (1997), pp. 901–926. ISSN: 0091-1798, 2168-894X. DOI: 10.1214/aop/1024404423.

[FKN16] R. L. Frank, R. Killip, and P. T. Nam. “Nonexistence of Large Nuclei in the Liquid Drop Model”. In: *Letters in Mathematical Physics* 106.8 (2016), pp. 1033–1036. ISSN: 1573-0530. DOI: 10.1007/s11005-016-0860-8.

[FNV18] R. L. Frank, P. T. Nam, and H. Van Den Bosch. “The Ionization Conjecture in Thomas–Fermi–Dirac–von Weizsäcker Theory”. In: *Communications on Pure and Applied Mathematics* 71.3 (2018), pp. 577–614. ISSN: 1097-0312. DOI: 10.1002/cpa.21717.

[FL89] J. Franklin and J. Lorenz. “On the Scaling of Multidimensional Matrices”. In: *Linear Algebra and its Applications*. Special Issue Dedicated to Alan J. Hoffman 114–115 (1989), pp. 717–735. ISSN: 0024-3795. DOI: 10.1016/0024-3795(89)90490-4.

[FP23] G. Friesecke and M. Penka. *Convergence Proof for the GenCol Algorithm in the Case of Two-Marginal Optimal Transport.* 2023. URL: http://arxiv.org/abs/2303.07137. preprint.

[FP22] G. Friesecke and M. Penka. *The GenCol Algorithm for High-Dimensional Optimal Transport: General Formulation and Application to Barycenters and Wasserstein Splines.* 2022. URL: http://arxiv.org/abs/2209.09081. preprint.

[FSV22] G. Friesecke, A. S. Schulz, and D. Vögler. "Genetic Column Generation: Fast Computation of High-Dimensional Multimarginal Optimal Transport Problems". In: *SIAM Journal on Scientific Computing* 44.3 (2022), A1632–A1654. ISSN: 1064-8275. DOI: 10.1137/21M140732X.

[Gal16] A. Galichon. *Optimal Transport Methods in Economics.* 1st ed. Princeton University Press, 2016.

[Gal+05] A. Gallegos, R. Carbó-Dorca, F. Lodier, E. Cancès, and A. Savin. "Maximal Probability Domains in Linear Molecules". In: *Journal of Computational Chemistry* 26.5 (2005), pp. 455–460. ISSN: 1096-987X. DOI: 10.1002/jcc.20180.

[GŚ98] W. Gangbo and A. Święch. "Optimal Maps for the Multidimensional Monge-Kantorovich Problem". In: *Communications on Pure and Applied Mathematics* 51.1 (1998), pp. 23–45. ISSN: 1097-0312. DOI: 10.1002/(SICI)1097-0312(199801)51:1<23::AID-CPA2>3.0.CO;2-H.

[Gar22] L. Garrigue. "Building Kohn–Sham Potentials for Ground and Excited States". In: *Archive for Rational Mechanics and Analysis* 245.2 (2022), pp. 949–1003. ISSN: 1432-0673. DOI: 10.1007/s00205-022-01804-1.

[GKR20] A. Gerolin, A. Kausamo, and T. Rajala. "Multi-Marginal Entropy-Transport with Repulsive Cost". In: *Calculus of Variations and Partial Differential Equations* 59.3 (2020), p. 90. ISSN: 1432-0835. DOI: 10.1007/s00526-020-01735-3.

[GHS20] M. Ghossoub, J. Hall, and D. Saunders. "Maximum Spectral Measures of Risk with given Risk Factor Marginal Distributions". 2020.

[GMP16] F. Golse, C. Mouhot, and T. Paul. "On the Mean Field and Classical Limits of Quantum Mechanics". In: *Communications in Mathematical Physics* 343.1 (2016), pp. 165–205. ISSN: 1432-0916. DOI: 10.1007/s00220-015-2485-7.

[Haa+21] I. Haasler, R. Singh, Q. Zhang, J. Karlsson, and Y. Chen. "Multi-Marginal Optimal Transport and Probabilistic Graphical Models". In: *IEEE Transactions on Information Theory* 67.7 (2021), pp. 4647–4668. DOI: 10.1109/TIT.2021.3077465.

[Hei02] H. Heinich. "Problème de Monge Pour n Probabilités". In: *Comptes Rendus Mathematique. Academie des Sciences. Paris* 334.9 (2002), pp. 793–795. ISSN: 1631-073X. DOI: 10.1016/S1631-073X(02)02341-5.

[HK64] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Physical review* 136 (3B 1964), B864.

[IN18] N. Igbida and V. T. Nguyen. "Augmented Lagrangian Method for Optimal Partial Transportation". In: *IMA Journal of Numerical Analysis* 38.1 (2018), pp. 156–183. ISSN: 0272-4979. DOI: 10.1093/imanum/drw077.

[IL15] B. Iooss and P. Lemaître. "A Review on Global Sensitivity Analysis Methods". In: *Uncertainty management in simulation-optimization of complex systems* (2015), pp. 101–122.

[JKT23] S. Jansen, T. Kuna, and D. Tsagkarogiannis. "Virial Inversion and Density Functionals". In: *Journal of Functional Analysis* 284.1 (2023), p. 109731. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2022.109731.

[JLM23] M. Jex, M. Lewin, and P. S. Madsen. "Classical Density Functional Theory: Representability and Universal Bounds". In: *Journal of Statistical Physics* 190.4 (2023), p. 73. ISSN: 1572-9613. DOI: 10.1007/s10955-023-03086-7.

[JKO98] R. Jordan, D. Kinderlehrer, and O. Otto. "The Variational Formulation of the Fokker-Planck Equation". In: *SIAM journal on mathematical analysis* 29.1 (1998), pp. 1–17.

[Kel] H. G. Kellerer. *Duality Theorems for Marginal Problems | SpringerLink*. URL: https://link.springer.com/article/10.1007/BF00532047.

[Kho+20] Y. Khoo, L. Lin, M. Lindsey, and L. Ying. "Semidefinite Relaxation of Multimarginal Optimal Transport for Strictly Correlated Electrons in Second Quantization". In: *SIAM Journal on Scientific Computing* 42.6 (2020), B1462–B1489. ISSN: 1064-8275. DOI: 10.1137/20M1310977.

[KY19] Y. Khoo and L. Ying. "Convex Relaxation Approaches for Strictly Correlated Density Functional Theory". In: *SIAM Journal on Scientific Computing* 41.4 (2019), B773–B795.

[KM10] Y.-H. Kim and R. J. McCann. "Continuity, Curvature, and the General Covariance of Optimal Transportation". In: *Journal of the European Mathematical Society* 12.4 (2010), pp. 1009–1040. ISSN: 1435-9855. DOI: 10.4171/jems/221.

[KP14a] Y.-H. Kim and B. Pass. "A General Condition for Monge Solutions in the Multi-Marginal Optimal Transport Problem". In: *Siam Journal On Mathematical Analysis* 46.2 (2014), pp. 1538–1550. ISSN: 0036-1410. DOI: 10.1137/130930443.

[KP15] Y.-H. Kim and B. Pass. "Multi-Marginal Optimal Transport on Riemannian Manifolds". In: *American Journal of Mathematics* 137.4 (2015), pp. 1045–1060. ISSN: 0002-9327. DOI: 10.1353/ajm.2015.0024.

[KP14b] Y.-H. Kim and B. Pass. "Wasserstein Barycenters over Riemannian Manfolds". 2014.

[KPS20] Y.-H. Kim, B. Pass, and D. J. Schneider. "Optimal Transport and Barycenters for Dendritic Measures". In: *Pure and Applied Analysis* 2.3 (2020), pp. 581–601.

[KP14c] J. Kitagawa and B. Pass. "The Multi-Marginal Optimal Partial Transport Problem". 2014.

[Lel22] R. Lelotte. *An External Dual Charge Approach to the Optimal Transport with Coulomb Cost*. 2022. URL: http://arxiv.org/abs/2208.14762. preprint.

[Léo14] C. Léonard. "A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport". In: *Discrete and Continuous Dynamical Systems* 34.4 (2014), p. 1533. DOI: 10.3934/dcds.2014.34.1533.

[Lev79] M. Levy. “Universal Variational Functionals of Electron Densities, First-Order Density Matrices, and Natural Spin-Orbitals and Solution of the v-Representability Problem”. In: *Proceedings of the National Academy of Sciences* 76.12 (1979), pp. 6062–6065. DOI: 10.1073/pnas.76.12.6062.

[Lew18] M. Lewin. “Semi-Classical Limit of the Levy-Lieb Functional in Density Functional Theory”. In: *Comptes Rendus Mathematique. Academie des Sciences. Paris* 356.4 (2018), pp. 449–455. DOI: 10.1016/j.crma.2018.03.002.

[LLS18] M. Lewin, E. H. Lieb, and R. Seiringer. “Statistical Mechanics of the Uniform Electron Gas”. In: *Journal de l'École polytechnique — Mathématiques* 5 (2018), pp. 79–116. ISSN: 2270-518X. DOI: 10.5802/jep.64.

[LLS19] M. Lewin, E. H. Lieb, and R. Seiringer. “The Local Density Approximation in Density Functional Theory”. In: *Pure and Applied Analysis* 2.1 (2019), pp. 35–73. ISSN: 2578-5885. DOI: 10.2140/paa.2020.2.35.

[LLS23] M. Lewin, E. H. Lieb, and R. Seiringer. “Universal Functionals in Density Functional Theory”. In: *Density Functional Theory: Modeling, Mathematical Analysis, Computational Methods, and Applications*. Mathematics and Molecular Modeling. Cham: Springer International Publishing, 2023, pp. 115–182. ISBN: 978-3-031-22340-2. DOI: 10.1007/978-3-031-22340-2_3.

[Lie83] E. H. Lieb. “Density Functionals for Coulomb Systems”. In: *International Journal of Quantum Chemistry* 24.3 (1983), pp. 243–277.

[Lie84] E. H. Lieb. “Bound on the Maximum Negative Ionization of Atoms and Molecules”. In: *Physical Review A* 29.6 (1984), pp. 3018–3028. DOI: 10.1103/PhysRevA.29.3018.

[LS10] E. H. Lieb and R. Seiringer. *The Stability of Matter in Quantum Mechanics*. Cambridge University Press, 2010. 310 pp. ISBN: 978-0-521-19118-0.

[MS23] H. Malamut and M. Sylvestre. *Convergence Rates of the Regularized Optimal Transport : Disentangling Suboptimality and Entropy*. 2023. URL: http://arxiv.org/abs/2306.06940.

[MG20] S. D. Marino and A. Gerolin. “An Optimal Transport Approach for the Schrödinger Bridge Problem and Convergence of Sinkhorn Algorithm”. In: *Journal of Scientific Computing* 85.2 (2020), p. 27. ISSN: 1573-7691. DOI: 10.1007/s10915-020-01325-7.

[McC97] R. J. McCann. “A Convexity Principle for Interacting Gases”. In: *Advances in Mathematics* 128.1 (1997), pp. 153–179.

[MP20] R. J. McCann and B. Pass. “Optimal Transportation Between Unequal Dimensions”. In: *Archive for Rational Mechanics and Analysis* 238.3 (2020), pp. 1475–1520. ISSN: 1432-0673. DOI: 10.1007/s00205-020-01569-5.

[MFE15] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Revised. Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2015. xix+699. ISBN: 978-0-691-16627-8.

[MSS21] Q. Mérigot, F. Santambrogio, and C. SARRAZIN. “Non-Asymptotic Convergence Bounds for Wasserstein Approximation Using Point Clouds”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 12810–12821.

[MSG13] A. Mirtschink, M. Seidl, and P. Gori-Giorgi. “Derivative Discontinuity in the Strong-Interaction Limit of Density-Functional Theory”. In: *Physical Review Letters* 111.12 (2013), p. 126402. DOI: 10.1103/PhysRevLett.111.126402.

[MP17] A. Moameni and B. Pass. “Solutions to Multi-Marginal Optimal Transport Problems Concentrated on Several Graphs”. In: *ESAIM. Control, Optimisation and Calculus of Variations* 23.2 (2017), pp. 551–567. ISSN: 1292-8119. DOI: 10.1051/cocv/2016003.

[Mon81] G. Monge. *Mémoire Sur La Théorie Des Déblais et Des Remblais*. De l’Imprimerie Royale, 1781.

[Nen16] L. Nenna. “Numerical Methods for Multi-Marginal Optimal Transportation”. PhD thesis. Université Paris sciences et lettres, 2016.

[NP22] L. Nenna and B. Pass. “A Note on Cournot-Nash Equilibria and Optimal Transport between Unequal Dimensions”. In: *Optimal Transport Statistics for Economics and Related Topics*. Vol. 483. Studies in Systems, Decision and Control. arXiv, 2022. DOI: 10.48550/arXiv.2209.14888.

[NP23a] L. Nenna and B. Pass. *An ODE Characterisation of Multi-Marginal Optimal Transport with Pairwise Cost Functions*. 2023. URL: http://arxiv.org/abs/2212.12492. preprint.

[NP23b] L. Nenna and B. Pass. “Transport Type Metrics on the Space of Probability Measures Involving Singular Base Measures”. In: *Applied Mathematics & Optimization* 87.2 (2023), p. 28.

[NP20] L. Nenna and B. Pass. “Variational Problems Involving Unequal Dimensional Optimal Transport”. In: *Journal de Mathématiques Pures et Appliquées* 139 (2020), pp. 83–108.

[NP23c] L. Nenna and P. Pegon. “Convergence Rate of Entropy-Regularized Multi-Marginal Optimal Transport Costs”. In: *to appear on Canadian Journal of Mathematics* arXiv:2307.03023 (2023). DOI: 10.48550/arXiv.2307.03023. preprint.

[NW22] M. Nutz and J. Wiesel. “Stability of Schrödinger Potentials and Convergence of Sinkhorn’s Algorithm”. 2022. DOI: 10.48550/arXiv.2201.10059.

[Pas15] B. Pass. “Multi-Marginal Optimal Transport: Theory and Applications”. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 49.6 (6 2015), pp. 1771–1790. ISSN: 0764-583X, 1290-3841. DOI: 10.1051/m2an/2015020.

[Pas12] B. Pass. “On the Local Structure of Optimal Measures in the Multi-Marginal Optimal Transportation Problem”. In: *Calculus of Variations and Partial Differential Equations* 43.3 (2012), pp. 529–536. ISSN: 1432-0835. DOI: 10.1007/s00526-011-0421-z.

[Pas11a] B. Pass. “Structural Results on Optimal Transportation Plans”. PhD thesis. University of Toronto, 2011.

[Pas11b] B. Pass. “Uniqueness and Monge Solutions in the Multimarginal Optimal Transportation Problem”. In: *SIAM Journal on Mathematical Analysis* 43.6 (2011), pp. 2758–2775. ISSN: 0036-1410. DOI: 10.1137/100804917.

[PV22] B. Pass and A. Vargas-Jiménez. “Monge Solutions and Uniqueness in Multi-Marginal Optimal Transport: Weaker Conditions on the Cost, Stronger Conditions on the Marginals”. 2022.

[PV21] B. W. Pass and A. Vargas-Jiménez. “Multi-Marginal Optimal Transportation Problem for Cyclic Costs”. In: *Siam Journal On Mathematical Analysis* 53.4 (2021), pp. 4386–4400. ISSN: 0036-1410. DOI: 10.1137/19M130889X.

[PC19] G. Peyré and M. Cuturi. “Computational Optimal Transport: With Applications to Data Science”. In: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000073.

[Puc13] G. Puccetti. “Sharp Bounds on the Expected Shortfall for a Sum of Dependent Random Variables”. In: *Statistics & Probability Letters* 83.4 (2013), pp. 1227–1232.

[PR13] G. Puccetti and L. Rüschendorf. “Sharp Bounds for Sums of Dependent Risks”. In: *Journal of Applied Probability* 50.1 (2013), pp. 42–53. ISSN: 0021-9002. DOI: 10.1239/jap/1363784423.

[Rab+11] J. Rabin, G. Peyré, J. Delon, and M. Bernot. “Wasserstein Barycenter and Its Application to Texture Mixing”. In: *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer. 2011, pp. 435–446.

[RC98] J.-C. Rochet and P. Choné. “Ironing, Sweeping, and Multidimensional Screening”. In: *Econometrica* 66.4 (1998), pp. 783–826. ISSN: 0012-9682. DOI: 10.2307/2999574.

[Rue99] Ruelle. “The States of Statistical Mechanics”. In: *Statistical Mechanics*. CO-PUBLISHED WITH IMPERIAL COLLEGE PRESS, 1999, pp. 168–197. ISBN: 978-981-02-3862-9. DOI: 10.1142/9789812815286_0007.

[RT98] L. Rüschendorf and W. Thomsen. “Closedness of Sum Spaces and the Generalized Schrödinger Problem”. In: *Theory of Probability & Its Applications* 42.3 (1998), pp. 483–494. ISSN: 0040-585X. DOI: 10.1137/S0040585X97976301.

[Rüs06] L. Rüschendorf. “Law Invariant Convex Risk Measures for Portfolio Vectors”. In: *Statistics & Decisions. International Journal for Statistical Theory and Related Fields* 24.1 (2006), pp. 97–108. ISSN: 0721-2631. DOI: 10.1524/stnd.2006.24.1.97.

[Rüs13] L. Rüschendorf. *Mathematical Risk Analysis*. Springer Series in Operations Research and Financial Engineering. Springer, Heidelberg, 2013. xii+408. ISBN: 978-3-642-33589-1 978-3-642-33590-7. DOI: 10.1007/978-3-642-33590-7.

[San15] F. Santambrogio. *Optimal Transport for Applied Mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and Their Applications. Cham: Springer International Publishing, 2015. ISBN: 978-3-319-20828-2. DOI: 10.1007/978-3-319-20828-2.

[Sar22] C. Sarrazin. “Lagrangian Discretization of Variational Mean Field Games”. In: *SIAM Journal on Control and Optimization* 60.3 (2022), pp. 1365–1392. ISSN: 0363-0129. DOI: 10.1137/20M1377291.

[Sei+17] M. Seidl, S. Di Marino, A. Gerolin, L. Nenna, K. J. H. Giesbertz, and P. Gori-Giorgi. “The Strictly-Correlated Electron Functional for Spherically Symmetric Systems Revisited”. In: *Physical Review A: Atomic, Molecular, and Optical Physics* (2017).

[SGS07] M. Seidl, P. Gori-Giorgi, and A. Savin. “Strictly Correlated Electrons in Density-Functional Theory: A General Formulation with Applications to Spherical Densities”. In: *Physical Review A* 75.4 (2007), p. 042511. DOI: `10.1103/PhysRevA.75.042511`.

[Sim11] B. Simon. “Convexity”. In: *(No Title)* (2011). DOI: `10.1017/cbo9780511910135`.

[Sim00] B. Simon. “Schrödinger Operators in the Twenty-First Century”. In: *Mathematical physics* 2000 (2000), pp. 283–288.

[Sin64] R. Sinkhorn. “A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices”. In: *The Annals of Mathematical Statistics* 35.2 (1964), pp. 876–879. ISSN: 0003-4851, 2168-8990. DOI: `10.1214/aoms/1177703591`.

[Sin67] R. Sinkhorn. “Diagonal Equivalence to Matrices with Prescribed Row and Column Sums”. In: *The American Mathematical Monthly* 74.4 (1967), pp. 402–405.

[TJK22] N. G. Trillos, M. Jacobs, and J. Kim. *The Multimarginal Optimal Transport Formulation of Adversarial Multiclass Classification.* arXiv, 2022. DOI: `10.48550/ARXIV.2204.12676`. URL: `https://arxiv.org/abs/2204.12676`.

[Vil09] C. Villani. *Optimal Transport: Old and New.* Vol. 338. Grundlehren Der Mathematischen Wissenschaften. Springer, 2009.

[Vil03] C. Villani. *Topics in Optimal Transportation.* Vol. 58. Graduate Studies in Mathematics. American Mathematical Society, Providence, 2003.

[WW11] B. Wang and R. Wang. “The Complete Mixability and Convex Minimization Problems with Monotone Marginal Densities”. In: *Journal of Multivariate Analysis* 102.10 (2011), pp. 1344–1360. ISSN: 0047-259X. DOI: `10.1016/j.jmva.2011.05.002`.

[Wan+13] W. Wang, D. Slepčev, S. Basu, J. A. Ozolek, and G. K. Rohde. “A Linear Optimal Transportation Framework for Quantifying and Visualizing Variations in Sets of Images”. In: *International Journal of Computer Vision* 101.2 (2013), pp. 254–269. ISSN: 0920-5691. DOI: `10.1007/s11263-012-0566-z`.

[Wee18] J. Weed. “An Explicit Analysis of the Entropic Penalty in Linear Programming”. In: *Proceedings of the 31st Conference On Learning Theory.* Conference On Learning Theory. PMLR, 2018, pp. 1841–1855.

[You82] L.-S. Young. “Dimension, Entropy and Lyapunov Exponents”. In: *Ergodic Theory and Dynamical Systems* 2.1 (1982), pp. 109–124. ISSN: 1469-4417, 0143-3857. DOI: `10.1017/S0143385700009615`.

**Titre :** Sur quelques généralisations du problème du transport optimal

**Mots clés :** Transport optimal multi-matinales, TFD, Grand canonique, Fonctionnelle de Lieb, métriques, transport optimal

**Résumé :** Ce manuscrit est consacré à l'étude de quelques généralisations du problème de transport optimal, ainsi qu'à des applications dans les domaines de la finance mathématique, de la théorie des jeux et de la physique quantique. Plus précisément, nous commençons par étudier le problème du transport optimal multi-marginales (TOM) en affinant la théorie existante pour le cas unidimensionnel. Nous fournissons une caractérisation explicite des mesures de risque spectrales. Nous étudions également la régularisation entropique de MMOT et en particulier, nous dérivons des taux de convergence du coût entropique vers le coût multi-marginal lorsque le paramètre de régularisation tend vers zéro. De plus, dans le cadre du transport entropique, nous proposons une caractérisation du problème discret en utilisant une équation différentielle. En ce qui concerne l'application du transport optimal à la théorie de la fonctionnelle de densité (TFD), nous proposons une généralisation du TOMM à l'ensemble canonique grand, c'est-à-dire que nous autorisons le nombre de marginales à varier. De plus, nous nous concentrons également sur une approximation parcimonieuse de la fonctionnelle de Lieb via des contraintes de moments. Enfin, nous étudions le problème du transport optimal entre dimensions différentes et certains problèmes variationnels issus de la théorie des jeux et de l'économie, tels que les équilibres de Cournot-Nash ou le problème de hedonic pricing. La dernière partie du manuscrit est consacrée au développement de la théorie de certaines métriques impliquant des mesures singulières.

**Title :** On some generalisations of Optimal Transport problem

**Keywords :** Multi-Marginal Optimal Transport, DFT, Grand Canonical, Lieb functional, metrics, Optimal Transport

**Abstract :** This manuscript is devoted to the study of some generalisation of Optimal Transport problem as well as some applications arising in Mathematical Finance, Game Theory and Quantum Physics. More specifically, we first consider the Multi-Marginal Optimal Transport (MMOT) problem by a careful refinement of the existing theory for the one dimensional case, we provide an explicit characterisation of spectral risk measures. We also study the entropic regularization of MMOT and in particular derive rate of convergence of the entropic cost to the multi-marginal cost as the regularization parameter vanishes. Moreover, in the framework of entropic transport we provide an ODE characterisation of the discrete problem. Concerning the application of Optimal Transport to Density Functional Theory (DFT) we propose a generalisation of MMOT to the grand canonical ensemble, that is we allow the number of marginals to vary. Furthermore, we also focus on a sparse approximation of the Lieb functional via moments constraints. Finally, we study unequal dimensional Optimal Transport and some related variational problems arising in Game Theory and Economics as, for instance, Cournot-Nash equilibria or the hedonic pricing problem. The last part of the manuscript is then devoted to develop the theory of some metrics involving singular measures