Multidimensional martingale optimal transport.

Hadrien de March

université
PARIS-SACLAY

NNT : 2018SACLX042

Thèse de doctorat

Transport optimal de martingale multidimensionnel

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'École Polytechnique

Ecole doctorale n°142 Ecole doctorale de mathématiques Hadamard (EDMH)
Spécialité de doctorat : Mathématiques appliquées

Thèse présentée et soutenue à Palaiseau, le 29 juin 2018, par

HADRIEN DE MARCH

Composition du Jury :

Guillaume Carlier Professeur, Université Paris-Dauphine (CEREMADE)	Président
Benjamin Jourdain Professeur, École des Ponts Paritech (CERMICS)	Rapporteur
Walter Schachermayer Professeur, University of Vienna	Rapporteur
Pierre Henry-Labordère Chercheur, Société Générale (SGCIB)	Examinateur
Sylvie Méléard Professeur, École Polytechnique (CMAP)	Examinateur
Nizar Touzi Professeur, École Polytechnique (CMAP)	Directeur de thèse

Transport Optimal de Martingale Multidimensionnel

Hadrien De March

CMAP
École Polytechnique

This dissertation is submitted for the degree of
Docteur en mathématiques appliquées de l'École Polytechnique

Université Paris-Saclay December 2018

À Alix et Maubert…

Déroulement de la thèse

De tous temps les hommes se sont posé des questions. Certaines pouvant être qualifiées d'utiles, d'autres moins. Quand on s'engage dans la noble voie des mathématiques, on reste généralement éloigné de la première catégorie. Je n'ai pas fait exception à cette règle.

Aux origines de ces travaux figure une question à l'apparence simple.

"Vois-tu Hadrien, en une dimension et sous la condition de Spence-Mirlees, on montre que le transport optimal conditionnel se concentre sur deux graphes. Je voudrais que tu trouves une condition en dimension supérieure qui garantisse que le transport se concentre sur $d+1$ points. On sent que c'est vrai car sous un modèle optimal, on est en marché complet."

Suite à cet entretien initial avec mon directeur de thèse, Nizar Touzi, le labeur commença. D'arrachages de cheveux en épluchage de bibliographies, j'explorai de nombreuses pistes, toutes infructueuses. Les mois passèrent mais le problème restait entier, résistant à toutes formes d'assaut. En dimension $d = 2$, un maudit quatrième point semblait toujours s'inviter et gâcher la fête. Je rencontrai alors Tongseok Lim lors d'une école d'été au Mans. Tongseok était un jeune doctorant, élève de Nassif Ghoussoub à la University of British Columbia. Il venait d'achever un papier qui annonçait décrire la structure du transport optimal martingale multidimensionnel. Je ressentis à sa lecture la crainte terrible connue de tout chercheur : avait-il réussi à résoudre ce problème avant moi, réduisant ainsi à néant mes efforts ? À la stressante lecture de son papier je constatai qu'il avait prouvé le résultat dans le cas particulier où la mesure cible était atomique. Je réalisai que sa démonstration était très spécifique à ce cas particulier. Je redoublai alors d'efforts et crus plusieurs fois être parvenu à une preuve qui s'avérait toujours fausse.

Après 6 mois de thèse, Nizar accepta d'acter ma défaite et de me donner un autre problème à étudier. Il s'agissait d'étendre un résultat qu'il avait prouvé à un horizon en temps infini. J'étudiai ce problème avec le peu de passion que suscite l'idée d'emprunter les sentiers battus pour n'ajouter qu'une moindre pierre à l'édifice. Ainsi, une part de mon esprit continuait de ruminer le transport. Un jour, face à la difficulté de prouver le résultat de transport optimal martingale demandé par Nizar, je décidai de me simplifier la vie en prouvant qu'il était faux. Ce fut alors explosif, la fécondité gigantesque de cette nouvelle approche me submergea tant, que je trouvai un nouveau résultat chaque semaine. Cependant, je les présentai si mal à Nizar qu'il ne comprit pas leur intérêt. Il me proposa d'étudier un nouveau problème.

Ainsi commença l'étude qui allait mener aux Chapitres 2 et 3 de cette thèse : l'étude de la dualité. Nizar me montra ce qui se passait en une dimension, il m'incombait alors de l'étendre à la dimension supérieure. Je compris très tôt que pour comprendre le phénomène de partitionnement de l'espace pour les plans de transport, il fallait observer l'action de la différence des lois marginales sur les fonctions convexes. Cette action avait un effet dual, d'une part sur les lieux d'accessibilité des noyaux des plans de transport, d'autre part sur le contrôle des fonctions duales. Bien que cette idée soit la bonne, sa mise en œuvre technique posait une myriade de problèmes. C'est ainsi qu'après une année entière, la taille surcritique de cette œuvre m'obligea à la couper en deux parties : une première traitant uniquement de la décomposition de l'espace et la seconde traitant de la dualité.

C'est dans la rédaction de ce premier article que j'ai le plus travaillé avec Nizar. Malgré ses nombreuses responsabilités à ce moment dans le laboratoire, il parvint à me consacrer un temps certain. Au plus fort de la tempête, j'allai jusqu'à le rejoindre en Tunisie pour commencer la rédaction de l'article.

"Voilà comment nous nous organiserons : nous travaillerons ensemble le matin pendant que les enfants dorment et je m'occuperai de ma famille l'après-midi."

Puis après la rentrée, il m'accueillit dans l'atelier de sculpteur au fond de son jardin de Malakoff plusieurs samedis matin pour finaliser le travail, loin de la sollicitation incessante de ses responsabilités au laboratoire.

Ainsi le premier article fut écrit. Une première version fut mise en ligne en février 2017. Cette opération ne se fit pas sans inquiétude. Lors d'une conférence en janvier, nous échangeâmes avec Jan Obloj. Jan était un grand expert en transport martingale, professeur à l'université d'Oxford. Il affirmait que dans un travail conjoint avec Pietro Siorpaes, un jeune assistant professeur italien à l'Imperial College, il démontrait également l'existence des composantes irréductibles. Nous convînmes de mettre en ligne nos travaux en même temps en se citant mutuellement, belle pratique de gentlemen de la recherche, traditionnelle dans le domaine des mathématiques financières.

"Nous sommes si peu nombreux à travailler sur ces sujets, si on commence à se fâcher avec quelques-uns, on perd alors la moitié de notre auditoire !"

Finalement, les travaux de nos collègues étaient à un stade beaucoup moins avancé et il leur manquait quelques ingrédients cruciaux pour arriver à un résultat aussi abouti que le nôtre. Mon travail était sauf.

Suite à cette première mise en ligne, je fus invité à présenter cette œuvre au séminaire de calcul stochastique de Vienne où je commençai à travailler avec Mathias Beiglböck, professeur à l'université technologique de Vienne et expert internationnal reconnu du transport optimal martingale. Je fus également invité au séminaire de l'École des Ponts où je pus échanger avec Aurélien Alfonsi et Benjamin Jourdain.

Travailler sur ce projet avec Nizar m'a énormément appris, en particulier sur la rédaction à adopter et la quantité d'information à inclure dans les démonstrations. J'ai ainsi appris à structurer, qualité essentielle pour présenter des résultats à la technicité retorse. Fort de cette expérience je m'attelais à deux nouveaux articles : le Chapitre 3 qui concerne la dualité et le Chapitre 4 qui reprenait et nettoyait mes travaux illisibles du début de thèse qui donnaient la structure des noyaux de transport martingale optimaux. Cet ordre était naturel car le résultat de dualité s'appuyait sur la décomposition en composantes irréductibles et le résultat de structure prenait sa source dans la dualité.

En travaillant sur le résultat de structure, je fus amené à étudier des mathématiques très éloignées de mon domaine : la géomètrie algébrique. Je pus apprendre les rudiments de cet art avec des chercheurs du deuxième laboratoire de mathématiques de Polytechnique, le Centre de Mathématiques Laurent Schwartz (CMLS), opportunément situé à proximité. J'échangeai avec René Mboro, élève de ma promotion spécialisé en géomètrie algébrique des corps finis. Il m'exposa patiemment les contre-intuitives notions de géomètrie algébrique, le théorème de Bézout ainsi que les grandes références classiques. Il m'orienta ensuite vers Erwan Brugallé, chargé de recherche à l'X en géométrie algébrique réelle qui répondit aimablement à mes premières interrogations. Il posa également mes questions plus délicates à la communauté de la géométrie algébrique réelle parisienne. J'échangeai également sur ce passionnant sujet avec Guillaume Cloître, alors doctorant à Villetaneuse, qui m'a initié à la beauté de la notion de schéma. Plus tard Nguyen-Bac Dang du CMLS relut également mon papier et me poussa à rédiger une preuve supplémentaire.

Deux années plus tard je les contactai suite à une erreur découverte par Benjamin Jourdain dans le Chapitre 4 de ce manuscrit. L'erreur semblait superficielle mais était en réalité profonde et portait sur des définitions de géométrie algébrique. Ils surent alors m'apporter les outils nécessaires à la résolution de mon problème et à la formulation d'un résultat nouveau permettant de répondre franchement par la négative à la question initiale de Nizar. En effet, pour presque toute fonction de coût régulière, on peut trouver des marginales qui donnent un transport optimal qui se décompose sur $d+2$ points en dimension paire.

Revenons cependant en juin 2017. J'eus alors une discussion avec Nizar :

"Hadrien, tu as de beaux résultats dans ta thèse, mais ils se cantonnent à un sujet très précis, dit-il, il serait bien que tu abordes un autre sujet un peu plus appliqué."

Je décidai alors de m'attaquer à la résolution numérique du problème. Comment aurais-je pu prétendre à un doctorat en mathématiques appliquées sans avoir fait d'application ? Je commençai alors une étude de l'état de l'art en transport optimal classique. Je profitai une nouvelle fois de la proximité du CMLS en allant y rencontrer un des plus grands experts mondiaux du transport optimal : Yann Brénier. Il eut la grande gentillesse de me faire un panorama assez exhaustif des méthodes existantes en transport classique, me vantant en particulier l'approche entropique. J'arrivai à la première conclusion erronée que cette approche entropique ne pouvait fonctionner que dans le cas du transport classique.

Xiaolu Tan, ancien doctorant de Nizar, aujourd'hui assistant professeur à Dauphine, avait travaillé dans sa thèse sur des méthodes numériques appliquées au transport optimal martingale en temps continu. Il me proposa un algorithme astucieux basé sur une dualisation partielle du problème. Je passai les semaines suivantes à coder cet algorithme, travaillant comme un forçat pour dénicher les bugs et trouver des méthodes d'optimisation numériques. Je passai ainsi l'école des probabilités de Saint-Flour à coder hors des cours, ne dormant que quelques heures chaque nuit. L'algorithme semblait converger, mais sa lenteur était désespérante. Je décidai d'utiliser le calcul parallèle dont l'efficacité m'avait laissé d'agréables souvenirs lors de mon expérience en banque. Je reçus alors l'aide de Romain Poncet, doctorant du CMAP qui participait à la même conférence à Téhéran et me partagea son expérience du numérique et du calcul parallèle. Cet art nécessitait une part d'ingénierie : un ensemble de recettes qui fonctionnent pour une raison qui peut être inconnue. De retour à l'X je pus bénéficier de puissants calculateurs multi-cœurs pouvant éxécuter jusqu'à 16 tâches en parallèle.

Malgré ces ressources démultipliées, la résolution demeurait trop lente. Pour trouver le meilleur algorithme à employer pour minimiser ma fonction convexe irrégulière, Éric Moulines me conseilla d'aller poser la question à Marco Cuturi, expert de la résolution numérique des problèmes de transport optimaux. Son laboratoire venait justement d'emménager avec l'ENSAE sur le plateau de Saclay à deux pas de Polytechnique. Sa rencontre me donna du grain à moudre et me relança sur la piste de l'approche entropique, car le meilleur moyen pour optimiser une fonction irrégulière est de l'approcher par une fonction régulière, comme me confirma un échange avec Charles-Albert Lehalle, professionnel habitué aux problèmes pratiques. Je finis par réaliser que l'approche entropique était applicable au cas martingale par une petite dose d'astuce. Elle s'avéra même d'une impressionnante efficacité. Elle permettait en effet cette régularisation que je recherchais. Après l'avoir confirmé expérimentalement, je poussai l'étude vers la recherche de meilleurs algorithmes sous la forme d'algorithmes de Newton. Je recodai même le package Python de ces algorithmes sous la recommandation de Bruno Levy pour mieux en choisir et contrôler le paramétrage. Ainsi naquit le Chapitre 5 de cette thèse.

Une nouvelle fois, j'eus la terrible surprise de constater, lorsqu'il m'invita à Oxford pour discuter de mes projets théoriques, que Jan Obloj travaillait sur le même sujet. Il collaborait avec mon grand frère de thèse Gaoyue Guo, parti en postdoctorat à Oxford. Fort heureusement je constatai qu'ils n'avaient guère exploré la direction entropique en pratique. Je pus ainsi, en réarticulant mes résultats, obtenir que mon travail prolonge le leur. Pour avoir de nouveaux résultats, je décidai de mettre au jour les vitesses de convergence des schémas numériques. Je repérai alors un étrange phénomène sur les simulations, l'erreur sur l'approximation entropique semblait beaucoup plus faible que ce que la théorie habituelle prévoyait. Je décidai alors d'étudier théoriquement ce phénomène et vis par une approche à la physicienne que la borne était universelle : elle ne dépendait ni de la fonction de coût, ni des lois marginales. La preuve de ce résultat extrêmement technique et le juste calibrage des hypothèses le concernant prit beaucoup plus de temps que je ne le pensais initialement, la preuve rigoureuse n'est pas encore terminée et nécéssite encore quelques précisions. Mais tout mathématicien sait que le diable se cache dans ce genre de détails. Il peut nous maintenir en échec sur des périodes souvent sous-estimées. La preuve de ce même résultat, n'existant pas dans la littérature pour le transport classique, aurait été beaucoup plus facile.

En conclusion cette thèse fut une expérience incroyablement enrichissante, aussi bien du point de vue scientifique, qu'humain. L'affrontement avec la difficulté irréductible. Le travail d'horloger qui doit sans cesse recréer et réorganiser son mécanisme. Ce fut une aventure qui me permit d'aller au fond des choses et de mobiliser toutes mes ressources tout en gérant les difficultés inhérentes à la frustration et à la lenteur. La phrase la plus présente dans mon esprit durant ces trois années fut "il est à présent temps d'en finir avec ce projet". Mais demeurent toujours des erreurs dissimulées dans des détails anodins. Si bien qu'à l'instant où un travail est effectivement achevé, le soulagement en devient incomparable.

Naviguant de l'exaltation à la dépression profonde, le seul moyen d'affronter cette maniaco-dépression endémique a été pour moi d'accueillir la contingence avec détachement et constance, tout en continuant à chercher et à me battre. Comme scandait Rudyard Kipling : "If you can meet with Triumph and Disaster, and treat those two impostors just the same".

Remerciements

Malgré la solitude qui lui est intrinsèque, le difficile travail de thèse repose sur l'interaction essentielle avec d'autres acteurs, comme le bref récit qui précède a voulu en témoigner. Je tiens ici à les remercier, en espérant n'oublier personne.

Je tiens en premier lieu à remercier Nizar Touzi, mon directeur de thèse, que j'avais rencontré pour la première fois à son cours de deuxième année de chaînes de Markov et martingale, un des cours que j'avais trouvé vraiment exaltants à l'X. L'année suivante je passai par son réseau pour obtenir un stage de recherche académique. "Je voudrais faire des maths dans un endroit anglophone". Mon vœu fut exaucé et je partis au Canada avec Matthieu Vermersch, un autre étudiant de ma promotion, pour faire un stage de recherche passionnant sur le risque systémique et les graphes aléatoires, sous la direction de Tom Hurd et Matheus Grasselli. Je remercie une nouvelle fois Nizar, Tom, Matheus et Matthieu pour cette expérience qui m'a donné un goût certain pour la recherche. Nizar m'a également permis de faire cette thèse malgré son démarrage chaotique lié au timing de mes stages de master 2 qui sont légèrement sortis des clous. Enfin pendant la thèse en elle-même Nizar a été très présent et m'a beaucoup aidé et tiré vers le haut, malgré ses nombreuses responsabilités et le fait qu'il parte chaque semaine aux quatre coins du monde. J'ai été très sensible au fait qu'il passe au moins deux jours chaque semaine à l'X pour travailler avec ses doctorants et ses postdoctorants et pour s'endormir à cause du décalage horaire devant leurs explications ennuyeuses.

Je me dois également de remercier ceux qui ont accepté d'être mes rapporteurs. Tout d'abord Walter Schachermayer dont la gentillesse et le style raffiné m'ont frappé dès notre première rencontre informelle au Bachelier Colloquium de Métabief. J'ai été par la suite impressionné par la clarté et l'intérêt communicatif que dégageaient ses exposés. Finalement je n'oublierai jamais le moment où, à la suite à mon exposé au séminaire de Vienne il m'avait glissé "this is beautiful" au sujet de mon travail. Je remercie également Benjamin Jourdain qui par sa relecture a trouvé de sensibles points d'amélioration pour les Chapitres 4 et 5. Je le remercie également pour son invitation au séminaire des Ponts où j'avais présenté le chapitre 2. À cette occasion, il me montra des résultats numériques qu'il avait obtenus qui m'inspirèrent une conjecture sur le nombre de graphes maximal sur lequel le transport optimal martingale se concentre en pratique.

Je remercie également ceux qui me font l'honneur d'être les membres de mon jury de thèse. En premier lieu Sylvie Méléard, connue de tous les Polytechniciens car elle est LA professeur de mathématiques appliquées du tronc commun. Son cours fut ma première vraie rencontre avec les probabilités et m'a immédiatement séduit. Plus tard je la revis au moment d'une conférence à Berlin ayant pour objectif de mêler biologie et finance. Je fut alors frappé par l'effort qu'elle fit pour comprendre mon poster, pourtant éloigné de son domaine habituel et j'en tirai un profond respect pour elle. Je remercie aussi Pierre Henry-Labordère, connu en tant qu'analyste quantitatif disruptant les mathématiques financières par ses connaissances en théorie des cordes. Il est un des créateurs initiaux du transport optimal martingale. Il a toujours foisonné d'idées originales en apparence insensées. J'ai eu l'honneur d'échanger avec lui au sujet des Chapitres 4 et 5 et d'une variante quantique du transport optimal de son invention. Il a aussi su poser les bonnes questions sur les Chapitres 2 et 3 : "mais en fait ça sert à quoi tout ça ?" Enfin je remercie Guillaume Carlier, co-créateur des barycentres de Wasserstein, sorte de "bonne notion" de moyenne pour plusieurs distributions de probabilité. Nos échanges concernant le transport optimal martingale se sont limités à la marge des pots de thèse de Gaoyue Guo et d'Habilitation à Diriger des Recherches de Xiaolu Tan, je suis donc heureux d'approfondir cette relation en le comptant dans mon jury.

Mes remerciements suivants vont à ceux qui m'ont aidé à comprendre la géométrie algébrique : René Mboro, Erwan Brugallé, Guillaume Cloître et Nguyen-Bac Dang. J'ai également une pensée pour Xiaolu Tan qui m'a aidé à étudier la sélection mesurable et qui m'a donné un algorithme pour résoudre le transport optimal numériquement. Je remercie également Yann Brénier, Éric Moulines, Marco Cuturi, Charles-Albert Lehalle et Olivier Levy pour m'avoir apporté leur aide concernant le numérique. Je remercie par ailleurs Daniel Bartl qui m'a montré l'extraordinaire notion qu'est la limite médiale. Merci à Nicole Spillane pour ses explications sur le préconditionnement.

Merci à Mathias Beiglböck, Nicolas Juillet, Aurélien Alfonsi et Jan Obloj pour m'avoir permis de présenter mes travaux lors de séminaires et de m'avoir encouragé à poursuivre dans cette voie. Merci à Marcel Nutz, Nacif Ghoussoub, Tongseok Lim, Mete Soner, Alexander Cox, Pietro Siorpaes, Martin Huesmann et Christian Léonard pour leur gentillesse et l'intérêt qu'ils ont porté à mes travaux lors de conversations privées. Je remercie aussi Beatrice Acciaio qui m'a expliqué pendant toute une soirée arrosée à Oxford que mon idée de principe de non-arbitrage n'intéresserait personne. Finalement un grand merci à Julio Backhoff qui m'a fait part de sa stratégie pour démontrer une version multidimensionnelle du théorème de Kellerer. Merci à Frédéric Bonnans pour ses cours de qualité et ses explications sur l'enveloppe convexe fonctionnelle. Merci à Cédric Villani pour la carté de son ouvrage de référence sur le transport.

Je remercie ensuite le personnel administratif, sans qui la vie serait beaucoup moins facile. Je remercie en premier lieu Nasséra Naar pour son éternelle gentillesse et bienveillance, ainsi que pour sa grande disponibilité, malgré la montagne de travail qu'elle a à abattre. De plus je suis conscient de ce que je lui dois quant au démarrage tardif de ma thèse. Je remercie Vincent Chapelier qui s'occupait de mes missions et m'a confié que j'étais le doctorant qui partait le plus, malgré son refus de me rembourser un reçu russe au motif des mentions "Bodka" et "Moxito" qui y figuraient. Je remercie bien entendu Wilfried Edmond pour son aide, mais aussi pour son calme, sa sérénité et son swag. Je tiens à remercier Alexandra Noiret pour la confiance absolue qu'elle m'a témoigné lors de mes dernières missions. Une pensée enfin pour Anne de Bouard qui a assumé pendant presque toute ma thèse le rôle de présidente du laboratoire. Je remercie également le reste des personnels administratifs avec lesquels j'ai moins interagi mais dont la contribution essentielle m'a naturellement impacté. Hors du laboratoire, je remercie Emmanuel Fullenwarth pour son aide à l'organisation de la soutenance, le centre Poly-Média et en particulier William Théodon, pour l'impression de ce manuscrit, ainsi que Audrey Lemaréchal, pour sa réactivité et son professionalisme.

Tout ceci ne serait pas arrivé si j'avais fait des choix différents. Merci à Younes Kchia, ancien doctorant de Nizar travaillant en tant qu'analyste quantitatif pour m'avoir dit que son seul regret sur sa thèse était de l'avoir terminée en deux ans et de ne pas en avoir profité une année de plus. Je remercie également Jelper Striet mon chef dans la banque pour ses mots emplis de sagesse : "les seules choses que nous regrettons sont celles que nous ne faisons pas". Je remercie Lorenzo Bergomi et Xavier Menguy de m'avoir encouragé dans mes décisions. Je dois aussi beaucoup à Goldman Sachs qui m'a énormément appris en un temps record sur des thèmes variés tout en m'offrant de quoi passer mon doctorat dans de bonnes conditions. Merci à Jayakiran Akurathi pour ses "shut the fuck up, s'il-vous-plaît", à Gavin Duffy pour avoir été "very sorry", à Takashi Hatanaka pour son whisky japonais, à Alan Zagury pour son restaurant, à Alain Griveau pour "junior trader is 100% of crap, but junior quant is only 50% of crap" et à François Rigou pour avoir cru en mon projet de doctorat. Je remercie également mes jeunes collègues Oriol Lozano et Hugo Forget, le premier pour avoir payé les 50 euros d'impôt pour lesquels l'administration fiscale Hong-Kongaise m'a poursuivi, et le second pour m'avoir lancé dans un projet de détection de bulles financières formateur. Je remercie mes camarades de master qui m'ont indirectement soutenu en faisant le même choix : Florentin Damiens et Simon Clinet.

Je remercie Therry Bodineau pour son cours de Saint-Flour et pour avoir pris garde au cours de ces trois ans que je ne dépasse aucune deadline.

Mes remerciements vont à ceux qui m'ont enseigné les mathématiques au fur et à mesure de ma scolarité, M. Dessertenne en première, Mme Robillard en terminale, Jean-Marc Wachter en première année de classe préparatoire et enfin Serge Francinou en mathématiques spécialités, qui n'a jamais su trouver son égal pour me lancer des chiffons remplis de craie au visage quand je mangeais en classe. Vincent Bansaye qui sans le savoir a éveillé un véritable goût chez moi pour les probabilités appliquées et fondamentales par le biais de son projet de première année "Pourquoi on attend autant le bus ?" Amandine Veber qui a encadré mon PSC sur la conquête de l'Ouest et à qui j'ai emprunté un livre traitant de statistiques sur une durée de 5 ans. Je remercie Stéphane Gaubert avec qui j'ai fait un projet d'Enseignement d'Approfondissement sur le rayon spectral tropical d'ensembles de matrices. Enfin je remercie mes professeurs de master 2, en particulier Gilles Pagès pour sa causticité, Philippe Bougerol pour son identité, Jean Bertoin pour sa lettre de recommandation inégalable "Je recommande Adrien De March qui a eu 20 à mon partiel, ce qui montre qu'il a parfaitement assimilé les notions de bases de la théorie des processus de Levy", Mathieu Rosenbaum pour son attachement à la ponctualité, Nicole El Karoui pour ses cours inoubliables et Emmanuel Gobet pour m'avoir autorisé à faire n'importe quoi.

Je dis également un grand merci à ma famille de thèse, le grand frère ainé Bruno Bouchard pour le jogging sous le soleil tunisien, Dylan Possamai pour avoir su montrer la voie et pour ses conseils, Zhen-Jie Ren pour sa vision positive des mathématiques, Gaoyue Guo pour ses enseignements sur le transport optimal martingale, Kaitong Hu pour sa coiffure exceptionnelle et enfin merci à Heythem Farhat pour sa réceptivité à mon enseignement et à ma sagesse sans borne, ainsi que pour sa relation fusionnelle avec les escabots dans les caves des boîtes shanghaiennes.

Finalement que serait une thèse sans des compagnons d'infortune ? Big up au bureau 20 16 : Massil Achab, parti trop tôt dans le bunker du big data, Antoine Hocquet et Aymeric Maury, illustres anciens de caractère, happés par l'achèvement de leur thèse. Merci à Perle Geoffroy, toujours prête à écouter mes plaintes et à m'apporter son soutien total et inconditionnel dans les moments difficiles. Jean-Bernard Eytard pour la qualité de sa conversation en des temps raisonnablement éloignés des compétitions sportives, pour ses brillantes participations à l'émission "des chiffres et des lettres" qui en fait un exemple à suivre pour nous tous. Florian Feppon pour son amour de l'Inde et des jeux de mots laids. Paul Thévenin pour ses calembours et son rire inimitable. Matthieu Kohli pour son complotisme amical et son amour de Jacques Attali. Paul Jusselin pour sa sympathie et sa performance au 10km. Cheikh Touré pour son inépuisable bonne humeur. Merci à Jing-jing Hao pour son art inimitable du "awkward moment". Raphael Forien pour sa force tranquille et son regard expressif, ainsi que pour le jeu de mot que je n'ai que peu osé dire tout haut : "Raphael, il ne Fo rien". Uladislau Stazhynski pour son don d'invisibilité. Cédric Rommel pour ses origines glorieuses. Julie Tourniaire pour ses belles tenues de sport. Nicolas Augier pour son implication. Merci à Pamela pour son entrain tout libanais, son talent pour le body combat et sa surveillance des vilains traders haute fréquence. Enfin mention spéciale à Pierre Cordesse pour son magistral ragequit du bureau pour cause de bruit.

Évidemment n'oublions pas l'équipe mathématiques financières du laboratoire, Mathieu Rosenbaum pour son aptitude à créer le traquenard en soirée et pour avoir remarqué que j'utilisait le logo du Centre de Médiation et d'Arbitrage de Paris pour mes présentations, Stefano De Marco pour son charme italien désuet, son nom qui sonne bien, son regard qui traversera de nombreuses générations d'étudiants, ainsi que sa réplique "je confirme !" le jour où il me surprit en train de dire que d'habitude je m'habillais n'importe comment. Merci à Thibaut Mastrolia pour son esprit mal tourné et son rat de compagnie. A Jocelyne Bion-Nadal pour les discussions autour du café du matin. Je n'oublie pas non plus les postdoctorants, merci à Julien Claisse pour son détachement, sa sympathie inconditionnelle et côté Guillaume Meurice, merci à Ankhush Agarwal pour sa maturité et son accent indien, merci à Mauro Rosestolato pour avoir écouté mes élucubrations sur la bonne utilisation de la récurrence transfinie sans broncher. À Omar El Euch pour sa goguenardise et ses incessants "alors, tu sais ce que tu fais après ta thèse ?". Merci à Othmane Mounjid pour sa bonne humeur perpétuelle et pour sa mesure en temps réel de mon IMC "tu as grossi/minci Hadrien". Merci au petit Thomas Ozello qui a su mettre un sel tout particulier dans l'école d'été en Russie. Merci à Gang Liu pour sa gentillesse dans ce corps de brute. Merci à Sigrid Kaalblad pour sa joie de vivre apparente, sa révélation de nombreux potins et sa tentative de nous emmener dans un restaurant végétarien gastronomique avec l'argent du séminaire de Vienne. Merci à Yiqing Lin pour la conférence à Shanghai, à Junjian Yang pour sa sociabilité poussée à l'extrémisme et à Alexandre Richard pour les longues conversations berlinoises. Merci également aux professionnels d'EDF, toujours rayonnants et intéressés par mon travail : René Aïd et Clémence Alasseur.

Mentionnons également dans le laboratoire Vianney Bœuf pour sa passion de l'administration, de la politique et du traditionnalisme, Joon Kwon pour sa capacité à lui donner la réplique, son poste d'agronome et sa passion pour les claviers mécaniques, Loïc Richier pour sa gentillesse et sa passion pour le motocrotte, Gustaw Matulewicz pour son besoin de cruncher la data avec des devices Apple, Tristan Roget pour son style inimitable et son accent du sud. Romain Poncet pour son sourire charmeur, ses connaissances en informatique et ses vidéos de condensats de Bose-Einstein. Merci à Martin Averseng pour son grand banditisme du café, ses paris stupides et son optimisme que même une troisième guerre mondiale au napalm ne saurait entacher. Céline Bonnet pour sa sollicitude. Merci à Ludovic Saccelli pour m'avoir pardonné d'avoir oublié son nom au moins huit fois. Mathilde Boissier pour son prénom idoine. Maxime Grangereau pour ses plaisanteries amusantes. Je remercie aussi Carl Graham de m'avoir rappelé une bonne dixaine de fois au petit déjeuner en voyant mes céréales de la marque Dukan que le docteur Dukan qui leur avait donné son nom avait été radié de l'ordre des médecins pour pratique illégale de la médecine. Un grand merci à Behlal Karimi pour ses unboxings de contrefaçons de la marque SUPREME, à Alexei pour montrer qu'il est possible de mener deux thèses en parallèle en publiant 7 articles, à Fedor Goncharov pour m'avoir excusé de parler français trop vite, à Geneviève Robin pour son amour de Sheryl Sandberg, à Kevish Napal pour son volume de cheveux exceptionnel, à Aline Marguet pour sa sérénité et à Jaouad Mourtada pour les belles couleurs de ses présentations. Merci à Giovanni Conforti pour nos échanges sur le transport optimal entropique. Sarah Kakai pour sa conversation agréable et passionnante, ainsi que les problèmes de probabilités qu'elle m'a soumis. Antoine Havet pour être parvenu à m'extorquer une rente de 5 euros par ans pour le GICS. Juliette Chevallier pour ses attrape-rêves. Frédéric Logé-Munerel pour ses apparitions aléatoires. Corentin Caillaud pour avoir écouté mes preuves fausses. Il en est de toute évidence d'autres que j'oublie, mais je ne les oublie pas.

Merci à mon bureau secondaire improvisé de Jussieu : merci à Guillermo Durand pour tous les drafts et pour m'avoir introduit aux groupes neurchi. Merci à Romain Mismer pour avoir encaissé toutes ces insultes contre les facqueux avec un sourire méprisant mais bienveillant. Merci à Léa Longepierre pour venir préparer le dîner pendant les drafts et merci à Maximilien Baudry de perdre. Merci à Yating Liu pour son dur labeur. Merci à Nicolas Gilliers pour son ragequit surprise du bureau qui frise l'esthétisme absolu et qui tutoie la passif-agressif magistral.

Je n'oublie pas les doctorants du reste du monde. Merci à Alexandre Zhou pour son invitation au séminaire des doctorants du CERMICS la semaine d'après l'invitation au séminaire de ce même laboratoire. Je remercie Loïc Frazza pour son aide pour mon déménagement. Merci à Alkéos Michaïl pour son soutien et pour sa classe naturelle. Merci à Lénaïc Chizat pour la passionnante discussion sur le numérique pour le transport. Je dois aussi à Raphaël Butez pour sa générosité et la qualité de son pot de thèse dont j'ai probablement absorbé la moitié. Emma Hubert pour ses plaisanteries potaches. Adrien Matricon pour ses opinions politiques absurdes impliquant des robots qui auraient des droits. Jérémy Fensch pour le fait d'être lui-même et sa superbe soutenance. Benjamain Carantino pour ses pavés politiques éclairants. Carina Beering pour sa gentillesse et son soutien. Delphin Sénizergues pour sa coolitude. Moritz Voss pour son air de jeune musicien virtuose. Antoine Kornprobst pour son style. Je remercie Nicolas Baradel pour la qualité des contenus internet qu'il consomme et partage. Merci à Timothée Bernard pour ses garden parties. Merci à Marc Savel pour ses encouragements. Merci finalement à Annemarie Grass d'avoir écouté mes jérémiades avec patience. Merci à et al.

Comment résister à une telle épreuve sans le soutien d'amis fidèles ? Merci à Su Yang pour toutes ces soirées à prétendre aller faire du sport qui ont fini à la place au fast food avec une glace géante aux m&m's. Merci à Hanghao Hu pour m'avoir poussé à atteindre le niveau 38 de Pokémon Go et essayer d'en faire de même avec Yu-Gi-Oh duel links. Merci à Andrew Aylward d'être toujours resté carré et péchu. Merci aux passionnés avec qui j'ai partagé de magnifiques randonnées : Marc-Olivier Renou pour son invitation à se geler en Suisse, Dorian Nogneng pour ses questions inutiles et impossibles à résoudre et merci à Svyatoslav Covanov pour m'avoir démontré qu'il était possible de traverser les Pyrénées en tongs à condition de s'armer de courage et de dire "non" à la douleur. Merci à Solène Durieux de m'avoir fait découvrir les open mics de Bastille. Merci aux raffeurs : Jean-Denis Zafar pour avoir fait un rituel debout, Xavier-Thomas Starkloff pour son vol de données consenti, Tristan Sylvain pour son amour des chiens et des données et Augustin Desurmont pour son travail. La team paou : merci à Xavier Muraine pour être un chiard mavra, Maxime Vachier-Lagrave pour m'avoir appris qu'on pouvait réussir sans se laver les cheveux et malgré de multiples échecs, Olivier Breux pour ses afterworks de la mort, Laurent Rousseau pour m'avoir invité à son mariage avec une princesse dans un château et enfin merci à Julien Clouet pour ses récits de combat avec des crocodiles à mains nues. Merci à Marion Chauprade pour m'avoir accepté dans sa demeure au moment où j'en avais besoin et pour nous empoisonner tous et dans les ténèbres nous lier. Merci à la clique, en particulier Tsy-Jon Lau pour les wars. Merci à Jean-Christophe Dietrich et Olivier Pradere pour les soirées à thème II et III. Adrien Laroche pour m'avoir permis de m'investir sans bornes au GICS. Merci à Clément Choukroun de m'avoir lâché seul au cœur de la crémaillère neurchi. Merci à Arsalan Shoysoronov et Seung-Ho Lee pour les soirées shark. Merci à Jean-Damien Foata et Seung-Ho pour la révolution Omni Price. Merci à Olga Chashchina et à Bruno Martinaud pour m'avoir appris à pitcher l'innovation de demain. Merci à Marie, Laura et Cyril dont les brunchs et les fêtes d'anniversaire m'ont donné les nutriments nécessaires pour entretenir une mémoire compétitive. Merci à

Julie Coniglio et Pierre Holleaux pour m'avoir invité à leur mariage. Merci aux "noyaux durs" d'Hossegor, Julien Christophe pour "Mr Robot c'est l'histoire d'un hacker autiste. Un peu comme toi, sauf que toi t'es pas un hacker !" et Christophe Mongin pour ses 11 défaites d'affilée face au deck gobelin.

Merci aux élèves de mes différents travaux dirigés pour leur enthousiasme à apprendre, leur sérieux et l'énergie qu'ils m'ont donné. Merci également aux équipes pédagogiques, professeurs ou moniteurs avec qui ce fut un plaisir de travailler.

Je remercie de la même façon l'équipe et les étudiants de Prep'Dau pour cette superbe expérience, en particulier Mimoun El Alami, Abdelhamid Zakaria ainsi que Xavier Houtarde qui nous a mis en relation.

Merci aux nombreux arbres qui m'ont permis d'imprimer mes multiples brouillons et articles. Merci à l'École Polytechnique pour m'avoir fourni une coupure générale de courant prolongée au boût de 30 minutes de soutenance. Merci à Arthur Toussaint du JTX 2016 pour m'avoir fourni la seule caméra du monde fonctionnant sur secteur pour filmer cette soutenance.

Je remercie le soutien inconditionnel et adorable du petit Maubert, ainsi que celui des regrettés Alix et Floppy dont la mémoire ne sera pas oubliée.

Je terminerai par ma famille, je remercie ma grand-mère Gisèle Tournier qui fut la première à m'enseigner les additions. Elle éveilla mon sens critique en me montrant qu'on nous mentait à l'école en prétendant qu'on ne pouvait pas soustraire 6 à 3. Merci à la regrettée Geneviève De March "mamie Ginette" qui aurait été très déçue de constater que je n'avais pas fait HEC, contrairement à sa prédiction sur mon berceau. Merci au regretté Roger Tournier pour m'avoir de nombreuses fois emmené voir les étoiles. Merci à Pauline Tournier pour m'avoir invité à son beau mariage. Je remercie mon frère Tristan De March qui me rappelle pourquoi je ne dois pas me reposer sur mes lauriers. Merci à ma mère Isabelle Tournier qui essaye toujours aujourd'hui de m'inculquer les bonnes manières. Mon père François De March qui m'a aimablement logé pendant deux ans à Paris.

Abstract

Nous étudions dans cette thèse divers aspects du transport optimal martingale en dimension plus grande que 1, de la dualité à la structure locale, puis nous proposons finalement des méthodes d'approximation numérique.

On prouve d'abord l'existence de composantes irréductibles intrinsèques aux transport martingale entre deux mesures données, ainsi que la canonicité de ces composantes. Nous avons ensuite prouvé un résultat de dualité pour le transport optimal martingale en dimension quelconque, la dualité points par points n'est plus vraie mais une forme de dualité quasi-sûre est démontrée. Cette dualité permet de démontrer la possibilité de décomposer le transport optimal quasi-sûr en une série de sous-problèmes de transports optimaux points par points sur chaque composante irréductible. On utilise enfin cette dualité pour démontrer un principe de monotonie martingale, analogue au célèbre principe de monotonie du transport optimal classique. Nous étudions ensuite la structure locale des transports optimaux, déduite de considérations différentielles. On obtient ainsi une caractérisation de cette structure en utilisant des outils de géométrie algébrique réelle. On en déduit la structure des transports optimaux martingales dans le cas des coûts puissances de la norme euclidienne, ce qui permet de résoudre une conjecture qui date de 2015. Finalement, nous avons comparé les méthodes numériques existantes et proposé une nouvelle méthode qui s'avère plus efficace et permet de traiter un problème intrinsèque de la contrainte martingale qu'est le défaut d'ordre convexe. On donne également des techniques pour gérer en pratique les problèmes numériques.

Mots clés. Transport optimal martingale, composantes irréductibles, dualité, structure locale, numérique, non-arbitrage, finance robuste, modèle, régularisation entropique.

Table of contents

Chapter 1

Introduction

1.1 Motivation: la finance robuste

1.1.1 Méthode classique de pricing

L'enjeu principal des mathématiques financières du XXe siècle a été l'évaluation du prix de produits dérivés. Étant donné un payoff, i.e. un ensemble de flux financiers déterminés par un contrat, l'enjeu est de déterminer la juste valeur de vente de ce payoff. Bachelier [12] fut le premier à utiliser le mouvement Brownien pour modéliser l'évolution des prix par un Mouvement Brownien, ce processus à l'évolution incertaine qui oscille tant tout au long de sa trajectoire qu'il n'est nulle part dérivable. Deux tiers de siècle plus tard, Black, Scholes et Merton [34] apportaient une révolution qui sera à l'origine des mathématiques financières modernes. Ils découvrent qu'en faisant l'hypothèse que le cours d'une action suit la loi d'un mouvement Brownien géométrique, il est possible de trouver une stratégie de couverture dynamique : c'est à dire qu'en achetant à chaque instant une quantité prescrite de sous-jacent, on parvient à obtenir précisément le bon payoff à la maturité du produit. Ainsi on peut répliquer des options d'achat (call) ou des options de vente (put) à partir d'un portefeuille dynamique constitué de cash et de sous-jacent. On obtient la célèbre formule de Black-Scholes qui donne le prix de l'option call en fonction de ses caractéristiques, son strike et sa maturité, ainsi qu'en fonction de caractéristiques intrinsèques à la dynamique de marché : le taux d'intérêt, la valeur actuelle du sous-jacent et sa volatilité.

Cette notion de prix est déduite de l'hypothèse fondamentale de non-arbitrage. Il s'agit d'une hypothèse selon laquelle il n'est pas possible sur un marché de gagner de l'argent à coup sûr sans risque. Cette hypothèse est justifiée par la présence sur le marché d'arbitragistes professionnels dont le métier consiste à chercher ces arbitrages pour faire du profit. Leur action va donc faire naturellement disparaître ces arbitrages en influant sur les prix pour qu'ils se rapprochent d'un système de prix sans arbitrage. Le corollaire de ce principe de non-arbitrage est que s'il existe une stratégie autofinancée qui réplique le payoff d'un produit sans faille, alors le seul prix sans arbitrage possible pour ce produit est la valeur initiale de ce portefeuille réplicateur. C'est par ce raisonnement que la découverte par Black, Scholes et Merton [34] de cette stratégie de couverture parfaite a conduit à une parfaite notion de prix.

Ainsi les mathématiques financières ont connu un essor. Il a cependant rapidement été constaté que le modèle log-normal n'était pas satisfaisant car il ne couvrait pas une série de phénomènes observés sur les cours de la bourse. Plus fondamentalement, ses évolutions ne pouvaient être réduites à un simple paramètre de volatilité constante. De nombreux modèles plus complexes ont été étudiés, permettant de capturer plus efficacement les comportements des prix des actifs. On pourra citer le modèle d'Ornstein-Uhlenbeck [154] le modèle de Heston [87], célèbre pour sa formule quasi-fermée, les modèles à volatilité locale [64] et à volatilité stochastique [81]. Tous ces modèles sont basés sur des processus diffusifs, guidés par un mouvement Brownien. Ils ont le défaut de ne pas prendre en compte l'instabilité locale des processus réels de prix [115]. Pour parer à ce problème fondamental, de nouveaux modèles ont été trouvés, on peut citer en premier lieu les processus de Lévy [138], incluant des sauts imprévisibles dans les cours, observés dans les cours réels lors d'événements macroéconomiques majeurs, ainsi que parfois sans raison apparente, laissant lieu à des spéculations interprétatives a posteriori pas les analystes financiers. Les modèles les plus à la mode actuellement utilisent un processus continu mais plus erratique que le mouvement Brownien classique, il s'agit du mouvement Brownien fractionnaire [116], dont le mouvement Brownien est un cas particulier. Ce mouvement Brownien fractionnaire possède une caractéristique appelée exposant de Hurst $0 < \alpha < 1$. Cet exposant exprime la "rugosité" de la trajectoire du mouvement Brownien fractionnaire, exprimée par une relation du type $\mathbb{E}\left[|\mathrm{BF}_t - \mathrm{BF}_s|^2\right] = (t-s)^{2\alpha}$. Le mouvement Brownien classique correspond à $\alpha = \frac{1}{2}$.

Malgré l'amélioration de ces modèles et certains de leurs bons résultats, ils ne sont jamais parfaits, car l'économie est rarement une science exacte et l'imprévisible est par essence difficile à modéliser parfaitement. Pire, il est difficile de quantifier le "risque de modèle", c'est à dire la taille caractéristique de l'erreur que l'on fait en choisissant un modèle plutôt qu'un autre. Ce problème est fondamentalement mal posé, car la seule manière existante de gérer l'imprévisible est la modélisation elle-même. Ainsi évaluer l'imprévisibilité liée au risque de modèle consisterait à créer un super-modèle englobant des modèles possibles, ayant chacun une probabilité associée. Ainsi on ne résout pas le problème car on se ramène toujours à un modèle.

Une question fondamentale motivant les mathématiciens financiers, analystes quantitatifs et traders est la question du prix de vente et d'achat de ces produits. La technique adoptée en pratique consiste à trouver un modèle pour les prix des sous-jacents, puis à estimer la juste valeur d'un produit à l'espérance mathématique sous le modèle choisi des flux financiers de son payoff. Pour choisir le bon modèle il y a deux critères à vérifier. En premier lieu, par hypothèse de non-arbitrage, le modèle doit donner son prix de marché à tout produit liquide, i.e. facilement disponible sur le marché (ou de manière équivalente si on peut parler de son prix, c'est-à-dire si la différence entre son prix d'achat et son prix de vente sur le marché est négligeable). Par exemple, dans le modèle, le prix actuel d'une action doit être égal à la valeur initiale de l'action dans le modèle. De manière plus élaborée, si on suppose que le taux d'intérêt exponentiel est constant égal à r et que l'on fixe un actif de prix S_t au temps t, pour toute fonction mesurable bornée f, si on fixe deux temps $t_1 < t_2$ dans le futur, alors le payoff au temps t_2 égal à $h(S_{t_1})S_{t_2} - h(S_{t_1})S_{t_1}e^{r(t_2-t_1)}$ est autofinancé en empruntant $h(S_{t_1})S_{t_1}$ au temps t_1 et en achetant $h(t_1)$ unités d'actif grâce à cet emprunt, puis en revendant ces actifs au prix $h(S_{t_1})S_{t_2}$ au temps t_2. Nous en déduisons que les payoffs de la forme $h(S_{t_1})(S_{t_2} - S_{t_1}e^{r(t_2-t_1)})$, appelés hedges, ont un prix nul sous tout modèle sans arbitrage. On obtient donc un résultat classique: sous un modèle sans arbitrage, un hedge est gratuit. Ainsi par dualité, cela implique que sous un modèle sans arbitrage, un actif que l'on peut acheter sur le marché doit être martingale entre tous les temps d'achat possibles, si on le renormalise par le taux d'actualisation engendré par le taux d'intérêt. Pour simplifier l'analyse, nous supposerons dans toute la suite que $r = 0$, ce qui simplifiera grandement la présentation. Le deuxième critère est que les paramètres fixes du modèle doivent évoluer peu dans le temps. Ce dernier critère doit être vérifié statistiquement mais n'est pas fiable.

Mais un problème apparaît, les "bons" modèles sont nombreux et deux modèles différents donnent dans les faits deux prix différents. On raconte que certains praticiens utilisent plusieurs modèles puis font une moyenne arithmétique des prix obtenus. Plusieurs auteurs analysent le risque de modèle en laissant des degrés de liberté aux paramètres, voir Denis et Martini [61], Soner, Touzi et Zhang [143], Nutz et Neufeld [123] ou Possamaï, Royer et Touzi [130]. Cependant il est possible d'aller beaucoup plus loin, une approche intéressante pour comprendre ce risque de modèle consisterait à regarder quel spectre de prix on peut obtenir en parcourant l'ensemble des modèles admissibles sous l'hypothèse de non arbitrage.

1.1.2 Réplication et sur-réplication des payoffs

Le calcul des prix des actifs se base donc sur l'existence de stratégies de réplication autofinancée. En pratique, les coûts de transactions et les légers retards entre observations des prix et décision de transaction font qu'il n'est pas possible de procéder au hedging dynamique prescrit par la théorie de Black-Scholes ou par ses dérivés. Ainsi il est en pratique difficile de parler de prix certain pour les produits dérivés, car ces prix dépendent du modèle employé. Pour certains produits liquides, le prix est déterminé par le marché. Les acteurs estiment le risque qu'ils sont prêts à prendre et la loi de l'offre et de la demande se charge de fixer un prix. Une classe de produits liquides classique est celle des produits "vanilles", qualificatif d'origine américaine désignant les choses simples. Les produits financiers "vanilles" sont en pratique les produits dont le payoff ne dépend que du cours d'un unique actif très échangé, à une unique maturité très échangée sur les marchés.

L'approche aujourd'hui utilisée par les praticiens consiste à utiliser les produits vanilles, faciles à acheter, et au prix connu, pour couvrir les risques des produits "exotiques" (i.e. non vanille). La technique employée consiste à fixer un modèle sans arbitrage de calcul du prix, déterminer les sensibilités (i.e. dérivées partielles du prix) du produit exotique par rapport aux différents facteurs observables, puis annuler la sensibilité globale du portefeuille en achetant la bonne quantité de produit vanille pour la compenser. Cette méthode est appelée [nom de la sensibilité]-hedging. L'exemple le plus simple est le Delta-hedging. En pratique, on appelle Delta d'un portefeuille de produits la dérivée partielle de la valeur du portefeuille P_t par rapport au prix d'une action S_t au temps t : $\Delta_t = \frac{\partial P_t}{\partial S_t}$. Ainsi pour annuler la sensibilité du portefeuille à une variation du prix de l'action à l'ordre 1, il faudra vendre Δ_t actions. Ainsi le nouveau delta du portefeuille sera : $\frac{(\partial P_t - \Delta_t S_t)}{\partial S_t} = \Delta_t - \Delta_t = 0$.

Une autre méthode utilisée en pratique consiste à couvrir les parties très irrégulières des payoffs par une sur-couverture statique. Une couverture vise à répliquer un payoff tandis qu'une sur-couverture vise à obtenir des flux financiers strictement supérieurs aux flux du payoff. Par exemple dans le cas des options digitales au payoff très irrégulier, les praticiens utilisent $\frac{1}{dK}$ options d'achat de strike $K - dK$ et vendent $\frac{1}{dK}$ options d'achat de strike K. Rendre dK trop petit rendrait le hedging trop instable, la solution consiste donc à ne pas rendre dK trop petit et à placer la différence avec l'option digitale dans un portefeuille d'overhedge, qui contient une part négligeable mais surtout strictement positive, donc ne posant aucun problème car couverte par 0.

Il est possible de généraliser cette approche. On peut sur-couvrir son portefeuille à l'aide de produits liquides, ainsi le choix du modèle n'a aucune importance, car on couvre le payoff indépendamment des événements à venir. L'objectif est alors de trouver la sur-couverture au prix le plus faible possible.

1.1.3 Dualité entre les approches

Ici, nous nous concentrerons sur la situation de marché suivante : le but est de couvrir un produit dérivé structuré de maturité t_2 sur $d \geq 1$ actifs risqués sous-jacents, dont le payoff est donné par une fonction $c : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$ calculée en le prix des sous-jacents aux temps futurs t_1 et t_2. Ainsi, ce produit verse un flux $c\Big((S^1_{t_1},...,S^d_{t_1}),(S^1_{t_2},...,S^d_{t_2})\Big)$ au temps t_2, où S^i_t est le prix de l'actif i au temps t. On suppose également qu'on dispose d'un marché liquide sans arbitrage et sans taux d'intérêt où peuvent être achetés tous les sous-jacents, ainsi que tout produit dérivé des sous-jacents dont le payoff ne dépend que d'une unique maturité. En termes mathématiques, on peut acheter sur ce marché tout produit de la forme $\varphi\Big((S^1_{t_1},...,S^d_{t_1})\Big)$, et tout produit de la forme $\psi\Big((S^1_{t_2},...,S^d_{t_2})\Big)$, ainsi que n'importe quel hedge $\sum_{i=1}^d h_i\Big((S^1_{t_1},...,S^d_{t_1})\Big)(S^i_{t_2}-S^i_{t_1}) = h(S_{t_1})\cdot(S_{t_2}-S_{t_1})$, où $h := (h_1,...,h_d)$ et $S_t := (S^1_t,...,S^d_t)$. Pour poser le problème simplement et faire le rapport aisément avec le transport optimal, on introduit les vecteurs aléatoires $X := S_{t_1}$ et $Y := S_{t_2}$. On rappelle que les hedges sont gratuits sous taux d'intérêt nul, quant aux payoffs $\varphi(X)$ et $\psi(Y)$, leurs prix sont donnés par le marché. La fonction de prix qui à une fonction φ associe le prix du payoff $\varphi(X)$ est une forme linéaire croissante telle que le prix de 1 est 1, elle peut donc être représentée par une mesure de probabilité. On appelle μ cette probabilité et ν la probabilité associée au temps t_2. Ainsi si $\varphi,\psi : \mathbb{R}^d \longrightarrow \mathbb{R}$ et $h : \mathbb{R}^d \longmapsto \mathbb{R}^d$, le prix du payoff $\varphi(X)+\psi(Y)+h(X)\cdot(Y-X)$ est $\mu[\varphi]+\nu[\psi]$. On notera $\mathbb{L}^1(\mu)$ l'ensemble des fonction $\mu-$intégrables, $\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$ l'ensemble des fonction $\mathbb{R}^d \longrightarrow \mathbb{R}^d$ boréliennes et $\mathcal{D}_{\mu,\nu}(c) := \{(\varphi,\psi,h) \in \mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d) : \varphi\oplus\psi+h^{\otimes} \geq c\}$ l'ensemble des sur-couvertures du payoff c, où $h^{\otimes}$ désigne $h(X)\cdot(Y-X)$.

Dans ces conditions, un modèle probabiliste sur (X,Y) donnant les bons prix aux produits liquides du marché est une mesure de probabilité $\mathbb{P}$ sur $\Omega := \mathbb{R}^d\times\mathbb{R}^d$ telle que la loi de X est donnée par μ et la loi de Y est donnée par ν. De plus, comme évoqué précédemment, la gratuité des hedges est équivalente pour $\mathbb{P}$ à vérifier la contrainte martingale, i.e. $\mathbb{E}^{\mathbb{P}}[Y|X] = X$, $\mu-$presque sûrement. On notera $\mathcal{M}(\mu,\nu)$, l'ensemble des modèles vérifiant ces contraintes.

On peut donc poser deux problèmes qui s'avèreront être duaux. Premièrement, le problème du modèle donnant le prix le plus élevé au payoff c tout en donnant les bons prix aux produits liquides du marché, qui s'avèrera être le problème de transport optimal martingale :

$$\mathbf{S}_{\mu,\nu}(c) := \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{E}^{\mathbb{P}}[c]. \tag{1.1.1}$$

Deuxièmement, on introduit le problème de la sur-couverture la moins onéreuse, étant également le problème dual du problème de transport optimal martingale:

$$\mathbf{I}_{\mu,\nu}(c) := \inf_{(\varphi,\psi,h)\in\mathcal{D}_{\mu,\nu}(c)} \mu[\varphi]+\nu[\psi]. \tag{1.1.2}$$

Beiglböck, Henry-Labordère et Penckner [18] ont prouvé qu'il y avait une relation de dualité dans le cas $d=1$, entre le problème du modèle sans arbitrage qui donne le prix le plus élevé et le problème de la sur-couverture au prix le moins élevé. Ce supremum et cet infimum sont égaux : $\mathbf{S}_{\mu,\nu}(c)=\mathbf{I}_{\mu,\nu}(c)$. Quitte à appliquer ce résultat à l'opposé du payoff, on a également égalité entre la sous-couverture au prix le plus élevé et le modèle qui donne le prix le moins élevé. Ce résultat permet de comprendre la pertinence de la valorisation par utilisation de modèles sans arbitrage, ces modèles donnant tous les prix sans arbitrage possible, et le problème dual donnant des arbitrages pour les prix sortant de ces bornes.

Il est classique d'utiliser des relations de dualité pour lier problématiques de modélisation et absence d'arbitrage. Dans le cas où il existe une probabilité universelle indiquant les événements impossibles en pratique, le célèbre "théorème fondamental de la valorisation d'actif" dit que l'absence d'arbitrage est équivalent à l'existence d'un modèle martingale qui détermine les prix (voir Delbaen et Schachermayer [59], Föllmer et Schied [70] ou El Karoui et Quenez [66] dans le cas d'un marché incomplet). Dans le cas où il n'existe pas une telle probabilité dominante en temps discret, Bouchard et Nutz [35] ont apporté un résultat de dualité exploitant un type de limite universelle compatible avec les probabilités. Cette procédure a été étendue au temps continu par [31]. On citera également [40] qui donne un résultat similaire dans un cadre légèrement différent. Les autres exemples de travaux démontrant le même type de dualité sont nombreux, voir [1, 63, 69, 78, 95].

Ces bornes ont l'intérêt pratique de montrer quel est le risque de modèle associé à un payoff. Ce problème de sur-couverture robuste avait été introduit par Hobson [94], et donnait des solutions pour des exemples particuliers de produits dérivés exotiques à l'aide de solutions correspondantes du problème de plongement de Skorokhod, voir [51, 92, 93], et l'étude [91]. On peut également citer des résultats analogues obtenus dans le cas du temps continu pour des options asiatiques [50, 145], pour des options américaines [4, 60] ou même pour des options sur le temps local [47], voir également [84, 98]. Beiglböck, Cox et Huessman [17] montrent même que le problème de plongement de Skorokhod optimal peut être vu comme un problème de transport optimal et que par nature, il optimise l'espérance d'un payoff, mettant ainsi en lien les travaux précédemment évoqués.

Cette relation de dualité est l'analogue de la célèbre relation de dualité de Kantorovitch [102], et le problème de l'espérance maximale du payoff est un problème de transport optimal sous contrainte martingale. La relation de dualité s'obtient de manière similaire grâce au théorème de minimax [105].

Nous avons fait ici l'hypothèse que le marché permettait d'acheter les payoffs de la forme $\varphi(S_{t_1})$ de façon liquide. Dans le cas $d=1$, on peut en pratique acheter des options d'achat, ou call (de payoff $(S_{t_1}-K)_+$, où on appelle K le prix d'exercice, ou strike) et des options de vente, ou put (de payoff $(K-S_{t_1})_+$, où K est le strike) sur le marché. Breeden et Litzenberger [37] prouvent que sous l'hypothèse de non arbitrage, et si le marché offre des calls ou puts liquides à tous strikes $K \geq 0$, alors il existe une unique probabilité μ représentant leur prix, et elle est obtenue en prenant la dérivée seconde au sens des distributions des prix de call par rapport au strike : $\mu := \partial_K^2 Call(t_1,K) = \partial_K^2 Put(t_1,K)$. Réciproquement, tout payoff régulier peut s'écrire comme une intégrale de calls et de puts d'après la formule de Carr-Madan [43], ainsi il est raisonnable d'estimer pouvoir acheter n'importe quel payoff fonction de S_{t_1} ou S_{t_2}. En pratique les calls et puts sont liquides sur certaines grandes échéances de temps à un nombre de prix d'exercice dépendant de la taille du marché de l'actif. On approxime souvent ce nombre discret par un continuum.

Si on suppose à présent que $d>1$, il faut supposer que l'on a accès à un grand nombre d'options basket de payoffs $(\lambda_1 S_{t_1}^1 + ... + \lambda_d S_{t_1}^d - K)_+$, pour des strikes K et des coefficients $\lambda_1,...,\lambda_d \geq 0$, qui somment à 1. On peut par une formule de type Carr-Madan reconstituer une portion de la fonction génératrice exponentielle de S_{t_1} et en reconstituer la loi. En pratique, cette hypothèse est peu réaliste, mais le problème reste intéressant. Il est plus naturel de ne pas supposer connaître la copule entre les S^i (voir [114]) ou de résoudre un problème hybride où on connaît les lois marginales des actifs à un temps donné tout en ajoutant comme contraintes les espérances de quelques payoffs basket.

1.2 Transport optimal

1.2.1 Le problème de Monge

Le transport optimal martingale est une variante récente du transport optimal tiré de la finance robuste. Cependant le transport optimal classique possède des racines beaucoup plus anciennes et une évolution qui en a fait un des sujets incontournables des probabilités. Le transport optimal est apparu au XVIIIe siècle, posé initialement par Gaspard Monge [121]. A la suite de la Révolution française, le temps était à la reconstruction et le problème de transports optimal est apparu comme un problème concret lié aux travaux publics : comment déplacer des déblais vers les lieux de remblais en parcourant le moins de distance possible ? Le problème initial consistait à trouver une fonction de transport T qui indique le lieu $T(x)$ où envoyer une part de déblais partis d'un point x. On peut l'écrire comme suit

$$\inf_{\substack{T:\mathcal{X}\longrightarrow\mathcal{Y}\\ T\#\mu=\nu}} \int_{\mathcal{X}} |T(x)-x|\,\mu(dx). \tag{1.2.3}$$

Plus d'un siècle plus tard, Kantorovitch redécouvrit le problème de transport optimal en lui donnant une forme probabiliste [102]. On peut l'exprimer comme suit

$$\inf_{\mathbb{P}\in\mathcal{P}(\mu,\nu)} \mathbb{E}^{\mathbb{P}}\left[c(X,Y)\right], \tag{1.2.4}$$

où $\mathcal{P}(\mu,\nu)$ est l'ensemble des couplages entre μ et ν, c'est à dire l'ensemble des probabilités $\mathbb{P}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ aux marginales prescrites $\mathbb{P}\circ X^{-1}=\mu$ et $\mathbb{P}\circ Y^{-1}=\nu$, avec X la variable aléatoire fondamentale de projection sur $\mathcal{X}$ et Y la projection sur $\mathcal{Y}$. Il fit plus tard le lien avec le problème de Monge [100]. La différence entre les deux problèmes réside dans le fait qu'un grain de matière n'a plus nécessairement une unique destination, mais peut avoir une masse de destination. En devenant probabiliste, l'ensemble d'optimisation acquiert une structure compacte pour la topologie faible qui permet de montrer qu'un minimum existe dans le cas semi-continu inférieurement (voir le chapitre 4 de [157]) ou alors en général, modulo une rectification de la fonction de coût [23]. Cette nouvelle version du problème permet également de définir la distance de Kantorovitch-Rubinstein [101, 99], ou distance de Wasserstein [155] entre deux mesures de probabilité qui consiste en la solution du transport optimal avec ces deux marginales, et avec la fonction de coût $|X-Y|^p$ pour $p\geq 1$. Cette métrique est très naturelle et possède de bonnes propriétés : elle donne à l'ensemble des mesures de probabilité sur un espace polonais une structure d'espace polonais (voir Section 14 de [62]).

1.2.2 Dualité de Kantorovitch et ses conséquences

Certains des résultats du transport optimal classique notables auront des extensions au transport optimal martingale. Le premier est la célèbre dualité de Kantorovitch, il s'agit d'un résultat de dualité de type Kuhn-Tucker [108] en dimension infinie. Kantorovitch [102] montre que le problème suivant

$$\sup_{\varphi \oplus \psi \leq c} \mu[\varphi] + \nu[\psi], \tag{1.2.5}$$

est le dual du problème 1.2.4 et possède la même valeur. De plus ce problème possède un optimiseur si la fonction de coût c est continue. Le problème dual du transport optimal est un problème d'optimisation sur le dual des mesures, c'est à dire des fonctions. Si on limite le problème à la dimension finie, les fonctions duales ont l'intérêt d'être de dimension très inférieure à la probabilité, ce qui présente des avantages de nombreux points de vue.

Une de ses conséquences est la monotonie (voir [132]). On peut montrer qu'il existe un ensemble Γ qui caractérise l'optimalité des transports, i.e. un transport $\mathbb{P} \in \mathcal{P}(\mu,\nu)$ est optimal pour le problème 1.2.4 si et seulement si $\mathbb{P}$ est concentré sur Γ. L'ensemble de contact entre la fonction de coût à deux variables et la somme de fonctions duales optimales d'une variable $\Gamma := \{c = \varphi \oplus \psi\}$ est un choix possible pour cet ensemble monotone.

Sous certaines hypothèses de régularité sur c, on peut prouver que φ est localement Lipschitz et donc dérivable presque partout (voir [68]). Ainsi raisonnons formellement, soit $\Delta(X,Y) := c(X,Y) - \varphi(X) - \psi(Y)$. Nous avons vu que, les probabilités optimales pour 1.2.4 sont concentrées sur $\Gamma := \{\Delta = 0\}$, et par définition du dual, $\Delta \geq 0$. Ainsi Δ atteint son minimum sur Γ, ce qui donne une équation utile en utilisant la condition d'optimalité du premier ordre. On obtient

$$\partial_x c(x,y) = \nabla\varphi(x), \quad \text{for} \quad (x,y) \in \Gamma. \tag{1.2.6}$$

Dans le cas où l'application $y \longmapsto \partial_x c(x,y)$ est injective (comme c'est le cas pour $c := |X-Y|^2$), on remarque que y est fixé quand x est fixé : $y = \partial_x c(x,\cdot)^{-1}\Big(\nabla\varphi(x)\Big)$ pout tout $x \in \mathcal{X}$. Ainsi les transports optimaux obtenus en ce cas sont déterministes, voir [135]. Cela permet également de montrer qu'il y a unicité en utilisant le fait que le problème est linéaire et que toute combinaison convexe de transports optimaux est encore optimale. Dans le cas particulier du coût distance au carré, un célèbre résultat de Yann Brénier [39] montre que l'application de transport optimal T est le gradient d'une fonction potentielle convexe.

Dans le cas du coût distance qui était initialement présent pour le problème de Monge (1.2.3), l'équation (1.2.6) indique que les transports sont envoyés sur des droites. Mais prouver l'existence d'une fonction de transport solution du problème (1.2.3) rigoureusement est très compliqué. Sudakov [147] a cru avoir résolu le problème mais il a pensé à tort que la mesure de Lebesgue se désintégrait de manière régulière sur les lignes de transport. Ambrosio corrigea sa preuve plus tard [7]. Dans [10], est exhibé un contre-exemple, inspiré d'un contre-exemple paradoxal de Davies [54], qui sera important pour le transport optimal martingale. Il s'agit de l'exemple de l'ensemble de Nikodym N, un ensemble presque partout égal à un cube tridimensionnel tel qu'un continuum de lignes deux à deux disjointes qui intersectent tout N, n'en intersectent qu'un point à la fois. Ainsi la mesure de Lebesgue sur le cube se désintègre comme des masses de Dirac sur les lignes du continuum. Le problème (1.2.3) avait cependant été résolu entre temps par Evans et Gangbo [67] sous de fortes hypothèses, par Trudinger et Wang [152] et enfin par Cafarelli, Feldman et McCann [41]. Plus tard, [32] étudie des propriétés d'extrémalité et d'unicité de ces transports et [33] étend ce résultat à des espaces géodésiques.

La dualité a aussi des applications pour des coûts non réguliers. L'application du théorème du minimax qui donne la dualité requiert au moins que c soit semi-continue inférieurement. Cependant Kellerer [103] est parvenu à étendre ce résultat de dualité à des fonctions seulement mesurables grâce au puissant théorème de capacitabilité de Choquet [46] basé sur son extension des ensembles analytiques [45]. Cette dualité permet d'obtenir un résultat très difficile à prouver autrement : la structure des ensembles boréliens polaires. Un ensemble $N \subset \mathcal{X} \times \mathcal{Y}$ est dit $\mathcal{P}(\mu,\nu)-$polaire si $\mathbb{P}[N] = 0$ pour tout $\mathbb{P} \in \mathcal{P}(\mu,\nu)$. Ainsi en appliquant la dualité de Kantorovitch-Kellerer au coût $c := -\mathbf{1}_N$, on obtient le résultat suivant: soit $\mathcal{N}_\mu$ l'ensemble des ensembles $\mu-$négligeables et $\mathcal{N}_\nu$ l'ensemble des ensembles $\nu-$négligeables.

Theorem 1.2.1. *Soit un ensemble borélien $N \subset \mathcal{X} \times \mathcal{Y}$, alors N est $\mathcal{P}(\mu,\nu)-$polaire si et seulement si*

$$N \subset (N_\mu \times \mathbb{R}^d) \cup (\mathbb{R}^d \times N_\nu), \quad \textit{pour un certain couple} \quad (N_\mu, N_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu.$$

Un autre résultat notable que la dualité permet de démontrer est le fait que la restriction d'un transport optimal est toujours optimale, cf Théorème 5.19 de [157]. Pour des monographes de référence sur le transport optimal, voir Villani [156, 157], Ambrosio, Gigli et Savaré [9, 8] et Santambrosio [136].

1.2.3 Techniques de résolution numérique et applications

Pour résoudre le problème numériquement il semble difficile de s'affranchir d'une discrétisation des marginales. La façon naturelle de procéder est de remplacer μ et ν par des sommes de masses de Dirac sur des grilles finies. On peut soit passer par une grille régulière, soit utiliser un tirage de Monte Carlo pour approximer les mesures μ et ν, pour combattre la malédiction de la dimension en dimension plus grande que 2. Les méthodes de résolution ont énormément évolué, se sont drastiquement améliorées et sont encore un domaine de recherche très actif étant donnée l'utilité du transport optimal dans divers domaines tels que l'analyse d'image par exemple.

Historiquement, ce problème était naturellement résolu par des méthodes de programmation linéaire. On pourra citer la méthode hongroise [107], l'algorithme des enchères, le simplexe [3], on évoquera également [76]. Cependant, les méthodes de programmation linéaire ont un coût polynomial, voir [141] et [144]. Plus tard, Benhamou et Brénier [25] relancèrent la course à la performance en découvrant une autre méthode pour résoudre le problème en le transformant en problème de programmation dynamique avec une pénalisation finale sur la différence entre la loi finale du processus et la mesure cible. Pour des cas particuliers, il est aussi possible de passer par l'équation de Monge-Ampère [80]. Dans le cas du coût distance au carré, le potentiel convexe de Brénier [39] u satisfait une équation provenant de l'équation de conservation des masses par changement de variable. Quand en particulier les mesures de départ et d'arrivée possèdent une densité par rapport à la mesure de Lebesgue, on peut prouver que u est solution de l'équation de Monge-Ampère $\det D^2 u = \frac{g \circ \partial_x c(X,\cdot)^{-1} \circ \nabla u}{f}$, où f est la densité de μ et g est la densité de ν. Cette équation satisfait un principe du maximum, ce qui permet de la résoudre en pratique, voir [28] et [27]. D'autre travaux comme [153] résolvent des équations de Monge-Ampère plus générales qui permettent d'envisager de prouver que des problèmes de transport optimal pour des coûts plus généraux la satisfont. Nous nous devons également de mentionner Mérigot [118], qui résout le problème en utilisant une formulation semi-discrète (i.e. une des mesures est approximée par une somme de mesures de masses de Dirac et l'autre par une mesure à densité affine par morceaux par rapport à la mesure de Lebesgue). Lévy [111] a introduit une méthode de Newton utilisant la dérivée seconde qui permet de résoudre le problème semi-discret extrêmement rapidement.

Ces solutions sont cependant limitées car soit elles ne permettent pas de résoudre des problèmes assez grands, soit elle ne permettent de résoudre que des cas particuliers de coûts, bien que ceux-ci soient très pertinents. C'est alors que Leonard [110] prouva la Gamma-convergence d'un problème de transport avec pénalisation entropique vers le problème original de transport optimal en marge d'un article traitant du problème de Schrödinger. La Gamma-convergence est la convergence adaptée dans le cas de l'étude des limites des problèmes d'optimisation. Elle montre la convergence de la valeur du problème de transport optimal entropique vers la valeur du problème (1.2.4), et de plus toute suite d'optimiseurs convergente converge vers un optimiseur de (1.2.4). Voir [42] et [48] pour des études plus précises et quantitatives de ces convergences dans des cas particuliers. Il avait été observé par [106] que la formulation entropique était particulièrement utile en pratique pour résoudre des problèmes numériquement, étant donné que cela permettait d'employer le célèbre algorithme de Sinkhorn [140]. La puissance de cette technique a été redécouverte par Marco Cuturi [52], et largement adoptée par la communauté, voir [142, 131, 151]. Le principe de cette méthode a déjà été adapté pour résoudre plusieurs variantes du transport optimal, tels que les barycentres de Wasserstein [2] et le transport optimal multi-marginales [26], des problèmes de flots de gradient [128], le transport optimal déséquilibré [44], et le transport optimal martingale en dimension 1 [77].

Le travail remarquable de Schmitzer [137], très orienté vers le praticien, donne des considérations très pratiques et des astuces pour stabiliser et faire converger l'algorithme de Sinkhorn beaucoup plus vite que par une implémentation naïve. Cuturi et Peyré [53] ont utilisé une méthode de quasi-Newton pour résoudre le problème de transport optimal. Leur conclusion semble être que l'algorithme de Sinkhorn reste plus efficace. Cependant, [36] utilise une méthode de Newton inexacte (i.e. utilisant également l'expression de la dérivée d'ordre 2) et parvient à surperformer l'algorithme de Sinkhorn. Il nous faut également mentionner [6] qui introduit un "Greenkhorn algorithm" qui surperforme l'algorithme de Sinkhorn d'après leurs expériences numériques, et de la même manière [150] introduit une version relâchée de l'algorithme de Sinkhorn qui passe l'exposant de convergence linéaire au carré, accélérant sensiblement sa convergence.

1.3 Transport optimal martingale

1.3.1 Résultats de la littérature en une dimension

Le transport optimal martingale est une variante du transport optimal incluant une contrainte martingale. Ce problème a été introduit comme le dual d'un problème de couverture robuste d'un produit dérivé exotique en mathématiques financières, voir Beiglböck, Henry-Labordère et Penkner [18] en temps discret, et Galichon, Henry-Labordère et Touzi [73] en temps continu. En temps discret, le problème inclut une temporalité et des contraintes martingales. On se place dans le cas $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ pour $d \geq 1$, la variable aléatoire Y est considérée comme postérieure à la variable aléatoire X et on introduit une contrainte martingale en relation : $\mathbb{E}^{\mathbb{P}}[Y|X] = X$. On écrit $\mathcal{M}(\mu,\nu) := \{\mathbb{P} \in \mathcal{P}(\mu,\nu) : \mathbb{E}^{\mathbb{P}}[Y|X] = X\}$, l'ensemble des transports optimaux martingales. Contrairement à l'ensemble $\mathcal{P}(\mu,\nu)$ qui avait la particularité de ne jamais être vide, contenant $\mu\otimes\nu$, par exemple, l'ensemble $\mathcal{M}(\mu,\nu)$ peut être vide, par exemple si $\mu := \delta_x$ et $\nu := \delta_y$ avec $x \neq y$. Il existe une caractérisation basée sur le théorème de Hahn-Banach de la vacuité ou non de $\mathcal{M}(\mu,\nu)$. On remarque que si $f : \mathbb{R}^d \longmapsto \mathbb{R}$ est convexe et $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, alors

$$\mathbb{E}^{\mathbb{P}}[f(Y)|X] \geq f\Big(\mathbb{E}^{\mathbb{P}}[Y|X]\Big) = f(X), \tag{1.3.7}$$

par l'inégalité de Jensen. En intégrant (1.3.7) par rapport à la mesure μ, on obtient

$$\nu[f] \geq \mu[f]. \tag{1.3.8}$$

Quand l'équation (1.3.8) est vérifiée pour toute fonction convexe f, on dit que μ est plus petite que ν pour l'ordre convexe, nous venons donc de montrer que si $\mathcal{M}(\mu,\nu)$ est non vide, alors $\mu \leq \nu$ pour l'ordre convexe. Strassen [146] a prouvé qu'il s'agit d'une équivalence. On saisit dès à présent la forte intrication entre la structure de $\mathcal{M}(\mu,\nu)$ et l'action de l'opérateur $(\nu-\mu)$ sur les fonctions convexes. Nous exploiterons de nouveau par la suite la fécondité de cette relation.

Qu'en est-il de la dualité ? [18] prouve une relation de dualité pour des fonctions de couplage semi-continues supérieurement et ne montre pas qu'il existe un optimiseur pour le problème dual. Beiglböck, Nutz et Touzi [22] montrent qu'en une dimension il faut considérer des fonctions duales dominant $\mathcal{M}(\mu,\nu)-$quasi-sûrement (i.e. $\mathbb{P}-$presque sûrement pour tout $\mathbb{P} \in \mathcal{M}(\mu,\nu)$) la fonction de couplage c plutôt que uniformément. Il faut également prendre garde au cas où on a $\mu[\varphi]+\nu[\psi] = -\infty+\infty$, auquel cas il faut utiliser un modérateur convexe χ et étendre la définition de $\mu[\varphi]+\nu[\psi]$ comme suit :

$$\mu[\varphi]+\nu[\psi]:=\mu[\varphi+\chi]+\nu[\psi-\chi]+(\nu-\mu)[\chi],$$

où l'opérateur $(\nu-\mu)$ est étendu de manière spécifique pour les fonctions convexes. Avec cette définition étendue, la dualité est obtenue en remplaçant le problème (1.1.2) par le problème

$$\mathbf{I}^{qs}_{\mu,\nu}(c):=\inf_{(\varphi,\psi,h)\in\mathcal{D}^{qs}_{\mu,\nu}(c)}\mu[\varphi]+\nu[\psi], \tag{1.3.9}$$

où $\mathcal{D}^{qs}_{\mu,\nu}(c):=\{(\varphi,\psi,h):\varphi\oplus\psi+h^{\otimes}\geq c,\mathcal{M}(\mu,\nu)-\text{q.s.}\}$.

[22] caractérise pour ce faire le fait qu'une propriété soit quasi-sûre, ils présentent un phénomène étonnant initialement découvert par [20], l'existence de composantes irréductibles laissées stables par tous les transports martingales entre μ et ν. Ces composantes sont des intervalles ouverts disjoints $(I_k)_k$, en conséquence au plus dénombrables. On a alors la propriété suivante : $\mathbb{P}[Y\in\operatorname{cl}I_k|X\in I_k]=1$ pour tout k. [22] prouve même que ces composantes permettent de caractériser les ensembles $\mathcal{M}(\mu,\nu)-$polaires par le théorème suivant, qui peut être vu comme une extension du Théorème 1.2.1. Soit $J_k=\operatorname{conv}\operatorname{supp}\mathbb{P}_k$, où $\mathbb{P}_k:=\mathbb{P}_{|I_k\times\mathbb{R}}$, pour $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, dont le choix n'influe pas la valeur de J_k.

Theorem 1.3.1. *Soit $N\subset\mathbb{R}^2$ Borel, alors N est $\mathcal{M}(\mu,\nu)-$polaire si et seulement si*

$$N\subset(N_\mu\times\mathbb{R})\cup(\mathbb{R}\times N_\nu)\cup(\cup_k I_k\times J_k\cup\{X=Y\})^c,$$
$$\textit{pour un certain couple } (N_\mu,N_\nu)\in\mathcal{N}_\mu\times\mathcal{N}_\nu.$$

Les J_k sont des ensembles convexes compris au sens de l'inclusion entre I_k et $\operatorname{cl}I_k$, c'est à dire avec zéro, un ou deux points supplémentaires. Les auteurs montrent aussi que le problème se décompose en sous-problèmes de transport optimal martingale sur chaque composante.

Beiglböck, Nutz et Touzi donnent aussi d'utiles contre-exemples montrant que même dans des cas très simples, soit la dualité n'était pas vérifiée pour des fonctions non semi-continues supérieurement, soit il n'existait pas de dual ne nécessitant pas de modérateur convexe qui minimise le problème dual.

Nutz, Stebegg et Tan [124] ont généralisé le résultat précédent à un nombre fini de temps. Beiglböck, Lim et Obloj [21] prouvent qu'en dimension 1, sous des hypothèses de régularité sur la fonction de couplage c, on peut montrer qu'un résultat de dualité uniforme est vraie, malgré la présence de composantes irréductibles.

Il existe aussi des résultat de structure, Beiglböck et Juillet [20] introduisent le "left-curtain" couplage martingale tel que les noyaux conditionnés en X sont toujours concentrés sur deux points au plus, et [85] montrent que ce couplage particulier est toujours solution si la dérivée partielle en x de la fonction de couplage c est strictement convexe en y. Beiglböck et Juillet [20] prouvent également que dans le cas du coût distance, les noyaux conditionnels sont également concentrés sur deux points au plus.

Le problème en temps continu, qui ne sera pas traité par cette thèse, a été introduit par [73] et supplante l'outil habituellement utilisé, le plongement de Skorokhod qui consiste à trouver un temps d'arrêt τ pour un mouvement Brownien $(W_t)_{t\geq 0}$ de loi initiale μ, tel que le mouvement Brownien arrêté W_τ possède la loi prescrite ν. Ce problème a été introduit par Monroe [122] pour la loi initiale $\mu = \delta_0$, puis généralisé à μ quelconque par Cox [49]. On trouve de nombreux exemples dans [90]. Par la suite [19] établit une nouvelle fois cette relation entre plongement de Skorokhod et transport optimal martingale dans le cas précis de l'étude du couplage "left curtain". Une dualité ainsi qu'une étroitesse sont démontrées dans le cas continu par [79], tandis que [13] étend le résultat d'interpolation de Brénier au cas martingale. Voir aussi l'étude [126].

1.3.2 Résultats de la littérature en dimension supérieure

La propriété de dualité de [18] s'étend sans difficulté à la dimension supérieure même si ils ne l'écrivent pas. On peut obtenir le résultat à partir de Zaev [160] comme cas particulier où les contraintes linéaires sont la contrainte martingale. Des résultats généraux de dualité, où l'ensemble dual conserve une forme abstraite peuvent être trouvés dans [65] ou [14].

Lim [113] a été le premier à essayer de comprendre la structure des solutions en dimension supérieure. Basé sur un problème à symmétrie radiale, il considère un cas particulier permettant de se ramener à la dimension 1. Ghoussoub, Kim et Lim [74] sont les premiers à avoir apporté des résultats vraiment intrinsèques à la dimension supérieure. Ils prouvent l'existence de composantes sur lesquelles la dualité a lieu. Ils utilisent ces composantes pour prouver un résultat de structure en dimensions 1 et 2 et conjecturent que ce résultat reste vrai en dimension supérieure. Ce résultat de structure stipule que dans le cas du problème de maximisation du transport optimal pour le coût distance, les noyaux du transport optimal se concentrent sur l'enveloppe de Choquet (i.e. les points extrêmes de l'enveloppe convexe) de leur support. Les composantes irréductibles de [74] présentent l'inconvénient de ne pas être universelles, elles peuvent dépendre du support choisi, de plus elles peuvent être en quantité non dénombrable, et donc des propriétés de mesurabilité sont nécessaires pour parvenir à décomposer correctement le problème "composante par composante" et cette mesurabilité semble difficile à démontrer dans le cadre de leur définition des composantes. Ils relèvent une difficulté qu'ils ne résolvent qu'en dimension 2 pour leur résultat de structure, il s'agit du problème causé par le fait que la décomposition des marginales sur les composantes irréductibles peut ne pas être diffuse, comme dans l'exemple de l'ensemble de Nikodym [11].

Obloj et Siorpaes [127] définissent des composantes irréductibles intrinsèques aux fonctions convexes dont la minimalité n'est pas avérée.

Finalement, nous mentionnons [114] où Lim traite un problème légèrement différent où chaque marginale $1-$dimensionnelle est prescrite et la contrainte martingale est imposée. Ce problème est plus facile que le transport optimal martingale en dimension supérieure mais semble plus adapté à la finance en cela qu'il peut être peu réaliste de supposer que l'on connaît parfaitement la copule entre deux actifs à un temps donné.

1.3.3 Littérature pour la résolution numérique du transport optimal martingale

Pierre Henry-Labordère résout le problème de transport optimal martingale par de simples méthodes de programmation linéaire [83]. Il utilise des payoffs à structure particulière, ce qui lui permet de réduire les contraintes sur les fonctions duales, et donc lui permet de résoudre un problème pour une grille de taille 10000 en moins d'une minute. Tan et Touzi [148] résolvent une forme continue du transport optimal avec une méthode proche de [25].

Alfonsi, Corbetta et Jourdain [5] utilisent également une version discrétisée des marginales mais se heurtent au problème de la nécessité que les marginales soient en ordre convexe pour qu'un transport martingale existe et que le problème linéaire n'ait pas une solution infinie. Ils résolvent ce problème en présentant des résultats d'existence de marginales en ordre convexe qui approche pour des distances de Wassertstein la mesure cible, et ils calculent ces marginales optimales.

Guo et Obloj [77] résolvent le problème précédent en relâchant la contrainte martingale dans leur approximation. Ils proposent ensuite un algorithme d'optimisation semi-duale implicite non-régulière et un algorithme de projection entropique de Bregman sans préciser leur efficacité respective.

On peut également mentionner Henry-Labordère et Touzi [85] qui utilisent la structure du transport martingale "left curtain" pour le déterminer numériquement.

1.4 Nouveaux résultats en dimension supérieure

1.4.1 Composantes irréductibles

Dans le Chapitre 2, on démontre que l'on peut définir des composantes irréductibles correspondant aux transports martingales. Il s'agit de l'article [58]. En dimension 1 ([22]), les composantes sont "révélées" par les potentiels des marginales. Ces potentiels sont définis par $u_\mu(x) := \int_{\mathbb{R}} |y-x|\mu(dy)$ et la même définition pour u_ν. Par Strassen [146], les marginales sont en ordre convexe, on a donc $u_\mu \leq u_\nu$. La fonction $u_\nu - u_\mu$ est continue car Lipschitz. L'ensemble $(u_\nu - u_\mu)^{-1}(\mathbb{R}_+^*)$ est donc un ouvert de $\mathbb{R}$, et peut donc classiquement être décomposé de manière unique en une union dénombrable d'intervalles ouverts disjoints. On remarque que si $u_\mu(x) = u_\nu(x)$ pour un certain $x \in \mathbb{R}$, on est dans le cas d'égalité de l'inégalité de Jensen dans (1.3.7). Ainsi la fonction $y \mapsto |y-x|$ est vue comme affine par tour les transports martingales. D'où la propriété de stabilité par les transports des composantes. En parallèle, on remarque que si $(\nu-\mu)[\chi_n]$ est borné avec $(\chi_n)_{n\geq 1}$ une suite de fonctions convexes, χ_n possède des propriétés de compacité sur chaque composante irréductible. Ainsi sur chaque composante irréductible, il est possible de trouver une limite pour les fonctions duales. On a alors un optimiseur dual sur chaque composante. Il est cependant difficile d'étendre cette technique à la dimension supérieure car les fonctions convexes extrêmes ont une structure beaucoup plus compliquée que les $y \longmapsto |y-x|$ de la dimension 1, voir [97].

On a vu que pour le résultat de Strassen [146] ainsi que pour [22], l'action de $(\nu-\mu)$ sur les fonctions convexes permet de comprendre la structure de $\mathcal{M}(\mu,\nu)$. En utilisant cette considération, on prouve dans le Chapitre 2 l'existence de composantes irréductibles intrinsèquement liées à (μ,ν), couple de probabilités dans l'ordre convexe. Pour ceci ils définissent les fonctions convexes tangentes, qui sont les limites de fonctions $\mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}_+$, définies par $\theta := f(Y) - f(X) - p(X)\cdot(Y-X) \geq 0$, pour $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ convexe et $p(x) \in \partial f(x)$, tels que $(\nu-\mu)[f] = \mathbb{E}^{\mathbb{P}}[\theta] \leq 1$, pour tout $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. On remarque que θ est une fonction que l'on trouve naturellement dans l'espace dual dans le cas du transport optimal. On appelle cet ensemble de limites $\mathcal{T}(\mu,\nu)$, et pour $\theta \in \mathcal{T}(\mu,\nu)$, on a $\mathbb{E}^{\mathbb{P}}[\theta] \leq 1$. Ainsi $\mathbb{P}[Y \in \mathrm{cl}\,\mathrm{dom}\theta(X,\cdot)] = 1$, où cl désigne la fermeture d'un ensemble et $\mathrm{dom}\theta$ est le domaine de θ, i.e. son lieu de finitude. Étant donné que formellement, on a $\mathrm{ri}\,\mathrm{dom}\theta(X,\cdot)$ qui est convexe relativement ouvert si on désigne par ri l'ouvert relatif d'un ensemble, et que de plus, on peut montrer que $\{\mathrm{ri}\,\mathrm{dom}\theta(x,\cdot) : x \in \mathbb{R}^d\}$ est une partition de $\mathbb{R}^d$. Ainsi les ensembles $\mathrm{ri}\,\mathrm{dom}\theta(x,\cdot)$ pour $x \in \mathbb{R}^d$ sont des candidats naturels pour être les composantes irréductibles. La stratégie consiste donc à résoudre un problème de minimisation de la taille de ces domaines. On obtient ainsi les composantes irréductibles. On définit G, une fonction qui mesure la taille d'un ensemble convexe, puis on minimise une quantité qui correspond formellement à $\int_{\mathbb{R}^d} G\Big(\mathrm{dom}\theta(x,\cdot)\Big)\mu(dx)$. La minimisation est possible grâce à une propriété de mesurabilité analytique des fonctions à valeur ensemblistes $\mathrm{ri}\,\mathrm{dom}\theta(X,\cdot)$. Cette mesurabilité permettra aussi de démontrer la possibilité de désintégrer les problèmes de transport composante par composante.

On a remarqué que les fonctions de $\mathcal{T}(\mu,\nu)$ pouvaient servir de parties de fonctions duales. Ainsi on obtient de la compacité dans le problème dual et on peut utiliser le théorème de capacitabilité de Choquet [46] de la même manière que dans [103] et [22] pour montrer un théorème de dualité général pour les fonctions mesurables. Ce théorème de dualité permet de montrer que ces composantes irréductibles sont les plus petites possibles. En effet on peut montrer qu'il existe une probabilité dont le noyau remplit toutes les composantes irréductibles. En termes mathématiques, si on note I l'application qui à $x \in \mathbb{R}^d$ associe la composante irréductible le contenant $I(x)$, il existe $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ tel que $\mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\,\widehat{\mathbb{P}}_X = \mathrm{cl}\,I(X)$, $\mu-$presque sûrement. Ce résultat est montré de manière non constructive en passant par la dualité. On regarde un problème de maximisation sur la surface recouverte par les transports martingales $\mathbb{P} \in \mathcal{M}(\mu,\nu)$: $\int_{\mathbb{R}^d} G\Big(\mathrm{conv}\,\mathrm{supp}\mathbb{P}_x\Big)\mu(dx)$. Soit $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ le maximiseur, on remarque que $N := \{Y \notin \mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\,\widehat{\mathbb{P}}_X\}$ est un ensemble $\mathcal{M}(\mu,\nu)-$polaire. Ainsi en appliquant le résultat de dualité à $\mathbf{1}_N$, on obtient l'inclusion difficile $\mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\,\widehat{\mathbb{P}}_X \supset \mathrm{cl}\,I(X)$, $\mu-$presque sûrement.

On remarque via ce qui précède que l'on peut obtenir une caractérisation des ensembles polaires de $\mathcal{M}(\mu,\nu)$ similaire au Théorème 1.3.1.

1.4.2 Dualité

Le travail sur les composantes irréductibles et sur les ensembles polaires de [58] permet de servir de base pour obtenir des résultats de dualité. Le Chapitre 3, présente un résultat de dualité quasi-sûre et quelques résultats corollaires, tels qu'une désintégration du problème de transport optimal sur les composantes irréductibles en sous-problèmes de transport optimal, ainsi qu'un théorème de monotonie.

Le chapitre précédent use d'un résultat de dualité imparfait, car les limites des fonctions duales sont en partie des limites inférieures. Le lieu problématique est la frontière relative des composantes irréductibles. Dans [57] on utilise différentes hypothèses pour dominer ces parties manquantes du lieu de la convergence. On fait l'hypothèse que les composantes sont de dimension 1 ou moins (maîtrise de la frontière qui est constituée de deux points au maximum), sont de dimension d (elles sont alors en quantité dénombrable) ou alors ne chargent pas leur frontière. Une autre hypothèse qui permet d'obtenir la dualité est l'existence des limites médiales, impliquée par l'axiome du continu. Les limites médiales, inventées par Mokobodzkyi [120] (voir [119] et [149]) sont des objets qui généralisent les limites. Toute suite bornée inférieurement possède une limite médiale, quand elle possède une limite, elle est alors égale à la limite médiale et la limite médiale de fonctions universellement mesurables est encore universellement mesurable (i.e. mesurable par rapport à toutes les mesures boréliennes).

Une fois que la convergence a lieu, une grande difficulté est de décomposer à nouveau la limite comme somme d'une fonction de X, d'une fonction de Y et d'un hedge. Pour obtenir ce résultat on montre qu'avoir la bonne forme est équivalent par dualité à un principe de monotonie sur les fonctions limites. Pour obtenir ce résultat, une autre étape consiste à obtenir des propriétés fines sur la structure des ensembles polaires.

Ce résultat de dualité permet d'obtenir que le transport optimal martingale se décompose en sous-problèmes de transport optimal martingales sur chaque composante irréductible, et sur chaque composante, la dualité est uniforme. Ce théorème possède une version plus faible qui s'affranchit du besoin des hypothèses.

De la même façon, la dualité permet d'obtenir un principe de monotonie, i.e. on prouve qu'il existe un ensemble Γ qui concentre tous les transports optimaux et qui est monotone. Un résultat très difficile et encore ouvert est l'implication inverse, le fait de savoir si le fait d'être concentré sur l'ensemble Γ implique l'optimalité du transport martingale.

Ce chapitre est également l'occasion de donner des exemples montrant qu'il ne peut y avoir de régularité suffisante de la fonction de couplage c pour permettre à un optimiseur dual d'exister quel que soient les lois μ et ν. On y trouve également un exemple où la restriction du problème à une composante irréductible n'est plus irréductible après son isolation des autres composantes.

1.4.3 Structure locale

Dans le Chapitre 4 qui correspond à [56], on s'intéresse à la structure locale de chaque noyau d'un transport optimal. On commence par démontrer un théorème qui donne la structure générale de ces noyaux sous certaines hypothèses techniques, il s'agit d'intersections entre le gradient partiel en x du coût c et une fonction affine dépendante de x. On peut utiliser la notation d'événements pour ces intersections :

$$S_0 = \{\partial_x c(x_0, Y) = A(Y)\}, \tag{1.4.10}$$

avec $A : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, affine.

La suite du chapitre consiste à expliciter les informations sur la structure des ensembles (1.4.10) que l'on peut obtenir pour différentes fonctions de couplage c. On montre en premier lieu un lien entre la géométrie algébrique réelle et la finitude des transports optimaux conditionnels, ainsi que avec le nombre maximal de fonctions de transport atteignables. En effet, si la fonction c est suffisamment régulière, son gradient partiel $\partial_x c$ se comporte localement comme un vecteur de d polynômes à d variables, c'est donc également le cas de $\partial_x c - A$. Ainsi trouver les zéros de $\partial_x c - A$ est un exercice qui se rapproche localement de trouver les zéros communs dans $\mathbb{R}^d$ de d polynômes de d variables, ce qui est le problème fondamental de la géométrie algébrique. On peut utiliser par exemple le théorème de Bézout, ici exprimé en termes formels :

Theorem 1.4.1 (Bézout)**.** *Soit $d \in \mathbb{N}$ et $(P_1, ..., P_d)$, d polynômes "complets" dans $\mathbb{R}[X_1, ..., X_d]$. Alors $|\{(P_1, ..., P_d) = 0\}| = deg(P_1)...deg(P_d)$, où les racines sont comptées avec multiplicité et incluent des racines "à l'infini".*

Voir [82]. Des précisions sur les notions de complétude, de racines à l'infini et de multiplicité seront apportées par le Chapitre 4. On montre en particulier grâce à ces notions que pour "presque toute" fonction de coût régulière, si la première marginale μ possède une densité par rapport à la mesure de Lebesgue, alors le noyau de tout transport optimal se concentre sur un ensemble discret et on peut toujours trouver des lois marginales μ et ν qui sont absolument continues par rapport à la mesure de Lebesgue qui donnent un transport optimal concentré sur $d+2$ graphes quand d est pair.

On donne également des résultats sur la structure des supports (1.4.10) dans le cas des coûts puissance de la norme Euclidienne. On montre ainsi que le nombre naturel de graphes de transport atteints dans le cas de la norme distance, pour la maximisation ou pour la minimisation est $2d$, contrairement à la Conjecture 2 de [74]. On montre également que le résultat de structure qui peut être obtenu est beaucoup plus précis que le résultat de [74] qui se contente de dire que le transport optimal conditionnel est concentré sur la frontière de Choquet de l'enveloppe convexe de son support.

1.4.4 Méthodes numériques pour le transport optimal martingale

Enfin au-delà des considérations théoriques, les problèmes de type transport optimal sont des problèmes qui ont vocation à être résolus en pratique. Il était donc naturel d'explorer l'aspect numérique des choses. Le Chapitre 5 traite de ce sujet, il s'agit de [55]. Il compare et évalue la performance théorique de plusieurs algorithmes de résolution existant et en propose un nouveau, beaucoup plus performant et résolvant du même coup le problème de marginales qui perdent leur ordre convexe à la suite de leur discrétisation, relevé par [5].

Les algorithmes existants sont, la programmation linéaire [83], l'optimisation duale semi-implicitée introduite par [77], la projection itérative de Bregman, inspirée de [52] et suggérée par [77]. On présente finalement un nouvel algorithme de Newton implicité, qui peut être vu comme une version régularisée de l'optimisation duale semi-implicitée et qui surperforme les autres.

Un autre apport qui nous semble majeur est un théorème sur la précision de l'approximation entropique du transport optimal. La littérature fait généralement mention d'une précision de l'ordre de la pénalisation entropique maximale, c'est à dire $\varepsilon\Big(\ln(N)-1\Big)$, où N est la taille de la grille de discrétisation. Cependant, sous couvert de faire des hypothèses de régularité difficilement vérifiables, on peut établir que le saut de dualité est plus faible qu'une quantité tendant vers $\varepsilon\frac{d}{2}$, valeur qui possède une surprenante universalité. Une remarque indique qu'un résultat identique peut être obtenu pour le transport classique. Malgré la difficulté à démontrer la régularité nécessaire pour la preuve du résultat, on constate ce taux de convergence sur les expériences numériques. Ainsi avec un travail substantiel, il est probablement possible de l'établir de manière rigoureuse en utilisant la régularité prodiguée par l'équation de Monge-Ampère.

Comme expliqué dans la Section 1.2.3, la résolution numérique est généralement obtenue en ajoutant une pénalisation entropique, pondérée par un poids $\varepsilon \geq 0$. Le problème dual correspondant consiste à minimiser la fonction duale V_ε définie par

$$V_\varepsilon(\varphi,\psi,h) := \mu[\varphi]+\nu[\psi]+\varepsilon\int_{\mathcal{X}\times\mathcal{Y}}\exp\left(-\frac{\varphi\oplus\psi+h^{\otimes}-c}{\varepsilon}\right)dm_0. \tag{1.4.11}$$

La minimisation partielle en fonction de chacune des variables duales implique le respect des variables primales. Dans le problème entropique primal, la mesure optimale est donnée par $d\mathbb{P} := \exp\left(-\frac{\varphi\oplus\psi+h^{\otimes}-c}{\varepsilon}\right)dm_0$. La minimisation de V_ε en φ entraîne le fait que $\mathbb{P}\circ X^{-1}=\mu$, la minimisation en ψ entraîne $\mathbb{P}\circ Y^{-1}=\nu$ et la minimisation en h donne la contrainte martingale. L'algorithme de type projection de Bregman exploite le fait que la minimisation partielle de V_ε en φ et en ψ est explicite, et la minimisation en h est semi-explicite. Ainsi il est naturel pour minimiser globalement cette fonction d'optimiser alternativement en φ, puis en ψ, puis en h et de recommencer. Notons que la minimisation en φ ne brise pas la martingalité, il est donc naturel de minimiser en h puis en φ, dans cet ordre.

Cette stratégie de minimisation par blocs n'exploite pas toute la régularité de la fonction et la connaissance explicite de sa dérivée seconde. Ainsi on propose un algorithme de Newton. Face à l'instabilité du terme exponentiel pouvant exploser facilement pour de faibles valeurs de ε, on utilise le fait que $\overline{V}_\varepsilon(\psi) := \inf_{\varphi,h} V_\varepsilon(\varphi,\psi,h)$ est également deux fois dérivable, de dérivées d'ordre 1 et 2 connues, mais avec la stabilité apportée par la minimisation en φ.

On présente également dans ce chapitre plusieurs astuces pratiques qui permettent d'améliorer la stabilité, la vitesse de convergence, ainsi que la vitesse de calcul effective de l'algorithme. Une de ces méthodes permet par la même occasion de résoudre le problème de défaut d'ordre convexe entre μ et ν remarqué par [5]. Il s'agit de l'idée de minimiser $\overline{V}_\varepsilon^\alpha := \overline{V}_\varepsilon + \alpha f$, où f est une fonction strictement convexe, sur-linéaire et $p-$homogène, avec $p > 1$. On montre que si $\alpha \longrightarrow 0$, et si on note $\mathbb{P}_\alpha$, l'unique optimiseur de $\overline{V}_\varepsilon^\alpha$, alors $\nu_\alpha := \mathbb{P}_\alpha \circ Y^{-1}$ converge vers ν_l, plus grand que μ dans l'ordre convexe et minimisant $f^*(\nu_l - \nu)$, où f^* est la conjuguée de Fenchel de f. Il suffit donc de minimiser $\overline{V}_1^\alpha$ avec $c = 0$ en faisant tendre α vers 0.

Chapter 2

Irreducible Convex Paving for Decomposition of Martingale Transport Plans

Martingale transport plans on the line are known from Beiglböck & Juillet [20] to have an irreducible decomposition on a (at most) countable union of intervals. We provide an extension of this decomposition for martingale transport plans in $\mathbb{R}^d$, $d \geq 1$. Our decomposition is a partition of $\mathbb{R}^d$ consisting of a possibly uncountable family of relatively open convex components, with the required measurability so that the disintegration is well-defined. We justify the relevance of our decomposition by proving the existence of a martingale transport plan filling these components. We also deduce from this decomposition a characterization of the structure of polar sets with respect to all martingale transport plans.
Key words. Martingale optimal transport, irreducible decomposition, polar sets.

2.1 Introduction

The problem of martingale optimal transport was introduced as the dual of the problem of robust (model-free) superhedging of exotic derivatives in financial mathematics, see Beiglböck, Henry-Labordère & Penkner [18] in discrete time, and Galichon, Henry-Labordère & Touzi [73] in continuous-time. The robust superhedging problem was introduced by Hobson [94], and was addressing specific examples of exotic derivatives by means of corresponding solutions of the Skorokhod embedding problem, see [51, 92, 93], and the survey [91].

Given two probability measures μ, ν on $\mathbb{R}^d$, with finite first order moment, martingale optimal transport differs from standard optimal transport in that the set of all coupling probability measures $\mathcal{P}(\mu, \nu)$ on the product space is reduced to the subset $\mathcal{M}(\mu, \nu)$ restricted by the martingale condition. We recall from Strassen [146] that $\mathcal{M}(\mu, \nu) \neq \emptyset$ if and only if $\mu \preceq \nu$ in the convex order, i.e. $\mu(f) \leq \nu(f)$ for all convex functions f. Notice that the inequality $\mu(f) \leq \nu(f)$ is a direct consequence of the Jensen inequality, the reverse implication follows from the Hahn-Banach theorem.

This paper focuses on the critical observation by Beiglböck & Juillet [20] that, in the one-dimensional setting $d = 1$, any such martingale interpolating probability measure $\mathbb{P}$ has a canonical decomposition $\mathbb{P} = \sum_{k \geq 0} \mathbb{P}_k$, where $\mathbb{P}_k \in \mathcal{M}(\mu_k, \nu_k)$ and μ_k is the restriction of μ to the so-called irreducible components I_k, and $\nu_k := \int_{x \in I_k} \mathbb{P}(dx, \cdot)$, supported in J_k, $k \geq 0$, is independent of the choice of $\mathbb{P}_k$. Here, $(I_k)_{k \geq 1}$ are open intervals, $I_0 := \mathbb{R} \setminus (\cup_{k \geq 1} I_k)$, and J_k is an augmentation of I_k by the inclusion of either one of the endpoints of I_k, depending on whether they are charged by the distribution $\mathbb{P}_k$. Remarkably, the irreducible components $(I_k, J_k)_{k \geq 0}$ are independent of the choice of $\mathbb{P} \in \mathcal{M}(\mu, \nu)$. To understand this decomposition, notice that convex functions in one dimension are generated by the family $f_{x_0}(x) := |x - x_0|$, $x_0 \in \mathbb{R}$. Then, in terms of the potential functions $U^\mu(x_0) := \mu(f_{x_0})$, and $U^\nu(x_0) := \nu(f_{x_0})$, $x_0 \in \mathbb{R}$, we have $\mu \preceq \nu$ if and only if $U^\mu \leq U^\nu$ and μ, ν have same mean. Then, at any contact points x_0, of the potential functions, $U^\mu(x_0) = U^\nu(x_0)$, we have equality in the underlying Jensen's equality, which means that the singularity x_0 of the underlying function f_{x_0} is not seen by the measure. In other words, the point x_0 acts as a barrier for the mass transfer in the sense that martingale transport maps do not cross the barrier x_0. Such contact points are precisely the endpoints of the intervals I_k, $k \geq 1$.

The decomposition into irreducible components plays a crucial role for the quasi-sure formulation introduced by Beiglböck, Nutz, and Touzi [22], and represents an important difference between martingale transport and standard transport. Indeed, while the martingale transport problem is affected by the quasi-sure formulation, the standard optimal transport problem is not changed. We also refer to Ekren & Soner [65] for further functional analytic aspects of this duality.

Our objective in this paper is to extend the last decomposition to an arbitrary $d-$dimensional setting, $d \geq 1$. The main difficulty is that convex functions do not have anymore such a simple generating family. Therefore, all of our analysis is based on the set of convex functions. A first extension of the last decomposition to the multi-dimensional case was achieved by Ghoussoub, Kim & Lim [74]. Motivated by the martingale monotonicity principle of Beiglböck & Juillet [20] (see also Zaev [160] for higher dimension and general linear constraints), their strategy is to find a monotone set $\Gamma \subset \mathbb{R}^d \times \mathbb{R}^d$, where the robust superhedging holds with equality, as a support of the optimal martingale transport in $\mathcal{M}(\mu,\nu)$. Denoting $\Gamma_x := \{y : (x,y) \in \Gamma\}$, this naturally induces the relation $x \,\mathrm{Rel}\, x'$ if $x \in \mathrm{ri}\,\mathrm{conv}(\Gamma_{x'})$, which is then completed to an equivalence relation $\sim$. The corresponding equivalence classes define their notion of irreducible components.

Our subsequent results differ from [74] from two perspectives. First, unlike [74], our decomposition is universal in the sense that it is not relative to any particular martingale measure in $\mathcal{M}(\mu,\nu)$ (see example 2.2.2). Second, our construction of the irreducible convex paving allows to prove the required measurability property, thus justifying completely the existence of a disintegration of martingale plans.

Finally, during the final stage of writing the present paper, we learned about the parallel work by Jan Obłój and Pietro Siorpaes [127]. Although the results are close, our approach is different from theirs. We are grateful to them for pointing to us the notions of "convex face" and "Wijsmann topology" and the relative references, which allowed us to streamline our presentation. In an earlier version of this work we used instead a topology that we called the compacted Hausdorff distance, defined as the topology generated by the countable restrictions of the space to the closed balls centered in the origin with integer radii; the two are in our case the same topologies, as the Wijsman topology is locally equivalent to the Hausdorff topology in a locally compact set. We also owe Jan and Pietro special thanks for their useful remarks and comments on a first draft of this paper privately exchanged with them.

The paper is organized as follows. Section 3.3 contains the main results of the paper, namely our decomposition into irreducible convex paving, and shows the identity with the Beiglböck & Juillet [20] notion in the one-dimensional setting. Section 5.2 collects the main technical ingredients needed for the statement of our main results, and gives the structure of polar sets. In particular, we introduce the new notions of relative face and tangent convex functions, together with the required topology on the set of such functions. The remaining sections contain the proofs of these results. In particular, the measurability of our irreducible convex paving is proved in Section 2.7.

Notation We denote by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ the completed real line, and similarly denote $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. We fix an integer $d \geq 1$. For $x \in \mathbb{R}^d$ and $r \geq 0$, we denote $B_r(x)$ the closed ball for the Euclidean distance, centered in x with radius r. We denote for simplicity $B_r := B_r(0)$. If $x \in \mathcal{X}$, and $A \subset \mathcal{X}$, where $(\mathcal{X}, \mathrm{d})$ is a metric space, $\mathrm{dist}(x,A) := \inf_{a\in A}\mathrm{d}(x,a)$. In all this paper, $\mathbb{R}^d$ is endowed with the Euclidean distance.

If V is a topological affine space and $A\subset V$ is a subset of V, $\mathrm{int}A$ is the interior of A, $\mathrm{cl}\,A$ is the closure of A, $\mathrm{aff}A$ is the smallest affine subspace of V containing A, $\mathrm{conv}A$ is the convex hull of A, $\dim(A) := \dim(\mathrm{aff}A)$, and $\mathrm{ri}A$ is the relative interior of A, which is the interior of A in the topology of $\mathrm{aff}A$ induced by the topology of V. We also denote by $\partial A := \mathrm{cl}\,A\setminus \mathrm{ri}A$ the relative boundary of A, and by λ_A the Lebesgue measure of $\mathrm{aff}A$.

The set $\mathcal{K}$ of all closed subsets of $\mathbb{R}^d$ is a Polish space when endowed with the Wijsman topology[1] (see Beer [16]). As $\mathbb{R}^d$ is separable, it follows from a theorem of Hess [86] that a function $F:\mathbb{R}^d\longrightarrow\mathcal{K}$ is Borel measurable with respect to the Wijsman topology if and only if its associated multifunction is Borel measurable, i.e.

$$F^-(V) := \{x\in\mathbb{R}^d : F(x)\cap V\neq\emptyset\} \qquad \text{is Borel measurable}$$
$$\text{for all open subset } V\subset\mathbb{R}^d.$$

The subset $\overline{\mathcal{K}}\subset\mathcal{K}$ of all the convex closed subsets of $\mathbb{R}^d$ is closed in $\mathcal{K}$ for the Wijsman topology, and therefore inherits its Polish structure. Clearly, $\overline{\mathcal{K}}$ is isomorphic to $\mathrm{ri}\overline{\mathcal{K}} := \{\mathrm{ri}K : K\in\overline{\mathcal{K}}\}$ (with reciprocal isomorphism cl). We shall identify these two isomorphic sets in the rest of this text, when there is no possible confusion.

We denote $\Omega := \mathbb{R}^d\times\mathbb{R}^d$ and define the two canonical maps

$$X : (x,y)\in\Omega\longmapsto x\in\mathbb{R}^d \quad\text{and}\quad Y : (x,y)\in\Omega\longmapsto y\in\mathbb{R}^d.$$

For $\varphi,\psi:\mathbb{R}^d\longrightarrow\bar{\mathbb{R}}$, and $h:\mathbb{R}^d\longrightarrow\mathbb{R}^d$, we denote

$$\varphi\oplus\psi := \varphi(X)+\psi(Y), \quad\text{and}\quad h^\otimes := h(X)\cdot(Y-X),$$

with the convention $\infty-\infty=\infty$.

For a Polish space $\mathcal{X}$, we denote by $\mathcal{B}(\mathcal{X})$ the collection of Borel subsets of $\mathcal{X}$, and $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\big(\mathcal{X},\mathcal{B}(\mathcal{X})\big)$. For $\mathbb{P}\in\mathcal{P}(\mathcal{X})$, we denote by $\mathcal{N}_\mathbb{P}$ the collection of all $\mathbb{P}-$null sets, $\mathrm{supp}\,\mathbb{P}$ the smallest closed support of $\mathbb{P}$, and $\overline{\mathrm{supp}}\mathbb{P} := \mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\mathbb{P}$ the smallest convex closed support of $\mathbb{P}$. For a measurable function $f:\mathcal{X}\longrightarrow\mathbb{R}$, we denote $\mathrm{dom}\,f := \{|f|<\infty\}$, and we use again the convention

[1]The Wijsman topology on the collection of all closed subsets of a metric space $(\mathcal{X},\mathrm{d})$ is the weak topology generated by $\{\mathrm{dist}(x,\cdot) : x\in\mathcal{X}\}$.

$\infty - \infty = \infty$ to define its integral, and denote

$$\mathbb{P}[f] := \mathbb{E}^{\mathbb{P}}[f] = \int_{\mathcal{X}} f d\mathbb{P} = \int_{\mathcal{X}} f(x)\mathbb{P}(dx) \quad \text{for all} \quad \mathbb{P} \in \mathcal{P}(\mathcal{X}).$$

Let $\mathcal{Y}$ be another Polish space, and $\mathbb{P} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The corresponding conditional kernel[2] $\mathbb{P}_x$ is defined $\mu-$a.e. by:

$$\mathbb{P}(dx, dy) = \mu(dx) \otimes \mathbb{P}_x(dy), \text{ where } \mu := \mathbb{P} \circ X^{-1}.$$

We denote by $\mathbb{L}^0(\mathcal{X}, \mathcal{Y})$ the set of Borel measurable maps from $\mathcal{X}$ to $\mathcal{Y}$. We denote for simplicity $\mathbb{L}^0(\mathcal{X}) := \mathbb{L}^0(\mathcal{X}, \bar{\mathbb{R}})$ and $\mathbb{L}^0_+(\mathcal{X}) := \mathbb{L}^0(\mathcal{X}, \bar{\mathbb{R}}_+)$. Let $\mathcal{A}$ be a $\sigma-$algebra of $\mathcal{X}$, we denote by $\mathbb{L}^{\mathcal{A}}(\mathcal{X}, \mathcal{Y})$ the set of $\mathcal{A}-$measurable maps from $\mathcal{X}$ to $\mathcal{Y}$. For a measure m on $\mathcal{X}$, we denote $\mathbb{L}^1(\mathcal{X}, m) := \{f \in \mathbb{L}^0(\mathcal{X}) : m[|f|] < \infty\}$. We also denote simply $\mathbb{L}^1(m) := \mathbb{L}^1(\bar{\mathbb{R}}, m)$ and $\mathbb{L}^1_+(m) := \mathbb{L}^1_+(\bar{\mathbb{R}}_+, m)$.

We denote by $\mathfrak{C}$ the collection of all finite convex functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}$. We denote by $\partial f(x)$ the corresponding subgradient at any point $x \in \mathbb{R}^d$. We also introduce the collection of all measurable selections in the subgradient, which is nonempty by Lemma 2.9.2,

$$\partial f := \left\{ p \in \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d) : p(x) \in \partial f(x) \text{ for all } x \in \mathbb{R}^d \right\}.$$

We finally denote $\underline{f}_\infty := \liminf_{n\to\infty} f_n$, for any sequence $(f_n)_{n\geq 1}$ of real numbers, or of real-valued functions.

2.2 Main results

Throughout this paper, we consider two probability measures μ and ν on $\mathbb{R}^d$ with finite first order moment, and $\mu \preceq \nu$ in the convex order, i.e. $\nu(f) \geq \mu(f)$ for all $f \in \mathfrak{C}$. Using the convention $\infty - \infty = \infty$, we may define $(\nu - \mu)(f) \in [0, \infty]$ for all $f \in \mathfrak{C}$.

We denote by $\mathcal{M}(\mu, \nu)$ the collection of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathbb{P} \circ X^{-1} = \mu$ and $\mathbb{P} \circ Y^{-1} = \nu$. Notice that $\mathcal{M}(\mu, \nu) \neq \emptyset$ by Strassen [146].

An $\mathcal{M}(\mu, \nu)-$polar set is an element of $\cap_{\mathbb{P} \in \mathcal{M}(\mu,\nu)} \mathcal{N}_{\mathbb{P}}$. A property is said to hold $\mathcal{M}(\mu, \nu)-$quasi surely (abbreviated as q.s.) if it holds on the complement of an $\mathcal{M}(\mu, \nu)-$polar set.

[2]The usual definition of a kernel requires that the map $x \mapsto \mathbb{P}_x[B]$ is Borel measurable for all Borel set $B \in \mathcal{B}(\mathbb{R}^d)$. In this paper, we only require this map to be analytically measurable.

2.2.1 The irreducible convex paving

The next first result shows the existence of a maximum support martingale transport plan, i.e. a martingale interpolating measure $\widehat{\mathbb{P}}$ whose disintegration $\widehat{\mathbb{P}}_x$ has a maximum convex hull of supports among all measures in $\mathcal{M}(\mu,\nu)$.

Theorem 2.2.1. *There exists $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ such that*

$$\text{for all } \mathbb{P} \in \mathcal{M}(\mu,\nu), \quad \overline{\operatorname{supp}}\,\mathbb{P}_X \subset \overline{\operatorname{supp}}\,\widehat{\mathbb{P}}_X, \quad \mu - a.s. \tag{2.2.1}$$

Furthermore $\overline{\operatorname{supp}}\,\widehat{\mathbb{P}}_X$ is $\mu-$a.s. unique, and we may choose this kernel so that
(i) $x \longmapsto \overline{\operatorname{supp}}\,\widehat{\mathbb{P}}_x$ *is analytically measurable*[3] $\mathbb{R}^d \longrightarrow \overline{\mathcal{K}}$,
(ii) $x \in I(x) := \operatorname{ri}\overline{\operatorname{supp}}\,\widehat{\mathbb{P}}_x$, *for all* $x \in \mathbb{R}^d$, *and* $\left\{I(x), x \in \mathbb{R}^d\right\}$ *is a partition of* $\mathbb{R}^d$.

This Theorem will be proved in Subsection 2.6.3. The ($\mu-$a.s. unique) set valued map $I(X)$ paves $\mathbb{R}^d$ by its image by (ii) of Theorem 2.2.1. By (2.2.1), this paving is stable by all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$:

$$Y \in \operatorname{cl} I(X), \quad \mathcal{M}(\mu,\nu) - \text{q.s.} \tag{2.2.2}$$

Finally, the measurability of the map I in the Polish space $\overline{\mathcal{K}}$ allows to see it as a random variable, which allows to condition probabilistic events to $X \in I$, even when these components are all $\mu-$negligible when considered apart from the others. Under the conditions of Theorem 2.2.1, we call such $I(X)$ the irreducible convex paving associated to (μ,ν).

Now we provide an important counterexample proving that for some (μ,ν) in dimension larger than 1, particular couplings in $\mathcal{M}(\mu,\nu)$ may define different pavings.

Example 2.2.2. In $\mathbb{R}^2$, we introduce $x_0 := (0,0)$, $x_1 := (1,0)$, $y_0 := x_0$, $y_{-1} := (0,-1)$, $y_1 := (0,1)$, and $y_2 := (2,0)$. Then we set $\mu := \frac{1}{2}(\delta_{x_0} + \delta_{x_1})$ and $\nu := \frac{1}{8}(4\delta_{y_0} + \delta_{y_{-1}} + \delta_{y_1} + 2\delta_{y_2})$. We can show easily that $\mathcal{M}(\mu,\nu)$ is the nonempty convex hull of $\mathbb{P}_1$ and $\mathbb{P}_2$ where

$$\mathbb{P}_1 := \frac{1}{8}\Big(4\delta_{x_0,y_0} + 2\delta_{x_1,y_2} + \delta_{x_1,y_1} + \delta_{x_1,y_{-1}}\Big)$$

and

$$\mathbb{P}_2 := \frac{1}{8}\Big(2\delta_{x_0,y_0} + \delta_{x_0,y_1} + \delta_{x_0,y_{-1}} + 2\delta_{x_1,y_0} + 2\delta_{x_1,y_2}\Big)$$

[3] Analytically measurable means measurable with respect to the smallest $\sigma-$algebra containing the analytic sets. All Borel sets are analytic and all analytic sets are universally measurable, i.e. measurable with respect to all Borel measures (see Proposition 7.41 and Corollary 7.42.1 in [30]).

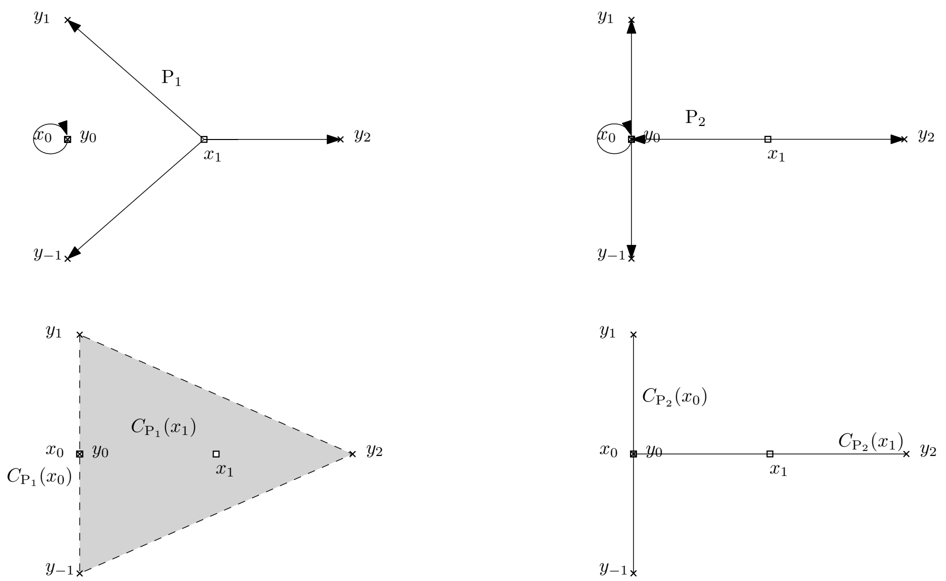


Fig. 2.1 The extreme probabilities and associated irreducible paving.

(i) *The Ghoussoub-Kim-Lim [74] (GKL, hereafter) irreducible convex paving.* Let $c_1 = \mathbf{1}_{\{X=Y\}}$, $c_2 = 1 - c_1 = \mathbf{1}_{\{X \neq Y\}}$, and notice that $\mathbb{P}_i$ is the unique optimal martingale transport plan for c_i, $i = 1,2$. Then, it follows that the corresponding $\mathbb{P}_i-$irreducible convex paving according to the definition of [74] are given by

$$C_{\mathbb{P}_1}(x_0) = \{x_0\},\ C_{\mathbb{P}_1}(x_1) = \mathrm{ri}\,\mathrm{conv}\{y_1, y_{-1}, y_2\},$$
$$\text{and } C_{\mathbb{P}_2}(x_0) = \mathrm{ri}\,\mathrm{conv}\{y_1, y_{-1}\},\ C_{\mathbb{P}_2}(x_1) = \mathrm{ri}\,\mathrm{conv}\{y_0, y_2\}.$$

Figure 2.1 shows the extreme probabilities $\mathbb{P}_1$ and $\mathbb{P}_2$, and their associated irreducible convex pavings map $C_{\mathbb{P}_1}$ and $C_{\mathbb{P}_2}$.

(ii) *Our irreducible convex paving.* The irreducible components are given by

$$I(x_0) = \mathrm{ri}\,\mathrm{conv}(y_1, y_{-1}) \quad \text{and} \quad I(x_1) = \mathrm{ri}\,\mathrm{conv}(y_1, y_{-1}, y_2).$$

To see this, we use the characterization of Proposition 2.2.4. Indeed, as $\mathcal{M}(\mu,\nu) = \mathrm{conv}(\mathbb{P}_1, \mathbb{P}_2)$, for any $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, $\mathbb{P} \ll \widehat{\mathbb{P}} := \frac{\mathbb{P}_1+\mathbb{P}_2}{2}$, and $\mathrm{supp}\,\mathbb{P}_x \subset \mathrm{conv}\Big(\mathrm{supp}\,\widehat{\mathbb{P}}_x\Big)$ for $x = x_0, x_1$. Then $I(x) = \mathrm{riconv}\Big(\mathrm{supp}\,\widehat{\mathbb{P}}_x\Big)$ for $x = x_0, x_1$ (i.e. μ-a.s.) by Proposition 2.2.4.

Remark 2.2.3. *In the one dimensional case, a convex paving which is invariant with respect to some $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ is automatically invariant with respect to all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. Given a particular coupling $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, the finest convex paving which is $\mathbb{P}-$invariant* *roughly corresponds to the GKL convex paving constructed in [74]. Then Example 2.2.2 shows that this does not hold any more in dimension greater than one.*

Furthermore, in dimension one the "restriction" $\nu_I := \int_I \mathbb{P}(dx,\cdot)$ does not depend on the choice of the coupling $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. Once again Example 2.2.2 shows that it does not hold in higher dimension. Conditions guaranteeing that this property still holds in higher dimension will be investigated in [57].

2.2.2 Behavior on the boundary of the components

For a probability measure $\mathbb{P}$ on a topological space, and a Borel subset A, $\mathbb{P}|_A := \mathbb{P}[\cdot\cap A]$ denotes its restriction to A.

Proposition 2.2.4. *We may choose $\widehat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$ in Theorem 2.2.1 so that for all $\mathbb{P}\in\mathcal{M}(\mu,\nu)$ and $y\in\mathbb{R}^d$,*

$$\mu\Big[\mathbb{P}_X[\{y\}]>0\Big]\le\mu\Big[\widehat{\mathbb{P}}_X[\{y\}]>0\Big],$$
$$\text{and}\quad \operatorname{supp}\left(\mathbb{P}_X|_{\partial I(X)}\right)\subset\overline{\operatorname{supp}}\left(\widehat{\mathbb{P}}_X|_{\partial I(X)}\right),\ \mu-a.s.$$

(i) *The set-valued maps $\underline{J}(X) := I(X)\cup\Big\{y\in\mathbb{R}^d:\nu[y]>0,\text{ and }\widehat{\mathbb{P}}_X\Big[\{y\}\Big]>0\Big\}$, and $\bar{J}(X) := I(X)\cup\overline{\operatorname{supp}}\,\widehat{\mathbb{P}}_X|_{\partial I(X)}$ are $\mu-$a.s. independent of the choice of $\widehat{\mathbb{P}}$, and $Y\in\bar{J}(X)$, $\mathcal{M}(\mu,\nu)-$q.s.*
(ii) *We may chose the kernel $\widehat{\mathbb{P}}_X$ so that the map $\bar{J}$ is convex valued, $I\subset\underline{J}\subset\bar{J}\subset\operatorname{cl}I$, and both $\underline{J}$ and $\bar{J}$ are constant on $I(x)$, for all $x\in\mathbb{R}^d$.*

The proof is reported in Subsection 2.6.3.

2.2.3 Structure of polar sets

Here we state the structure of polar sets that will be made more precise by Theorem 2.3.18.

Theorem 2.2.5. *A Borel set $N\in\mathcal{B}(\Omega)$ is $\mathcal{M}(\mu,\nu)-$polar if and only if*

$$N\subset\{X\in N_\mu\}\cup\{Y\in N_\nu\}\cup\{Y\notin J(X)\},$$

for some $(N_\mu,N_\nu)\in\mathcal{N}_\mu\times\mathcal{N}_\nu$ and a set valued map J such that $\underline{J}\subset J\subset\bar{J}$, the map J is constant on $I(x)$ for all $x\in\mathbb{R}^d$, $I(X)\subset\operatorname{conv}(J(X)\setminus N'_\nu)$, $\mu-$a.s. for all $N'_\nu\in\mathcal{N}_\nu$, and $Y\in J(X)$, $\mathcal{M}(\mu,\nu)-$q.s.

2.2.4 The one-dimensional setting

In the one-dimensional case, the decomposition into irreducible components and the structure of $\mathcal{M}(\mu,\nu)-$polar sets were introduced in Beiglböck & Juillet [20] and Beiglböck, Nutz & Touzi [22], respectively.

Let us see how the results of this paper reduce to the known concepts in the one dimensional case. First, in the one-dimensional setting, $I(x)$ consists of open intervals (at most countable number) or single points. Following [20] Proposition 2.3, we denote the full dimension components $(I_k)_{k\geq 1}$.

We also have $\underline{J}=\bar{J}$ (see Proposition 2.2.6 below) therefore, Theorem 2.2.5 is equivalent to Theorem 3.2 in [22]. Similar to $(I_k)_{k\geq 1}$, we introduce the corresponding sequence $(J_k)_{k\geq 1}$, as defined in [22]. Similar to [20], we denote by μ_k the restriction of μ to I_k, and $\nu_k := \int_{x\in I_k}\mathbb{P}[dx,\cdot]$ is independent of the choice of $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. We define the Beiglböck & Juillet (BJ)-irreducible components

$$\left(I^{BJ},J^{BJ}\right): x\mapsto \begin{cases} (I_k,J_k) & \text{if } x\in I_k, \text{ for some } k\geq 1,\\ \big(\{x\},\{x\}\big) & \text{if } x\notin \cup_k I_k.\end{cases}$$

Proposition 2.2.6. *Let $d=1$. Then $I=I^{BJ}$, and $\bar{J}=\underline{J}=J^{BJ}$, $\mu-a.s.$*

Proof. By Proposition 2.2.4 (i)-(ii), we may find $\widehat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$ such that $\overline{\operatorname{supp}}\widehat{\mathbb{P}}_X=\operatorname{cl} I(X)$, and $\overline{\operatorname{supp}}\widehat{\mathbb{P}}_X|_{\partial I(X)}=\bar{J}(X)$, $\mu-$a.s. Notice that as $\bar{J}\setminus I(\mathbb{R})$ only consists of a countable set of points, we have $\underline{J}=\bar{J}$. By Theorem 3.2 in [22], we have $Y\in J^{BJ}(X)$, $\mathcal{M}(\mu,\nu)-$q.s. Therefore, $Y\in J^{BJ}(X)$, $\widehat{\mathbb{P}}-$a.s. and we have $\bar{J}(X)\subset J^{BJ}(X)$, $\mu-$a.s.

On the other hand, let $k\geq 1$. By the fact that $u_\nu-u_\mu>0$ on I_k, together with the fact that $J_k\setminus I_k$ is constituted with atoms of ν, for any $N_\nu\in\mathcal{N}_\nu$, $J_k\subset\operatorname{conv}(J_k\setminus N_\nu)$. As $\mu=\nu$ outside of the components,

$$J^{BJ}(X)\subset\operatorname{conv}(J^{BJ}(X)\setminus N_\nu),\quad \mu-\text{a.s.} \tag{2.2.3}$$

Then by Theorem 3.2 in [22], as $\{Y\notin\bar{J}(X)\}$ is polar, we may find $N_\nu\in\mathcal{N}_\nu$ such that $J^{BJ}(X)\setminus N_\nu\subset\bar{J}(X)$, $\mu-$a.s. The convex hull of this inclusion, together with (2.2.3) gives the remaining inclusion $J^{BJ}(X)\subset\bar{J}(X)$, $\mu-$a.s.

The equality $I(X)=I^{BJ}(X)$, $\mu-$a.s. follows from the relative interior taken on the previous equality. □

2.3 Preliminaries

The proof of these results needs some preparation involving convex analysis tools.

2.3.1 Relative face of a set

For a subset $A \subset \mathbb{R}^d$ and $a \in \mathbb{R}^d$, we introduce the face of A relative to a (also denoted $a-$relative face of A):

$$\mathrm{rf}_a A := \Big\{ y \in A : (a - \varepsilon(y-a), y + \varepsilon(y-a)) \subset A, \text{ for some } \varepsilon > 0 \Big\}. \tag{2.3.1}$$

Figure 2.2 illustrates examples of relative faces of a square S, relative to some points.

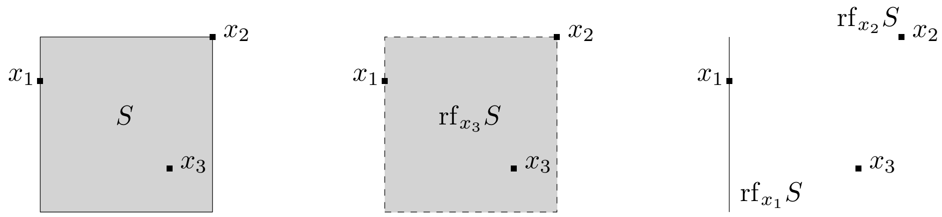


Fig. 2.2 Examples of relative faces.

For later use, we list some properties whose proofs are reported in Section 2.9. [4]

Proposition 2.3.1. (i) *For $A, A' \subset \mathbb{R}^d$, we have $\mathrm{rf}_a(A \cap A') = \mathrm{rf}_a(A) \cap \mathrm{rf}_a(A')$, and $\mathrm{rf}_a A \subset \mathrm{rf}_a A'$ whenever $A \subset A'$. Moreover, $\mathrm{rf}_a A \neq \emptyset$ iff $a \in \mathrm{rf}_a A$ iff $a \in A$.*
(ii) *For a convex A, $\mathrm{rf}_a A = \mathrm{ri} A \neq \emptyset$ iff $a \in \mathrm{ri} A$. Moreover, $\mathrm{rf}_a A$ is convex relatively open, $A \setminus \mathrm{cl}\,\mathrm{rf}_a A$ is convex, and if $x_0 \in A \setminus \mathrm{cl}\,\mathrm{rf}_a A$ and $y_0 \in A$, then $[x_0, y_0) \subset A \setminus \mathrm{cl}\,\mathrm{rf}_a A$. Furthermore, if $a \in A$, then $\dim(\mathrm{rf}_a \mathrm{cl}\, A) = \dim(A)$ if and only if $a \in \mathrm{ri} A$. In this case, we have $\mathrm{cl}\,\mathrm{rf}_a \mathrm{cl}\, A = \mathrm{cl}\,\mathrm{ri}\,\mathrm{cl}\, A = \mathrm{cl}\, A = \mathrm{cl}\,\mathrm{rf}_a A$.*

2.3.2 Tangent Convex functions

Recall the notation (2.3.1), and denote for all $\theta : \Omega \to \bar{\mathbb{R}}$:

$$\mathrm{dom}_x \theta \quad := \quad \mathrm{rf}_x \mathrm{conv}\, \mathrm{dom}\, \theta(x, \cdot).$$

[4] $\mathrm{rf}_a A$ is equal to the only relative interior of face of A containing a, where we extend the notion of face to non-convex sets. A face F of A is a nonempty subset of A such that for all $[a,b] \subset A$, with $(a,b) \cap F \neq \emptyset$, we have $[a,b] \subset F$. It is discussed in Hiriart-Urruty-Lemaréchal [89] as an extension of Proposition 2.3.7 that when A is convex, the relative interior of the faces of A form a partition of A, see also Theorem 18.2 in Rockafellar [132].

For $\theta_1, \theta_2 : \Omega \longrightarrow \mathbb{R}$, we say that $\theta_1 = \theta_2$, $\mu\otimes$pw, if

$$\mathrm{dom}_X\theta_1 = \mathrm{dom}_X\theta_2, \quad \text{and} \quad \theta_1(X,\cdot) = \theta_2(X,\cdot) \text{ on } \mathrm{dom}_X\theta_1, \; \mu-\text{a.s.}$$

The crucial ingredient for our main result is the following.

Definition 2.3.2. *A measurable function $\theta : \Omega \to \overline{\mathbb{R}}_+$ is a tangent convex function if*

$$\theta(x,\cdot) \text{ is convex, and } \theta(x,x) = 0, \text{ for all } x \in \mathbb{R}^d.$$

We denote by Θ the set of tangent convex functions, and we define

$$\Theta_\mu \;:=\; \Big\{\theta \in \mathbb{L}^0(\Omega, \overline{\mathbb{R}}_+) : \theta = \theta', \; \mu\otimes\text{pw}, \text{and } \theta \geq \theta', \text{ for some } \theta' \in \Theta\Big\}.$$

In order to introduce our main example of such functions, let

$$\mathbf{T}_p f(x,y) := f(y) - f(x) - p^{\otimes}(x,y) \geq 0, \quad \text{for all} \quad f \in \mathfrak{C}, \text{ and } p \in \partial f.$$

Then, $\mathbf{T}(\mathfrak{C}) := \{\mathbf{T}_p f : f \in \mathfrak{C}, p \in \partial f\} \subset \Theta \subset \Theta_\mu$.

Example 2.3.3. *The second inclusion is strict. Indeed, let $d = 1$, and consider the convex function $f := \infty \mathbf{1}_{(-\infty,0)}$. Then $\theta' := f(Y-X) \in \Theta$. Now let $\theta = \theta' + \sqrt{|Y-X|}$. Notice that since $\mathrm{dom}_X\theta' = \mathrm{dom}_X\theta = \{X\}$, we have $\theta' = \theta$, $\mu\otimes$pw for any measure μ, and $\theta \geq \theta'$. Therefore $\theta \in \Theta_\mu$. However, for all $x \in \mathbb{R}^d$, $\theta(x,\cdot)$ is not convex, and therefore $\theta \notin \Theta$.*

In higher dimension we may even have $X \in \mathrm{ri}\,\mathrm{dom}\theta(X,\cdot)$, and $\theta(X,\cdot)$ is not convex. Indeed, for $d = 2$, let $f : (y_1,y_2) \longmapsto \infty(\mathbf{1}_{\{|y_1|>1\}} + \mathbf{1}_{\{|y_2|>1\}})$, so that $\theta := f(Y-X) \in \Theta$. Let $x_0 := (1,0)$ and $\theta := \theta' + \mathbf{1}_{\{Y=X+x_0\}}$. Then, $\theta = \theta'$, $\mu\otimes$pw for any measure μ, and $\theta \geq \theta'$. Therefore $\theta \in \Theta_\mu$. However, $\theta \notin \Theta$ as $\theta(x,\cdot)$ is not convex for all $x \in \mathbb{R}^d$.

Proposition 2.3.4. (i) *Let $\theta \in \Theta_\mu$, then $\mathrm{dom}_X\theta = \mathrm{rf}_X\mathrm{dom}\theta(X,\cdot) \subset \mathrm{dom}\theta(X,\cdot)$, $\mu-$a.s.*
(ii) *Let $\theta_1, \theta_2 \in \Theta_\mu$, then $\mathrm{dom}_X(\theta_1+\theta_2) = \mathrm{dom}_X\theta_1 \cap \mathrm{dom}_X\theta_2$, $\mu-$a.s.*
(iii) *Θ_μ is a convex cone.*

Proof. <u>(i)</u> It follows immediately from the fact that on $\mathrm{dom}_X\theta$, we have that $\theta(X,\cdot)$ is convex and finite, $\mu-$a.s. by definition of Θ_μ. Then $\mathrm{dom}_X\theta \subset \mathrm{rf}_X\mathrm{dom}\theta(X,\cdot)$. On the other hand, as $\mathrm{dom}\theta(X,\cdot) \subset \mathrm{conv}\,\mathrm{dom}\theta(X,\cdot)$, the monotony of rf_x gives the other inclusion: $\mathrm{rf}_X\mathrm{dom}\theta(X,\cdot) \subset \mathrm{dom}_X\theta$.
<u>(ii)</u> As $\theta_1, \theta_2 \geq 0$, $\mathrm{dom}(\theta_1+\theta_2) = \mathrm{dom}\theta_1 \cap \mathrm{dom}\theta_2$. Then, for $x \in \mathbb{R}^d$, $\mathrm{conv}\,\mathrm{dom}(\theta_1(x,\cdot)+$

$\theta_2(x,\cdot)) \subset \mathrm{conv\,dom}\theta_1(x,\cdot) \cap \mathrm{conv\,dom}\theta_2(x,\cdot)$. By Proposition 2.3.1 (i),

$$\mathrm{dom}_x(\theta_1+\theta_2) \subset \mathrm{dom}_x\theta_1 \cap \mathrm{dom}_x\theta_2, \quad \text{for all} \quad x \in \mathbb{R}^d.$$

As for the reverse inclusion, notice that (i) implies that $\mathrm{dom}_X\theta_1 \cap \mathrm{dom}_X\theta_2 \subset \mathrm{dom}\theta_1(X,\cdot) \cap \mathrm{dom}\theta_2(X,\cdot) = \mathrm{dom}\big(\theta_1(X,\cdot)+\theta_2(X,\cdot)\big) \subset \mathrm{conv\,dom}\big(\theta_1(X,\cdot)+\theta_2(X,\cdot)\big)$, $\mu-$a.s. Observe that $\mathrm{dom}_x\theta_1 \cap \mathrm{dom}_x\theta_2$ is convex, relatively open, and contains x. Then,

$$\begin{aligned}
\mathrm{dom}_X\theta_1 \cap \mathrm{dom}_X\theta_2 &= \mathrm{rf}_X\big(\mathrm{dom}_X\theta_1 \cap \mathrm{dom}_X\theta_2\big) \\
&\subset \mathrm{rf}_X\Big(\mathrm{conv\,dom}\big(\theta_1(X,\cdot)+\theta_2(X,\cdot)\big)\Big) \\
&= \mathrm{dom}_X(\theta_1+\theta_2) \quad \mu-\text{a.s.}
\end{aligned}$$

<u>(iii)</u> Given (ii), this follows from direct verification. □

Definition 2.3.5. *A sequence* $(\theta_n)_{n\geq 1} \subset \mathbb{L}^0(\Omega)$ *converges* $\mu\otimes\mathrm{pw}$ *to some* $\theta \in \mathbb{L}^0(\Omega)$ *if*

$$\mathrm{dom}_X\left(\underline{\theta}_\infty\right) = \mathrm{dom}_X\theta \quad \textit{and} \quad \theta_n(X,\cdot) \longrightarrow \theta(X,\cdot), \textit{ pointwise on } \mathrm{dom}_X\theta, \ \mu-a.s.$$

Notice that the $\mu\otimes$pw-limit is $\mu\otimes$pw unique. In particular, if θ_n converges to θ, $\mu\otimes$pw, it converges as well to $\underline{\theta}_\infty$.

Proposition 2.3.6. *Let* $(\theta_n)_{n\geq 1} \subset \Theta_\mu$, *and* $\theta : \Omega \longrightarrow \bar{\mathbb{R}}_+$, *such that* $\theta_n \underset{n\to\infty}{\longrightarrow} \theta$, $\mu\otimes\mathrm{pw}$,
(i) $\mathrm{dom}_X\theta \subset \liminf_{n\to\infty}\mathrm{dom}_X\theta_n$, *$\mu-a.s.$*
(ii) *If* $\theta_n' = \theta_n$, $\mu\otimes\mathrm{pw}$, *and* $\theta_n' \geq \theta_n$, *then* $\theta_n' \underset{n\to\infty}{\longrightarrow} \theta$, $\mu\otimes\mathrm{pw}$;
(iii) $\underline{\theta}_\infty \in \Theta_\mu$.

Proof. <u>(i)</u> Let $x \in \mathbb{R}^d$, such that $\theta_n(x,\cdot)$ converges on $\mathrm{dom}_x\theta$ to $\theta(x,\cdot)$. Let $y \in \mathrm{dom}_x\theta$, let $y' \in \mathrm{dom}_x\theta$ such that $y' = x-\epsilon(y-x)$, for some $\epsilon > 0$. As $\theta_n(x,y) \underset{n\to\infty}{\longrightarrow} \theta(x,y)$, and $\theta_n(x,y') \underset{n\to\infty}{\longrightarrow} \theta(x,y')$, then for n large enough, both are finite, and $y \in \mathrm{dom}_x\theta_n$. $y \in \liminf_{n\to\infty}\mathrm{dom}_x\theta_n$, and $\mathrm{dom}_x\theta \subset \liminf_{n\to\infty}\mathrm{dom}_x\theta_n$. The inclusion is true for $\mu-$a.e. $x \in \mathbb{R}^d$, which gives the result.
<u>(ii)</u> By (i), we have $\mathrm{dom}_X\theta \subset \liminf_{n\to\infty}\mathrm{dom}_X\theta_n = \liminf_{n\to\infty}\mathrm{dom}_X\theta_n'$, $\mu-$a.s. As $\theta_n \leq \theta_n'$, $\mathrm{dom}_X\underline{\theta'}_\infty \subset \mathrm{dom}_X\underline{\theta}_\infty \subset \liminf_{n\to\infty}\mathrm{dom}_X\theta_n$, $\mu-$a.s. We denote $N_\mu \in \mathcal{N}_\mu$, the set on which $\theta_n(X,\cdot)$ does not converge to $\theta(X,\cdot)$ on $\mathrm{dom}_X\theta(X,\cdot)$. For $x \notin N_\mu$, for $y \in \mathrm{dom}_x\theta$, $\theta_n(x,y) = \theta_n'(x,y)$, for n large enough, and $\theta_n'(x,y) \underset{n\to\infty}{\longrightarrow} \theta(x,y) < \infty$. Then $\mathrm{dom}_X\theta = \mathrm{dom}_X\underline{\theta'}_\infty$, and $\theta_n'(X,\cdot)$ converges to $\theta(X,\cdot)$, on $\mathrm{dom}_X\theta$, $\mu-$a.s. We proved that $\theta_n' \underset{n\to\infty}{\longrightarrow} \theta$, $\mu\otimes$pw.
<u>(iii)</u> has its proof reported in Subsection 2.8.2 due to its length and technicality. □

The next result shows the relevance of this notion of convergence for our setting.

Proposition 2.3.7. *Let $(\theta_n)_{n\geq 1} \subset \Theta_\mu$. Then, we may find a sequence $\widehat{\theta}_n \in \mathrm{conv}(\theta_k, k \geq n)$, and $\widehat{\theta}_\infty \in \Theta_\mu$ such that $\widehat{\theta}_n \longrightarrow \widehat{\theta}_\infty$, $\mu\otimes\mathrm{pw}$ as $n \to \infty$.*

The proof is reported in Subsection 2.8.2.

Definition 2.3.8. (i) *A subset $\mathcal{T} \subset \Theta_\mu$ is $\mu\otimes\mathrm{pw}$-Fatou closed if $\underline{\theta}_\infty \in \mathcal{T}$ for all $(\theta_n)_{n\geq 1} \subset \mathcal{T}$ converging $\mu\otimes\mathrm{pw}$ (in particular, Θ_μ is $\mu\otimes\mathrm{pw}-$Fatou closed by Proposition 2.3.6* (iii)*).*
(ii) *The $\mu\otimes\mathrm{pw}-$Fatou closure of a subset $A \subset \Theta_\mu$ is the smallest $\mu\otimes\mathrm{pw}-$Fatou closed set containing A:*

$$\widehat{A} := \bigcap \Big\{ \mathcal{T} \subset \Theta_\mu : \; A \subset \mathcal{T}, \textit{ and } \mathcal{T} \; \mu\otimes\mathrm{pw}\textit{-Fatou closed} \Big\}.$$

We next introduce for $a \geq 0$ the set $\mathfrak{C}_a := \Big\{ f \in \mathfrak{C} : (\nu - \mu)(f) \leq a \Big\}$, and

$$\widehat{\mathcal{T}}(\mu,\nu) := \bigcup_{a\geq 0} \widehat{\mathcal{T}}_a, \text{ where } \widehat{\mathcal{T}}_a := \widehat{\mathbf{T}(\mathfrak{C}_a)}, \text{ and } \mathbf{T}\Big(\mathfrak{C}_a\Big) := \Big\{ \mathbf{T}_p f : f \in \mathfrak{C}_a, p \in \partial f \Big\}.$$

Proposition 2.3.9. *$\widehat{\mathcal{T}}(\mu,\nu)$ is a convex cone.*

Proof. We first prove that $\widehat{\mathcal{T}}(\mu,\nu)$ is a cone. We consider $\lambda, a > 0$, as we have $\lambda\mathfrak{C}_a = \mathfrak{C}_{\lambda a}$, and as convex combinations and inferior limit commute with the multiplication by λ, we have $\lambda\widehat{\mathcal{T}}_a = \widehat{\mathcal{T}}_{\lambda a}$. Then $\widehat{\mathcal{T}}(\mu,\nu) = \mathrm{cone}(\widehat{\mathcal{T}}_1)$, and therefore it is a cone.

We next prove that $\widehat{\mathcal{T}}_a$ is convex for all $a \geq 0$, which induces the required convexity of $\widehat{\mathcal{T}}(\mu,\nu)$ by the non-decrease of the family $\{\widehat{\mathcal{T}}_a, a \geq 0\}$. Fix $0 \leq \lambda \leq 1$, $a \geq 0$, $\theta_0 \in \widehat{\mathcal{T}}_a$, and denote $\mathcal{T}(\theta_0) := \Big\{ \theta \in \widehat{\mathcal{T}}_a : \lambda\theta_0 + (1-\lambda)\theta \in \widehat{\mathcal{T}}_a \Big\}$. In order to complete the proof, we now verify that $\mathcal{T}(\theta_0) \supset \mathbf{T}\Big(\mathfrak{C}_a\Big)$ and is $\mu\otimes\mathrm{pw}-$Fatou closed, so that $\mathcal{T}(\theta_0) = \widehat{\mathcal{T}}_a$.

To see that $\mathcal{T}(\theta_0)$ is Fatou-closed, let $(\theta_n)_{n\geq 1} \subset \mathcal{T}(\theta_0)$, converging $\mu\otimes\mathrm{pw}$. By definition of $\mathcal{T}(\theta_0)$, we have $\lambda\theta_0 + (1-\lambda)\theta_n \in \widehat{\mathcal{T}}_a$ for all n. Then, $\lambda\theta_0 + (1-\lambda)\theta_n \longrightarrow \liminf_{n\to\infty} \lambda\theta_0 + (1-\lambda)\underline{\theta}_n$, $\mu\otimes\mathrm{pw}$, and therefore $\lambda\theta_0 + (1-\lambda)\underline{\theta}_\infty \in \widehat{\mathcal{T}}_a$, which shows that $\underline{\theta}_\infty \in \mathcal{T}(\theta_0)$.

We finally verify that $\mathcal{T}(\theta_0) \supset \mathbf{T}\Big(\mathfrak{C}_a\Big)$. First, for $\theta_0 \in \mathbf{T}\Big(\mathfrak{C}_a\Big)$, this inclusion follows directly from the convexity of $\mathbf{T}\Big(\mathfrak{C}_a\Big)$, implying that $\mathcal{T}(\theta_0) = \widehat{\mathcal{T}}_a$ in this case. For general $\theta_0 \in \widehat{\mathcal{T}}_a$, the last equality implies that $\mathbf{T}\Big(\mathfrak{C}_a\Big) \subset \mathcal{T}(\theta_0)$, thus completing the proof. □

Notice that even though $\mathbf{T}(\mathfrak{C}_a) \subset \Theta$, the functions in $\widehat{\mathcal{T}}(\mu,\nu)$ may not be in Θ as they may not be convex in y on $(\mathrm{dom}_x\theta)^c$ for some $x \in \mathbb{R}^d$ (see Example 2.3.3). The following result shows that some convexity is still preserved.

Proposition 2.3.10. *For all $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, we may find $N_\mu \in \mathcal{N}_\mu$ such that for $x_1, x_2 \notin N_\mu$, $y_1, y_2 \in \mathbb{R}^d$, and $\lambda \in [0,1]$ with $\bar{y} := \lambda y_1 + (1-\lambda) y_2 \in \mathrm{dom}_{x_1}\theta \cap \mathrm{dom}_{x_2}\theta$, we have:*

$$\begin{aligned}\lambda\theta(x_1,y_1) + (1-\lambda)\theta(x_1,y_2) - \theta(x_1,\bar{y}) &= \lambda\theta(x_2,y_1) + (1-\lambda)\theta(x_2,y_2) \\ &\quad -\theta(x_2,\bar{y}) \geq 0.\end{aligned}$$

The proof of this claim is reported in Subsection 2.8.1. We observe that the statement also holds true for a finite number of points $y_1, ..., y_k$.[5]

2.3.3 Extended integral

We now introduce the extended $(\nu-\mu)-$integral:

$$\nu \widehat{\ominus} \mu[\theta] := \inf\left\{a \geq 0 : \theta \in \widehat{\mathcal{T}}_a\right\} \quad \text{for} \quad \theta \in \widehat{\mathcal{T}}(\mu,\nu).$$

Proposition 2.3.11. (i) $\mathbb{P}[\theta] \leq \nu\widehat{\ominus}\mu[\theta] < \infty$ *for all $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$ and $\mathbb{P} \in \mathcal{M}(\mu,\nu)$.*
(ii) $\nu\widehat{\ominus}\mu[\mathbf{T}_p f] = (\nu-\mu)[f]$ *for $f \in \mathfrak{C}\cap\mathbb{L}^1(\nu)$ and $p \in \partial f$.*
(iii) $\nu\widehat{\ominus}\mu$ *is homogeneous and convex.*

Proof. (i) For $a > \nu\widehat{\ominus}\mu[\theta]$, set $S^a := \left\{F \in \Theta_\mu : \mathbb{P}[F] \leq a \text{ for all } \mathbb{P} \in \mathcal{M}(\mu,\nu)\right\}$. Notice that S^a is $\mu\otimes$pw$-$Fatou closed by Fatou's lemma, and contains $\mathbf{T}(\mathfrak{C}_a)$, as for $f \in \mathfrak{C}\cap\mathbb{L}^1(\nu)$ and $p \in \partial f$, $\mathbb{P}[T_p f] = (\nu-\mu)[f]$ for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. Then S^a contains $\widehat{\mathcal{T}}_a$ as well, which contains θ. Hence, $\theta \in S^a$ and $\mathbb{P}[\theta] \leq a$ for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. The required result follows from the arbitrariness of $a > \nu\widehat{\ominus}\mu[\theta]$.
(ii) Let $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. For $p \in \partial f$, notice that $T_p f \in \mathbf{T}(\mathfrak{C}_a) \subset \widehat{\mathcal{T}_a}$ for some $a = (\nu-\mu)[f]$, and therefore $(\nu-\mu)[f] \geq \nu\widehat{\ominus}\mu[T_p f]$. Then, the result follows from the inequality $(\nu-\mu)[f] = \mathbb{P}[T_p f] \leq \nu\widehat{\ominus}\mu[T_p f]$.
(iii) Similarly to the proof of Proposition 2.3.9, we have $\lambda\widehat{\mathcal{T}}_a = \widehat{\mathcal{T}}_{\lambda a}$, for all $\lambda, a > 0$. Then with the definition of $\nu\widehat{\ominus}\mu$ we have easily the homogeneity.

To see that the convexity holds, let $0 < \lambda < 1$, and $\theta, \theta' \in \widehat{\mathcal{T}}(\mu,\nu)$ with $a > \nu\widehat{\ominus}\mu[\theta]$, $a' > \nu\widehat{\ominus}\mu[\theta']$, for some $a, a' > 0$. By homogeneity and convexity of $\widehat{\mathcal{T}}_1$, $\lambda\theta + (1-\lambda)\theta' \in \widehat{\mathcal{T}}_{\lambda a + (1-\lambda)a'}$, so that $\nu\widehat{\ominus}\mu[\lambda\theta + (1-\lambda)\theta'] \leq \lambda a + (1-\lambda)a'$. The required convexity property now follows from arbitrariness of $a > \nu\widehat{\ominus}\mu[\theta]$ and $a' > \nu\widehat{\ominus}\mu[\theta']$. □

The following compacteness result plays a crucial role.

[5] This is not a direct consequence of Proposition 2.3.10, as the barycentre $\bar{y}$ has to be in $\mathrm{dom}_{x_1}\theta \cap \mathrm{dom}_{x_2}\theta$.

Lemma 2.3.12. *Let* $(\theta_n)_{n\geq 1} \subset \widehat{\mathcal{T}}(\mu,\nu)$ *be such that* $\sup_{n\geq 1} \nu\widehat{\ominus}\mu(\theta_n) < \infty$*. Then we can find a sequence* $\widehat{\theta}_n \in \mathrm{conv}(\theta_k, k\geq n)$ *such that*

$$\underline{\widehat{\theta}}_\infty \in \widehat{\mathcal{T}}(\mu,\nu), \quad \widehat{\theta}_n \longrightarrow \underline{\widehat{\theta}}_\infty, \ \mu\otimes\mathrm{pw}, \quad \textit{and} \quad \nu\widehat{\ominus}\mu(\underline{\widehat{\theta}}_\infty) \leq \liminf_{n\to\infty} \nu\widehat{\ominus}\mu(\theta_n).$$

Proof. By possibly passing to a subsequence, we may assume that $\lim_{n\to\infty}(\nu\widehat{\ominus}\mu)(\theta_n)$ exists. The boundedness of $\nu\widehat{\ominus}\mu(\theta_n)$ ensures that this limit is finite. We next introduce the sequence $\widehat{\theta}_n$ of Proposition 2.3.7. Then $\widehat{\theta}_n \longrightarrow \widehat{\theta}_\infty$, $\mu\otimes\mathrm{pw}$, and therefore $\underline{\widehat{\theta}}_\infty \in \widehat{\mathcal{T}}(\mu,\nu)$, because of the convergence $\widehat{\theta}_n \longrightarrow \underline{\widehat{\theta}}_\infty$, $\mu\otimes\mathrm{pw}$. As $(\nu\widehat{\ominus}\mu)(\widehat{\theta}_n) \leq \sup_{k\geq n}(\nu\widehat{\ominus}\mu)(\theta_k)$ by Proposition 2.3.11 (iii), we have that $\infty > \lim_{n\to\infty}(\nu\widehat{\ominus}\mu)(\theta_n) = \lim_{n\to\infty}\sup_{k\geq n}(\nu\widehat{\ominus}\mu)(\theta_k) \geq \limsup_{n\to\infty}(\nu\widehat{\ominus}\mu)(\widehat{\theta}_n)$. Set $l := \limsup_{n\to\infty}\nu\widehat{\ominus}\mu(\widehat{\theta}_n)$. For $\epsilon > 0$, we consider $n_0 \in \mathbb{N}$ such that $\sup_{k\geq n_0}\nu\widehat{\ominus}\mu(\widehat{\theta}_k) \leq l+\epsilon$. Then for $k \geq n_0$, $\widehat{\theta}_k \in \widehat{\mathcal{T}}_{l+2\epsilon}(\mu,\nu)$, and therefore $\underline{\widehat{\theta}}_\infty = \liminf_{k\geq n_0}\widehat{\theta}_k \in \widehat{\mathcal{T}}_{l+2\epsilon}(\mu,\nu)$, implying $\nu\widehat{\ominus}\mu(\widehat{\theta}) \leq l+2\epsilon \longrightarrow l$, as $\epsilon \to 0$. Finally, $\liminf_{n\to\infty}(\nu\widehat{\ominus}\mu)(\theta_n) \geq \nu\widehat{\ominus}\mu(\underline{\widehat{\theta}}_\infty)$. □

2.3.4 The dual irreducible convex paving

Our final ingredient is the following measurement of subsets $K \subset \mathbb{R}^d$:

$$G(K) := \dim(K) + g_K(K) \text{ where } g_K(dx) := \frac{e^{-\frac{1}{2}|x|^2}}{(2\pi)^{\frac{1}{2}\dim K}}\lambda_K(dx), \tag{2.3.2}$$

Notice that $0 \leq G \leq d+1$ and, for any convex subsets $C_1 \subset C_2$ of $\mathbb{R}^d$, we have

$$G(C_1) = G(C_2) \quad \text{iff} \quad \mathrm{ri}C_1 = \mathrm{ri}C_2 \quad \text{iff} \quad \mathrm{cl}\,C_1 = \mathrm{cl}\,C_2. \tag{2.3.3}$$

For $\theta \in \mathbb{L}^0_+(\Omega), A \in \mathcal{B}(\mathbb{R}^d)$, we introduce the following map from $\mathbb{R}^d$ to the set $\overline{\mathcal{K}}$ of all relatively open convex subsets of $\mathbb{R}^d$:

$$K_{\theta,A}(x) := \mathrm{rf}_x\mathrm{conv}(\mathrm{dom}\theta(x,\cdot)\setminus A) = \mathrm{dom}_X(\theta + \infty 1_{\mathbb{R}^d\times A}), \tag{2.3.4}$$

for all $x \in \mathbb{R}^d$. We recall that a function is universally measurable if it is measurable with respect to every complete probability measure that measures all Borel subsets.

Lemma 2.3.13. *For* $\theta \in \mathbb{L}^0_+(\Omega)$ *and* $A \in \mathcal{B}(\mathbb{R}^d)$*, we have:*
(i) $\mathrm{cl\,conv\,dom}\theta(X,\cdot) : \mathbb{R}^d \longmapsto \overline{\mathcal{K}}$, $\mathrm{dom}_X\theta : \mathbb{R}^d \longmapsto \mathrm{ri}\overline{\mathcal{K}}$, *and* $K_{\theta,A} : \mathbb{R}^d \longmapsto \mathrm{ri}\overline{\mathcal{K}}$ *are universally measurable;*
(ii) $G : \overline{\mathcal{K}} \longrightarrow \mathbb{R}$ *is Borel measurable;*

(iii) *if* $A \in \mathcal{N}_\nu$, *and* $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, *then up to a modification on a* $\mu-$*null set,* $K_{\theta,A}(\mathbb{R}^d) \subset \mathrm{ri}\,\overline{\mathcal{K}}$ *is a partition of* $\mathbb{R}^d$ *with* $x \in K_{\theta,A}(x)$ *for all* $x \in \mathbb{R}^d$.

The proof is reported in Subsections 2.4.2 for (iii), 2.7.1 for (ii), and 2.7.2 for (i). The following property is a key-ingredient for our dual decomposition into irreducible convex paving.

Proposition 2.3.14. *For all* $(\theta, N_\nu) \in \widehat{\mathcal{T}}(\mu,\nu) \times \mathcal{N}_\nu$, *we have the inclusion* $Y \in \mathrm{cl}\,K_{\theta,N_\nu}(X)$, $\mathcal{M}(\mu,\nu)-$*q.s.*

Proof. For an arbitrary $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, we have by Proposition 2.3.11 that $\mathbb{P}[\theta] < \infty$. Then, $\mathbb{P}\Big[\mathrm{dom}\theta \setminus (\mathbb{R}^d \times N_\nu)\Big] = 1$ i.e. $\mathbb{P}[Y \in D_X] = 1$ where $D_x := \mathrm{conv}(\mathrm{dom}\theta(x,\cdot) \setminus N_\nu)$. By the martingale property of $\mathbb{P}$, we deduce that

$$X \;=\; \mathbb{E}^{\mathbb{P}}[Y \mathbf{1}_{Y \in D_X} | X] \;=\; (1-\Lambda)E_K + \Lambda E_D, \quad \mu-\text{a.s.}$$

Where $\Lambda := \mathbb{P}_X[Y \in D_X \setminus \mathrm{cl}\,K_{\theta,N_\nu}(X)]$, $E_D := \mathbb{E}^{P_X}[Y | Y \in D_X \setminus \mathrm{cl}\,K_{\theta,N_\nu}(X)]$, $E_K := \mathbb{E}^{\mathbb{P}_X}[Y | Y \in \mathrm{cl}\,K_{\theta,N_\nu}(X)]$, and $\mathbb{P}_X$ is the conditional kernel to X of $\mathbb{P}$. We have $E_K \in \mathrm{cl}\,\mathrm{rf}_X D_X \subset D_X$ and $E_D \in D_X \setminus \mathrm{cl}\,\mathrm{rf}_X D_X$ because of the convexity of $D_X \setminus \mathrm{cl}\,\mathrm{rf}_X D_X$ given by Proposition 2.3.1 (ii) (D_X is convex). The lemma also gives that if $\Lambda \neq 0$, then $\mathbb{E}^{\mathbb{P}}[Y|X] = \Lambda E_D + (1-\Lambda)E_K \in D_X \setminus \mathrm{cl}\,K_{\theta,N_\nu}(X)$. This implies that

$$\begin{aligned}
\{\Lambda \neq 0\} \subset \{\mathbb{E}^{\mathbb{P}}[Y|X] \in D_X \setminus \mathrm{cl}\,K_{\theta,N_\nu}(X)\} &\subset \{\mathbb{E}^{\mathbb{P}}[Y|X] \notin K_{\theta,N_\nu}(X)\} \\
&\subset \{\mathbb{E}^{\mathbb{P}}[Y|X] \neq X\}.
\end{aligned}$$

Then $\mathbb{P}[\Lambda \neq 0] = 0$, and therefore $\mathbb{P}\Big[Y \in D_X \setminus \mathrm{cl}\,K_{\theta,N_\nu}(X)\Big] = 0$. Since $\mathbb{P}[Y \in D_X] = 1$, this shows that $\mathbb{P}[Y \in \mathrm{cl}\,K_{\theta,N_\nu}(X)] = 1$. □

In view of Proposition 2.3.14 and Lemma 2.3.13 (iii), we introduce the following optimization problem which will generate our irreducible convex paving decomposition:

$$\inf_{(\theta,N_\nu) \in \widehat{\mathcal{T}}(\mu,\nu) \times \mathcal{N}_\nu} \mu[G(K_{\theta,N_\nu})]. \tag{2.3.5}$$

The following result gives another possible definition for the irreducible paving.

Proposition 2.3.15. (i) *We may find a* μ*-a.s. unique universally measurable minimizer* $\widehat{K} := K_{\widehat{\theta},\widehat{N}_\nu} : \mathbb{R}^d \to \overline{\mathcal{K}}$ *of* (2.3.5), *for some* $(\widehat{\theta}, \widehat{N}_\nu) \in \widehat{\mathcal{T}}(\mu,\nu) \times \mathcal{N}_\nu$;

(ii) *for all* $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$ *and* $N_\nu \in \mathcal{N}_\nu$, *we have* $\widehat{K}(X) \subset K_{\theta,N_\nu}(X)$, μ*-a.s;*

(iii) *we have the equality* $\widehat{K}(X) = I(X)$, $\mu-$*a.s.*

In item (i), the measurability of $\widehat{K}$ is induced by Lemma 2.3.13 (i). Existence and uniqueness, together with (ii), are proved in Subsection 2.4.1. finally, the proof of (iii) is reported in Subsection 2.6.3, and is a consequence of Theorem 2.3.18 below. Proposition 2.3.15 provides a characterization of the irreducible convex paving by means of an optimality criterion on $\left(\widehat{\mathcal{T}}(\mu,\nu),\mathcal{N}_\nu\right)$.

Remark 2.3.16. We illustrate how to get the components from optimization Problem (2.3.5) in the case of Example 2.2.2. A $\widehat{\mathcal{T}}(\mu,\nu)$ function minimizing this problem (with $N_\nu := \emptyset \in \mathcal{N}_\nu$) is $\widehat{\theta} := \liminf_{n\to\infty} \mathbf{T}_{p_n} f_n$, where $f_n := nf$, $p_n := np$ for some $p \in \partial f$, and

$$f(x) := \operatorname{dist}\Big(x, \operatorname{aff}(y_1, y_{-1})\Big) + \operatorname{dist}\Big(x, \operatorname{aff}(y_1, y_2)\Big) + \operatorname{dist}\Big(x, \operatorname{aff}(y_2, y_{-1})\Big).$$

One can easily check that $\mu[f] = \nu[f]$ for any $n \geq 1$: $f, f_n \in \mathfrak{C}_0$. These functions separate $I(x_0)$, $I(x_1)$ and $\Big(I(x_0) \cup I(x_1)\Big)^c$.

Notice that in this example, we may as well take $\theta := 0$, and $N_\nu := \{y_{-1}, y_0, y_1, y_2\}^c$, which minimizes the optimization problem as well.

2.3.5 Structure of polar sets

Let $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, we denote the set valued map $J_\theta(X) := \operatorname{dom}\theta(X,\cdot) \cap \bar{J}(X)$, where $\bar{J}$ is introduced in Proposition 2.2.4.

Remark 2.3.17. Let $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, up to a modification on a $\mu-$null set, we have

$$\begin{gathered} Y \in J_\theta(X),\ \mathcal{M}(\mu,\nu)-\text{q.s},\ \ \underline{J} \subset J_\theta \subset \bar{J}, \\ \text{and}\ \ J_\theta \text{ constant on } I(x), \ \ \text{for all}\ \ x \in \mathbb{R}^d. \end{gathered}$$

These claims are a consequence of Proposition 2.6.2 together with Lemma 2.6.6.

Our second main result shows the importance of these set-valued maps:

Theorem 2.3.18. *A Borel set $N \in \mathcal{B}(\Omega)$ is $\mathcal{M}(\mu,\nu)-$polar if and only if*

$$N \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J_\theta(X)\},$$

for some $(N_\mu, N_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$ and $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$.

The proof is reported in Section 2.6.3. This Theorem is an extension of the one-dimensional characterization of polar sets given by Theorem 3.2 in [22], indeed in dimension one $\underline{J} = J_\theta = \bar{J}$ by Proposition 2.2.6, together with the inclusion in Remark 2.3.17.

We conclude this section by reporting a duality result which will be used for the proof of Theorem 2.3.18. We emphasize that the primal objective of the accompanying paper De March [57] is to push further this duality result so as to be suitable for the robust superhedging problem in financial mathematics.

Let $c:\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}_+$, and consider the martingale optimal transport problem:

$$\mathbf{S}_{\mu,\nu}(c) \;:=\; \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{P}[c]. \tag{2.3.6}$$

Notice from Proposition 2.3.11 (i) that $\mathbf{S}_{\mu,\nu}(\theta)\le\nu\widehat{\ominus}\mu(\theta)$ for all $\theta\in\widehat{\mathcal{T}}$. We denote by $\mathcal{D}^{mod}_{\mu,\nu}(c)$ the collection of all (φ,ψ,h,θ) in $\mathbb{L}^1_+(\mu)\times\mathbb{L}^1_+(\nu)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)\times\widehat{\mathcal{T}}(\mu,\nu)$ such that

$$\mathbf{S}_{\mu,\nu}(\theta)=\nu\widehat{\ominus}\mu(\theta), \quad \text{and} \quad \varphi\oplus\psi+h^{\otimes}+\theta \;\ge\; c, \text{ on } \{Y\in\operatorname{aff}K_{\theta,\{\psi=\infty\}}(X)\}.$$

The last inequality is an instance of the so-called robust superhedging property. The dual problem is defined by:

$$\mathbf{I}^{mod}_{\mu,\nu}(c) \;:=\; \inf_{(\varphi,\psi,h,\theta)\in\mathcal{D}^{mod}_{\mu,\nu}(c)} \mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu(\theta).$$

Notice that for any measurable function $c:\Omega\longrightarrow\mathbb{R}_+$, any $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, and any $(\varphi,\psi,h,\theta)\in\mathcal{D}^{mod}_{\mu,\nu}(c)$, we have $\mathbb{P}[c]\le\mu[\varphi]+\nu[\psi]+\mathbb{P}[\theta]\le\mu[\varphi]+\nu[\psi]+\mathbf{S}_{\mu,\nu}(\theta)$, as a consequence of the above robust superhedging inequality, together with the fact that $Y\in\operatorname{aff}K_{\theta,\{\psi=\infty\}}(X)$, $\mathcal{M}(\mu,\nu)$-q.s. by Proposition 2.3.14 This provides the weak duality:

$$\mathbf{S}_{\mu,\nu}(c) \;\le\; \mathbf{I}^{mod}_{\mu,\nu}(c). \tag{2.3.7}$$

The following result states that the strong duality holds for upper semianalytic functions. We recall that a function $f:\mathbb{R}^d\to\mathbb{R}$ is upper semianalytic if $\{f\ge a\}$ is an analytic set for any $a\in\mathbb{R}$. In particular, a Borel function is upper semianalytic.

Theorem 2.3.19. *Let $c:\Omega\to\overline{\mathbb{R}}_+$ be upper semianalytic. Then we have*
(i) $\mathbf{S}_{\mu,\nu}(c)=\mathbf{I}^{mod}_{\mu,\nu}(c)$;
(ii) *If in addition $\mathbf{S}_{\mu,\nu}(c)<\infty$, then existence holds for the dual problem $\mathbf{I}^{mod}_{\mu,\nu}(c)$.*

Remark 2.3.20. *By allowing h to be infinite in some directions, orthogonal to $\operatorname{aff}K_{\theta,\{\psi=\infty\}}(X)$, together with the convention $\infty-\infty=\infty$, we may reformulate the robust superhedging inequality in the dual set as $\varphi\oplus\psi+h^{\otimes}+\theta\ge c$ pointwise.*

2.3.6 One-dimensional tangent convex functions

For an interval $J \subset \mathbb{R}$, we denote $\mathfrak{C}(K)$ the set of convex functions on K.

Proposition 2.3.21. *Let $d=1$, then*

$$\widehat{\mathcal{T}}(\mu,\nu) = \Big\{ \sum_k \mathbf{1}_{\{X\in I_k\}} \mathbf{T}_{p_k} f_k : f_k \in \mathfrak{C}(J_k),\ p_k \in \partial f_k,\ \sum_k (\nu_k - \mu_k)[f_k] < \infty \Big\},$$

$\mathcal{M}(\mu,\nu)-$q.s. Furthermore, for all such $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$ and its corresponding $(f_k)_k$, we have $\nu \widehat{\ominus} \mu(\theta) = \sum_k (\nu_k - \mu_k)[f_k]$.

Proof. As all functions we consider are null on the diagonal, equality on $\cup_k I_k \times J_k$ implies $\mathcal{M}(\mu,\nu)-$q.s. equality by Theorem 3.2 in [22]. Let $\mathcal{L}$ be the set on the right hand side.

Step 1: We first show $\subset$, for $a \geq 0$, we denote $\mathcal{L}_a := \{\theta \in \mathcal{L} : \sum_k (\nu_k - \mu_k)[f_k] \leq a\}$. Notice that $\mathcal{L}_a$ contains $\mathbf{T}(\mathfrak{C}_a)$ modulo $\mathcal{M}(\mu,\nu)-$q.s. equality. We intend to prove that $\mathcal{L}_a$ is $\mu\otimes$pw$-$Fatou closed, so as to conclude that $\widehat{\mathcal{T}}_a \subset \mathcal{L}_a$, and therefore $\widehat{\mathcal{T}}(\mu,\nu) \subset \mathcal{L}$ by the arbitrariness of $a \geq 0$.

Let $\theta_n = \sum_k \mathbf{1}_{\{X\in I_k\}} \mathbf{T}_{p_k^n} f_k^n \in \mathcal{L}_a$ converging $\mu\otimes$pw. By Proposition 2.3.6, $\theta_n \longrightarrow \theta := \underline{\theta}_\infty$, $\mu\otimes$pw. For $k \geq 1$, let $x_k \in I_k$ be such that $\theta_n(x_k,\cdot) \longrightarrow \theta(x_k,\cdot)$ on $\mathrm{dom}_{x_k}\theta$, and set $f_k := \theta(x_k,\cdot)$. By Proposition 5.5 in [22], f_k is convex on I_k, finite on J_k, and we may find $p_k \in \partial f_k$ such that for $x \in I_k$, $\theta(x,\cdot) = \mathbf{T}_{p_k} f_k(x,\cdot)$. Hence, $\theta = \sum_k \mathbf{1}_{\{X\in I_k\}} \mathbf{T}_{p_k} f_k$, and $\sum_k (\nu_k - \mu_k)[f_k] \leq a$ by Fatou's Lemma, implying that $\theta \in \mathcal{L}_a$, as required.

Step 2: To prove the reverse inclusion $\supset$, let $\theta = \sum_k \mathbf{1}_{\{X\in I_k\}} \mathbf{T}_{p_k} f_k \in \mathcal{L}$. Let f_k^ϵ be a convex function defined by $f_k^\epsilon := f_k$ on $J_k^\epsilon = J_k \cap \{x \in J_k : \mathrm{dist}(x, J_k^c) \geq \epsilon\}$, and f_k^ϵ affine on $\mathbb{R} \setminus J_k^\epsilon$. Set $\epsilon_n := n^{-1}$, $\bar{f}_n = \sum_{k=1}^n f_k^{\epsilon_n}$, and define the corresponding subgradient in $\partial \bar{f}_n$:

$$\bar{p}_n := p_k + \nabla(\bar{f}_n - f_k^{\varepsilon_n}) \ \text{ on } \ J_k^{\varepsilon_n},\ k \geq 1, \ \text{ and } \ \bar{p}_n := \nabla \bar{f}_n \ \text{ on } \ \mathbb{R} \setminus \Big(\cup_k J_k^{\varepsilon_n} \Big).$$

We have $(\nu-\mu)[\bar{f}_n] = \sum_{k=1}^n (\nu_k - \mu_k)[f_k^{\epsilon_n}] \leq \sum_k (\nu_k - \mu_k)[f_k] < \infty$. By definition, we see that $\mathbf{T}_{\bar{p}_n} \bar{f}_n$ converges to θ pointwise on $\cup_k (I_k)^2$ and to $\theta_*(x,y) := \liminf_{\bar{y}\to y} \theta(x,\bar{y})$ on $\cup_k I_k \times \mathrm{cl}\, I_k$ where, using the convention $\infty - \infty = \infty$, $\theta' := \theta - \theta_* \geq 0$, and $\theta' = 0$ on $\cup_k (I_k)^2$. For $k \geq 1$, set $\Delta_k^l := \theta'(x_k, l_k)$, and $\Delta_k^r := \theta'(x_k, l_k)$ where $I_k = (l_k, r_k)$, and we fix some $x_k \in I_k$. For positive $\epsilon < \frac{r_k - l_k}{2}$, and $M \geq 0$, consider the piecewise affine function $g_k^{\epsilon,M}$ with break points $l_k + \epsilon$ and $r_k - \epsilon$, and:

$$g_k^{\epsilon,M}(l_k) = M \wedge \Delta_k^l, \quad g_k^{\epsilon,M}(r_k) = M \wedge \Delta_k^r, \quad g_k^{\epsilon,M}(l_k+\epsilon) = 0, \text{ and } g_k^{\epsilon,M}(r_k - \epsilon) = 0.$$

Notice that $g_k^{\epsilon,M}$ is convex, and converges pointwise to $g_k^M := M \wedge \theta'\left(\frac{l_k+r_k}{2},\cdot\right)$ on J_k, as $\epsilon \to 0$, with

$$\begin{aligned}(\nu_k-\mu_k)(g_k^M) &= \nu_k[\{l_k\}](M\wedge\Delta_k^l)+\nu_k[r_k](M\wedge\Delta_k^r)\\ &\leq (\nu_k-\mu_k)[f_k]-(\nu_k-\mu_k)[(f_k)_*] \leq (\nu_k-\mu_k)[f_k],\end{aligned}$$

where $(f_k)_*$ is the lower semi-continuous envelop of f_k. Then by the dominated convergence theorem, we may find positive $\epsilon_k^{n,M} < \frac{r_k-l_k}{2n}$ such that

$$(\nu_k-\mu_k)\left(g_k^{\epsilon_k^{n,M},M}\right) \leq (\nu_k-\mu_k)[f_k]+2^{-k}/n.$$

Now let $\bar g_n = \sum_{k=1}^n g_k^{\epsilon_k^{n,n},n}$, and $\bar p_n' \in \partial \bar g_n$. Notice that $\mathbf{T}_{\bar p_n'} g_n \longrightarrow \theta'$ pointwise on $\cup_k I_k \times J_k$, furthermore, $(\nu-\mu)(\bar g_n) \leq \sum_k(\nu_k-\mu_k)[f_k]+1/n \leq \sum_k(\nu_k-\mu_k)[f_k]+1<\infty$.

Then we have $\theta_n := \mathbf{T}_{\bar p_n}\bar f_n + \mathbf{T}_{\bar p_n'}\bar g_n$ converges to θ pointwise on $\cup_k I_k\times J_k$, and therefore $\mathcal{M}(\mu,\nu)-$q.s. by Theorem 3.2 in [22]. Since $(\nu-\mu)(\bar f_n+\bar g_n)$ is bounded, we see that $(\theta_n)_{n\geq 1}\subset \mathbf{T}(\mathfrak{C}_a)$ for some $a\geq 0$. Notice that θ_n may fail to converge $\mu\otimes$pw. However, we may use Proposition 2.3.7 to get a sequence $\widehat\theta_n \in \mathrm{conv}(\theta_k, k\geq n)$, and $\widehat\theta_\infty \in \Theta_\mu$ such that $\widehat\theta_n \longrightarrow \widehat\theta_\infty$, $\mu\otimes$pw as $n\to\infty$, and satisfies the same $\mathcal{M}(\mu,\nu)-$q.s. convergence properties than θ_n. Then $\underline{\widehat\theta}_\infty \in \widehat{\mathcal{T}}(\mu,\nu)$, and $\underline{\widehat\theta}_\infty = \theta$, $\mathcal{M}(\mu,\nu)-$q.s. □

2.4 The irreducible convex paving

2.4.1 Existence and uniqueness

Proof of Proposition 2.3.15 (i) The measurability follows from Lemma 2.3.13. We first prove the existence of a minimizer for the problem (2.3.5). Let m denote the infimum in (2.3.5), and consider a minimizing sequence $(\theta_n, N_\nu^n)_{n\in\mathbb{N}} \subset \widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu$ with $\mu[G(K_{\theta_n,N_\nu^n})] \leq m+1/n$. By possibly normalizing the functions θ_n, we may assume that $\nu\widehat\ominus\mu(\theta_n)\leq 1$. Set

$$\widehat\theta := \textstyle\sum_{n\geq 1} 2^{-n}\theta_n \quad \text{and} \quad \widehat{N}_\nu := \cup_{n\geq 1} N_\nu^n \in \mathcal{N}_\nu.$$

Notice that $\widehat\theta$ is well-defined as the pointwise limit of a sequence of the nonnegative functions $\widehat\theta_N := \sum_{n\leq N} 2^{-n}\theta_n$. Since $\nu\widehat\ominus\mu\left[\widehat\theta_N\right] \leq \sum_{n\geq 1}2^{-n}<\infty$ by convexity of $\nu\widehat\ominus\mu$, $\widehat\theta_N \longrightarrow \widehat\theta$, pointwise, and $\widehat\theta \in \widehat{\mathcal{T}}(\mu,\nu)$ by Lemma 2.3.12, since any convex extraction of $(\widehat\theta_n)_{n\geq 1}$ still converges to $\widehat\theta$. Since $\theta_n^{-1}(\{\infty\}) \subset \widehat\theta^{-1}(\{\infty\})$, it follows from the

definition of $\widehat{N}_\nu$ that $m+1/n \geq \mu[G(K_{\theta_n,N_\nu^n})] \geq \mu[G(K_{\widehat{\theta},\widehat{N}_\nu})]$, hence $\mu[G(K_{\widehat{\theta},\widehat{N}_\nu})]=m$ as $\widehat{\theta}\in\widehat{\mathcal{T}}(\mu,\nu)$, and $\widehat{N}_\nu\in\mathcal{N}_\nu$.
(ii) For an arbitrary $(\theta,N_\nu)\in\widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu$, we define $\bar{\theta}:=\theta+\widehat{\theta}\in\widehat{\mathcal{T}}(\mu,\nu)$ and $\bar{N}_\nu:=\widehat{N}_\nu\cup N_\nu$, so that $K_{\bar{\theta},\bar{N}_\nu}\subset K_{\widehat{\theta},\widehat{N}_\nu}$. By the non-negativity of θ and $\widehat{\theta}$, we have $m\leq \mu[G\big(K_{\bar{\theta},\bar{N}_\nu}\big)]\leq \mu[G\big(K_{\widehat{\theta},\widehat{N}_\nu}\big)]=m$. Then $G\big(K_{\bar{\theta},\bar{N}_\nu}\big)=G\big(K_{\widehat{\theta},\widehat{N}_\nu}\big)$, μ-a.s. By (2.3.3), we see that, $\mu-$a.s. $K_{\bar{\theta},\bar{N}_\nu}=K_{\widehat{\theta},\widehat{N}_\nu}$ and $K_{\bar{\theta},\bar{N}_\nu}=K_{\widehat{\theta},\widehat{N}_\nu}=\widehat{K}$. This shows that $\widehat{K}\subset K_{\theta,N_\nu}$, μ-a.s. □

2.4.2 Partition of the space in convex components

This section is dedicated to the proof of Lemma 2.3.13 (iii), which is an immediate consequence of Proposition 2.4.1 (ii).

Proposition 2.4.1. *Let $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$, and $A\in\mathcal{B}(\mathbb{R}^d)$. We may find $N_\mu\in\mathcal{N}_\mu$ such that:*
(i) *for all $x_1,x_2\notin N_\mu$ with $K_{\theta,A}(x_1)\cap K_{\theta,A}(x_2)\neq\emptyset$, we have $K_{\theta,A}(x_1)=K_{\theta,A}(x_2)$;*
(ii) *if $A\in\mathcal{N}_\nu$, then $x\in K_{\theta,A}(x)$ for $x\notin N_\mu$, and up to a modification of $K_{\theta,A}$ on N_μ, $K_{\theta,A}(\mathbb{R}^d)$ is a partition of $\mathbb{R}^d$ such that $x\in K_{\theta,A}(x)$ for all $x\in\mathbb{R}^d$.*

Proof. (i) Let N_μ be the $\mu-$null set given by Proposition 2.3.10 for θ. For $x_1,x_2\notin N_\mu$, we suppose that we may find $\bar{y}\in K_{\theta,A}(x_1)\cap K_{\theta,A}(x_2)$. Consider $y\in\mathrm{cl}\,K_{\theta,A}(x_1)$. As $K_{\theta,A}(x_1)$ is open in its affine span, $y':=\bar{y}+\frac{\epsilon}{1-\epsilon}(\bar{y}-y)\in K_{\theta,A}(x_1)$ for $0<\epsilon<1$ small enough. Then $\bar{y}=\epsilon y+(1-\epsilon)y'$, and by Proposition 2.3.10, we get

$$\epsilon\theta(x_1,y)+(1-\epsilon)\theta(x_1,y')-\theta(x_1,\bar{y})=\epsilon\theta(x_2,y)+(1-\epsilon)\theta(x_2,y')-\theta(x_2,\bar{y})$$

By convexity of $\mathrm{dom}_{x_i}\theta$, $K_{\theta,A}(x_i)\subset\mathrm{dom}_{x_i}\theta\subset\mathrm{dom}\theta(x_i,\cdot)$. Then $\theta(x_1,y')$, $\theta(x_1,\bar{y})$, $\theta(x_2,y')$, and $\theta(x_2,\bar{y})$ are finite and

$$\theta(x_1,y)<\infty \quad\text{if and only if}\quad \theta(x_2,y)<\infty.$$

Therefore $\mathrm{cl}\,K_{\theta,A}(x_1)\cap\mathrm{dom}\theta(x_1,\cdot)\subset\mathrm{dom}\theta(x_2,\cdot)$. We also have obviously the inclusion $\mathrm{cl}\,K_{\theta,A}(x_2)\cap\mathrm{dom}\theta(x_2,\cdot)\subset\mathrm{dom}\theta(x_2,\cdot)$. Subtracting A, we get

$$\Big(\mathrm{cl}\,K_{\theta,A}(x_1)\cap\mathrm{dom}\theta(x_1,\cdot)\setminus A\Big)\cup\Big(\mathrm{cl}\,K_{\theta,A}(x_2)\cap\mathrm{dom}\theta(x_2,\cdot)\setminus A\Big)\subset\mathrm{dom}\theta(x_2,\cdot)\setminus A.$$

Taking the convex hull and using the fact that the relative face of a set is included in itself, we see that $\mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big)\subset\mathrm{conv}\Big(\mathrm{dom}\theta(x_2,\cdot)\setminus A\Big)$. Notice that, as $K_{\theta,A}(x_2)$ is defined as the x_2-relative face of some set, either $x_2\in\mathrm{ri}K_{\theta,A}(x)$ or $K_{\theta,A}(x)=\emptyset$ by the properties of rf_{x_2}. The second case is excluded as we assumed that $K_{\theta,A}(x_1)\cap K_{\theta,A}(x_2)\neq\emptyset$. Therefore, as $K_{\theta,A}(x_1)$ and $K_{\theta,A}(x_2)$ are convex sets intersecting in relative interior points and $x_2\in\mathrm{ri}K_{\theta,A}(x_2)$, it follows from Lemma 2.9.1 that $x_2\in\mathrm{ri}\,\mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big)$. Then by Proposition 2.3.1 (ii),

$$\begin{aligned}\mathrm{rf}_{x_2}\mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big) &= \mathrm{ri}\,\mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big)\\ &= \mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big).\end{aligned}$$

Then, we have $\mathrm{conv}\Big(K_{\theta,A}(x_1)\cup K_{\theta,A}(x_2)\Big)\subset \mathrm{rf}_{x_2}\mathrm{conv}\Big(\mathrm{dom}\theta(x_2,\cdot)\setminus A\Big)=K_{\theta,A}(x_2)$, as rf_{x_2} is increasing. Therefore $K_{\theta,A}(x_1)\subset K_{\theta,A}(x_2)$ and by symmetry between x_1 and x_2, $K_{\theta,A}(x_1)=K_{\theta,A}(x_2)$.
(ii) We suppose that $A\in\mathcal{N}_\nu$. First, notice that, as $K_{\theta,A}(X)$ is defined as the $X-$relative face of some set, either $x\in K_{\theta,A}(x)$ or $K_{\theta,A}(x)=\emptyset$ for $x\in\mathbb{R}^d$ by the properties of rf_x. Consider $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. By Proposition 2.3.14, $\mathbb{P}[Y\in \mathrm{cl}\,K_{\theta,A}(X)]=1$. As $\mathrm{supp}(\mathbb{P}_X)\subset \mathrm{cl}\,K_{\theta,A}(X)$, μ-a.s., $K_{\theta,A}(X)$ is non-empty, which implies that $x\in K_{\theta,A}(x)$. Hence, $\{X\in K_{\theta,A}(X)\}$ holds outside the set $N^0_\mu:=\{\mathrm{supp}(\mathbb{P}_X)\not\subset \mathrm{cl}\,K_{\theta,A}(X)\}\in\mathcal{N}_\mu$. Then we just need to have this property to replace N_μ by $N_\mu\cup N^0_\mu\in\mathcal{N}_\mu$.

Finally, to get a partition of $\mathbb{R}^d$, we just need to redefine $K_{\theta,A}$ on N_μ. If $x\in\bigcup_{x'\notin N_\mu}K_{\theta,A}(x')$ then by definition of N_μ, the set $K_{\theta,A}(x')$ is independent of the choice of $x'\notin N_\mu$ such that $x\in K_{\theta,A}(x')$: indeed, if $x_1',x_2'\notin N_\mu$ satisfy $x\in K_{\theta,A}(x_1')\cap K_{\theta,A}(x_2')$, then in particular $K_{\theta,A}(x_1')\cap K_{\theta,A}(x_2')$ is non-empty, and therefore $K_{\theta,A}(x_1')=K_{\theta,A}(x_2')$ by (i). We set $K_{\theta,A}(x):=K_{\theta,A}(x')$. Otherwise, if $x\notin\bigcup_{x'\notin N_\mu}K_{\theta,A}(x')$, we set $K_{\theta,A}(x):=\{x\}$ which is trivially convex and relatively open. With this definition, $K_{\theta,A}(\mathbb{R}^d)$ is a partition of $\mathbb{R}^d$. $\square$

2.5 Proof of the duality

For simplicity, we denote $\mathrm{Val}(\xi):=\mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu(\theta)$, for $\xi:=(\varphi,\psi,h,\theta)\in\mathcal{D}^{mod}_{\mu,\nu}(c)$.

2.5.1 Existence of a dual optimizer

Lemma 2.5.1. *Let $c,c_n:\Omega\longrightarrow\overline{\mathbb{R}_+}$, and $\xi_n\in\mathcal{D}^{mod}_{\mu,\nu}(c_n)$, $n\in\mathbb{N}$, be such that*

$$c_n\longrightarrow c,\ \textit{pointwise, and}\ \mathrm{Val}(\xi_n)\longrightarrow \mathbf{S}_{\mu,\nu}(c)<\infty\ \textit{as}\ n\to\infty.$$

Then there exists $\xi\in\mathcal{D}^{mod}_{\mu,\nu}(c)$ such that $\mathrm{Val}(\xi_n)\longrightarrow\mathrm{Val}(\xi)$ as $n\to\infty$.

Proof. Denote $\xi_n := (\varphi_n, \psi_n, h_n, \theta_n)$, and observe that the convergence of $\mathrm{Val}(\xi_n)$ implies that the sequence $\big(\mu(\varphi_n), \nu(\psi_n), \nu\widehat{\ominus}\mu(\theta_n)\big)_n$ is bounded, by the non-negativity of φ_n, ψ_n and $\nu\widehat{\ominus}\mu(\theta_n)$. We also recall the robust superhedging inequality

$$\varphi_n \oplus \psi_n + h_n^{\otimes} + \theta_n \geq c_n, \quad \text{on } \{Y \in \mathrm{aff} K_{\theta_n,\{\psi_n=\infty\}}(X)\}, \quad n \geq 1. \tag{2.5.1}$$

<u>*Step 1.*</u> By Komlòs Lemma together with Lemma 2.3.12, we may find a sequence $(\widehat{\varphi}_n, \widehat{\psi}_n, \widehat{\theta}_n) \in \mathrm{conv}\{(\varphi_k, \psi_k, \theta_k), k \geq n\}$ such that

$$\widehat{\varphi}_n \longrightarrow \varphi := \underline{\widehat{\varphi}}_\infty, \ \mu-\text{a.s.}, \ \widehat{\psi}_n \longrightarrow \psi := \underline{\widehat{\psi}}_\infty, \ \nu-\text{a.s., and}$$
$$\widehat{\theta}_n \longrightarrow \widetilde{\theta} := \underline{\widehat{\theta}}_\infty \in \widehat{\mathcal{T}}(\mu,\nu), \ \mu\otimes\text{pw}.$$

Set $\varphi := \infty$ and $\psi := \infty$ on the corresponding non-convergence sets, and observe that $\mu[\varphi] + \nu[\psi] < \infty$, by the Fatou Lemma, and therefore $N_\mu := \{\varphi = \infty\} \in \mathcal{N}_\mu$ and $N_\nu := \{\psi = \infty\} \in \mathcal{N}_\nu$. We denote by $(\widehat{h}_n, \widehat{c}_n)$ the same convex extractions from $\{(h_k, c_k), k \geq n\}$, so that the sequence $\widehat{\xi}_n := (\widehat{\varphi}_n, \widehat{\psi}_n, \widehat{h}_n, \widehat{\theta}_n)$ inherits from (2.5.1) the robust superhedging property, as for $\theta_1, \theta_2 \in \widehat{\mathcal{T}}(\mu,\nu)$, $\psi_1, \psi_2 \in \mathbb{L}^1_+(\mathbb{R}^d)$, and $0 < \lambda < 1$, we have $\mathrm{aff} K_{\lambda\theta_1+(1-\lambda)\theta_2,\{\lambda\psi_1+(1-\lambda)\psi_2=\infty\}} \subset \mathrm{aff} K_{\theta_1,\{\psi_1=\infty\}} \cap \mathrm{aff} K_{\theta_2,\{\psi_2=\infty\}}$:

$$\widehat{\varphi}_n \oplus \widehat{\psi}_n + \widehat{\theta}_n + \widehat{h}_n^{\otimes} \geq \widehat{c}_n \geq 0, \quad \text{pointwise on } \mathrm{aff} K_{\widehat{\theta}_n,\{\widehat{\psi}_n=\infty\}}(X). \tag{2.5.2}$$

<u>*Step 2.*</u> Next, notice that $l_n := \left(\widehat{h}_n^{\otimes}\right)^- := \max\left(-\widehat{h}_n^{\otimes}, 0\right) \in \Theta$ for all $n \in \mathbb{N}$. By the convergence Proposition 2.3.7, we may find convex combinations $\widehat{l}_n := \sum_{k\geq n} \lambda_k^n l_k \longrightarrow l := \underline{\widehat{l}}_\infty$, $\mu\otimes$pw. Updating the definition of φ by setting $\varphi := \infty$ on the zero $\mu-$measure set on which the last convergence does not hold on $(\partial^x \mathrm{dom} l)^c$, it follows from (2.5.2), and the fact that $\mathrm{aff} K_{\bar{\theta},\{\psi=\infty\}} \subset \liminf_{n\to\infty} \mathrm{aff} K_{\widehat{\theta}_n,\{\widehat{\psi}_n=\infty\}}$, that

$$l = \underline{\widehat{l}}_\infty \leq \liminf_n \sum_{k\geq n} \lambda_k^n \left(\widehat{\varphi}_k \oplus \widehat{\psi}_k + \widehat{\theta}_k\right) \leq \varphi \oplus \psi + \bar{\theta},$$
$$\text{pointwise on } \left\{Y \in \mathrm{aff} K_{\bar{\theta},\{\psi=\infty\}}(X)\right\}.$$

where $\bar{\theta} := \liminf_n \sum_{k\geq n} \lambda_k^n \widehat{\theta}_k \in \widehat{\mathcal{T}}(\mu,\nu)$. As $\{\varphi = \infty\} \in \mathcal{N}_\mu$, by possibly enlarging N_μ, we assume without loss of generality that $\{\varphi = \infty\} \subset N_\mu$, we see that $\mathrm{dom}\, l \supset (N_\mu^c \times N_\nu^c) \cap \mathrm{dom}\bar{\theta} \cap \left\{Y \in \mathrm{aff} K_{\bar{\theta},\{\psi=\infty\}}(X)\right\}$, and therefore

$$K_{\bar{\theta},\{\psi=\infty\}}(X) \subset \mathrm{dom}_X l' \subset \mathrm{dom}\, l'(X,\cdot), \quad \mu\text{-a.s.} \tag{2.5.3}$$

Step 3. Let $\widehat{\widehat{h}}_n := \sum_{k\ge n}\lambda_k^n \widehat{h}_k$. Then $b_n := \widehat{\widehat{h}}_n^{\otimes} + \widehat{l}_n = \sum_{k\ge n}\lambda_k^n \left(\widehat{h}_k^{\otimes}\right)^+$ defines a non-negative sequence in Θ. By Proposition 2.3.7, we may find a sequence $\widehat{b}_n =: \widetilde{h}_n^{\otimes} + \widetilde{l}_n \in \mathrm{conv}(b_k, k\ge n)$ such that $\widehat{b}_n \longrightarrow b := \underline{\widehat{b}}_\infty$, $\mu\otimes\mathrm{pw}$, where b takes values in $[0,\infty]$. $\widehat{b}_n(X,\cdot) \longrightarrow b(X,\cdot)$ pointwise on $\mathrm{dom}_X b$, $\mu-$a.s. Combining with (2.5.3), this shows that

$$\widetilde{h}_n^{\otimes}(X,\cdot) \longrightarrow (b-l)(X,\cdot), \quad \text{pointwise on} \quad \mathrm{dom}_X b \cap K_{\bar\theta,\{\psi=\infty\}}(X), \quad \mu-\text{a.s.}$$

$(b-l)(X,\cdot) = \liminf_n \widetilde{h}_n^{\otimes}(X,\cdot)$, pointwise on $K_{\bar\theta,\{\psi=\infty\}}(X)$ (where l is a limit of l_n), $\mu-$a.s. Clearly, on the last convergence set, $(b-l)(X,\cdot) > -\infty$ on $K_{\bar\theta,\{\psi=\infty\}}(X)$, and we now argue that $(b-l)(X,\cdot) < \infty$ on $K_{\bar\theta,\{\psi=\infty\}}(X)$, therefore $K_{\bar\theta,\{\psi=\infty\}}(X) \subset \mathrm{dom}_X b$, so that we deduce from the structure of $\widetilde{h}_n^{\otimes}$ that the last convergence holds also on $\mathrm{aff}\, K_{\bar\theta,\{\psi=\infty\}}(X)$:

$$\begin{aligned} \widetilde{h}_n^{\otimes}(X,\cdot) \longrightarrow (b-l)(X,\cdot) \qquad & =: h^{\otimes}(X,\cdot), \\ \text{pointwise on} \quad & K_{\bar\theta,\{\psi=\infty\}}(X), \quad \mu-\text{a.s.} \end{aligned} \tag{2.5.4}$$

Indeed, let x be an arbitrary point of the last convergence set, and consider an arbitrary $y \in K_{\bar\theta,\{\psi=\infty\}}(x)$. By the definition of $K_{\bar\theta,\{\psi=\infty\}}$, we have $x \in \mathrm{ri}K_{\bar\theta,\{\psi=\infty\}}(x)$, and we may therefore find $y' \in K_{\bar\theta,\{\psi=\infty\}}(x)$ with $x = py + (1-p)y'$ for some $p\in(0,1)$. Then, $p\widetilde{h}_n^{\otimes}(x,y) + (1-p)\widetilde{h}_n^{\otimes}(x,y') = 0$. Sending $n\to\infty$, by concavity of the $\liminf$, this provides $p(b-l)(x,y) + (1-p)(b-l)(x,y') \le 0$, so that $(b-l)(x,y') > -\infty$ implies that $(b-l)(x,y) < \infty$.

Step 4. Notice that by dual reflexivity of finite dimensional vector spaces, (2.5.4) defines a unique $h(X)$ in the vector space $\mathrm{aff}\, K_{\bar\theta,\{\psi=\infty\}}(X) - X$, such that $(b-l)(X,\cdot) = h^{\otimes}(X,\cdot)$ on $\mathrm{aff}\, K_{\bar\theta,\{\psi=\infty\}}(X)$. At this point, we have proceeded to a finite number of convex combinations which induce a final convex combination with coefficients $(\bar\lambda_n^k)_{k\ge n\ge 1}$. Denote $\bar\xi_n := \sum_{k\ge n}\bar\lambda_n^k \xi_k$, and set $\theta := \underline{\bar\theta}_\infty$. Then, applying this convex combination to the robust superhedging inequality (2.5.1), we obtain by sending $n\to\infty$ that $(\varphi\oplus\psi + h^{\otimes} + \theta)(X,\cdot) \ge c(X,\cdot)$ on $\mathrm{aff}\, K_{\bar\theta,\{\psi=\infty\}}(X)$, $\mu-$a.s. and $\varphi\oplus\psi + h^{\otimes} + \theta = \infty$ on the complement μ null-set. As θ is the liminf of a convex extraction of $(\widehat{\theta}_n)$, we have $\theta \ge \underline{\widehat{\theta}}_\infty = \bar\theta$, and therefore $\mathrm{aff}\, K_{\theta,\{\psi=\infty\}} \subset \mathrm{aff}\, K_{\bar\theta,\{\psi=\infty\}}$. This shows that the limit point $\xi := (\varphi,\psi,h,\theta)$ satisfies the pointwise robust superhedging inequality

$$\varphi\oplus\psi + \theta + h^{\otimes} \ge c, \quad \text{on} \ \left\{Y \in \mathrm{aff}\, K_{\theta,\{\psi=\infty\}}(X)\right\}. \tag{2.5.5}$$

<u>*Step 5.*</u> By Fatou's Lemma and Lemma 2.3.12, we have

$$\mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu[\theta] \leq \liminf_n \mu[\varphi_n]+\nu[\psi_n]+\nu\widehat{\ominus}\mu[\theta_n] = \mathbf{S}_{\mu,\nu}(c). \tag{2.5.6}$$

By (2.5.5), we have $\mu[\varphi]+\nu[\psi]+\mathbb{P}[\theta]\geq\mathbb{P}[c]$ for all $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. Then, $\mu[\varphi]+\nu[\psi]+\mathbf{S}_{\mu,\nu}[\theta]\geq\mathbf{S}_{\mu,\nu}[c]$. By Proposition 2.3.11 (i), we have $\mathbf{S}_{\mu,\nu}[\theta]\leq\nu\widehat{\ominus}\mu[\theta]$, and therefore

$$\mathbf{S}_{\mu,\nu}[c]\leq\mu[\varphi]+\nu[\psi]+\mathbf{S}_{\mu,\nu}[\theta]\leq\mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu[\theta]\leq\mathbf{S}_{\mu,\nu}(c),$$

by (2.5.6). Then we have $\mathrm{Val}(\xi)=\mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu[\theta]=\mathbf{S}_{\mu,\nu}(c)$ and $\mathbf{S}_{\mu,\nu}[\theta]=\nu\widehat{\ominus}\mu[\theta]$, so that $\xi\in\mathcal{D}^{mod}_{\mu,\nu}(c)$. □

2.5.2 Duality result

We first prove the duality in the lattice USC_b of bounded upper semicontinuous fonctions $\Omega\longrightarrow\overline{\mathbb{R}_+}$. This is a classical result using the Hahn-Banach Theorem, the proof is reported for completeness.

Lemma 2.5.2. *Let $f\in\mathrm{USC}_b$, then $\mathbf{S}_{\mu,\nu}(f)=\mathbf{I}^{mod}_{\mu,\nu}(f)$.*

Proof. We have $\mathbf{S}_{\mu,\nu}(f)\leq\mathbf{I}^{mod}_{\mu,\nu}(f)$ by weak duality (4.1.5), let us now show the converse inequality $\mathbf{S}_{\mu,\nu}(f)\geq\mathbf{I}^{mod}_{\mu,\nu}(f)$. By standard approximation technique, it suffices to prove the result for bounded continuous f. We denote by $\mathrm{C}_l(\mathbb{R}^d)$ the set of continuous mappings $\mathbb{R}^d\to\mathbb{R}$ with linear growth at infinity, and by $\mathrm{C}_b(\mathbb{R}^d,\mathbb{R}^d)$ the set of continuous bounded mappings $\mathbb{R}^d\longrightarrow\mathbb{R}^d$. Define

$$\mathcal{D}(f):=\left\{(\bar\varphi,\bar\psi,\bar h)\in\mathrm{C}_l(\mathbb{R}^d)\times\mathrm{C}_l(\mathbb{R}^d)\times\mathrm{C}_b(\mathbb{R}^d,\mathbb{R}^d):\ \bar\varphi\oplus\bar\psi+\bar h^{\otimes}\geq f\right\},$$

and the associated $\mathbf{I}_{\mu,\nu}(f):=\inf_{(\bar\varphi,\bar\psi,\bar h)\in\mathcal{D}(f)}\mu(\bar\varphi)+\nu(\bar\psi)$. By Theorem 2.1 in Zaev [160], and Lemma 2.5.3 below, we have

$$\mathbf{S}_{\mu,\nu}(f)\ =\ \mathbf{I}_{\mu,\nu}(f)\ =\ \inf_{(\bar\varphi,\bar\psi,\bar h)\in\mathcal{D}(f)}\mu(\bar\varphi)+\nu(\bar\psi)\ \geq\ \mathbf{I}^{mod}_{\mu,\nu}(f),$$

which provides the required result. □

Proof of Theorem 2.3.19 The existence of a dual optimizer follows from a direct application of the compactness Lemma 2.5.1 to a minimizing sequence of robust superhedging strategies.

As for the extension of duality result of Lemma 2.5.2 to non-negative upper semianalytic functions, we shall use the capacitability theorem of Choquet, similar to [103] and [22]. Let $[0,\infty]^\Omega$ denote the set of all nonnegative functions $\Omega \to [0,\infty]$, and USA_+ the sublattice of upper semianalytic functions. Note that USC_b is stable by infimum.

Recall that a USC_b-capacity is a monotone map $\mathbf{C}:[0,\infty]^\Omega \longrightarrow [0,\infty]$, sequentially continuous upwards on $[0,\infty]^\Omega$, and sequentially continuous downwards on USC_b. The Choquet capacitability theorem states that a USC_b-capacity $\mathbf{C}$ extends to USA_+ by:

$$\mathbf{C}(f) \;=\; \sup\Big\{\mathbf{C}(g) : g \in \mathrm{USC}_b \text{ and } g \le f\Big\} \quad \text{for all} \quad f \in \mathrm{USA}_+.$$

In order to prove the required result, it suffices to verify that $\mathbf{S}_{\mu,\nu}$ and $\mathbf{I}^{mod}_{\mu,\nu}$ are USC_b-capacities. As $\mathcal{M}(\mu,\nu)$ is weakly compact, it follows from similar argument as in Prosposition 1.21, and Proposition 1.26 in Kellerer [103] that $\mathbf{S}_{\mu,\nu}$ is a USC_b-capacity. We next verify that $\mathbf{I}^{mod}_{\mu,\nu}$ is a USC_b-capacity. Indeed, the upwards continuity is inherited from $\mathbf{S}_{\mu,\nu}$ together with the compactness lemma 2.5.1, and the downwards continuity follows from the downwards continuity of $\mathbf{S}_{\mu,\nu}$ together with the duality result on USC_b of Lemma 2.5.2. □

Lemma 2.5.3. *Let $c:\Omega \to \overline{\mathbb{R}}_+$, and $(\bar\varphi,\bar\psi,\bar h) \in \mathcal{D}(c)$. Then, we may find $\xi \in \mathcal{D}^{mod}_{\mu,\nu}(c)$ such that* $\mathrm{Val}(\xi) = \mu[\bar\varphi] + \nu[\bar\psi]$.

Proof. Let us consider $(\bar\varphi,\bar\psi,\bar h) \in \mathcal{D}(c)$. Then $\bar\varphi \oplus \bar\psi + \bar h^{\otimes} \ge c \ge 0$, and therefore

$$\bar\psi(y) \ge f(y) := \sup_{x\in\mathbb{R}^d} -\bar\varphi(x) - \bar h(x)\cdot(y-x).$$

Clearly, f is convex, and $f(x) \ge -\bar\varphi(x)$ by taking value $x = y$ in the supremum. Hence $\bar\psi - f \ge 0$ and $\bar\varphi + f \ge 0$, implying in particular that f is finite on $\mathbb{R}^d$. As $\bar\varphi$ and $\bar\psi$ have linear growth at infinity, f is in $\mathbb{L}^1(\nu)\cap\mathbb{L}^1(\mu)$. We have $f \in \mathfrak{C}_a$ for $a = \nu[f] - \mu[f] \ge 0$. Then we consider $p \in \partial f$ and denote $\theta := \mathbf{T}_p f$. $\theta \in \mathbf{T}\Big(\mathfrak{C}_a\Big) \subset \widehat{\mathcal{T}}(\mu,\nu)$. Then denoting $\varphi := \bar\varphi + f$, $\psi := \bar\psi - f$, and $h := \bar h + p$, we have $\xi := (\varphi,\psi,h,\theta) \in \mathcal{D}^{mod}_{\mu,\nu}(c)$ and

$$\mu[\bar\varphi] + \nu[\bar\psi] = \mu[\varphi] + \nu[\psi] + (\nu-\mu)[f] = \mu[\varphi] + \nu[\psi] + \nu\widehat{\ominus}\mu[\theta] = \mathrm{Val}(\xi).$$

□

2.6 Polar sets and maximum support martingale plan

2.6.1 Boundary of the dual paving

Consider the optimization problems:

$$\inf_{(\theta,N_\nu)\in\widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu}\mu\Big[G(R_{\theta,N_\nu})\Big],\tag{2.6.1}$$

with $R_{\theta,N_\nu}:=\mathrm{cl\,conv}\Big(\mathrm{dom}\theta(X,\cdot)\cap\partial\widehat{K}(X)\cap N_\nu^c\Big)$, and for $y\in\mathbb{R}^d$ we consider

$$\inf_{(\theta,N_\nu)\in\widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu}\mu\Big[y\in\partial\widehat{K}(X)\cap\mathrm{dom}\theta(X,\cdot)\cap N_\nu^c\Big].\tag{2.6.2}$$

These problems are well defined by the following measurability result, whose proof is reported in Subsection 2.7.2.

Lemma 2.6.1. *Let $F:\mathbb{R}^d\longrightarrow\mathcal{K}$, $\gamma-$measurable. Then we may find $N_\gamma\in\mathcal{N}_\gamma$ such that $\mathbf{1}_{Y\in F(X)}\mathbf{1}_{X\notin N_\gamma}$ is Borel measurable, and if $X\in\mathrm{ri}F(X)$ convex, $\gamma-$a.s., then $\mathbf{1}_{Y\in\partial F(X)}\mathbf{1}_{X\notin N_\gamma}$ is Borel measurable as well.*

By the same argument than that of the proof of existence and uniqueness in Proposition 2.3.15, we see that the problem (2.6.1), (resp. (2.6.2) for $y\in\mathbb{R}^d$) has an optimizer $(\theta^*,N_\nu^*)\in\widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu$, (resp. $(\theta_y^*,N_{\nu,y}^*)\in\widehat{\mathcal{T}}(\mu,\nu)\times\mathcal{N}_\nu$). Furthermore, we have that the map $D:=R_{\theta^*,N_\nu^*}$, (resp. $D_y(x):=\{y\}$ if $y\in\partial\widehat{K}(x)\cap\mathrm{dom}\theta_y^*(x,\cdot)\cap N_{\nu,y}^*$, and $\emptyset$ otherwise, for $x\in\mathbb{R}^d$) does not depend on the choice of (θ^*,N_ν^*), (resp. θ_y^*) up to a $\mu-$negligible modification.

We define $\bar{K}:=D\cup\widehat{K}$, and $K_\theta(X):=\mathrm{dom}\theta(X,\cdot)\cap\bar{K}(X)$ for $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$. Notice that if $y\in\mathbb{R}^d$ is not an atom of ν, we may chose $N_{\nu,y}$ containing y, which means that Problem (2.6.2) is non-trivial only if y is an atom of ν. We denote atom(ν), the (at most countable) atoms of ν, and define the mapping $\underline{K}:=(\cup_{y\in\mathrm{atom}(\nu)}D_y)\cup\widehat{K}$,

Proposition 2.6.2. *Let $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$. Up to a modification on a $\mu-$null set, we have*
(i) *$\bar{K}$ is convex valued, moreover $Y\in\bar{K}(X)$, and $Y\in K_\theta(X)$, $\mathcal{M}(\mu,\nu)-$q.s.*
(ii) *$\widehat{K}\subset\underline{K}\subset K_\theta\subset\bar{K}\subset\mathrm{cl}\,\widehat{K}$,*
(iii) *$\underline{K}$, K_θ, and $\bar{K}$ are constant on $\widehat{K}(x)$, for all $x\in\mathbb{R}^d$.*

Proof. (i) For $x\in\mathbb{R}^d$, $\bar{K}(x)=D(x)\cup\widehat{K}(x)$. Let $y_1,y_2\in\bar{K}(x)$, $\lambda\in(0,1)$, and set $y:=\lambda y_1+(1-\lambda)y_2$. If $y_1,y_2\in\widehat{K}(x)$, or $y_1,y_2\in D(x)$, we get $y\in\bar{K}(x)$ by convexity of $\widehat{K}(x)$, or $D(x)$. Now, up to switching the indices, we may assume that $y_1 \in \widehat{K}(x)$, and $y_2 \in D(x)\setminus\widehat{K}(x)$. As $D(x)\setminus\widehat{K}(x) \subset \partial\widehat{K}(x)$, $y \in \widehat{K}(x)$, as $\lambda > 0$. Then $y \in \bar{K}(x)$. Hence, $\bar{K}$ is convex valued.

Since $\Big(\mathrm{dom}\theta^*(X,\cdot)\setminus N_\nu^*\Big)\cap\Big((\mathrm{cl}\,\widehat{K})\setminus\widehat{K}\Big) \subset R_{\theta^*,N_\nu^*}$, we have the inclusion $\Big(\mathrm{dom}\theta^*(X,\cdot)\setminus N_\nu^*\Big)\cap\mathrm{cl}\,\widehat{K} \subset R_{\theta^*,N_\nu^*}\cup\widehat{K} = \bar{K}$. Then, as $Y \in \mathrm{dom}\theta^*(X,\cdot)\setminus N_\nu^*$, and $Y \in \mathrm{cl}\,\widehat{K}(X)$, $Y \in \bar{K}(X)$, $\mathcal{M}(\mu,\nu)-$q.s.

Let $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, then $Y \in \mathrm{dom}\theta(X,\cdot)$, $\mathcal{M}(\mu,\nu)-$q.s. Finally we get $Y \in \mathrm{dom}\theta(X,\cdot)\cap\bar{K}(X) = K_\theta(X)$, $\mathcal{M}(\mu,\nu)-$q.s.

<u>(ii)</u> As $R_{\theta,N_\nu}(X) \subset \mathrm{cl}\,\mathrm{conv}\partial\widehat{K}(X) = \mathrm{cl}\,\widehat{K}(X)$, $\bar{K} \subset \mathrm{cl}\,\widehat{K}$. By definition, $K_\theta \subset \bar{K}$, and $\widehat{K} \subset \underline{K}$. For $y \in \mathrm{atom}(\nu)$, and $\theta_0 \in \widehat{\mathcal{T}}(\mu,\nu)$, by minimality,

$$D_y(X) \subset \mathrm{dom}\theta_0(X,\cdot)\cap\partial\widehat{K}(X), \quad \mu-\text{a.s.} \tag{2.6.3}$$

Applying (2.6.3) for $\theta_0 = \theta$, we get $D_y \subset \mathrm{dom}\theta(X,\cdot)$, and for $\theta_0 = \theta^*$, $D_y(X) \subset \bar{K}(X)$, $\mu-$a.s. Taking the countable union: $\underline{K} \subset K_\theta$, $\mu-$a.s. (This is the only inclusion that is not pointwise). Then we change $\underline{K}$ to $\widehat{K}$ on this set to get this inclusion pointwise.

<u>(iii)</u> For $\theta_0 \in \widehat{\mathcal{T}}(\mu,\nu)$, let $N_\mu \in \mathcal{N}_\mu$ from Proposition 2.3.10. Let $x \in N_\mu^c$, $y \in \partial\widehat{K}(x)$, and $y' := \frac{x+y}{2} \in \widehat{K}(x)$. Then for any other $x' \in \widehat{K}(x)\cap N_\mu^c$, $\frac{1}{2}\theta_0(x,y) - \theta_0(x,y') = \frac{1}{2}\theta_0(x',x) + \frac{1}{2}\theta_0(x',y) - \theta_0(x',y')$, in particular, $y \in \mathrm{dom}\theta(x,\cdot)$ if and only if $y \in \mathrm{dom}\theta(x',\cdot)$. Applying this result to θ, θ^*, and θ_y^* for all $y \in \mathrm{atom}(\nu)$, we get N_μ such that for any $x \in \mathbb{R}^d$, $\bar{K}$, K_θ, and $\underline{K}$ are constant on $\widehat{K}(x)\cap N_\mu^c$. To get it pointwise, we redefine these mappings to this constant value on $\widehat{K}(x)\cap N_\mu$, or to $\widehat{K}(x)$, if $\widehat{K}(x)\cap N_\mu^c = \emptyset$. The previous properties are preserved. $\square$

2.6.2 Structure of polar sets

Proposition 2.6.3. *A Borel set $N \in \mathcal{B}(\Omega)$ is $\mathcal{M}(\mu,\nu)-$polar if and only if for some $(N_\mu,N_\nu) \in \mathcal{N}_\mu\times\mathcal{N}_\nu$ and $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, we have*

$$N \subset \{X \in N_\mu\}\cup\{Y \in N_\nu\}\cup\{Y \notin K_\theta(X)\}.$$

Proof. One implication is trivial as $Y \in K_\theta(X)$, $\mathcal{M}(\mu,\nu)-$q.s. for all $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, by Proposition 2.6.2. We only focus on the non-trivial implication. For an $\mathcal{M}(\mu,\nu)$-polar set N, we have $\mathbf{S}_{\mu,\nu}(\infty\mathbf{1}_N) = 0$, and it follows from the dual formulation of Theorem 2.3.19 that $0 = \mathrm{Val}(\xi)$ for some $\xi = (\varphi,\psi,h,\theta) \in \mathcal{D}_{\mu,\nu}^{mod}(\infty\mathbf{1}_N)$. Then,

$$\varphi < \infty,\ \mu-\text{a.s.},\quad \psi < \infty,\ \nu-\text{a.s.}\quad \text{and } \theta \in \widehat{\mathcal{T}}(\mu,\nu),$$

As h is finite valued, and φ, ψ are non-negative functions, the superhedging inequality $\varphi \oplus \psi + \theta + h^{\otimes} \geq \infty \mathbf{1}_N$ on $\{Y \in \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X)\}$ implies that

$$\mathbf{1}_{\{\varphi=\infty\}} \oplus \mathbf{1}_{\{\psi=\infty\}} + \mathbf{1}_{\{(\mathrm{dom}\theta)^c\}} \geq \mathbf{1}_N \quad \text{on} \quad \{Y \in \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X)\}. \tag{2.6.4}$$

By Proposition 2.3.15 (ii), we have $\widehat{K}(X) \subset K_{\theta,\{\psi=\infty\}}(X)$, $\mu-$a.s. Then $\bar{K}(X) \subset \mathrm{aff}\widehat{K}(X) \subset \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X)$, which implies that

$$K_\theta(X) := \mathrm{dom}\theta(X,\cdot) \cap \bar{K}(X) \subset \mathrm{dom}\theta(X,\cdot) \cap \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X), \ \mu-\text{a.s.} \tag{2.6.5}$$

We denote $N_\mu := \{\varphi = \infty\} \cup \{K_\theta(X) \not\subset \mathrm{dom}\theta(X,\cdot) \cap \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X)\} \in \mathcal{N}_\mu$, and $N_\nu := \{\psi = \infty\} \in \mathcal{N}_\nu$. Then by (2.6.4), $\mathbf{1}_N = 0$ on $(\{\varphi = \infty\}^c \times \{\psi = \infty\}^c) \cap \{Y \in \mathrm{dom}\theta(X,\cdot) \cap \mathrm{aff} K_{\theta,\{\psi=\infty\}}(X)\}$, and therefore by (2.6.5), $N \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin K_\theta(X)\}$.
□

2.6.3 The maximal support probability

In order to prove the existence of a maximum support martingale transport plan, we introduce the maximization problem.

$$M := \sup_{\mathbb{P} \in \mathcal{M}(\mu,\nu)} \mu[G(\overline{\mathrm{supp}}\mathbb{P}_X)]. \tag{2.6.6}$$

where we rely on the following measurability result whose proof is reported in Subsection 2.7.2.

Lemma 2.6.4. *For $\mathbb{P} \in \mathcal{P}(\Omega)$, the map $\overline{\mathrm{supp}}\mathbb{P}_X$ is analytically measurable, and the map $\overline{\mathrm{supp}}\left(\mathbb{P}_X|_{\partial \widehat{K}(X)}\right)$ is $\mu-$measurable.*

Now we prove a first Lemma about the existence of a maximal support probability.

Lemma 2.6.5. *There exists $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ such that for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ we have the inclusion $\overline{\mathrm{supp}}\,\mathbb{P}_X \subset \overline{\mathrm{supp}}\,\widehat{\mathbb{P}}_X$, $\mu-$a.s.*

Proof. We proceed in two steps:
Step 1: We first prove existence for the problem (2.6.6). Let $(\mathbb{P}^n)_{n\geq 1} \subset \mathcal{M}(\mu,\nu)$ be a maximizing sequence. Then the measure $\widehat{\mathbb{P}} := \sum_{n\geq 1} 2^{-n}\mathbb{P}^n \in \mathcal{M}(\mu,\nu)$, and satisfies $\overline{\mathrm{supp}}\mathbb{P}^n_X \subset \overline{\mathrm{supp}}\widehat{\mathbb{P}}_X$ for all $n \geq 1$. Consequently $\mu[G(\overline{\mathrm{supp}}_X\mathbb{P}^n_X)] \leq \mu[G(\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X)]$, and therefore $M = \mu[G(\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X)]$.
Step 2: We next prove that $\overline{\mathrm{supp}}\mathbb{P}_X \subset \overline{\mathrm{supp}}\widehat{\mathbb{P}}_X$, μ-a.s. for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. Indeed,

the measure $\overline{\mathbb{P}} := \frac{\widehat{\mathbb{P}}+\mathbb{P}}{2} \in \mathcal{M}(\mu,\nu)$ satisfies $M \geq \mu[G(\overline{\mathrm{supp}}\overline{\mathbb{P}}_X)] \geq \mu[G(\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X)] = M$, implying that $G(\overline{\mathrm{supp}}\overline{\mathbb{P}}_X) = G(\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X)$, $\mu-$a.s. The required result now follows from the inclusion $\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X \subset \overline{\mathrm{supp}}\overline{\mathbb{P}}_X$. □

Proof of Proposition 2.3.15 (iii) Let $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ from Lemma 2.6.5, if we denote $S(X) := \overline{\mathrm{supp}}\widehat{\mathbb{P}}_X$, then we have $\mathrm{supp}(\mathbb{P}_X) \subset S(X)$, $\mu-$a.s. Then $\{Y \notin S(X)\}$ is $\mathcal{M}(\mu,\nu)-$polar. By Lemma 2.6.1, $\{Y \notin S(X)\} \cup \{X \notin N'_\mu\}$ is Borel for some $N'_\mu \in \mathcal{N}_\mu$. By Theorem 2.3.18, we see that $\{Y \notin S(X)\} \subset \{Y \notin S(X)\} \cup \{X \notin N'_\mu\} \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin K_\theta(X)\}$, and therefore

$$\{Y \in S(X)\} \quad \supset \quad \{X \notin N_\mu\} \cap \{Y \in K_\theta(X) \setminus N_\nu\},$$

for some $N_\mu \in \mathcal{N}_\mu$, $N_\nu \in \mathcal{N}_\nu$, and $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$. The last inclusion implies that $K_\theta(X) \setminus N_\nu \subset S(X)$, μ-a.s. However, by Proposition 2.3.15 (ii), $\widehat{K}(X) \subset \mathrm{conv}\Big(\mathrm{dom}\theta(X,\cdot) \setminus N_\nu\Big)$, $\mu-$a.s. Then, since $S(X)$ is closed and convex, we see that $\mathrm{cl}\,\widehat{K}(X) \subset S(X)$.

To obtain the reverse inclusion, we recall from Proposition 2.3.15 (i) that $\{Y \in \mathrm{cl}\,\widehat{K}(X)\}$, $\mathcal{M}(\mu,\nu)-$q.s. In particular $\widehat{\mathbb{P}}[Y \in \mathrm{cl}\,\widehat{K}(X)] = 1$, implying that $S(X) \subset \mathrm{cl}\,\widehat{K}(X)$, μ-a.s. as $\mathrm{cl}\,\widehat{K}(X)$ is closed convex. Finally, recall that by definition $I := \mathrm{ri}S$ and therefore $\widehat{K}(X) = \mathrm{cl}\,I(X)$, $\mu-$a.s. □

Lemma 2.6.6. *We may choose $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$ in Theorem 2.2.1 so that for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ and $y \in \mathbb{R}^d$,*

$$\mu\Big[\mathbb{P}_X[\{y\}] > 0\Big] \leq \mu\Big[\widehat{\mathbb{P}}_X[\{y\}] > 0\Big],$$
$$\textit{and} \;\; \mathrm{supp}\,\mathbb{P}_X|_{\partial I(X)} \subset \overline{\mathrm{supp}}\,\widehat{\mathbb{P}}_X|_{\partial I(X)}, \quad \mu - a.s.$$

In this case the maps $\underline{J}(X) := I(X) \cup \Big\{y \in \mathbb{R}^d : \nu[y] > 0$ and $\widehat{\mathbb{P}}_X\Big[\{y\}\Big] > 0\Big\}$, and $\bar{J}(X) := I(X) \cup \overline{\mathrm{supp}}\,\widehat{\mathbb{P}}_X|_{\partial I(X)}$ are unique $\mu-$a.s. Furthermore $\underline{J}(X) = \underline{K}(X)$, $\bar{J}(X) = \bar{K}(X)$, and $J_\theta(X) = K_\theta(X)$, $\mu-$a.s. for all $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$.

Proof. Step 1: By the same argument as in the proof of Lemma 2.6.5, we may find $\widehat{\mathbb{P}}' \in \mathcal{M}(\mu,\nu)$ such that

$$\begin{aligned} M' &:= \sup_{\mathbb{P} \in \mathcal{M}(\mu,\nu)} \mu\Big[G\Big(\overline{\mathrm{supp}}\Big(\mathbb{P}_X|_{\partial \widehat{K}(X)}\Big)\Big)\Big] \qquad (2.6.7) \\ &= \mu\Big[G\Big(\overline{\mathrm{supp}}\Big(\widehat{\mathbb{P}}'_X|_{\partial \widehat{K}(X)}\Big)\Big)\Big]. \end{aligned}$$

We also have similarly that $\overline{\mathrm{supp}}\Big(\mathbb{P}_X|_{\partial\widehat{K}(X)}\Big) \subset \overline{\mathrm{supp}}\Big(\widehat{\mathbb{P}}'_X|_{\partial\widehat{K}(X)}\Big)$, μ-a.s. for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. Then we prove similarly that $S'(X) := \overline{\mathrm{supp}}\Big(\widehat{\mathbb{P}}'_X|_{\partial\widehat{K}(X)}\Big) = D(X)$, $\mu-$a.s., where recall that D is the optimizer for (2.6.1). Indeed, by the previous step, we have $\overline{\mathrm{supp}}(\mathbb{P}_X|_{\partial\widehat{K}(X)}) \subset S'(X)$, $\mu-$a.s. Then we have $\{Y \notin S'(X) \cup \widehat{K}(X)\}$ is $\mathcal{M}(\mu,\nu)-$polar. By Theorem 2.3.18, we see that $\{Y \notin S'(X) \cup \widehat{K}(X)\} \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin K_\theta(X) \cup \widehat{K}(X)\}$, or equivalently

$$\{Y \in S'(X) \cup \widehat{K}(X)\} \quad \supset \quad \{X \notin N_\mu\} \cap \{Y \in K_\theta(X) \setminus N_\nu\}, \qquad (2.6.8)$$

for some $N_\mu \in \mathcal{N}_\mu$, $N_\nu \in \mathcal{N}_\nu$, and $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$. Similar to the previous analysis, we have $K_\theta(X) \setminus N_\nu \setminus \widehat{K}(X) \subset S'(X)$, μ-a.s. Then, since $S'(X)$ is closed and convex, we see that $D(X) \subset S'(X)$.

To obtain the reverse inclusion, we recall from Proposition 2.6.2 that $\{Y \in \bar{K}(X)\}$, $\mathcal{M}(\mu,\nu)-$q.s. In particular $\widehat{\mathbb{P}}'[Y \in \widehat{K}(X) \cup D(X)] = 1$, implying that $S'(X) \subset D(X)$, μ-a.s. By Proposition 2.3.15 (iii), we have $\bar{J}(X) = (I \cup S')(X) = (\widehat{K} \cup D)(X) = \bar{K}(X)$, $\mu-$a.s.

Finally, $\frac{\widehat{\mathbb{P}}+\widehat{\mathbb{P}}'}{2}$ is optimal for both problems (2.6.6), and (2.6.7). By definition, the equality $J_\theta(X) = K_\theta(X)$, $\mu-$a.s. for $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$ immediately follows.

<u>*Step 2:*</u> Let $y \in \mathrm{atom}(\nu)$, if y is an atom of $\gamma_1 \in \mathcal{P}(\mathbb{R}^d)$ and $\gamma_2 \in \mathcal{P}(\mathbb{R}^d)$, then y in an atom of $\lambda\gamma_1 + (1-\lambda)\gamma_2$ for all $0 < \lambda < 1$. By the same argument as in Step 1, we may find $\widehat{\mathbb{P}}^y \in \mathcal{M}(\mu,\nu)$ such that

$$\begin{aligned} M_y &:= \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mu\Big[\mathbb{P}_X\big[\{y\} \cap \mathrm{cl}\,\widehat{K}(X)\big] > 0\Big] \qquad (2.6.9)\\ &= \mu\Big[\widehat{\mathbb{P}}^y_X\big[\{y\} \cap \mathrm{cl}\,\widehat{K}(X)\big] > 0\Big]. \end{aligned}$$

We denote $S_y(X) := \mathrm{supp}\widehat{\mathbb{P}}^y_X|_{\mathrm{aff}\,\widehat{K}(X)\cap\{y\}}$. Recall that D_y is the notation for the optimizer of problem (2.6.2). We consider the set $N := \Big\{Y \notin (\mathrm{cl}\,\widehat{K}(X) \setminus \{y\}) \cup S_y(X)\Big\}$. N is polar as $Y \in \mathrm{cl}\,\widehat{K}(X)$, q.s., and by definition of S_y. Then $N \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin K_\theta(X)\}$, or equivalently,

$$\Big\{Y \notin (\mathrm{cl}\,\widehat{K}(X) \setminus \{y\}) \cup S_y(X)\Big\} \supset \{X \notin N_\mu\} \cap \{Y \in K_\theta(X) \setminus N_\nu\}, \quad (2.6.10)$$

for some $N_\mu \in \mathcal{N}_\mu$, $N_\nu \in \mathcal{N}_\nu$, and $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$. Then $D_y(X) \subset K_\theta(X) \setminus N_\nu \subset \mathrm{cl}\,\widehat{K}(X) \setminus \{y\} \cup S_y(X)$, $\mu-$a.s. Finally $D_y(X) \subset S_y(X)$, $\mu-$a.s.

On the other hand, $S_y \subset D_y$, $\mu-$a.s., as if $\widehat{\mathbb{P}}^y_X[\{y\}] > 0$, we have $\theta(X,y) < \infty$, $\mu-$a.s. at the corresponding points. Hence, $D_y(X) = S_y(X)$, $\mu-$a.s. Now if we sum up the countable optimizers for $y \in \mathrm{atom}(\nu)$, with the previous optimizers, then the probability $\widehat{\mathbb{P}}$ we get is an optimizer for (2.6.6), (2.6.7), and (2.6.9), for all $y \in \mathbb{R}^d$ (the optimum is 0 if it is not an atom of ν). Furthermore, the $\mu-$a.e. equality of the maps S_y and D_y for these countable $y \in \mathrm{atom}(\nu)$ is preserved by this countable union, then together with Proposition 2.3.15 (iii), we get $\underline{J} = \underline{K}$, $\mu-$a.s. □

As a preparation to prove the main Theorem 2.2.1, we need the following lemma, which will be proved in Subsection 2.7.2.

Lemma 2.6.7. *Let $F : \mathbb{R}^d \longrightarrow \mathrm{ri}\,\overline{\mathcal{K}}$ be a $\gamma-$measurable function for some $\gamma \in \mathcal{P}(\mathbb{R}^d)$, such that $x \in F(x)$ for all $x \in \mathbb{R}^d$, and $\{F(x) : x \in \mathbb{R}^d\}$ is a partition of $\mathbb{R}^d$. Then up to a modification on a $\gamma-$null set, F can be chosen in addition to be analytically measurable.*

Proof of Theorem 2.2.1 Existence holds by Lemma 2.6.5 above, (i) is a consequence of Lemma 2.6.4, and (ii) directly stems from Lemma 2.3.13 (iii) together with Proposition 2.3.15 (iii). Now we need to deal with the measurability issue. Lemma 2.6.7 allows to modify $\mathrm{ri}\,\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X$ to get (ii) while preserving its analytic measurability, we denote I its modification. However, we need to modify $\widehat{\mathbb{P}}_X$ to get the result. As $\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X$ is analytically measurable by Lemma 2.6.4, the set of modification $N_\mu := \{\overline{\mathrm{supp}}\widehat{\mathbb{P}}_X \neq \mathrm{cl}\,I(X)\} \in \mathcal{N}_\mu$ is analytically measurable. Then we may redefine $\widehat{\mathbb{P}}_X$ on N_μ, so as to preserve a kernel for $\widehat{\mathbb{P}}$. By the same arguments than the proof of Lemma 2.3.13 (ii), the measure-valued map $\kappa_X := g_{I(X)}$ is a kernel thanks to the analytic measurability of I, recall the definition of g_K given by (2.3.2). Furthermore, $\overline{\mathrm{supp}}\,\kappa_X = I(X)$ pointwise by definition. Then a suitable kernel modification from which the result follows is given by

$$\widehat{\mathbb{P}}'_X := \mathbf{1}_{\{X \in N_\mu\}}\kappa_X + \mathbf{1}_{\{X \notin N_\mu\}}\widehat{\mathbb{P}}_X.$$

□

Proof of Proposition 2.2.4 The existence and the uniqueness are given by Lemma 2.6.6 and the other properties follow from the identity between the J maps and the K maps, also given by the Lemma, together with Proposition 2.6.2. □

Proof of Theorem 2.3.18 We simply apply Lemma 2.6.6 to replace K_θ by J_θ in Proposition 2.6.3. □

2.7 Measurability of the irreducible components

2.7.1 Measurability of G

Proof of Lemma 2.3.13 (ii) As $\mathbb{R}^d$ is locally compact, the Wijsman topology is locally equivalent to the Hausdorff topology[6], i.e. as $n \to \infty$, $K_n \longrightarrow K$ for the Wijsman topology if and only if $K_n \cap B_M \longrightarrow K \cap B_M$ for the Hausdorff topology, for all $M \geq 0$.

We first prove that $K \longmapsto \dim \operatorname{aff} K$ is a lower semi-continuous map $\mathcal{K} \to \mathbb{R}$. Let $(K_n)_{n\geq 1} \subset \mathcal{K}$ with dimension $d_n \leq d' \leq d$ converging to K. We consider $A_n := \operatorname{aff} K_n$. As A_n is a sequence of affine spaces, it is homeomorphic to a $d+1$-uplet. Observe that the convergence of K_n allow us to chose this $d+1$-uplet to be bounded. Then up to taking a subsequence, we may suppose that A_n converges to an affine subspace A of dimension less than d'. By continuity of the inclusion under the Wijsman topology, $K \subset A$ and $\dim K \leq \dim A \leq d'$.

We next prove that the mapping $K \mapsto g_K(K)$ is continuous on $\{\dim K = d'\}$ for $0 \leq d' \leq d$, which implies the required measurability. Let $(K_n)_{n\geq 1} \subset \mathcal{K}$ be a sequence with constant dimension d', converging to a $d'-$dimensional subset, K in $\mathcal{K}$. Define $A_n := \operatorname{aff} K_n$ and $A := \operatorname{aff} K$, A_n converges to A as for any accumulation set A' of A_n, $K \subset A'$ and $\dim A' = \dim A$, implying that $A' = A$. Now we consider the map $\phi_n : A_n \to A$, $x \mapsto proj_A(x)$. For all $M > 0$, it follows from the compactness of the closed ball B_M that ϕ_n converges uniformly to identity as $n \to \infty$ on B_M. Then, $\phi_n(K_n) \cap B_M \longrightarrow K \cap B_M$ as $n \to \infty$, and therefore $\lambda_A[\phi_n(K_n \cap B_M) \setminus K] + \lambda_A[K \setminus \phi_n(K_n) \cap B_M] \longrightarrow 0$. As the Gaussian density is bounded, we also have

$$g_A[\phi_n(K_n \cap B_M)] \longrightarrow g_A[K \cap B_M].$$

We next compare $g_A[\phi_n(K_n \cap B_M)]$ to $g_{K_n}(K_n \cap B_M)$. As (ϕ_n) is a sequence of linear functions that converges uniformly to identity, we may assume that ϕ_n is a C^1-diffeomorphism. Furthermore, its constant Jacobian J_n converges to 1 as $n \to \infty$. Then,

$$\begin{aligned}
\int_{K_n \cap B_M} \frac{e^{-|\phi_n(x)|^2/2}}{(2\pi)^{d'/2}} \lambda_{K_n}(dx) &= \int_{\phi_n(K_n \cap B_M)} \frac{e^{-|y|^2/2} J_n^{-1}}{(2\pi)^{d'/2}} \lambda_A(dy) \\
&= J_n^{-1} g_A[\phi_n(K_n \cap B_M)].
\end{aligned}$$

[6]The Haussdorff distance on the collection of all compact subsets of a compact metric space $(\mathcal{X},\mathrm{d})$ is defined by $d_H(K_1,K_2) = \sup_{x\in\mathcal{X}} |\mathrm{dist}(x,K_1) - \mathrm{dist}(x,K_2)|$, for $K_1, K_2 \subset \mathcal{X}$, compact subsets.

As the Gaussian distribution function is 1-Lipschitz, we have

$$\left|\int_{K_n\cap B_M}\frac{e^{-|\phi_n(x)|^2/2}}{(2\pi)^{d'/2}}\lambda_{K_n}(dx)-g_{K_n}(K_n\cap B_M)\right| \leq \lambda_{K_n}[K_n\cap B_M]|\phi_n-Id_A|_\infty,$$

where $|\cdot|_\infty$ is taken on $K_n\cap B_M$. Now for arbitrary $\epsilon>0$, by choosing M sufficiently large so that $g_V[V\setminus B_M]\leq\epsilon$ for any $d'-$dimensional subspace V, we have

$$\begin{aligned}
|g_{K_n}[K_n]-g_K[K]| &\leq |g_{K_n}[K_n\cap B_M]-g_A[K\cap B_M]|+2\epsilon\\
&\leq \left|g_{K_n}[K_n\cap B_M]-\int_{K_n\cap B_M}C\exp\left(\frac{-|\phi_n(x)|^2}{2}\right)\lambda_{K_n}(dx)\right|\\
&\qquad +\left|J_n^{-1}g_A[\phi_n(K_n\cap B_M)]-g_A[K\cap B_M]\right|+2\epsilon\\
&\leq 4\epsilon,
\end{aligned}$$

for n sufficiently large, by the previously proved convergence. Hence $G_{d'}:=G\Big|_{\mathrm{dim}^{-1}\{d'\}}$ is continuous, implying that $G:K\longmapsto\sum_{d'=0}^{d}\mathbf{1}_{\mathrm{dim}^{-1}\{d'\}}(K)G_{d'}(K)$ is Borel-measurable. □

2.7.2 Further measurability of set-valued maps

This subsection is dedicated to the proof of Lemmas 2.3.13 (i), 2.6.1, and 2.6.4. In preparation for the proofs, we start by giving some lemmas on measurability of set-valued maps. Let $\mathcal{A}$ be a $\sigma-$algebra of $\mathbb{R}^d$. In practice we will always consider either the $\sigma-$algebra of Borel sets, the $\sigma-$algebra of analytically measurable sets, or the $\sigma-$algebra of universally measurable sets.

Lemma 2.7.1. *Let $(F_n)_{n\geq1}\subset\mathbb{L}^{\mathcal{A}}(\mathbb{R}^d,\mathcal{K})$. Then $\mathrm{cl}\cup_{n\geq1}F_n$ and $\cap_{n\geq1}F_n$ are $\mathcal{A}-$measurable.*

Proof. The measurability of the union is a consequence of Propositions 2.3 and 2.6 in Himmelberg [88]. The measurability of the intersection follows from the fact that $\mathbb{R}^d$ is σ-compact, together with Corollary 4.2 in [88]. □

Lemma 2.7.2. *Let $F\in\mathbb{L}^{\mathcal{A}}(\mathbb{R}^d,\mathcal{K})$. Then, $\mathrm{cl\,conv}F$, $\mathrm{aff}F$, and $\mathrm{cl\,rf}_X\mathrm{cl\,conv}F$ are $\mathcal{A}-$measurable.*

Proof. The measurability of $\mathrm{cl\,conv}F$ is a direct application of Theorem 9.1 in [88].

We next verify that $\mathrm{aff}F$ is measurable. Since the values of F are closed, we deduce from Theorem 4.1 in Wagner [158], that we may find a measurable $x\longmapsto y(x)$, such that $y(x)\in F(x)$ if $F(x)\neq\emptyset$, for all $x\in\mathbb{R}^d$. Then we may write $\mathrm{aff}F(x)=$

$\mathrm{cl\,conv\,cl}\ \cup_{q\in\mathbb{Q}}\Big(y(x)+q\,(F(x)-y(x))\Big)$ for all $x\in\mathbb{R}^d$. The measurability follows from Lemmas 2.7.1, together with the first step of the present proof.

We finally justify that $\mathrm{cl\,rf}_X\mathrm{cl\,conv}F$ is measurable. We may assume that F takes convex values. By convexity, we may reduce the definition of rf_x to a sequential form:

$$\begin{aligned}\mathrm{cl\,rf}_xF(x) &= \mathrm{cl}\cup_{n\geq 1}\Big\{y\in\mathbb{R}^d, y+\frac{1}{n}(y-x)\in F(x) \text{ and } x-\frac{1}{n}(y-x)\in F(x)\Big\}\\ &= \mathrm{cl}\cup_{n\geq 1}\Big[\Big\{y\in\mathbb{R}^d, y+\frac{1}{n}(y-x)\in F(x)\Big\}\cap\Big\{y\in\mathbb{R}^d, x-\frac{1}{n}(y-x)\in F(x)\Big\}\Big]\\ &= \mathrm{cl}\cup_{n\geq 1}\Big[\Big(\frac{1}{n+1}x+\frac{n}{n+1}F(x)\Big)\cap(-(n+1)x-nF(x))\Big],\end{aligned}$$

so that the required measurability follows from Lemma 2.7.1. □

We denote by $\mathcal{S}$ the set of finite sequences of positive integers, and Σ the set of infinite sequences of positive integers. Let $s\in\mathcal{S}$, and $\sigma\in\Sigma$. We shall denote $s<\sigma$ whenever s is a prefix of σ.

Lemma 2.7.3. *Let $(F_s)_{s\in\mathcal{S}}$ be a family of universally measurable functions $\mathbb{R}^d\longrightarrow\mathcal{K}$ with convex image. Then the mapping $\mathrm{cl\,conv}\Big(\cup_{\sigma\in\Sigma}\cap_{s<\sigma}F_s\Big)$ is universally measurable.*

Proof. Let $\mathcal{U}$ the collection of universally measurable maps from $\mathbb{R}^d$ to $\mathcal{K}$ with convex image. For an arbitrary $\gamma\in\mathcal{P}(\mathbb{R}^d)$, and $F:\mathbb{R}^d\longrightarrow\mathcal{K}$, we introduce the map

$$\gamma G^*[F] := \inf_{F\subset F'\in\mathcal{U}}\gamma G[F'], \quad \text{where} \quad \gamma G[F'] := \gamma\Big[G\Big(F'(X)\Big)\Big] \text{ for all } F'\in\mathcal{U}.$$

Clearly, γG and γG^* are non-decreasing, and it follows from the dominated convergence theorem that γG, and thus γG^*, are upward continuous.

<u>*Step 1:*</u> In this step we follow closely the line of argument in the proof of Proposition 7.42 of Bertsekas and Shreve [30]. Set $F:=\mathrm{cl\,conv}\Big(\cup_{\sigma\in\Sigma}\cap_{s<\sigma}F_s\Big)$, and let $(\bar F_n)_n$ a minimizing sequence for $\gamma G^*[F]$. Notice that $F\subset\bar F:=\cap_{n\geq 1}\bar F_n\in\mathcal{U}$, by Lemma 2.7.1. Then $\bar F$ is a minimizer of $\gamma G^*[F]$.

For $s,s'\in S$, we denote $s\leq s'$ if they have the same length $|s|=|s'|$, and $s_i\leq s'_i$ for $1\leq i\leq|s|$. For $s\in S$, let

$$R(s) := \mathrm{cl\,conv}\cup_{s'\leq s}\cup_{\sigma>s'}\cap_{s''<\sigma}F_{s''} \quad\text{and}\quad K(s) := \mathrm{cl\,conv}\cup_{s'\leq s}\cap_{j=1}^{|s'|}F_{s'_1,\ldots,s'_j}.$$

Notice that $K(s)$ is universally measurable, by Lemmas 2.7.1 and 2.7.2, and

$$R(s)\subset K(s),\ \mathrm{cl}\cup_{s_1\geq 1}R(s_1)=F,\ \text{and}\ \mathrm{cl}\cup_{s_k\geq 1}R(s_1,...,s_k)=R(s_1,...,s_{k-1}).$$

By the upwards continuity of γG^*, we may find for all $\epsilon>0$ a sequence $\sigma^\epsilon\in\Sigma$ s.t.

$$\gamma G^*[F]\leq\gamma G^*[R(\sigma_1^\epsilon)]+2^{-1}\epsilon,\quad\text{and}\quad\gamma G^*[R(\underline{\sigma}_{k-1})]\leq\gamma G^*[R(\underline{\sigma}_k)]+2^{-k}\epsilon,$$

for all $k\geq 1$, with the notation $\underline{\sigma}_k^\varepsilon:=(\sigma_1^\epsilon,\dots,\sigma_k^\varepsilon)$. Recall that the minimizer $\overline{F}$ and $K(s)$ are in $\mathcal{U}$ for all $s\in\mathcal{S}$. We then define the sequence $K_k^\epsilon:=\overline{F}\cap K(\underline{\sigma}_k^\epsilon)\in\mathcal{U}$, $k\geq 1$, and we observe that

$$\begin{gathered}(K_k^\epsilon)_{k\geq 1}\ \text{ decreasing},\ \underline{F}^\epsilon:=\cap_{k\geq 1}K_k^\epsilon\subset F,\\ \text{and }\gamma G[K_k^\epsilon]\geq\gamma G^*[F]-\epsilon=\gamma G[\overline{F}]-\epsilon,\end{gathered}\tag{2.7.1}$$

by the fact that $R(\underline{\sigma}_k^\epsilon)\subset K_k^\epsilon$. We shall prove in Step 2 that, for an arbitrary $\alpha>0$, we may find $\varepsilon=\varepsilon(\alpha)\leq\alpha$ such that (2.7.1) implies that

$$\gamma G[\underline{F}^\epsilon]\geq\inf_{k\geq 1}\gamma G[K_k^\epsilon]-\alpha\geq\gamma G[\overline{F}]-\epsilon-\alpha.\tag{2.7.2}$$

Now let $\alpha=\alpha_n:=n^{-1}$, $\varepsilon_n:=\epsilon(\alpha_n)$, and notice that $\underline{F}:=\mathrm{cl\,conv}\cup_{n\geq 1}\underline{F}^{\epsilon_n}\in\mathcal{U}$, with $\underline{F}^{\epsilon_n}\subset\underline{F}\subset F\subset\overline{F}$, for all $n\geq 1$. Then, it follows from (2.7.2) that $\gamma G[\underline{F}]=\gamma G[\overline{F}]$, and therefore $\underline{F}=F=\overline{F}$, $\gamma-$a.s. In particular, F is $\gamma-$measurable, and we conclude that $F\in\mathcal{U}$ by the arbirariness of $\gamma\in\mathcal{P}(\mathbb{R}^d)$.

<u>Step 2:</u> It remains to prove that, for an arbitrary $\alpha>0$, we may find $\varepsilon=\varepsilon(\alpha)\leq\alpha$ such that (2.7.1) implies (2.7.2). This is the point where we have to deviate from the argument of [30] because γG is not downwards continuous, as the dimension can jump down.

Set $A_n:=\{G\big(\overline{F}(X)\big)-\dim\overline{F}(X)\leq 1/n\}$, and notice that $\cap_{n\geq 1}A_n=\emptyset$. Let $n_0\geq 1$ such that $\gamma[A_{n_0}]\leq\frac{1}{2}\frac{\alpha}{d+1}$, and set $\epsilon:=\frac{1}{2}\frac{1}{n_0}\frac{\alpha}{d+1}>0$. Then, it follows from (2.7.1) that

$$\begin{aligned}\gamma\Big[\inf_n G(K_n^\epsilon)-\dim\overline{F}\leq 0\Big] &\leq \gamma\Big[\inf_n G(K_n^\epsilon)-G(\overline{F})\leq n_0^{-1}\Big]\\ &\quad+\gamma\Big[G(\overline{F})-\dim\overline{F}\leq -n_0^{-1}\Big]\\ &\leq n_0\Big(\gamma\Big[G(\overline{F})\Big]-\gamma\Big[\inf_n G(K_n^\epsilon)\Big]\Big)+\gamma\Big[A_{n_0}\Big]\\ &= n_0\Big(\gamma\Big[G(\overline{F})\Big]-\inf_n\gamma\Big[G(K_n^\epsilon)\Big]\Big)+\gamma\Big[A_{n_0}\Big]\\ &\leq n_0\epsilon+\frac{1}{2}\frac{\alpha}{d+1}\ =\ \frac{\alpha}{d+1},\end{aligned}\tag{2.7.3}$$

where we used the Markov inequality and the monotone convergence theorem. Then:

$$\begin{aligned}
\gamma\Big[\inf_n G(K_n^\epsilon) - G\Big(\underline{F}^\epsilon\Big)\Big] &\le \gamma\Big[\mathbf{1}_{\{\inf_n G(K_n^\epsilon)-\dim\overline{F}\le 0\}}\Big(\inf_n G(K_n^\epsilon) - G\Big(\underline{F}^\epsilon\Big)\Big)\\
&\qquad +\mathbf{1}_{\{\inf_n G(K_n^\epsilon)-\dim\overline{F}> 0\}}\Big(\inf_n G(K_n^\epsilon) - G\Big(\underline{F}^\epsilon\Big)\Big)\Big]\\
&\le \gamma\Big[(d+1)\mathbf{1}_{\{\inf_n G(K_n^\epsilon)-\dim\overline{F}\le 0\}}\\
&\qquad +\mathbf{1}_{\{\inf_n G(K_n^\epsilon)-\dim\overline{F}> 0\}}\Big(\inf_n G(K_n^\epsilon) - G\Big(\underline{F}^\epsilon\Big)\Big)\Big].
\end{aligned}$$

We finally note that $\inf_n G(K_n^\epsilon) - G\Big(\underline{F}^\epsilon\Big) = 0$ on $\{\inf_n G(K_n^\epsilon) - \dim\overline{F} > 0\}$. Then (2.7.2) follows by substituting the estimate in (2.7.3). $\square$

Proof of Lemma 2.3.13 (i) We consider the mappings $\theta:\Omega\to\bar{\mathbb{R}}_+$ such that $\theta = \sum_{k=1}^n \lambda_k 1_{C_k^1\times C_k^2}$ where $n\in\mathbb{N}$, the λ_k are non-negative numbers, and the C_k^1, C_k^2 are closed convex subsets of $\mathbb{R}^d$. We denote the collection of all these mappings $\mathcal{F}$. Notice that $\mathrm{cl}\,\mathcal{F}$ for the pointwise limit topology contains all $\mathbb{L}_+^0(\Omega)$. Then for any $\theta\in\mathbb{L}_+^0(\Omega)$, we may find a family $(\theta_s)_{s\in\Sigma}\subset\mathcal{F}$, such that $\theta = \inf_{\sigma\in\Sigma}\sup_{s<\sigma}\theta_s$. For $\theta\in\mathbb{L}_+^0(\Omega)$, and $n\ge 0$, we denote $F_\theta : x\longmapsto \mathrm{cl\,conv\,dom}\theta(x,\cdot)$, and $F_{\theta,n}: x\longmapsto \mathrm{cl\,conv}\,\theta(x,\cdot)^{-1}([0,n])$. Notice that $F_\theta = \mathrm{cl}\cup_{n\ge1}F_{\theta,n}$. Notice as well that $F_{\theta,n}$ is Borel measurable for $\theta\in\mathcal{F}$, and $n\ge 0$, as it takes values in a finite set, from a finite number of measurable sets. Let $\theta\in\mathbb{L}_+^0(\Omega)$, we consider the associated family $(\theta_s)_{s\in\Sigma}\subset\mathcal{F}$, such that $\theta = \inf_{\sigma\in\Sigma}\sup_{s<\sigma}\theta_s$. Notice that $F_{\theta,n} = \mathrm{cl\,conv}\Big(\cup_{\sigma\in\Sigma}\cap_{s<\sigma}F_{\theta_s,n}\Big)$ is universally measurable by Lemma 2.7.3, thus implying the universal measurability of $F_\theta = \mathrm{cl\,dom}\theta(X,\cdot)$ by Lemma 2.7.1.

In order to justify the measurability of $\mathrm{dom}_X\theta$, we now define

$$F_\theta^0 := F_\theta \quad\text{and}\quad F_\theta^k := \mathrm{cl\,conv}(\mathrm{dom}\theta(X,\cdot)\cap\mathrm{aff\,rf}_X F_\theta^{k-1}),\ k\ge 1.$$

Note that $F_\theta^k = \mathrm{cl}\cup_{n\ge1}\Big(\mathrm{cl\,conv}\cup_{\sigma\in\Sigma}\cap_{s<\sigma}F_{\theta_s,n}\cap\mathrm{aff\,rf}_x F_\theta^{k-1}\Big)$. Then, as F_θ^0 is universally measurable, we deduce that $\Big(F_\theta^k\Big)_{k\ge1}$ are universally measurable, by Lemmas 2.7.2 and 2.7.3.

As $\mathrm{dom}_X\theta$ is convex and relatively open, the required measurability follows from the claim:

$$F_\theta^d \;=\; \mathrm{cl\,dom}_X\theta.$$

To prove this identity, we start by observing that $F_\theta^k(x) \supset \mathrm{cl}\,\mathrm{dom}_x\theta$. Since the dimension cannot decrease more than d times, we have $\mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x) = \mathrm{aff}\, F_\theta^d(x)$ and

$$\begin{aligned} F_\theta^{d+1}(x) &= \mathrm{cl}\,\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big) \\ &= \mathrm{cl}\,\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\mathrm{rf}_x F_\theta^{d-1}(x)\Big) \;=\; F_\theta^d(x). \end{aligned}$$

i.e. $(F_\theta^{d+1})_k$ is constant for $k \geq d$. Consequently,

$$\begin{aligned} \dim \mathrm{rf}_x \mathrm{conv}(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)) &= \dim F_\theta^d(x) \\ &\geq \dim \mathrm{conv}(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)). \end{aligned}$$

As $\dim \mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big) \geq \dim \mathrm{rf}_x \mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big)$, we have equality of the dimension of $\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big)$ with its rf_x. Then it follows from Proposition 2.3.1 (ii) that $x \in \mathrm{ri}\,\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big)$, and therefore:

$$\begin{aligned} F_\theta^d(x) &= \mathrm{cl}\,\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big) \\ &= \mathrm{cl}\,\mathrm{ri}\,\mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big) \\ &= \mathrm{cl}\,\mathrm{rf}_x \mathrm{conv}\Big(\mathrm{dom}\theta(x,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_x F_\theta^d(x)\Big) \;\subset\; \mathrm{cl}\,\mathrm{dom}_x\theta. \end{aligned}$$

Hence $F_\theta^d(x) = \mathrm{cl}\,\mathrm{dom}_x\theta$.

Finally, $K_{\theta,A} = \mathrm{dom}_X(\theta + \infty 1_{\mathbb{R}^d \times A})$ is universally measurable by the universal measurability of dom_X. $\square$

Proof of Lemma 2.6.1 We may find $(F_n)_{n\geq 1}$, Borel-measurable with finite image, converging $\gamma-$a.s. to F. We denote $N_\gamma \in \mathcal{N}_\gamma$, the set on which this convergence does not hold. For $\epsilon > 0$, we denote $F_k^\epsilon(X) := \{y \in \mathbb{R}^d : dist\Big(y, F_k(X)\Big) \leq \epsilon\}$, so that

$$F(x) = \cap_{i\geq 1} \liminf_{n\to\infty} F_n^{1/i}(x), \quad \text{for all} \quad x \notin N_\gamma.$$

Then, as $\mathbf{1}_{Y\in F(X)}\mathbf{1}_{X\notin N_\gamma} = \inf_{i\geq 1} \liminf_{n\to\infty} \mathbf{1}_{Y\in F_n^{1/i}(X)}\mathbf{1}_{X\notin N_\gamma}$, the Borel-measurability of this function follows from the Borel-measurability of each $\mathbf{1}_{Y\in F_n^{1/i}(X)}$.

Now we suppose that $X \in \mathrm{ri}F(X)$ convex, $\gamma-$a.s. Up to redefining N_γ, we may suppose that this property holds on N_γ^c, then $\partial F(x) = \cap_{n\geq 1} F(x) \setminus \Big(x + \frac{n}{n+1}(F(x) - x)\Big)$, for $x \notin N_\gamma$. We denote $a := \mathbf{1}_{Y\in F(X)}\mathbf{1}_{X\notin N_\gamma}$. The result follows from the identity

$\mathbf{1}_{Y\in\partial F(X)}\mathbf{1}_{X\notin N_\gamma} = a - \sup_{n\geq 1} a\Big(X, X+\frac{n}{n+1}(Y-X)\Big)$. □

Proof of Lemma 2.6.4 Let $\mathcal{K}_\mathbb{Q} := \{K = \mathrm{conv}(x_1,\ldots,x_n) : n\in\mathbb{N}, (x_i)_{i\leq n}\subset\mathbb{Q}^d\}$. Then

$$\overline{\mathrm{supp}}\mathbb{P}_x = \mathrm{cl}\cup_{N\geq 1}\cap\{K\in\mathcal{K}_\mathbb{Q} : \overline{\mathrm{supp}}\mathbb{P}_x\cap B_N\subset K\} = \mathrm{cl}\cup_{N\geq 1}\cap_{K\in\mathcal{K}_\mathbb{Q}} F_K^N(x),$$

where $F_K^N(x) := K$ if $\mathbb{P}_x[B_N\cap K] = \mathbb{P}_x[B_N]$, and $F_K^N(x) := \mathbb{R}^d$ otherwise. As for any $K\in\mathcal{K}_\mathbb{Q}$ and $N\geq 1$, the map $\mathbb{P}_X[B_N\cap K]-\mathbb{P}_X[B_N]$ is analytically measurable, then F_K^N is analytically measurable. The required measurability result follows from lemma 2.7.1.

Now, in order to get the measurability of $\overline{\mathrm{supp}}(\mathbb{P}_X|_{\partial I(X)})$, we have in the same way

$$\overline{\mathrm{supp}}(\mathbb{P}_X|_{\partial I(X)}) \quad = \quad \mathrm{cl}\cup_{n\geq 1}\cap_{K\in\mathcal{K}_\mathbb{Q}} F_K'^N(x),$$

where $F_K'^N(x) := K$ if $\mathbb{P}_x[\partial I(x)\cap B_N\cap K] = \mathbb{P}_x[\partial I(x)\cap B_N]$, and $F_K'^N(x) := \mathbb{R}^d$ otherwise. As $\mathbb{P}_X[\partial I(X)\cap B_N\cap K] = \mathbb{P}_X[\mathbf{1}_{Y\in\partial I(X)}\mathbf{1}_{X\notin N_\mu}\mathbf{1}_{Y\notin B_N\cap K}]$, $\mu-$a.s., where $N_\mu\in\mathcal{N}_\mu$ is taken from Lemma 2.6.1, $\mathbb{P}_X[\partial I(X)\cap B_N\cap K]$ is $\mu-$measurable, as equal $\mu-$a.s. to a Borel function. Then similarly, $\mathbb{P}_X[\partial I(X)\cap B_N\cap K]-\mathbb{P}_X[\partial I(X)\cap B_N]$ is $\mu-$measurable, and therefore $\overline{\mathrm{supp}}(\mathbb{P}_X|_{\partial I(X)})$ is $\mu-$measurable. □

Proof of Lemma 2.6.7 By $\gamma-$measurability of F, we may find a Borel function $F_B : \mathbb{R}^d \longrightarrow \mathrm{ri}\overline{\mathcal{K}}$ such that $F = F_B$, $\gamma-$a.s. Let a Borel $N_\gamma\in\mathcal{N}_\gamma$ such that $F = F_B$ on N_γ^c. By the fact that $\mathrm{ri}\overline{\mathcal{K}}$ is Polish, we may find a sequence $(F_n)_{n\geq 1}$ of Borel functions with finite image converging pointwise towards F_B when $n\longrightarrow\infty$. We will give an explicit expression for F_n that will be useful later in the proof. Let $(K_n)_{n\geq 1}\subset\mathrm{ri}\overline{\mathcal{K}}$ a dense sequence,

$$F_n(x) \quad := \quad \underset{K\in(K_i)_{i\leq n}}{\mathrm{argmin}}\ dist\Big(F_B(x), K\Big), \tag{2.7.4}$$

Where $dist$ is the distance on $\mathrm{ri}\overline{\mathcal{K}}$ that makes it Polish, and we chose the K with the smallest index in case of equality.

We fix $n\geq 1$, let $K\in F_n(N_\gamma^c)$, the image of F_n outside of N_γ, and $A_K := F_n^{-1}\Big(\{K\}\Big)$. We will modify the image of F_n so that it is the same for all $x'\in F_B(x) = F(x)$, for all $x\in N_\gamma^c\cap A_K$. Then we consider the set $A'_K := \cup_{x\in N_\gamma^c\cap A_K} F_B(x)$, we now prove that this set in analytic. By Theorem 4.2 (b) in [158], $GrF_B := \{Y\in\mathrm{cl}\,F_B(X)\}$ is a Borel set. Let $\lambda > 0$, we define the affine deformation $f_\lambda : \Omega\longrightarrow\Omega$ by $f_\lambda(X,Y) := \Big(X, X+\lambda(Y-X)\Big)$. By the fact that for $k\geq 1$, $f_{1-1/k}(GrF_B)$ is Borel together with the fact that $x\in f_B(x)$

for $x \notin N_\gamma$, we have

$$\{Y \in F_B(X)\} \cap \{X \notin N_\gamma\} = \cup_{k\geq 1} f_{1-1/k}(GrF_B) \cap \{X \notin N_\gamma\}.$$

Therefore, $\{Y \in F_B(X)\} \cap \{X \notin N_\gamma\}$ is Borel, and so is $\{Y \in F_B(X)\} \cap \{X \in N_\gamma^c \cap A_K\}$. Finally,

$$A'_K = Y\Big(\{Y \in F_B(X)\} \cap \{X \in N_\gamma^c \cap A_K\}\Big),$$

therefore, A'_K is the projection of a Borel set, which is one of the definitions of an analytic set (see Proposition 7.41 in [30]). Now we define a suitable modification of F_n by $F'_n(x) := K$ for all $x \in A'_K$, we do this redefinition for all $K \in F_B(N_\gamma^c)$. Notice that thanks to the definition (2.7.4) and the fact that $F_B(x) = F_B(x')$ if $x, x' \notin N_\gamma$ and $x' \in F_B(x) = F(x)$, we have the inclusion $A'_K \subset A_K \cup N_\gamma$. Then the redefinitions of F_n only hold outside of N_γ, furthermore for different $K_1, K_2 \in F_n(N_\gamma^c)$, $A'_{K_1} \cap A'_{K_2} = \emptyset$ as the value of $F_n(x)$ only depends on the value of $F_B(x)$ by (2.7.4). Notice that

$$N'_\gamma := \Big(\cup_{K \in F_n(N_\gamma^c)} A'_K\Big)^c = \Big(\cup_{x \notin N_\gamma} F_B(x)\Big)^c \subset N_\gamma, \tag{2.7.5}$$

is analytically measurable, as the complement of an analytic set, and does not depend on n. For $x \in N'_\gamma$, we define $F'_n(x) := \{x\}$. Notice that F'_n is analytically measurable as the modification of a Borel Function on analytically measurable sets.

Now we prove that F'_n converges pointwise when $n \longrightarrow \infty$. For $x \in N'_\gamma$, $F'_n(x)$ is constant equal to $\{x\}$, if $x \notin N'_\gamma$, by (2.7.5) $x \in \cup_{x \notin N_\gamma} F_B(x)$, and therefore $F'_n(x) = F_B(x') = F(x')$ for some $x \in N_\gamma^c$, for all $n \geq 1$. Then as $F'_n(x')$ converges to $F(x')$, $F'_n(x)$ converges to $F(x)$. Let F' be the pointwise limit of F'_n. the maps F'_n are analytically measurable, and therefore, so does F'. For all $n \geq 1$, $F'_n = F_n$, $\gamma-$a.e. and therefore $F' = F_B = F$, $\gamma-$a.e. Finally, $F'(N_\gamma^c) = F(N_\gamma^c)$, and $\cup F(N_\gamma^c) = (N'_\gamma)^c$. By property of F, $F'(N_\gamma^c)$ is a partition of $(N'_\gamma)^c$ such that $x \in F'(x)$ for all $x \notin N'_\gamma$. On N'_γ, this property is trivial as $F'(x) = \{x\}$ for all $x \in N'_\gamma$. □

2.8 Properties of tangent convex functions

2.8.1 x-invariance of the y-convexity

We first report a convex analysis lemma.

Lemma 2.8.1. *Let $f:\mathbb{R}^d\to\bar{\mathbb{R}}$ be convex finite on some convex open subset $U\subset\mathbb{R}^d$. We denote $f_*:\mathbb{R}^d\to\bar{\mathbb{R}}$ the lower-semicontinuous envelop of f on U, then*

$$f_*(y)=\lim_{\epsilon\searrow 0} f\Big(\epsilon x+(1-\epsilon)y\Big),\quad \textit{for all}\quad (x,y)\in U\times\mathrm{cl}U.$$

Proof. f_* is the lower semi-continuous envelop of f on U, i.e. the lower semi-continuous envelop of $f':=f+\infty\mathbf{1}_{U^c}$. Notice that f' is convex $\mathbb{R}^d\longrightarrow\mathbb{R}\cup\{\infty\}$. Then by Proposition 1.2.5 in Chapter IV of [89], we get the result as $f=f'$ on U. □

Proof of Proposition 2.3.10 The result is obvious in $\mathbf{T}(\mathfrak{C}_1)$, as the affine part depending on x vanishes. We may use $N_\nu=\emptyset$. Now we denote $\mathcal{T}$ the set of mappings in Θ_μ such that the result from the proposition holds. Then we have $\mathbf{T}(\mathfrak{C}_1)\subset\mathcal{T}$.

We prove that $\mathcal{T}$ is $\mu\otimes$pw$-$Fatou closed. Let $(\theta_n)_n$ be a sequence in $\mathcal{T}$ converging $\mu\otimes$pw to $\theta\in\Theta_\mu$. Let $n\geq 1$, we denote N_μ, the set in $\mathcal{N}_\mu$ from the proposition applied to θ_n, and let $N^0_\mu\in\mathcal{N}_\mu$ corresponding to the $\mu\otimes$pw convergence of θ_n to θ. We denote $N_\mu:=\cup_{n\in\mathbb{N}}N^n_\mu\in\mathcal{N}_\mu$. Let $x_1,x_2\notin N_\mu$, and $\bar y\in\mathrm{dom}_{x_1}\theta\cap\mathrm{dom}_{x_2}\theta$. Let $y_1,y_2\in\mathrm{dom}_{x_1}\theta$, such that we have the convex combination $\bar y=\lambda y_1+(1-\lambda)y_2$, and $0\leq\lambda\leq 1$. Then for $i=1,2$, $\theta_n(x_1,y_i)\longrightarrow\theta(x_1,y_i)$, and $\theta_n(x_1,\bar y)\longrightarrow\theta(x_1,\bar y)$, as $n\to\infty$. Using the fact that $\theta_n\in\mathcal{T}$, for all n, we have

$$\Delta_n:=\lambda\theta_n(x_i,y_1)+(1-\lambda)\theta_n(x_i,y_2)-\theta_n(x_i,\bar y)\geq 0,\ \text{and independent of } i=1,2. \tag{2.8.1}$$

Taking the limit $n\to\infty$ gives that $\underline{\theta}_\infty(x_2,y_i)<\infty$, and $y_i\in\mathrm{dom}\underline{\theta}_\infty(x_2,\cdot)$. $\bar y$ is interior to $\mathrm{dom}_{x_1}\theta$, then for any $y\in\mathrm{dom}_{x_1}\theta$, $y':=\bar y+\frac{\epsilon}{1-\epsilon}(\bar y-y)\in\mathrm{dom}_{x_1}\theta$ for $0<\epsilon<1$ small enough. Then $\bar y=\epsilon y+(1-\epsilon)y'$. As we may chose any $y\in\mathrm{dom}_{x_1}\theta$, we have $\mathrm{dom}_{x_1}\theta\subset\mathrm{dom}\underline{\theta}_\infty(x_2,\cdot)$. Then, we have

$$\mathrm{rf}_{x_2}\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)\subset\mathrm{rf}_{x_2}\mathrm{conv}\,\mathrm{dom}\Big(\underline{\theta}_\infty(x_2,\cdot)\Big)=\mathrm{dom}_{x_2}\theta. \tag{2.8.2}$$

By Lemma 2.9.1, as $\mathrm{dom}_{x_1}\theta\cap\mathrm{dom}_{x_2}\theta\neq\emptyset$, $\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)=\mathrm{ri}\,\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)$. In particular, $\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)$ is relatively open and contains x_2, and therefore $\mathrm{rf}_{x_2}\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)=\mathrm{conv}(\mathrm{dom}_{x_1}\theta\cup\mathrm{dom}_{x_2}\theta)$. Finally, by (2.8.2), $\mathrm{dom}_{x_1}\theta\subset\mathrm{dom}_{x_2}\theta$. As there is a symmetry between x_1, and x_2, we have $\mathrm{dom}_{x_1}\theta=\mathrm{dom}_{x_2}\theta$. Then we may go to the limit in equation (2.8.1):

$$\Delta_\infty:=\lambda\theta(x_i,y_1)+(1-\lambda)\theta(x_i,y_2)-\theta(x_i,\bar y)\geq 0, \tag{2.8.3}$$
$$\text{and independent of } i=1,2.$$

Now, let $y_1, y_2 \in \mathbb{R}^d$, such that we have the convex combination $\bar{y} = \lambda y_1 + (1-\lambda) y_2$, and $0 \le \lambda \le 1$. we have three cases to study.
<u>*Case 1:*</u> $y_i \notin \mathrm{cl}\,\mathrm{dom}_{x_1}\theta$ for some $i = 1, 2$. Then, as the average $\bar{y}$ of the y_i is in $\mathrm{dom}_{x_1}\theta$, by Proposition 2.3.1 (ii), me may find $i' = 1, 2$ such that $y_{i'} \notin \mathrm{conv}\,\mathrm{dom}\theta(x_1, \cdot)$, thus implying that $\theta(x_1, y_i) = \infty$. Then $\lambda\theta(x_1, y_1) + (1-\lambda)\theta(x_1, y_2) - \theta(x_1, \bar{y}) = \infty \ge 0$. As $\mathrm{dom}_{x_1}\theta = \mathrm{dom}_{x_2}\theta$, we may apply the same reasoning to x_2, we get $\lambda\theta(x_1, y_1) + (1-\lambda)\theta(x_2, y_2) - \theta(x_2, \bar{y}) = \infty \ge 0$. We get the result.
<u>*Case 2:*</u> $y_1, y_2 \in \mathrm{dom}_{x_1}\theta$. This case is (2.8.3).
<u>*Case 3:*</u> $y_1, y_2 \in \mathrm{cl}\,\mathrm{dom}_{x_1}\theta$. The problem arises here if some y_i is in the boundary $\partial\mathrm{dom}_{x_1}\theta$. Let $x \notin N_\mu$, we denote the lower semi-continuous envelop of $\theta(x, \cdot)$ in $\mathrm{cl}\,\mathrm{dom}_x\theta$, by $\theta_*(x, y) := \lim_{\epsilon \searrow 0} \theta(x, \epsilon x + (1-\epsilon)y')$, for $y \in \mathrm{cl}\,\mathrm{dom}_x\theta$, where the latest equality follows from Lemma 2.8.1 together with that fact that $\theta(x, \cdot)$ is convex on $\mathrm{dom}_x\theta$. Let $y \in \mathrm{cl}\,\mathrm{dom}_{x_1}\theta$, for $1 \ge \epsilon > 0$, $y^\epsilon := \epsilon x_1 + (1-\epsilon)y \in \mathrm{dom}_{x_1}\theta$. By (2.8.1), $(1-\epsilon)\theta_n(x_1, y) - \theta_n(x_1, y^\epsilon) = (1-\epsilon)\theta_n(x_2, y) - \theta_n(x_2, y^\epsilon)$. Taking the $\liminf$, we have $(1-\epsilon)\theta(x_1, y) - \theta(x_1, y^\epsilon) = (1-\epsilon)\theta(x_2, y) - \theta(x_2, y^\epsilon)$. Now taking $\epsilon \searrow 0$, we have $\theta(x_1, y) - \theta_*(x_1, y) = \theta(x_2, y) - \theta_*(x_2, y)$. Then the jump of $\theta(x, \cdot)$ in y is independent of $x = x_1$ or x_2. Now for $1 \ge \epsilon > 0$, by (2.8.3)

$$\lambda\theta(x_1, y_1^\epsilon) + (1-\lambda)\theta(x_1, y_2^\epsilon) - \theta(x_1, \bar{y}^\epsilon) = \lambda\theta(x_2, y_1^\epsilon) + (1-\lambda)\theta(x_2, y_2^\epsilon) - \theta(x_2, \bar{y}^\epsilon) \ge 0.$$

By going to the limit $\epsilon \searrow 0$, we get

$$\lambda\theta_*(x_1, y_1) + (1-\lambda)\theta_*(x_1, y_2) - \theta_*(x_1, \bar{y}) = \lambda\theta_*(x_2, y_1) + (1-\lambda)\theta_*(x_2, y_2) - \theta_*(x_2, \bar{y}) \ge 0.$$

As the (nonnegative) jumps do not depend on $x = x_1$ or x_2, we finally get

$$\lambda\theta(x_1, y_1) + (1-\lambda)\theta(x_1, y_2) - \theta(x_1, \bar{y}) = \lambda\theta(x_2, y_1) + (1-\lambda)\theta(x_2, y_2) - \theta(x_2, \bar{y}) \ge 0.$$

Finally, $\mathcal{T}$ is $\mu\otimes$pw$-$Fatou closed, and convex. $\widehat{\mathcal{T}}_1 \subset \mathcal{T}$. As the result is clearly invariant when the function is multiplied by a scalar, the Result is proved on $\widehat{\mathcal{T}}(\mu, \nu)$. □

2.8.2 Compactness

Proof of Proposition 2.3.7 We first prove the result for $\theta = (\theta_n)_{n \ge 1} \subset \Theta$. Denote $\mathrm{conv}(\theta) := \{\theta' \in \Theta^{\mathbb{N}} : \theta'_n \in \mathrm{conv}(\theta_k, k \ge n), n \in \mathbb{N}\}$. Consider the minimization problem:

$$m \;:=\; \inf_{\theta' \in \mathrm{conv}(\theta)} \mu[G(\mathrm{dom}_X \underline{\theta}'_\infty)], \tag{2.8.4}$$

where the measurability of $G(\mathrm{dom}_X \underline{\theta}'_\infty)$ follows from Lemma 2.3.13.
Step 1: We first prove the existence of a minimizer. Let $(\theta'^k)_{k\in\mathbb{N}} \in \mathrm{conv}(\theta)^{\mathbb{N}}$ be a minimizing sequence, and define the sequence $\widehat{\theta} \in \mathrm{conv}(\theta)$ by:

$$\widehat{\theta}_n := (1-2^{-n})^{-1} \textstyle\sum_{k=1}^n 2^{-k} \theta'^k_n, \; n \geq 1.$$

Then, $\mathrm{dom}(\underline{\widehat{\theta}}_\infty) \subset \bigcap_{k\geq 1} \mathrm{dom}(\underline{\theta}'^k_\infty)$ by the non-negativity of θ', and we have the inclusion $\left\{\widehat{\theta}_n \underset{n\to\infty}{\longrightarrow} \infty\right\} \subset \left\{\theta'^k_n \underset{n\to\infty}{\longrightarrow} \infty \text{ for some } k \geq 1\right\}$. Consequently,

$$\mathrm{dom}_x \underline{\widehat{\theta}}_\infty \;\subset\; \mathrm{conv}\Big(\textstyle\bigcap_{k\geq 1} \mathrm{dom}\underline{\theta}'^k_\infty(x,\cdot)\Big) \;\subset\; \textstyle\bigcap_{k\geq 1} \mathrm{dom}_x \underline{\theta}'^k_\infty \quad \text{for all} \quad x \in \mathbb{R}^d.$$

Since $(\theta'^k)_k$ is a minimizing sequence, and $\widehat{\theta} \in \mathrm{conv}(\theta)$, this implies that $\mu[G(\mathrm{dom}_X \underline{\widehat{\theta}}_\infty)] = m$.
Step 2: We next prove that we may find a sequence $(y_i)_{i\geq 1} \subset \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d)$ such that

$$\begin{aligned} y_i(X) &\in \mathrm{aff}(\mathrm{dom}_X \underline{\widehat{\theta}}_\infty), \\ \text{and} \quad (y_i(X))_{i\geq 1} &\text{ dense in } \mathrm{aff}\,\mathrm{dom}_X \underline{\widehat{\theta}}_\infty, \quad \mu-\text{a.s.} \end{aligned} \tag{2.8.5}$$

Indeed, it follows from Lemmas 2.3.13, and 2.7.2 that the map $x \mapsto \mathrm{aff}(\mathrm{dom}_x \underline{\widehat{\theta}}_\infty)$ is universally measurable, and therefore Borel-measurable up to a modification on a $\mu-$null set. Since its values are closed and nonempty, we deduce from the implication $(ii) \implies (ix)$ in Theorem 4.2 of the survey on measurable selection [158] the existence of a sequence $(y_i)_{i\geq 1}$ satisfying (2.8.5).
Step 3: Let $m(dx,dy) := \mu(dx) \otimes \sum_{i\geq 0} 2^{-i} \delta_{\{y_i(x)\}}(dy)$. By the Komlòs lemma (in the form of Lemma A1.1 in [59], similar to the one used in the proof of Proposition 5.2 in [22]), we may find $\widetilde{\theta} \in \mathrm{conv}(\widehat{\theta})$ such that $\widetilde{\theta}_n \longrightarrow \widetilde{\theta}_\infty \in \mathbb{L}^0(\Omega)$, $m-$a.s. Clearly, $\mathrm{dom}_x \underline{\widetilde{\theta}}_\infty \subset \mathrm{dom}_x \underline{\widehat{\theta}}_\infty$, and therefore $\mu\Big[G(\mathrm{dom}_X \underline{\widetilde{\theta}}_\infty)\Big] \leq \mu\Big[G(\mathrm{dom}_x \underline{\widehat{\theta}}_\infty)\Big]$, for all $x \in \mathbb{R}^d$. This shows that

$$G(\mathrm{dom}_X \underline{\widetilde{\theta}}_\infty) \;=\; G(\mathrm{dom}_X \underline{\widehat{\theta}}_\infty), \quad \mu-\text{a.s.} \tag{2.8.6}$$

so that $\widetilde{\theta}$ is also a solution of the minimization problem (2.8.4). Moreover, it follows from (2.3.3) that

$$\mathrm{ri}\,\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty = \mathrm{ri}\,\mathrm{dom}_X\underline{\widehat{\theta}}_\infty, \quad \text{and therefore} \quad \mathrm{aff}\,\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty = \mathrm{aff}\,\mathrm{dom}_X\underline{\widehat{\theta}}_\infty, \quad \mu-a.s.$$

<u>*Step 4:*</u> Notice that the values taken by $\widetilde{\theta}_\infty$ are only fixed on an $m-$full measure set. By the convexity of elements of Θ in the $y-$variable, $\mathrm{dom}_X\widetilde{\theta}_n$ has a nonempty interior in $\mathrm{aff}(\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty)$. Then as $\mu-$a.s., $\widetilde{\theta}_n(X,\cdot)$ is convex, the following definition extends $\widetilde{\theta}_\infty$ to Ω:

$$\widetilde{\theta}_\infty(x,y) := \sup\Big\{a\cdot y + b : (a,b)\in\mathbb{R}^d\times\mathbb{R},\ a\cdot y_n(x)+b \le \widetilde{\theta}_\infty(x,y_n(x)) \text{ for all } n\ge 0\Big\}.$$

This extension coincides with $\widetilde{\theta}_\infty$, in $(x,y_n(x))$ for $\mu-$a.e. $x\in\mathbb{R}^d$, and all $n\ge 1$ such that $y_n(x)\notin\partial\mathrm{dom}_X\widetilde{\theta}_k$ for some $k\ge 1$ such that $\mathrm{dom}_x\widetilde{\theta}_n$ has a nonempty interior in $\mathrm{aff}(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$. As for k large enough, $\partial\mathrm{dom}_X\widetilde{\theta}_k$ is Lebesgue negligible in $\mathrm{aff}(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$, the remaining $y_n(x)$ are still dense in $\mathrm{aff}(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$. Then, for $\mu-$a.e. $x\in\mathbb{R}^d$, $\widetilde{\theta}_n(x,\cdot)$ converges to $\widetilde{\theta}_\infty(x,\cdot)$ on a dense subset of $\mathrm{aff}(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$. We shall prove in Step 6 below that

$$\mathrm{dom}\,\widetilde{\theta}_\infty(X,\cdot) \text{ has nonempty interior in } \mathrm{aff}(\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty), \quad \mu-\text{a.s.} \tag{2.8.7}$$

Then, by Theorem 2.9.3, we have the convergence $\widetilde{\theta}_n(X,\cdot)\longrightarrow\widetilde{\theta}_\infty(X,\cdot)$, pointwise on $\mathrm{aff}(\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty)\setminus\partial\mathrm{dom}\widetilde{\theta}_\infty(X,\cdot)$, $\mu-$a.s. Since $\mathrm{dom}_X\theta_\infty = \mathrm{dom}_X\underline{\theta}_\infty$, and $\widetilde{\theta}$ converges to θ_∞ on $\mathrm{dom}_X\theta_\infty$, $\mu-$a.s., $\widetilde{\theta}$ converges to $\theta_\infty\in\Theta$, $\mu\otimes$pw.

<u>*Step 5:*</u> Finally for general $(\theta_n)_{n\ge1}\subset\Theta_\mu$, we consider θ'_n, equal to θ_n, $\mu\otimes$pw, such that $\theta'_n\le\theta_n$, for $n\ge 1$, from the definition of Θ_μ. Then we may find λ^k_n, coefficients such that $\widehat{\theta}'_n := \sum_{k\ge n}\lambda^k_n\theta'_k\in\mathrm{conv}(\theta')$ converges $\mu\otimes$pw to $\widehat{\theta}_\infty\in\Theta$. We denote $\widehat{\theta}_n := \sum_{k\ge n}\lambda^k_n\theta_k\in\mathrm{conv}(\theta)$, $\widehat{\theta}_n = \widehat{\theta}'_n$, $\mu\otimes$pw, and $\widehat{\theta}_n\ge\widehat{\theta}'_n$. By Proposition 2.3.6 (iii), $\widehat{\theta}$ converges to $\widehat{\theta}_\infty$, $\mu\otimes$pw. The Proposition is proved.

<u>*Step 6:*</u> In order to prove (2.8.7), suppose to the contrary that there is a set A such that $\mu[A]>0$ and $\mathrm{dom}\widetilde{\theta}_\infty(x,\cdot)$ has an empty interior in $\mathrm{aff}(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$ for all $x\in A$. Then, by the density of the sequence $(y_n(x))_{n\ge1}$ stated in (2.8.5), we may find for all $x\in A$ an index $i(x)\ge 0$ such that

$$\widehat{y}(x) := y_{i(x)}(x)\in\mathrm{ri}\,\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty, \quad \text{and} \quad \widetilde{\theta}_\infty(x,\widehat{y}(x)) = \infty. \tag{2.8.8}$$

Moreover, since $i(x)$ takes values in $\mathbb{N}$, we may reduce to the case where $i(x)$ is a constant integer, by possibly shrinking the set A, thus guaranteeing that $\widehat{y}$ is measurable.

Define the measurable function on Ω:

$$\theta_n^0(x,y) := \mathrm{dist}(y, L_x^n), \qquad \text{with} \quad L_x^n := \Big\{ y \in \mathbb{R}^d : \widetilde{\theta}_n(x,y) < \widetilde{\theta}_n(x,\widehat{y}(x)) \Big\}. \tag{2.8.9}$$

Since L_x^n is convex, and contains x for n sufficiently large by (2.8.8), we see that

$$\theta_n^0 \;\text{ is convex in } y \text{ and } \;\theta_n^0(x,y) \le |x-y|, \;\text{ for all }\; (x,y) \in \Omega. \tag{2.8.10}$$

In particular, this shows that $\theta_n^0 \in \Theta$. By Komlòs Lemma, we may find

$$\widehat{\theta}_n^0 := \textstyle\sum_{k\ge n} \lambda_k^n \theta_k^0 \in \mathrm{conv}(\theta^0) \;\text{ such that }\; \widehat{\theta}_n^0 \longrightarrow \widehat{\theta}_\infty^0, \;\; m-\text{a.s.}$$

for some non-negative coefficients $(\lambda_k^n, k \ge n)_{n\ge 1}$ with $\sum_{k\ge n} \lambda_k^n = 1$. By convenient extension of this limit, we may assume that $\widehat{\theta}_\infty^0 \in \Theta$. We claim that

$$\widehat{\theta}_\infty^0 > 0 \text{ on } H_x := \{h(x)\cdot(y-\widehat{y}(x)) > 0\}, \text{ for some } h(x) \in \mathbb{R}^d. \tag{2.8.11}$$

We defer the proof of this claim to Step 7 below and we continue in view of the required contradiction. By definition of θ_n^0 together with (2.8.10), we compute that

$$\begin{aligned}
\theta_n^1(x,y) := \sum_{k\ge n} \lambda_k^n \widetilde{\theta}_k(x,y) &\ge \sum_{k\ge n} \lambda_k^n \widetilde{\theta}_k(x,\widehat{y}(x)) \mathbf{1}_{\{\theta_n^0>0\}} \\
&\ge \sum_{k\ge n} \lambda_k^n \widetilde{\theta}_k(x,\widehat{y}(x)) \frac{\theta_k^0(x,y)}{|x-y|} \\
&\ge \frac{\widehat{\theta}_n^0(x,y)}{|x-y|} \inf_{k\ge n} \widetilde{\theta}_k(x,\widehat{y}(x)).
\end{aligned}$$

By (2.8.8) and (2.8.11), this shows that the sequence $\theta^1 \in \mathrm{conv}(\theta)$ satisfies

$$\theta_n^1(x,\cdot) \longrightarrow \infty, \quad \text{on} \quad H_x, \quad \text{for all} \quad x \in A.$$

We finally consider the sequence $\widetilde{\theta}^1 := \frac{1}{2}(\widetilde{\theta} + \theta^1) \in \mathrm{conv}(\theta)$. Clearly, $\mathrm{dom}\underline{\widetilde{\theta}}_\infty^1(X,\cdot) \subset \mathrm{dom}\underline{\widetilde{\theta}}_\infty(X,\cdot)$, and it follows from the last property of θ^1 that $\mathrm{dom}\underline{\widetilde{\theta}}_\infty^1(x,\cdot) \subset H_x^c \cap \mathrm{dom}\underline{\widetilde{\theta}}_\infty(x,\cdot)$ for all $x \in A$. Notice that $\widehat{y}(x)$ lies on the boundary of the half space H_x and, by (2.8.8), $\widehat{y}(x) \in \mathrm{ridom}_x\underline{\widetilde{\theta}}_\infty$. Then $G(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty^1) < G(\mathrm{dom}_x\underline{\widetilde{\theta}}_\infty)$ for all $x \in A$ and, since $\mu[A] > 0$, we deduce that $\mu\Big[G(\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty^1)\Big] < \mu\Big[G(\mathrm{dom}_X\underline{\widetilde{\theta}}_\infty)\Big]$, contradicting the optimality of $\widetilde{\theta}$, by (2.8.6), for the minimization problem (2.8.4).

Step 7: It remains to justify (2.8.11). Since $\widetilde{\theta}_n(x,\cdot)$ is convex, it follows from the Hahn-Banach separation theorem that:

$$\widetilde{\theta}_n(x,\cdot) \geq \widetilde{\theta}_n(x,\widehat{y}(x)) \text{ on } H^n_x := \Big\{ y \in \mathbb{R}^d : h^n(x)\cdot(y-\widehat{y}(x)) > 0 \Big\},$$

for some $h^n(x) \in \mathbb{R}^d$, so that it follows from (2.8.9) that $L^n_x \subset (H^n_x)^c$, and

$$\theta^0_n(x,y) \geq \mathrm{dist}\Big(y,(H^n_x)^c\Big) = \Big[\Big(y-\widehat{y}(x)\Big)\cdot h^n(x)\Big]^+.$$

Denote $g_x := g_{\mathrm{dom}_x \widehat{\underline{\theta}}_\infty}$ the Gaussian kernel restricted to the affine span of $\mathrm{dom}_x \widehat{\underline{\theta}}_\infty$, and $B_r(x_0)$ the corresponding ball with radius r, centered at some point x_0. By (2.8.8), we may find r^x so that $B^x_r := B_r(\widehat{y}(x)) \subset \mathrm{ri}\,\mathrm{dom}_x \widetilde{\underline{\theta}}_\infty$ for all $r \leq r^x$, and

$$\begin{aligned} \int_{B^x_r} \theta^0_n(x,y) g_x(y) dy &\geq \int_{B^x_r} \Big[\Big(y-\widehat{y}(x)\Big)\cdot h^n(x)\Big]^+ g_x(y)dy \\ &\geq \min_{B^x_r} g_x \int_{B_r(0)} (y\cdot e_1)^+ dy =: b^r_x > 0, \end{aligned}$$

where e_1 is an arbitrary unit vector of the affine span of $\mathrm{dom}_x \widehat{\underline{\theta}}_\infty$. Then we have the inequality $\int_{B^x_r} \widehat{\theta}^0_n(x,y) g_x(y) dy \geq b^r_x$, and since $\widehat{\theta}^0_n$ has linear growth in y by (2.8.10), it follows from the dominated convergence theorem that $\int_{B^r_x} \widehat{\theta}^0_\infty(x,y) g(dy) \geq b^r_x > 0$, and therefore $\widehat{\theta}^0_\infty(x,y^r_x) > 0$ for some $y^r_x \in B^r_x$. From the arbitrariness of $r \in (0,r_x)$, We deduce (2.8.11) as a consequence of the convexity of $\widehat{\theta}^0(x,\cdot)$. □

Proof of Proposition 2.3.6 (iii) We need to prove the existence of some

$$\theta' \in \Theta \quad \text{such that} \quad \underline{\theta}_\infty = \theta', \ \ \mu\otimes\mathrm{pw}, \ \ \text{and} \ \ \underline{\theta}_\infty \geq \theta'. \tag{2.8.12}$$

For simplicity, we denote $\theta := \underline{\theta}_\infty$. Let

$$F^1 := \mathrm{cl}\,\mathrm{conv}\,\mathrm{dom}\theta(X,\cdot), \quad F^k := \mathrm{cl}\,\mathrm{conv}\Big(\mathrm{dom}\theta(X,\cdot)\cap \mathrm{aff}\,\mathrm{rf}_X F^{k-1}\Big), \ k \geq 2,$$
$$\text{and} \quad F := \cup_{n\geq 1}(F^n \setminus \mathrm{cl}\,\mathrm{rf}_X F^n) \cup \mathrm{cl}\,\mathrm{dom}_X \theta.$$

Fix some sequence $\varepsilon_n \searrow 0$, and denote $\theta_* := \liminf_{n\to\infty} \theta\Big(X, \varepsilon_n X + (1-\varepsilon_n)Y\Big)$, and

$$\theta' := \Big[\infty \mathbf{1}_{Y\notin F(X)} + \mathbf{1}_{Y\in \mathrm{cl}\,\mathrm{dom}_X\theta}\theta_*\Big]\mathbf{1}_{X\notin N_\mu},$$

where $N_\mu \in \mathcal{N}_\mu$ is chosen such that $\mathbf{1}_{Y\in F^k(X)}\mathbf{1}_{X\notin N_\mu}$ are Borel measurable for all k from Lemma 2.6.1, and $\theta(x,\cdot)$ (resp. $\theta_n(x,\cdot)$) is convex finite on $\mathrm{dom}_x\theta$ (resp. $\mathrm{dom}_x\theta_n$), for $x \notin N_\mu$. Consequently, θ' is measurable. In the following steps, we verify that θ' satisfies (2.8.12).

Step 1: We prove that $\theta' \in \Theta$. Indeed, $\theta' \in \mathbb{L}^0_+(\Omega)$, and $\theta'(X,X)=0$. Now we prove that $\theta'(x,\cdot)$ is convex for all $x \in \mathbb{R}^d$. For $x \in N_\mu$, $\theta'(x,\cdot)=0$. For $x \notin N_\mu$, $\theta(x,\cdot)$ is convex finite on $\mathrm{dom}_x\theta$, then by the fact that $\mathrm{dom}_x\theta$ is a convex relatively open set containing x, it follows from Lemma 2.8.1 that $\theta_*(x,\cdot)=\lim_{n\to\infty}\theta\Big(x,\varepsilon_n x+(1-\varepsilon_n)\cdot\Big)$ is the lower semi-continuous envelop of $\theta(x,\cdot)$ on $\mathrm{cl\,dom}_x\theta$. We now prove the convexity of $\theta'(x,\cdot)$ on all $\mathbb{R}^d$. We denote $\widehat{F}(x) := F(x)\setminus \mathrm{cl\,dom}_x\theta$ so that $\mathbb{R}^d = F(x)^c \cup \widehat{F}(x) \cup \mathrm{cl\,dom}_x\theta$. Now, let $y_1, y_2 \in \mathbb{R}^d$, and $\lambda \in (0,1)$. If $y_1 \in F(x)^c$, the convexity inequality is verified as $\theta'(x,y_1)=\infty$. Moreover, $\theta'(x,\cdot)$ is constant on $\widehat{F}(x)$, and convex on $\mathrm{cl\,dom}_x\theta$. We shall prove in Steps 4 and 5 below that

$$F(x) \text{ is convex, and } \mathrm{rf}_x F(x) = \mathrm{dom}_x\theta. \tag{2.8.13}$$

In view of Proposition 2.3.1 (ii), this implies that the sets $\widehat{F}(x)$ and $\mathrm{cl\,dom}_x\theta$ are convex. Then we only need to consider the case when $y_1 \in \widehat{F}(x)$, and $y_2 \in \mathrm{cl\,dom}_x\theta$. By Proposition 2.3.1 (ii) again, we have $[y_1,y_2) \subset \widehat{F}(x)$, and therefore $\lambda y_1 + (1-\lambda)y_2 \in \widehat{F}(x)$, and $\theta'(x,\lambda y_1+(1-\lambda)y_2)=0$, which guarantees the convexity inequality.

Step 2: We next prove that $\theta=\theta'$, $\mu\otimes$pw. By the second claim in (2.8.13), it follows that $\theta_*(X,\cdot)$ is convex finite on $\mathrm{dom}_X\theta$, $\mu-$a.s. Then as a consequence of Proposition 2.3.4 (ii), we have $\mathrm{dom}_X\theta' = \mathrm{dom}_X(\infty\mathbf{1}_{Y\notin F(X)}) \cap \mathrm{dom}_X(\theta_*\mathbf{1}_{Y\in\mathrm{cl\,dom}_X\theta})$, $\mu-$a.s. The first term in this intersection is $\mathrm{rf}_X F(X) = \mathrm{dom}_X\theta$. The second contains $\mathrm{dom}_X\theta$, as it is the dom_X of a function which is finite on $\mathrm{dom}_X\theta$, which is convex relatively open, containing X. Finally, we proved that $\mathrm{dom}_X\theta = \mathrm{dom}_X\theta'$, $\mu-$a.s. Then $\theta'(X,\cdot)$ is equal to $\theta_*(X,\cdot)$ on $\mathrm{dom}_X\theta$, and therefore, equal to $\theta(X,\cdot)$, $\mu-$a.s. We proved that $\theta=\theta'$, $\mu\otimes$pw.

Step 3: We finally prove that $\theta' \le \theta$ pointwise. We shall prove in Step 6 below that

$$\mathrm{dom}\theta(X,\cdot) \quad \subset \quad F. \tag{2.8.14}$$

Then, $\infty\mathbf{1}_{Y\notin F(X)}\mathbf{1}_{X\notin N_\mu} \le \theta$, and it remains to prove that

$$\theta(x,y) \ge \theta_*(x,y) \quad \text{for all} \quad y \in \mathrm{cl\,dom}_x\theta, \quad x \notin N_\mu.$$

To see this, let $x \notin N_\mu$. By definition of N_μ, $\theta_n(x,\cdot) \longrightarrow \theta(x,\cdot)$ on $\mathrm{dom}_x\theta$. Notice that $\theta(x,\cdot)$ is convex on $\mathrm{dom}_x\theta$, and therefore as a consequence of Lemma 2.8.1,

$$\theta_*(x,y) = \lim_{\epsilon\searrow 0}\theta\Big(x, \epsilon x + (1-\epsilon)y\Big), \quad \text{for all} \quad y \in \mathrm{cl}\,\mathrm{dom}_x\theta.$$

Then $y^\epsilon := (1-\epsilon)y + \epsilon x \in \mathrm{dom}_x\theta_n$, for $\varepsilon \in (0,1]$, and n sufficiently large by (i) of this Proposition, and therefore $(1-\epsilon)\theta_n(x,y) - \theta_n(x,y_\epsilon) \geq (1-\epsilon)\theta'_n(x,y) - \theta'_n(x,y_\epsilon) \geq 0$, for $\theta'_n \in \Theta$ such that $\theta'_n = \theta_n$, $\mu\otimes\mathrm{pw}$, and $\theta_n \geq \theta'_n$. Taking the $\liminf$ as $n \to \infty$, we get $(1-\epsilon)\theta(x,y) - \theta(x,y_\epsilon) \geq 0$, and finally $\theta(x,y) \geq \lim_{\epsilon\searrow 0}\theta\Big(x,\epsilon x + (1-\epsilon)y\Big) = \theta'(x,y)$, by sending $\epsilon \searrow 0$.

<u>Step 4:</u> (First claim in (2.8.13)) Let $x_0 \in \mathbb{R}^d$, let us prove that $F(x_0)$ is convex. Indeed, let $x,y \in F(x_0)$, and $0<\lambda<1$. Since $\mathrm{cl}\,\mathrm{dom}_x\theta$ is convex, and $F^n(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_X F^n(x_0)$ is convex by Proposition 2.3.1 (ii), we only examine the following non-obvious cases:

- Suppose $x \in F^n(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_{x_0}F^n(x_0)$, and $y \in F^p(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_{x_0}F^p(x_0)$, with $n < p$. Then as $F^p(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_{x_0}F^p(x_0) \subset \mathrm{cl}\,\mathrm{rf}_{x_0}F^n(x_0)$, we have $\lambda x + (1-\lambda)y \in F^n(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_{x_0}F^n(x_0)$ by Proposition 2.3.1 (ii).
- Suppose $x \in F^n(x_0)\setminus \mathrm{cl}\,\mathrm{rf}_{x_0}F^n(x_0)$, and $y \in \mathrm{cl}\,\mathrm{dom}_{x_0}\theta$, then as $\mathrm{cl}\,\mathrm{dom}_{x_0}\theta \subset \mathrm{cl}\,\mathrm{rf}_{x_0}F^n(x_0)$, this case is handled similar to previous case.

<u>Step 5:</u> (Second claim in (2.8.13)). We have $\mathrm{dom}_X\theta \subset F(X)$, and therefore $\mathrm{dom}_X\theta \subset \mathrm{rf}_X F(X)$. Now we prove by induction on $k \geq 1$ that $\mathrm{rf}_X F(X) \subset \cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta$. The inclusion is trivially true for $k=1$. Let $k \geq 1$, we suppose that the inclusions holds for k, hence $\mathrm{rf}_X F(X) \subset \cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta$. As $\cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta \subset F^k$. Applying rf_X gives

$$\begin{aligned}
\mathrm{rf}_X F(X) &\subset \mathrm{rf}_X\Big[\cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta\Big]\\
&= \mathrm{rf}_X\Big[F^k\cap\cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta\Big]\\
&= \mathrm{rf}_X F^k\cap \mathrm{rf}_X\Big[\cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta\Big]\\
&\subset \mathrm{cl}\,\mathrm{rf}_X F^k\cap\cup_{n\geq k}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta\\
&\subset \cup_{n\geq k+1}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n)\cup \mathrm{cl}\,\mathrm{dom}_X\theta.
\end{aligned}$$

Then the result is proved for all k. In particular we apply it for $k = d+1$. Recall from the proof of Lemma 2.3.13 that for $n \geq d+1$, F^n is stationary at the value $\mathrm{cl}\,\mathrm{dom}_X\theta$. Then $\cup_{n\geq d+1}(F^n\setminus \mathrm{cl}\,\mathrm{rf}_X F^n) = \emptyset$, and $\mathrm{rf}_X F(X) \subset \mathrm{rf}_X \mathrm{cl}\,\mathrm{dom}_X\theta = \mathrm{dom}_X\theta$. The result is proved.

<u>*Step 6:*</u> We finally prove (2.8.14). Indeed, $\mathrm{dom}\theta(X,\cdot) \subset F^1$ by definition. Then

$$
\begin{aligned}
\mathrm{dom}\theta(X,\cdot) &\subset F^1 \setminus \mathrm{aff} F^1 \cup \Big(\cup_{2\le k\le d+1} (\mathrm{dom}\theta(X,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_X F^{k-1}) \setminus \mathrm{aff} F^k \Big) \cup F^{d+1} \\
&\subset F^1 \setminus \mathrm{cl}\, F^1 \cup \Big(\cup_{k\ge 2}\, \mathrm{cl}\,\mathrm{conv}(\mathrm{dom}\theta(X,\cdot) \cap \mathrm{aff}\,\mathrm{rf}_X F^{k-1}) \setminus \mathrm{cl}\, F^k \Big) \cup \mathrm{cl}\,\mathrm{dom}_X\theta \\
&= \cup_{k\ge 1} F^k \setminus \mathrm{cl}\, F^k \cup \mathrm{cl}\,\mathrm{dom}_X \theta = F.
\end{aligned}
$$

□

2.9 Some convex analysis results

As a preparation, we first report a result on the union of intersecting relative interiors of convex subsets which was used in the proof of Proposition 2.4.1. We shall use the following characterization of the relative interior of a convex subset K of $\mathbb{R}^d$:

$$
\begin{aligned}
\mathrm{ri}K &= \Big\{ x \in \mathbb{R}^d : x - \epsilon(x'-x) \in K \text{ for some } \epsilon > 0, \text{ for all } x' \in K \Big\} && (2.9.1)\\
&= \Big\{ x \in \mathbb{R}^d : x \in (x', x_0], \text{ for some } x_0 \in \mathrm{ri}K, \text{ and } x' \in K \Big\}. && (2.9.2)
\end{aligned}
$$

We start by proving the required properties of the notion of relative face.

Proof of Proposition 2.3.1 <u>(i)</u> The proofs of the first properties raise no difficulties and are left as an exercise for the reader. We only prove that $\mathrm{rf}_a A = \mathrm{ri}A \neq \emptyset$ iff $a \in \mathrm{ri}A$. We assume that $\mathrm{rf}_a A = \mathrm{ri}A \neq \emptyset$. The non-emptiness implies that $a \in A$, and therefore $a \in \mathrm{rf}_a A = \mathrm{ri}A$. Now we suppose that $a \in \mathrm{ri}A$. Then for $x \in \mathrm{ri}A$, $\big[x, a - \epsilon(x-a)\big] \subset \mathrm{ri}A \subset A$, for some $\epsilon > 0$, and therefore $\mathrm{ri}A \subset \mathrm{rf}_a A$. On the other hand, by (2.9.2), $\mathrm{ri}A = \{x \in \mathbb{R}^d : x \in (x', x_0], \text{ for some } x_0 \in \mathrm{ri}A, \text{ and } x' \in A\}$. Taking $x_0 := a \in \mathrm{ri}A$, we have the remaining inclusion $\mathrm{rf}_a A \subset \mathrm{ri}A$.
<u>(ii)</u> We now assume that A is convex.

<u>*Step 1:*</u> We first prove that $\mathrm{rf}_a A$ is convex. Let $x, y \in \mathrm{rf}_a A$ and $\lambda \in [0,1]$. We consider $\epsilon > 0$ such that $\big(a - \epsilon(x-a), x + \epsilon(x-a)\big) \subset A$ and $\big(a - \epsilon(y-a), y + \epsilon(y-a)\big) \subset A$. Then if we write $z = \lambda x + (1-\lambda) y$, we have $\big(a - \epsilon(z-a), z + \epsilon(z-a)\big) \subset A$ by convexity of A, because $a, x, y \in A$.

<u>*Step 2:*</u> In order to prove that $\mathrm{rf}_a A$ is relatively open, we consider $x, y \in \mathrm{rf}_a A$, and we verify that $\big(x - \epsilon(y-x), y + \epsilon(y-x)\big) \subset \mathrm{rf}_a A$ for some $\epsilon > 0$. Consider the two alternatives:

<u>*Case 1:*</u> If a, x, y are on a line. If $a = x = y$, then the required result is obvious. Otherwise,

$$\Big(a - \epsilon(x-a), x + \epsilon(x-a)\Big) \cup \Big(a - \epsilon(y-a), y + \epsilon(y-a)\Big) \subset \mathrm{rf}_a A.$$

This union is open in the line and x and y are interior to it. We can find $\epsilon' > 0$ such that $\Big(x - \epsilon'(y-x), y + \epsilon'(y-x)\Big) \subset \mathrm{rf}_a A$.

<u>*Case 2:*</u> If a, x, y are not on a line. Let $\epsilon > 0$ be such that $\Big(a - 2\epsilon(x-a), x + 2\epsilon(x-a)\Big) \subset A$ and $\Big(a - 2\epsilon(y-a), y + 2\epsilon(y-a)\Big) \subset A$. Then $x + \epsilon(x-a) \in \mathrm{rf}_a A$ and $a - \epsilon(y-a) \in \mathrm{rf}_a A$. Then, if we take $\lambda := \frac{\epsilon}{1+2\epsilon}$,

$$\lambda(a - \epsilon(y-a)) + (1-\lambda)(x + \epsilon(x-a)) = (1-\lambda)(1+\epsilon)x - \lambda\epsilon y = x + \lambda\epsilon(x-y).$$

Then $x + \lambda\epsilon(x-y) \in \mathrm{rf}_a A$ and symmetrically, $y + \lambda\epsilon(y-x) \in \mathrm{rf}_a A$ by convexity of $\mathrm{rf}_a A$. And still by convexity, we have that $\Big(x - \epsilon'(y-x), y + \epsilon'(y-x)\Big) \subset \mathrm{rf}_a A$, for $\epsilon' := \frac{\epsilon^2}{1+2\epsilon} > 0$.

<u>*Step 3:*</u> Now we prove that $A \setminus \mathrm{cl}\,\mathrm{rf}_a A$ is convex, and that if $x_0 \in A \setminus \mathrm{cl}\,\mathrm{rf}_a A$ and $y_0 \in A$, then $[x_0, y_0) \subset A \setminus \mathrm{cl}\,\mathrm{rf}_a A$. We will prove these two results by an induction on the dimension of the space d. First if $d = 0$ the results are trivial. Now we suppose that the result is proved for any $d' < d$, let us prove it for dimension d.

<u>*Case 1:*</u> $a \in \mathrm{ri} A$. This case is trivial as $\mathrm{rf}_a A = \mathrm{ri} A$ and $A \subset \mathrm{cl}\,\mathrm{ri} A = \mathrm{cl}\,\mathrm{rf}_a A$ because of the convexity of A. Finally $A \setminus \mathrm{cl}\,\mathrm{rf}_a A = \emptyset$ which makes it trivial.

<u>*Case 2:*</u> $a \notin \mathrm{ri} A$. Then $a \in \partial A$ and there exists a hyperplan support H to A in a because of the convexity of A. We will write the equation of E, the corresponding half-space containing A, $E : c \cdot x \leq b$ with $c \in \mathbb{R}^d$ and $b \in \mathbb{R}$. As $x \in \mathrm{rf}_a A$ implies that $[a - \epsilon(x-a), x + \epsilon(x-a)] \subset A$ for some $\epsilon > 0$, we have $(a - \epsilon(x-a)) \cdot c \leq b$ and $(x + \epsilon(x-a)) \cdot c \leq b$. These equations are equivalent using that $a \in H$ and thus $a \cdot c = b$ to $-\epsilon(x-a) \cdot c \leq 0$ and $(1+\epsilon)(x-a) \cdot c \leq 0$. We finally have $(x-a) \cdot c = 0$ and $x \in H$. We proved that $\mathrm{rf}_a A \subset H$.

Now using (i) together with the fact that $\mathrm{rf}_a A \subset H$ and $a \in H$ affine, we have

$$\mathrm{rf}_a(A \cap H) = \mathrm{rf}_a A \cap \mathrm{rf}_a H = \mathrm{rf}_a A \cap H = \mathrm{rf}_a A.$$

Then we can now have the induction hypothesis on $A \cap H$ because $\dim H = d - 1$ and $A \cap H \subset H$ is convex. Then we have $A \cap H \setminus \mathrm{cl}\,\mathrm{rf}_a A$ which is convex and if $x_0 \in A \cap H \setminus \mathrm{cl}\,\mathrm{rf}_a(A \cap H)$, $y_0 \in A \cap H$ and if $\lambda \in (0,1]$ then $\lambda x_0 + (1-\lambda) y_0 \in A \setminus \mathrm{cl}\,\mathrm{rf}_a(A \cap H)$.

First $A\setminus \mathrm{cl}\,\mathrm{rf}_a A = (A\setminus H)\cup\Big((A\cap H)\setminus \mathrm{cl}\,\mathrm{rf}_a A\Big)$, let us show that this set is convex. The two sets in the union are convex ($A\setminus H = A\cap(E\setminus H)$), so we need to show that a non trivial convex combination of elements coming from both sets is still in the union. We consider $x\in A\setminus H$, $y\in A\cap H\setminus \mathrm{cl}\,\mathrm{rf}_a A$ and $\lambda>0$, let us show that $z:=\lambda x+(1-\lambda)y\in (A\setminus H)\cup(A\cap H\setminus \mathrm{cl}\,\mathrm{rf}_a A)$. As $x,y\in A$ ($\mathrm{cl}\,\mathrm{rf}_a A\subset A$ because A is closed), $z\in A$ by convexity of A. We now prove $z\notin H$,

$$z\cdot c = \lambda x\cdot c+(1-\lambda)y\cdot c = \lambda x\cdot c+(1-\lambda)b < \lambda b+(1-\lambda)b = b.$$

Then z is in the strict half space: $z\in E\setminus H$. Finally $z\in A\setminus H$ and $A\setminus \mathrm{cl}\,\mathrm{rf}_a A$ is convex. Let us now prove the second part: we consider $x_0\in A\setminus \mathrm{cl}\,\mathrm{rf}_a A$, $y_0\in \mathrm{cl}\,\mathrm{rf}_a A$ and $\lambda\in(0,1]$ and write $z_0:=\lambda x_0+(1-\lambda)y_0$.

<u>*Case 2.1:*</u> $x_0,y_0\in H$. We apply the induction hypothesis.

<u>*Case 2.2:*</u> $x_0,y_0\in A\setminus H$. Impossible because $\mathrm{rf}_a A\subset H$ and $\mathrm{cl}\,\mathrm{rf}_a A\subset \mathrm{cl}\,H = H$. $y_0\in H$.

<u>*Case 2.3:*</u> $x_0\in A\setminus H$ and $y_0\in H$. Then by the same computation than in Step 1,

$$z_0\in A\setminus H\subset A\setminus \mathrm{cl}\,\mathrm{rf}_a A.$$

<u>*Step 4:*</u> Now we prove that if $a\in A$, then $\dim(\mathrm{rf}_a \mathrm{cl}\,A)=\dim(A)$ if and only if $a\in \mathrm{ri}A$, and that in this case, we have $\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A = \mathrm{cl}\,\mathrm{ri}\,\mathrm{cl}\,A = \mathrm{cl}\,A = \mathrm{cl}\,\mathrm{rf}_a A$. We first assume that $a\in\mathrm{ri}A$. As by the convexity of A, $\mathrm{ri}A = \mathrm{ri}\,\mathrm{cl}\,A$, $\mathrm{rf}_a\mathrm{cl}\,A = \mathrm{ri}\,\mathrm{cl}\,A$, and therefore $\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A = \mathrm{cl}\,A$. Finally, taking the dimension, we have $\dim(\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A)=\dim(A)$. In this case we proved as well that $\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A = \mathrm{cl}\,\mathrm{ri}\,\mathrm{cl}\,A = \mathrm{cl}\,A = \mathrm{cl}\,\mathrm{rf}_a A$, the last equality coming from the fact that $\mathrm{ri}A = \mathrm{rf}_a A$ as $a\in\mathrm{ri}A$.

Now we assume that $a\notin\mathrm{ri}A$. Then $a\in\partial\mathrm{cl}\,A$, and $\mathrm{rf}_a\mathrm{cl}\,A\subset\partial\mathrm{cl}\,A$. Taking the dimension (in the local sense this time), and by the fact that $\dim\partial\mathrm{cl}\,A = \dim\partial A<\dim A$, we have $\dim(\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A)<\dim(A)$ (as $\mathrm{cl}\,\mathrm{rf}_a\mathrm{cl}\,A$ is convex, the two notions of dimension coincide). $\square$

Lemma 2.9.1. *Let $K_1,K_2\subset\mathbb{R}^d$ be convex with $\mathrm{ri}K_1\cap\mathrm{ri}K_2\neq\emptyset$. Then $\mathrm{conv}(\mathrm{ri}K_1\cup\mathrm{ri}K_2) = \mathrm{ri}\,\mathrm{conv}(K_1\cup K_2)$.*

Proof. We fix $y\in\mathrm{ri}K_1\cap\mathrm{ri}K_2$.

Let $x\in\mathrm{conv}(\mathrm{ri}K_1\cup\mathrm{ri}K_2)$, we may write $x=\lambda x_1+(1-\lambda)x_2$, with $x_1\in\mathrm{ri}K_1$, $x_2\in\mathrm{ri}K_2$, and $0\le\lambda\le1$. If λ is 0 or 1, we have trivially that $x\in\mathrm{ri}\,\mathrm{conv}(K_1\cap K_2)$. Let us now treat the case $0<\lambda<1$. Then for $x'\in\mathrm{conv}(K_1\cup K_2)$, we may write $x'=\lambda' x_1'+(1-\lambda')x_2'$, with $x_1'\in K_1$, $x_2'\in K_2$, and $0\le\lambda'\le1$. We will use y as a center as it is in both the sets. For all the variables, we add a bar on it when we subtract y, for example $\bar{x} := x - y$. The geometric problem is the same when translated with y,

$$\bar{x} - \epsilon(\bar{x}' - \bar{x}) \;=\; \lambda\left(\bar{x}_1 - \epsilon\left(\frac{\lambda'}{\lambda}\bar{x}_1' - \bar{x}_1\right)\right) + (1-\lambda)\left(\bar{x}_2 - \epsilon\left(\frac{1-\lambda'}{1-\lambda}\bar{x}_2' - \bar{x}_2\right)\right). \quad (2.9.3)$$

However, as $\bar{x}_1$ and $\bar{x}_1'$ are in $K_1 - y$, as 0 is an interior point, $\epsilon(\frac{\lambda'}{\lambda}\bar{x}_1' - \bar{x}_1) \in K_1 - y$ for ϵ small enough. Then as $\bar{x}_1$ is interior to $K_1 - y$ as well, $\bar{x}_1 - \epsilon(\frac{\lambda'}{\lambda}\bar{x}_1' - \bar{x}_1) \in K_1 - y$ as well. By the same reasoning, $\bar{x}_2 - \epsilon(\frac{1-\lambda'}{1-\lambda}\bar{x}_2' - \bar{x}_2) \in K_2 - y$. Finally, by (2.9.3), for ϵ small enough, $x - \epsilon(x' - x) \in \mathrm{conv}(K_1 \cup K_2)$. By (2.9.1), $x \in \mathrm{ri\,conv}(K_1 \cup K_2)$.

Now let $x \in \mathrm{ri\,conv}(K_1 \cup K_2)$. We use again y as an origin with the notation $\bar{x} := x - y$. As $\bar{x}$ is interior, we may find $\epsilon > 0$ such that $(1+\epsilon)\bar{x} \in \mathrm{conv}(K_1 \cup K_2)$. We may write $(1+\epsilon)\bar{x} = \lambda\bar{x}_1 + (1-\lambda)\bar{x}_2$, with $\bar{x}_1 \in K_1 - y$, $\bar{x}_2 \in K_2 - y$, and $0 \le \lambda \le 1$. Then $\bar{x} = \lambda\frac{1}{1+\epsilon}\bar{x}_1 + (1-\lambda)\frac{1}{1+\epsilon}\bar{x}_2$. By (2.9.2), $\frac{1}{1+\epsilon}\bar{x}_1 \in \mathrm{ri}K_1$, and $\frac{1}{1+\epsilon}\bar{x}_2 \in \mathrm{ri}K_2$. $\bar{x} \in \mathrm{conv}\Big(\mathrm{ri}(K_1 - y) \cup \mathrm{ri}(K_2 - y)\Big)$, and therefore $x \in \mathrm{conv}(\mathrm{ri}K_1 \cup \mathrm{ri}K_2)$. □

Now we use the measurable selection theory to establish the non-emptiness of ∂f.

Lemma 2.9.2. *For all $f \in \mathfrak{C}$, we have $\partial f \neq \emptyset$.*

Proof. By the fact that f is continuous, we may write $\partial f(x) = \cap_{n\ge1} F_n(x)$ for all $x \in \mathbb{R}^d$, with $F_n(x) := \{p \in \mathbb{R}^d : f(y_n) - f(x) \ge p\cdot(y_n - x)\}$ where $(y_n)_{n\ge1} \subset \mathbb{R}^d$ is some fixed dense sequence. All F_n are measurable by the continuity of $(x,p) \longmapsto f(y_n) - f(x) - p\cdot(y_n - x)$ together with Theorem 6.4 in [88]. Therefore the mapping $x \longmapsto \partial f(x)$ is measurable by Lemma 2.7.1. Moreover, the fact that this mapping is closed nonempty-valued is a well-known property of the subgradient of finite convex functions in finite dimension. Then the result holds by Theorem 4.1 in [158]. □

We conclude this section with the following result which has been used in our proof of Proposition 2.3.7. We believe that this is a standard convex analysis result, but we could not find precise references. For this reason, we report the proof for completeness.

Theorem 2.9.3. *Let $f_n, f : \mathbb{R}^d \to \bar{\mathbb{R}}$ be convex functions with $\mathrm{int\,dom}f \neq \emptyset$. Then $f_n \longrightarrow f$ pointwise on $\mathbb{R}^d \setminus \partial\mathrm{dom}f$ if and only if $f_n \longrightarrow f$ pointwise on some dense subset $A \subset \mathbb{R}^d \setminus \partial\mathrm{dom}f$.*

Proof. We prove the non-trivial implication "if". We first prove the convergence on $\mathrm{int\,dom}f$. f_n converges to f on a dense set. The reasoning will consist in proving that the f_n are Lipschitz, it will give a uniform convergence and then a pointwise convergence. First we consider $K \subset \operatorname{int}\operatorname{dom} f$ compact convex with nonempty interior. We can find $N \in \mathbb{N}$ and $x_1, \ldots x_N \in A \cap (\operatorname{int}\operatorname{dom} f \setminus K)$ such that $K \subset \operatorname{int}\operatorname{conv}(x_1, \ldots, x_N)$. We use the pointwise convergence on A to get that for n large enough, $f_n(x) \leq M$ for $x \in \operatorname{conv}(x_1, \ldots, x_N)$, $M > 0$ (take $M = \max_{1 \leq k \leq N} f(x_k) + 1$). Then we will prove that f_n is bounded from below on K. We consider $a \in A \cap K$ and $\delta_0 := \sup_{x \in K} |x - a|$. For n large enough, $f_n(a) \geq m$ for any $a \in A$ (take for example $m = f(a) - 1$). We write $\delta_1 := \min_{(x,y) \in K \times \partial\operatorname{conv}(x_1, \ldots, x_N)} |x - y|$. Finally we write $\delta_2 := \sup_{x,y \in \operatorname{conv}(x_1, \ldots, x_N)} |x - y|$. Now, for $x \in K$, we consider the half line $x + \mathbb{R}_+(a - x)$, it will cut $\partial\operatorname{conv}(x_1, \ldots, x_N)$ in one only point $y \in \partial\operatorname{conv}(x_1, \ldots, x_N)$. Then $a \in [x, y]$, and therefore $a = \frac{|a-y|}{|x-y|} x + \frac{|a-x|}{|x-y|} y$. By the convex inequality, $f_n(a) \leq \frac{|a-y|}{|x-y|} f_n(x) + \frac{|a-x|}{|x-y|} f_n(y)$. Then $f_n(x) \geq -\frac{|a-x|}{|a-y|} M + \frac{|x-y|}{|a-y|} m \geq -\frac{\delta_0}{\delta_1} M + \frac{\delta_2}{\delta_1} m$. Finally, if we write $m_0 := -\frac{\delta_0}{\delta_1} M + \frac{\delta_2}{\delta_1} m$,

$$M \geq f_n \geq m_0, \quad \text{on } K.$$

This will prove that f_n is $\frac{M - m0}{\delta_1}$-Lipschitz. We consider $x \in K$ and a unit direction $u \in \mathcal{S}^{d-1}$ and $f_n' \in \partial f_n(x)$. For a unique $\lambda > 0$, $y := x + \lambda u \in \partial\operatorname{conv}(x_1, \ldots, x_N)$. As u is a unit vector, $\lambda = |y - x| \geq \delta_1$. By the convex inequality, $f_n(y) \geq f_n(x) + f_n'(x) \cdot (y - x)$. Then $M - m_0 \geq \delta_0 |f_n' \cdot u|$ and finally $|f_n' \cdot u| \leq \frac{M - m0}{\delta_1}$ as this bound does not depend on u, $|f_n'| \leq \frac{M - m0}{\delta_1}$ for any such subgradient. For n large enough, the f_n are uniformly Lipschitz on K, and so in f. The convergence is uniform on K, it is then pointwise on K. As this is true for any such K, the convergence is pointwise on $\operatorname{int}\operatorname{dom} f$.

Now let us consider $x \in (\operatorname{cl}\operatorname{dom} f)^c$. The set $\operatorname{conv}(x, \operatorname{int}\operatorname{dom} f) \setminus \operatorname{dom} f$ has a nonempty interior because $\operatorname{dist}(x, \operatorname{dom} f) > 0$ and $\operatorname{int}\operatorname{dom} f \neq \emptyset$. As A is dense, we can consider $a \in A \cap \operatorname{conv}(x, \operatorname{int}\operatorname{dom} f) \setminus \operatorname{dom} f$. By definition of $\operatorname{conv}(x, \operatorname{int}\operatorname{dom} f)$, we can find $y \in \operatorname{int}\operatorname{dom} f$ such that $a = \lambda y + (1 - \lambda) x$. We have $\lambda < 1$ because $a \notin \operatorname{dom} f$. If $\lambda = 0$, $f_n(x) = f_n(a) \underset{n \to \infty}{\longrightarrow} \infty$. Otherwise, by the convexity inequality, $f_n(a) \leq \lambda f_n(y) + (1 - \lambda) f_n(x)$. Then, as $f_n(a) \underset{n \to \infty}{\longrightarrow} \infty$, and $f_n(y) \underset{n \to \infty}{\longrightarrow} f(y) < \infty$, we have $f_n(x) \underset{n \to \infty}{\longrightarrow} \infty$. $\square$

Chapter 3

Quasi-sure duality for multi-dimensional martingale optimal transport

Based on the multidimensional irreducible paving of De March & Touzi [58], we provide a multi-dimensional version of the quasi sure duality for the martingale optimal transport problem, thus extending the result of Beiglböck, Nutz & Touzi [22]. Similar to [22], we also prove a disintegration result which states a natural decomposition of the martingale optimal transport problem on the irreducible components, with pointwise duality verified on each component. As another contribution, we extend the martingale monotonicity principle to the present multi-dimensional setting. Our results hold in dimensions 1, 2, and 3 provided that the target measure is dominated by the Lebesgue measure. More generally, our results hold in any dimension under an assumption which is implied by the Continuum Hypothesis. Finally, in contrast with the one-dimensional setting of [21], we provide an example which illustrates that the smoothness of the coupling function does not imply that pointwise duality holds for compactly supported measures.

Key words. Martingale optimal transport, duality, disintegration, monotonicity principle.

3.1 Introduction

The problem of martingale optimal transport was introduced as the dual of the problem of robust (model-free) superhedging of exotic derivatives in financial mathematics, see Beiglböck, Henry-Labordère & Penkner [18] in discrete time, and Galichon, Henry-Labordère & Touzi [73] in continuous-time. This robust superhedging problem was introduced by Hobson [94], and was addressing specific examples of exotic derivatives by means of corresponding solutions of the Skorokhod embedding problem, see [51, 92, 93], and the survey [91].

Given two probability measures μ,ν on $\mathbb{R}^d$, with finite first order moment, martingale optimal transport differs from standard optimal transport in that the set of all interpolating probability measures $\mathcal{P}(\mu,\nu)$ on the product space is reduced to the subset $\mathcal{M}(\mu,\nu)$ restricted by the martingale condition. We recall from Strassen [146] that $\mathcal{M}(\mu,\nu)\neq\emptyset$ if and only if $\mu\preceq\nu$ in the convex order, i.e. $\mu(f)\leq\nu(f)$ for all convex functions f. Notice that the inequality $\mu(f)\leq\nu(f)$ is a direct consequence of the Jensen inequality, the reverse implication follows from the Hahn-Banach theorem.

This paper focuses on proving that quasi-sure duality holds in higher dimension, thus extending the results by Beiglböck, Nutz and Touzi [22] who prove that quasi-sure duality holds by identifying the polar sets. The structure of these polar sets is given by the critical observation by Beiglböck & Juillet [20] that, in the one-dimensional setting $d=1$, any such martingale interpolating probability measure $\mathbb{P}$ has a canonical decomposition $\mathbb{P}=\sum_{k\geq 0}\mathbb{P}_k$, where $\mathbb{P}_k\in\mathcal{M}(\mu_k,\nu_k)$ and μ_k is the restriction of μ to the so-called irreducible components I_k, and $\nu_k:=\int_{x\in I_k}\mathbb{P}(dx,\cdot)$, supported in J_k for $k\geq 0$, is independent of the choice of $\mathbb{P}\in\mathcal{M}(\mu,\nu)$. Here, $(I_k)_{k\geq 1}$ are open intervals, $I_0:=\mathbb{R}\setminus(\cup_{k\geq 1}I_k)$, and J_k is an augmentation of I_k by the inclusion of either one of the endpoints of I_k, depending on whether they are charged by the distribution $\mathbb{P}_k$.

In [22], this irreducible decomposition gives a form of compactness of the convex functions on each components, and plays a crucial role for the quasi-sure formulation, and represents an important difference between martingale transport and standard transport. Indeed, while the martingale transport problem is affected by the quasi-sure formulation, the standard optimal transport problem is not changed. We also refer to Ekren & Soner [65] for further functional analytic aspects of this duality.

Our objective in this paper is to extend the quasi-sure duality, find a disintegration on the components, and a monotonicity principle for an arbitrary $d-$dimensional setting, $d\geq 1$. The main difficulty is that convex functions may lose information when converging. A first attempt to find such duality results was achieved by Ghoussoub, Kim & Lim [74]. Their strategy consists in finding the largest sets on which pointwise monotonicity holds, and prove that it implies a pointwise existence of dual optimisers.

The paper is organized as follows. Section 3.2 collects the main technical ingredients needed for the definition of the relaxed dual problem in view of the statement of our main results. Section 3.3 contains the main results of the paper, namely the duality for the relaxed dual problem, the disintegration of the problem in the irreducible components identified in [58], and a monotonicity principle. In all the cases there are some claims that hold without any need of assumption, and a second part using Assumption 3.2.6 defined in the beginning of the section. Section 3.4 shows the identity with the Beiglböck, Nutz & Touzi [20] duality theorems in the one-dimensional setting, and provides non-intuitive examples, in particular Example 3.4.1 showing that there is no hope of having pointwise duality. The remaining sections contain the proofs of these results. In particular, Section 5.6 contains the proofs of the main results, and Section 3.6 checks the situations in which Assumption 3.2.6 holds.

Notation We denote by $\bar{\mathbb{R}}$ the completed real line $\mathbb{R}\cup\{-\infty,\infty\}$, and similarly denote $\overline{\mathbb{R}}_+ := \mathbb{R}_+\cup\{\infty\}$. We fix an integer $d\geq 1$. If $x\in\mathcal{X}$, and $A\subset\mathcal{X}$, where $(\mathcal{X},\mathrm{d})$ is a metric space, $\mathrm{dist}(x,A) := \inf_{a\in A}\mathrm{d}(x,a)$. In all this paper, $\mathbb{R}^d$ is endowed with the Euclidean distance.

If V is a topological affine space and $A\subset V$ is a subset of V, $\mathrm{int}A$ is the interior of A, $\mathrm{cl}\,A$ is the closure of A, $\mathrm{aff}A$ is the smallest affine subspace of V containing A, $\mathrm{conv}A$ is the convex hull of A, $\dim(A) := \dim(\mathrm{aff}A)$, and $\mathrm{ri}A$ is the relative interior of A, which is the interior of A in the topology of $\mathrm{aff}A$ induced by the topology of V. We also denote by $\partial A := \mathrm{cl}\,A\setminus\mathrm{ri}A$ the relative boundary of A. If A is an affine subspace of $\mathbb{R}^d$, we denote by proj_A the orthogonal projection on A, and ∇A is the vector space associated to A (i.e. $A-a$ for $a\in A$, independent of the choice of a). We finally denote $\mathrm{Aff}(V,\mathbb{R})$ the collection of affine maps from V to $\mathbb{R}$.

The set $\mathcal{K}$ of all closed subsets of $\mathbb{R}^d$ is a Polish space when endowed with the Wijsman topology[1] (see Beer [16]). As $\mathbb{R}^d$ is separable, it follows from a theorem of Hess [86] that a function $F:\mathbb{R}^d\longrightarrow\mathcal{K}$ is Borel measurable with respect to the Wijsman topology if and only if

$$F^-(V) := \{x\in\mathbb{R}^d : F(x)\cap V\neq\emptyset\}\quad\text{is Borel for each open subset } V\subset\mathbb{R}^d.$$

[1]The Wijsman topology on the collection of all closed subsets of a metric space $(\mathcal{X},\mathrm{d})$ is the weak topology generated by $\{\mathrm{dist}(x,\cdot) : x\in\mathcal{X}\}$.

The subset $\overline{\mathcal{K}} \subset \mathcal{K}$ of all the convex closed subsets of $\mathbb{R}^d$ is closed in $\mathcal{K}$ for the Wijsman topology, and therefore inherits its Polish structure. Clearly, $\overline{\mathcal{K}}$ is isomorphic to $\mathrm{ri}\overline{\mathcal{K}} := \{\mathrm{ri}K : K \in \overline{\mathcal{K}}\}$ (with reciprocal isomorphism cl). We shall identify these two isomorphic sets in the rest of this text, when there is no possible confusion.

We denote $\Omega := \mathbb{R}^d \times \mathbb{R}^d$ and define the two canonical maps

$$X : (x,y) \in \Omega \longmapsto x \in \mathbb{R}^d \quad \text{and} \quad Y : (x,y) \in \Omega \longmapsto y \in \mathbb{R}^d.$$

For $\varphi, \psi : \mathbb{R}^d \longrightarrow \bar{\mathbb{R}}$, and $h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, we denote

$$\varphi \oplus \psi := \varphi(X) + \psi(Y), \quad \text{and} \quad h^{\otimes} := h(X) \cdot (Y - X),$$

with the convention $\infty - \infty = \infty$. Finally, for $A \subset \Omega$, and $x \in \mathbb{R}^d$, we denote $A_x := \{y \in \mathbb{R}^d : (x,y) \in A\}$, and $A_x^c := \{y \in \mathbb{R}^d : (x,y) \notin A\}$.

For a Polish space $\mathcal{X}$, we denote by $\mathcal{B}(\mathcal{X})$ the collection of Borel subsets of $\mathcal{X}$, and $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\big(\mathcal{X}, \mathcal{B}(\mathcal{X})\big)$. For $\mathbb{P} \in \mathcal{P}(\mathcal{X})$, we denote by $\mathcal{N}_{\mathbb{P}}$ the collection of all $\mathbb{P}-$null sets, $\mathrm{supp}\,\mathbb{P}$ the smallest closed support of $\mathbb{P}$, and $\overline{\mathrm{supp}}\,\mathbb{P} := \mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\,\mathbb{P}$ the smallest convex closed support of $\mathbb{P}$. For a measurable function $f : \mathcal{X} \to \mathbb{R}$, we use again the convention $\infty - \infty = \infty$ to define its integral, and we denote

$$\mathbb{P}[f] := \mathbb{E}^{\mathbb{P}}[f] = \int_{\mathcal{X}} f d\mathbb{P} = \int_{\mathcal{X}} f(x)\mathbb{P}(dx) \quad \text{for all} \quad \mathbb{P} \in \mathcal{P}(\mathcal{X}).$$

Let $\mathcal{Y}$ be another Polish space, and $\mathbb{P} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The corresponding conditional kernel $\mathbb{P}_x$ is defined by:

$$\mathbb{P}(dx, dy) = \mu(dx) \otimes \mathbb{P}_x(dy), \text{ where } \mu := \mathbb{P} \circ X^{-1}.$$

We denote by $\mathbb{L}^0(\mathcal{X}, \mathcal{Y})$ the set of Borel measurable maps from $\mathcal{X}$ to $\mathcal{Y}$. We denote for simplicity $\mathbb{L}^0(\mathcal{X}) := \mathbb{L}^0(\mathcal{X}, \bar{\mathbb{R}})$ and $\mathbb{L}^0_+(\mathcal{X}) := \mathbb{L}^0(\mathcal{X}, \bar{\mathbb{R}}_+)$. For a measure m on $\mathcal{X}$, we denote $\mathbb{L}^1(\mathcal{X}, m) := \{f \in \mathbb{L}^0(\mathcal{X}) : m[|f|] < \infty\}$. We also denote simply $\mathbb{L}^1(m) := \mathbb{L}^1(\bar{\mathbb{R}}, m)$ and $\mathbb{L}^1_+(m) := \mathbb{L}^1_+(\bar{\mathbb{R}}_+, m)$.

We denote by $\mathfrak{C}$ the collection of all finite convex functions $f : \mathbb{R}^d \longrightarrow \mathbb{R}$. We denote by $\partial f(x)$ the corresponding subgradient at any point $x \in \mathbb{R}^d$. We also introduce the collection of all measurable selections in the subgradient, which is nonempty (see e.g. Lemma 9.2 in [58]),

$$\partial f := \Big\{p \in \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d) : p(x) \in \partial f(x) \text{ for all } x \in \mathbb{R}^d\Big\}.$$

Let $f:\mathbb{R}^d\longrightarrow\overline{\mathbb{R}}$, $f_{conv}(x):=\sup\{g(x)$ such that $g:\mathbb{R}^d\longrightarrow\overline{\mathbb{R}}$ is convex and $g\le f\}$ denotes the lower convex envelop of f. We also denote $\underline{f}_\infty:=\liminf_{n\to\infty}f_n$, for any sequence $(f_n)_{n\ge1}$ of real number, or of real-valued functions.

Let $I:\mathbb{R}^d\longmapsto\overline{\mathcal{K}}$ be the irreducible components mapping defined in [58], which is the $\mu-$a.s. unique mapping such that for some $\widehat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$, $\mathrm{ri\,conv\,supp}\,\widehat{\mathbb{P}}_X=I(X)\supset\mathrm{ri\,conv\,supp}\,\mathbb{P}_X$, $\mu-$a.s. for all $\mathbb{P}\in\mathcal{M}(\mu,\nu)$.

3.2 The relaxed dual problem

3.2.1 Preliminaries

Throughout this paper, we consider two probability measures μ and ν on $\mathbb{R}^d$ with finite first order moment, and $\mu\preceq\nu$ in the convex order, i.e. $\nu(f)\ge\mu(f)$ for all $f\in\mathfrak{C}$. Using the convention $\infty-\infty=\infty$, we may then define $(\nu-\mu)(f)\in[0,\infty]$ for all $f\in\mathfrak{C}$.

We denote by $\mathcal{M}(\mu,\nu)$ the collection of all probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals $\mathbb{P}\circ X^{-1}=\mu$ and $\mathbb{P}\circ Y^{-1}=\nu$. Notice that $\mathcal{M}(\mu,\nu)\neq\emptyset$ by Strassen [146].

An $\mathcal{M}(\mu,\nu)-$polar set is an element of $\mathcal{N}_{\mu,\nu}:=\cap_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\mathcal{N}_{\mathbb{P}}$. A property is said to hold $\mathcal{M}(\mu,\nu)-$quasi surely (abbreviated as q.s.) if it holds on the complement of an $\mathcal{M}(\mu,\nu)-$polar set.

For a derivative contract defined by a non-negative cost function $c:\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}_+$, the martingale optimal transport problem is defined by:

$$\mathbf{S}_{\mu,\nu}(c) \quad := \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\mathbb{P}[c]. \tag{3.2.1}$$

The corresponding robust superhedging problem is

$$\mathbf{I}_{\mu,\nu}(c) \quad := \inf_{(\varphi,\psi,h)\in\mathcal{D}_{\mu,\nu}(c)}\mu(\varphi)+\nu(\psi), \tag{3.2.2}$$

where

$$\mathcal{D}_{\mu,\nu}(c) \quad := \quad \Big\{(\varphi,\psi,h)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^0(\mu,\mathbb{R}^d):\ \varphi\oplus\psi+h^\otimes\ge c\Big\}. \tag{3.2.3}$$

The following inequality is immediate:

$$\mathbf{S}_{\mu,\nu}(c)\le\mathbf{I}_{\mu,\nu}(c). \tag{3.2.4}$$

This inequality is the so-called weak duality. For upper semi-continuous cost function, Beiglböck, Henry-Labordère, and Penckner [18] proved that there is no duality gap, i.e. $\mathbf{S}_{\mu,\nu}(c) = \mathbf{I}_{\mu,\nu}(c)$. See also Zaev [160]. The objective of this paper is to establish a similar duality result for general measurable positive cost functions, thus extending the findings of Beiglböck, Nutz, and Touzi [22].

For a probability $\mathbb{P} \in \mathcal{P}(\Omega)$, we say that $\mathbb{P}' \in \mathcal{P}(\Omega)$ is a competitor to $\mathbb{P}$ if $\mathbb{P} \circ X^{-1} = \mathbb{P}' \circ X^{-1}$, $\mathbb{P} \circ Y^{-1} = \mathbb{P}' \circ Y^{-1}$, and $\mathbb{P}[Y|X] = \mathbb{P}'[Y|X]$. Let $f : \Omega \longrightarrow \bar{\mathbb{R}}$, we say that a set $A \subset \Omega$ is $f-$martingale monotone if for all probability $\mathbb{P}$ having a finite support in A, and for all competitor $\mathbb{P}'$ to $\mathbb{P}$, we have $\mathbb{P}[f] \geq \mathbb{P}'[f]$.

3.2.2 Tangent convex functions

Definition 3.2.1. *Let $\theta : \Omega \to \overline{\mathbb{R}}_+$ be a universally measurable function, and a Borel set $N \in \mathcal{N}_{\mu,\nu}$ with $\{X = Y\} \subset N^c$. We say that θ is a $N-$tangent convex function if*
(i) *$\theta(x,x) = 0$, and $\theta(x,\cdot)$ is partially convex in y on N^c_x;*
(ii) *N^c is $\theta-$martingale monotone;*
(iii) *for all $\mathbb{P}$ with finite support in N^c, and any competitor $\mathbb{P}'$ to $\mathbb{P}$ such that* $\operatorname{supp}\mathbb{P}' \cap N$ *is a singleton, we have $\mathbb{P}'[N] = 0$;*
(iv) *$A := \{X \notin N_\mu\} \cap \{Y \in I(X)\} \subset N^c$, and $\mathbf{1}_A\theta$ is finite Borel measurable, for some $N_\mu \in \mathcal{N}_\mu$.*

We denote by $\Theta_{\mu,\nu}$ the collection of all functions θ which are $N-$tangent convex for some N as above. Clearly, $\Theta_{\mu,\nu} \supset \{\mathbf{T}_p f : f \in \mathfrak{C}, p \in \partial f\}$, where

$$\mathbf{T}_p f(x,y) := f(y) - f(x) - p^{\otimes}(x,y), \quad \text{for all} \quad f : \mathbb{R}^d \longmapsto \bar{\mathbb{R}}, \text{ and } p : \mathbb{R}^d \longmapsto \mathbb{R}^d.$$

Indeed, for $f \in \mathfrak{C}$, and $p \in \partial f$, $\mathbf{T}_p f$ is convex in the second variable, thus satisfying (i) with $N = \emptyset$. For all $\mathbb{P}_0$ with finite support in $N^c = \Omega$, and $\mathbb{P}'$ competitor to $\mathbb{P}_0$, $\mathbb{P}_0[f(X)] = \mathbb{P}'[f(X)]$, $\mathbb{P}_0[f(Y)] = \mathbb{P}'[f(Y)]$, and $\mathbb{P}_0[p(X)\cdot(Y-X)] = \mathbb{P}_0[p(X)\cdot(\mathbb{P}_0[Y|X]-X)] = \mathbb{P}'[p(X)\cdot(Y-X)]$, and therefore $\mathbb{P}_0[\mathbf{T}_p f] = \mathbb{P}'[\mathbf{T}_p f]$.

Definition 3.2.2. *We say that a sequence $(\theta_n)_{n\geq 1} \subset \Theta_{\mu,\nu}$ generates some $\theta \in \Theta_{\mu,\nu}$ (and we denote $\theta_n \rightsquigarrow \theta$) if*

$$\underline{\theta}_\infty \leq \theta, \quad \textit{and} \quad \mathbb{P}[\theta] \leq \limsup_{n\to\infty} \mathbb{P}[\theta_n], \ \textit{for all} \ \mathbb{P} \in \mathcal{P}(\Omega).$$

Notice that some sequences in $\Theta_{\mu,\nu}$ may generate infinitely many elements of $\Theta_{\mu,\nu}$. For example, for any nonzero $\theta \in \Theta_{\mu,\nu}$, the sequence $(\theta_n)_{n\in\mathbb{N}} := (0,\theta,0,\theta,...)$ generates any $\theta' \in \Theta_{\mu,\nu}$ which is smaller than θ. In particular $\theta_n \rightsquigarrow x\theta$, as n goes to infinity, for all $0 \le x \le 1$, which are uncountably many.

Definition 3.2.3. (i) *A subset $\mathcal{T} \subset \Theta_{\mu,\nu}$ is semi-closed if $\theta \in \mathcal{T}$ for all $(\theta_n)_{n\ge1} \subset \mathcal{T}$ generating θ (in particular, $\Theta_{\mu,\nu}$ is semi-closed).*
(ii) *The semi-closure of a subset $A \subset \Theta_{\mu,\nu}$ is the smallest semi-closed set containing A:*

$$\widetilde{A} := \bigcap \Big\{ \mathcal{T} \subset \Theta_{\mu,\nu} : \; A \subset \mathcal{T}, \text{ and } \mathcal{T} \text{ semi-closed} \Big\}.$$

We next introduce for $a \ge 0$ the set $\mathfrak{C}_a := \Big\{ f \in \mathfrak{C} : (\nu - \mu)(f) \le a \Big\}$, and

$$\widetilde{\mathcal{T}}(\mu,\nu) := \bigcup_{a \ge 0} \widetilde{\mathcal{T}}_a, \text{ where } \mathcal{T}_a := \Big\{ \mathbf{T}_p f : f \in \mathfrak{C}_a, p \in \partial f \Big\}.$$

Remark 3.2.4. *Notice that even though the construction of $\widetilde{\mathcal{T}}(\mu,\nu)$ is very similar to the construction of $\widehat{\mathcal{T}}(\mu,\nu)$ in [58], these objects may be different, see Lemma 3.5.4 below.*

Proposition 3.2.5. *$\widetilde{\mathcal{T}}(\mu,\nu)$ is a convex cone.*

Proof. The proof is similar to the proof of Proposition 2.9 in [58], using the fact that for $\theta, \theta_n, \theta_\infty \in \Theta_{\mu,\nu}$, the generation $\theta_n \rightsquigarrow \theta_\infty$ implies the generation $\theta_n + \theta \rightsquigarrow \theta_\infty + \theta$. □

3.2.3 Structure of polar sets

The main results of this paper require the following assumption.

Assumption 3.2.6. (i) *For all $(\theta_n)_{n\ge1} \subset \widetilde{\mathcal{T}}_1$, we may find $\theta \in \widetilde{\mathcal{T}}_1$ such that $\theta_n \rightsquigarrow \theta$.*
(ii) *$I(X) \in \mathcal{C} \cup \mathcal{D} \cup \mathcal{R}$, $\mu-$a.s. for some subsets $\mathcal{C}, \mathcal{D}, \mathcal{R} \subset \overline{\mathcal{K}}$ with $\mathcal{C}$ well ordered, $\dim(\mathcal{D}) \subset \{0,1\}$, and $\cup_{K \ne K' \in \mathcal{R}} \Big[K \times (\operatorname{cl} K \cap \operatorname{cl} K') \Big] \in \mathcal{N}_{\mu,\nu}$.*

The condition $\cup_{K \ne K' \in \mathcal{R}} \Big[K \times (\operatorname{cl} K \cap \operatorname{cl} K') \Big] \in \mathcal{N}_{\mu,\nu}$ means that the probabilities in $\mathcal{M}(\mu,\nu)$ do not charge the intersections between frontiers of elements in $\mathcal{R}$, see Figure 3.1.

We provide in Section 3.3.4 some simple sufficient conditions for the last assumption to hold true. In particular, Assumption 3.2.6 holds true in dimensions $d = 1,2$, in dimension 3 with ν dominated by the Lebesgue measure, and in arbitrary dimension under the continuum hypothesis.

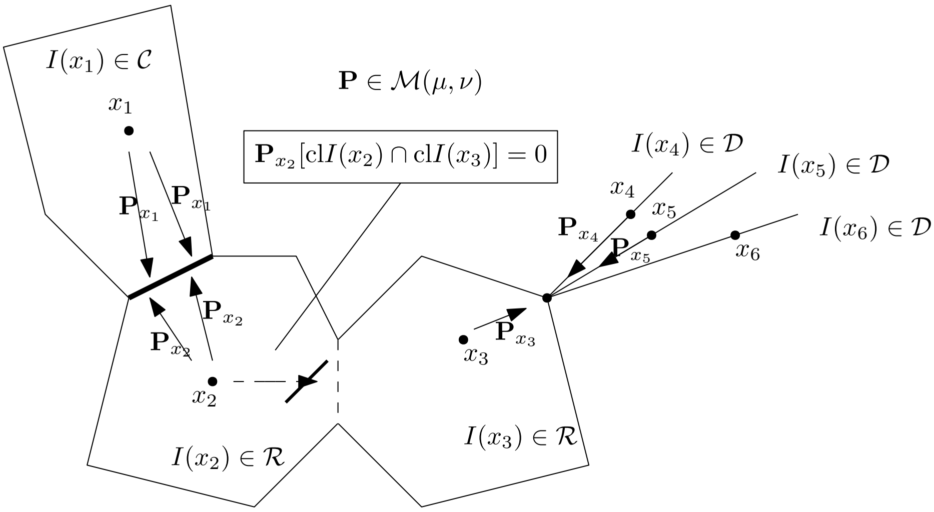


Fig. 3.1 No communication between frontiers of elements in $\mathcal{R}$.

Recall that by Theorem 3.7 in [58], a Borel set $N \in \mathcal{B}(\Omega)$ is $\mathcal{M}(\mu,\nu)-$polar if and only if

$$N \subset \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J_\theta(X)\}, \text{ for some } (N_\mu, N_\nu, \theta) \in \mathcal{N}_\mu \times \mathcal{N}_\nu \times \widehat{\mathcal{T}}(\mu,\nu) \quad (3.2.5)$$

with $J_\theta := \mathrm{dom}\theta(X,\cdot) \cap \bar{J}$, for some $I \subset \bar{J} \subset \mathrm{cl}\, I$, characterized $\mu-$a.s. by $\overline{\mathrm{supp}}\mathbb{P}_{X|\partial I(X)} \subset \bar{J}(X) \setminus I(X) = \overline{\mathrm{supp}}\widehat{\mathbb{P}}_{X|\partial I(X)}$, $\mu-$a.s., for some $\widehat{\mathbb{P}} \in \mathcal{M}(\mu,\nu)$, for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. The definition of $\widehat{\mathcal{T}}(\mu,\nu) \subset \mathbb{L}^0_+(\Omega)$ is reported to Subsection 3.5.2. By Remark 3.5 in [58], J_θ is constant on $I(x)$ for all $x \in \mathbb{R}^d$. Then the random variable J_θ is $I-$measurable. Notice as well that by this remark we have

$$I \subset \underline{J} \subset J_\theta \subset \bar{J} \subset \mathrm{cl}\, I, \quad \mu-\text{a.s.}$$

Where $\underline{J}$ is characterized in Proposition 2.4 in [58]. These sets J_θ are very important for characterising the polar sets. However they are not satisfactory as they may not be convex. We extend the notion in next proposition. Let $A \subset \Omega$, we say that A is martingale monotone if for all finitely supported $\mathbb{P} \in \mathcal{P}(\Omega)$, and all competitor $\mathbb{P}'$ to $\mathbb{P}$, $\mathbb{P}[A] = 1$ if and only if $\mathbb{P}'[A] = 1$. Notice that A is martingale monotone if and only if A is $\mathbf{1}_{A^c}-$martingale monotone.

Proposition 3.2.7. *Under Assumption 3.2.6, for any $N-$tangent convex $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$, we may find $\theta \le \theta' \in \widetilde{\mathcal{T}}(\mu,\nu)$ and $(N^0_\mu, N^0_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$ such that for all $(N^0_\mu, N^0_\nu) \subset (N_\mu, N_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$, the maps I, $\underline{J}$, and $\bar{J}$ from [58] may be chosen so that $J(X) := \mathrm{conv}(\mathrm{dom}\theta'(X,\cdot) \setminus N_\nu) \cap \mathrm{aff}\, I(X)$ satisfies, up to a modification on N_μ:*
(i) $J(X) = \mathrm{conv}\Big(J(X) \setminus N_\nu\Big)$*, and on N^c_μ, we have* $J(X) \subset \mathrm{dom}\theta(X,\cdot)$*;*

(ii) *$N \subset N' := \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J(X)\} \in \mathcal{N}_{\mu,\nu}$ and N'^c is martingale monotone;*
(iii) *the set-valued map $J^\circ(x) := \cup_{x' \in J(x)\setminus N_\mu} I(x') \cup (J(x)\setminus N_\nu) \cup \{x\}$ satisfies $\underline{J} \subset J^\circ \subset J \subset \bar{J}$, furthermore J and J° are constant on I(x), for all $x \in \mathbb{R}^d$.*

The proof of Proposition 3.2.7 is reported in Subsection 3.5.4. We denote by $\mathcal{J}(\mu,\nu)$ (resp. $\mathcal{J}^\circ(\mu,\nu)$) the set of these modified set-valued mappings J $\Big($resp. $J^\circ\Big)$ from Proposition 3.2.7.

Remark 3.2.8. *Let $J \in \mathcal{J}(\mu,\nu)$, $N_\nu \in \mathcal{N}_\nu$, and $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$ from Proposition 3.2.7. The following holds for $\widetilde{J} \in \{J, J^\circ, J \setminus N_\nu\}$. Let $x, x' \in \mathbb{R}^d$,*
(i) *$Y \in \widetilde{J}(X)$, $\mathcal{M}(\mu,\nu)-q.s.$;*
(ii) *$\widetilde{J}(x) \cap \widetilde{J}(x') = \operatorname{aff}\left(\widetilde{J}(x) \cap \widetilde{J}(x')\right) \cap \widetilde{J}(x)$;*
(iii) *$J(x) \cap J(x') = \operatorname{conv}\left(\widetilde{J}(x) \cap \widetilde{J}(x')\right)$;*
(iv) *if $I(x') \cap \widetilde{J}(x) \neq \emptyset$, then $\widetilde{J}(x') \subset \widetilde{J}(x)$.*

Remark 3.2.8 will be justified in Subsection 3.5.4. We next introduce a subset of polar sets which play an important role.

Definition 3.2.9. *We say that $N \in \mathcal{N}_{\mu,\nu}$ is canonical if $N = \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J(X)\}$, for some $(N_\mu, N_\nu, J) \in \mathcal{N}_\mu \times \mathcal{N}_\nu \times \mathcal{J}(\mu,\nu)$ from Proposition 3.2.7 for some $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$.*

Theorem 3.2.10. *Under Assumption 3.2.6, an analytic set $N \subset \Omega$ is $\mathcal{M}(\mu,\nu)-$polar if and only if it is contained in a canonical $\mathcal{M}(\mu,\nu)-$polar set.*

The proof of Theorem 3.2.10 is reported in Subsection 3.5.4.

Remark 3.2.11. *For a fixed $x \in \mathbb{R}^d$, even though $J(x)$ is convex for $J \in \mathcal{J}(\mu,\nu)$, it may not be Borel anymore, unlike $J_\theta(x)$ when $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$. The same holds for $J^\circ(x)$, with $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$ or for a canonical polar sets, they may not be Borel but only universally measurable (i.e. $\mathbb{P}-$measurable[2] for all $\mathbb{P} \in \mathcal{P}(\Omega)$). Similar to J_θ for $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, the invariance of $J \in \mathcal{J}(\mu,\nu)$ and $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$ on $I(x)$ for each $x \in \mathbb{R}^d$ proves that J is $I-$measurable.*

3.2.4 Weakly convex functions

We see from [22] 4.2 that the integral of the dual functions needs to be compensated by a convex (concave in [22]) moderator to deal with the case $\mu[\varphi] + \nu[\psi] = -\infty + \infty$.

[2] A set A is said to be $\mathbb{P}-$measurable if $\mathbb{P}\big[(A \cup B) \setminus (A \cap B)\big] = 0$ for some Borel set $B \subset \Omega$.

However, they need to define a new concave moderator for each irreducible component before summing them up on the countable components. In higher dimension, as the components may not be countable there may be measurability issues arising. We need to store all these convex moderators in one single moderator which is convex on each component, but that may not be globally convex (see Example 3.2.14).

Definition 3.2.12. *A function $f:\mathbb{R}^d \longrightarrow \mathbb{R}$ is said to be $\mathcal{M}(\mu,\nu)$-convex or weakly convex if there exists a tangent convex function $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$ such that*

$$\mathbf{T}_p f = \theta, \quad \text{on } \{Y \in J^\circ(X), X \notin N_\mu\}, \quad \text{for some } p:\mathbb{R}^d \to \mathbb{R}^d, \\ \text{and } (N_\mu, J^\circ) \in \mathcal{N}_\mu \times \mathcal{J}^\circ(\mu,\nu).$$

Under these conditions, we write that $\theta \approx \mathbf{T}_p f$. Notice that by Remark 3.2.8, $Y \in J^\circ(X)$, $\mathcal{M}(\mu,\nu)-$q.s., whence $\theta \approx \mathbf{T}_p f$ implies that $\theta = \mathbf{T}_p f$, $\mathcal{M}(\mu,\nu)-$q.s. We denote by $\mathfrak{C}_{\mu,\nu}$ the collection of all $\mathcal{M}(\mu,\nu)$-convex functions. Similarly to convex functions, we introduce a convenient notion of subgradient:

$$\partial^{\mu,\nu} f \quad := \quad \Big\{p:\mathbb{R}^d \longmapsto \mathbb{R}^d : \mathbf{T}_p f \approx \theta \in \widetilde{\mathcal{T}}(\mu,\nu)\Big\},$$

which is by definition non-empty. A key ingredient for all the results of this paper is that the sets $\Theta_{\mu,\nu}$ and $\mathfrak{C}_{\mu,\nu}$ turn out to be in one-to-one relationship.

Proposition 3.2.13. *Under Assumption 3.2.6,*

$$\widetilde{\mathcal{T}}(\mu,\nu) = \{\theta \approx \mathbf{T}_p f, \text{ for some } f \in \mathfrak{C}_{\mu,\nu}, \text{ and } p \in \partial^{\mu,\nu} f\}.$$

The proof of this proposition is reported in Subsection 3.5.6.

Example 3.2.14. *[$\mathcal{M}(\mu,\nu)-$convex function in dimension one] Let $\mu := \frac{1}{2}(\delta_{-1} + \delta_1)$, and $\nu(dy) := \frac{1}{8}\Big(\mathbf{1}_{[-2,2]}(y)dy + \delta_{-2}(dy) + 2\delta_0(dy) + \delta_2(dy)\Big)$. For these measures, one can easily check that the irreducible components from [20], [22], and [58] are given by $I(-1) = (-2,0)$, and $I(1) = (0,2)$, and the associated $\bar{J}$ mapping is given by $\bar{J}(-1) = [-2,0]$, and $\bar{J}(1) = [0,2]$. By Example 3.2.17 in this paper, $f:\mathbb{R} \longrightarrow \mathbb{R}$ is $\mathcal{M}(\mu,\nu)-$convex if it is convex on each irreducible components. See Figure 3.2.*

The next result shows that the weakly convex functions are convex on each irreducible component. Let $\eta := \mu \circ I^{-1}$, and recall that any $J \in \mathcal{J}(\mu,\nu)$ is $I-$measurable by Remark 3.2.11.

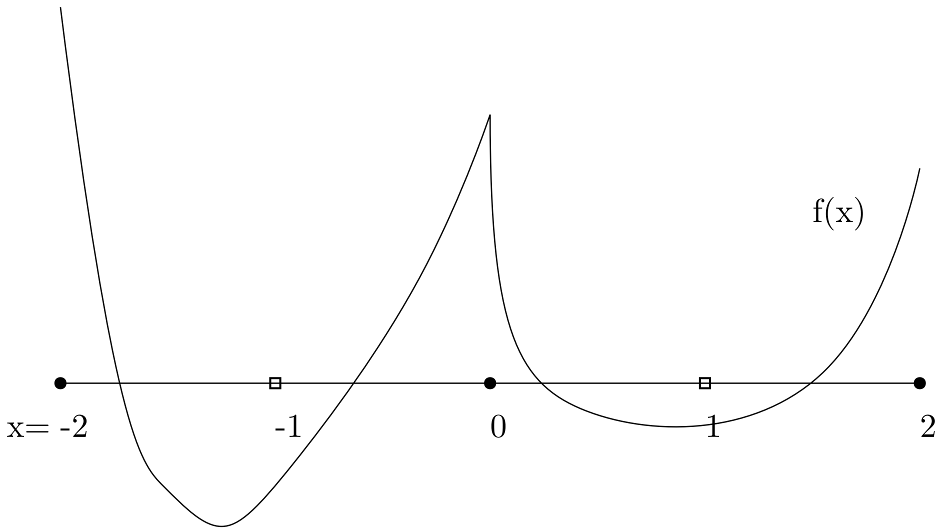


Fig. 3.2 Example of a $\mathcal{M}(\mu,\nu)-$convex function.

Proposition 3.2.15. *Let $f \in \mathfrak{C}_{\mu,\nu}$ and $p \in \partial^{\mu,\nu} f$. Then f is convex on J°, and $\mathrm{proj}_{\nabla \mathrm{aff} J^\circ}(p)(X) \in \partial f|_{J^\circ}(X)$, $\mu-$a.s. Furthermore, we may find $\widetilde{f} \in \mathfrak{C}_{\mu,\nu}$ and $\widetilde{p} \in \partial^{\mu,\nu}\widetilde{f}$ such that $f = \widetilde{f}$, $\mu+\nu-$a.s., $\widetilde{p} = \mathrm{proj}_{\nabla \mathrm{aff} J^\circ}(p)$, $\mu-$a.s., and $\widetilde{f}$ is convex on J with $\widetilde{p} \in \partial \widetilde{f}|_I$, η-a.s. for some $J \in \mathcal{J}(\mu,\nu)$.*

The proof of this proposition is reported in Subsection 3.5.6.

3.2.5 Extended integrals

The following integral is clearly well-defined:

$$(\nu-\mu)[f] = \mathbb{P}[\mathbf{T}_p f] \quad \text{for all} \quad \mathbb{P} \in \mathcal{M}(\mu,\nu),\ f \in \mathfrak{C} \cap \mathbb{L}^1(\nu), p \in \partial f. \qquad (3.2.6)$$

Similar to Beiglböck, Nutz & Touzi [22], we need to introduce a convenient extension of this integral. For $f \in \mathfrak{C}_{\mu,\nu}$, define:

$$\nu \overline{\ominus} \mu[f] \quad := \quad \inf\Big\{a \geq 0 : \mathbf{T}_p f \approx \theta \in \widetilde{\mathcal{T}}_a, \text{ for some } p \in \partial^{\mu,\nu} f\Big\} \qquad (3.2.7)$$

$$\nu \underline{\ominus} \mu[f] := \mathbf{S}_{\mu,\nu}(\mathbf{T}_p f), \quad \text{for} \quad p \in \partial^{\mu,\nu} f, \qquad (3.2.8)$$

where the last value is not impacted by the choice of $p \in \partial^{\mu,\nu} f$, whenever $\nu \underline{\ominus} \mu[f] < \infty$. Indeed, if $p_1, p_2 \in \partial^{\mu,\nu} f$ such that $\mathbb{P}[\mathbf{T}_{p_1} f] < \infty$ and $\mathbb{P}[\mathbf{T}_{p_2} f] < \infty$, then $\mathbf{T}_{p_1} f - \mathbf{T}_{p_2} f = (p_2 - p_1)^{\otimes} \in \mathbb{L}^1(\mathbb{P})$, and it follows from the Fubini theorem that $\mathbb{P}[\mathbf{T}_{p_1} f - \mathbf{T}_{p_2} f] = \mathbb{P}[(p_2-p_1)^{\otimes}] = \mathbb{P}[\mathbb{P}[(p_2-p_1)^{\otimes}|X]] = 0$.

We also abuse notation and define for $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$, $\nu\overline{\ominus}\mu[\theta] := \inf\left\{a \geq 0 : \theta \in \widetilde{\mathcal{T}}_a\right\}$.

Proposition 3.2.16. *For $f \in \mathfrak{C}_{\mu,\nu}$ and $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$, we have*
(i) $\nu\overline{\ominus}\mu[f] \geq \nu\underline{\ominus}\mu[f] \geq 0$*, and* $\nu\overline{\ominus}\mu[\theta] \geq \mathbf{S}_{\mu,\nu}(\theta) \geq 0$*;*
(ii) *if* $f \in \mathfrak{C} \cap \mathbb{L}^1(\nu)$*, then* $\nu\overline{\ominus}\mu[f] = \nu\underline{\ominus}\mu[f] = \nu\overline{\ominus}\mu[\mathbf{T}_p f] = (\nu-\mu)[f]$*, for all* $p \in \partial f$*;*
(iii) $\nu\underline{\ominus}\mu$ *and* $\nu\overline{\ominus}\mu$ *are homogeneous and convex.*

Proof. The proof is similar to the proof of Proposition 2.11 in [58]. □

We can prove the next simple characterization of $\widetilde{\mathcal{T}}(\mu,\nu)$, $\mathcal{C}(\mu,\nu)$ and $\widehat{\mathcal{T}}(\mu,\nu)$ in the one-dimensional setting. In dimension 1, by Beiglböck, Nutz & Touzi [22], there are only countably many irreducible components of full dimension. The other components are points. Then we can write these components I_k for $k \in \mathbb{N}$ like in [22] Proposition 2.3. We also have uniqueness of the $J(x)$ from Theorem 3.7 in [58], that is equivalent in dimension 1 to Theorem 3.2. We denote them J_k as well. We also take another notation from the paper, μ_k and ν_k the restrictions of μ and ν to I_k and J_k, and $(\nu_k - \mu_k)$ extending their Definition 4.2 to non integrable convex functions, which corresponds to the operator $\nu\ominus\mu$ in this paper.

Example 3.2.17. *If $d = 1$,*

$$\begin{aligned}
\mathfrak{C}_{\mu,\nu} &= \left\{f : \mathbb{R}^d \to \mathbb{R} : f_{|J_k} \text{ is convex for all } k\right\},\\
\widetilde{\mathcal{T}}(\mu,\nu) &= \Big\{\theta = \sum_k \mathbf{1}_{X\in I_k}\mathbf{T}_{p_k} f_k : f_k \text{ convex finite on } J_k, p_k \in \partial f_k,\\
&\qquad \text{and } \sum_k (\nu_k - \mu_k)(f_k) < \infty\Big\},\\
&\qquad \text{and}\\
\nu\overline{\ominus}\mu[f] &= \nu\underline{\ominus}\mu[f] = \sum_k (\nu_k - \mu_k)(f_{|J_k}), \quad \text{for all } f \in \mathfrak{C}_{\mu,\nu}.
\end{aligned}$$

This characterization follows from the same argument than the proof of Proposition 3.11 in [58].

3.2.6 Problem formulation

Definition 3.2.18. *Let $\varphi, \psi : \mathbb{R}^d \longrightarrow \mathbb{R}$ and $f \in \mathfrak{C}_{\mu,\nu}$. We say that f is a convex moderator for (φ,ψ) if*

$$\varphi + f \in \mathbb{L}^1_+(\mu), \quad \psi - f \in \mathbb{L}^1_+(\nu), \quad \text{and} \quad \nu\ominus\mu[f] := \nu\overline{\ominus}\mu[f] = \nu\underline{\ominus}\mu[f] < \infty.$$

We denote by $\widehat{\mathbb{L}}(\mu,\nu)$ the collection of triplets (φ,ψ,h) such that (φ,ψ) has some convex moderator f with $h+p\in\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$ for some $p\in\partial^{\mu,\nu}f$.

We now introduce the objective function of the robust superhedging problem for a pair $(\varphi,\psi)\in\widehat{\mathbb{L}}(\mu,\nu)$ with convex moderator f:

$$\mu[\varphi]\oplus\nu[\psi] \;\; := \;\; \mu[\varphi+f]+\nu[\psi-f]+\nu\ominus\mu[f]. \tag{3.2.9}$$

We observe immediately that this definition does not depend on the choice of the convex moderator. Indeed, if f_1,f_2 are two convex moderators for (φ,ψ), it follows that $f_1-f_2\in\mathbb{L}^1(\mu)\cap\mathbb{L}^1(\nu)$, and consequently $\mu\underline{\ominus}\nu[f_1]=\mu\underline{\ominus}\nu[f_2]+(\nu-\mu)[f_1-f_2]$ by Proposition 3.2.16. This implies that

$$\mu[\varphi+f_1]+\nu[\psi-f_1]+\nu\ominus\mu[f_1] \;\; = \;\; \mu[\varphi+f_2]+\nu[\psi-f_2]+\nu\ominus\mu[f_2].$$

For a cost function $c:\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}_+$, the relaxed robust superhedging problem is

$$\mathbf{I}^{qs}_{\mu,\nu}(c) \;\; := \inf_{(\varphi,\psi,h)\in\mathcal{D}^{qs}_{\mu,\nu}(c)} \mu[\varphi]\oplus\nu[\psi], \tag{3.2.10}$$

where

$$\mathcal{D}^{qs}_{\mu,\nu}(c) \;\; := \;\; \Big\{(\varphi,\psi,h)\in\widehat{\mathbb{L}}(\mu,\nu):\; \varphi\oplus\psi+h^{\otimes}\geq c,\; \mathcal{M}(\mu,\nu)-\text{q.s.}\Big\}. \tag{3.2.11}$$

Remark 3.2.19. *This dual problem depends on the primal variables $\mathcal{M}(\mu,\nu)$. However this issue is solved by the fact that Theorem 3.7 in [58] gives an intrinsic description of the polar sets. See also Theorem 3.2.10.*

We also introduce the pointwise version of the robust superhedging problem:

$$\mathbf{I}^{\text{pw}}_{\mu,\nu}(c) \;\; := \inf_{(\varphi,\psi,h)\in\mathcal{D}^{\text{pw}}_{\mu,\nu}(c)} \mu[\varphi]\oplus\nu[\psi], \tag{3.2.12}$$

where

$$\mathcal{D}^{\text{pw}}_{\mu,\nu}(c) \;\; := \;\; \Big\{(\varphi,\psi,h)\in\widehat{\mathbb{L}}(\mu,\nu):\; \varphi\oplus\psi+h^{\otimes}\geq c\Big\}. \tag{3.2.13}$$

The following inequalities extending the classical weak duality (4.1.5) are immediate,

$$\mathbf{S}_{\mu,\nu}(c)\leq\mathbf{I}^{qs}_{\mu,\nu}(c)\leq\mathbf{I}^{\text{pw}}_{\mu,\nu}(c). \tag{3.2.14}$$

3.3 Main results

Remark 3.3.1. *All the results in this section are given for $c \geq 0$. The extension to the case $c \geq \varphi_0 \oplus \psi_0 + h_0^{\otimes}$ with $(\varphi_0, \psi_0, h_0) \in \mathbb{L}^1(\mu) \times \mathbb{L}^1(\nu) \times \mathbb{L}^1(\mu, \mathbb{R}^d)$, is immediate by applying all results to $c - \varphi_0 \oplus \psi_0 - h_0^{\otimes} \geq 0$.*

3.3.1 Duality and attainability

We recall that an upper semianalytic function is a function $f : \mathbb{R}^d \to \mathbb{R}$ such that $\{f \geq a\}$ is an analytic set for any $a \in \mathbb{R}$. In particular, a Borel function is upper semianalytic.

Theorem 3.3.2. *Let $c : \Omega \to \overline{\mathbb{R}}_+$ be upper semianalytic. Then, under Assumption 3.2.6, we have*
(i) $\mathbf{S}_{\mu,\nu}(c) = \mathbf{I}^{qs}_{\mu,\nu}(c)$;
(ii) *If in addition $\mathbf{S}_{\mu,\nu}(c) < \infty$, then existence holds for the quasi-sure dual problem $\mathbf{I}^{qs}_{\mu,\nu}(c)$.*

This Theorem will be proved in Subsection 3.5.3.

Remark 3.3.3. For an upper-semicontinuous coupling function c, we observe that the duality result $\mathbf{S}_{\mu,\nu}(c) = \mathbf{I}^{qs}_{\mu,\nu}(c) = \mathbf{I}^{pw}_{\mu,\nu}(c)$ holds true, together with the existence of an optimal martingale interpolating measure for the martingale optimal transport problem $\mathbf{S}_{\mu,\nu}(c)$, without any need to Assumptions 3.2.6. This is an immediate extension of the result of Beiglböck, Henry-Labordère & Penckner [18], see also Zaev [160]. However, dual optimizers may not exist in general, see the counterexamples in Beiglböck, Henry-Labordère & Penckner and in Beiglböck, Nutz & Touzi [22]. Observe that in the one-dimensional case, Beiglböck, Lim & Obłój [21] proved that pointwise duality, and integrability hold for C^2 cost functions together with compactly supported μ, and ν. We show in Example 3.4.1 below that this result does not extend to higher dimension.

Remark 3.3.4. An existence result for the robust superhedging problem was proved by Ghoussoub, Kim & Lim [74]. We emphasize that their existence result requires strong regularity conditions on the coupling function c and duality, and is specific to each component of the decomposition in irreducible convex pavings, see Subsection 3.3.2 below. In particular, their construction does not allow for a global existence result because of non-trivial measurability issues. Our existence result in Theorem 3.3.2 (ii) by-passes these technical problems, provides global existence of a dual optimizer, and does not require any regularity of the cost function c.

3.3.2 Decomposition on the irreducible convex paving

The measurability of the map I stated in Theorem 2.1 (i) in [58], induces a decomposition of any function on the irreducible paving by conditioning on I. We shall denote $\eta := \mu \circ I^{-1}$, and set μ_I the law of X, conditionally to I. Then for any measurable $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, non-negative or $\mu-$integrable, we have $\int_{\mathbb{R}^d} f(x)\mu(dx) = \int_{I(\mathbb{R}^d)} \left(\int_i f(x)\mu_i(dx)\right)\eta(di)$.

Similar to the one-dimensional context of Beiglböck, Nutz & Touzi [22], it turns out that the martingale transport problem reduces to componentwise irreducible martingale transport problems for which the quasi-sure formulation and the pointwise one are equivalent. For $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, we shall denote $\nu_I^{\mathbb{P}} := \mathbb{P} \circ (Y|X \in I)^{-1}$ and $\mathbb{P}_I := \mathbb{P} \circ ((X,Y)|X \in I)^{-1}$.

Theorem 3.3.5. *Let $c : \Omega \to \overline{\mathbb{R}}_+$ be upper semianalytic with $\mathbf{S}_{\mu,\nu}(c) < \infty$. Then we have:*

$$\mathbf{S}_{\mu,\nu}(c) = \sup_{\mathbb{P} \in \mathcal{M}(\mu,\nu)} \int_{I(\mathbb{R}^d)} \mathbf{S}_{\mu_i,\nu_i^{\mathbb{P}}}(c)\eta(di). \tag{3.3.1}$$

Furthermore, we may find functions $(\varphi, h) \in \mathbb{L}^0(\mathbb{R}^d) \times \mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$, and $(\psi_K)_{K \in I(\mathbb{R}^d)} \subset \mathbb{L}^0_+(\mathbb{R}^d)$ with $\psi_{I(X)}(Y) \in \mathbb{L}^0_+(\Omega)$, and $\mathrm{dom}\,\psi_I = J_\theta$, $\eta-$a.s. for some $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, such that
(i) *$c \le \bar{c} := \varphi(X) + \psi_{I(X)}(Y) + h^\otimes$, and $\mathbf{S}_{\mu,\nu}(c) = \mathbf{S}_{\mu,\nu}\Big(\bar{c}\Big)$.*
(ii) *If the supremum* (3.3.1) *has an optimizer $\mathbb{P}^* \in \mathcal{M}(\mu,\nu)$, then we may chose $\Big(\varphi, h, (\psi_K)_K\Big)$ so that $(\varphi, \psi_I, h) \in \mathcal{D}^{pw}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c_{|I\times J_\theta})$, and*
$\mathbf{S}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c) = \mathbf{I}^{\mathrm{pw}}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c) = \mu_I[\varphi] \oplus \nu_I^{\mathbb{P}^*}[\psi_I]$, *$\eta-a.s.$*
(iii) *If Assumption 3.2.6 holds, we may find $J \in \mathcal{J}(\mu,\nu)$, and $(\varphi',\psi',h') \in \mathcal{D}^{qs}_{\mu,\nu}(c)$ optimizer for $\mathbf{I}^{qs}_{\mu,\nu}(c)$ such that $c \le \varphi' \oplus \psi' + h'^\otimes$, on $\{Y \in J(X)\}$.*
(iv) *Under the conditions of* (ii) *and* (iii), *we may find $(\varphi',\psi',h') \in \mathcal{D}^{pw}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c_{|I\times J})$, such that $\mathbf{S}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c) = \mu_I[\varphi'] \oplus \nu_I^{\mathbb{P}^*}[\psi']$, $\eta-a.s.$*

Theorem 3.3.5 will be proved in Subsection 3.5.5

Remark 3.3.6. *Notice that $(\mu_I, \nu_I^{\mathbb{P}^*})$ may not be irreducible. See Example 3.4.2. This is an important departure from the one-dimensional case.*

Remark 3.3.7. *Existence holds for the maximization problem* (3.3.1) *(and therefore (ii) in Theorem 3.3.5 holds) under any of the following assumptions:*
(i) *$\nu_I := \nu_I^{\mathbb{P}}$ is independent of $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ (see Remark 3.3.12 for some sufficient conditions);*

(ii) *There exists a primal optimizer for the problem* $\mathbf{S}_{\mu,\nu}(c)$.

3.3.3 Martingale monotonicity principle

As a consequence of the last duality result, we now provide the martingale version of the monotonicity principle which extends the corresponding result in standard optimal transport theory, see Theorem 5.10 in Villani [157]. The following monotonicity principle states that the optimality of a martingale measure reduces to a property of the corresponding support.

The one-dimensional martingale monotonicity principle was introduced by Beiglböck & Juillet [20], see also Zaev [160], and Beiglböck, Nutz & Touzi [22].

Theorem 3.3.8. *Let* $c:\Omega\to\overline{\mathbb{R}}_+$ *be upper semianalytic with* $\mathbf{S}_{\mu,\nu}(c)<\infty$.
(i) *Then we may find a Borel set* $\Gamma\subset\Omega$ *such that:*

(a) *Any solution* $\mathbb{P}$ *of* $\mathbf{S}_{\mu,\nu}(c)$, *is concentrated on* Γ;

(b) *we may find* $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$ *and* $(\Gamma_K)_{K\in I(\mathbb{R}^d)}$ *such that* $\Gamma=\cup_{K\in I(\mathbb{R}^d)}\Gamma_K$ *with* $\Gamma_I\subset I\times J_\theta$, Γ_I *is c-martingale monotone, and for any optimizer* $\mathbb{P}^*$ *of* $\mathbf{S}_{\mu,\nu}(c)$, *we have that any optimizer* $\mathbb{P}\in\mathcal{M}(\mu_I,\nu_I^{\mathbb{P}^*})$ *of* $\mathbf{S}_{\mu_I,\nu_I^{\mathbb{P}^*}}(c)$, *is concentrated on* Γ_I.
(ii) *if Assumption 3.2.6 holds, we may find a universally measurable* $\Gamma'\subset N^c$, *for some canonical* $N\in\mathcal{N}_{\mu,\nu}$, *satisfying* (a) *and* (b), *such that* Γ' *is c-martingale monotone.*

Proof. Let functions $(\varphi,h)\in\mathbb{L}^0(\mathbb{R}^d)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$ and functions $(\psi_K)_{K\in I(\mathbb{R}^d)}\subset\mathbb{L}^0_+(\mathbb{R}^d)$ with $\psi_{I(X)}(Y)\in\mathbb{L}^0_+(\Omega)$ from Theorem 3.3.5. Recall that pointwise we have $c\le\varphi(X)+\psi_{I(X)}(Y)+h^\otimes$. We set $\Gamma:=\{c=\varphi(X)+\psi_{I(X)}(Y)+h^\otimes<\infty\}$.

(i) If $\mathbb{P}^*$ is optimal for the primal problem then,

$$\infty>\mathbb{P}^*[c]=\mathbb{P}^*[\varphi(X)+\psi_{I(X)}(Y)+h^\otimes]=\mathbf{S}_{\mu,\nu}(c)\quad\textit{and}\quad\mathbb{P}^*[\varphi(X)+\psi_{I(X)}(Y)+h^\otimes-c]=0$$

As $\varphi(X)+\psi_{I(X)}(Y)+h^\otimes-c\ge 0$, and the expectation of c is finite, and therefore $\mathbb{P}^*[c<\infty]=1$, it follows that $\mathbb{P}^*$ is concentrated on Γ.

(ii) Let $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$ such that $J_\theta=\mathrm{dom}\psi_{I(X)}$ from Theorem 3.3.5. For $K\in I(\mathbb{R}^d)$, let $\Gamma_K:=\Gamma\cap K\times\mathbb{R}^d$. Then we have $\Gamma_{I(x)}\subset I(x)\times J_\theta(x)$ for all $x\in\mathbb{R}^d$, $\Gamma_{I(x)}$ is c-martingale monotone because of the pointwise duality on each component, and $\Gamma=\cup_{x\in\mathbb{R}^d}\Gamma_{I(x)}$ by definition because $I(\mathbb{R}^d)$ is a partition of $\mathbb{R}^d$.

If Assumption 3.2.6 holds, we consider $(\varphi',\psi',h')\in\mathcal{D}^{qs}_{\mu,\nu}(c)$ from the second part of Theorem 3.3.5. Let a canonical $N\in\mathcal{N}_{\mu,\nu}$ be such that $c=\varphi'\oplus\psi'+\bar{h}'^\otimes$ on N^c. $\Gamma:=N^c\cap\{c=\varphi'\oplus\psi'+\bar{h}'^\otimes\}$. Similarly, (i) and (ii) hold.

(iii) By definition of $\Theta_{\mu,\nu}$, for $\mathbb{P}_0$ with finite support, supported on $\Gamma \subset N^c$, and $\mathbb{P}'$ competitor to $\mathbb{P}_0$. As N^c is canonical, it is martingale monotone by definition. Then $\mathbb{P}'[N^c] = 1$, and therefore $\mathbb{P}'[c] \leq \mathbb{P}'[\varphi' \oplus \psi' + h'^{\otimes}] = \mathbb{P}[\varphi' \oplus \psi' + h'^{\otimes}] = \mathbb{P}_0[c]$.

Finally, by definition we have $\Gamma \subset N^c$. □

Remark 3.3.9. *Let $(\varphi, \psi, h) \in \mathcal{D}^{qs}_{\mu,\nu}(c)$ be a minimizer of $\mathbf{I}^{q.s.}_{\mu,\nu}(c)$. Assume that $\mathbb{P}[\varphi \oplus \psi + h^{\otimes}]$ does not depend on the choice of $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ (e.g. if $(\varphi,\psi) \in \mathbb{L}^1(\mu) \times \mathbb{L}^1(\nu)$, or if $d = 1$). Then we may chose Γ such that a measure $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ is optimal for $\mathbf{S}_{\mu,\nu}(c)$ if and only if it is concentrated on Γ. Indeed, with the notations from the previous proof, if $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ is concentrated on Γ, $\mathbb{P}[\varphi \oplus \psi + h^{\otimes} - c] = 0$ and as $\mathbb{P}[\varphi \oplus \psi + h^{\otimes}] = \mu[\varphi] \oplus \nu[\psi]$ because of the invariance,*

$$\mathbb{P}(c) = \mathbb{P}[\varphi \oplus \psi + h^{\otimes}] = \mathbf{I}^{qs}_{\mu,\nu}(c) = \mathbf{S}_{\mu,\nu}(c).$$

3.3.4 On Assumption 3.2.6

Proposition 3.3.10. *Assumption 3.2.6 holds true under either one of the following conditions:*
(i) *$Y \notin \partial I(X)$, $\mathcal{M}(\mu,\nu)$-q.s. or equivalently $\mu \circ I^{-1} = \nu \circ I^{-1}$.*
(ii) $\dim I(X) \in \{0,1,d\}$, *$\mu-$a.s.*
(iii) *ν is dominated by the Lebesgue measure and* $\dim I(X) \in \{0,1,d-1,d\}$, *$\mu-$a.s.*
(iv) *$I(X) \in \mathcal{C} \cup \mathcal{D} \cup \mathcal{R}$, $\mu-$a.s. for some subsets $\mathcal{C},\mathcal{D},\mathcal{R} \subset \overline{\mathcal{K}}$ with $\mathcal{C}$ countable,* $\dim(\mathcal{D}) \subset \{0,1\}$, *and $\cup_{K \in \mathcal{R}} K \times \partial K \in \mathcal{N}_{\mu,\nu}$.*
Furthermore, (iv) *is implied by either one of* (i), (ii), *and* (iii).

This proposition is proved in Subsection 3.6.1.

Remark 3.3.11. *Assumption 3.2.6 holds in dimension* 1 *by Proposition 3.3.10. Theorem 3.3.2 is equivalent to [22] Theorem 7.4 and the monotonicity principle Theorem 3.3.8 is equivalent to [22] Corollary 7.8.*

Remark 3.3.12. *Notice that under either one of (i) or (iii) of Proposition 3.3.10, or in dimension one, the disintegration $\nu^{\mathbb{P}}_I := \mathbb{P} \circ (Y|X \in I)^{-1}$ is independent of the choice of $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. See Subsection 3.6.1 for a justification of this claim.*

Remark 3.3.13. *Proposition 3.3.10 may be applied in particular in the trivial case in which there is a unique irreducible component. We state here that any pair of measures $\mu,\nu \in \mathbb{P}(\mathbb{R}^d)$ in convex order may be approximated by pairs of measures that have a unique irreducible component, and therefore satisfy Assumption 3.2.6. We may then* *use a stability result like in Guo & Obłój [77] to use the approximation* $(\mu_\epsilon,\nu_\epsilon)$ *of* (μ,ν) *in practice.*

Let $\mu'\preceq\nu'$ *in convex order with* (μ',ν') *irreducible, and* $\mathrm{supp}\,\nu\subset\mathrm{ri\,conv\,supp}\,\nu'$. *Then* $(\mu_\epsilon,\nu_\epsilon):=\frac{1}{1+\epsilon}(\mu+\epsilon\mu',\nu+\epsilon\nu')$ *is irreducible for all* $\epsilon>0$. *Indeed by Proposition 3.4 in [58], we may find* $\hat{\mathbb{P}}\in\mathcal{M}(\mu',\nu')$ *such that* $\mathrm{conv\,supp}\,\hat{\mathbb{P}}_X=\mathrm{ri\,conv\,supp}\,\nu'$, $\mu'-a.s.$ *Then,* $\frac{1}{1+\epsilon}(\mathbb{P}+\varepsilon\hat{\mathbb{P}})\in\mathcal{M}(\mu_\epsilon,\nu_\epsilon)$ *for all* $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, *and* $\mathrm{ri\,conv\,supp}\,\nu'\subset I(X)$ *on a set charged by* μ_ϵ, *which proves that* $I=\mathrm{ri\,conv\,supp}\,\nu'\supset\mathrm{supp}\,\nu$, *preventing other components from appearing on the boundary. Thus* $(\mu_\epsilon,\nu_\epsilon)$ *is irreducible.*

Convenient measures to consider are for example $\mu':=\delta_0$ *or* $\mu':=\mathcal{N}(0,1)$, *and* $\nu':=\mathcal{N}(0,2)$. *For finitely supported* μ *and* ν *we may consider* $y_1,...,y_k\in\mathbb{R}^d$ *for some* $k\geq1$ *such that* $\mathrm{supp}\,\nu\subset\mathrm{int\,conv}(y_1,...,y_k)$, $\nu':=\frac{\delta_{y_1}+...+\delta_{y_k}}{n}$, *and* $\mu':=\delta_{\frac{y_1+...+y_n}{n}}$.

Proposition 3.3.14. *Assumption 3.2.6 holds if we assume existence of medial limits and Axiom of choice for* $\mathbb{R}$.

We prove this Proposition in Subsection 3.6.2.

Remark 3.3.15. *Notice that existence of medial limits and Axiom of choice for* $\mathbb{R}$ *is implied by Martin's axiom and Axiom of choice for* $\mathbb{R}$, *which is implied by the continuum hypothesis. Furthermore, all these axiom groups are undecidable under either the Theory ZF nor the Theory ZFC. See Subsection 3.6.2.*

3.3.5 Measurability and regularity of the dual functions

In the main theorem, only $\varphi\oplus\psi+h^\otimes$ has some measurability. However, we may get some measurability on the separated dual optimizers.

Proposition 3.3.16. *For all* $(\varphi,\psi,h)\in\widehat{\mathbb{L}}(\mu,\nu)$,
(i) $\big(\varphi,\psi,proj_{\nabla\mathrm{aff}I}(h)\big)\in\mathbb{L}^0(I)\times\mathbb{L}^0(I)\times\mathbb{L}^0(I,\nabla\mathrm{aff}I)$;
(ii) *under any one of the conditions of Proposition 3.3.10, we may find* $(\varphi',\psi',h')\in\widehat{\mathbb{L}}(\mu,\nu)$ *such that* $\varphi\oplus\psi+h^\otimes=\varphi'\oplus\psi'+h'^\otimes$, *q.s. and* $(\varphi',\psi',h')\in\mathbb{L}^0\big(\mathbb{R}^d\big)^2\times\mathbb{L}^0\big(\mathbb{R}^d,\mathbb{R}^d\big)$. *Furthermore, the canonical set from Theorem 3.2.10, and the set* Γ' *from Theorem 3.3.8 may be chosen to be Borel measurable, and* $\{Y\in J(X)\}$ *(resp.* $\{Y\in J^\circ(X)\}$*) for* $J\in\mathcal{J}(\mu,\nu)$ *(resp.* $J^\circ\in\mathcal{J}^\circ(\mu,\nu)$*) may be chosen to be analytically measurable.*

The proof of this proposition is reported to Subsection 3.5.6. We may as well prove some regularity of the dual functions, provided that the cost function has some appropriate regularity. This Lemma is very close to Theorem 2.3 (1) in [74].

Lemma 3.3.17. *Let $c : \Omega \to \overline{\mathbb{R}}_+$ be upper semi-analytic. We assume that $x \longmapsto c(x,y)$ is locally Lipschitz in x, uniformly in y, and that $\mathbf{S}_{\mu,\nu}(c) = \mathbf{S}_{\mu,\nu}(\varphi \oplus \psi + h^\otimes) < \infty$, with $\varphi : \mathbb{R}^d \longmapsto \mathbb{R} \cup \{\infty\}$, $\psi : \mathbb{R}^d \longmapsto \mathbb{R} \cup \{\infty\}$, and $h : \mathbb{R}^d \longmapsto \mathbb{R}^d$ such that $c \le \varphi \oplus \psi + h^\otimes$, pointwise. Then, we may find $(\varphi', h') = (\varphi, h)$, $\mu-$a.e. such that $c \le \varphi' \oplus \psi + h'^\otimes \le \varphi \oplus \psi + h'^\otimes$, φ' is locally Lipschitz, and h' is locally bounded on* $\mathrm{ri}\,\mathrm{conv}\,\mathrm{dom}\,\psi$.

The proof of Lemma 3.3.17 is reported in Subsection 3.5.7.

3.4 Examples

3.4.1 Pointwise duality failing in higher dimension

In the one-dimensional case, Beiglböck, Lim & Obłój [21] proved that pointwise duality, and integrability hold for C^2 cost functions together with compactly supported μ, and ν. We believe that integrability may hold in higher dimension, and strong monotonicity holds. However the following example shows that dual attainability does not hold with such generality for a dimension higher than 2.

Example 3.4.1. *Let $y_{--} := (-1,-1)$, $y_{-+} := (-1,1)$, $y_{+-} := (1,-1)$, $y_{++} := (1,1)$, $y_{0-} := (0,-1)$, $y_{0+} := (0,1)$, $y_{00} := (0,0)$, $y_{+0} := (1,0)$, $C := \mathrm{conv}(y_{--}, y_{-+}, y_{+-}, y_{++})$, $x_1 := (-\frac{1}{2}, 0)$, $x_2 := (\frac{1}{2}, \frac{1}{2})$, $x_3 := (\frac{1}{2}, -\frac{1}{2})$, $\mu := \frac{1}{2}\delta_{x_1} + \frac{1}{4}\delta_{x_2} + \frac{1}{4}\delta_{x_3}$, and $\nu := \frac{1}{4}\mathbf{1}_C\,Vol$. We can prove that for these marginals, the irreducible components are given by*

$$I(x_1) := \mathrm{ri}\,\mathrm{conv}(y_{--}, y_{-+}, y_{0+}, y_{0-}), \quad I(x_2) := \mathrm{ri}\,\mathrm{conv}(y_{0+}, y_{++}, y_{+0}, y_{00}),$$
$$\text{and} \quad I(x_3) := \mathrm{ri}\,\mathrm{conv}(y_{00}, y_{+0}, y_{+-}, y_{0-}),$$

and $\mathcal{M}(\mu,\nu)$ is a singleton $\{\mathbb{P}\}$, with

$$\mathbb{P}(dx,dy) := \frac{1}{4}\Big(2\delta_{x_1}(dx)\mathbf{1}_{y\in I(x_1)} + \delta_{x_2}(dx)\mathbf{1}_{y\in I(x_2)} + \delta_{x_3}(dx)\mathbf{1}_{y\in I(x_3)}\Big) \otimes Vol(dy).$$

Now we define a cost function c such that $c\Big(x_1,\cdot\Big)$ is 0 on $\mathrm{cl}\,I(x_1)$, $c\Big(x_2,\cdot\Big)$ is 0 on $\mathrm{cl}\,I(x_2)$, and $c\Big(x_3,\cdot\Big)$ is 0 on $\mathrm{cl}\,I(x_3)$. However we also require $c(x_2, y_{+-}) = 1$. We may have these conditions satisfied with $c \ge 0$, and C^∞. Let (φ, ψ, h) be pointwise dual optimizers, then $\varphi \oplus \psi + h^\otimes = c$, $\mathbb{P}-$a.s. then ψ is affine on each irreducible components: $\psi(y) = c(x_i, y) - \varphi(x_i) - h(x_1)\cdot(y - x_i) = -\varphi(x_i) - h(x_1)\cdot(y-x_i)$, Lebesgue-a.e. on $I(x_i)$, for $i = 1,2,3$. By the last equality, we deduce that $\varphi(x_i) = -\psi(x_i)$, and $h(x_i) = -\nabla\psi(x_i)$. Now by the superhedging inequality, $\psi(y) - \psi(x_i) - \nabla\psi(x_i)\cdot(y - x_i) \ge c(x_i,y) \ge 0$. Therefore ψ is a.e. equal to a convex function, piecewise affine on the components.

However a convex function that is affine on $I(x_1)$, $I(x_2)$, and $I(x_3)$ is affine on $\mathrm{cl}\,I(x_2)\cup\mathrm{cl}\,I(x_3)$ *(it follows from the verification at the angles between the regions where ψ has nonzero curvature). Then $c(x_2,y)\le\psi(y)-\psi(x_2)-\nabla\psi(x_2)=0$ for a.e. $y\in\mathrm{cl}\,I(x_3)\subset\mathrm{cl}\,I(x_2)\cup\mathrm{cl}\,I(x_3)$. This is the required contradiction as $c(x_2,y_{+-})=1$ and c is continuous, and therefore nonzero on a non-negligible neighborhood of (x_2,y_{+-}).*

Notice that in this example, μ is not dominated by the Lebesgue measure for simplicity, however this example also holds when δ_{x_i} is replaced by $\frac{1}{\pi\epsilon^2}\mathbf{1}_{B_\epsilon(x_i)}Vol$ for $\epsilon>0$ small enough.

3.4.2 Disintegration on an irreducible component is not irreducible

Example 3.4.2. *Let $x_0:=(-1,0)$, $x_1:=(\frac12,\frac12)$, $x_{-1}:=(\frac12,-\frac12)$, $y_1=(0,1)$, $y_2:=(2,0)$, $y_{-1}:=-y_1$, $y_{-2}:=-y_2$, and $y_0:=0$. Let the probabilities*

$$\mu:=\frac13(\delta_{x_0}+\delta_{x_1}+\delta_{x_{-1}}),\quad \textit{and}\quad \nu:=\frac16(\delta_{y_{-2}}+\delta_{y_2}+\delta_{y_0})+\frac14(\delta_{y_1}+\delta_{y_{-1}}).$$

We can prove that for these marginals, the irreducible components are given by

$$I(x_0)=\mathrm{ri}\,\mathrm{conv}(y_{-2},y_1,y_{-1}),\quad \textit{and } I(x_1)=I(x_{-1})=\mathrm{ri}\,\mathrm{conv}(y_2,y_1,y_{-1}),$$

indeed, $\mathcal{M}(\mu,\nu)=\mathrm{conv}(\mathbb{P}_1,\mathbb{P}_2)$, with

$$\begin{aligned}\mathbb{P}_1:=\frac16\delta_{(x_0,y_{-2})}+\frac16\delta_{(x_0,y_0)} &+ \frac1{12}\delta_{(x_1,y_2)}+\frac3{16}\delta_{(x_1,y_1)}+\frac1{16}\delta_{(x_1,y_{-1})}\\ &+\frac1{12}\delta_{(x_{-1},y_2)}+\frac3{16}\delta_{(x_{-1},y_{-1})}+\frac1{16}\delta_{(x_{-1},y_1)},\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}_2:=\frac16\delta_{(x_0,y_{-2})}+\frac1{12}\delta_{(x_0,y_1)}+\frac1{12}\delta_{(x_0,y_{-1})} &+ \frac16\delta_{(x_1,y_1)}+\frac1{12}\delta_{(x_1,y_0)}+\frac1{12}\delta_{(x_1,y_2)}\\ &+\frac16\delta_{(x_{-1},y_{-1})}+\frac1{12}\delta_{(x_{-1},y_0)}+\frac1{12}\delta_{(x_{-1},y_2)}.\end{aligned}$$

(See Figure 3.3). Let c be smooth, equal to 1 in the neighborhood of (x_0,y_1) and 0 at a distance higher than $\frac12$ from this point, $\mathbb{P}_2$ is the only optimizer for the martingale optimal transport problem $\mathbf{S}_{\mu,\nu}(c)$. However, $\mu_{I(x_1)}=\frac12(\delta_{x_1}+\delta_{x_{-1}})$, and $\nu^{\mathbb{P}_2}_{I(x_1)}=$

$\frac{1}{4}(\delta_{y_2} + \delta_{y_0} + \delta_{y_1} + \delta_{y_{-1}})$, *and the associated irreducible components are*

$$I_{\mu_{I(x_1)},\nu^{\mathbb{P}_2}_{I(x_1)}}(x_1) = \mathrm{ri\,conv}(y_0, y_1, y_2), \quad \textit{and} \quad I_{\mu_{I(x_1)},\nu^{\mathbb{P}_2}_{I(x_1)}}(x_{-1}) = \mathrm{ri\,conv}(y_0, y_{-1}, y_2),$$

and therefore, the couple $\left(\mu_{I(x_1)}, \nu^{\mathbb{P}_2}_{I(x_1)}\right)$ *obtained from the disintegration of the optimal probability* $\mathbb{P}_2$ *in the irreducible component* $I(x_1) = I_1$ *can be reduced again in two irreducible sub-components.*

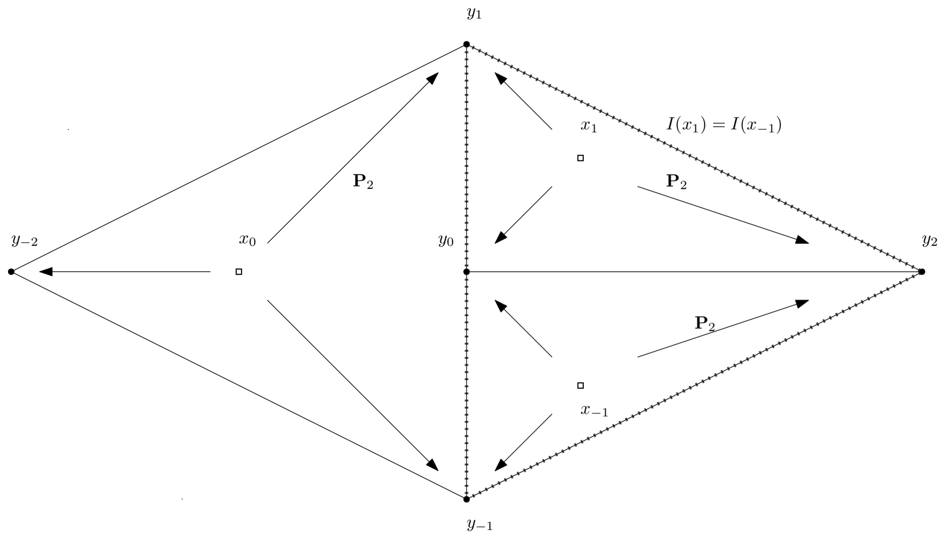


Fig. 3.3 Disintegration on an irreducible component is not irreducible.

3.4.3 Coupling by elliptic diffusion

Assumption 3.2.6 holds when ν is obtained from an Elliptic diffusion from μ.

Remark 3.4.3. *Notice that* (iii) *in Proposition 3.3.10 holds if* ν *is the law of* $X_\tau := X_0 + \int_0^t \sigma_s dW_s$, *where* $X_0 \sim \mu$, W *a* $d-$*dimensional Brownian motion independent of* X_0, τ *is a positive bounded stopping time, and* $(\sigma_t)_{t\geq 0}$ *is a bounded cadlag process with values in* $\mathcal{M}_d(\mathbb{R})$ *adapted to the* $W-$*filtration with* σ_0 *invertible. We observe that the strict positivity of the stopping time is essential, see Example 3.4.4.*

We justify Remark 3.4.3 in Subsection 3.6.1.

Example 3.4.4. *Let* $C := [-1,1] \times [0,2] \times [-1,1]$, $F := \{0\} \times [-1,1] \times [-1,1]$, $x_0 := (0,0,0)$, $x_1 := (0,1,0)$ $\mu := \frac{1}{2}\delta_{x_0} + \frac{1}{2}\delta_{x_1}$, *a* $\mathcal{F}-$*Brownian motion* W, *and* X *a random variable* $\mathcal{F}_0-$*measurable with* $X_0 \sim \mu$. *Consider the bounded stopping time* $\tau := 1 \wedge \inf\{t \geq 0 : W_t \in \partial C\}$, *and* ν, *the law of* $X_0 + W_\tau$. *We have* $\mu \preceq \nu$ *in convex order, as* *the law $\mathbb{P}$ of $(X,Y) := (X_0, X_0 + W_\tau)$ is clearly a martingale coupling. However, observe that $p := \mathbb{P}[X = x_1, Y \in C] > 0$, and that by symmetry $\mathbb{P}[Y|X = x_1, Y \in C] = x_0$. Let ν_C be the law of Y, conditioned on $\{X = x_1, Y \in C\}$. Then $\mathbb{P}' := \mathbb{P} + p\Big((\delta_{x_0} - \delta x_1)\otimes \nu_C - (\delta_{x_0} - \delta x_1)\otimes \delta_{x_0}\Big)$ is also in $\mathcal{M}(\mu,\nu)$. We may prove that the irreducible components are* ri$C$, *and* ri$F$, *and therefore (iii) of Proposition 3.3.10 does not hold. This proves the importance of the strict positivity of the stopping time τ in Remark 3.4.3. In dimension* 4, *we may find an example in which (v) of Proposition 3.3.10 does not hold either, by replacing F by a continuum of translated F in the fourth variable, thus introducing an orthogonal curvature in the lower face of C to avoid the copies of F to communicate with each other.*

3.5 Proof of the main results

3.5.1 Moderated duality

Let $c \geq 0$, we define the moderated dual set of c by

$$\widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c) \;:= \Big\{ \; (\bar\varphi, \bar\psi, \bar h, \theta) \in \mathbb{L}^1_+(\mu)\times\mathbb{L}^1_+(\nu)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)\times\widetilde{\mathcal{T}}(\mu,\nu) : \\ c \leq \bar\varphi \oplus \bar\psi + \bar h^{\otimes} + \theta, \text{ on } \{Y \in \operatorname{aff}\operatorname{rf}_X \operatorname{conv}\operatorname{dom}(\theta + \bar\psi)\}\Big\}.$$

We then define for $(\bar\varphi, \bar\psi, \bar h, \theta) \in \widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)$, $Val(\bar\varphi, \bar\psi, \bar h, \theta) := \mu[\bar\varphi] + \nu[\bar\psi] + \nu\widetilde{\ominus}\nu[\theta]$, and the moderated dual problem $\widetilde{\mathbf{I}^{mod}_{\mu,\nu}}(c) := \inf_{\xi \in \widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)} Val(\xi)$.

Theorem 3.5.1. *Let $c : \Omega \to \overline{\mathbb{R}}_+$ be upper semianalytic. Then, under Assumption 3.2.6, we have*
(i) $\mathbf{S}_{\mu,\nu}(c) = \widetilde{\mathbf{I}^{mod}_{\mu,\nu}}(c)$;
(ii) *If in addition $\mathbf{S}_{\mu,\nu}(c) < \infty$, then existence holds for the moderated dual problem $\widetilde{\mathbf{I}^{mod}_{\mu,\nu}}(c)$.*

This Theorem will be proved in Subsection 3.5.3.

3.5.2 Definitions

We first need to recall some concepts from [58]. For a subset $A \subset \mathbb{R}^d$ and $a \in \mathbb{R}^d$, we introduce the face of A relative to a (also denoted $a-$relative face of A): $\operatorname{rf}_a A := \Big\{ y \in$

$A:(a-\varepsilon(y-a),y+\varepsilon(y-a))\subset A,$ for some $\varepsilon>0\Big\}$. Now denote for all $\theta:\Omega\to\bar{\mathbb{R}}$:

$$\mathrm{dom}_x\theta \;\; := \;\; \mathrm{rf}_x\mathrm{conv}\,\mathrm{dom}\,\theta(x,\cdot).$$

For $\theta_1,\theta_2:\Omega\longrightarrow\mathbb{R}$, we say that $\theta_1=\theta_2$, $\mu\otimes\mathrm{pw}$, if

$$\mathrm{dom}_X\theta_1=\mathrm{dom}_X\theta_2, \;\; \text{and} \;\; \theta_1(X,\cdot)=\theta_2(X,\cdot) \text{ on } \mathrm{dom}_X\theta_1, \; \mu-\text{a.s.}$$

The main ingredient for our extension is the following.

Definition 3.5.2. *A measurable function $\theta:\Omega\to\overline{\mathbb{R}}_+$ is a tangent convex function if*

$$\theta(x,\cdot) \textit{ is convex, and } \theta(x,x)=0, \; \textit{for all } x\in\mathbb{R}^d.$$

We denote by Θ the set of tangent convex functions, and we define

$$\Theta_\mu \;\; := \;\; \Big\{\theta\in\mathbb{L}^0(\Omega,\overline{\mathbb{R}}_+):\theta=\theta', \; \mu\otimes\mathrm{pw}, \textit{and } \theta\geq\theta', \; \textit{for some } \theta'\in\Theta\Big\}.$$

Definition 3.5.3. *A sequence $(\theta_n)_{n\geq1}\subset\mathbb{L}^0(\Omega)$ converges $\mu\otimes\mathrm{pw}$ to some $\theta\in\mathbb{L}^0(\Omega)$ if*

$$\mathrm{dom}_X\left(\underline{\theta}_\infty\right)=\mathrm{dom}_X\theta \;\; \textit{and} \;\; \theta_n(X,\cdot)\longrightarrow\theta(X,\cdot), \textit{pointwise on } \mathrm{dom}_X\theta, \; \mu-\textit{a.s.}$$

(i) *A subset $\mathcal{T}\subset\Theta_\mu$ is $\mu\otimes\mathrm{pw}$-Fatou closed if $\underline{\theta}_\infty\in\mathcal{T}$ for all $(\theta_n)_{n\geq1}\subset\mathcal{T}$ converging $\mu\otimes\mathrm{pw}$.*
(ii) *The $\mu\otimes\mathrm{pw}-$Fatou closure of a subset $A\subset\Theta_\mu$ is the smallest $\mu\otimes\mathrm{pw}-$Fatou closed set containing A:*

$$\widehat{A}:=\bigcap\Big\{\mathcal{T}\subset\Theta_\mu: \;A\subset\mathcal{T}, \; \textit{and } \mathcal{T} \; \mu\otimes\mathrm{pw}\textit{-Fatou closed}\Big\}.$$

Recall the definition for $a\geq0$, of the set $\mathfrak{C}_a:=\Big\{f\in\mathfrak{C}:(\nu-\mu)(f)\leq a\Big\}$, we introduce

$$\widehat{\mathcal{T}}(\mu,\nu):=\bigcup_{a\geq0}\widehat{\mathcal{T}}_a, \text{ where } \widehat{\mathcal{T}}_a:=\widehat{\mathbf{T}(\mathfrak{C}_a)}, \text{ and } \mathbf{T}\Big(\mathfrak{C}_a\Big):=\Big\{\mathbf{T}_pf:f\in\mathfrak{C}_a,p\in\partial f\Big\}.$$

Similar to $\nu\overline{\ominus}\mu$ for $\widetilde{\mathcal{T}}(\mu,\nu)$, we now introduce the extended $(\nu-\mu)-$integral:

$$\nu\widehat{\ominus}\mu[\theta]:=\inf\Big\{a\geq0:\theta\in\widehat{\mathcal{T}}_a\Big\} \;\; \text{for} \;\; \theta\in\widehat{\mathcal{T}}(\mu,\nu).$$

3.5.3 Duality result

As a preparation for the proof of Theorem 3.5.1, we prove the following Lemma.

Lemma 3.5.4. *Let $\widehat{\theta} \in \widehat{\mathcal{T}}(\mu,\nu)$, under Assumption 3.2.6, we may find $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$ such that $\theta \geq \widehat{\theta}$ and $\nu \overline{\ominus} \mu[\theta] \leq \nu \widehat{\ominus} \mu[\widehat{\theta}]$.*

Proof. Let $a > 0$, we consider $\mathcal{T}$ the collection of $\widehat{\theta} \in \Theta_\mu$ such that we may find $\theta \in \widetilde{\mathcal{T}}_a$ with $\theta \geq \widehat{\theta}$. First we have easily $\mathbf{T}(\mathfrak{C}_a) \subset \mathcal{T}$, as $\mathbf{T}(\mathfrak{C}_a) \subset \widetilde{\mathcal{T}}_a$. Now we consider $(\widehat{\theta}_n)_{n\geq 1} \subset \mathcal{T}$ converging $\mu\otimes$pw to $\widehat{\underline{\theta}}_\infty$. For each $n \geq 1$, we may find $\theta_n \in \widetilde{\mathcal{T}}_a$ such that $\theta_n \geq \widehat{\theta}_n$ and $\nu \overline{\ominus} \mu[\theta_n] \leq a$. Now we may use Assumption 3.2.6, we may find $\theta \in \widetilde{\mathcal{T}}_a$ such that $\theta_n \rightsquigarrow \theta$ by the fact that $\widetilde{\mathcal{T}}_a = a\widetilde{\mathcal{T}}_1$. By the generation properties, $\theta \geq \underline{\theta}_\infty \geq \widehat{\underline{\theta}}_\infty$, which implies that $\widehat{\underline{\theta}}_\infty \in \mathcal{T}$. $\mathcal{T}$ is $\mu\otimes$pw$-$Fatou closed, and therefore $\widehat{\mathcal{T}}_a \subset \mathcal{T}$.

Now let $\widehat{\theta} \in \widehat{\mathcal{T}}(\mu,\nu)$, with $l := \nu \widehat{\ominus} \mu[\widehat{\theta}]$. By what we did above, for all $n \geq 1$, we may find $\theta_n \in \widetilde{\mathcal{T}}_{l+1/n}$ such that $\widehat{\theta} \leq \theta_n$. We use again Assumption 3.2.6 to get $\theta_n \rightsquigarrow \theta$, by properties of generation, $\theta \geq \underline{\theta}_\infty \geq \widehat{\theta}$. By construction, $\nu \overline{\ominus} \mu[\theta] \leq l = \nu \widehat{\ominus} \mu[\widehat{\theta}]$. □

Proof of Theorem 3.5.1 By Theorem 3.8 in [58], we may find $(\bar{\varphi}, \bar{\psi}, \bar{h}, \widehat{\theta}) \in \mathbb{L}^1_+(\mu) \times \mathbb{L}^1_+(\nu) \times \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d) \times \widehat{\mathcal{T}}(\mu,\nu)$ such that $c \leq \bar{\varphi} \oplus \bar{\psi} + \bar{h}^\otimes + \widehat{\theta}$ on $\{Y \in \operatorname{aff}\operatorname{rf}_X \operatorname{conv}\operatorname{dom}(\widehat{\theta}(X,\cdot) + \bar{\psi})\}$, furthermore, $\mathbf{S}_{\mu,\nu}(c) = \mu[\bar{\varphi}] + \nu[\bar{\psi}] + \nu \widehat{\ominus} \mu[\widehat{\theta}]$ and $\mathbf{S}_{\mu,\nu}(\widehat{\theta}) = \nu \widehat{\ominus} \mu[\widehat{\theta}] < \infty$. By lemma 3.5.4, we may find $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$ such that $\widehat{\theta} \leq \theta$ and $\nu \overline{\ominus} \mu[\theta] \leq \nu \widehat{\ominus} \mu[\widehat{\theta}]$.

We have that $c \leq \bar{\varphi} \oplus \bar{\psi} + \bar{h}^\otimes + \widehat{\theta} \leq \bar{\varphi} \oplus \bar{\psi} + \bar{h}^\otimes + \theta$ on $\{Y \in \operatorname{aff}\operatorname{rf}_X \operatorname{conv}\operatorname{dom}(\theta(X,\cdot) + \bar{\psi})\}$ which is included in $\{Y \in \operatorname{aff}\operatorname{rf}_X \operatorname{conv}\operatorname{dom}(\widehat{\theta}(X,\cdot) + \bar{\psi})\}$.

As $\widehat{\theta} \leq \theta$, we have $\mathbf{S}_{\mu,\nu}(\theta) \geq \mathbf{S}_{\mu,\nu}(\widehat{\theta}) = \nu \widehat{\ominus} \mu[\widehat{\theta}] \geq \nu \overline{\ominus} \mu[\theta]$. From Proposition 3.2.16 (i), we get that $\mathbf{S}_{\mu,\nu}(\theta) = \nu \overline{\ominus} \mu[\theta] = \nu \widehat{\ominus} \mu[\widehat{\theta}] < \infty$. As $\theta \geq \widehat{\theta}$, we have $(\bar{\varphi}, \bar{\psi}, \bar{h}, \theta) \in \widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)$. Finally, as $Val(\bar{\varphi}, \bar{\psi}, \bar{h}, \theta) = \mu[\bar{\varphi}] + \nu[\bar{\psi}] + \nu \overline{\ominus} \mu[\theta] = \mu[\bar{\varphi}] + \nu[\bar{\psi}] + \nu \widehat{\ominus} \mu[\widehat{\theta}] = \mathbf{S}_{\mu,\nu}(c)$, the result is proved. □

Proof of Theorem 3.3.2 By Theorem 3.5.1, we may find $(\bar{\varphi}, \bar{\psi}, \bar{h}, \theta) \in \widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)$ such that $\mu[\bar{\varphi}] + \nu[\bar{\psi}] + \nu \widehat{\ominus} \mu[\theta] = \mathbf{S}_{\mu,\nu}(c)$. As Assumption 3.2.6 holds, by Proposition 3.2.13, we get $f \in \mathfrak{C}_{\mu,\nu}$ and $p \in \partial^{\mu,\nu} f$ such that $\mathbf{T}_p f = \theta$, q.s. Therefore, by definition we have $\nu \overline{\ominus} \mu[f] \leq \nu \widehat{\ominus} \mu[\theta]$. Then we denote $\varphi := \bar{\varphi} - f$, $\psi := \bar{\psi} + f$, and $h := \bar{h} - p$. As $\varphi \oplus \psi + h^\otimes = \bar{\varphi} \oplus \bar{\psi} + \bar{h}^\otimes + \theta \geq c$, q.s., (as $Y \in \operatorname{aff}\operatorname{rf}_X \operatorname{conv}\operatorname{dom}(\widehat{\theta}(X,\cdot) + \bar{\psi})$, q.s.) $\mathbf{S}_{\mu,\nu}(\mathbf{T}_p f) = \nu \widehat{\ominus} \mu[\theta] \geq \nu \overline{\ominus} \mu[f]$. As $\nu \underline{\ominus} \mu[f] := \mathbf{S}_{\mu,\nu}(\mathbf{T}_p f) \leq \nu \overline{\ominus} \mu[f]$ by Proposition 3.2.16 (i), we have $\nu \ominus \mu[f] := \nu \underline{\ominus} \mu[f] = \nu \overline{\ominus} \mu[f]$, and therefore f is a $\mathcal{M}(\mu,\nu)-$convex moderator for (φ, ψ), and as $\mu[\varphi + f] + \mu[\psi - f] + \nu \ominus \mu[f] = \mathbf{S}_{\mu,\nu}(c)$, the duality result, and attainment are proved. □

3.5.4 Structure of polar sets

Proof of Proposition 3.2.7 Step 1: Let a Borel $N \in \mathcal{N}_{\mu,\nu}$ such that θ is a $N-$tangent convex function. Then $c := \infty \mathbf{1}_N$ is Borel measurable and non-negative. Notice that $\mathbf{S}_{\mu,\nu}(c) = 0$. By Theorem 3.5.1, we may find $(\varphi_1, \psi_1, h_1, \theta_1) \in \widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)$ such that $\mu[\varphi_1] + \nu[\psi_1] + \nu\widehat{\ominus}\mu[\theta_1] = \mathbf{S}_{\mu,\nu}(c) = 0$. Then by the pointwise inequality $\infty \mathbf{1}_N \le \varphi_1 \oplus \psi_1 + h_1^{\otimes} + \theta_1$ on $\{Y \in \mathrm{aff}\,\mathrm{rf}_X \mathrm{conv}\, D(X)\}$, with $D(X) := \mathrm{dom}\big(\theta_1(X,\cdot) + \psi_1\big)$, (the convention is $0 \times \infty = 0$).

By Subsection 6.1 in [58], we may find $N'_\mu \in \mathcal{N}_\mu$, $N_\nu \in \mathcal{N}_\nu$, and $\widehat{\theta} \in \widehat{\mathcal{T}}(\mu,\nu)$ such that $I(X) \subset D(X)$, $\mathrm{rf}_X \mathrm{conv}(\mathrm{dom}\widehat{\theta}(X,\cdot) \setminus N_\nu) = I(X)$, and $\mathrm{dom}\widehat{\theta}(X,\cdot) \setminus N_\nu \subset \bar{J}(X)$, on N'^c_μ. By Lemma 3.5.4 we may find $\widehat{\theta} \le \widetilde{\theta} \in \widetilde{\mathcal{T}}(\mu,\nu)$. Up to adding $\mathbf{1}_{N'_\mu}$ to ϕ_1, $\mathbf{1}_{N_\nu}$ to ψ_1, and $\widetilde{\theta}$ to θ_1, we may assume that $\mathbf{1}_{N'_\mu} \le \phi_1$, $\mathbf{1}_{N_\nu} \le \psi_1$, and $\widetilde{\theta} \le \theta_1$. We get that

$$
\begin{aligned}
N &\subset \{\varphi_1(X) = \infty\} \cup \{\psi_1(Y) = \infty\} \cup \{Y \notin \mathrm{dom}\theta_1(X,\cdot)\} \cup \{Y \notin \mathrm{aff}\,\mathrm{rf}_X \mathrm{conv}\, D(X)\} \\
&= \{\varphi_1(X) = \infty\} \cup \Big\{Y \notin D(X) \cap \mathrm{aff}\,\mathrm{rf}_X \mathrm{conv}\, D(X)\Big\} \\
&= \{\varphi_1(X) = \infty\} \cup \Big\{Y \notin D(X) \cap \mathrm{aff}\, I(X)\Big\}.
\end{aligned}
$$

We have

$$
N \subset \mathrm{dom}\Big(\theta_1 + \varphi_1 \oplus \psi_1\Big)^c, \quad \text{and} N \subset \{X \in N_\mu\} \cup \{Y \in N_\mu\} \cup \{Y \notin \bar{J}(X)\} \tag{3.5.1}
$$

Notice that as $\mu[\varphi_1] + \nu[\psi_1] = 0$, $\{\varphi_1 = \infty\} \in \mathcal{N}_\mu$ and $\{\psi_1 = \infty\} \in \mathcal{N}_\nu$. We also have $\nu\overline{\ominus}\mu[\theta_1] < \infty$. We may replace φ_1 by $\infty \mathbf{1}_{\varphi_1 = \infty}$, ψ_1 by $\infty \mathbf{1}_{\psi_1 = \infty}$, and θ_1 by $\infty \mathbf{1}_{\theta_1 = \infty} \in \widetilde{\mathcal{T}}(\mu,\nu)$, where the fact that $\infty \mathbf{1}_{\theta_1 = \infty} \in \widetilde{\mathcal{T}}(\mu,\nu)$ stems from the fact that $\frac{1}{n}\theta_1 \rightsquigarrow \infty \mathbf{1}_{\theta_1 = \infty} \in \widetilde{\mathcal{T}}(\mu,\nu)$, proving as well that

$$
\nu\overline{\ominus}\mu[\infty \mathbf{1}_{\theta_1 = \infty}] = 0. \tag{3.5.2}
$$

Thanks to these modifications, φ_1, ψ_1, and θ_1 only take the values 0 or ∞.
Step 2: Now let a Borel set $N_1 \in \mathcal{N}_{\mu,\nu}$ be such that θ_1 is a N_1-tangent convex function. Then similar to what was done for N, we may find $(\varphi_2, \psi_2, \theta_2) \in \mathbb{L}^1_+(\mu) \times \mathbb{L}^1_+(\nu) \times \widetilde{\mathcal{T}}(\mu,\nu)$ such that

$$
N_1 \subset \mathrm{dom}\Big(\theta_2 + \varphi_2 \oplus \psi_2\Big)^c.
$$

Iterating this process for all $k \geq 2$, we define $(N_k, \varphi_k, \psi_k, \theta_k)$ for all $k \geq 1$. Now let

$$\varphi_\infty := \sum_{k\geq 1} \varphi_k \in \mathbb{L}^1_+(\mu), \quad \psi_\infty := \textstyle\sum_{k\geq 1} \psi_k \in \mathbb{L}^1_+(\nu), \quad \text{and } \theta' := \theta + \displaystyle\sum_{k\geq 1} \theta_k \in \widetilde{\mathcal{T}}(\mu,\nu) \geq \theta.$$

Let $N^0_\mu := (\mathrm{dom}\varphi_\infty)^c$, and $N^0_\nu := (\mathrm{dom}\psi_\infty)^c$. Notice that $\mu[\varphi_\infty] = \nu[\psi_\infty] = 0$, and therefore, $(N^0_\mu, N^0_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$. We now fix $(N^0_\mu, N^0_\nu) \subset (N_\mu, N_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$, and denote $\varphi := \infty \mathbf{1}_{N_\mu}$, and $\psi := \infty \mathbf{1}_{N_\nu}$.

Recall that $J(X) := \mathrm{conv}\,\mathrm{dom}\Big(\theta'(X,\cdot) + \psi\Big) \cap \mathrm{aff}\, I(X) = \mathrm{conv}\, D_\infty(X) \cap \mathrm{aff}\, D_\infty(X)$, where we denote $D_\infty(X) := \mathrm{dom}\Big(\theta'(X,\cdot) + \psi\Big)$ By Proposition 2.1 (ii) in [58], $\mathrm{conv}\, D_\infty(x) \setminus \mathrm{rf}_x \mathrm{conv}\, D_\infty(x)$ is convex for $x \in \mathbb{R}^d$. Therefore, we may find $u(x) \in (\mathrm{aff}\,\mathrm{rf}_x \mathrm{conv}\, D_\infty(x) - x)^\perp$ such that $y \in \mathrm{conv}\, D_\infty(x) \setminus \mathrm{rf}_x \mathrm{conv}\, D_\infty(x)$ implies that $u(x) \cdot (y - x) > 0$ by the Hahn-Banach theorem, so that

$$J(X) = D_\infty(X) \cap \mathrm{aff}\,\mathrm{rf}_X \mathrm{conv}\, D_\infty(X) = \mathrm{dom}\Big((\theta' + \infty u^\otimes)(X,\cdot) + \psi\Big),$$

with the convention $\infty - \infty = \infty$. Finally,

$$\begin{aligned} N' &= \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J(X)\} \\ &= \mathrm{dom}(\varphi \oplus \psi + \infty u + \theta')^c \\ &\supset \cup_{k\geq 1} N_k \cup (\mathrm{dom}\theta_k)^c \cup N \qquad (3.5.3) \\ &\supset N. \end{aligned}$$

We proved the inclusion from (ii).

Step 3: Now we prove that N'^c is martingale monotone, which is the end of (ii). Let $\mathbb{P}$ with finite support such that $\mathbb{P}[N'^c] = 1$, and $\mathbb{P}'$ a competitor to $\mathbb{P}$. Let $k \geq 1$, we have $\mathbb{P}[N_k^c] = 1$ by (3.5.3), therefore, as θ_k is a N_k-tangent convex function, $\mathbb{P}'[\theta_k] \leq \mathbb{P}[\theta_k]$, therefore, as by (3.5.3) we have that $\mathbb{P}[dom\theta_k] = 1$, we also have that $\mathbb{P}'[\mathrm{dom}\theta_k] = 1$. As this holds for all $k \geq 1$, and for N and the $N-$tangent convex function θ, we have $\mathbb{P}'[\mathrm{dom}\theta'] = 1$. Now as $\mathbb{P}[\mathrm{dom}\varphi \times \mathrm{dom}\psi] = 1$, we clearly have $\mathbb{P}'[\mathrm{dom}\varphi \times \mathrm{dom}\psi] = 1$. Recall that by construction, $\mathrm{dom}\Big(\theta' + \varphi \oplus \psi\Big) = \Big(\infty \mathbf{1}_{\theta'=\infty} + \varphi \oplus \psi\Big)^{-1}(0)$, therefore, $\mathbb{P}[\infty \mathbf{1}_{\theta'=\infty} + \varphi \oplus \psi] = \mathbb{P}'[\infty \mathbf{1}_{\theta'=\infty} + \varphi \oplus \psi] = 0$. Let $n \geq 1$, $\mathbb{P}[\infty \mathbf{1}_{\theta'=\infty} + n u^\otimes + \varphi \oplus \psi] = \mathbb{P}'[\infty \mathbf{1}_{\theta'=\infty} + n u^\otimes + \varphi \oplus \psi] = 0$. As $u^\otimes$ is negative only where the rest of the function is infinite, $\infty \mathbf{1}_{\theta'=\infty} + n u^\otimes + \varphi \oplus \psi \geq 0$ for all $n \geq 1$. Then by monotone convergence theorem, $\mathbb{P}'[\infty \mathbf{1}_{\theta'=\infty} + \infty u^\otimes + \varphi \oplus \psi] = \mathbb{P}[\infty \mathbf{1}_{\theta'=\infty} + \infty u^\otimes + \varphi \oplus \psi] = 0$. Therefore, $\mathbb{P}'[N'] = 0$, proving that N'^c is martingale monotone.

Step 4: Now we prove that $J(X) = \mathrm{conv}\Big(J(X)\setminus N_\nu\Big) \subset \mathrm{dom}\theta(X,\cdot)$, which is the first part of (i).

$$\mathrm{dom}\Big((\theta' + \infty u^\otimes)(X,\cdot) + \psi\Big) \subset J(X) \cap \mathrm{dom}\psi_\infty \subset J(X).$$

Passing to the convex hull, we get $J(X) = \mathrm{conv}\Big(J(X)\cap\mathrm{dom}\psi\Big) = \mathrm{conv}\Big(J(X)\setminus N_\nu\Big)$ as $N_\nu = \{\psi = \infty\}$.

Step 5: Now we prove that $J(X) \subset \mathrm{dom}\theta'(X,\cdot) \subset \mathrm{dom}\theta(X,\cdot)$, which is the second part of (i). Let $x \in \mathbb{R}^d$, and $y \in J(x)$. Then $y = \sum_i \lambda_i y_i$, convex combination, with $(y_i) \subset \mathrm{dom}\theta'(x,\cdot)\cap\mathrm{dom}\psi$. Let $\mathbb{P} := \sum_i \lambda_i \delta_{(x,y_i)} + \delta_{(y,y)}$. Let $k \geq 1$, $\mathbb{P}[N_k^c \cup \{X = Y\}] = 1$, $\mathbb{P}[\theta_k] < \infty$, and therefore, as $\mathbb{P}' := \sum_i \lambda_i \delta_{(y,y_i)} + \delta_{(x,y)}$ is a competitor to $\mathbb{P}$, $\mathbb{P}'[\theta_k] \leq \mathbb{P}[\theta_k] < \infty$, and $y \in \mathrm{dom}\theta_k(x,\cdot)$. $J(x) \subset \mathrm{dom}\theta_k(x,\cdot)$ for all $k \geq 1$, $J(X) \subset \mathrm{dom}\theta'(X,\cdot)$ on N_μ^c.

Step 6: Now we prove that up to choosing well I, and up to a modification of J on a $\mu-$null set, $I \subset J \subset \bar{J} \subset \mathrm{cl}\, I$, and J is constant on $I(x)$, for all $x \in \mathbb{R}^d$, which is the part concerning J of the end of (iii).

We have that $\{I(x), x \in \mathbb{R}^d\}$ is a partition of $\mathbb{R}^d$, $I \subset \bar{J} \subset \mathrm{cl}\, I$, and $\bar{J}$ is constant on $I(x)$ for all x. By looking at the proof of Theorem 2.1 in [58], we may enlarge the $\mu-$null set $N_\mu^I \in \mathcal{N}_\mu$ such that $I = \{X\}$ on $\Big(\cup_{x' \notin N_\mu^I} I(x')\Big)^c$. We do so by requiring that $N_\mu \subset N_\mu^I$. Now we prove that J is constant on $I(X)$, $\mu-$a.s. Let $x_1, x_2 \in \mathrm{dom}\varphi_\infty$, and $y \in \mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_1,\cdot)\cap\mathrm{dom}\psi_\infty$, then $y - \epsilon(y - x_2) \in I(x_2)$ for $\epsilon > 0$ small enough, as $x_2 \in \mathrm{ri} I(x_2) = \mathrm{ri} I(x_1)$, and $y \in \mathrm{cl}\, I(x_1)$, as $J(X) \subset \bar{J}(X) \subset \mathrm{cl}\, I(X)$ by (3.5.1). Then we may find $x_1 = \sum_i \lambda_i y_i + \lambda y$, convex combination, with $(y_i) \subset \mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_1,\cdot)\cap \mathrm{dom}\psi_\infty$, and $\lambda > 0$. Then let $\mathbb{P} := \sum_i \lambda_i \delta_{(x_1,y_i)} + \lambda\delta_{(x_1,y)} + \delta_{(x_2,x_2)}$. For all $k \geq 1$, notice that $\mathbb{P}[N_k^c \cup \{X = Y\}] = 1$, as $x_2 \in (N_k^c)_{x_2}$. Notice furthermore that $\mathbb{P}[\theta_k] < \infty$, and that $\mathbb{P}' := \sum_i \lambda_i \delta_{(x_2,y_i)} + \lambda\delta_{(x_2,y)} + \delta_{(x_1,x_2)}$ is a competitor to $\mathbb{P}$. Then as θ_k is a N_k-tangent convex function, $\mathbb{P}'[\theta_k] \leq \mathbb{P}[\theta_k] < \infty$, and therefore, as $\lambda > 0$, $\theta_k(x_2, y) < \infty$. We proved that

$$\mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_1,\cdot)\cap\mathrm{dom}\psi_\infty \subset \mathrm{dom}\theta_k(x_2,\cdot).$$

Therefore, $\mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_1,\cdot)\cap\mathrm{dom}\psi_\infty \subset \mathrm{dom}\theta_\infty(x_2,\cdot)$. As the other ingredients of J do not depend on x, and as we can exchange x_1, and x_2 in the previous reasoning,

$$\mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_1,\cdot)\cap\mathrm{dom}\psi_\infty = \mathrm{dom}(\theta_\infty + \infty u_\infty^\otimes)(x_2,\cdot)\cap\mathrm{dom}\psi_\infty.$$

Taking the convex hull, we get $J(x_1) = J(x_2)$.

Step 7: Now we prove that thanks to the modification of I and J, we have that J° is constant on all $I(x)$, for $x \in \mathbb{R}^d$, and that $\underline{J} \subset J^\circ \subset J$, which is the remaining part of (iii). By its definition, we see that the dependence of J° in x stems from a direct dependence in $J(x)$. The map J is constant on each $I(x)$, $x\in\mathbb{R}^d$, whence the same property for J°. Now for $I(x)\notin I(N_\mu^c)$, all these maps are equal to $\{x\}$, whence the inclusions and the constance.

Now we claim that for $x,x'\in\mathbb{R}^d$ such that $x'\in J(x)$, we have $J(x')\subset J(x)$. This claim will be justified in (iii) of the proof of Remark 3.2.8 above. Now if $x'\in J(x)\setminus N_\mu\subset J(x)$, we have as a consequence that $J(x')\subset J(x)$, and therefore $I(x')\subset J(x)$. We proved that $J^\circ\subset J$.

Finally by Proposition 2.4 in [58], we may find $\widehat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$ such that $\underline{J}(X)\setminus I(X)\subset\{y:\widehat{\mathbb{P}}_X[\{y\}]\}$, on N_μ. Then $\underline{J}\subset J^\circ$ on N_μ^c. Otherwise, these maps are again equal to $\{X\}$, whence the result. □

Proof of Remark 3.2.8 (i) Recall that, with the notations from Proposition 3.2.7, $J(X):=\text{conv}(\text{dom}\theta'(X,\cdot)\setminus N_\nu)\cap\bar{J}(X)$. $\theta'\in\widetilde{\mathcal{T}}(\mu,\nu)$, then $\mathbf{S}_{\mu,\nu}(\theta)<\infty$ and $Y\in\text{dom}\theta'(X,\cdot)$, $\mathcal{M}(\mu,\nu)-$q.s. Recall that $Y\in\bar{J}(X)$, and $Y\notin N_\nu$, $\mathcal{M}(\mu,\nu)-$q.s. All these ingredients prove that $Y\in J(X)$, q.s. and $Y\in J(X)\setminus N_\nu$, q.s. The result for J° is a consequence of the inclusion

$$J\setminus N_\nu\subset J^\circ\subset J. \tag{3.5.4}$$

(ii) Let $x,x'\in\mathbb{R}^d$, we prove that $J(x)\cap J(x')=\text{aff}\Big(J(x)\cap J(x')\Big)\cap J(x)$. The direct inclusion is trivial, let us prove the indirect inclusion. We first assume that $x,x'\in N_\mu^c$. We claim that

$$J(x)\cap J(x')=\text{conv}\Big(J(x)\cap J(x')\setminus N_\nu'\Big). \tag{3.5.5}$$

This claim will be proved in (iii). If $J(x)\cap J(x')=\emptyset$, the assertion is trivial, we assume now that this intersection is non-empty. Let $y_1,...,y_k\in J(x)\cap J(x')\setminus N_\nu'$ with $k\ge 1$, spanning $\text{aff}\Big(J(x)\cap J(x')\setminus N_\nu'\Big)$. Let $y\in\text{aff}\Big(J(x)\cap J(x')\Big)\cap J(x)$, and $y'=\frac{1}{k}\sum_i y_k$. We have $y'\in\text{ri}\,\text{conv}(y_1,...,y_k)$ and $y\in\text{aff}\,\text{conv}(y_1,...,y_k)$, therefore, for $\varepsilon>0$ small enough, $\varepsilon y+(1-\varepsilon)y'\in\text{ri}\,\text{conv}(y_1,...,y_k)\subset J(x)\cap J(x')\subset J(x')=\text{conv}\Big(J(x')\setminus N_\nu\Big)$ by (i). Then, for ε small enough, $\varepsilon y+(1-\varepsilon)y'=\sum_i\lambda_i y_i'$, convex combination, with $(y_i)_i\subset J(x')\setminus N_\nu$. Then $\mathbb{P}=\frac{1}{2}\varepsilon\delta_{(x,y)}+\frac{1}{2k}(1-\varepsilon)\sum_i\delta(x,y_i)+\frac{1}{2}\sum_i\lambda_i\delta(x',y_i')$ is concentrated on N'^c, and by (iv) we have that its competitor $\mathbb{P}'=\frac{1}{2}\varepsilon\delta_{(x',y)}+\frac{1}{2k}(1-\varepsilon)\sum_i\delta(x',y_i)+\frac{1}{2}\sum_i\lambda_i\delta(x,y_i')$ is also concentrated on N'^c. Therefore $y\in J(x')$, and as $y\in J(x)$, we proved the reverse inclusion: $J(x)\cap J(x')=\text{aff}\Big(J(x)\cap J(x')\Big)\cap J(x)$.

Now if $x, x' \in \cup_{x'' \notin N_\mu} I(x'')$, we may find $x_1, x_2 \in N_\mu^c$ such that $J(x) = J(x_1)$, and $J(x') = J(x_2)$, whence the result from what precedes. Finally if x or x' is not in $\cup_{x'' \notin N_\mu} I(x'')$, If it is x, then $I(x) = J(x) = \{x\}$, and if $x \in J(x')$, then the result is $\{x\} = \{x\}$, else it is $\emptyset = \emptyset$. If it is x', then if $x' \in J(x)$, the result is $\{x'\} = \{x'\}$, otherwise, it is again $\emptyset = \emptyset$. In all the cases, the result holds.

Finally we extend this result to J°. Notice that by (3.5.5) together with (3.5.4), we have $\mathrm{aff}\Big(J(x) \cap J(x')\Big) = \mathrm{aff}\Big(J^\circ(x) \cap J^\circ(x')\Big)$. Now consider the equation that we previously proved $J(x) \cap J(x') = \mathrm{aff}\Big(J(x) \cap J(x')\Big) \cap J(x)$, subtracting $N_\nu \setminus \cup_{x'' \notin N_\mu} I(x'')$ and replacing $\mathrm{aff}\Big(J(x) \cap J(x')\Big)$, we get $J^\circ(x) \cap J^\circ(x') = \mathrm{aff}\Big(J^\circ(x) \cap J^\circ(x')\Big) \cap J^\circ(x)$.
<u>(iii)</u> Let $y \in J(x) \cap J(x')$. By (i), $\mathrm{conv}\Big(J(x) \setminus N_\nu\Big) = J(x)$, and the same holds for x'. Then we may find $y_1, ..., y_k \in J(x) \setminus N_\nu$ and $y_1', ..., y_{k'}' \in J(x') \setminus N_\nu$ with $\sum_i \lambda_i y_i = \sum_i \lambda_i' y_i' = y$, where the (λ_i) and (λ_i') are non-zero coefficients such that the sums are convex combinations. Now notice that $\mathbb{P} := \frac{1}{2}\sum_i \lambda_i \delta_{(x,y_i)} + \frac{1}{2}\sum_i \lambda_i' \delta_{(x',y_i')}$ is supported in N'^c. By (iv), its competitor $\mathbb{P}' := \frac{1}{2}\sum_i \lambda_i \delta_{(x',y_i)} + \frac{1}{2}\sum_i \lambda_i' \delta_{(x,y_i')}$ is also supported on N'^c. Therefore, $y_1, ..., y_k, y_1', ..., y_{k'}' \in J(x) \cap J(x') \setminus N_\nu$. We proved that $J(x) \cap J(x') \subset \mathrm{conv}\Big(J(x) \cap J(x') \setminus N_\nu'\Big)$, and therefore as the other inclusion is easy, we have $J(x) \cap J(x') = \mathrm{conv}\Big(J(x) \cap J(x') \setminus N_\nu'\Big)$. The extension of this result for J° is again a consequence of the inclusion (3.5.4).
<u>(iv)</u> Now we assume additionally that $I(x') \cap J(x) \neq \emptyset$, let us prove that then $J(x') \subset J(x)$. If $x' \notin \cup_{x'' \notin N_\mu} I(x'')$, then $J(x') = \{x'\}$ and the result is trivial. If $x \notin \cup_{x'' \notin N_\mu} I(x'')$, then the result is similarly trivial. By constance of J and I on $I(x)$ for all x, we may assume now that $x, x' \in N_\mu^c$. Then let $y \in I(x') \cap J(x) \subset \mathrm{conv}\Big(J(x') \setminus N_\nu\Big) \cap \mathrm{conv}\Big(J(x) \setminus N_\nu\Big)$. Let $y' \in J(x') \setminus N_\nu$, for $\varepsilon > 0$ small enough, $y - \varepsilon(y' - y) \in I(y')$ by the fact that $I(y')$ is open in $\mathrm{aff} J(x')$. Then $y - \varepsilon(y' - y) = \sum_i \lambda_i y_i$, and $y = \sum_i \lambda_i' y_i'$, convex combinations where $(y_i)_i \subset J(x') \setminus N_\nu$, and $(y_i)_i \subset J(x') \setminus N_\nu$. Then $\mathbb{P} := \frac{1}{2}\frac{\varepsilon}{1+\varepsilon}\delta_{(x,y')} + \frac{1}{2}\frac{1}{1+\varepsilon}\sum_i \lambda_i \delta_{(x,y_i)} + \frac{1}{2}\sum_i \lambda_i' \delta_{(x',y_i')}$ is concentrated on N'^c, and by (iv), so does its competitor $\mathbb{P}' := \frac{1}{2}\frac{\varepsilon}{1+\varepsilon}\delta_{(x',y')} + \frac{1}{2}\frac{1}{1+\varepsilon}\sum_i \lambda_i \delta_{(x',y_i)} + \frac{1}{2}\sum_i \lambda_i' \delta_{(x,y_i')}$. Then in particular, $y' \in J(x)$. Finally, $J(x') \setminus N_\nu \subset J(x)$, passing to the convex hull, we get that $J(x') \subset J(x)$.

Finally, if $I(x') \cap J^\circ(x) \neq \emptyset$, then $I(x') \cap J(x) \neq \emptyset$, and $J(x') \subset J(x)$. Subtracting $N_\nu \setminus \cup_{x'' \notin N_\mu} I(x'')$ on both sides, we get $J^\circ(x') \subset J^\circ(x)$. □

Proof of Theorem 3.2.10 Let $(N_\mu, N_\nu) \in \mathcal{N}_\mu \times \mathcal{N}_\nu$, and $J \in \mathcal{J}(\mu, \nu)$. The "if" part holds as $Y \in J(X)$, $X \notin N_\mu$, and $Y \notin N_\nu$ q.s.

Now, consider an analytic set $N \in \mathcal{N}_{\mu,\nu}$. Then $c := \infty \mathbf{1}_N$ is upper semi-analytic non-negative. Notice that $\mathbf{S}_{\mu,\nu}(c) = 0$. By Theorem 3.5.1, we may find $(\varphi, \psi, h, \theta) \in$

$\widetilde{\mathcal{D}^{mod}_{\mu,\nu}}(c)$ such that $\mu[\varphi]+\nu[\psi]+\nu\widehat{\ominus}\mu[\theta]=\mathbf{S}_{\mu,\nu}(c)=0$. Then by the pointwise inequality $\infty\mathbf{1}_N\le\varphi\otimes\psi+h^{\otimes}+\theta$, on $\{Y\in\mathrm{aff}\,\mathrm{rf}_X\mathrm{conv}\,D(X)\}$, with $D(X)=\mathrm{dom}\big(\theta(X,\cdot)+\psi\big)$, we get that

$$\begin{aligned} N &\subset \{\varphi(X)=\infty\}\cup\{\psi(Y)=\infty\}\cup\{Y\notin\mathrm{dom}\theta(X,\cdot)\}\cup\{Y\notin\mathrm{aff}\,\mathrm{rf}_X\mathrm{conv}\,D(X)\}\\ &= \{\varphi(X)=\infty\}\cup\{\psi(Y)=\infty\}\cup\Big\{Y\notin\mathrm{dom}\theta(X,\cdot)\cap\mathrm{aff}\,\mathrm{rf}_X\mathrm{conv}\,D(X)\Big\},\end{aligned}$$

Let $J\in\mathcal{J}(\mu,\nu)$ from Proposition 3.2.7 for θ, and $N_\nu:=\mathrm{dom}\psi^c\in\mathcal{N}_\nu$. We have $J(X)\subset\mathrm{aff}\,\mathrm{rf}_X\mathrm{conv}\,D(X)$ and $J(X)\subset\bar{J}(X)\subset\mathrm{dom}\theta(X,\cdot)$, $\mu-$a.s. Therefore, we have

$$N \subset N_0:=\{X\in N_\mu\}\cup\{Y\in N_\nu\}\cup\{Y\notin J(X)\},$$

for some $N_\mu\in\mathcal{N}_\mu$, and $N_\nu\subset N_\nu\in\mathcal{N}_\nu$. By Proposition 3.2.7 (i) and (iv), N_0 may be chosen canonical up to enlarging N_μ. □

3.5.5 Decomposition in irreducible martingale optimal transports

In order to prove theorem 3.3.5, we first need to establish the following lemma.

Lemma 3.5.5. *Let $\theta\in\widehat{\mathcal{T}}(\mu,\nu)$ and $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, we may find $\theta'\in\widehat{\mathcal{T}}(\mu,\nu)$ such that $\theta\le\theta'$, $\nu\widehat{\ominus}\mu[\theta']\le\nu\widehat{\ominus}\mu[\theta]$, and $\int_{I(\mathbb{R}^d)}\nu_i^{\mathbb{P}}\widehat{\ominus}\mu_i[\theta']\eta(di)\le\nu\widehat{\ominus}\mu[\theta]$. Furthermore under Assumption 3.2.6, we may find $f\in\mathfrak{C}_{\mu,\nu}$ and $p\in\partial^{\mu,\nu}f$ such that $\theta\le\mathbf{T}_pf$, q.s., $\nu\overline{\ominus}\mu[f]\le\nu\widehat{\ominus}\mu[\theta]$, and $\int_{I(\mathbb{R}^d)}\nu_i^{\mathbb{P}}\overline{\ominus}\mu_i[f]\eta(di)\le\nu\widehat{\ominus}\mu[\theta]$.*

Proof. Let $a>0$, we consider $\mathcal{T}$ the collection of $\widehat{\theta}\in\Theta_\mu$ such that we may find $\theta'\in\widehat{\mathcal{T}}_a$ with $\theta'\ge\widehat{\theta}$, $\nu\widehat{\ominus}\mu[\theta']\le a$, and $\int_{I(\mathbb{R}^d)}\nu_i^{\mathbb{P}}\widehat{\ominus}\mu_i[\theta']\eta(di)\le a$. First we have easily $\mathbf{T}(\mathfrak{C}_a)\subset\mathcal{T}$, as $\mathbf{T}(\mathfrak{C}_a)\subset\widetilde{\mathcal{T}}_a$, and $\int_{I(\mathbb{R}^d)}\nu_i^{\mathbb{P}}\widehat{\ominus}\mu_i[\theta']\eta(di)=\int_{I(\mathbb{R}^d)}(\nu_i^{\mathbb{P}}-\mu_i)[\theta']\eta(di)=(\nu-\mu)[\theta']$, for $\theta'\in\mathbf{T}(\mathfrak{C}_a)$. Now we consider $(\widehat{\theta}_n)_{n\ge1}\subset\mathcal{T}$ converging $\mu\otimes$pw to $\underline{\widehat{\theta}}_\infty$. For each $n\ge1$, we may find $\theta_n\in\widehat{\mathcal{T}}_a$ such that $\theta_n\ge\widehat{\theta}_n$, $\nu\widehat{\ominus}\mu[\theta_n]\le a$, and $\int_{I(\mathbb{R}^d)}\nu_i^{\mathbb{P}}\widehat{\ominus}\mu_i[\theta_n]\eta(di)\le a$. By the Komlós Lemma on $I\longmapsto\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_n]$ under the probability η together with Lemma 2.12 in [58], we may find convex combination coefficients $(\lambda_k^n)_{1\le n\le k}$ such that $\sum_{k=n}^\infty\lambda_k^n\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_k]$ converges $\eta-$a.s. and $\theta_n':=\sum_{k=n}^\infty\lambda_k^n\theta_k$ converges $\mu\otimes$pw to $\theta':=\underline{\theta'}_\infty$, as $n\longrightarrow\infty$, and moreover $\nu\widehat{\ominus}\mu[\theta']\le a$. As θ_n' is a convex extraction of $\widehat{\theta}_n$, we have $\theta':=\underline{\theta'}_\infty\ge\underline{\widehat{\theta}}_\infty$. Moreover, by convexity of $\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I$, we have $\sum_{k=n}^\infty\lambda_k^n\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_k]\ge\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_n']$, and therefore

$$\liminf_{n\to\infty}\sum_{k=n}^\infty\lambda_k^n\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_k]=\limsup_{n\to\infty}\sum_{k=n}^\infty\lambda_k^n\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_k]\ge\limsup_{n\to\infty}\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta_n']\ge\nu_I^{\mathbb{P}}\widehat{\ominus}\mu_I[\theta']$$

$\eta-$a.s. Integrating this inequality with respect to η, and using Fatou's Lemma, we get

$$\int_{I(\mathbb{R}^d)} \nu_i^{\mathbb{P}} \widehat{\ominus} \mu_i[\theta'] \eta(di) \leq a.$$

Then $\underline{\widehat{\theta}}_\infty \in \mathcal{T}$. Hence, $\mathcal{T}$ is $\mu\otimes$pw$-$Fatou closed, and therefore $\widehat{\mathcal{T}}_a \subset \mathcal{T}$.

Now let $\widehat{\theta} \in \widehat{\mathcal{T}}(\mu,\nu)$, with $l := \nu\widehat{\ominus}\mu[\widehat{\theta}]$. By the previous step, for all $n \geq 1$, we may find $\theta'_n \in \widehat{\mathcal{T}}_{l+1/n}$ with $\int_{I(\mathbb{R}^d)} \nu_i^{\mathbb{P}} \widehat{\ominus} \mu_i[\theta'_n]\eta(di) \leq l + 1/n$ such that $\widehat{\theta} \leq \theta'_n$. Similar to the proof of Lemma 3.5.4, we get $\theta' \in \widehat{\mathcal{T}}(\mu,\nu)$ such that $\theta \leq \theta'$, $\nu\widehat{\ominus}\mu[\theta'] \leq l$, and $\int_{I(\mathbb{R}^d)} \nu_i^{\mathbb{P}} \widehat{\ominus} \mu_i[\theta'] \eta(di) \leq l$, thus proving the result.

We prove the second part of the Lemma similarly, using Assumption 3.2.6 instead of Lemma 2.12 in [58]. □

For the proof of next result, we need the following lemma:

Lemma 3.5.6. *Let $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, $m_X := \mu[X|I(X)]$, and $f_X(\cdot) := \theta(m_X,\cdot)$. Then we may find a $\mu-$unique measurable $\widehat{p}(X) \in \mathrm{aff} I(X) - X$ such that for some $N_\mu \in \mathcal{N}_\mu$,*

$$\theta = f_X(Y) - f_X(X) - \widehat{p}(X)\cdot(Y - X), \quad \text{on } \{Y \in \mathrm{aff}\,\mathrm{dom}_X\theta\} \cap \{X \notin N_\mu\}. \quad (3.5.6)$$

Proof. We consider $N_\mu \in \mathcal{N}_\mu$ from Proposition 2.10 in [58], so that for $x_1, x_2 \notin N_\mu$, $y_1, y_2 \in \mathbb{R}^d$, and $\lambda \in [0,1]$ with $\bar{y} := \lambda y_1 + (1-\lambda)y_2 \in \mathrm{dom}_{x_1}\theta \cap \mathrm{dom}_{x_2}\theta$, we have:

$$\lambda\theta(x_1,y_1) + (1-\lambda)\theta(x_1,y_2) - \theta(x_1,\bar{y}) = \lambda\theta(x_2,y_1) + (1-\lambda)\theta(x_2,y_2) - \theta(x_2,\bar{y}) \geq 0. \quad (3.5.7)$$

By possibly enlarging N_μ, we may suppose in addition that $I(x) \subset \mathrm{dom}_x\theta$ for all $x \in N_\mu^c$. For $x \in N_\mu^c$ and $y \in \mathrm{dom}_x\theta$, we define $H_x(y) := f_x(y) - f_x(x) - \theta(x,y)$. By (3.5.7), H_x is affine on $\mathrm{aff}\,\mathrm{dom}_x\theta \cap \mathrm{dom}\theta(x,\cdot)$. Indeed let $y_1 \in \mathrm{aff}\,\mathrm{dom}_x\theta \cap \mathrm{dom}\theta(x,\cdot)$, $y_2 \in \mathrm{dom}_x\theta$, and $0 \leq \lambda \leq 1$, then $\bar{y} := \lambda y_1 + (1-\lambda)y_2 \in \mathrm{dom}_x\theta$ and

$$\begin{aligned} H_x(\bar{y}) &= \theta(m_x,\bar{y}) - \theta(m_x,x) - \theta(x,\bar{y}) \\ &= \lambda\theta(m_x,y_1) + (1-\lambda)\theta(m_x,y_2) - \lambda\theta(x,y_1) - (1-\lambda)\theta(x,y_2) - \theta(m_x,x) \\ &= \lambda H_x(y_1) + (1-\lambda)H_x(y_2) \end{aligned}$$

We notice as well that $H_x(x) = 0$. Then we may find a unique $\widehat{p}(x) \in \mathrm{aff} I(x) - x$ so that for $y \in \mathrm{dom}_x\theta$, $H_x(y) = \widehat{p}(x)\cdot(y-x)$. $\widehat{p}(X)$ is measurable and unique on N_μ^c, and therefore $\mu-$a.e. unique. For $y \in \mathrm{aff}\,\mathrm{dom}_x\theta \cap \mathrm{dom}\theta(x,\cdot)$, it gives the desired equality (3.5.6). Now for $y \in \mathrm{aff}\,\mathrm{dom}_x\theta \cap \mathrm{dom}\theta(x,\cdot)^c$, let $0 < \lambda < 1$ such that $\bar{y} := \lambda x + (1-\lambda)y \in \mathrm{dom}_x\theta$. By (3.5.7), $\lambda\theta(x,x) + (1-\lambda)\theta(x,y) - \theta(x,\bar{y}) = \lambda f_x(x) + (1-\lambda)f_x(y) - f_x(\bar{y})$, and therefore $\theta(x,y)$ is finite if and only if $f_x(y)$ is finite. This proves that (3.5.6) holds for $y\in\operatorname{aff}\operatorname{dom}_x\theta$. $\square$

Proof of Theorem 3.3.5 For $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, $I_0\in I(\mathbb{R}^d)$, we have by definition of the supremum,

$$\mathbb{P}_{I_0}[c]\le \mathbf{S}_{\mathbb{P}_{I_0}\circ X^{-1},\mathbb{P}_{I_0}\circ Y^{-1}}(c)=\mathbf{S}_{\mu_{I_0},\nu^{\mathbb{P}}_{I_0}}(c),$$

where we denote by $\mathbb{P}_I$ a conditional disintegration of $\mathbb{P}$ with respect to the random variable I. Now we consider a minimizer for the dual problem $(\bar\varphi,\bar\psi,\bar h,\widehat\theta)\in\mathcal{D}^{mod}_{\mu,\nu}(c)$ and $\theta'\in\widehat{\mathcal{T}}(\mu,\nu)$ such that $\theta\le\theta'$, $\nu\widehat\ominus\mu[\theta']\le\nu\widehat\ominus\mu[\theta]$, and $\int_{I(\mathbb{R}^d)}\nu^{\mathbb{P}}_i\widehat\ominus\mu_i[\theta']\eta(di)\le\nu\widehat\ominus\mu[\theta]$ from Lemma 3.5.5. Recall the notation $m_X:=\mu[X|I(X)]=\mathbb{P}[Y|I(X)]$, by the martingale property, and let $f_X(Y):=\theta'(m_X,Y)$. From Lemma 3.5.6, we have $\theta'(X,Y)=f_X(Y)-f_X(X)-p_X(X)\cdot(Y-X)$, with $p_X\in\partial f_X(X)$, $\mathcal{M}(\mu,\nu)-$q.s. Then let $\varphi:=\bar\varphi-f_X$, $\psi_I(X):=\bar\psi(Y)+f_X(Y)$, $h:=\bar h-p_X$.

$$\mu_I[\bar\varphi]+\nu^{\mathbb{P}}_I[\bar\psi]+\nu^{\mathbb{P}}_I\widehat\ominus\mu_I[\theta']\ge\mu_I[\varphi]\oplus\nu^{\mathbb{P}}_I[\psi]\ge\mathbf{I}_{\mu_I,\nu^{\mathbb{P}}_I}(c)\ge\mathbf{S}_{\mu_I,\nu^{\mathbb{P}}_I}(c).$$

Integrating with respect to η, we get:

$$\begin{aligned}\mathbf{I}^{mod}_{\mu,\nu}(c)&\ge\int_{I(\mathbb{R}^d)}\mu_i[\bar\varphi]+\nu^{\mathbb{P}}_i[\bar\psi]+\nu^{\mathbb{P}}_i\widehat\ominus\mu_i[\theta']\eta(di)\ge\int_{I(\mathbb{R}^d)}\mathbf{I}_{\mu_i,\nu^{\mathbb{P}}_i}(c)\eta(di)\\&\ge\int_{I(\mathbb{R}^d)}\mathbf{S}_{\mu_i,\nu^{\mathbb{P}}_i}(c)\eta(di)\ge\mathbb{P}[c].\end{aligned}$$

Taking the supremum over $\mathbb{P}$:

$$\mathbf{I}^{mod}_{\mu,\nu}(c)\ge\sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\int_{I(\mathbb{R}^d)}\mathbf{I}_{\mu_i,\nu^{\mathbb{P}}_i}(c)\eta(di)\ge\sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\int_{I(\mathbb{R}^d)}\mathbf{S}_{\mu_i,\nu^{\mathbb{P}}_i}(c)\eta(di)\ge\mathbf{S}_{\mu,\nu}(c)$$

Then all the inequalities are equalities by the duality Theorem 3.8 in [58].

We consider $\mathbb{P}^*$ such that $\mathbb{P}^*[c]=\mathbf{S}_{\mu,\nu}(c)=\mathbf{I}^{mod}_{\mu,\nu}(c)$ gives us that there is an optimizer.

$$\begin{aligned}\mathbf{S}_{\mu,\nu}(c)&=\mathbb{P}^*[c]=\int_{I(\mathbb{R}^d)}\mathbb{P}^*_i[c]\eta(di)\le\int_{I(\mathbb{R}^d)}\mathbf{S}_{\mu_i,\nu^{\mathbb{P}^*}_i}(c)\eta(di)\le\int_{I(\mathbb{R}^d)}\mathbf{I}^{mod}_{\mu_i,\nu^{\mathbb{P}^*}_i}(c)\eta(di)\\&\le\int_{I(\mathbb{R}^d)}\mu_i[\bar\varphi]+\nu^{\mathbb{P}^*}_i[\bar\psi]+\nu^{\mathbb{P}^*}_i\widehat\ominus\mu_i[\theta']\eta(di)\le\mathbf{I}^{mod}_{\mu,\nu}(c).\end{aligned}$$

Then all these inequalities are equalities by duality.

The second part is proved similarly, using the second part of Lemma 3.5.5. $\square$

3.5.6 Properties of the weakly convex functions

The proof of Proposition 3.2.13 is very technical and requires several lemmas as a preparation.

Lemma 3.5.7. *Let $N_\mu \in \mathcal{N}_\mu$, we may find $N_\mu \subset N'_\mu \in \mathcal{N}_\mu$, and a Borel mapping* $\mathrm{ri}\overline{\mathcal{K}} \ni K \longmapsto m_K$ *such that $m_{I(X)} \in I(X) \setminus N'_\mu$ on $\{X \notin N'_\mu\}$.*

Proof. We may approximate N^c_μ from inside by a countable non-decreasing sequence of compacts $(K_n)_{n\geq 1}$: $\cup_{n\geq 1} K_n \subset N^c_\mu$, and $\mu[\cup_{n\geq 1} K_n] = 1$. Let $N'_\mu := (\cup_{n\geq 1} K_n)^c \in \mathcal{N}_\mu$. For $n \in \mathbb{N}$, the mapping $I_n : x \mapsto x + (1-1/n)\big(\mathrm{cl}\, I(x) - x\big) \cap K_n$ is measurable with closed values. Then we deduce from Theorem 4.1 of the survey on measurable selection [158] that we may find a measurable selection $m^n : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that $m^n(x) \in I_n(x)$ for all $x \in \mathbb{R}^d$. Define

$$m'(x) := m^{n(x)}(x) \quad \text{where} \quad n(x) := \inf\{n \geq 1 : I_n(x) \neq \emptyset\}, \quad x \in \mathbb{R}^d,$$

and $m^\infty := 0$. Then for all $x \notin N'_\mu$, we have the inclusion $\emptyset \neq \{x\} \cap N'^c_\mu \subset I(x) \cap N'^c_\mu = \cup_{n\geq 1} I_n(x)$, so that $n(x) < \infty$ and $m'(x) \in I(x) \setminus N'_\mu$. However, we want to find a map from $\overline{\mathcal{K}}$ to $\mathbb{R}^d$. Consider again the map $\bar m_I := \mathbb{E}^\mu[X|I]$. Notice that $\bar m_I \in I$ by the convexity of I, and that it is constant on $I(x)$, for all $x \in \mathbb{R}^d$. Then the map $m_I := m'(\bar m_I)$ satisfies the requirements of the lemma. $\square$

We fix a $N-$tangent convex function $\theta \in \widetilde{\mathcal{T}}(\mu,\nu)$. Let $N^0 := \{X \in N^0_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J(X)\} \in \mathcal{N}_{\mu,\nu}$, a canonical polar set such that $(N^0)^c \subset N^c \cap \mathrm{dom}\theta$ from Proposition 3.2.7. Consider the map m_I given by Lemma 3.5.7 for N^0_μ, let $\mathcal{N}_\mu \ni N_\mu \supset N^0_\mu$ such that $m_{I(X)} \in I(X) \setminus N_\mu$ on $\{X \notin N_\mu\}$. By Proposition 3.2.7 together with the fact that $N_\mu \supset N^0_\mu$, we may chose the map I so that $N' := \{X \in N_\mu\} \cup \{Y \in N_\nu\} \cup \{Y \notin J(X)\} \in \mathcal{N}_{\mu,\nu}$, a canonical polar set such that $N'^c \subset N^c \cap \mathrm{dom}\theta$. For $K \in I(\mathbb{R}^d) := \{I(x) : x \in \mathbb{R}^d\}$ we denote $f_K := \theta(m_K, \cdot)$.

Lemma 3.5.8. *We may find $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$ such that $\{Y \in J^\circ(X), X \notin N_\mu\} \subset N^c \cap$ $\mathrm{dom}\theta$, $\mathrm{conv}J^\circ = J = \mathrm{conv}\big(J \setminus N_\nu\big)$, and $\mathrm{conv}\big(J^\circ(x) \cap J^\circ(x')\big) = J(x) \cap J(x') = \mathrm{conv}\big(J(x) \cap J(x') \setminus N_\nu\big)$ for all $x, x' \in \mathbb{R}^d$.*

Proof. The map defined by $J^\circ(x) := \cup_{x' \in J(x)\setminus N_\mu} I(x') \cup J(x) \setminus N_\nu$ is in $\mathcal{J}^\circ(\mu,\nu)$. By Proposition 3.2.7, $J^\circ \subset J$, therefore $J^\circ \subset J = \mathrm{conv}(J \setminus N_\nu) \subset \mathrm{conv}\,\mathrm{dom}\theta(X,\cdot) = \mathrm{dom}\theta(X,\cdot)$ on N^c_μ, whence the inclusion $\{Y \in J^\circ(X), X \notin N_\mu\} \subset \mathrm{dom}\theta$.

Now we prove that $\{Y \in J^\circ(X), X \notin N_\mu\} \subset N^c$. Recall that $N' = \{Y \in J(X) \setminus N_\nu, X \notin N_\mu\} \subset N^c$. Let $x \notin N_\mu$, and $x' \in J(x) \setminus N_\mu$, then $I(x') \subset N^c_{x'}$. Let $y \in I(x') \subset$

$N'^c_{x'}$, by Proposition 3.2.7, $y \in J(x) \cap J(x') = \mathrm{conv}\Big(J(x) \cap J(x') \setminus N_\nu\Big)$. Then we may find $y_1, ..., y_k \in J(x) \cap J(x') \setminus N_\nu$ such that $y = \sum_i \lambda_i y_i$, convex combination. We also have $y \in N^c_{x'}$, then $\mathbb{P} := \frac{1}{2}\sum_i \lambda_i \delta_{(x,y_i)} + \frac{1}{2}\delta_{x',y}$, and $\mathbb{P}' := \frac{1}{2}\sum_i \lambda_i \delta_{(x',y_i)} + \frac{1}{2}\delta_{x,y}$ are competitors such that the only point in their support that may not be in N^c is (x,y), then by Definition 3.2.1 (iii), $(x,y) \in N^c$. We proved that $\{Y \in J^\circ(X), X \notin N_\mu\} \subset N^c$.

The other properties are direct consequences of Remark 3.2.8. □

Let $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$ and $N_\mu \in \mathcal{N}_\mu$ from Lemma 3.5.8.

Lemma 3.5.9. *We have $\theta = \mathbf{T}_{\widehat{p}} f_I(X)$ on $\{Y \in J^\circ(X), X \notin N_\mu\}$ for some $\widehat{p} \in \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d)$, and $J^\circ \in \mathcal{J}^\circ(\mu,\nu)$.*

Proof. Let $a_x := f_{I(x)} - f_{I(x)}(x) - \theta(x,\cdot)$. We claim that a_x is affine on $J^\circ(x)$, for all $x \notin N_\mu$, i.e. we may find a measurable map $\widehat{p}$ on N^c_μ such that, by the above definition of a_x together with the fact that $a_x(x) = 0$,

$$\theta = f_{I(X)}(Y) - f_{I(X)}(X) - \widehat{p}(X)\cdot(Y - X), \quad \text{on } \{Y \in J^\circ(X), X \notin N_\mu\}.$$

Now we prove the claim. Let $x \notin N_\mu$, and $y, y_1, ..., y_k \in J^\circ(x)$, for some $k \in \mathbb{N}$, such that $y = \sum_i \lambda_i y_i$, convex combination. Now consider

$$\mathbb{P} := \sum_i \delta_{(m_{I(x)}, y_i)} + \delta_{x,y}, \quad \text{and} \quad \mathbb{P}' := \sum_i \delta_{(x,y_i)} + \delta_{m_{I(x)},y}.$$

Notice that $\mathbb{P}$, and $\mathbb{P}'$ are competitors with finite supports, concentrated on N^c, by the fact that $m_{I(x)} \notin N_\mu$, together with Lemma 3.5.8, and the fact that J° is constant on $I(x)$ by Proposition 3.2.7. Therefore

$$\sum_i \lambda_i \theta(m_{I(x)}, y_i) + \theta(x,y) \quad = \quad \sum_i \lambda_i \theta(x, y_i) + \theta(m_{I(x)}, y), \tag{3.5.8}$$

from Definition 3.2.1 (ii). Then the proof that a_x is affine is similar to the proof of Lemma 3.5.6.

Let $\widehat{p}(x)$ be a vector in $\nabla \mathrm{aff} I(x)$ representing this linear form. By the fact that a_x is linear and finite on $J^\circ(x)$, we have the identity

$$\theta(x,y) = f_{I(x)}(y) - f_{I(x)}(x) - \widehat{p}(x)\cdot(y - x), \quad \text{for all } (x,y) \in \{Y \in J^\circ(X), X \notin N_\mu\}. \tag{3.5.9}$$

□

Recall that we want to find $f:\mathbb{R}^d\longrightarrow\mathbb{R}$, and $p:\mathbb{R}^d\longrightarrow\mathbb{R}^d$ such that $\theta=\mathbf{T}_pf$ on $\{Y\in J^\circ(X), X\notin N_\mu\}$. A good candidate for f would be f_I, in view of (3.5.9). However f defined this way could mismatch at the interface between two components. We now focus on the interface between components. Let $K,K'\in I(\mathbb{R}^d)$, we denote $\mathrm{interf}(K,K'):=J^\circ(m_K)\cap J^\circ(m_{K'})$ if $m_K,m_{K'}\notin N_\mu$, and $\emptyset$ otherwise.

Lemma 3.5.10. *Let* $(A_K)_{K\in I(\mathbb{R}^d)}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ *be such that*

$$A_K(y)-A_{K'}(y)=f_{K'}(y)-f_K(y),\quad \textit{for all } y\in\mathrm{interf}(K,K'),\quad \textit{and } K,K'\in I(\mathbb{R}^d) \tag{3.5.10}$$

Then $f(y):=f_K(y)+A_K(y)$ *does not depend of the choice of* K *such that* $y\in J^\circ(m_K)$*, and if we set* $p(y):=\widehat{p}(y)+\nabla A_{I(y)}$*, we have*

$$\theta=\mathbf{T}_pf,\quad \textit{on}\quad \{Y\in J^\circ(X),X\notin N_\mu\}.$$

Proof. Let $K,K'\in I(\mathbb{R}^d)$ such that $y\in J^\circ(m_K)\cap J^\circ(m_{K'})=\mathrm{interf}(K,K')$. Then $f_K(y)+A_K(y)-\Big(f_{K'}(y)+A_{K'}(y)\Big)=0$ by (3.5.10). The first point is proved.

Then $\mathbf{T}_pf=\mathbf{T}_{\widehat{p}+\nabla A_I}(f_I+A_I)=\mathbf{T}_{\widehat{p}}f_I+\mathbf{T}_{\nabla A_I}A_I=\mathbf{T}_{\widehat{p}}f_I$, where the last equality comes from the fact the A_I is affine in y. Then Lemma 3.5.9 concludes the proof. □

We now use Assumption 3.2.6 (ii) to prove the existence of a family $(A_K)_K$ satisfying the conditions of Lemma 3.5.10. Let $\mathcal{C}\subset\overline{\mathcal{K}}$, $\mathcal{D}\subset\overline{\mathcal{K}}$, and $\mathcal{R}\subset\overline{\mathcal{K}}$ from Assumption 3.2.6 such that $I(X)\in\mathcal{C}\cup\mathcal{D}\cup\mathcal{R}$, $\mu-$a.s. with $\mathcal{C}$ well ordered, $\dim(\mathcal{D})\subset\{0,1\}$, and $\cup_{K\neq K'\in\mathcal{R}}\Big[K\times(\mathrm{cl}\,K\cap\mathrm{cl}\,K')\Big]\in\mathcal{N}_{\mu,\nu}$.

Lemma 3.5.11. *We assume Assumption 3.2.6, and the existence of* $(T_K^{K'})_{K,K'\in\mathcal{C}\cup\mathcal{R}}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ *such that*
(i) $T_K^{K'}+T_{K'}^{K''}+T_{K''}^{K}=0$*, for all* $K,K',K''\in\mathcal{C}\cup\mathcal{R}$*;*
(ii) $T_K^{K'}(y)=f_{K'}(y)-f_K(y)$*, for all* $y\in\mathrm{interf}(K,K')$*,* $K,K'\in\mathcal{C}\cup\mathcal{R}$*.*

Then we may find $(A_K)_{K\in I(\mathbb{R}^d)}$ *satisfying the conditions of Lemma 3.5.10.*

Proof. We define A_K by for $K\in\mathcal{C}\cup\mathcal{R}$. If this set is non-empty, we fix $K_0\in\mathcal{C}\cup\mathcal{R}$. Let $K\in\mathcal{C}$, we set $A_K:=-T_{K_0}^{K}$.

Now for $K\in\mathcal{D}$, K has at most two end-points, let $x\in J^\circ(m_K)$ be an end-point of K. If $x\in J^\circ(m_{K'})$ for some $K'\in\mathcal{C}\cup\mathcal{R}$, then we set $A_K(x):=A_{K'}(x)+f_{K'}(x)-f_K(x)$. If $x\in J^\circ(m_{K'})$ for some $K'\in\mathcal{D}$, then we set $A_K(x):=-f_K(x)$. Otherwise, we set $A_K(x):=0$, and set A_K to be the only affine function on K that has the right values at the endpoints, and has a derivative orthogonal to K, which exists as K is at most one-dimensional.

We define $A_K = 0$ for all the remaining $K \in I(\mathbb{R}^d)$.

Now we check that $(A_K)_K$ satisfies the right conditions at the interfaces. Let $K, K' \in I(\mathbb{R}^d)$ such that $\mathrm{interf}(K, K') \neq \emptyset$. If $K \in \mathcal{D}$, or $K' \in \mathcal{D}$, the value at endpoints has been adapted to get the desired value. Now we treat the remaining case, we assume that $K, K' \in \mathcal{C} \cup \mathcal{R}$. We have $A_K - A_{K'} = -T_{K_0}^K + T_{K_0}^{K'}$. Property (i) applied to (K, K, K) implies that $T_K^K = 0$, and therefore, (i) applied to (K_0, K, K) gives that $T_{K_0}^K = T_K^{K_0}$. Finally, (i) applied to (K, K_0, K') gives that $A_K - A_{K'} = T_K^{K'}$. Finally, by (iii), we get that $A_K - A_{K'} = f_{K'}(y) - f_K(y)$ for all $y \in \mathrm{interf}(K, K')$. □

Lemma 3.5.12. *Let $K, K' \in I(\mathbb{R}^d)$, we have that $f_{K'} - f_K$ is affine finite on* $\mathrm{interf}(K, K')$.

Proof. First, by the fact that $\mathrm{interf}(K, K') \subset \mathrm{dom}\theta(m_K, \cdot) \cap \mathrm{dom}\theta(m_{K'}, \cdot)$, $a := f_{K'} - f_K$ is finite on $\mathrm{interf}(K, K')$. Now we prove that this map is affine, let $y_1, ..., y_k, y'_1, ..., y'_{k'} \in \mathrm{interf}(K, K')$ such that $y = \sum_i \lambda_i y_i = \sum_i \lambda'_i y'_i$, convex combinations. Then $\mathbb{P} := \frac{1}{2}\sum_i \lambda_i \delta_{(m_K, y_i)} + \frac{1}{2}\sum_i \lambda'_i \delta_{(m_{K'}, y'_i)}$, and $\mathbb{P}' := \frac{1}{2}\sum_i \lambda_i \delta_{(m_{K'}, y_i)} + \frac{1}{2}\sum_i \lambda'_i \delta_{(m_K, y'_i)}$ are competitors that are concentrated on $\{Y \in J^\circ(X), X \notin N_\mu\} \subset N^c$ by Lemma 3.5.8. Therefore, by Definition 3.2.1 (ii) we have $\sum_i \lambda_i \theta(m_K, y_i) + \sum_i \lambda'_i \theta(m_{K'}, y'_i) = \sum_i \lambda_i \theta(m_{K'}, y_i) + \sum_i \lambda'_i \theta(m_K, y'_i)$, which gives

$$\sum_i \lambda_i a(y_i) = \sum_i \lambda'_i a(y'_i).$$

Similar to the proof of Lemma 3.5.6, we have that a is affine on $\mathrm{interf}(K, K')$. □

Let $K, K' \in I(\mathbb{R}^d)$, by the preceding lemma $f_{K'} - f_K$ is affine finite on $\mathrm{interf}(K, K')$. If this set is not empty, let the unique $a_K^{K'} \in \nabla \mathrm{aff}\, \mathrm{interf}(K, K')$ and $b_K^{K'} \in \mathbb{R}$ such that

$$f_{K'}(y) - f_K(y) = a_K^{K'} \cdot y + b_K^{K'}, \quad \text{for} \quad y \in \mathrm{interf}(K, K').$$

We denote $H_K^{K'} : y \longmapsto a_K^{K'} \cdot y + b_K^{K'} \in \mathrm{Aff}(\mathbb{R}^d, \mathbb{R})$. If $\mathrm{interf}(K, K') = \emptyset$, we set $H_K^{K'} := 0$.

Lemma 3.5.13. *We may find $(T_K^{K'})_{K,K' \in \mathcal{C}\cup\mathcal{R}} \subset \mathrm{Aff}(\mathbb{R}^d, \mathbb{R})$ satisfying (i), and (ii) from Lemma 3.5.11 if and only if we may find $(\overline{H}_K^{K'})_{K,K' \in \mathcal{C}\cup\mathcal{R}} \subset \mathrm{Aff}(\mathbb{R}^d, \mathbb{R})$ such that $\overline{H}_K^{K'} = 0$ on* $\mathrm{interf}(K, K')$ *for all $K, K' \in \mathcal{C} \cup \mathcal{R}$, and for all triplet $(K_i)_{i=1,2,3} \in (\mathcal{C} \cup \mathcal{R})^3$ such that with the convention $K_4 = K_1$, we have*

$$\sum_{i=1}^{3} H_{K_i}^{K_{i+1}} + \overline{H}_{K_i}^{K_{i+1}} = 0. \tag{3.5.11}$$

Proof. We start with the necessary condition, let $(T_K^{K'})_{K,K' \in \mathcal{C}\cup\mathcal{R}} \subset \mathrm{Aff}(\mathbb{R}^d, \mathbb{R})$ satisfying (i), and (ii) from Lemma 3.5.11. Then for $K, K' \in \mathcal{C} \cup \mathcal{R}$, we introduce

$\overline{H}_K^{K'} := T_K^{K'} - H_K^{K'}$. By (ii), together with the definition of $H_K^{K'}$, we have $\overline{H}_K^{K'} = 0$ on $\operatorname{interf}(K,K')$. Now let a finite $(K,K',K'') \subset \mathcal{C} \cup \mathcal{R}$, by (ii) we have

$$H_K^{K'} + \overline{H}_K^{K'} + H_{K'}^{K''} + \overline{H}_{K'}^{K''} + H_{K''}^{K} + \overline{H}_{K''}^{K} = T_K^{K'} + T_{K'}^{K''} + T_{K''}^{K} = 0.$$

Now we prove the sufficiency. Let $(\overline{H}_K^{K'})_{K,K' \in \mathcal{C}\cup\mathcal{R}} \subset \operatorname{Aff}(\mathbb{R}^d,\mathbb{R})$ such that $\overline{H}_K^{K'} = 0$ on $\operatorname{interf}(K,K')$ for all $K,K' \in \mathcal{C}\cup\mathcal{R}$, and for all finite set $\mathcal{F} \subset \mathcal{C}\cup\mathcal{R}$, and all triplet $(K_i)_{i=1,2,3} \in (\mathcal{C}\cup\mathcal{R})^3$ such that we have $\sum_{i=1}^3 H_{K_i}^{K_{i+1}} + \overline{H}_{K_i}^{K_{i+1}} = 0$.

Then for $K,K' \in \mathcal{C}\cup\mathcal{R}$, let $T_K^{K'} := H_K^{K'} + \overline{H}_K^{K'}$. The property (ii) of $(T_K^{K'})$ follows from the fact that $T_K^{K'} = H_K^{K'} + \overline{H}_K^{K'} = H_K^{K'} = f_{K'} - f_K$ on $\operatorname{interf}(K,K')$.

Property (i) is a direct consequence of (3.5.11) with $(K,K',K'') \in (\mathcal{C}\cup\mathcal{R})^3$. □

Lemma 3.5.14. *Let $\mathcal{F} \subset I(N_\mu^c)$ finite, we may find $(\overline{H}_K^{K'})_{K,K' \in \mathcal{F}} \subset \operatorname{Aff}(\mathbb{R}^d,\mathbb{R})$ such that $\overline{H}_K^{K'} = 0$ on $\operatorname{interf}(K,K')$ for all $K,K' \in \mathcal{F}$, and for all triplet $(K_i)_{i=1,2,3} \in \mathcal{F}^3$ such that with the convention $K_4 = K_1$, we have $\sum_{i=1}^3 H_{K_i}^{K_{i+1}} + \overline{H}_{K_i}^{K_{i+1}} = 0$.*

Proof. Let $p \in \mathcal{F}^2$, we denote $H_p := H_{p_1}^{p_2}$, $\operatorname{interf}(p) := \operatorname{interf}(p_1,p_2)$, and the linear map $g_p : A \in \operatorname{Aff}(\mathbb{R}^d,\mathbb{R}) \longmapsto A_{|\operatorname{aff}\operatorname{interf}(p)} \in \operatorname{Aff}(\operatorname{aff}\operatorname{interf}(p),\mathbb{R})$. Let the linear map

$$g : (A_p)_{p\in\mathcal{F}^2} \in \operatorname{Aff}(\mathbb{R}^d,\mathbb{R})^{\mathcal{F}^2} \longmapsto \Big(g_p(A_p)\Big)_{p\in\mathcal{F}^2} \in \prod_{p\in\mathcal{F}^2} \operatorname{Aff}(\operatorname{aff}\operatorname{interf}(p),\mathbb{R}),$$

and if we denote $t_{i,j} := (t_i,t_j) \in \mathcal{F}^2$ for $t \in \mathcal{F}^3$ and $i,j \in \{1,2,3\}$, let the other linear map

$$f : (A_p)_{p\in\mathcal{F}^2} \in \operatorname{Aff}(\mathbb{R}^d,\mathbb{R})^{\mathcal{F}^2} \longmapsto \Big(A_{t_{1,2}} + A_{t_{2,3}} + A_{t_{3,1}}\Big)_{t\in\mathcal{F}^3} \in \operatorname{Aff}(\mathbb{R}^d,\mathbb{R})^{\mathcal{F}^3}.$$

Notice that the result may be written in terms of f and g as

$$f\Big((H_p)_{p\in\mathcal{F}^2}\Big) \in f(\ker g). \tag{3.5.12}$$

We prove this statement by using the monotonicity principle (ii) of Definition 3.2.1. Let the canonical basis $(e_j)_{1\le j\le d}$ of $\mathbb{R}^d$, and $e_0 := 0$ so that $(e_j)_{0\le j\le d}$ is an affine basis of $\mathbb{R}^d$, and the scalar product on $\operatorname{Aff}(\mathbb{R}^d,\mathbb{R})^{\mathcal{F}^3}$ defined by $\Big\langle (A_t)_{t\in\mathcal{F}^3}, (A'_t)_{t\in\mathcal{F}^3}\Big\rangle := \sum_{t\in\mathcal{F}^3, 0\le j\le d} A_t(e_j)A'_t(e_j)$. As the dimensions are finite, (3.5.12) is equivalent with the inclusion $f(\ker g)^\perp \subset \Big\{f\Big((H_p)_{p\in\mathcal{F}^2}\Big)\Big\}^\perp$.

Let $(A_t)_{t\in\mathcal{F}^3}\in f(\ker g)^\perp$, we now prove that $(A_t)_{t\in\mathcal{F}^3}\in\Big\{f\Big((H_p)_{p\in\mathcal{F}^2}\Big)\Big\}^\perp$, i.e. that

$$\begin{aligned}\Big\langle (A_t)_{t\in\mathcal{F}^3}, f\Big((H_p)_{p\in\mathcal{F}^2}\Big)\Big\rangle &= \sum_{t\in\mathcal{F}^3, 0\le j\le d} A_t(e_j)\Big(H_{t_{1,2}}(e_j)+H_{t_{2,3}}(e_j)+H_{t_{3,1}}(e_j)\Big)\\ &= 0.\end{aligned}$$

Let $p\in\mathcal{F}^2$, $\mathbf{P}_p := \mathrm{proj}_{\mathrm{aff}\,\mathrm{interf}(p)}$, and $0\le j\le d$. By the fact that $\mathbf{P}_p(e_j)\in\mathrm{aff}\,\mathrm{interf}(p) = \mathrm{aff}\Big(J(m_{p_1})\cap J(m_{p_2})\setminus N_\nu\Big)$ by Remark 3.2.8, we may find $(y_{i,j,p})_{1\le i\le d+1}\subset J(m_{p_1})\cap J(m_{p_2})\setminus N_\nu$, and $(\lambda_{i,j,p})_{1\le i\le d+1}\subset\mathbb{R}$ such that $\mathbf{P}_p(e_j) = \sum_{i=1}^{d+1}\lambda_{i,j,p}y_{i,j,p}$, affine combination, and $\sum_{i=1}^{d+1}\lambda_{i,j,p}=1$. Then with these ingredients we may give the expression of $H_p(e_j)$ as a function of values of θ:

$$\begin{aligned}H_p(e_j)=\sum_{i=1}^{d+1}\lambda_{i,j,p}H_p(y_{i,j,p}) &= \sum_{i=1}^{d+1}\lambda_{i,j,p}\left[\theta(m_{p_2},y_{i,j,p})-\theta(m_{p_1},y_{i,j,p})\right]\\ &= L_p^j[\theta],\end{aligned}$$

where $L_p^j := \sum_{i=1}^{d+1}\lambda_{i,j,p}\Big[\delta_{(m_{p_2},y_{i,j,p})}-\delta_{(m_{p_1},y_{i,j,p})}\Big]$ is a signed measure with finite support in $\{Y\in J(X)\setminus N_\nu, X\notin N_\mu\}$. We now study the marginals of L_p^j: we have obviously from its definition that $L_p^j[Y=y]=0$ for all $y\in\mathbb{R}^d$. For the X-marginals, $L_p^j[X=m_{p_2}] = -L_p^j[X=m_{p_1}] = \sum_{i=1}^{d+1}\lambda_{i,j,p}=1$, and $L_p^j[X=x]=0$ for all other $x\in\mathbb{R}^d$. Finally we look at its conditional barycenter:

$$L_p^j[Y|X=m_{p_2}] = -L_p^j[Y|X=m_{p_1}] = \sum_{i=1}^{d+1}\lambda_{i,j,p}y_{i,j,p} = \mathbf{P}_p(e_j). \tag{3.5.13}$$

Now let $t\in\mathcal{F}^3$, we denote $L_t^j := L_{t_{1,2}}^j + L_{t_{2,3}}^j + L_{t_{3,1}}^j$. We still have $L_t^j[Y=y]=0$ for all $y\in\mathbb{R}^d$ by linearity. Now

$$\begin{aligned}L_t^j[X=t_1] &= L_{t_{1,2}}^j[X=t_1]+L_{t_{2,3}}^j[X=t_1]+L_{t_{3,1}}^j[X=t_1]\\ &= -\mathbf{1}_{t_1=t_1}+\mathbf{1}_{t_2=t_1}-\mathbf{1}_{t_2=t_1}+\mathbf{1}_{t_3=t_1}-\mathbf{1}_{t_3=t_1}+\mathbf{1}_{t_3=t_1}\\ &= 0.\end{aligned}$$

Similar, $L_t^j[X=t_2]=L_t^j[X=t_3]=0$, and $L_t^j[X=x]=0$ for all $x\in\mathbb{R}^d$.

Notice that $\Big\langle (A_t)_{t\in\mathcal{F}^3}, f\Big((H_p)_{p\in\mathcal{F}^2}\Big)\Big\rangle = L[\theta]$, with $L:=\sum_{t\in\mathcal{F}^3,0\le j\le d}A_t(e_j)L_t^j$. By linearity, we have that

$$L[X=x]=L[Y=x]=0, \quad \text{for all} \quad x\in\mathbb{R}^d. \tag{3.5.14}$$

Furthermore, L is supported on $\{Y \in J(X) \setminus N_\nu, X \notin N_\mu\} \subset N^c$ like each L_p^j. We claim that $L[Y|X] = 0$, this claim will be justified at the end of this proof. Then we consider the Jordan decomposition $L = L_+ - L_-$ with L_+ the positive part of L and L_- its negative part. By the fact that $L[\mathbb{R}^d] = 0$, we have the decomposition $L = C(\mathbb{P}_+ - \mathbb{P}_-)$, for $C = L_+[R^d] = -L_-[R^d] \geq 0$. Then $\mathbb{P}_+$ and $\mathbb{P}_-$ are two finitely supported probabilities concentrated on N^c. By the fact that $L[X] = L[Y] = L[Y|X] = 0$, $\mathbb{P}_+$, and $\mathbb{P}_-$ are furthermore competitors, then by Definition 3.2.1 (ii), $\mathbb{P}_+[\theta] = \mathbb{P}_-[\theta]$, and therefore $\Big\langle (A_t)_{t\in\mathcal{F}^3}, f\Big((H_p)_{p\in\mathcal{F}^2}\Big)\Big\rangle = L[\theta] = 0$, which concludes the proof.

It remains to prove the claim that $L[Y|X] = 0$. Recall that $(A_t)_{t\in\mathcal{F}^3} \in f(\ker g)^\perp$. Let $K \in \mathcal{F}$ and $p \in \mathcal{F}^2$ such that $p_1 = K$, and $u \in \mathbb{R}^d$, the map $\xi_p : x \longmapsto u \cdot \Big(x - \mathbf{P}_p(x)\Big)$ is in $\ker g_p$. For all the other $p' \in \mathcal{F}^2$, we set $\xi_{p'} := 0 \in \ker g_{p'}$. Then $(\xi_p)_{p\in\mathcal{F}^2} \in \ker g$, and therefore $\Big\langle (A_t)_{t\in\mathcal{F}^3}, f\Big((\xi_p)_{p\in\mathcal{F}^2}\Big)\Big\rangle = 0$, we have

$$\begin{aligned} 0 &= \sum_{t\in\mathcal{F}^3, 0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} \mathbf{1}_{p_1=K} u \cdot \Big(e_j - \mathbf{P}_p(e_j)\Big) \\ &= u \cdot \sum_{t\in\mathcal{F}^3, 0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} \mathbf{1}_{p_1=K} \Big(e_j - \mathbf{P}_p(e_j)\Big). \end{aligned}$$

As this holds for all $u \in \mathbb{R}^d$, we have $\sum_{t\in\mathcal{F}^3,0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} \mathbf{1}_{p_1=K}\Big(e_j - \mathbf{P}_p(e_j)\Big) = 0$. Similarly, we have $\sum_{t\in\mathcal{F}^3,0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} \mathbf{1}_{p_2=K}\Big(e_j - \mathbf{P}_p(e_j)\Big) = 0$. Combining these two equations, and using (3.5.13) together with the definition of L we get

$$\begin{aligned} L[Y|X = m_K] &= \sum_{t\in\mathcal{F}^3, 0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} (\mathbf{1}_{p_2=K} - \mathbf{1}_{p_1=K}) \mathbf{P}_p(e_j) \\ &= \sum_{t\in\mathcal{F}^3, 0\leq j\leq d} A_t(e_j) \sum_{p=t_{1,2},t_{2,3},t_{3,1}} (\mathbf{1}_{p_2=K} - \mathbf{1}_{p_1=K}) e_j \\ &= L[X = m_K] e_j = 0, \end{aligned}$$

by (3.5.14) together with the definition of L. We conclude that $L[Y|X = m_K] = 0$, the claim is proved. □

Lemma 3.5.15. *Under Assumption 3.2.6, we may find* $(\overline{H}_K^{K'})_{K,K'\in\mathcal{C}\cup\mathcal{R}} \subset \mathrm{Aff}(\mathbb{R}^d, \mathbb{R})$ *such that* $\overline{H}_K^{K'} = 0$ *on* $\mathrm{interf}(K, K')$ *for all* $K, K' \in \mathcal{C}\cup\mathcal{R}$, *and for all triplet* $(K_i)_{i=1,2,3} \in (\mathcal{C}\cup\mathcal{R})^3$ *such that with the convention* $K_4 = K_1$, *we have* $\sum_{i=1}^3 H_{K_i}^{K_{i+1}} + \overline{H}_{K_i}^{K_{i+1}} = 0$.

Proof. We use the well-order of $\mathcal{C}$ from Assumption 3.2.6 to extend the result of Lemma 3.5.14 to the possibly infinite number of components. By the fact that $\mathcal{C}$ is well ordered, we have that $\mathcal{C}^2$ is also well ordered (we may use for example the lexicographic order based on the well-order of $\mathcal{C}$). We shall argue by transfinite induction on $\mathcal{C}^2$. For $(K,K')\in\mathcal{C}^2$, we denote $\mathcal{C}(K,K'):=\{(K_1,K_2)\in\mathcal{C}^2:(K_1,K_2)<(K,K')\}$. Finally we fix $\|\cdot\|$, a euclidean norm on the finite dimensional space $\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$, and for $(K,K')\in\mathcal{C}^2$, we define an order relation $\preceq_{K,K'}$ on $\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})^{\mathcal{C}(K,K')}$ which is the lexicographical order induced by $(\mathcal{C}(K,K'),\leq)$, and by the order on affine function $(\mathrm{Aff}(\mathbb{R}^d,\mathbb{R}),\preceq)$, defined by $A\preceq A'$ if $\|A\|\leq\|A'\|$. Our induction hypothesis is:

$\mathcal{H}(K,K')$: we may find a unique $(\overline{H}_{K_1}^{K_2})_{(K_1,K_2)\in\mathcal{C}(K)}$ such that:

(i) for all finite $\mathcal{F}\subset\mathcal{C}\cup\mathcal{R}$, we may find $(\widetilde{H}_{K_1}^{K_2})_{K_1,K_2\in\mathcal{F}}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ such that $\widetilde{H}_{K_1}^{K_2}=0$ on $\mathrm{interf}(K_1,K_2)$ for all $K_1,K_2\in\mathcal{F}$, such that for all triplet $(K_i)\in\mathcal{F}^3$ we have $\sum_{i=1}^3 H_{K_i}^{K_{i+1}}+\widetilde{H}_{K_i}^{K_{i+1}}=0$, and finally such that $\overline{H}_{K_1}^{K_2}=\widetilde{H}_{K_1}^{K_2}$ for all $(K_1,K_2)\in\mathcal{F}^2\cap\mathcal{C}(K,K')$;

(ii) for all $(K'',K''')\leq(K,K')$, $(\overline{H}_{K_1}^{K_2})_{K_1,K_2\in\mathcal{C}(K'',K''')}$ is the minimal vector satisfying (i) of $\mathcal{H}(K'',K''')$, for the order $\preceq_{K'',K'''}$.

Similar to the ordinals, we consider $\mathcal{C}^2$ as the upper bound of all the elements it contains, which gives a meaning to $\mathcal{H}(\mathcal{C}^2)$. The transfinite induction works similarly to a classical structural induction: let $(K_0,K_0')\in\mathcal{C}^2$ be the smallest element of $\mathcal{C}$, then the fact that $\mathcal{H}(K_0,K_0')$ holds, together with the fact that for all $(K,K')\in\mathcal{C}$, we have that $\mathcal{H}(K'',K''')$ holding for all $(K'',K''')<(K,K')$ implies that $\mathcal{H}(K,K')$ holds, then the transfinite induction principle implies that $\mathcal{H}(\mathcal{C}^2)$ holds.

The initialization is a direct consequence of Lemma 3.5.14 as $\mathcal{C}(K_0,K_0')=\emptyset$. Now let $(K,K')\in\mathcal{C}$, we assume that $\mathcal{H}(K',K'')$ holds for all $(K'',K''')<(K,K')$. Let $(K_1,K_1')<(K_2,K_2')<(K,K')$. As $\mathcal{H}(K_1,K_1')$, and $\mathcal{H}(K_2,K_2')$ hold, we may find unique $(\overline{H}_{K'}^{1,K''})_{K',K''\in\mathcal{C}(K_1,K_1')}$, and $(\overline{H}_{K'}^{2,K''})_{K',K''\in\mathcal{C}(K_2,K_2')}$ satisfying the conditions of the induction hypothesis. The restriction $(\overline{H}_{K'}^{2,K''})_{K',K''\in\mathcal{C}(K_1,K_1')}$ satisfies the conditions of $\mathcal{H}(K_1,K_1')$ by $\mathcal{H}(K_2,K_2')$, and by the fact that for the lexicographic order, if a word is minimal then all its prefixes are minimal as well for the sub-lexicographic orders. Therefore, by uniqueness in $\mathcal{H}(K_1,K_1')$, $(\overline{H}_{K'}^{1,K''})_{K',K''\in\mathcal{C}(K_1,K_1')}=(\overline{H}_{K'}^{2,K''})_{K',K''\in\mathcal{C}(K_1,K_1')}$. For all $(K'',K''')<(K,K')$ which are not predecessors of (K,K') (i.e. such that we may find $(K_{int},K_{int}')\in\mathcal{C}^2$ with $(K'',K''')<(K_{int},K_{int}')<(K,K')$), let $\overline{H}_{K'}^{K''}$ be the $(K',K'')-$th affine function of $(\overline{H}_{K_1}^{K_2})_{K_1,K_2\in\mathcal{C}(K_{int},K_{int}')}$ satisfying $\mathcal{H}(K_{int},K_{int}')$, which is unique by the preceding reasoning. If (K,K') has no predecessor, then $\overline{H}:=(\overline{H}_{K'}^{K'''})_{K'',K'''\in\mathcal{C}(K,K')}$ solves $\mathcal{H}(K,K')$. Now we treat the case in which we may find a predecessor $(K_{pred},K_{pred}')\in\mathcal{C}^2$ to (K,K'). In this case this predecessor is unique because $\mathcal{C}^2$ is well ordered. Then we consider $\overline{H}:=(\overline{H}_{K_1}^{K_2})_{K_1,K_2\in\mathcal{C}(K_{pred},K_{pred}')}$ from $\mathcal{H}(K_{pred},K_{pred}')$. Now we need to complete $\overline{H}$ by defining $\overline{H}_{K_{pred}}^{K_{pred}'}$.

For all finite $\mathcal{F}\subset\mathcal{C}\cup\mathcal{R}$, we define the affine subset $A_\mathcal{F}\subset \mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ of all $H\in \mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ such that $(\widetilde{H}_{K_1}^{K_2})_{K_1,K_2\in\mathcal{C}(K,K')}$ satisfies (i) of $\mathcal{H}(K,K')$, with $\widetilde{H}_{K_1}^{K_2}=\overline{H}_{K_1}^{K_2}$ if $(K_1,K_2)<(K_{pred},K'_{pred})$ and $\widetilde{H}_{K_{pred}}^{K'_{pred}}$. By $\mathcal{H}(K_{pred},K'_{pred})$ (i) applied to $\mathcal{F}\cup\{(K_{pred},K'_{pred})\}$, we have that $A_\mathcal{F}$ is non-empty for all $\mathcal{F}$. Then the intersection taken on finite sets $A:=\cap_{\mathcal{F}\subset\mathcal{C}\cup\mathcal{R}}A_\mathcal{F}$ is also non-empty as we intersect finite dimensional always non-empty affine spaces that have the property $A_{\mathcal{F}_1}\cap A_{\mathcal{F}_2}=A_{\mathcal{F}_1\cup\mathcal{F}_2}$. Then if we chose $H_{K_{pred}}^{K'_{pred}}\in A$, $\mathcal{H}(K_{pred},K'_{pred})$ will be verified, except for the minimality. To have the minimality, we chose the minimal $H\in A$ for the norm $\|\cdot\|$, which is unique as A is affine and the norm is Euclidean. This uniqueness, together with the uniqueness from the induction hypothesis gives the uniqueness for $\mathcal{H}_(K_{pred},K'_{pred})$ by properties of the lexicographic order. We proved $\mathcal{H}_(K_{pred},K'_{pred})$, and therefore $\mathcal{H}(\mathcal{C}^2)$ holds.

Finally, we need to include $\mathcal{R}$ in the indices of $\overline{H}$. Let the unique $(\overline{H}_K^{K'})_{(K,K')\in\mathcal{C}^2}$ from $\mathcal{H}(\mathcal{C}^2)$. Let $K\in\mathcal{R}$, $K'\in\mathcal{C}$. Similar to the step in the induction (K_{pred},K'_{pred}) to (K,K'), we may find a unique $\overline{H}_K^{K'}$ which satisfies the right relations and is minimal for the norm $\|\cdot\|$. As we may do it independently for all $(K,K')\in\mathcal{R}\times\mathcal{C}$ by the property of $\mathcal{R}$ in Assumption 3.2.6. For $K\in\mathcal{C}$ and $K'\in\mathcal{R}$, we set $\overline{H}_K^{K'}:=-\overline{H}_{K'}^{K}$. Finally for $K,K'\in\mathcal{R}$, if $\mathcal{C}=\emptyset$, then we set $\overline{H}_K^{K'}:=0$, else we set $\overline{H}_K^{K'}:=\overline{H}_{K_0}^{K'}-\overline{H}_{K_0}^{K}$ for some $K_0\in\mathcal{C}$. We may prove thanks to $\mathcal{H}(\mathcal{C}^2)$ that this definition does not depend on the choice of $K_0\in\mathcal{C}$, and that $\overline{H}$ defined this way on $(\mathcal{C}\cup\mathcal{R})^2$ satisfies the right conditions. □

Proof of Proposition 3.2.13 The inclusion $\supset$ is obvious from the definition of $\partial^{\mu,\nu}f$. We now prove the reverse inclusion by using Assumption 3.2.6. Then by Lemma 3.5.15, we may find $(\overline{H}_K^{K'})_{K\sim_1K'\in\mathcal{C}\cup\mathcal{R}}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ such that for all finite set $\mathcal{F}\subset\mathcal{C}\cup\mathcal{R}$, and all permutation $\sigma\in\mathcal{S}_\mathcal{F}$ such that $K\sim_1\sigma(K)$ for all $K\in\mathcal{F}$, we have $\sum_{K\in\mathcal{F}}H_K^{\sigma(K)}+\overline{H}_K^{\sigma(K)}=0$. Then, by Lemma 3.5.13, we may find $(T_K^{K'})_{K,K'\in I(\mathbb{R}^d):K\sim K'}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ satisfying (i), (ii), and (iii) from Lemma 3.5.11. Then we may apply Lemma 3.5.11: we may find $(A_K)_{K=I(x),x\in\mathbb{R}^d}\subset\mathrm{Aff}(\mathbb{R}^d,\mathbb{R})$ such that $A_K(y)-A_{K'}(y)=f_{K'}(y)-f_K(y)$ for all $y\in\mathrm{interf}(K,K')$, and for all $K,K'\in I(\mathbb{R}^d)$. Finally, by Lemma 3.5.10, $f(y):=f_K(y)+A_K(y)$ does not depend of the choice of K such that $y\in J^\circ(m_K)$, and if we set $p(y):=\widehat{p}(y)+\nabla A_{I(y)}$, we have

$$\theta=\mathbf{T}_pf,\quad\text{on}\quad\{X\notin N_\mu\}\cap\{Y\in J^\circ(X)\}.$$

Therefore, $\theta\approx\mathbf{T}_pf$, whence $f\in\mathcal{C}_{\mu,\nu}$ and we proved the reverse inclusion. □

Now, we prove the convexity of the functions in $\mathfrak{C}_{\mu,\nu}$ on each components.

Proof of Proposition 3.2.15 Let $p\in\partial^{\mu,\nu}f$, and $\theta\in\widetilde{\mathcal{T}}(\mu,\nu)$ such that $\mathbf{T}_pf=\theta$ on $\{Y\in J^\circ(X),X\notin N_\mu\}$ for a $N-$tangent convex function $\theta\in\widetilde{\mathcal{T}}(\mu,\nu)$, $N_\mu\in\mathcal{N}_\mu$, and $J^\circ(\mu,\nu)$. By proposition 3.2.7, we may chose N_μ and J° such that $\{Y\in J^\circ(X),X\notin N_\mu\}\subset\mathrm{dom}\theta\cap N^c$. For all $x\notin N_\mu$, and $y\in J^\circ(x)$, $f(y)=f(x)+p(x)\cdot(y-x)+\theta(x,y)$, which is clearly convex in y for x fixed. The function f is convex on J°, $\eta-$a.s.

For all $x\in N_\mu^c$ and $y\in J^\circ(x)$, we have $f(y)-f(x)-\mathrm{proj}_{\nabla\mathrm{aff}J^\circ}(p)(x)\cdot(y-x)=f(y)-f(x)-p(x)\cdot(y-x)=\theta(x,y)\geq 0$. Then by definition, $\mathrm{proj}_{\nabla\mathrm{aff}J^\circ}(p)(x)\in\partial f|_{J^\circ}(x)$ for all $x\notin N_\mu$.

For $x\in N_\mu^c$, we define $\widetilde{f}:=(f\mathbf{1}_{J^\circ(x)})_{conv}$ on $J(x)=\mathrm{conv}\Big(J^\circ(x)\Big)$, where the equality comes from Proposition 3.2.7 (i) together with the fact that $J\setminus N_\nu\subset J^\circ$. We also define $\widetilde{f}:=f$ on $\cap_{x\in N_\mu^c}J^\circ(x)^c\in\mathcal{N}_\mu\cap\mathcal{N}_\nu=\mathcal{N}_{\mu+\nu}$. These definitions are not interfering as if $x'\in J(x)$ then $J(x')\subset J(x)$ by Remark 3.2.8. Therefore, the convex envelops $(f\mathbf{1}_{J^\circ(x)})_{conv}$ and $(f\mathbf{1}_{J^\circ(x')})_{conv}$ coincide on $J(x')$.

Then the map $p(X)\cdot(Y-X)=f(Y)-f(X)-\theta(X,Y)$ is Borel measurable on $I(x)\times I(x)$ for all $x\in N_\mu^c$. Let $x\notin N_\mu$, $d_x:=\dim I(x)$, and $\Big(y_i\Big)_{1\leq i\leq d_x+1}\in I(x)$, affine basis of $\mathrm{aff}I(x)$. Therefore, $\mathrm{proj}_{\nabla\mathrm{aff}I(x)}\Big(p(x')\Big)=M^{-1}\Big(p(x')\cdot\Big(y_i-y_{d_x+1}\Big)\Big)_{1\leq i\leq d_x}$, with $M:=\Big(y_i-y_{d_x+1}\Big)_{1\leq i\leq d_x}$, where everything is expressed in the basis $(y_i-y_{d_x+1})_{1\leq i\leq d_x}$, is Borel measurable on $I(x)$. Then as it is a subgradient of $f_{|I(x)}$ on $I(x)$ by the fact that $\theta(x,y)=f(y)-f(x)-\mathrm{proj}_{\nabla\mathrm{aff}I(x)}\Big(p(x)\Big)\cdot(y-x)\geq 0$ for all $x,y\in I(x)$, we have the result.

Finally, notice that $\mathbf{T}_{\widetilde{p}}\widetilde{f}=\mathbf{T}_pf=\theta$ on $\{Y\in J^\circ(X),X\notin N_\mu\}$, which proves that $\widetilde{f}\in\mathfrak{C}_{\mu,\nu}$ and $\widetilde{p}\in\partial^{\mu,\nu}\widetilde{f}$. $\square$

Proof of Proposition 3.3.16 (i) Let $(\varphi,\psi,h)\in\widehat{\mathbb{L}}(\mu,\nu)$, and let f be its q.s.-convex moderator, and $p\in\partial^{\mu,\nu}f$. By Proposition 3.2.15, f is convex and finite on I, and $\mathrm{proj}_{\nabla\mathrm{aff}I}(p)\in\partial f_{|I}$, $\eta-$a.s. Then $\psi=(\psi-f)+f$ is Borel measurable on I, $\varphi=(\varphi+f)-f$ is Borel measurable on I, and $\mathrm{proj}_{\nabla\mathrm{aff}J}(h)=\mathrm{proj}_{\nabla\mathrm{aff}J}(h+p)-\mathrm{proj}_{\nabla\mathrm{aff}J}(p)$ is Borel measurable on I, $\eta-$a.s.

(ii) If one of the conditions in Proposition 3.3.10 holds, then condition (iv) holds by Proposition 3.3.10. Then the transfinite induction from the proof of Proposition 3.2.13 becomes a countable induction, thus preserving the measurability. The process of subtracting lines for the one dimensional components is also measurable. $\square$

3.5.7 Consequences of the regularity of the cost in x

Proof of Lemma 3.3.17 We have for all $x,y\in\mathbb{R}^d$, $\varphi(x)+\psi(y)+h(x)\cdot(y-x)\geq c(x,y)$. Then $\varphi(x)\geq\varphi'(x):=-(\psi-c(x,\cdot))_{conv}(x)$. For all $x\in\mathbb{R}^d$, $f_x:=(\psi-c(x,\cdot))_{conv}$ is convex and finite on $D:=\mathrm{ri}\,\mathrm{conv}\,\mathrm{dom}\,\varphi$, let $-h':\mathbb{R}^d\longmapsto\mathbb{R}^d$ be a measurable selection in its subgradient on D (then in $\mathrm{aff} D-x_0$ for some $x_0\in D$). Then for all $y\in\mathbb{R}^d$,

$$-\varphi'(x)-h'(x)\cdot(y-x)\leq f_x(y):=(\psi-c(x,\cdot))_{conv}(y)\leq\psi(y)-c(x,y).$$

Then $c\leq\varphi'\oplus\psi+h'^{\otimes}$, and therefore, $\mathbb{P}[\varphi'\oplus\psi+h'^{\otimes}]\geq\mathbb{P}[c]$ is well defined. Subtracting $\mathbb{P}[\varphi\oplus\psi+h^{\otimes}]<\infty$, we get

$$\mu[\varphi'-\varphi]=\mathbb{P}[(\varphi'-\varphi)(X)+(h'-h)^{\otimes}]\geq\mathbb{P}[c]-\mathbb{P}[\varphi\oplus\psi+h^{\otimes}].$$

Finally, taking the supremum over $\mathbb{P}$, we get $\mu[\varphi'-\varphi]\geq\mathbf{S}_{\mu,\nu}(c)-\mathbf{S}_{\mu,\nu}(\varphi\oplus\psi+h^{\otimes})=0$. As $\varphi'-\varphi\leq 0$, this shows that $\varphi'=\varphi$, $\mu-$a.e. Now

$$f_x(y)=-\inf\left\{\sum_{i=1}^r\lambda_i\Big(\psi(y_i)-c(x,y_i)\Big):\sum_{i=1}^r\lambda_iy_i=y,\text{ and } r\geq 1\right\}\tag{3.5.15}$$

For $r\geq 1$, and $y=\sum_{i=1}^r\lambda_iy_i$, $x\longmapsto\sum_{i=1}^r\lambda_i\Big(\psi(y_i)-c(x,y_i)\Big)$ is locally Lipschitz. By taking the infimum, we get that for $x\in D$, $f_x(y)$ is uniformly Lipschitz in x. Furthermore, f_x is convex on the relative interior of its domain D, and therefore locally Lipschitz on it. We claim that for the convex function f_x, the Lipschitz constant on a compact $K\subset D$ is bounded by $\frac{\max_{K'}f_x-\min_K f_x}{\delta}$, where $\delta=\inf_{(x,y)\in K\times K'}|x-y|$, for any compact $K'\subset D$ such that $K\subset\mathrm{ri}\,K'$ (cf proof of Theorem 9.3 in [58]). Then if we fix K and K', the Lipschitz constant of f_x is dominated on K as $x\longmapsto(\max'_K f_x,\min_K f_x)$ is Locally Lipschitz. Then for $K\subset D$ compact, we may find L, and L', Lipschitz constants for both variables. Finally, for $x_1,x_2\in B$,

$$|\varphi'(x_1)-\varphi'(x_2)|\leq|f_{x_1}(x_1)-f_{x_1}(x_2)|+|f_{x_1}(x_2)-f_{x_2}(x_2)|\leq(L+L')|x_1-x_2|.$$

In the proof of Theorem 9.3 in [58], the bound $\frac{\max'_K f_x-\min_K f_x}{\delta}$ is in fact a bound for the subgradients of f_x. As $-h'$ is a subgradient of f_x in x, its component in $\mathrm{aff} D-x_0$ (for some $x_0\in D$) is bounded in K.

3.6 Verification of Assumptions 3.2.6

3.6.1 Marginals for which the assumption holds

In preparation to prove Proposition 3.3.10, we first need to prove two lemmas.

Lemma 3.6.1. *Assume that there exists $\mathbb{Q} \in \mathcal{P}(\Omega)$ such that*

$$(\theta_n)_{n\geq 1} \subset \widetilde{\mathcal{T}}_1, \text{ converges } \mathcal{M}(\mu,\nu)-q.s. \text{ whenever } (\theta_n)_{n\geq 1}, \text{ converges } \mathbb{Q}-a.s. \quad (3.6.1)$$

Then for all $(\theta_n)_{n\geq 1} \subset \widetilde{\mathcal{T}}_1$, we may find $\theta \in \widetilde{\mathcal{T}}_1$ such that $\theta_n \rightsquigarrow \theta$.

Proof. Let $\mathbb{Q} \in \mathcal{P}(\Omega)$ satisfying (3.6.1). Let $\mathbb{Q}' := \frac{1}{2}\mathbb{Q} + \frac{1}{2}\mu(dx) \otimes \sum_{n\geq 1} 2^{-n}\delta_{f_n(x)}(dy)$, where $(f_n)_{n\geq 1} \subset \mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$ is chosen such that $\{f_n(x) : n \geq 1\} \subset \operatorname{aff} I(x)$ is dense in $\operatorname{aff} I(x)$ for all $x \in \mathbb{R}^d$ (see Step 2 in the proof of Proposition 2.7 in [58]). Then by Komlós lemma, we may find $\widehat{\theta}_n \in \operatorname{conv}(\theta_k, k \geq n)$ such that $\widehat{\theta}_n$ converges $\mathbb{Q}'-$a.s. Therefore, $\widehat{\theta}_n$ converges q.s. to $\theta := \underline{\widehat{\theta}}_\infty$. As $\widehat{\theta}_n \in \operatorname{conv}(\theta_k, k \geq n)$, we have the inequality $\underline{\widehat{\theta}}_\infty \geq \underline{\theta}_\infty$. We also have by Fatou's lemma $\mathbb{P}[\underline{\widehat{\theta}}_\infty] \leq \liminf_{n\to\infty} \mathbb{P}[\widehat{\theta}_n] \leq \limsup_{n\to\infty} \mathbb{P}[\theta_n]$, for all $\mathbb{P} \in \mathcal{P}(\Omega)$. Finally we need to prove that $\theta \in \Theta_{\mu,\nu}$. For $n \geq 1$, let $N_n \in \mathcal{N}_{\mu,\nu}$ be the set from Definition 3.2.1 for θ_n, and let $N_{cvg} \in \mathcal{N}_{\mu,\nu}$ be the set where $\widehat{\theta}_n$ does not converge. We set $N := \cup_{n\geq 1} N_n \cup N_{cvg} \in \mathcal{N}_{\mu,\nu}$. As $\theta_n(X,X) = 0$ for all $n \geq 1$, we have obviously $\{X = Y\} \subset N_{cvg}^c$, and $\{X = Y\} \subset N^c$. By convexity of $\theta_n(x,\cdot)$, the $\mu(dx) \otimes \sum_{n\geq 1} 2^{-n}\delta_{f_n(x)}(dy)-$convergence implies pointwise convergence of $\theta(X,\cdot)$ on $I(X)$, $\mu-$a.s. as in the case of $\mu\otimes$pw$-$convergence. Then $\theta(x,\cdot)$ is convex on N_x^c by passing to the limit, $I(X) \subset N_X^c$, $\mu-$a.s. By Lemma 6.1 in [58], we may chose $N_\mu \in \mathcal{N}_\mu$ so that if $\mathcal{N}_\mu \ni N'_\mu \supset N_\mu$, then $\{Y \in I(X)\} \cap \{X \in N'^c_\mu\}$ is a Borel set, and therefore, the function $\mathbf{1}_{\{Y\in I(X)\}\cap\{X\in N'^c_\mu\}}\theta_\infty$ is Borel and Definition 3.2.1 (iv) holds.

For $\mathbb{P}$ with finite support on N^c, and $\mathbb{P}'$ competitor to $\mathbb{P}$, $\mathbb{P}[\theta] = \lim_{n\to\infty} \mathbb{P}[\widehat{\theta}_n]$, and $\mathbb{P}'[\theta] \leq \liminf_{n\to\infty} \mathbb{P}'[\widehat{\theta}_n]$ by Fatou's Lemma. As for all $n \geq 1$, $\mathbb{P}[\widehat{\theta}_n] \geq \mathbb{P}'[\widehat{\theta}_n]$, we get the inequality $\mathbb{P}[\theta] \geq \mathbb{P}'[\theta]$. Furthermore, if we suppose to the contrary that $\{\omega\} := \operatorname{supp}\mathbb{P}' \cap N$ is a singleton, $\omega \notin N_n$ for all $n \geq 1$ by Definition 3.2.1 (iii). Then for all $n \geq 1$, $\mathbb{P}[\theta_n] = \mathbb{P}'[\theta_n]$, and $\mathbb{P}'[\omega]\theta_n(\omega) = \mathbb{P}[\theta_n] - \mathbb{P}'[\theta_n \mathbf{1}_{\Omega\setminus\{\omega\}}]$. Then as the term on the right of this equality converges, $\theta_n(\omega)$ converges as well, and $\omega \in N^c$. We got the contradiction, (iii) of Definition 3.2.1 holds. □

Lemma 3.6.2. *Assume that ν is dominated by the Lebesgue measure. Then $Y \notin \partial I(X)$ whenever $\dim I(X) \geq d-1$, $\mathcal{M}(\mu,\nu)-q.s.$*

Proof. First the components of dimension d are at most countable, and their boundary is Lebesgue negligible as they are convex. Then, if we enumerate the countable

$d-$dimensional components $(I_k)_{k\geq 1}$, we have $Y \notin \cup_{k\geq 1}\partial I_k$, $\nu-$a.s. and therefore $\mathcal{M}(\mu,\nu)-$q.s.

Now we deal with the $(d-1)-$dimensional components. I is a Borel map, and therefore by Lusin theorem (see Theorem 1.14 in [68]), for all $\epsilon > 0$, we may find $K_\epsilon \subset \{\dim I(X) = d-1\}$ with $\mu[K_\epsilon] \geq \mu[\dim I(X) = d-1] - \epsilon$, on which I is continuous. We may also assume that K_ϵ is compact. Then for all $x \in K_\epsilon$ such that $\dim I(x) = d-1$, $I(x)$ contains a closed $d-1-$dimensional ball $B_x := I(x) \cap B_{r_x}(x)$ for some $r_x > 0$. As I is continuous on K_ϵ, we may find $\epsilon_x > 0$ such that for $x' \in B_{\epsilon_x}(x)$, $B_x \subset \mathrm{proj}_{\mathrm{aff}I(x)}\Big(I(x')\Big)$, and such that the angle between the normals of $I(x)$ and $I(x')$ is smaller than $\eta := \pi/4 < \pi/2$. We denote l_x the line from x, normal to $I(x)$. The balls $B_{\epsilon_x}(x)$ cover K_ϵ, then by the compactness of K_ϵ, we may consider $x_1,...,x_k \in K_\epsilon$ for $k \geq 1$ such that $K_\epsilon \subset \cup_{i=1}^k B_{\epsilon_{x_i}}(x_i)$. Let $1 \leq i \leq k$, by Lemma C.1. in [74], we may find a bi-Lipschitz flattening map $F : \cup_{x'\in A_i} I(x') \longrightarrow \mathbb{R}^d = \mathrm{aff}I(x_i) \times l_{x_i}$, where $A_i := B_{\epsilon_{x_i}}(x_i) \cap l_{x_i}$, such that for all $x' \in A_i$ and all $(v,w) \in I(x')$, $F(v,w) = (v,x')$. Notice that for all $x' \in B_{\epsilon_{x_i}}(x_i)$, $I(x') \cap A_i \neq \emptyset$. Then for all $x' \in A_i$, $F\Big(I(x')\Big) \subset \mathrm{aff}I(x_i) \times \{x'\}$. Now, let λ be the Lebesgue measure. By the Fubini theorem, $\lambda\Big[F\Big(\cup_{x'\in A_i} \partial I(x')\Big)\Big] = \int_{l_x} \mathbf{1}_{x'\in A_i} \lambda_{x'}\Big[F\Big(\partial I(x')\Big)\Big]dx'$. By the facts that F is bi-Lipschitz, $\partial I(x')$ is Lebesgue-negligible in $\mathrm{aff}I(x')$, and $\lambda_{x'}$ is a $d-1-$dimensional Lebesgue measure, we have $\lambda_{x'}\Big[F\Big(\partial I(x')\Big)\Big] = 0$, $\mathbf{1}_{x'\in A_i} dx'-$a.e. Therefore, $\lambda\Big[F\Big(\cup_{x'\in A_i} \partial I(x')\Big)\Big] = 0$, and as F is bi-Lipschitz, $\lambda[\cup_{x'\in A_i}\partial I(x')] = 0$. Then summing up on all the $1 \leq i \leq k$ and by the fact that ν is dominated by the Lebesgue measure, we get $\nu[\cup_{x\in K_\epsilon}\partial I(x)] = 0$, so that for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, we have

$$\mathbb{P}[Y \in \partial I(X), \dim I(X) = d-1] \leq \mathbb{P}[X \notin K_\epsilon, \dim I(X) = d-1] + \mathbb{P}[Y \in \cup_{x\in K_\epsilon}\partial I(x)] \leq \epsilon.$$

As this holds for all $\epsilon > 0$ and for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, the lemma is proved. □

Proof of Proposition 3.3.10 Let us first prove the equivalence from (i). First for $\mathbb{P} \in \mathcal{M}(\mu,\nu)$. As $Y \in I(X)$, $\mathbb{P}$-a.s., we have $I(X) = I(Y)$, $\mathbb{P}-$a.s., and therefore, for all $A \in \mathcal{B}(\mathcal{K})$,

$$\nu \circ I^{-1}[A] = \mathbb{P}[I(Y) \in A] = \mathbb{P}[I(X) \in A] = \mu[I(X) \in A] = \mu \circ I^{-1}[A]$$

Conversely, suppose that $\mu \circ I^{-1} = \nu \circ I^{-1}$. We will prove by backward induction on $0 \leq k \leq d+1$ that $Y \in I(X)$, $\mathcal{M}(\mu,\nu)$-q.s., conditionally to $\dim I(X) \geq k$. For $k = d+1$ this is trivial because the dimension is lower than d. Now for $k \in \mathbb{N}$ we suppose that the property is true for $k' > k$. Then conditionally to $\dim I(x) = k$, we have that

$Y\in\operatorname{cl}I(X)$, q.s. Then for $\mathbb{P}\in\mathcal{M}(\mu,\nu)$,

$$\mathbb{P}[\dim I(Y)=k]=\mathbb{P}[Y\in I(X) \text{ and } \dim I(X)=k]+\mathbb{P}[Y\in\partial I(X) \text{ and } \dim I(X)>k]$$

By the induction hypothesis, $\mathbb{P}[Y\in\partial I(X) \text{ and } \dim I(X)>k]=0$. (i) gives that $\mathbb{P}[\dim I(Y)=k]=\mathbb{P}[\dim I(X)=k]$. Then

$$\mathbb{P}[\dim I(X)=k]=\mathbb{P}[Y\in I(X) \text{ and } \dim I(X)=k],$$

implying that $\mathbb{P}-a.s.$, $\dim I(X)=k \implies Y\in I(X)$. As holds true for all $\mathbb{P}\in\mathcal{M}(\mu,\nu)$, combined with the induction hypothesis, we proved the result at rank k. By induction, $Y\in I(X)$, q.s. The equivalence is proved.

It remains to show that (iv) is implied by all the other conditions. If (i) holds, $\cup_{x\in\mathbb{R}^d}I(x)\times\partial I(x)\in\mathcal{N}_{\mu,\nu}$ then (iv) holds with $\mathcal{C}=\mathcal{D}=\emptyset$, and $\mathcal{R}:=I(\mathbb{R}^d)$. If (ii) holds, as $I(\mathbb{R}^d)$ is a partition of $\mathbb{R}^d$, there can be at most countably many components with full dimension. Therefore (iv) holds with $\mathcal{C}:=I(\{\dim I=d\})$, and $\mathcal{D}:=\{\dim I\le 1\}$.

Now we suppose (iii), by Lemma 3.6.2, $Y\notin\partial I(X)$ if $\dim I(X)=d-1$, $\mathcal{M}(\mu,\nu)-$q.s. Then we just set $\mathcal{D}:=\{\dim I\le 1\}$, $\mathcal{C}:=\{\dim I=d\}$, and $\mathcal{R}:=\{\dim I=d-1\}$. Now we prove the claim.

We suppose that (iv) holds. The second part of the proposition follows from the fact that a countable set can be well ordered. Now let us deal with the first part. According to Lemma 3.6.1, we just need to find a probability measure $\mathbb{Q}$ that implies the quasi-sure convergence of functions in $\widetilde{\mathcal{T}}_1$. This is possible thanks to the convexity of these functions in the second variable: the interior of the components can be dealt with $\mu(dx)\otimes\sum_{n\ge1}2^{-n}\delta_{f_n(x)}(dy)$, where $(f_n)_{n\ge1}\subset\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$ is chosen such that $\{f_n(x):n\ge1\}\subset\operatorname{aff}I(x)$ is dense in $\operatorname{aff}I(x)$ for all $x\in\mathbb{R}^d$ (see the proof of Lemma 3.6.1).

For the boundaries, the measure $\mu\otimes\nu$ will deal with the countable components of $\mathcal{C}$. Indeed, let $K\in\mathcal{C}$ such that $\eta[K]>0$. Let $(\theta_n)_n\subset\widetilde{\mathcal{T}}_1$, converging $\mu(dx)\otimes\sum_{n\ge1}2^{-n}\delta_{f_n(x)}(dy)+\mu\otimes\nu-$a.s. to some function θ. We already have that $\theta_n(x,\cdot)\longrightarrow\theta(x,\cdot)$ on K for all $x\in N_\mu^c\cap K$, for some $N_\mu\in\mathcal{N}_\mu$ by the previous step. For all $n\ge1$, let $N_n\in\mathcal{N}_{\mu,\nu}$ be such that θ_n is a N_n-tangent convex function. By (3.2.5) and by possibly enlarging the $\mu-$null set N_μ, we may assume that we may find $(N_\nu,\theta)\in\mathcal{N}_\nu\times\widehat{\mathcal{T}}(\mu,\nu)$ such that $N_\mu^c\times N_\nu^c\cap\{Y\in J_\theta(X)\}\subset N^c:=(\cup_{n\ge1}N_n)^c$, and that $N_\mu^c\times\{Y\in I(X)\}\subset N^c$. Then for $x,x'\in N_\mu^c\cap K$, $x_0\in K$, and $y\in J_\theta(x)\setminus(K\cup N_\nu)$, let the probability measures

$$4\mathbb{P}:=\delta_{x,x_0}+\delta_{x,y}+2\delta_{x',y'},\quad\text{and}\quad 4\mathbb{P}':=\delta_{x',x_0}+\delta_{x',y}+2\delta_{x,y'}$$

with $y' := \frac{1}{2}(y+x_0)$. Let $n \geq 1$, notice that $\mathbb{P}$ and $\mathbb{P}'$ are competitors and concentrated on N_n, then by θ_n-martingale monotonicity of N_n, we have

$$\theta_n(x,x_0)+\theta_n(x,y)+2\theta_n(x',y') = \theta_n(x',x_0)+\theta_n(x',y)+2\theta_n(x,y').$$

We re-order the terms

$$\theta_n(x,y)-2\theta_n(x,y')+\theta_n(x,x_0) = \theta_n(x',y)-2\theta_n(x',y')+\theta_n(x',x_0). \qquad (3.6.2)$$

Then $\theta_n(x,y)-2\theta_n(x,y')+\theta_n(x,x_0)$ does not depend on the choice of $x \in K \cap N_\mu^c$. As we assumed that θ_n converges $\mu(dx)\otimes\sum_{n\geq 1}2^{-n}\delta_{f_n(x)}(dy)+\mu\otimes\nu-$a.s. by possibly enlarging N_μ, without loss of generality, we may assume that for all $x \in N_\mu^c$, $\theta_n(x,\cdot)$ converges pointwise to θ on $I(x)$, and $\theta_n(x,Y)$ converges $\nu-$a.s. Let $x' \in N_\mu^c \cap K$, up to enlarging N_ν, we may assume that $\theta_n(x',y)$ converges to $\theta(x',y)$ for all $y \in N_\nu^c$. Then if $x,y \in (K\cap N_\mu^c)\times N_\nu^c$, and $x \in K$, identity (3.6.2) implies that $\theta_n(x,y)$ converges, as all the other terms have a limit, and $\theta(x,y')$ and $\theta(x',y')$ are finite. Now for $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, $\mathbb{P}[(K\cap N_\mu^c)\times N_\nu^c] = \eta[K]$. Then θ_n converges $\mathbb{P}-$a.s. on $K\times\mathbb{R}^d$. This holds for all $K \in \mathcal{C}$, and $\mathbb{P} \in \mathcal{M}(\mu,\nu)$.

For the 1-dimensional components of $\mathcal{D}$, if we call $a(x)$ and $b(x)$ their (measurably selected) endpoints, the measure $\mu(dx)\otimes\frac{\delta_{a(x)}+\delta_{b(x)}}{2}$ will fit. Finally, in the case of the components in $\mathcal{R}$, for all probability $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, $\mathbb{P}_x$ does not send mass to ∂K for $\mu-$a.e. $x \in K \in \mathcal{R}$ by assumption. We take

$$\mathbb{Q}(dx,dy) := \mu(dx)\sum_{n\geq 1}2^{-n}\delta_{f_n(x)}(dy)+\mu(dx)\nu(dy)+\mu(dx)\frac{\delta_{a(x)}+\delta_{b(x)}}{2}(dy).$$

the convergence of θ_n, $\mathbb{Q}-$a.s. implies its convergence $\mathcal{M}(\mu,\nu)-$q.s. Assumption 3.2.6 holds. $\square$

Proof of Remark 3.3.12 The fact the $\nu_I^{\mathbb{P}}$ is independent of $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ for $d=1$ is proved by Beiglböck & Juillet [20].

Now we assume that (i) in Proposition 3.3.10 holds. If $Y \in I(X)$, $\mathcal{M}(\mu,\nu)-$q.s., then by symmetry as $\{I(x): x\in\mathbb{R}^d\}$ is a partition of $\mathbb{R}^d$, we have $X \in I(Y)$, $\nu-$a.s. Then similar to μ_I, $\nu_I := \nu_I^{\mathbb{P}} := \mathbb{P}\circ(Y|X\in I)^{-1}$ does not depend on the choice of $\mathbb{P} \in \mathcal{M}(\mu,\nu)$.

Now in the case of (iii) in Proposition 3.3.10, let $\nu_I := \nu\circ I^{-1}$. On $\{\dim I(X) \geq d-1\}$, $Y \notin \partial I(X)$, q.s. by Lemma 3.6.2, so that for all $\mathbb{P} \in \mathcal{M}(\mu,\nu)$, $\nu_I^{\mathbb{P}} = \nu_I$ on $\{\dim I(X) \geq d-1\}$. Now on $\{\dim I(X) = 0\}$, $\mu_I = \nu_I^{\mathbb{P}}$ is also independent of $\mathbb{P}$. Finally, on $\{\dim I(X)=1\}$, by the fact that there is not mass coming from higher dimensional components, we have $\nu_I^{\mathbb{P}}=\nu_I+\lambda_1(I)\delta_{a(I)}+\lambda_2(I)\delta_{b(I)}$, where $\lambda_1(I),\lambda_2(I)\geq 0$, and $a(I),b(I)$ are measurable selections of the boundary of I. Then $\mu_I-\nu_I=\lambda_1(I)+\lambda_2(I)$, and $\mu_I[X]-\nu_I[Y]=\lambda_1(I)a(I)+\lambda_2(I)b(I)$. Therefore, λ_1 and λ_2 depend only on μ_I and ν_I, therefore, $\nu_I^{\mathbb{P}}$ does not depend on the choice of $\mathbb{P}$. □

Proof of Remark 3.4.3 We consider τ the stopping time, and write $\mathbb{Q}$ the probability measure associated with the diffusion. We claim that the components $\overline{\text{supp}}\mathbb{P}_{X_0}\subset I(X_0)$, $\mu-$a.s. have dimension d, μ-a.s, where $\mathbb{P}\in\mathcal{M}(\mu,\nu)$ is the joint law of (X_0,X_τ). Then (iii) of Proposition 3.3.10 holds, which proves the remark.

Now we prove the claim. Let $p>0$. For $x\in\mathbb{R}^d$, we consider τ_x, the stopping time τ conditional to $X_0=x$, and σ_t^x, which is σ_t conditional to $X_0=x$. Now we fix $x\in\mathbb{R}^d$. As σ_0 has rank d, $\|\sigma_0^x\|:=\inf_{|u|=1}|u^t\sigma_0^x|>0$, a.s. Then we may find $\alpha>0$ such that $\mathbb{Q}[\|\sigma_0^x\|\leq\alpha]\leq p$. Similarly, we consider $\delta>0$ small enough so that

$$\mathbb{Q}[\tau<\delta]\leq p. \tag{3.6.3}$$

Finally, by the fact that σ_t^x is right-continuous in 0, a.s, we may lower $\delta>0$ so that $\mathbb{Q}\left[\sup_{t\leq\delta}|\sigma_t^x-\sigma_0^x|^2>\beta\right]\leq p$ for some $\beta>0$ that we will fix later. Now we use these ingredients to prove that (X_t) "spreads out in all directions" for t close to 0. Let $u\in\mathbb{R}^d$ with $|u|=1$ and $\lambda>0$,

$$\mathbb{Q}[u\cdot\sigma_0^x W_\delta\geq\lambda\alpha\sqrt{\delta}]\geq\mathbb{Q}[v\cdot W_1\geq\lambda]-p\geq\frac{1}{2}-2p, \tag{3.6.4}$$

with $v=u\cdot\sigma_0^x/|u\cdot\sigma_0^x|$, for λ small enough, independent of α and δ. Now recall that $\mathbb{Q}\left[\sup_{t\leq\delta}|\sigma_t^x-\sigma_0^x|^2>\beta\right]\leq p$. As a consequence, the stopping time $\tilde\tau=\inf\{t\geq 0:|\sigma_t^x-\sigma_0^x|^2\geq\beta\}$ satisfies

$$\mathbb{Q}[\tilde\tau<\delta]\leq p. \tag{3.6.5}$$

Now, stopping X_t, we get, conditionally to $X=x$: $\mathbb{E}^{\mathbb{Q}}[(\int_0^{\delta\wedge\tilde\tau}(\sigma_t^x-\sigma_0^x)dW_t)^2]\leq\delta\beta$ by Itô isometry, and therefore, by the Markov inequality, $\mathbb{Q}\left[\left|\int_0^{\delta\wedge\tilde\tau}(\sigma_t^x-\sigma_0^x)dW_t\right|\geq\alpha\lambda\sqrt{\delta}/2\right]\leq\frac{4\delta\beta}{\alpha^2\lambda^2\delta}$. Then if we chose $\beta=p\frac{\alpha^2\lambda^2}{4}$ (not depending on δ), we finally get that

$$\mathbb{Q}\left[\left|\int_0^{\delta\wedge\tilde\tau}(\sigma_t^x-\sigma_0^x)dW_t\right|\geq\alpha\lambda\sqrt{\delta}/2\right]\leq p. \tag{3.6.6}$$

Therefore $\mathbb{Q}[(X_{t\wedge\tau}-x)\cdot u \geq \alpha\lambda\sqrt{\delta}/2|X=x]$ is greater than

$$\mathbb{Q}\left[\sigma_0^x W_{t\wedge\tau}\cdot u \geq \alpha\lambda\sqrt{\delta},\text{ and }\left|\int_0^\delta(\sigma_t^x-\sigma_0^x)dW_t\right| \leq \alpha\lambda\sqrt{\delta}/2,\text{ and }\tilde{\tau}\geq\delta,\text{ and }\tau\geq\delta|X=x\right]$$
$$\geq \mathbb{Q}[u\cdot\sigma_0^x W_\delta \geq \lambda\alpha\sqrt{\delta}]-3p \geq \frac{1}{2}-5p,$$

by (3.6.3), (3.6.4), (3.6.5), and (3.6.6). Then by setting $p=\frac{1}{12}$, for all u of norm 1, we get

$$\mathbb{Q}[(X_{t\wedge\tau}-x)\cdot u \geq \alpha_0|X_0=x] \geq p_0, \tag{3.6.7}$$

with $\alpha_0 := \alpha\lambda\sqrt{\delta}/2>0$, and $p_0 := \frac{1}{12}>0$.

We use (3.6.7) to prove that $\overline{\text{supp}}\mathbb{P}_x$ is d dimensional. Indeed, we suppose for contradiction that $\overline{\text{supp}}\mathbb{P}_x\subset H$, where H is a hyperplane. H contains 0, as it contains $\overline{\text{supp}}\mathbb{P}_x$. Let u be a unit normal vector to H, by (3.6.7), we have $\mathbb{Q}[(X_{t\wedge\tau}-x)\cdot u\geq \alpha_0|X=x]\geq p_0$. Then by the martingale property (the volatility is bounded) combined with the boundedness of τ, we have $\mathbb{E}^{\mathbb{Q}}[X_\tau|\mathcal{F}_{t\wedge\tau}]=X_{t\wedge\tau}$. Therefore, $\mathbb{P}_x[Y\cdot u\geq\alpha_0/2]=\mathbb{Q}[X_\tau\cdot u\geq\alpha_0/2|X=x]>0$, which contradicts the inclusion of the support of $\mathbb{P}_x$ in H. □

3.6.2 Medial limits

Medial limits, introduced by Mokobodzki [120] (see also Meyer [119]), are powerful instruments. It is an operator from the set of real bounded sequences l^∞ to $\mathbb{R}$ satisfying the following properties:

Definition 3.6.3. *A linear operator* $\mathfrak{m}: l^\infty\to\mathbb{R}$ *is a medial limit if*
(i) $\mathfrak{m}$ *is nonnegative: if* $u\geq 0$ *then* $\mathfrak{m}(u)\geq 0$.
(ii) $\mathfrak{m}$ *is invariant by translation: if* $\mathcal{T}$ *is the translation operator (*$\mathcal{T}:(u_n)_n\mapsto(u_{n+1})_n$*) then* $\mathfrak{m}(\mathcal{T}u)=\mathfrak{m}(u)$.
(iii) $\mathfrak{m}((1)_n)=1$.
(iv) $\mathfrak{m}$ *is universally measurable on the unit ball* $[0,1]^{\mathbb{N}}$.
(v) $\mathfrak{m}$ *is measure linear: for any sequence of Borel-measurable functions* $f_n:[0,1]\to[0,1]$, *if we write* $f:=\mathfrak{m}((f_n)_n)$ *(defined pointwise), then for any Borel measure* λ *on* $[0,1]$, f *is* λ*-measurable and*

$$\int f d\lambda = \int \mathfrak{m}(f_n)d\lambda = \mathfrak{m}\left(\int f_n d\lambda\right).$$

We can extend any medial limit $\mathfrak{m}$ to $\mathbb{R}_+^{\mathbb{N}}$ by setting $\mathfrak{m}(u) := \sup_{N\in\mathbb{N}} \mathfrak{m}((u_n \wedge N)_n)$. It keeps the same properties, except (v) which becomes a kind of Fatou's Lemma: for any sequence of Borel-measurable functions $f_n : [0,1] \to \mathbb{R}_+$, then for any Borel measure λ on $[0,1]$,

$$\int \mathfrak{m}(f_n) d\lambda \;\; \leq \;\; \mathfrak{m}\Big(\int f_n d\lambda\Big). \tag{3.6.8}$$

The existence of medial limits is implied by Martin's axiom. Notice that Martin's axiom is implied by the continuum hypothesis (See Chapter I of Volume 5 of [72]). Kurt Gödel [75] provides 6 paradoxes implied by the continuum hypothesis, Martin's axiom implies only 3 of these paradoxes. All these axioms are undecidable either under ZF and under ZFC, indeed Paul Larson [109] proved that if ZFC is consistent, then ZFC+"there exists no medial limits" is also consistent (Corollary 3.3 in [109]). See [149] for a complete survey.

Proof of Proposition 3.3.14 Axiom of choice on $\mathbb{R}$ implies that $\mathbb{R}$ can be well-ordered, which proves that Assumption 3.2.6 (ii) holds. Now let us prove the first part. For $(\theta_n)_{n\geq 1} \subset \widetilde{\mathcal{T}}_1$, we denote $\theta := \mathfrak{m}(\theta_n)$. The Proposition is proved if we show that $\theta_n \rightsquigarrow \theta$. $\theta = \mathfrak{m}(\theta_n) \geq \underline{\theta}_\infty$ by linearity of a medial limit together with Definition 3.6.3 (i) and (ii). Let $\mathbb{P} \in \mathcal{P}(\Omega)$, $\mathbb{P}[\theta] \leq \mathfrak{m}(\mathbb{P}[\theta_n]) \leq \limsup_{n\to\infty} \mathbb{P}[\theta_n]$ by (3.6.8). Finally the linearity combined with Definition 3.6.3 (i) give that $\theta \in \Theta_{\mu,\nu}$, as it is a property of comparison of linear combinations of values of θ, θ is a $\emptyset-$tangent convex function. Finally, we prove that we may have (iv) in Definition 3.2.1. Up to assuming that we applied the Komlós Lemma to $(\theta_n)_{n\geq 1}$ (which only reduces the superior limits and increase the inferior limits, thus preserving the previous properties) under the probability $\mu(dx) \otimes \sum_{n\geq 1} 2^{-n} \delta_{f_n(x)}(dy)$, where $(f_n)_{n\geq 1} \subset \mathbb{L}^0(\mathbb{R}^d, \mathbb{R}^d)$ is chosen such that $\{f_n(x) : n \geq 1\} \subset \mathrm{aff} I(x)$ is dense in $\mathrm{aff} I(x)$ for all $x \in \mathbb{R}^d$ as in the proof of Lemma 3.6.1, we may assume without loss of generality that (θ_n) converges pointwise on $\{X \in N_\mu'^c\} \cap \{Y \in I(X)\}$. Then let $N_\mu^n \in \mathcal{N}_\mu$ be from Definition 3.2.1 (iv) for θ_n. Let $N_\mu = \cup_{n\geq 1} N_\mu^n \cup N_\mu'$. Let $A := \{X \in N_\mu^c\} \cap \{Y \in I(X)\}$, $\mathbf{1}_A \theta$ is Borel measurable as the pointwise limit of Borel measurable functions $\mathbf{1}_A \theta_n$, as the medial limit coincides with the real limit when convergence holds. $\square$

Chapter 4

Local structure of multi-dimensional martingale optimal transport

This paper analyzes the support of the conditional distribution of optimal martingale transport couplings between marginals in $\mathbb{R}^d$ for arbitrary dimension $d \geq 1$. In the context of a distance cost in dimension larger than 2, previous results established by Ghoussoub, Kim & Lim [74] show that this conditional distribution is concentrated on its own Choquet boundary. Moreover, when the target measure is atomic, they prove that the support of this distribution is concentrated on $d+1$ points, and conjecture that this result is valid for arbitrary target measure.

We provide a structure result of the support of the conditional distribution for general Lipschitz costs. Using tools from algebraic geometry, we provide sufficient conditions for finiteness of this conditional support, together with (optimal) lower bounds on the maximal cardinality for a given cost function. More results are obtained for specific examples of cost functions based on distance functions. In particular, we show that the above conjecture of Ghoussoub, Kim & Lim is not valid beyond the context of atomic target distributions.

Key words. Martingale optimal transport, local structure, differential structure, support.

4.1 Introduction

The problem of martingale optimal transport was introduced as the dual of the problem of robust (model-free) superhedging of exotic derivatives in financial mathematics, see Beiglböck, Henry-Labordère & Penkner [18] in discrete time, and Galichon, Henry-Labordère & Touzi [73] in continuous-time. Previously the robust superhedging problem was introduced by Hobson [94], and was addressing specific examples of exotic derivatives by means of corresponding solutions of the Skorokhod embedding problem, see [51, 92, 93], and the survey [91].

Our interest in the present paper is on the multi-dimensional martingale optimal transport. Given two probability measures μ,ν on $\mathbb{R}^d$, with finite first order moment, martingale optimal transport differs from standard optimal transport in that the set of all interpolating probability measures $\mathcal{P}(\mu,\nu)$ on the product space is reduced to the subset $\mathcal{M}(\mu,\nu)$ restricted by the martingale condition. We recall from Strassen [146] that $\mathcal{M}(\mu,\nu)\neq\emptyset$ if and only if $\mu\preceq\nu$ in the convex order, i.e. $\mu(f)\leq\nu(f)$ for all convex functions f. Notice that the inequality $\mu(f)\leq\nu(f)$ is a direct consequence of the Jensen inequality, the reverse implication follows from the Hahn-Banach theorem.

This paper focuses on showing the differential structure of the support of optimal probabilities for the martingale optimal transport Problem. In the case of optimal transport, a classical result by Rüschendorf [134] states that if the map $y\longmapsto c_x(x_0,y)$ is injective, then the optimal transport is unique and supported on a graph, i.e. we may find $T:\mathcal{X}\longrightarrow\mathcal{Y}$ such that $\mathbb{P}^*[Y=T(X)]=1$ for all optimal coupling $\mathbb{P}^*\in\mathcal{P}(\mu,\nu)$. The corresponding result in the context of the one-dimensional martingale transport problem was obtained by Beiglböck-Juillet [20], and further extended by Henry-Labordère & Touzi [85]. Namely, under the so-called martingale Spence-Mirrlees condition, c_x strictly convex in y, the left-curtain transport plan is optimal and concentrated on two graphs, i.e. we may find $T_d,T_u:\mathcal{X}\longrightarrow\mathcal{Y}$ such that $\mathbb{P}^*[Y\in\{T_d(X),T_u(X)\}]=1$ for all optimal coupling $\mathbb{P}^*\in\mathcal{M}(\mu,\nu)$. In this case we get similarly the uniqueness by a convexity argument.

An important issue in optimal transport is the existence and the characterization of optimal transport maps. Under the so-called twist condition (also called Spence-Mirrlees condition in the economics litterature) it was proved that the optimal transport is supported on one graph. In the context of martingale optimal transport on the line, Beiglböck & Juillet introduced the left-monotone martingale interpolating measure as a remarkable transport plan supported on two graphs, and prove its optimality for some classes of cost functions. Ghoussoub, Kim & Lim conjectured that in higher dimensional Martingale Optimal Transport for distance cost, the optimal plans will be supported on $d+1$ graphs. We prove here that there is no hope of extending this property beyond the case of atomic measure. This is obtained using the reciprocal property of the structure theorem of this paper, which serves as a counterexample generator. We further prove that for "almost all" smooth cost function, the optimal coupling are always concentrated on a finite number of graphs, and we may always find densities μ and ν that are dominated by the Lebesgue measure such that the optimal coupling is concentrated on $d+2$ maps for d even.

A first such study in higher dimension was performed by Lim [113] under radial symmetry that allows in fact to reduce the problem to one-dimension. A more "higher-dimensional specific" approach was achieved by Ghoussoub, Kim & Lim [74]. Their main structure result is that for the Euclidean distance cost, the supports of optimal kernels will be concentrated on their own Choquet boundary (i.e. the extreme points of the closure of their convex hull).

Our subsequent results differ from [74] from two perspectives. First, we prove that with the same techniques we can easily prove much more precise results on the local structure of the optimal Kernel, in particular, we prove that they are concentrated on $2d$ (possibly degenerate) graphs, which is much more precise than a concentration on the Choquet boundary. Our main structure result states that the optimal kernels are supported on the intersection of the graph of the partial gradient $c_x(x_0,\cdot)$ with the graph of an affine function $A_{x_0}\in \mathrm{Aff}_d$. Second, we prove a reciprocal property, i.e. that for any subset of such graph intersection $\{c_x(x_0,Y)=A(Y)\}$ for $A\in\mathrm{Aff}_d$, we may find marginals such that this set is an optimizer for these marginals. Thanks to this reciprocal property we prove that Conjecture 2 in [74] that we mentioned above is wrong. They prove this conjecture in the particular case in which the second marginal ν is atomic, however in view of our results it only works in this particular case, as we produce counterexamples in which μ and ν are dominated by the Lebesgue measure. Indeed, we prove that the support of the conditional kernel is characterized by an algebraic structure independent from the support of ν, then when this support is atomic, very particular phenomena happen. Thus the intuition suggests that finding this kind of solution for an atomic approximation of a non-atomic ν is not a stable approach, as in the limit there are generally $2d$ points in the kernel.

The paper is organized as follows. Section 4.2 gives the main results: Subsection 4.2.1 states the Assumption and the main structure theorem, Subsection 4.2.2 applies this theorem to show the relation between finiteness of the conditional support and the algebraic geometry of its derivatives, Subsection 4.2.3 gives the maximal cardinality that is universally reachable for the support up to choosing carefully the marginals, and finally Subsection 4.2.4 shows how the structure theorem applied to classical costs like powers of the Euclidean distance allows to give precise descriptions and properties of the conditional supports of optimal plans. Finally Section 4.3 contains all the proofs to the results in the previous sections, and Section 5.7 provides some numerical experiments.

Notation We fix an integer $d \geq 1$. For $x \in \mathbb{R}$, we denote $sg(x) := \mathbf{1}_{x>0} - \mathbf{1}_{x<0}$. If $f : \mathbb{R} \longrightarrow \mathbb{R}$ we denote by $\mathbf{fix}(f)$ the set of fixed points of f. A function $f : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is said to be super-linear if $\lim_{|y|\to\infty} \frac{|f(y)|}{|y|} = \infty$. Let a function $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ and $x_0 \in \mathbb{R}^d$, we say that f is super-differentiable (resp. sub-differentiable) at x_0 if we may find $p \in \mathbb{R}^d$ such that $f(x) - f(x_0) \leq p \cdot (x - x_0) + o(x - x_0)$ (resp. $\geq$) when $x \longmapsto x_0$, in this condition, we say that p belongs to the super-gradient $\partial^+ f(x_0)$ (resp. sub-gradient $\partial^- f(x_0)$) of f at x_0. This local notion extends the classical global notion of super-differential (resp. sub) for concave (resp. convex) functions.

For $x \in \mathbb{R}^d$, $r \geq 0$, and V an affine subspace of dimension d' containing x, we denote $\mathcal{S}_V(x,r)$ the $\dim V - 1$ dimensional sphere in the affine space V for the Euclidean distance, centered in x with radius r. We denote by Aff_d the set of Affine maps from $\mathbb{R}^d$ to itself. Let $A \in \mathrm{Aff}_d$, notice that its derivative ∇A is constant over $\mathbb{R}^d$, we abuse notation and denote ∇A for the matrix representation of this derivative. Let $M \in \mathcal{M}_d(\mathbb{R})$, a real matrix of size d, we denote $\det M$ the determinant of M, $\ker M$ is the kernel of M, $\mathrm{Im} M$ is the image of this matrix, and $Sp(M)$ is the set of all complex eigenvalues of M. We also denote $Com(M)$ the comatrix of M: for $1 \leq i,j \leq d$, $Com(M)_{i,j} = (-1)^{i+j} \det M^{i,j}$, where $M^{i,j}$ is the matrix of size $d-1$ obtained by removing the i^{th} line and the j^{th} row of M. Recall the useful comatrix formula:

$$Com(M)^t M = M Com(M)^t = (\det M) I_d. \tag{4.1.1}$$

As a consequence, whenever M is invertible, $M^{-1} = \frac{1}{\det M} Com(M)^t$. Throughout this paper, $\mathbb{R}^d$ is endowed with the Euclidean structure, the Euclidean norm of $x \in \mathbb{R}^d$ will be denoted $|x|$, the $p-$norm of x will be denoted $|x|_p := \left(\sum_{i=1}^d |x_i|^p\right)^{\frac{1}{p}}$. We denote $(e_i)_{1 \leq i \leq d}$ the canonical basis of $\mathbb{R}^d$. Let $B \subset E$ with E a vector space, we denote $B^* := B \setminus \{0\}$, and $|B|$ the possibly infinite cardinal of B. If V is a topological affine space and $B \subset V$ is a subset of V, $\mathrm{int} B$ is the interior of B, $\mathrm{cl}\, B$ is the closure of B, $\mathrm{aff} B$ is the smallest affine subspace of V containing B, $\mathrm{conv} B$ is the convex hull of B, $\dim(B) := \dim(\mathrm{aff} B)$, and $\mathrm{ri} B$ is the relative interior of B, which is the interior of B in the topology of $\mathrm{aff} B$ induced by the topology of V. We also denote by $\partial B := \mathrm{cl}\, B \setminus \mathrm{ri} B$

the relative boundary of B, and if V is endowed with a euclidean structure, we denote by $proj_B(x)$ the orthogonal projection of $x \in V$ on aff$B$. A set B is said to be discrete if it consists of isolated points.

We denote $\Omega := \mathbb{R}^d \times \mathbb{R}^d$ and define the two canonical maps

$$X : (x,y) \in \Omega \longmapsto x \in \mathbb{R}^d \quad \text{and} \quad Y : (x,y) \in \Omega \longmapsto y \in \mathbb{R}^d.$$

For $\varphi, \psi : \mathbb{R}^d \longrightarrow \bar{\mathbb{R}}$, and $h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, we denote

$$\varphi \oplus \psi := \varphi(X) + \psi(Y), \quad \text{and} \quad h^{\otimes} := h(X) \cdot (Y - X),$$

with the convention $\infty - \infty = \infty$.

For a Polish space $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\big(\mathcal{X}, \mathcal{B}(\mathcal{X})\big)$. For $\mathbb{P} \in \mathcal{P}(\mathcal{X})$, we denote by supp$\mathbb{P}$ the smallest closed support of $\mathbb{P}$. Let $\mathcal{Y}$ be another Polish space, and $\mathbb{P} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. The corresponding conditional kernel $\mathbb{P}_x$ is defined by:

$$\mathbb{P}(dx, dy) = \mu(dx)\mathbb{P}_x(dy), \text{ where } \mu := \mathbb{P} \circ X^{-1}.$$

Let $n \geq 0$ and a field $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$ in this paper), we denote $\mathbb{K}_n[X]$ the collection of all polynomials on $\mathbb{K}$ of degree at most n. The set $\mathbb{C}^{hom}[X]$ is the collection of homogeneous polynomials of $\mathbb{C}[X]$. Similarly for $k \geq 1$, we define $\mathbb{K}_n[X_1, ..., X_d]$ the collection of multivariate polynomials on $\mathbb{K}$ of degree at most n. We denote the monomial $X^\alpha := X_1^{\alpha_1} ... X_d^{\alpha_d}$, and $|\alpha| = \alpha_1 + ... + \alpha_d$ for all integer vector $\alpha \in \mathbb{N}^d$. For two polynomial P and Q, we denote gcd(P,Q) their greatest common divider. Finally, we denote $\mathbb{P}^d := \left(\mathbb{C}^{d+1}\right)^* / \mathbb{C}^*$ the projective plan of degree d.

The martingale optimal transport problem Throughout this paper, we consider two probability measures μ and ν on $\mathbb{R}^d$ with finite first order moment, and $\mu \preceq \nu$ in the convex order, i.e. $\nu(f) \geq \mu(f)$ for all integrable convex f. We denote by $\mathcal{M}(\mu,\nu)$ the collection of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mathbb{P} \circ X^{-1} = \mu$ and $\mathbb{P} \circ Y^{-1} = \nu$. Notice that $\mathcal{M}(\mu,\nu) \neq \emptyset$ by Strassen [146].

An $\mathcal{M}(\mu,\nu)-$polar set is an element of $\cap_{\mathbb{P} \in \mathcal{M}(\mu,\nu)} \mathcal{N}_{\mathbb{P}}$. A property is said to hold $\mathcal{M}(\mu,\nu)-$quasi surely (abbreviated as q.s.) if it holds on the complement of an $\mathcal{M}(\mu,\nu)-$polar set.

For a derivative contract defined by a non-negative cost function $c:\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}_+$, the martingale optimal transport problem is defined by:

$$\mathbf{S}_{\mu,\nu}(c) \quad := \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{P}[c]. \tag{4.1.2}$$

The corresponding robust superhedging problem is

$$\mathbf{I}_{\mu,\nu}(c) \quad := \inf_{(\varphi,\psi,h)\in\mathcal{D}_{\mu,\nu}(c)} \mu(\varphi)+\nu(\psi), \tag{4.1.3}$$

where

$$\mathcal{D}_{\mu,\nu}(c) \quad := \quad \Big\{(\varphi,\psi,h)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^1(\mu,\mathbb{R}^d):\ \varphi\oplus\psi+h^\otimes\geq c\Big\}. \tag{4.1.4}$$

The following inequality is immediate:

$$\mathbf{S}_{\mu,\nu}(c)\leq\mathbf{I}_{\mu,\nu}(c). \tag{4.1.5}$$

This inequality is the so-called weak duality. For upper semi-continuous cost, Beiglböck, Henry-Labordère, and Penckner [18], and Zaev [160] proved that strong duality holds, i.e. $\mathbf{S}_{\mu,\nu}(c)=\mathbf{I}_{\mu,\nu}(c)$. For any Borel cost function, De March [57] extended the quasi sure duality result to the multi-dimensional context, and proved the existence of a dual minimizer.

4.2 Main results

4.2.1 Main structure theorem

An important question in optimal transport theory is the structure of the support of the conditional distribution of optimal transport plans. Theorem 4.2.2 below gives a partial structure to this question. As a preparation we introduce a technical assumption.

We denote $\overline{\mathcal{K}}$ the collection of closed convex subsets of $\mathbb{R}^d$, which is a Polish space when endowed with the Wijsman topology (see Beer [16]). De March & Touzi [58] proved that we may find a Borel mapping $I:\mathbb{R}^d\longmapsto\overline{\mathcal{K}}$ such that $\{I(x):x\in\mathbb{R}^d\}$ is a partition of $\mathbb{R}^d$, $Y\in\mathrm{cl}\,I(X)$, $\mathcal{M}(\mu,\nu)-$a.s. and $\mathrm{cl}\,I(X)=\mathrm{cl}\,\mathrm{conv}\,\mathrm{supp}\hat{\mathbb{P}}_X$, $\mu-$a.s. for some $\hat{\mathbb{P}}\in\mathcal{M}(\mu,\nu)$. As the map I is Borel, $I(X)$ is a random variable, let $\eta:=\mu\circ I^{-1}$ be the push forward of μ by I. It was proved in [57] that the optimal transport disintegrates on all the "components" $I(X)$. The following conditions are needed throughout this paper.

Assumption 4.2.1. (i) *$c:\Omega\longmapsto\mathbb{R}$ is upper semi-analytic, $\mu\preceq\nu$ in convex order in $\mathcal{P}(\mathbb{R}^d)$, $c\geq\alpha\oplus\beta+\gamma^{\otimes}$ for some $(\alpha,\beta,\gamma)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$, and $\mathbf{S}_{\mu,\nu}(c)<\infty$.*
(ii) *The cost c is locally Lipschitz and sub-differentiable in the first variable $x\in I$, uniformly in the second variable $y\in\mathrm{cl}\, I$, $\eta-$a.s.*
(iii) *The conditional probability $\mu_I:=\mu\circ\Big(X|I\Big)^{-1}$ is dominated by the Lebesgue measure on I, $\eta-$a.s.*

The statements (i) and (ii) of Assumption 4.2.1 are verified for example if c is differentiable and if μ and ν are compactly supported. On another hand, the statement (iii) is much more tricky. It is well known that Sudakov [147] thought that he had solved the Monge optimal transport problem by using the (wrong) fact that the disintegration of the Lebesgue measure on a partition of convex sets would be dominated by the Lebesgue measure on each of these convex sets. However, [10], provides a counterexample inspired from another paradoxal counterexample by Davies [54]. This Nikodym set N is equal to the tridimensional cube up to a Lebesgue negligible set. Furthermore it is designed so that a continuum of mutually disjoint lines which intersect all N in one singleton each. Thus the Lebesgue measure on the cube disintegrates on this continuum of lines into Dirac measures on each lines.

Statement (iii) is implied for example by the domination of μ by the Lebesgue measure together with the fact that $\dim I(X)\in\{0,d-1,d\}$, $\mu-$a.s. (see Lemma C.1 of [74] implying that the Lebesgue measure disintegrates in measures dominated by Lebesgue on the $d-1-$dimensional components), in particular together with the fact that $d\leq 2$, or together with the fact that ν is the law of $X_\tau:=X_0+\int_0^t\sigma_s dW_s$, where $X_0\sim\mu$, W a $d-$dimensional Brownian motion independent of X_0, τ is a positive bounded stopping time, and $(\sigma_t)_{t\geq 0}$ is a bounded cadlag process with values in $\mathcal{M}_d(\mathbb{R})$ adapted to the $W-$filtration with σ_0 invertible. See the proof of Remark 4.3 in [57].

Theorem 4.2.2. (i) *Under Assumption 4.2.1 we may find $(A_x)_{x\in\mathbb{R}^d}\subset\mathrm{Aff}_d$ such that for all $\mathbb{P}^*\in\mathcal{M}(\mu,\nu)$ optimal for (4.1.2),*

$$x\in\mathrm{ri\,conv\,supp}\,\mathbb{P}^*_x,\quad \textit{and}\quad \mathrm{supp}\,\mathbb{P}^*_x\subset\{c_x(x,Y)=A_x(Y)\}\quad \textit{for } \mu-\textit{a.e. } x\in\mathbb{R}^d.$$

(ii) *Conversely, let a compact $S_0\subset\{c_x(x_0,Y)=A(Y)\}$ for some $x_0\in\mathbb{R}^d$ and $A\in\mathrm{Aff}_d$, be such that $x_0\in\mathrm{int\,conv}\,S_0$, c is $\mathrm{C}^{2,0}\cap\mathrm{C}^{1,1}$ in the neighborhood of $\{x_0\}\times S_0$, and $c_{xy}(S_0)-\nabla A\subset GL_d(\mathbb{R})$, then S_0 has a finite cardinal $k\geq d+1$ and we may find*

$\mu_0, \nu_0 \in \mathcal{P}(\mathbb{R}^d)$ *with* C^1 *densities such that*

$$\mathbb{P}^*(dx,dy) := \mu_0(dx)\sum_{i=1}^k \lambda_i(x)\delta_{T_i(x)}(dy)$$

is the unique solution to (4.1.2)*, with* $(T_i)_{1\le i\le k} \subset \mathrm{C}^1(\operatorname{supp}\mu_0, \mathbb{R}^d)$ *such that* $S_0 = \{T_i(x_0)\}_{1\le i\le k}$*, and* $(\lambda_i)_{1\le i\le k} \subset \mathrm{C}^1(\operatorname{supp}\mu_0)$.

Remark 4.2.3. *We have* $\nabla A_x(x) = \nabla\varphi(x) - h(x)$ *in Theorem 4.2.2 from its proof. Under the stronger assumption that* φ *and* h *are* C^1*, we can get this result much easier. As for* $(x,y) \in \mathbb{R}^d$,

$$\varphi(x) + \psi(y) + h(x)\cdot(y-x) - c(x,y) \ge 0,$$

with equality for $(x,y) \in \Gamma$*. When* y_0 *is fixed,* x_0 *such that* $(x_0,y_0) \in \Gamma$ *is a critical point of* $x \mapsto \varphi(x) + \psi(y_0) + h(x)\cdot(y_0 - x) - c(x,y_0)$*. Then we get* $c_x(x_0,y_0) = \nabla h(x_0)(y_0 - x_0) + \nabla\varphi(x_0) - h(x_0)$ *by the first order condition.*

We see that we have in this case $A_{x_0}(y) := \nabla h(x_0)(y - x_0) + \nabla\varphi(x_0) - h(x_0)$*, and* $\Gamma_{x_0} \subset \{c_x(x_0,Y) = A_{x_0}(Y)\}$*, for* $\mu-$*a.e.* $x_0 \in \mathbb{R}^d$.

Remark 4.2.4. *Even though the set* $S_0 := \{c_x(x_0,Y) = A(Y)\}$ *for* $x_0 \in \mathbb{R}^d$ *and* $A \in \mathrm{Aff}_d$ *may contain more than* $d+1$ *points, it is completely determined by* $d+1$ *affine independent points* $y_1, ..., y_{d+1} \in S_0$*, as the equations* $c_x(x_0,y_i) = A(y_i)$ *determine completely the affine map* A.

Proof of Theorem 4.2.2 (i) By Theorem 3.5 (i) in [57], (and using the notation therein), the quasi-sure robust super-hedging problem may be decomposed in pointwise robust super-hedging separate problems attached to each components, and we may find functions $(\varphi,h) \in \mathbb{L}^0(\mathbb{R}^d)\times\mathbb{L}^0(\mathbb{R}^d,\mathbb{R}^d)$, and $(\psi_K)_{K\in I(\mathbb{R}^d)} \subset \mathbb{L}^0_+(\mathbb{R}^d)$ with $\psi_{I(X)}(Y) \in \mathbb{L}^0_+(\Omega)$, and $\operatorname{dom}\psi_I = J_\theta$, $\eta-$a.s. for some $\theta \in \widehat{\mathcal{T}}(\mu,\nu)$, such that $c \le \varphi(X) + \psi_{I(X)}(Y) + h^\otimes$, and $\mathbf{S}_{\mu,\nu}(c) = \mathbf{S}_{\mu,\nu}\Big(\varphi(X) + \psi_{I(X)}(Y) + h^\otimes\Big)$. Then applying the theorem to $c' := \varphi(X) + \psi_{I(X)}(Y) + h^\otimes$, $\mathbf{S}_{\mu,\nu}(c) = \mathbf{S}_{\mu,\nu}\Big(\varphi(X) + \psi_{I(X)}(Y) + h^\otimes\Big) = \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\mathbf{S}_{\mu_I,\nu_I^{\mathbb{P}}}\Big(\varphi(X) + \psi_{I(X)}(Y) + h^\otimes\Big)$. Then if $\mathbb{P}\in\mathcal{M}(\mu,\nu)$ is optimal for $\mathbf{S}_{\mu,\nu}(c)$, then $\mathbb{P}_I[c = \varphi\oplus\psi_I + h^\otimes] = 1$, $\eta-$a.s. By Lemma 3.17 in [57] the regularity of c in Assumption 4.2.1 (ii) guarantees that we may chose φ to be locally Lipschitz on I, and h locally bounded on I. In view of Assumption 4.2.1 (iii), φ is differentiable μ_I-a.e. by the Rademacher Theorem. Then after possibly restricting to an irreducible component, we may suppose that we have the following duality: for any $x,y \in \mathbb{R}^d$,

$$\varphi(x) + \psi(y) + h(x)\cdot(y-x) - c(x,y) \;\;\ge\;\; 0, \qquad (4.2.6)$$

with equality if and only if $(x,y)\in\Gamma:=\{\varphi\oplus\psi+h^{\otimes}=c<\infty\}$, concentrating all optimal coupling for $\mathbf{S}_{\mu,\nu}(c)$.

Let $x_0\in\mathrm{ri\,conv\,dom}\,\psi$ such that φ is differentiable in x_0. Let $y_1,...,y_k\in\Gamma_{x_0}$ such that $\sum_{i=1}^k\lambda_iy_i=x_0$, convex combination. We complete $(y_1,...,y_k)$ in a barycentric basis $(y_1,...,y_k,y_{k+1},...,y_l)$ of $\mathrm{ri\,conv\,dom}\,\psi$. Let $x\in\mathrm{ri\,conv\,dom}\,\psi$ in the neighborhood of x_0, and let (λ_i') such that $x=\sum_{i=1}^l\lambda_i'y_i$, convex combination. We apply (4.2.6), both in the equality and in the inequality case:

$$\varphi(x)+\sum_{i=1}^l\lambda_i'\psi(y_i)\geq\sum_{i=1}^l\lambda_i'c(x,y_i),\quad \varphi(x_0)+\textstyle\sum_{i=1}^l\lambda_i'\psi(y_i)+h(x_0)\cdot(x-x_0)=\sum_{i=1}^l\lambda_i'c(x_0,y_i).$$

By subtracting these equations, we get

$$\varphi(x)-\varphi(x_0)-h(x_0)\cdot(x-x_0)\quad\geq\quad\sum_{i=1}^l\lambda_i'\Big(c(x,y_i)-c(x_0,y_i)\Big).$$

As c is Lipschitz in x, and $\lambda_i'\longrightarrow\lambda_i$ when $x\to x_0$, we get:

$$\Big(\nabla\varphi(x_0)-h(x_0)\Big)\cdot(x-x_0)+o(x-x_0)\geq\sum_{i=1}^k\lambda_i\Big(c(x,y_i)-c(x_0,y_i)\Big).$$

Then, $x\longmapsto\sum_{i=1}^k\lambda_ic(x,y_i)$ is super-differentiable at x_0, and $\nabla\varphi(x_0)-h(x_0)$ belongs to its super-gradient. As $x\longmapsto c(x,y)$ is sub-differentiable by Assumption 4.2.1 (ii), it implies that $x\longmapsto c(x,y_i)$ is differentiable at x_0 for all i such that $\lambda_i>0$, and therefore

$$\nabla\varphi(x_0)-h(x_0)\quad=\quad\sum_{i=1}^k\lambda_ic_x(x_0,y_i).\tag{4.2.7}$$

Now we want to prove that we may find $A_x\in\mathrm{Aff}_d$ such that $A_x(y)=c_x(x,y)$ for all $y\in\Gamma_x$.

Let $y_1^0,...,y_m^0\in\Gamma_{x_0}$ generating $\mathrm{aff}\Gamma_{x_0}$ and such that $x\in\mathrm{ri\,conv}(y_1^0,...,y_m^0)$, let $y\in\Gamma_{x_0}$. A_x is defined in a unique way if $\nabla A=0$ on $(\mathrm{aff}\Gamma_{x_0}-x_0)^\perp$ by its values on $(y_1^0,...,y_m^0)$. Now we prove that $A_x(y)=c_x(x_0,y)$. As $y\in\mathrm{aff}(y_1^0,...,y_m^0)$, we may find (μ_i) so that $\sum_{i=1}\mu_iy_i^0=y$, and $\sum_{i=1}\mu_i=1$. For $\varepsilon>0$ small enough, $x_0-\varepsilon(y-x_0)\in\mathrm{ri\,conv}(y_1^0,...,y_m^0)$. Then $x_0-\varepsilon(y-x_0)=\sum_{i=1}\lambda_i^0y_i$ with $\lambda_i^0>0$. We take the convex combination: $x_0=\frac{1}{1+\varepsilon}(x_0-\varepsilon(y-x_0))+\frac{\varepsilon}{1+\varepsilon}y$, and $x_0=\sum_{i=1}\Big(\frac{1}{1+\varepsilon}\lambda_i^0+\frac{\varepsilon}{1+\varepsilon}\mu_i\Big)y_i^0$. We suppose that ε is small enough so that $\lambda_i^\varepsilon:=\frac{1}{1+\varepsilon}\lambda_i^0+\frac{\varepsilon}{1+\varepsilon}\mu_i>0$. Then applying (4.2.7) for $(y_i) = (y_i^0)$ and $(\lambda_i) = (\lambda_i^\varepsilon)$,

$$\nabla\varphi(x_0) - h(x_0) = \sum_{i=1}^{l} \lambda_i^\varepsilon c_x(x_0, y_i) = \sum_{i=1}^{l} \frac{1}{1+\varepsilon}\lambda_i c_x(x_0, y_i) + \frac{\varepsilon}{1+\varepsilon} c_x(x_0, y).$$

By subtracting, we get $c_x(x_0, y) = A_{x_0}\Big(\frac{1+\varepsilon}{\varepsilon}\sum_{i=1}^{l}(\lambda_i^\varepsilon - \frac{1}{1+\varepsilon}\lambda_i)y_i\Big) = A_{x_0}(y)$. Now doing this for all $x \in \mathbb{R}^d$ so that φ is differentiable in x, by domination of μ_I by Lebesgue, this holds for μ_I-a.e. $x \in \mathbb{R}^d$, $\eta-$a.s. and therefore $\mu-$a.s.

(ii) Now we prove the converse statement. Let $S_0 \subset \{A(Y) = c_x(x_0, Y)\}$ be a closed bounded subset of Ω for some $x_0 \in \mathbb{R}^d$, and $A \in \mathrm{Aff}_d$ such that $x_0 \in \operatorname{int}\operatorname{conv} S_0$, c is $\mathrm{C}^{2,0} \cap \mathrm{C}^{1,1}$ in the neighborhood of S_0, and $c_{xy}(S_0) - \nabla A \subset GL_d(\mathbb{R})$. First, we show that S_0 is finite. Indeed, we suppose to the contrary that $|S_0| = \infty$, we can find a sequence $(y_n)_{n\geq 1} \subset S_0$ with distinct elements. As S_0 is closed bounded, and therefore compact, we may extract a subsequence $(y_{\varphi(n)})$ converging to $y_l \in S_0$. We have $c_x(x_0, y_{\varphi(n)}) = A(y_{\varphi(n)})$, and $c_x(x_0, y_l) = A(y_l)$. We subtract and get $c_x(x_0, y_{\varphi(n)}) - c_x(x_0, y_l) - \nabla A(y_{\varphi(n)} - y_l) = 0$, and using Taylor-Young around y_l, $c_{xy}(x_0, y_l)(y_{\varphi(n)} - y_l) + o(|y_{\varphi(n)} - y_l|) - \nabla A(y_{\varphi(n)} - y_l) = 0$. As $y_\varphi(n) \neq y_l$ for n large enough , we may write $u_n := \frac{y_{\varphi(n)} - y_l}{|y_{\varphi(n)} - y_l|}$. As u_n stands in the unit sphere which is compact, we can extract a subsequence $(u_{\psi(n)})$, converging to a unit vector u. As we have $c_{xy}(x_0, y_l)u_{\psi(n)} + o(1) - \nabla A u_{\psi(n)} = 0$, we may pass to the limit $n \to \infty$, and get:

$$(c_{xy}(x_0, y_l) - \nabla A)u = 0.$$

As $u \neq 0$, we get the contradiction: $c_{xy}(x_0, y) - \nabla A \notin GL_d(\mathbb{R})$.

Now, we denote $S_0 = \{y_i\}_{1\leq i\leq k}$ where $k := |S_0|$. For $r > 0$ small enough, the balls $\overline{B}\Big((x_0, y_i), r\Big)$ are disjoint, $c_{xy}(\cdot) - \nabla A \subset GL_d(\mathbb{R})$ on these balls by continuity of the determinant, and c is $\mathrm{C}^{2,0} \cap \mathrm{C}^{1,1}$ on these balls. Now we define appropriate dual functions. Let $M > 0$ large enough so that on the balls, $(M-1)I_d - (\nabla A + \nabla A^t) - c_{xx}$ is positive semidefinite.

We set $h(X) := \nabla A(X - x_0) - A(x_0)$, and $\varphi(X) := \frac{1}{2}M|X - x_0|^2$. Now for $1 \leq i \leq k$, $c_x(x_0, y_i) - \nabla A \cdot (y_i - x_0) = \nabla\varphi(x_0) - h(x_0)$, $(x, y) \longmapsto c_x(x, y) - \nabla A \cdot (y - x)$ is C^1, and its partial derivative with respect to y, $c_{xy} - \nabla A$ is invertible on the balls. Then by the implicit functions Theorem, we may find a mapping $T_i \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ such that for $x \in \mathbb{R}^d$ in the neighborhood of x_0,

$$c_x\Big(x, T_i(x)\Big) - \nabla A \cdot \Big(T_i(x) - x\Big) = \nabla\varphi(x) - h(x). \tag{4.2.8}$$

Its gradient at x_0 is given by $\nabla T_i(x_0) = \Big(c_{xy}(x_0,y_i) - \nabla A\Big)^{-1}\Big(MI_d - (\nabla A + \nabla A^t) - c_{xx}(x_0,y_i)\Big)$. This matrix is invertible, and therefore by the local inversion theorem T_i is a C^1-diffeomorphism in the neighborhood of x_0. We shrink the radius r of the balls so that each T_i is a diffeomorphism on $B := X\Big(\overline{B}\Big((x_0,y_i),r\Big)\Big)$ (independent of i). Let $B_i := T_i(B)$, for $y \in B_i$, let $\psi(y) := c\Big(T_i^{-1}(y),y\Big) - \varphi\Big(T_i^{-1}(y)\Big) - h\Big(T_i^{-1}(y)\Big)\cdot\Big(y - T_i^{-1}(y)\Big)$. These definitions are not interfering, as we supposed that the balls B_i are not overlapping.

Let $\Gamma := \{(x,T_i(x)) : x \in B, 1 \le i \le k\}$. By definition of ψ, $c = \varphi \oplus \psi + h^{\otimes}$ on Γ. Now let $(x,y) \in B \times B_i$, for some i. $(x_0,y) \in \Gamma$, for some $x_0 \in B$. Let $F := \varphi \oplus \psi + h^{\otimes} - c$, we prove now that $F(x,y) \ge 0$, with equality if and only if $x = x_0$ (i.e. $(x,y) \in \Gamma$). $F(x_0,y) = 0$, and $F_x(x_0,y) = 0$ by (4.2.8). However, $F_{xx}(X,Y) = MI_d - (\nabla A + \nabla A^t) - c_{xx}(X,Y)$ which is positive definite on $B \times B_i$, and therefore we get

$$\begin{aligned} F(x,y) = F(x,y) - F(x_0,y) &= \int_{x_0}^{x} F_x(z,y)\cdot dz = \int_{x_0}^{x}\Big(F_x(z,y) - F_x(x_0,y)\Big)\cdot dz \\ &= \int_{x_0}^{x}\int_{x_0}^{z} dw\cdot F_{xx}(w,y)\cdot dz \ge 0. \end{aligned}$$

Where the last inequality follows from the fact that F_{xx} is positive definite and dw and dz are two vectors collinear with $(x - x_0)$. It also proves that $F(x,y) = 0$ if and only if $(x,y) \in \Gamma$.

Now, we define C^1 mappings $\lambda_i : B \longrightarrow (0,1]$ such that $\sum_{i=1}^k \lambda_i(x)T_i(x) = x$. We may do this because we assumed that $x \in \operatorname{int}\operatorname{conv} S_0$, and therefore, by continuity, up to reducing B again, $x \in \operatorname{int}\operatorname{conv}\{T_1(x),...,T_k(x)\}$ for all $x \in B$. Finally let $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$ such that $\operatorname{supp}\mu_0 = B$ with C^∞ density f (take for example a well chosen wavelet). Now for $1 \le i \le k$, we define ν_0 on B_i by $\nu_0(dy) = \lambda_i\Big(T_i^{-1}(y)\Big)f\Big(T_i^{-1}(y)\Big)\Big|\det\nabla T_i\Big(T^{-1}(y)\Big)\Big|^{-1}$. Then $\mathbb{P}^*(dx,dy) := \mu_0(dx)\otimes\sum_{i=1}^k \lambda_i(x)\delta_{T_i(x)}(dy)$ is supported on Γ, is in $\mathcal{M}(\mu_0,\nu_0)$. As φ, and ψ are continuous, and therefore bounded, and as μ_0 and ν_0 are compactly supported, $\mathbb{P}^*[c] = \mu_0[\varphi] + \nu_0[\psi]$, and therefore $\mathbb{P}^*$ is an optimizer for $\mathbf{S}_{\mu_0,\nu_0}(c)$.

Now we prove that this is the only optimizer. Let $\mathbb{P}$ be an optimizer for $\mathbf{S}_{\mu_0,\nu_0}(c)$. Then $\mathbb{P}[\Gamma] = 1$, and therefore $\mathbb{P}(dx,dy) = \mu_0(dx)\otimes\sum_{i=1}^k \gamma_i(x)\delta_{T_i(x)}(dy)$, for some mappings γ_i. Let $1 \le i \le k$, as for $y \in B_i$, there is only one $x := T_i^{-1}(y) \in B$ such that $(x,y) \in \Gamma$. Then we may apply the Jacobian formula: $\nu_0(dy) = \gamma_i\Big(T_i^{-1}(y)\Big)f\Big(T_i^{-1}(y)\Big)\Big|\det\nabla T_i\Big(T^{-1}(y)\Big)\Big|^{-1}$. As this density in also equal to $\nu_0(dy) = \lambda_i\Big(T_i^{-1}(y)\Big)f\Big(T_i^{-1}(y)\Big)\Big|\det\nabla T_i\Big(T^{-1}(y)\Big)\Big|^{-1}$, and as $f\Big(T_i^{-1}(y)\Big)\Big|\det\nabla T_i\Big(T^{-1}(y)\Big)\Big|^{-1} > 0$, we deduce that $\lambda_i\Big(T^{-1}(Y)\Big) = \gamma_i\Big(T^{-1}(Y)\Big)$, ν_0-a.s. and $\lambda_i = \gamma_i$, μ_0-a.s. and therefore $\mathbb{P} = \mathbb{P}^*$. □

The statement (i) of Theorem 4.2.2 is well known, it is already used in [85] (to establish Theorem 5.1), [20] (see Theorem 7.1), and [74] (for Theorem 5.5). However, the converse implication (ii) is new and we will show in the next subsections how it gives crucial information about the structure of martingale optimal transport for classical cost functions. This converse implication will serve as a counterexample generator, similar to counterexample 7.3.2 in [20], which could have been found by an immediate application of the converse implication (ii) in Theorem 4.2.2.

Beiglböck & Juillet [20] and Henry-Labordère & Touzi [85] solved the problem in dimension 1 for the distance cost or for costs satisfying the "Spence-Mirless condition" (i.e. $\frac{\partial^3}{\partial x \partial y^2} c > 0$), in these particular cases, the support of the optimal probabilities is contained in two points in y for x fixed. See also Beiglböck, Henry-Labordère & Touzi [19]. Some more precise results have been provided by Ghoussoub, Kim, and Lim [74]: they show that for the distance cost, the image can be contained in its own Choquet boundary, and in the case of minimization, they show that in some particular cases the image consists of $d+1$ points, which provides uniqueness. They conjecture that this remains true in general. The subsequent theorem will allow us to prove that this conjecture is wrong, and that the properties of the image can be found much more precisely.

4.2.2 Algebraic geometric finiteness criterion

Completeness at infinity of multivariate polynomial families

Algebraic geometry is the study of algebraic varieties, which are the sets of zeros of families of multivariate polynomials. When the cost c is smooth, the set $\{c_x(x_0, Y) = A(Y)\}$ for $x_0 \in \mathbb{R}^d$ and $A \in \mathrm{Aff}_d$, behaves locally as an algebraic variety. This statement is illustrated by Proposition 4.2.12 and Theorem 4.2.18.

Let $k, d \in \mathbb{N}$ and $(P_1, ..., P_k)$ be k polynomials in $\mathbb{R}[X_1, ..., X_d]$. We denote $\langle P_1, ..., P_{i-1}\rangle$ the ideal generated by $(P_1, ..., P_{i-1})$ in $\mathbb{R}[X_1, ..., X_d]$ with the convention $\langle \emptyset \rangle = \{0\}$, and P^{hom} denotes the sum of the terms of P which have degree $\deg(P)$:

$$\text{If } P(X) = \sum_{|\alpha| \le \deg P} a_\alpha X^\alpha, \quad \text{then} \quad P^{hom}(X) := \sum_{|\alpha| = \deg P} a_\alpha X^\alpha.$$

Definition 4.2.5. *Let $k,d \in \mathbb{N}$ and $(P_1,...,P_k)$ be k multivariate polynomials in $\mathbb{R}[X_1,...,X_d]$. We say that the family $(P_1,...,P_k)$ is complete at infinity if*

$$QP_i^{hom} \notin \langle P_1^{hom},...,P_{i-1}^{hom}\rangle, \text{ for all } Q \notin \langle P_1^{hom},...,P_{i-1}^{hom}\rangle, \text{ for } 1 \le i \le k.$$[1]

Remark 4.2.6. *This notion actually means that the intersection of the zeros of the polynomials P_i in the points at infinity in the projective space has dimension $d-k-1$ (with the convention that all negative dimensions correspond to $\emptyset$), or equivalently by the correspondance from Corollary 1.4 of [82], that $P_1^{hom},...,P_d^{hom}$ is a regular sequence of $\mathbb{R}[X_1,...,X_d]$, see page 184 of [82]. See Proposition 4.3.3 to understand why $P_1^{hom},...,P_d^{hom}$ may be seen as the projections of $P_1,...,P_d$ at infinity. The algebraic geometers rather say that the algebraic varieties defined by the polynomials intersect completely at infinity. The ordering of the polynomials in Definition 4.2.5 does not matter. Notice that $P_1,...,P_d$ is a regular sequence if $P_1^{hom},...,P_d^{hom}$ is a regular sequence, therefore the completeness at infinity of $(P_i)_{1\le i\le k}$ implies that the intersection of the zeros of the polynomials in the points in the projective space has dimension $d-k$.*

Remark 4.2.7. *Notice that in Definition 4.2.5, we restrict to $\mathbb{R}[X_1,...,X_d]$, whereas the algebraic geometry results that we will use apply with the same definition where we need to replace $\mathbb{R}[X_1,...,X_d]$ by $\mathbb{C}[X_1,...,X_d]$. However, the families (P_i) that we will consider here stem from Taylor series of smooth cost functions. Therefore we only consider $(P_i) \subset \mathbb{R}[X_1,...,X_d]$, and we notice that in this case, Definition 4.2.5 is equivalent with $\mathbb{R}[X_1,...,X_d]$ or with $\mathbb{C}[X_1,...,X_d]$, up to projecting on the real or on the imaginary part of the equations.*

Example 4.2.8. *If $d \in \mathbb{N}^*$ and $k \in (\mathbb{N}^*)^d$ Then $(X_1^{k_1},...,X_d^{k_d})$ is complete. Indeed, let $1 \le i \le d$, $\langle X_1^{k_1},...,X_{i-1}^{k_{i-1}}\rangle = \{X_1^{k_1}P_1 + ... + X_{i-1}^{k_{i-1}}P_{i-1}, P_1,...,P_{i-1} \in \mathbb{R}[X_1,...,X_d]\}$. Notice that for this family of polynomials, $P \in \langle X_1^{k_1},...,X_{i-1}^{k_{i-1}}\rangle$ is equivalent to $\partial_{X^l}P(X_1 = 0,...,X_{i-1} = 0, X_i,...,X_d) = 0$ for all $l \in \mathbb{N}^d$ such that $l_j < k_j$ for $j < i$, and $l_j = 0$ for $j \ge i$. Let $Q \in \mathbb{R}[X_1,...,X_d]$ such that $QX_i^{k_i} \in \langle X_1^{k_1},...,X_{i-1}^{k_{i-1}}\rangle$, then for all such $l \in \mathbb{N}^d$, we have $\partial_{X^l}(QX_i^{k_i})(X_1 = 0,...,X_{i-1} = 0, X_i,...,X_d) = X_i^{k_i}\partial_{X^l}Q(X_1 = 0,...,X_{i-1} = 0, X_i,...,X_d) = 0$, and therefore $\partial_{X^l}Q(X_1 = 0,...,X_{i-1} = 0, X_i,...,X_d) = 0$, implying that $Q \in \langle X_1^{k_1},...,X_{i-1}^{k_{i-1}}\rangle$.*

The notion is also invariant by linear change of variables. For example, $(X^3 + XY + 3, Y^3 - X^2 + X)$ is complete at infinity because the homogeneous polynomial family (X^3, Y^3) is complete at infinity by Example 4.2.8 above.

[1] In algebraic terms this means that P_i^{hom} is not a divider of zero in the quotient ring $\mathbb{R}[X_1,...,X_d]/\langle P_1^{hom},...,P_{i-1}^{hom}\rangle$.

Example 4.2.9. *Let $d\in\mathbb{N}$ and (P_1,P_2) be 2 homogeneous polynomials in $\mathbb{R}[X_1,...,X_d]\setminus\mathbb{R}$, then (P_1,P_2) is complete at infinity if and only if* $\gcd(P_1,P_2)=1$*. Indeed, if* $\gcd(P_1,P_2)=\Pi\neq 1$*, then $P_1/\Pi\notin\langle P_1\rangle$ but $P_1/\Pi P_2=P_1P_2/\Pi\in\langle P_1\rangle$, and therefore (P_1,P_2) is not complete at infinity. Conversely, if (P_1,P_2) is non complete at infinity, we may find $P',Q\in\mathbb{R}[X_1,...,X_d]$ such that $Q\notin\langle P_1\rangle$ and $QP_2=P_1P'$. We assume for contradiction that* $\gcd(P_1,P_2)=1$*, then P_1 is a divider of Q, and $Q\in\langle P_1\rangle$, whence the contradiction.*

Let $k,d\in\mathbb{N}$ and $(P_1,...,P_k)$ be k homogeneous polynomials in $\mathbb{R}[X_0,X_1,...,X_d]$, we define the set of common zeros of $(P_1,...,P_k)$: $Z(P_1,...,P_k)=\{x\in\mathbb{P}^d: P_i(x)=0, \text{ for all } 1\le i\le k\}$. An element $x\in\mathbb{R}^d$ is a single common root of $P_1,...,P_k$ if $x\in Z(P_1,...,P_k)$, and the vectors $\nabla P_i(x)$ are linearly independent in $\mathbb{R}^d$.

Remark 4.2.10. *Let $k\in(\mathbb{N}^*)^d$. It is well known by algebraic geometers that we may find a polynomial equation system $T\in\mathbb{R}\Big[(X_{i,j})_{1\le i\le d,j\in(\mathbb{N}^*)^d:|j|\le k_i}\Big]$ such that for all $(P_1,...,P_d)\in\prod_{i=1}^d\mathbb{R}_{k_i}[X_1,...,X_d]$ with $P_i=\sum_{j\in(\mathbb{N}^*)^d:|j|\le k_i}a_{i,j}X_1^{j_1}...X_d^{j_d}$, we have the equivalence*

$$T\Big((a_{i,j})_{1\le i\le d,j\in(\mathbb{N}^*)^d:|j|\le k_i}\Big)\neq 0 \iff (P_1,...,P_d) \text{ is complete at infinity.}$$

We provide a proof of this statement in Subsection 4.3.1. Furthermore, not all multivariate polynomials families $(P_1,...,P_d)\in\prod_{i=1}^d\mathbb{R}_{k_i}[X_1,...,X_d]$ are solution of T as shows Example 4.2.8. As a consequence T is non-zero and we have that almost all (in the sense of the Lebesgue measure) homogeneous polynomial family is complete at infinity.

Criteria for finite support of conditional optimal martingale transport

We start with the one dimensional case. We emphasize that the sufficient condition (i) below corresponds to a local version of [85].

Theorem 4.2.11. *Let $d=1$ and let $S_0=\{c_x(x_0,Y)=A(Y)\}$, for some $A\in\mathrm{aff}(\mathbb{R},\mathbb{R})$, such that $x_0\in\mathrm{ri\,conv}S_0$, and $c:\Omega\longmapsto\mathbb{R}$.*
(i) *If $y\mapsto c_x(x_0,y)$ is strictly convex or strictly concave for some $x_0\in\mathbb{R}$, then $|S_0|\le 2$.*
(ii) *If for all $y_0\in\mathbb{R}$, we can find $k(y_0)\ge 2$ such that $y\mapsto c_x(x_0,y_0)$ is $k(y_0)$ times differentiable in y_0 and $c_{xy^{k(y_0)}}(x_0,y_0)\neq 0$, then S_0 is discrete. If furthermore $c_x(x_0,\cdot)$ is super-linear in y, then S_0 is finite.*

Proof. (i) The intersection of a strictly convex or concave curve with a line is two points or one if they intersect.

(ii) We suppose that S_0 is not discrete. Then we have $(y_n) \in S_0^{\mathbb{N}}$ a sequence of distinct elements converging to $y_0 \in \mathbb{R}$. In y_0, $f : y \mapsto c_x(x_0, y)$ is k times differentiable for some $k \geq 2$ and $f^{(k)}(y_0) = c_{xy^k}(x_0, y_0) \neq 0$. We have $f(y_n) = A(y_n)$. Passing to the limit $y_n \to y_0$; we get $f(y_0) = A(y_0)$. Now we subtract and get $f(y_n) - f(y_0) = \nabla A(y_n - y_0)$. We finally apply Taylor-Young around y_0 to get

$$(f'(y_0) - \nabla A)(y_n - y_0) + \sum_{i=2}^{k} \frac{f^{(i)}(y_0)}{i!}(y_n - y_0)^i + o(|y_n - y_0|^k) = 0$$

This is impossible for y_n close enough to y_0, as one of the terms of the expansion at least is nonzero. If furthermore $c_x(x_0, \cdot)$ is superlinear in y, S_0 is bounded, and therefore finite. □

Our next result is a weaker version of Theorem 4.2.11 (i) in higher dimension.

Proposition 4.2.12. *Let $x_0 \in \mathbb{R}^d$ such that for $y \in \mathbb{R}^d$, $c_x(x_0, y) = \sum_{i=1}^d P_i(y) u_i$, with for $1 \leq i \leq d$, $P_i \in \mathbb{R}[Y_1, .., Y_d]$ and $(u_i)_{1 \leq i \leq d}$ a basis of $\mathbb{R}^d$. We suppose that the P_i have degrees $deg(P_i) \geq 2$ and are complete at infinity. Then if $S_0 = \{c_x(x_0, Y) = A(Y)\}$ for some $x_0 \in \mathbb{R}^d$, and $A \in \mathrm{Aff}_d$, we have*

$$|S_0| \leq \deg(P_1) ... \deg(P_d).$$

The proof of this proposition is reported in Subsection 4.3.1.

Remark 4.2.13. *This bound is optimal as we see with the example: $P_i = (Y_i - 1)(Y_i - 2)...(Y_i - k_i)$, for $1 \leq i \leq d$. Then $\{1, 2, ..., k_1\} \times ... \times \{1, ..., k_d\} = \{c_x(x_0, Y) = A(Y)\}$. (For $A = 0$) And this set has cardinal $k_1 ... k_d = deg(P_1) ... deg(P_d)$. But this bound is not always reached when we fix the polynomials as we can see in the example $d = 1$ and $P = X^4$, we can add any affine function to it, it will never have more than 2 real zeros even if its degree is 4.*

The following example illustrates this theorem in dimension 2.

Example 4.2.14. *Let $d = 2$ and $c : (x, y) \in \mathbb{R}^2 \times \mathbb{R}^2 \longmapsto x_1(y_1^2 + 2y_2^2) + x_2(2y_1^2 + y_2^2)$. Then $c_x(x, y) = (y_1^2 + 2y_2^2)e_1 + (2y_1^2 + y_2^2)e_2$ for all (x, y), where (e_1, e_2) is the canonical basis of $\mathbb{R}^2$. Let $A \in \mathrm{Aff}_2$, $A = A_1 e_1 + A_2 e_2$. The equation $c_x(x_0, y) = A(y)$ can be written*

$$\begin{cases} y_1^2 + 2y_2^2 = & A_1(e_1)y_1 + A_1(e_2)y_2 + A_1(0) \\ 2y_1^2 + y_2^2 = & A_2(e_1)y_1 + A_2(e_2)y_2 + A_2(0). \end{cases}$$

These equations are equations of ellipses $\mathcal{C}_1$ of axes ratio $\sqrt{2}$ oriented along e_1, and $\mathcal{C}_2$ of axes ratio $\sqrt{2}$ oriented along e_2. Then we see visualy on Figure 4.1 that in *the nondegenerate case, $\mathcal{C}_1$ and $\mathcal{C}_2$ are determined by three affine independent points $y_1, y_2, y_3 \in \{c_x(x_0,Y) = A(Y)\}$, and that a fourth point y' naturally appears in the intersection of the ellipses.*

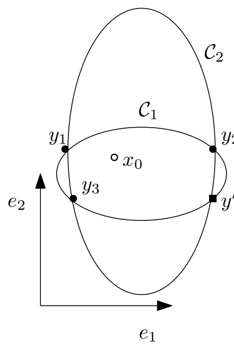


Fig. 4.1 Solution of $c_x(x_0,Y) = A(Y)$ for $c(x,y) = x_1(y_1^2+2y_2^2)+x_2(2y_1^2+y_2^2)$.

Now we give a general result. If $k \geq 1$, we denote

$$c_{x_i,y^k}(x_0,y_0)[Y^k] \quad := \sum_{1\leq j_1,...,j_k\leq d} \partial^{k+1}_{x_i,y_{j_1},...,y_{j_k}} c(x_0,y_0)Y_{j_1}...Y_{j_k}, \tag{4.2.9}$$

the homogeneous multivariate polynomial of degree k associated to the Taylor term of the expansion of the map $c_{x_i}(x_0,\cdot)$ around y_0 for $1 \leq i \leq d$.

We now provide e the extension of Theorem 4.2.11 (ii) to higher dimension.

Theorem 4.2.15. *Let $x_0 \in \mathbb{R}^d$ and $S_0 = \{c_x(x_0,Y) = A(Y)\}$ for some $A \in \mathrm{Aff}_d$. Assume that for all $y_0 \in \mathbb{R}^d$ and any $1 \leq i \leq d$, $c_{x_i}(x_0,\cdot)$ is $k_i \geq 2$ times differentiable at the point y_0 and that $\left(c_{x_i,y^{k_i}}(x_0,y_0)[Y^{k_i}]\right)_{1\leq i\leq d}$ is a complete at infinity family of $\mathbb{R}[Y_1,...,Y_d]$, then S_0 consists of isolated points. If furthermore $c_x(x_0,\cdot)$ is super-linear in y, then S_0 is finite.*

The proof of this theorem is reported in Subsection 4.3.1.

4.2.3 Largest support of conditional optimal martingale transport plan

The previous section provides a bound on the cardinal of the set S_0 in the polynomial case, which could be converted to a local result for a sufficiently smooth function, as it behaves locally like a multivariate polynomial. However, with the converse statement (ii) of the structure Theorem 4.2.2, we may also bound this cardinality from below.

Let c be a $\mathrm{C}^{1,2}$ cost function, and $x_0 \in \mathbb{R}^d$, we denote

$$N_c(x_0) := \sup_{P \in \mathbb{R}_1[Y_1,...,Y_d]^d} \left| Z^1_{\mathbb{R}}(H_c(x_0)+P) \right|, \quad \text{where} \quad H_c(x_0) := \left(c_{x_i,y^2}(x_0,x_0)[Y^2] \right)_{1 \le i \le d}.$$

where we denote by $Z^1_{\mathbb{R}}(Q_1,...,Q_d)$ the set of real (finite) single common zeros of the multivariate polynomials $Q_1,...,Q_d \in \mathbb{R}[Y_1,...,Y_d]$.

Definition 4.2.16. *We say that c is second order complete at infinity at $x_0 \in \mathbb{R}^d$ if c is differentiable at $x = x_0$ and twice differentiable at $y = x_0$, and $H_c(x_0)$ is a complete at infinity family of $\mathbb{R}_2[Y_1,...,Y_d]$.*

Remark 4.2.17. *Recall that by Remark 4.2.10, this property holds for almost all cost function. We highlight here that this consideration should be taken with caution, indeed cost functions of importance which are $c := f(|X-Y|)$ with f smooth fail to be second order complete at infinity, even in the case of c smooth at (x_0,x_0), as the sets $\{c_x(x_0,Y) = A(Y)\}$ for $A \in \mathrm{Aff}_d$ may be infinite and contradict Theorem 4.2.15, as they may contain balls, see Theorem 4.2.20 below.*

Theorem 4.2.18. *Let $c : \Omega \longrightarrow \mathbb{R}$ be second order complete at infinity and $\mathrm{C}^{2,0} \cap \mathrm{C}^{1,2}$ in the neighborhood of (x_0,x_0) for some $x_0 \in \mathbb{R}^d$. Then, we may find $\mu_0, \nu_0 \in \mathcal{P}(\mathbb{R}^d)$ with C^1 densities, and a unique $\mathbb{P}^* \in \mathcal{M}(\mu_0,\nu_0)$ such that*

$$\mathbf{S}_{\mu_0,\nu_0}(c) = \mathbb{P}^*[c] \quad \textit{and} \quad |\mathrm{supp}\, \mathbb{P}^*_X| = N_c(x_0),\, \mu - a.s.$$

The proof of this result is reported in subsection 4.3.2. Theorem 4.2.18 shows the importance of the determination of the numbers $N_c(x_0)$. We know by Remark 4.2.13 that for some cost $c : \Omega \longrightarrow \mathbb{R}$, the upper bound is reached: $N_c(x_0) = 2^d$. We conjecture that this bound is reached for all cost which is second order complete at infinity at x_0. An important question is whether there exists a criterion on cost functions to have the differential intersection limited to $d+1$ points, similarly to the Spence-Mirless condition in one dimension. It has been conjectured in [74] in the case of minimization for the distance cost. Theorem 4.2.22 together with (ii) of Theorem 4.2.2 proves that this conjecture is wrong. Now we prove that even for much more general second order complete at infinity cost functions, there is no hope to find such a criterion for d even.

Theorem 4.2.19. *Let $x_0 \in \mathbb{R}^d$, and c be second order complete at infinity and $\mathrm{C}^{1,2}$ at (x_0,x_0), then*

$$d+1+\mathbf{1}_{\{d \ even\}} \le N_c(x_0) \le 2^d.$$

The proof of Theorem 4.2.19 is reported in Subsection 4.3.2.

4.2.4 Support of optimal plans for classical costs

Euclidean distance based cost functions

Theorem 4.2.2 shows the importance of sets $S_0 = \{c_x(x_0,Y) = A(Y)\}$ for $x_0 \in \mathrm{ri\,conv}\, S_0$, and $A \in \mathrm{Aff}_d$. We can characterize them precisely when $c: (x,y) \in \mathbb{R}^d \times \mathbb{R}^d \longmapsto f(|x-y|)$ for some $f \in \mathrm{C}^1(\mathbb{R}_+,\mathbb{R})$. In view of Remark 4.2.4, the following result gives the structure of S_0 as a function of $d+1$ known points in this set. Let $g : t > 0 \longmapsto -f'(t)/t$, notice that

$$c_x(x,y) = g(|y-x|)(y-x), \quad \text{on} \quad \{X \neq Y\}.$$

Furthermore, $c(x,y)$ is differentiable in $x = y$ if and only if $f'(0) = 0$, in this case $c_x(x,x) = 0$. We fix $S_0 := \{c_x(x_0,Y) = A(Y)\}$, for some $x_0 \in \mathrm{int\,conv}\, S_0$, and $A \in \mathrm{Aff}_d$. The next theorem gives S_0 as a function of A and x_0. For $a \notin Sp(\nabla A)$, let $\mathbf{y}(a) := x_0 + (aI_d - \nabla A)^{-1}A(x_0)$. For $a \in Sp(\nabla A)$, if the limit exists, we write $|\mathbf{y}(a)| < \infty$ and denote $\mathbf{y}(a) := \lim_{t\to a}\mathbf{y}(t)$.

Theorem 4.2.20. *Let $S_0 := \{c_x(x_0,Y) = A(Y)\}$ for $x_0 \in \mathrm{ri\,conv}\, S_0$, and $A \in \mathrm{Aff}_d$. Then*

$$S_0 = \cup_{(t,\rho)\in\mathcal{A}} S_t^\rho \cup \Big\{\mathbf{y}(t) : t \in \mathbf{fix}(g \circ |\mathbf{y} - x_0|)\Big\},$$

where $S_t^\rho := \mathcal{S}_{V_t}\left(p_t, \sqrt{\rho^2 - |p_t - x_0|^2}\right)$, with $V_t := \mathbf{y}(t) + \ker(tI_d - \nabla A)$, $p_t := proj_{V_t}(x_0)$, and $\mathcal{A} := \Big\{(t,\rho) : t \in Sp(\nabla A), |\mathbf{y}(t)| < \infty, g(\rho) = t,$ and $\rho \geq |p_t - x_0|\Big\}$.

(i) The elements in the spheres $S_{t_0}^\rho$ for all ρ from Theorem 4.2.20 will be said to be $2d_{t_0}$ degenerate points, where $d_{t_0} := \dim V_{t_0}$. This convention corresponds to the degree $2d_{t_0}$ of their associated root t_0 of the extended polynomial $\chi(t) := \det(tI_d - \nabla A)^2 g^{-1}(t)^2 - |Com(tI_d - \nabla A)^t A(0)|^2)$. Notice that in the case $d_{t_0} = 1$, the sphere $S_{t_0}^\rho$ is a $0-$dimensional sphere, which consists in $2d_{t_0} = 2$ points.
(ii) We say that $\mathbf{y}(t_0) \in S_0$ is double for $t_0 \in \mathbb{R}$ if $\min_t \Big\{g\Big(|\mathbf{y}(t) - x_0|\Big) - t\Big\} = 0$ (attained at t_0) where the minimum is taken in the neighborhood of t_0. Notice that then in the smooth case, t_0 is a double root of χ.

Corollary 4.2.21. *S_0 contains at least $2d$ possibly degenerate points counted with multiplicity.*

The proofs of Theorem 4.2.20 and Corollary 4.2.21 are reported in Subsection 4.3.4.

Powers of Euclidean distance cost

In this section we provide calculations in the case where f is a power function. The particular cases $p=0,2$ are trivial, for other values, we have the following theorems.

Theorem 4.2.22. *Let* $c := |X-Y|^p$. *Let* $S_0 := \{c_x(x_0,Y) = A(Y)\}$, *for some* $x_0 \in \operatorname{int}\operatorname{conv} S_0$, *and* $A \in \mathrm{Aff}_d$. *Then if* $p \le 1$, S_0 *contains* $2d$ *possibly degenerate points counted with multiplicity, and if* $1 < p < 2-\frac{2}{5}$ *or* $p > 2+\frac{2}{3}$, S_0 *contains* $2d+1$ *possibly degenerate points counted with multiplicity.*

The proof of this theorem is reported in Subsection 4.3.4.

Remark 4.2.23. *In both cases, for almost all choice of* $y_0,\ldots,y_d \in \mathbb{R}^d$ *as the first elements of* S_0, *determining the Affine mapping* A, *we have* $d_i = 0$ *for all* i, *and* $c_{xy}(x_0,S_0) - \nabla A \subset GL_d(\mathbb{R}^d)$. *Then for* $-\infty < p \le 1$, *and* $p \neq 0$, $|S_0| = 2d$, *and for* $1 < p < 2-\frac{2}{5}$ *or* $p > 2+\frac{2}{3}$, $|S_0| = 2d+1$. *Therefore, by (ii) of Theorem 4.2.2, we may find* $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ *with* C^1 *densities such that the associated optimizer* $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ *of the MOT problem* (4.1.2) *satisfies* $|\operatorname{supp}\mathbb{P}_X| = 2d$, $\mu-$*a.s. if* $p \le 1$, *and* $|\operatorname{supp}\mathbb{P}_X| = 2d+1$, $\mu-$*a.s. if* $p > 1$.

Remark 4.2.24. *Based on numerical experiments, we conjecture that the result of Theorem 4.2.22 still holds for* $2-\frac{2}{5} \le p \le 2+\frac{2}{3}$, *and* $p \neq 2$. *See Section 5.7.*

Remark 4.2.25. *Assumption 4.2.1 implies that* c *is subdifferentiable. Then we can deal with cost functions* $c := -|X-Y|^p$ *with* $0 < p \le 1$ *only by evacuating the problem on* $\{X = Y\}$. *If* $0 < p \le 1$, *it was proved by Lim [113] that in this case the value* $\{X = Y\}$ *is preferentially chosen by the problem: Theorem 4.2 in [113] states that the mass* $\mu \wedge \nu$ *stays put (i.e. this common mass of* μ *and* ν *is concentrated on the diagonal* $\{X = Y\}$ *by the optimal coupling) and the optimization reduces to a minimization with the marginals* $\mu - \mu\wedge\nu$ *and* $\nu - \mu\wedge\nu$. *Therefore,* c *is differentiable on all the points concerned by this other optimization, and the supports are given by* $\operatorname{supp}\mathbb{P}_x \subset \{c_x(x,Y) = A_x(Y)\} \cup \{x\}$, *for* $\mu-$*a.e.* $x \in \mathbb{R}^d$. *Then the supports are exactly given by the ones from the maximisation case with eventually adding the diagonal.*

Notice that Remark 4.2.25 together with (ii) of Theorem 4.2.2 and Theorem 4.2.22 prove that Conjecture 2 in [74] is wrong, and explains the counterexample found by Lim [114], Example 2.9.

One and infinity norm cost

For $\varepsilon \in \mathcal{E}^1 := \{-1,1\}^d$, we denote $\mathcal{Q}^1_\varepsilon := \prod_{1\le i\le d}\varepsilon_i(0,\infty)$ the quadrant corresponding to the sign vector ε. Similarly, for $\varepsilon \in \mathcal{E}^\infty := \{\pm e_i\}_{1\le i\le d}$, we denote $\mathcal{Q}^\infty_\varepsilon := \{y \in \mathbb{R}^d : \varepsilon\cdot y > |y-(\varepsilon\cdot y)\varepsilon|_\infty\}$ the quadrant corresponding to the signed basis vector ε.

Proposition 4.2.26. *Let $c := |X-Y|_p$ with $p \in \{1,\infty\}$, and $S_0 := \{c_x(x_0,Y) = A(Y)\}$ for some $x_0 \in \operatorname{ri}\operatorname{conv} S_0$, and $A \in \operatorname{Aff}_d$, with $r := \operatorname{rank}\nabla A$. Then, we may find $2 \le k \le \mathbf{1}_{p=1}2^r + \mathbf{1}_{p=\infty}2r$, $\varepsilon_1,...,\varepsilon_k \in \mathcal{E}^p$, and $y_1,...,y_k \in \mathbb{R}^d$ such that*

$$S_0 = \cup_{i=1}^k (x_0 + \mathcal{Q}^p_{\varepsilon_i}) \cap (y_i + \ker\nabla A).$$

In particular, S_0 is concentrated on the boundary of its convex hull.

This Proposition will be proved in Subsection 4.3.3. The case $r = d$ is of particular interest.

Remark 4.2.27. *Notice that the gradient of c is locally constant where it exists (i.e. if c is differentiable at (x_0,y_0), then c is differentiable at (x,y) and $\nabla c(x,y) = \nabla c(x_0,y_0)$ for (x,y) in the neighborhood of (x_0,y_0)). Then if $r = d$, $c_{xy}(x_0,S_0) - \nabla A = -\nabla A \in GL_d(\mathbb{R})$, S_0 is finite and $|S_0| \le \mathbf{1}_{p=1}2^d + \mathbf{1}_{p=\infty}2d$. The bound is sharp (consider for example $A := x_0 + I_d$). Therefore, by (ii) of Theorem 4.2.2, we may find $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ with* C^1 *densities such that the associated optimizer $\mathbb{P} \in \mathcal{M}(\mu,\nu)$ of the MOT problem* (4.1.2) *satisfies $|\operatorname{supp}\mathbb{P}_X| = \mathbf{1}_{p=1}2^d + \mathbf{1}_{p=\infty}2d$, $\mu-$a.s.*

Concentration on the Choquet boundary

Recall that a set S_0 is included in its own Choquet boundary if $S_0 \subset Ext\big(\operatorname{cl}\operatorname{conv}(S_0)\big)$, i.e. any point of S_0 is extreme in $\operatorname{cl}\operatorname{conv}(S_0)$. A result showed in [74] is that the image of the optimal transport is concentrated in its own Choquet boundary for distance cost. We prove that this is a consequence of (i) of the structure Theorem 4.2.2, and we generalize this observation to some other cases.

Proposition 4.2.28. *Let $c : \Omega \longrightarrow \mathbb{R}$ be a cost function, $A \in \operatorname{Aff}_d$, $S_0 \subset \{c_x(x_0,Y) = A(Y)\}$, and $x_0 \in \operatorname{ri}\operatorname{conv} S_0$. S_0 is concentrated in its own Choquet boundary in the following cases:*

(i) *the map $y \mapsto c_x(x_0,y)\cdot u$ is strictly convex for some $u \in \mathbb{R}^d$;*

(ii) *$c : (x,y) \longmapsto |x-y|_p$, with $1 < p < \infty$;*

(iii) *$c : (x,y) \longmapsto |x-y|^p$, with $-\infty < p \le 1$;*

(iv) $c:(x,y)\longmapsto |x-y|^p$, *with* $1<p<2-\frac{2}{5}$ *or* $p>2+\frac{2}{3}$, *and* $p\left(\min_{y\in S_0}|y-x_0|\right)^{p-2}$ *is a double root of the polynomial* $\det(\nabla A - XI_d)^2 - |p|^{\frac{2}{2-p}}X^{\frac{2}{2-p}}|Com(\nabla A - XI_d)^t A(0)|^2$.

Furthermore, if $c:(x,y)\longmapsto |x-y|^p$, *with* $1<p<2-\frac{2}{5}$ *or* $p>2+\frac{2}{3}$, *and* S_0 *is not concentrated on its own Choquet boundary, then we may find a unique* $y_0\in S_0$ *such that* $|y_0-x_0|=\min_{y\in S_0}|y-x_0|$, *and* $S_0\setminus\{y_0\}$ *is concentrated on its own Choquet boundary.*

The proof of this proposition is reported in Subsection 4.3.5.

Remark 4.2.29. *If* $p=1$ *or* $p=\infty$, *there are counterexamples to Proposition 4.2.28 (ii), as* S_0 *may contain a non-trivial face of itself , see Proposition 4.2.26.*

4.3 Proofs of the main results

4.3.1 Proof of the support cardinality bounds

We first introduce some notions of Algebraic geometry. Recall $\mathbb{P}^d := \left(\mathbb{C}^{d+1}\right)^*/\mathbb{C}^*$, the $d-$dimensional projective space which complements the space with points at infinity. Recall that there is an isomorphism $\mathbb{P}^d\approx\mathbb{C}^d\cup\mathbb{P}^{d-1}$, where $\mathbb{P}^{d-1}$ are the "points at infinity". Then we may consider the points for which $x_0=0$ as "at infinity" because the surjection of $\mathbb{P}^d$ in $\mathbb{C}^d$ is given by $(x_0,x_1,...,x_d)\longmapsto(x_1/x_0,...,x_d/x_0)$ so that when $x_0=0$, we formally divide by zero and then consider that the point is sent to infinity. The isomorphism $\mathbb{P}^d\approx\mathbb{C}^d\cup\mathbb{P}^{d-1}$ follows from the easy decomposition:

$$
\begin{aligned}
\mathbb{P}^d &= \{(x_0,...,x_d)\in\mathbb{C}^{d+1}, x_0\neq 0\}/\mathbb{C}^*\cup\{(0,x_1,...,x_d),(x_1,...,x_d)\in\mathbb{C}^d\setminus\{0\}\}/\mathbb{C}^*\\
&= \{(1,x_1/x_0,...,x_d/x_0),(x_0,...,x_d)\in\mathbb{C}^{d+1},x_0\neq 0\}\\
&\quad \cup\Big\{(0,x_1,...,x_d),(x_1,...,x_d)\in\left(\mathbb{C}^d\right)^*\Big\}/\mathbb{C}^*\\
&\approx \mathbb{C}^d\cup\left(\mathbb{C}^d\right)^*/\mathbb{C}^*\approx\mathbb{C}^d\cup\mathbb{P}^{d-1}.
\end{aligned}
$$

The points in the projective space $\mathbb{P}^d$ in the equivalence class of $\{x_0=0\}$ are called points at infinity.

Definition 4.3.1. *The map*

$$
P=\sum_{n\in\mathbb{N}^d,|n|\leq\deg(P)} a_n X^n \longmapsto P^{proj} := \sum_{n\in\mathbb{N}^d,|n|\leq\deg(P)} a_n X^n X_0^{deg(P)-|n|},
$$

defines an isomorphism between $\mathbb{C}[X_1,...,X_d]$ *and* $\mathbb{C}^{hom}[X_0,X_1,...,X_d]$. *Let* $(P_1,...,P_k)$ *be* $k\geq 1$ *polynomials in* $\mathbb{R}[X_1,...,X_d]$, *we define the set of common projective zeros of* $(P_1,...,P_k)$ *by* $Z^{proj}(P_1,...,P_k) := Z(P_1^{proj},...,P_k^{proj})$.

This allows us to define the zeros of a nonhomogeneous polynomial in the projective space.

We finally report the following well-known result which will be needed for the proofs of Proposition 4.2.12 and Theorem 4.2.19.

Theorem 4.3.2 (Bezout)**.** *Let $d \in \mathbb{N}$ and $P_1, ..., P_d \in \mathbb{R}[X_1, ..., X_d]$ be complete at infinity. Then $|Z^{proj}(P_1, ..., P_d)| = deg(P_1)...deg(P_d)$, where the roots are counted with multiplicity.*

Proof. By Corollary 7.8 of Hartshorne [82] extended to $\mathbb{P}^d$ and d curves, we have

$$\sum_{V \in Irr\left(Z^{proj}(P_1,...,P_d)\right)} i\Big(Z^{proj}(P_1), ..., Z^{proj}(P_d), V\Big) = deg(P_1)...deg(P_d), \tag{4.3.10}$$

where $i\Big(Z^{proj}(P_1), ..., Z^{proj}(P_d), V\Big)$ is the multiplicity of the intersection of $Z^{proj}(P_1)$,..., and $Z^{proj}(P_d)$ along V, and $Irr\Big(Z^{proj}(P_1, ..., P_d)\Big)$ is the collection of irreducible components of $Z^{proj}(P_1, ..., P_d)$. By Remark 4.2.6, $Z^{proj}(P_1, ..., P_d)$ has dimension $d - d = 0$ by the fact that $(P_1, ..., P_d)$ is complete at infinity. Therefore, its irreducible components (in the algebraic sense) are singletons, and (4.3.10) proves the result. □

Notice that we have the identity $P^{hom} = P^{proj}(X_0 = 0)$. Then P^{hom} may be interpreted as the restriction to infinity of P^{proj} and we deduce the following characterization of completeness at infinity that justifies the name we gave to this notion. We believe that this is a standard algebraic geometry result, but we could not find precise references. For this reason, we report the proof for completeness. For $P_1, ..., P_d \in \mathbb{R}[X_1, ..., X_d]$, we denote $Z^{\text{aff}}(P_1, ..., P_d) := \{x \in \mathbb{C}^d : P_1(x) = ... = P_d(x) = 0\}$ the set of their common affine zeros.

Proposition 4.3.3. *Let $P_1, ..., P_d \in \mathbb{R}[X_1, ..., X_d]$, Then the following assertions are equivalent:*
(i) *$(P_1, ..., P_d)$ is complete at infinity;*
(ii) *$Z^{proj}(P_1, ..., P_d)$ contains no points at infinity;*
(iii) *$Z^{\text{aff}}(P_1^{hom}, ..., P_d^{hom}) = \{0\}$.*

Proof. We first prove $(iii) \implies (ii)$, let $x \in \mathbb{P}^d$ at infinity, i.e. such that $x_0 = 0$. Then by definition of the projective space, $x' := (x_1, ..., x_d) \neq 0$, and by (iii) we have that $P_i^{hom}(x') \neq 0$ for some i. Notice that $P_i^{hom}(x') = P_i^{proj}(x)$, and therefore $P_i^{proj}(x) \neq 0$ and $x \notin Z^{proj}(P_1, ..., P_d)$.

Now we prove $(i) \implies (iii)$. By definition of completeness at infinity, we have that $(P_1^{hom},...,P_d^{hom})$ is complete at infinity by the fact that $(P_1,...,P_d)$ is complete at infinity. By Theorem 4.3.2, $(P_1^{hom},...,P_d^{hom})$ has exactly $\deg P_1^{hom}...\deg P_d^{hom}$ common projective roots counted with multiplicity. However, by their homogeneity property, $(1,0,...,0)$ is a projective root of order $\deg P_1^{hom}...\deg P_d^{hom}$, therefore it is the only common projective root of these multivariate polynomials, in particular 0 is their only affine common root.

Finally we prove that $(ii) \implies (i)$. In order to prove this implication, we assume to the contrary that (i) does not hold. Then by Remark 4.2.6, we have that the dimension of this projective variety is higher than $d-(d-1)=1$. Then we may find some $x \in Z(P_1^{hom},...,P_{d-1}^{hom})$ which is different from $z=(1,0,...,0)$, as if z was the only zero, the dimension of $Z(P_1^{hom},...,P_{d-1}^{hom})$ would be 0. Now we consider $x' := (0,x_1,...,x_d) \in \mathbb{P}^d$. As $x'_0 = 0$, x' is at infinity and $P_i^{hom}(x) = P_i^{hom}(x') = P_i^{proj}(x')$. Therefore, $x' \in Z(P_1^{proj},...,P_d^{proj})$, contradicting (ii) by the fact that x' is at infinity. □

Proof of Remark 4.2.10 Let $\mathcal{X} := \{x \in \mathbb{P}^d : x_0 = 0\}$ the subset of points of $\mathbb{P}^d$ at infinity, and $\mathcal{Y} := H_{k_1} \times ... \times H_{k_d}$, with H_n the set of homogeneous polynomials of degree n for $n \in \mathbb{N}$. The set $\mathcal{X}$ is a projective variety as the set of zeros of the polynomial X_0, and the set $\mathcal{Y}$ is a quasi-projective variety as it is an affine space. The set $A := \{(p,P_1,...,P_d) \in \mathcal{X} \times \mathcal{Y} : P_1(p) = ... = P_d(p) = 0\}$ is a set of zeros of polynomials in $\mathcal{X} \times \mathcal{Y}$ (also called closed set for the Zariski topology by algebraic geometers). Notice that the set of non-complete at infinity polynomials in $\mathbb{R}[X_1,...,X_d]$ is exactly the projection of A on $\mathcal{Y}$ by Proposition 4.3.3, and therefore this set is characterized by a polynomial equation system on the coefficients of the P_i by Theorem 1.11 in [139], which states that the projection of closed sets for the Zariski topology in $\mathcal{X} \times \mathcal{Y}$ stays closed for the Zariski topology of $\mathcal{Y}$. □

Proof of Proposition 4.2.12 For $1 \le i \le d$, let $A_i := u_i \cdot A \in \mathrm{aff}(\mathbb{R}^d, \mathbb{R})$. If for each $1 \le i \le d$ we project this equation onto $Vect(u_i)$ along $Vect(u_j, j \neq i)$, we get:

$$P_i(y) = A_i(y), \quad i = 1,..,d.$$

Thanks to the completeness at infinity of (P_i), the $\widetilde{P}_i$ which are defined for $1 \le i \le d$ by

$$\widetilde{P}_i(Z_0,Z_1,...,Z_d) := P_i^{proj}(Z_1,...,Z_d) + \nabla A_i Z_0^{k-1} + A_i(0) Z_0^k$$

are also complete at infinity as for all i, we have $\widetilde{P}_i^{hom} = P_i^{hom}$. By Bezout Theorem 4.3.2 there are $\deg(P_1)...\deg(P_d)$ common projective roots to these polynomial. These roots may be complex, infinite, or multiple, therefore the set S_0 which is the set of these common roots that are finite and real has its cardinal bounded by $\deg(P_1)...\deg(P_d)$. □

Proof of Theorem 4.2.15 We suppose that S_0 is not discrete. Then we have $(y_n) \in S_0^{\mathbb{N}}$ a sequence of distinct elements converging to $y_0 \in \mathbb{R}^d$. We denote $P_i(Y_1,...,Y_d) := c_{x_i,y^{k_i}}(x_0,y_0)[Y^{k_i}]$ for $1 \le i \le d$. We know that $(P_i)_{1\le i\le d}$ is a complete at infinity family of $\mathbb{R}[Y_1,...,Y_d]$. We have $f(y_n) := c_x(x_0,y_n) = A(y_n)$. Passing to the limit $y_n \to y_0$, we get $f(y_0) = A(y_0)$. Now subtracting the terms, we get $f(y_n) - f(y_0) = \nabla A(y_n - y_0)$, and applying Taylor-Young around y_0, we get

$$(\nabla f(y_0) - \nabla A)\cdot(y_n - y_0) + \sum_{i=2}^{k-1} \frac{f^{(i)}(y_0)}{i!}[(y_n - y_0)^i] + P(y_n - y_0) + o(|y_n - y_0|^{k_i}) = 0 \quad (4.3.11)$$

With $P = (P_1,...,P_d)$. By Proposition 4.2.12, the Taylor multivariate polynomial is locally nonzero around y_0 as it has a finite number of zeros on $\mathbb{R}^d$. This is in contradiction with (4.3.11) for y_n close enough to y_0.

If furthermore c is super-linear in the y variable at x_0, T is bounded, and therefore finite. □

4.3.2 Lower bound for a smooth cost function

As a preparation for the proof of Theorem 4.2.19, we need to prove the following lemma.

Lemma 4.3.4. *Let $(P_1,...,P_d)$ be a complete at infinity family in $\mathbb{R}_2[X_1,...,X_d]$. Then the multivariate polynomial $\det(\nabla P_1,...,\nabla P_d)$ is non-zero.*

Proof. We suppose to the contrary that $\det(\nabla P) = 0$, where we denote $P = (P_1,...,P_d)$. We claim that we may find $y_0 \in \mathbb{R}^d$, and a map $u : \mathbb{R}^d \longrightarrow \mathcal{S}_1(0)$ which is $\mathcal{C}^\infty$ in the neighborhood of y_0 and such that $u(y) \in \ker(\nabla P(y))$ for y in this neighborhood. Then we solve the differential equation $y'(t) = u(y(t))$ with initial condition $y(0) = y_0$. As a consequence of the regularity of u in the neighborhood of y_0, by the Cauchy-Lipschitz theorem, this dynamic system has a unique solution for t in a neighborhood $[-\varepsilon,\varepsilon]$ of 0, where $\varepsilon > 0$. However, we notice that $P(y(t))$ is constant in t, indeed, $\frac{d(P(y(t)))}{dt} = \nabla P(y(t))u(y(t)) = 0$. Since $|y'(t)| = 1$, this solution is non constant, then $P - P(y_0)$ has an infinity of roots: $y([-\varepsilon,\varepsilon])$. However, as P is non-constant, $P - P(y_0)$

is also complete at infinity, which is in contradiction with the fact that it has an infinity of zeros by the Bezout Theorem 4.3.2.

It remains to prove the existence of $y_0 \in \mathbb{R}^d$, and a map $u : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, $\mathcal{C}^\infty$ in the neighborhood of y_0, such that $u(y) \in \ker(\nabla P(y_0))$ for y in this neighborhood. For all $i < d$, we consider the determinants of submatrices of ∇P which have size i. Let $r \geq 0$ the biggest such i so that at least one of these determinants is not the zero polynomial. By the fact that $\det(\nabla P) = 0$, and that the polynomials are non-constant by completeness at infinity, we have $0 < r < d-1$. We fix one of these non-zero polynomial determinants. Let $x_0 \in \mathbb{R}^d$ such that this determinant is non-zero at y_0. As this determinant is continuous in y, it is non-zero in the neighbourhood of y_0. Therefore, ∇P has exacly rank r in the neighbourhood of y_0. Now we show that this consideration allows to find a continuous map $y \longmapsto u(y)$, such that $u(y)$ is a unit vector in $\ker(\nabla P)$. Notice that $\ker(\nabla P) = \mathrm{Im}(\nabla P^t)^\perp$. We consider r columns of ∇P^t that are used for the non-zero determinant. We apply the Gramm-Schmidt orthogonalisation algorithm on them. We get $u_1(y), ..., u_r(y)$, an orthonormal basis of $\mathrm{Im}(\nabla P(y)^t)$, defined and C^∞ in the neighbourhood of y_0. Then let $u_0 \in \ker(\nabla P(y_0))$, a unit vector. The map

$$u(y) := \frac{u_0 - \sum_{i=1}^r \langle u_0, u_i(y)\rangle u_i(y)}{|u_0 - \sum_{i=1}^r \langle u_0, u_i(y)\rangle u_i(y)|}$$

is well defined, C^∞, and in $\mathrm{Im}(\nabla P(y)^t)^\perp = \ker(\nabla P(y))$ in the neigbourhood of y_0, and therefore satisfies the conditions of the claim. □

Proof of Theorem 4.2.19

Step 1: Let $P_i := (X_1, ..., X_d) c_{x_i,yy}(x_0, x_0)(X_1, ..., X_d)^t$. Let $y_1, ..., y_{d+1} \in \mathbb{R}^d$, affine independent. We may find $A \in \mathrm{Aff}_d$ such that $A(y_i) = P(y_i)$ for all i, where we denote $P := (P_i)_{1 \leq i \leq d}$. Now we prove that $\nabla(P(y_i') - A)$ may be made invertible at points y_i' at the neighborhood of y_i. Recall that A is a function of the $d+1$ vectors y_i: $A = A(y_1, ..., y_{d+1})$. Then we look for an explicit expression of $\nabla A(y_1, ..., y_{d+1})$ (denoted ∇A for simplicity) as a function of the y_i. Let $Y = Mat(y_i - y_{d+1}, i = 1, ..., d)$, the matrix with columns $y_i - y_{d+1}$, using the equality $\nabla A y_i + A(0) = P(y_i)$, we get the identity $\nabla A Y = M$, where we denote $M := Mat(P(y_i) - P(y_{d+1}), i = 1, ..., d)$. Then we get the result $\nabla A = M Y^{-1}$ (Y is invertible as the y_i are affine independent). Then having $\nabla P(y_{d+1}) - \nabla A$ invertible is equivalent to having $\nabla P(y_{d+1})Y - M$ invertible. Notice that $\nabla P(y_{d+1})Y - M = -Mat(\tilde{P}(y_i), i = 1, ..., d)$, where $\tilde{P} = P - P(y_{d+1}) - \nabla P(y_{d+1}) \cdot (Y - y_{d+1})$, and that the multivariate polynomials $\tilde{P}_i$ are complete at infinity, as they only differ from the P_i by degree one polynomials. Consider the multivariate polynomial

$D := \det(\nabla\tilde{P})$. Let $1 \le i \le d$, by Lemma 4.3.4 we may find y_i' in the neighborhood of y_i such that $D(y_i') \neq 0$, and therefore $\nabla\tilde{P}(y_i')$ is invertible. Thanks to this invertibility, we may perturb the y_i' to make $M' := Mat(\tilde{P}(y_i'), i = 1, ..., d)$ invertible. As $Sp(M')$ is finite, for $\lambda > 0$ small enough, $M' + \lambda I_d$ is invertible. For $1 \le i \le d$, we may find y_i'' in the neighborhood of y_i' so that $\tilde{P}(y_i'') = \tilde{P}(y_i') + \lambda e_i + o(\lambda)$, thanks to the invertibility of $\nabla\tilde{P}(y_i')$. Then for λ small enough, $(P(y_i''), i = 1, ..., d) = M' + \lambda I_d + o(\lambda)$ is invertible.

We were able, by perturbing the y_i for $i \neq d+1$ to make $\nabla(P(y_{d+1}') - A)$ invertible. By continuity, this invertibility property will still hold if we perturb again sufficiently slightly the y_i. Then we redo the same process, replacing y_{d+1}' by another y_i'. We suppose that the perturbation is sufficiently small so that all the invertibilities hold in spite of the successive perturbations of the y_i. Finally, we found $y_1', ..., y_{d+1}'$ affine independent so that $P(y_i') = A(y_i')$ and $\nabla P(y_i') - \nabla A$ is invertible for all $1 \le i \le d+1$. Step 2: Then $N_c(x_0) \ge d+1$ because $y_1', ..., y_{d+1}'$ are $d+1$ single real roots of $P + A = H_c(x_0) + A$, and $A \in \mathrm{Aff}_d$, which may be identified to $\mathbb{R}_1[Y_1, ..., Y_d]^d$. As the $P_i - A_i$ are real multivariate polynomials, all non-real zeros have to be coupled with their complex conjugate. Recall that by Theorem 4.3.2, there are exactly 2^d zeros to this system. There are no zeros at infinity by Proposition 4.3.3, and there is an even number of non-real zeros by the invariance by conjugation observation. Then there must be an even number of real roots. As the y_i' are simple roots by invertibility of the derivative of $P - A$ at these points, there must be an even number of real roots, counted with multiplicity. If d is even, $d+1$ is odd, which proves the existence of a possibly multiple $d+2-$th zero y_0, distinct from the y_i. We assume, up to renumbering, that $y_0', ..., y_d'$ are affine independent, and we perturb again $y_0', ..., y_d'$ to make y_0 a single zero. We need to check that y_{d+1}' is still a single zero of $P - A$. Indeed, the map $(y_1', ..., y_{d+1}') \longmapsto A$ if locally a diffeomorphism around $(y_1, ..., y_{d+1})$, then by the implicit functions Theorem, we may write $y_{d+1}' = F(y_1', ..., y_d', A) = F\Big(y_1', ..., y_d', A(y_0', ..., y_d')\Big)$, where F is a local smooth function. Then y_{d+1}' remains a single zero if the perturbation of $y_0, ..., y_d$ is small enough. The result is proved, if d is even we may find $d+2$ single zeros to $P - A$.

The reverse inequality is a simple application of Proposition 4.2.12. □

As a preparation for the proof of Theorem 4.2.18, we introduce the two following lemmas:

Lemma 4.3.5. *Let $Q_1, ..., Q_d$, d complete at infinity multivariate polynomials of degree 2 and $x \in \mathbb{R}^d$. Then, for all $P_1, ..., P_d$ multivariate polynomials of degree 1, we may find $\widetilde{P}_1, ..., \widetilde{P}_d$, multivariate polynomials of degree 1 such that $|Z^1_{\mathbb{R}}(Q_1 + \widetilde{P}_1, ..., Q_d + \widetilde{P}_d)| \ge |Z^1_{\mathbb{R}}(Q_1 + P_1, ..., Q_d + P_d)|$ and $x \in \mathrm{int\,conv}\, Z^1_{\mathbb{R}}(Q_1 + \widetilde{P}_1, ..., Q_d + \widetilde{P}_d)$.*

Proof. Let $P_1,...,P_d$ multivariate polynomials of degree 1. We claim that we may find $R_1,...,R_d$ of degree 1 so that $Z^1_{\mathbb{R}}(Q_1+R_1,...,Q_d+R_d)$ has full dimension and contains $Z^1_{\mathbb{R}}(Q_1+P_1,...,Q_d+P_d)$. Then we may find $x' \in \operatorname{int}\operatorname{conv} Z^1_{\mathbb{R}}(Q_1+R_1,...,Q_d+R_d)$, and by the fact that all Q_i have degree 2, we may find $\widetilde{P}_1,...,\widetilde{P}_d$ of degree 1 such that $(Q+P)(X+x'-x) = Q+\widetilde{P}$. Finally, as the change of variables $X+x'-x$ does not change the number of roots of $Q+P$ nor their multiplicity, and by the fact that $x \in \operatorname{int}\operatorname{conv} Z^1_{\mathbb{R}}(Q_1+\widetilde{P}_1,...,Q_d+\widetilde{P}_d)$ by translation, $\widetilde{P}_1,...\widetilde{P}_d$ solves the problem.

Now we prove the claim. We prove by induction that we may add dimensions to $Z^1_{\mathbb{R}}(Q_1+R_1,...,Q_d+R_d)$ by changing the R_i. First by Theorem 4.2.19, we may assume that $Z^1_{\mathbb{R}}(Q_1+R_1,...,Q_d+R_d)$ is non-empty. Up to making a distance-preserving linear change of variables, we may assume that $Z^1_{\mathbb{R}}(Q_1,...,Q_d) \subset \{X_d = 0\}$ and that $0 \in Z^1_{\mathbb{R}}(Q_1,...,Q_d)$. We look for $D \in \mathbb{R}[X_1,...,X_d]^d$ in the form $D = X_d v$ for some $v \in \mathbb{R}^d$, so that $Q+D$ leaves $Z^1_{\mathbb{R}}(Q_1,...,Q_d)$ unchanged. In order to include some $y \in \{X_d \neq 0\}$, we set $D := -Q(y)/y_d X_d$. The constraint that we have now is to fix y is that $\nabla(Q+D)(y') \in GL_d(\mathbb{R})$ for all $y' \in Z^1_{\mathbb{R}}(Q_1,...,Q_d)$ and for $y' = y$. Notice that all these constraints have the form $\det\Big(\nabla(y_d Q - Q(y)X_d)(y')\Big) \neq 0$ if $y' \neq y$, and $\det\Big(\nabla(y_d Q - Q(y)X_d)(y)\Big) \neq 0$ for the case $y' = y$, therefore in all the cases this is a polynomial equation in y. We claim that each of these equations on y have a solution. Then as there is a finite number of such equations, the set of solutions is a dense open set, in particular it is non-empty and we may find $y \in \mathbb{R}^d$ so that $\{y\} \cup Z^1_{\mathbb{R}}(Q) \subset Z^1_{\mathbb{R}}(Q+D)$ and $\dim Z^1_{\mathbb{R}}(Q+D) > \dim Z^1_{\mathbb{R}}(Q)$. By induction, we may reach full dimension for $\dim Z^1_{\mathbb{R}}(Q+D)$, and the problem is solved.

Finally, we prove the claim that the solution set to $\det\Big(\nabla(y_d Q - Q(y)X_d)(y')\Big) \neq 0$ is non-empty.

<u>Case 1:</u> $y' \in Z^1_{\mathbb{R}}(Q)$. Then, up to applying a translation change of variables, we may assume that $y' = 0$. Then by the fact that Q has degree 2, the equation that we would like to satisfy is

$$\det\Big(y_d \nabla Q(0) - \Big(\nabla Q(0)y + \frac{1}{2}D^2Q(0)[y^2]\Big)e_d^t\Big) \neq 0.$$

We make it more tractable by making operations on the columns:

$$\begin{aligned}
& \det\left(y_d\nabla Q(0)-\left(\nabla Q(0)y+\frac{1}{2}D^2Q(0)[y^2]\right)e_d^t\right)\\
= & \det\left(y_d\sum_{i=1}^{d}\nabla Q_i(0)e_i^t-\left(\sum_{i=1}^{d}\nabla Q_i(0)y_i+\frac{1}{2}D^2Q(0)[y^2]\right)e_d^t\right)\\
= & \det\left(y_d\sum_{i=1}^{d-1}\nabla Q_i(0)e_i^t-\frac{1}{2}D^2Q(0)[y^2]e_d^t\right),
\end{aligned}$$

where we have subtracted the i^{th} column multiplied by y_i/y_d to the d^{th} column for all $1\le i\le d-1$. Now we prove that this multivariate polynomial is non-zero. We assume for contradiction that it is zero. Then for all $y\in\{X_d\neq 0\}$, $D^2Q(0)[y^2]\in H:= Vect(\nabla Q_i(0), 1\le i\le d-1)$, which is $d-1-$dimensional by the fact that $\nabla Q\in GL_d(\mathbb{R})$ by simplicity of the root 0. By continuousness, we have in fact that $D^2Q(0)[y^2]\in H$ for all $y\in\mathbb{R}^d$. Therefore, for all $y_1,y_2\in\mathbb{R}^d$, we have the equality $D^2Q(0)[y_1,y_2]=\frac{1}{2}\left(D^2Q(0)[(y_1+y_2)^2]-D^2Q(0)[y_1,y_1]-D^2Q(0)[y_2,y_2]\right)\in H$. Then we may find $u\in\mathbb{R}^d$ non-zero such that $\sum_{i=1}^d u_iD^2Q_i(0)=0$. Then $(Q_1^{hom},...,Q_d^{hom})$ is $d-1-$dimensional and $Z^{proj}(Q_1,...,Q_d)$ is at least $1-$dimensional, then it intersects the variety of points at infinity, which is a contradiction by Proposition 4.3.3 together with the fact that $(Q_1,...,Q_d)$ is a complete at infinity family.

Case 2: $y'=y$. Then the equation that we would like to satisfy is

$$\det\left(y_d\nabla Q(y)-\left(\nabla Q(0)y+\frac{1}{2}D^2Q(0)[y^2]\right)e_d^t\right)\neq 0,$$

which may be expanded thanks to the fact that Q has degree 2:

$$\det\left(y_d\left(\nabla Q(0)+D^2Q(0)y\right)-\left(\nabla Q(0)y+\frac{1}{2}D^2Q(0)[y^2]\right)e_d^t\right)\neq 0.$$

Similar than in the previous case, by the same operations on the columns we get:

$$\begin{aligned}
& \det\left(y_d\left(\nabla Q(0)+D^2Q(0)y\right)-\left(\nabla Q(0)y+\frac{1}{2}D^2Q(0)[y^2]\right)e_d^t\right)\\
= & \det\left(y_d\sum_{i=1}^{d}\left(\nabla Q_i(0)+D^2Q_i(0)y\right)e_i^t-\left(\sum_{i=1}^{d}\nabla Q_i(0)y_i+\frac{1}{2}D^2Q_i(0)yy_i\right)e_d^t\right)\\
= & \det\left(y_d\sum_{i=1}^{d-1}\left(\nabla Q_i(0)+D^2Q_i(0)y\right)e_i^t+\frac{1}{2}D^2Q(0)[y^2]e_d^t\right),
\end{aligned}$$

Now we assume for contradiction that this polynomial in y is zero. Then for all $y \in \{X_d \neq 0\}$ small enough so that $\nabla Q(0) + D^2Q(0)y \in GL_d(\mathbb{R})$, $D^2Q(0)[y^2] \in H_y := Vect(\nabla Q_i(0) + D^2Q_i(0)y, 1 \leq i \leq d-1)$. Notice that up to multiplying y by $\lambda > 0$, we have that $\lambda^2 D^2Q(0)[y^2] \in H_{\lambda y}$, and therefore $D^2Q(0)[y^2] \in H_{\lambda y}$. By passing to the limit $\lambda \longrightarrow 0$, we have $D^2Q(0)[y^2] \in H_0$ thanks to the fact that $\nabla Q \in GL_d(\mathbb{R})$. Therefore we obtain a contradiction similar to case 1. $\square$

Lemma 4.3.6. *Let $M > 0$, we may find $R(M)$ such that for all $F : \mathbb{R}^d \longmapsto \mathbb{R}^d$ and $x_0 \in \mathbb{R}^d$ such that on $B_{M^{-1}}(x_0)$, F is C^2 and we have that ∇F and D^2F is bounded by M, and $\det \nabla F \geq M^{-1}$, we have that F is a C^1-diffeomorphism on $B_{R(M)}(x_0)$.*

Proof. The determinant is a polynomial application, therefore it is Lipschitz when restricted to the compact of matrices bounded by M. Let $L(M)$ be its Lipschitz constant. Then on the neighbourhood $B_{R_0(M)}(x_0)$, we have that $\det \nabla F$ is bigger than $\frac{1}{2}M^{-1}$, with $R_0(M) = \min\Big(M^{-1}, \frac{1}{2L(M)M}\Big)$. We claim that F is injective on $B_{R_1(M)}(x_0)$ with $R_1(M) := \min\Big(M^{-1}, \frac{1}{4M^2C(M)}\Big)$, where $C(M)$ is a bound for the comatrices of matrices dominated by M. Then by the global inversion theorem, F is a C^1-diffeomorphism on $B_{R(M)}(x_0)$ with $R(M) = \min\Big(R_0(M), R_1(M)\Big)$.

Now we prove the claim that F is injective on $B_{R_1(M)}(x_0)$. Let $x, y \in B_{R_1(M)}(x_0)$,

$$\begin{aligned}
F(y) - F(x) &= \int_0^1 \nabla F(tx + (1-t)y)(y-x)dt \\
&= \nabla F(x)(y-x) + \int_0^1 \int_0^t D^2F(sx + (1-s)y)[(y-x)^2]dsdt \\
&= \nabla F(x)\left(y - x + \nabla F(x)^{-1} \int_0^1 (1-s)D^2F(sx + (1-s)y)[(y-x)^2]ds\right).
\end{aligned}$$

Then we assume that $F(y) = F(x)$. Then

$$\begin{aligned}
|y - x| &= \left|\nabla F(x)^{-1} \int_0^1 (1-s)D^2F(sx + (1-s)y)[(y-x)^2]ds\right| \\
&\leq \|\nabla F(x)^{-1}\| \frac{M}{2} |y-x|^2 \\
&\leq C(M)M^2|y-x|^2, \qquad (4.3.12)
\end{aligned}$$

where the last estimate comes from the comatrix formula (5.1.2). Then by the fact that $R_1(M) \leq \frac{1}{4M^2C(M)}$, we have $|y-x| < \frac{1}{M^2C(M)}$, and therefore $x = y$ by (4.3.12). The injectivity is proved. $\square$

Proof of Theorem 4.2.18 By Taylor expansion of c_x in y in the neighborhood of x_0, we get for $h\in\mathbb{R}^d$ and $\varepsilon>0$ small enough that

$$c_x(x_0,x_0+\varepsilon h)=c_x(x_0,x_0)+c_{xy}(x_0,x_0)\varepsilon h+Q(\varepsilon h)+\varepsilon^2 R_\varepsilon(h),$$

where, recalling the notation (4.2.9), $Q_i(Y):=\frac{1}{2}c_{x_iyy}(x_0,x_0)[Y^2]$ and the remainder

$$R_\varepsilon(h)=\int_0^1(1-t)\Big(c_{xyy}(x_0,x_0+\varepsilon th)-c_{xyy}(x_0,x_0)\Big)[h^2]dt.$$

Notice that $\nabla R_\varepsilon(h)=3\int_0^1(1-t)\Big(c_{xyy}(x_0,x_0+\varepsilon th)-c_{xyy}(x_0,x_0)\Big)[h]dt$. By Proposition 4.2.12, we see that $N_c(x_0)$ is finite by second order completeness at infinity of c at (x_0,x_0). We consider from the definition of $N_c(x_0)$ an affine map $A\in\mathrm{Aff}_d$ such that the $d-$tuple of multivariate polynomials of degree one $A(X_1,...,X_d)$ satisfies

$$\left|Z^1_{\mathbb{R}}\Big(Q_i+A(X_1,...,X_d)_i:1\le i\le d\Big)\right|\quad=n\quad:=N_c(x_0).$$

By Theorem By Lemma 4.3.5, let $P=(P_1,...,P_d)$, d multivariate polynomials of degree 1 such that $|Z^1_{\mathbb{R}}(Q_1+\widetilde{P}_1,...,Q_d+\widetilde{P}_d)|\ge n$ and $0\in\mathrm{int\,conv}\,Z^1_{\mathbb{R}}(Q_1+\widetilde{P}_1,...,Q_d+\widetilde{P}_d)$.

Let $A_\varepsilon(y):=-\varepsilon^2P(0)-\varepsilon\nabla P(y-x_0)+c_x(x_0,x_0)+c_{xy}(x_0,x_0)(y-x_0)$, we have that $c_x(x_0,x_0+\varepsilon h)=A_\varepsilon(x_0+\varepsilon h)$ and $c_{xy}(x_0,x_0+\varepsilon h)\in GL_d(\mathbb{R})$ if and only if $Q(h)+P(h)+R_\varepsilon(h)=0$ and $(\nabla Q+\nabla P)(h)+\nabla R_\varepsilon(h)\in GL_d(\mathbb{R})$.

Now let $h_1,...,h_n\in\mathbb{R}^d$ the n elements of $Z^1_{\mathbb{R}}(Q_1+\widetilde{P}_1,...,Q_d+\widetilde{P}_d)$. By continuousness of c_{xyy} in the neighborhood of (x_0,x_0), up to restricting to a compact neighborhood, c_{xyy} is uniformly continuous on this neighborhood. For $\varepsilon>0$ small enough, each $x_0+\varepsilon h_i$ in in the interior of this neighborhood. Therefore, by uniform continuousness R_ε, and ∇R_ε converges uniformly to 0 when $\varepsilon\longrightarrow 0$. Let $1\le i\le n$, we have $(Q+P+R_\varepsilon)(h_i)=R_\varepsilon(h_i)$, and $\nabla(Q+P)(h_i)\in GL_d(\mathbb{R})$ by the fact that h_i is a single root of $Q+P$, and therefore $\nabla(Q+P+R_\varepsilon)(h_i)\in GL_d(\mathbb{R})$ for ε small enough. Therefore we may apply Lemma 5.6.1 around h_i: $Q+P+R_\varepsilon$ is a diffeomorphism in a neighborhood of h_i depending only on the lower bounds of $\det\nabla(Q+P+R_\varepsilon)(h_i)$ and of the bounds for $\nabla(Q+P+R_\varepsilon)$ and $D^2(Q+P+R_\varepsilon)$, which may then work for all ε small enough. Then for ε small enough, we may find h_i^ε in this neighborhood of h_i such that $(Q+P+R_\varepsilon)(h_i^\varepsilon)=0$. Furthermore, by the fact that $\nabla(Q+P+R_\varepsilon)(h_i)\longrightarrow\nabla(Q+P)(h_i)$ when $\varepsilon\longrightarrow 0$, $|h_i^\varepsilon-h_i|\le 2\|\nabla(Q+P)^{-1}(h_i)\||R_\varepsilon(h_i)|$, and therefore $h_i^\varepsilon\longrightarrow h_i$ when $\varepsilon\longrightarrow 0$. Then for ε small enough, the h_i^ε are distinct, $0\in\mathrm{ri\,conv}(y_i^\varepsilon,1\le i\le n)$, $Q(h_i^\varepsilon)+P(h_i^\varepsilon)+R_\varepsilon(h_i^\varepsilon)=0$, and $(\nabla Q+\nabla P)(h_i^\varepsilon)+\nabla R_\varepsilon(h_i^\varepsilon)\in GL_d(\mathbb{R})$.

Now the theorem is just an application of (ii) of Theorem 4.2.2 to $S_0 := \{x_0 + \varepsilon h_i^\varepsilon, i = 1, ..., n\}$. □

4.3.3 Characterization for the p-distance

Fot $p \geq 1$ and $x \in \mathbb{R}^d$, we have $c(\cdot, y)$ differentiable on $(\mathbb{R}^d)^*$ with

$$c_x(x,y) = \frac{1}{|x-y|_p^{p-1}} \sum_{i=1}^{d} |x_i - y_i|^{p-1} \frac{x_i - y_i}{|x_i - y_i|} e_i$$

For $p = 1$ and $p = \infty$, it takes a simpler form.

If $p = 1$, $c(\cdot, y)$ is differentiable on $\prod_{i=1}^d (\mathbb{R} \setminus \{y_i\})$ and $c_x(x,y) = \sum_{i=1}^{d} \frac{x_i - y_i}{|x_i - y_i|} e_i$.

If $p = \infty$, $c(\cdot, y)$ is differentiable on $\{x' \in \mathbb{R}^d, |x'_i - y_i| > |x'_j - y_j|, j \neq i,$ for some $1 \leq i \leq d\}$, let $i := \text{argmax}_{1 \leq j \leq d}(|x_j - y_j|)$, we have $c_x(x,y) = \frac{x_i - y_i}{|x_i - y_i|} e_i$.

Proof of Proposition 4.2.26 We start with the case $p = 1$. We suppose without loss of generality that $x_0 = 0$. Recall that $c(\cdot, y)$ is differentiable on $(\mathbb{R}^*)^d$ and $c_x(0,y) = \sum_{i=1}^d \frac{y_i}{|y_i|} e_i$. Then the equation that we get is $A(y) = \sum_{i=1}^d sg(y_i) e_i$. Let $E := \{\sum_{i=1}^d sg(y_i) e_i : y \in S_0\} \subset \varepsilon \in \{-1,1\}^d$. We have $E \subset \text{Im}A$, which is an affine space of dimension r. Then there are r coordinates $i_1, ..., i_r$ that can be chosen arbitrarily in $\text{Im}A$, and the other coordinates are affine functions of the previous one. We denote $I := (i_1, ..., i_r)$ and $\bar{I} := (1, ..., d) \setminus I$. Thus, $\text{card}(\text{Im}A \cap \{-1,1\}^d) \leq \text{card}(\{-1,1\}^I) = 2^r$. As $0 \in \text{ri}S_0$, $r \geq 1$. Now for all $\varepsilon \in E$, let $y_\varepsilon \in S_0$ such that $c_x(0, y_\varepsilon) = \varepsilon$. Then if $y := y_\varepsilon + y_0 \in \mathcal{Q}_\varepsilon^1$ with $y_0 \in \ker \nabla A$, we have $A(y) = c_x(0,y)$, and therefore $y \in S_0$, proving the first part of the result.

Now we prove that $S_0 \subset \partial \text{conv}\, S_0$. Let us suppose to the contrary that $y \in \text{ri}\,\text{conv}\, S_0 \cap S_0$. Let $y_1, ..., y_n \in S_0$ such that $y = \sum_{i=1}^n \lambda_i y_i$, convex combination. Then $c_x(0,y) = \sum_{i=1}^n \lambda_i c_x(0, y_i)$. As $|c_x(0,y)| = \sum_{i=1}^n \lambda_i |c_x(0,y_i)| = \sqrt{d}$, we are in a case of equality in Cauchy-Schwartz inequality. $\varepsilon := c_x(0,y), c_x(0,y_1), ..., c_x(0,y_n)$ are all non-negative multiples of the same unit vector, and therefore all equal as they have the same norm. Then $y, y_1, ..., y_n \in \mathcal{Q}_\varepsilon^1$, and $y, y_1, ..., y_n \in y_\varepsilon + \ker \nabla A$. As we may apply the same to any $y' \in y_\varepsilon + \ker \nabla A$, these vectors cannot be written as convex combinations of elements of S_0 that do not belong to $y_\varepsilon + \ker \nabla A$. Therefore, $(y_\varepsilon + \ker \nabla A) \cap S_0 = (y_\varepsilon + \ker \nabla A) \cap \mathcal{Q}_\varepsilon^1$ is a face of $\text{conv}\, S_0$. As we assumed that $y \in \text{ri}\,\text{conv}\, S_0$, we have $(y_\varepsilon + \ker \nabla A) \cap \mathcal{Q}_\varepsilon^1 = \text{ri}\,\text{conv}\, S_0$, by the fact that $\text{ri}\,\text{conv}\, S_0$ and $(y_\varepsilon + \ker \nabla A) \cap \mathcal{Q}_\varepsilon^1$ are faces of $\text{conv}\, S_0$ (which constitute a partition of $\text{conv}\, S_0$, see Hiriart-Urruty-Lemaréchal [89]) both containing y. This is impossible as $0 \in \text{ri}\,\text{conv}\, S_0$ and $0 \notin \mathcal{Q}_\varepsilon^1$. Whence the required contradiction.

The proof of the case $p=\infty$ is similar to the proof of Proposition 4.2.26, replacing by $\mathrm{card}(\{-1,1\}(e_i)_{1\le i\le d})=2d$ instead of 2^d, and by $|c_x(0,y)|=1$ instead of $\sqrt{d}$. $\square$

4.3.4 Characterization for the Euclidean p-distance cost

By the fact that $\mathrm{int\,conv}\, S_0$ contains x_0, we may find $y_1,...,y_{d+1}\in S_0$ that are affine independent. Then we may find unique barycenter coefficients $(\lambda_i)_i$ such that $x_0=\sum_{i=1}^{d+1}\lambda_i y_i$. For some $y_1,...,y_{d+1}\in S_0$. For all $a\in\mathbb{R}$, we define

$$\mathbf{y}'(a):=G(a)\sum_{i=1}^{d+1}\frac{\lambda_i}{a-a_i}y_i,\quad\text{with}\quad G(a)=\left(\sum_{i=1}^{d+1}\frac{\lambda_i}{a-a_i}\right)^{-1},\ \text{and } a_i:=g(|y_i-x_0|),\tag{4.3.13}$$

where $\{b_1,...,b_r\}:=\{a_1,...,a_{d+1}\}$ with $r\le d+1$ and $b_1<...<b_r$, and $d_i:=\Big|\{j:a_j=b_i\}\Big|-1$, the multiplicity of each b_i for all i.

Proposition 4.3.7. *We have $\mathbf{y}'(a)=\mathbf{y}(a)$ for all $a\notin Sp(\nabla A)$. In particular the map $\mathbf{y}'$ is independent of the choice of $y_1,...,y_{d+1}\in S_0$. Furthermore, $G(a)=\frac{(a-a_1)...(a-a_{d+1})}{\det(aI_d-\nabla A)}=\frac{(a-b_1)...(a-b_r)}{(a-\gamma_1)...(a-\gamma_{r-1})}$ where $\gamma_1<...<\gamma_{r-1}$ are eigenvalues of ∇A. Finally if we have $x_0\in\mathrm{int\,conv}(y_1,...,y_{d+1})$, then we have $b_1<\gamma_1<b_2<...<\gamma_{r-1}<b_r$.*

Proof. We suppose that $x_0=0$ for simplicity. Let $a\notin Sp(\nabla A)$, $\mathbf{y}(a)$ is the unique vector such that

$$(aI_d-\nabla A)\,\mathbf{y}(a)=A(0)\tag{4.3.14}$$

We now find the barycentric coordinates of $\mathbf{y}(a)$. For any i, $A(y_i)=a_iy_i$ with $a_i:=g(|y_i|)$. As $(y_i)_i$ is a barycentric basis, we may find unique $(\lambda_i(a))_i\subset\mathbb{R}$ such that $\mathbf{y}(a)=\sum_i\lambda_i(a)y_i$, and $1=\sum_i\lambda_i(a)$. Then we apply A and get $A(\mathbf{y}(a))=\sum_i\lambda_i(a)A(y_i)$, so that $a\mathbf{y}(a)=\sum_i\lambda_i(a)a_iy_i$. Subtracting the previous equality on $\mathbf{y}(a)$, we get $0=\sum_i\lambda_i(a)(a-a_i)y_i$. As $(y_i)_i$ is a barycentric basis, it is a family or rank d. Then, by the fact that $\sum_{i=1}^{d+1}\lambda_iy_i=0$, we have $(\lambda_i)_{1\le i\le d+1}$ and $(\lambda_i(a)(a-a_i))_{1\le i\le d+1}$ are in the same $1-$dimensional kernel of the matrix $(y_1,...,y_{d+1})$. Then we may find $G(a)$ such that $\lambda_i(a)(a-a_i)=G(a)\lambda_i$. Now we assume that a is not part of the a_i, then we have $\lambda_i(a)=G(a)\frac{\lambda_i}{a-a_i}$, and $G(a)=\left(\sum_{i=1}^{d+1}\frac{\lambda_i}{a-a_i}\right)^{-1}$. Finally

$$\mathbf{y}(a)=\mathbf{y}'(a)=G(a)\sum_{i=1}^{d+1}\frac{\lambda_i}{a-a_i}y_i\quad\text{with}\quad G(a)=\left(\sum_{i=1}^{d+1}\frac{\lambda_i}{a-a_i}\right)^{-1}.\tag{4.3.15}$$

Now we prove that $G(a) = \frac{(a-a_1)...(a-a_{d+1})}{\det(aI_d-\nabla A)}$. We first assume that $a_1 < ... < a_{d+1}$ and that $x_0 \in \operatorname{int}\operatorname{conv}(y_1,...,y_{d+1})$ (i.e. $\lambda_1,...,\lambda_{d+1} > 0$). Then $G(a)^{-1}$ has $d+1$ single poles $a_1,...,a_{d+1}$, such that $\lim_{a\uparrow a_i} G(a)^{-1} = +\infty$, and $\lim_{a\downarrow a_i} G(a)^{-1} = -\infty$ for all i. Therefore, $G(\gamma_i)^{-1} = 0$ for some $a_i < \gamma_i < a_{i+1}$ for all $i \le d$. Then γ_i is a pole of G, and $|\mathbf{y}'(a)|$ goes to infinity when $a \to \gamma_i$, as the coefficient in the affine basis $(y_i)_i$ go to $\pm\infty$. Therefore, γ_i is an eigenvalue of ∇A, as there are d such eigenvalues, we have obtained all of them. Finally, by the fact that the rational fraction f has degree 1, as the set of its roots is restricted to the $d+1$ numbers a_i. Furthermore the γ_i are d poles, and $a^{-1}G(a) \longrightarrow (\sum_{i=1}^{d+1}\lambda_i)^{-1} = 1$, when $a \to \infty$, we deduce the rational fraction $G(X) = \frac{(X-a_1)...(X-a_{d+1})}{(X-\gamma_1)...(X-\gamma_d)} = \frac{(X-a_1)...(X-a_{d+1})}{\det(XI_d-\nabla A)}$.

Now if we chose other affine independent $(y_i)_{1\le i\le d+1}$ (this time not necessary with $x_0 \in \operatorname{conv}(y_i, 1 \le i \le d+1)$), let the associated barycenter coordinates $\lambda_1,...,\lambda_{d+1} \in \mathbb{R}^*$, we suppose that the $(a_i)_i$ are still distinct, the poles of $\mathbf{y}'(a)$ are still the d distinct eigenvalues of ∇A that are determined by the γ_i such that $\lim_{a\to\gamma_i}|\mathbf{y}(a)|$, independent of the choice of $(y_i)_i$ because $\mathbf{y}'(a) = (aI_d - \nabla A)^{-1} A(0)$ is independent of this choice. However, the numerator of the fraction can be determined in the same way than it is determined in the previous case.

Now we want to generalize this result to $\lambda_1,...,\lambda_{d+1} \in \mathbb{R}$, and any $(a_i)_i$. If we stay in the open set in which $(y_i)_i$ is an affine basis of $\mathbb{R}^d$, the mapping $(y_i,a_i)_i \longmapsto A$ is continuous, and so is the mapping $(y_i)_i \longmapsto (\lambda_i)_i$. Therefore, as $(y_i,a_i,\lambda_i)_i \longmapsto \sum_{i=0}^{d}\frac{\lambda_i}{X-a_i}$ is continuous as well, the identity remains true for all a_i, y_i such that $(y_i)_i$ is an affine basis and $\lambda_i \ge 0$.

Let us now focus on the multiple a_is. We consider $1 \le i \le r$ such that $d_i > 0$. By passing to the limit $n \to \infty$ with some distinct a_i^n converging to a_i for all $1 \le i \le d$, d_i eigen values of ∇A at least will be trapped between the a_is, as $a_i^n < \gamma_{i+1}^n < a_{i+1}^n < ... < \gamma_{i+k}^n < a_{i+k}^n$ becomes at the limit $a_i = \gamma_{i+1} = a_{i+1} = ... = \gamma_{i+k} = a_{i+k}$. Now we prove that no other eigenvalue is equal to a_i. Indeed, rewriting (4.3.15) that equation become

$$\mathbf{y}(a) = \mathbf{y}'(a) = G(a)\sum_{i=1}^{r}\frac{\lambda_i'}{a-b_i}y_i \quad \text{with} \quad G(a) = \left(\sum_{i=1}^{r}\frac{\lambda_i'}{a-b_i}\right)^{-1}. \qquad (4.3.16)$$

with $\lambda_i' := \sum_{a_j=b_i}\lambda_j$. And $G(a) = \frac{(X-b_1)^{d_1+1}...(X-b_r)^{d_r+1}}{\det(XI_d-\nabla A)}$. By a similar reasoning than when the $(a_i)_i$ are distinct, we may find $b_1 < \gamma_1 < b_2 < ... < \gamma_{r-1} < b_r$, eigenvalues of ∇A. Then, as $\deg\det(XI_d - \nabla A) = d$, and $(X-b_1)^{d_1}...(X-b_r)^{d_r}$ is a divider to $\det(XI_d - \nabla A)$, we have $\det(XI_d - \nabla A) = (X-\gamma_1)...(X-\gamma_{r-1})(X-b_1)^{d_1}...(X-b_r)^{d_r}$.

$\square$

Remark 4.3.8. *Notice that in Proposition 4.3.7, the eigenvalues of ∇A are given by the γ_i, and by each b_i such that $d_i > 0$, which has multiplicity d_i, in particular, these coefficients (up to their numbering) do not depend on the choice of $y_1, ..., y_{d+1}$.*

Proof of Theorem 4.2.20 We suppose again that $x_0 = 0$ for simplicity. We know that if $y \in S_0$, $c_x(0,y) = g(|y|)y = A(y)$. We denote $a := g(|y|)$ and get,

$$(aI_d - \nabla A)\, y = A(0) \tag{4.3.17}$$

Let $a \in \mathbf{fix}(g \circ |\mathbf{y} - x_0|)$, then $(aI_d - \nabla A)\mathbf{y}(a) = A(0)$, and $A\Big(\mathbf{y}(a)\Big) = a\mathbf{y}(a) = g\Big(|\mathbf{y}(a)|\Big)\mathbf{y}(a) = c_x\Big(0, \mathbf{y}(a)\Big)$, and therefore $\mathbf{y}(a) \in S_0$. Conversely, if $y \in S_0$ and $a := g(|y|)$ is not an eigenvalue of ∇A, $y = (aI_d - \nabla A)^{-1}A(0) = \mathbf{y}(a)$, and finally $g(|\mathbf{y}(a)|) = a$, hence $a \in \mathbf{fix}(g \circ |\mathbf{y} - x_0|)$.

Now let $t \in Sp(\nabla A)$ such that $|\mathbf{y}(t)| < \infty$. Let $y \in S_t^\rho$, we have $(tI_d - \nabla A)y = (tI_d - \nabla A)(y - \mathbf{y}(t)) + A(0) = A(0)$, by passing to the limit $a \longrightarrow t$ in the equation $(aI_d - \nabla A)\mathbf{y}(a) = A(0)$. Finally, as $|y|^2 = \sqrt{\rho^2 - |p_t|^2}^2 + |p_t|^2 = \rho^2$ by Pythagoras theorem, $A(y) = c_x(0,y)$, and therefore $y \in S_0$. Conversely, if $y \in S_0$ with $g(|y|) = t$, then we have $y - \mathbf{y}(t) \in \ker(tI_d - \nabla A)$, and $|y - p_t| = \sqrt{\rho^2 - |p_t|^2}$ by Pythagoras theorem: by definition $y \in S_t^\rho$. □

Proof of Corollary 4.2.21 We use the notations from Proposition 4.3.7 and assume that $x_0 \in \operatorname{int}\operatorname{conv}(y_1, ..., y_{d+1})$. By Theorem 4.2.20, S_0 contains $2\sum_{i=1}^r d_i$ degenerate points. Furthermore, for all $1 \le i \le r-1$, $\lim_{t \to \gamma_i} |\mathbf{y}(t) - x_0| = \infty$, therefore, as b_{i+1} is a root of $g\Big(|\mathbf{y}(t) - x_0|\Big) - t$ between γ_i and γ_{i+1}, there is another root b_i', possibly multiple equal to b_i, by continuity of g. Finally we have $2\sum_{i=1}^r d_i + r + (r-2) = 2d$ elements in S_0 at least, with possible degeneracy. □

Proof of Theorem 4.2.22 We assume again that $x_0 = 0$ for simplicity. We suppose again that $x_0 = 0$ for simplicity. By identity (5.1.2), if we multiply (4.3.14) by the comatrix, we get $\det(\lambda I_d - \nabla A)y = Com(\lambda I_d - \nabla A)^t A(0)$. Now taking the square norm, we get: $\det(\lambda I_d - \nabla A)^2 |p|^{\frac{2}{p-2}} \lambda^{\frac{2}{p-2}} - |Com(\lambda I_d - \nabla A)^t A(0)|^2 = 0$. The polynomial with real exponents $\chi := \det(H - XI_d)^2 - |p|^{\frac{2}{2-p}} \lambda^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$ is continuous in $(y_i)_i$, then similar to the proof of Theorem 4.2.20, we can pass to the limit from sequences of y_i^n converging to y_i for all i such that for all $n \ge 1$, the vectors y_i^n have distinct norms. It follows that b_i is a d_i-eigenvalue of ∇A, and a $(2d_i - 1)$-root of χ.

By Theorem 4.2.20, we have

$$S_i = \mathcal{S}_{V_i}\left(p_i, \sqrt{b_i^2 - |p_i|^2}\right) \subset \{c_x(0,Y) = A(Y)\}.$$

With the radius $\sqrt{b_i^2 - |p_i|^2} > 0$ as there are more than one elements in the sphere. We have a single sphere as the function g is monotonic, and therefore injective.

Now we prove that if $-\infty < p \leq 1$, then the polynomial with real exponents

$$\chi(X) := \det(XI_d - \nabla A)^2 - |p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2 \tag{4.3.18}$$

has exactly $2d$ positive roots, counted with multiplicity. By Corollary 4.2.21, it has at least $2d$ roots, counted with multiplicity. Now we prove that there are at most $2d$ roots.

By Theorem 4.2.20, the roots of $\det(XI_d - \nabla A)$ all have the same sign (same than p). Consequently, the coefficients of $\det(XI_d - \nabla A)$ are alternated or all have the same sign. The same happens for $\det(XI_d - \nabla A)^2$. Now we use the Descartes rule[2] for polynomials with non integer exponents in order to dominated the number of roots of χ. Recall that $\chi = \det(XI_d - \nabla A)^2 - |p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$. We saw that the coefficients from the part $\det(XI_d - \nabla A)^2$ are alternated or all of the same sign. The exponent sequences from $\det(XI_d - \nabla A)^2$, and from $|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$ have both integer differences between two exponents from the same sequence. Then the exponents of $|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$ are located between the ones of $\det(XI_d - \nabla A)^2$ in the exponent sequence of χ, i.e. the sequence of χ consists in one exponent from $\det(XI_d - \nabla A)^2$, then one exponent from $|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$, and so on. By the fact that $deg(\det(XI_d - \nabla A)^2) = 2d$ and $deg(|Com(XI_d - \nabla A)^t A(0)|^2) = 2d - 2$, and $0 < \frac{2}{2-p} \leq 2$. Then $\chi(X)$ has at most $2d$ alternations in its coefficients, and therefore it has at most $2d$ positive roots according to the Descartes rule.

Now, assume that $1 < p < 2 - \frac{2}{5}$ or $p > 2 + \frac{2}{3}$, then

$$\chi(X) := \det(XI_d - \nabla A)^2 - |p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2 \tag{4.3.19}$$

has exactly $2d+1$ positive roots counted with multiplicity.

Let us first prove that the polynomial has less than $2d+1$ roots. Similar to above, the coefficients of $\det(XI_d - \nabla A)$ are alternated. And the same happens

[2]The Descartes rule states that for a polynomial with possibly non integer real coefficients, the number of positive roots is dominated by the number of alternations of signs of its coefficients ordered by their associated exponents, see [96].

for $\det(XI_d - \nabla A)^2$. Using the Descartes rule for polynomials with non integer coefficients, by the fact that the coefficients of $|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$ are located between the ones of $\det(XI_d - \nabla A)^2$, except strictly less than 3, and as $deg(\det(XI_d - \nabla A)^2) = 2d$, it follows that $deg(|Com(XI_d - \nabla A)^t A(0)|^2) = 2d - 2$ and $-3 < \frac{2}{2-p} < 5$. Then $\chi(X)$ has at most $2d+2$ alternations in its coefficients by the same reasoning than the case $p \leq 1$. Furthermore, the sign of the coefficients in front of the extreme monomials are opposed (because χ is a difference of positive polynomials) then the maximum number of positive roots is odd, and therefore it has at most $2d+1$ positive roots according to Descartes rule.

By Corollary 4.2.21, we have $2d$ elements in S_0, more precisely, which range between b_1 and b_r. Furthermore, between 0 and b_1 we can find some $a \in D$:
Case 1: We assume that $p > 2$. Then $\chi(X) \to -\infty$ when $X \to 0$ as we have that $-|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |Com(XI_d - \nabla A)^t A(0)|^2$ becomes dominant.
Case 2: We assume that $p < 2$. Then $\chi(X) \to -\infty$ when $X \to +\infty$ as we have that $-|p|^{\frac{2}{2-p}} X^{\frac{2}{2-p}} |{}^t Com(XI_d - \nabla A)A(0)|^2$ becomes dominant.

Therefore there is one more real root, on the side where the polynomial goes to $-\infty$ as there is already one. Finally χ has $2d+1$ roots at least and less than $2d+1$ roots, it follows that it has exactly $2d+1$ roots. We proved the second part of the theorem. □

4.3.5 Concentration on the Choquet boundary for the p-distance

Proof of Proposition 4.2.28 (i) Let $y_0, y_1, ..., y_k \in S_0$ such that $y_0 = \sum_{i=1}^{k} \lambda_i y_i$, convex combination. Then as $c_x(x_0, y_i) \cdot u = {}^t u A(y_i - x_0)$, we have $\sum_{i=1}^k \lambda_i c_x(x_0, y_i) \cdot u = u^t A(y_0 - x_0) = c_x(x_0, y_0) \cdot u$. As $y \mapsto c_x(x_0, y) \cdot u$ is strictly convex, this imposes that $\lambda_i = 1$ and $y_i = y_0$ for some i. Finally, y_0 is extreme in S_0, S_0 is concentrated in its own Choquet boundary.
(ii) We know that for any $y \in S_0$ we have $c_x(x_0, y) = A(y)$. As the situation is invariant in x_0, we will assume $x_0 = 0$ for notations simplicity. We consider $1 < q < +\infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. For any $y \in (\mathbb{R}^d)^*$,

$$|c_x(0,y)|_q = \left| \frac{1}{|y|_p^{p-1}} \sum_{i=1}^{d} |y_i|^{p-1} \frac{y_i}{|y_i|} e_i \right|_q = \frac{1}{|y|_p^{p-1}} \left(\sum_{i=1}^{d} |y_i|^{(p-1)q} \right)^{\frac{1}{q}} = \frac{1}{|y|_p^{\frac{p}{q}}} |y|_p^{\frac{p}{q}} = 1,$$

as we know that $y \neq 0$ because c is superdifferentiable. Then for any $y \in S_0$, we have $|Hy + v|_q = 1$. We now assume that $y_0 = \sum_{i=1}^{k} \lambda_i y_i$ is a strict convex combination with

$(y_i)_{0\leq i\leq k} \in S_0^{k+1}$.

$$1 \;=\; |A(y_0)|_q = \left|\sum_{i=1}^{k} \lambda_i\Big(A(y_i)\Big)\right|_q \leq \sum_{i=1}^{k} \lambda_i |A(y_i)|_q = \sum_{i=1}^{k} \lambda_i = 1$$

We are in a case of equality for the triangular inequality for the norm $|\cdot|_q$. We know then that all the $\lambda_i A(y_i)$ and $A(y_0)$ are positively multiples. As we know that all their q-norm is $\lambda_i \neq 0$ and 1, therefore $A(y_0) = ... = A(y_k)$ and $\frac{1}{|y_0|_p^{p-1}} \sum_{i=1}^{d} |(y_0)_i|^{p-1} \frac{(y_0)_i}{|(y_0)_i|} e_i = ... = \frac{1}{|y_k|_p^{p-1}} \sum_{i=1}^{d} |(y_k)_i|^{p-1} \frac{(y_k)_i}{|(y_k)_i|} e_i$. Notice that for $y \in \mathbb{R}^d$, we have $\frac{1}{|y|_p^{p-1}} \sum_{i=1}^{d} |y_i|^{p-1} \frac{y_i}{|y_i|} e_i = f(y/|y|_p)$, where $f : y \longmapsto \sum_{i=1}^{d} |y_i|^{p-1} \frac{y_i}{|y_i|} e_i$ is bijective $\mathbb{R}^d \longrightarrow \mathbb{R}^d$ for $p > 1$. Then we have $\frac{y_0}{|y_0|_p} = ... = \frac{y_k}{|y_k|_p}$. It means that they all belong to the same semi straight line originated in 0. As we supposed that y_0 is not extreme, 0 can be included in the convex combination as we must have $1 \leq i \leq k$ such that $|y_k| > |y_0|$. Then increasing the corresponding λ_i while decreasing all the others, 0 can be included. As $0 \in \operatorname{ri}\operatorname{conv} S_0$, we can then put any element of S_0 in the convex combination and $S_0 \subset \{0\} + \frac{y_0}{|y_0|}\mathbb{R}_+$. As $0 \in \operatorname{ri}\operatorname{conv} S_0$, then $S_0 = \{0\}$ and $y_0 = 0$, which is the required contradiction because we supposed that y_0 is not extreme in S_0.
(iii) We use the notations from Theorem 4.2.22. We suppose again without loss of generality that $x_0 = 0$. Let $d := \dim S_0$, for any $y_1, ..., y_{d+1} \in S_0$ with full dimension d, we may find unique barycentric coordinates $(\lambda_i)_{1\leq i\leq d+1}$ such that $\sum_{i=0}^{d} \lambda_i y_i = 0$. Let $y \in S_0$ such that $p|y|^{p-2} = g(|y|) \notin Sp(\nabla A)$. By Proposition 4.3.7, y can be expressed as

$$y = G(X) \sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i} y_i \quad \text{with} \quad G(X) = \left(\sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i}\right)^{-1}.$$

with $X = p|y|^{p-2} > 0$. To have $y \in \operatorname{conv}(S_0)$ we then need to have all the $\frac{\lambda_i}{X - a_i}$ of the same sign. As we supposed that the $(a_i)_i$ is an increasing sequence, there must be a $0 \leq i_0 \leq d-1$ such that $\lambda_i < 0$ if $i \leq i_0$ and $\lambda \geq 0$ if $i \geq i_0 + 1$ (or $\lambda_i > 0$ if $i \leq i_0$ and $\lambda \leq 0$ if $i \geq i_0 + 1$ but we will only treat the first case as this one can be dealt with similarly). Then the idea consists in proving that χ defined by (4.3.18) has no zero in $]a_{i_0}, a_{i_0+1}[$.

First let us prove that G has no pole on $]a_{i_0}, a_{i_0+1}[$. G^{-1} can hit 0 at most d times (It is a polynomial of degree d divided by another polynomial). It hits 0 in any $]a_i, a_{i+1}[$ for $i \neq i_0$, as the limits on the bounds are $+\infty$ and $-\infty$. This provides $d-1$ zeros. If there where a zero in $]a_{i_0}, a_{i_0+1}[$, it would be double, as the infinity limits at $a_{i_0}^+$ and $a_{i_0+1}^-$ have the same sign. Which would be a contradiction.

Finally, as the poles of G are the eigenvalues of ∇A and do not depend on the choice of $y_1, ..., y_{d+1}$, we know that there are exactly two roots of χ between two poles. As a_{i_0} and a_{i_0+1} are two zeros surrounded by two consecutive poles, there are not other zeros between these two poles. χ has no zero on $]a_{i_0}, a_{i_0+1}[$.

If $X = a_{i_0}$ or $X = a_{i_0+1}$, then it is a zero of $a_{i_0} - X$, and all the elements in the convex combination have same size than y. By the fact that we are in the case of equality in the Cauchy-Schwartz inequality, this proves that the combination only contains one element. Hence, $y \in S_0$ has to be extreme in S_0.

Now if y corresponds to an eigenvalue of ∇A, let $b := g(|y|)$. We suppose that $y = \sum_{i=1}^{d+1} \mu_i y_i$, convex combination with $y_1, ..., y_{d+1} \in S_0$, affine basis. Recall that all $\mathbf{y}(a)$ for $a \notin Sp(\nabla A)$ can be written $\mathbf{y}(a) = G(a) \sum_{i=1}^{d+1} \frac{\lambda_i}{a - a_i} y_i = G(a) \sum_{i=1}^{r} \frac{\lambda_i'}{a - b_i} y_i'$ where $\lambda_i' = \sum_{a_j = b_i} \lambda_j$, and $y_i' = \sum_{a_j = b_i} \frac{\lambda_j}{\lambda_i'} y_j$. Let i_0 such that $b_{i_0} = b$, let $\{y_1', ..., y_{d_{i_0}}'\} := \{y' \in \{y_1, ..., y_{d+1}\} : g(|y'|) = b_{i_0}\}$. $y \in \text{aff}(y_1', ..., y_{d_{i_0}}')$, therefore $\mu_i = 0$ if $a_i \neq b$. As S_i is a sphere, it is concentrated on its own Choquet boundary, and therefore the convex combination $y = \sum_{i=1}^{d+1} \mu_i y_i$ is trivial, $y = y_i$ for some i and $\mu_i = 1$.

(iv) In the first case, if $p|y_0|^{p-2}$ is a double root of χ defined by (4.3.19), then if $p < 2 - \frac{2}{5}$ or $p > 2 + \frac{2}{3}$, χ has $2d+1$ roots and at most $2d$ distinct roots set around the poles of G in the same way than in the case $p \leq 1$ in the proof of (iii).

The same happens when we remove the smallest element y_0 of S_0. Similarly $S_0 \setminus \{y_0\}$ is concentrated on its own Choquet boundary.

Now we prove that S_0 is not concentrated on its own Choquet boundary. If $p|y_0|^{p-2}$ is a single root of χ, we select $y_1', ..., y_{d+1}' \in S_0$ such that 0 is in their convex hull. By Proposition 4.3.7, if $y \in S_0$ and $X := p|y|^{p-2}$, then

$$y = G(X) \sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i} y_i \quad \text{with} \quad G(X) = \left(\sum_{i=1}^{d+1} \frac{\lambda_i}{X - a_i} \right)^{-1}. \tag{4.3.20}$$

Case 1: We assume that $y_1' = y_0$. Then we apply (4.3.20) to $X := p|y|^{p-2}$ the second smallest zero of χ which is strictly smaller than the first pole by Theorem 4.2.22 (which also means that $G(X) \geq 0$): $y := G(X) \sum_{i=0}^{d} \frac{\lambda_i}{a_i - X} y_i \in S_0$, or written otherwise:

$$\frac{\lambda_0 G(X)}{X - a_0} y_0 = G(X) \sum_{i=2}^{d+1} \frac{\lambda_i}{X - a_i} y_i - y$$

G has its first zero at a_0 which is smaller than its first pole which is between a_1 and a_2 strictly, so that $G(X) > 0$. This gives the result, rewriting the barycenter equation, we get:

$$y_0 = \sum_{i=2}^{d+1} \frac{\lambda_i(X-a_0)}{\lambda_0(X-a_i)} y_i + \frac{G(X)}{\lambda_0} y$$

Therefore, $y_0 \in \mathrm{conv}(S_0 \setminus \{y_0\})$.

Case 2: Now we assume that $y_0' \neq y_0$. We write the barycenter equation for $X = p|y_0|^{p-2}$, we get:

$$y_0 = \sum_{i=0}^{d} \frac{\lambda_i G(X)}{X-a_i} y_i' \quad \text{with} \quad G(X) = \left(\sum_{i=0}^{d} \frac{\lambda_i}{X-a_i} \right)^{-1}.$$

Then for any i, $\frac{\lambda_i G(X)}{X-a_i} > 0$ as all the $\frac{\lambda_i}{X-a_i}$ have the same sign. Therefore $y_0 \in \mathrm{conv}(S_0 \setminus \{y_0\})$. □

4.4 Numerical experiment

In the particular example $c(X,Y) = |X-Y|^p$, the computations are easy as the important unknown parameter $\lambda = p|y|^{p-2}$ is one-dimensional. We coded a solver that generates random $y_1, ..., y_{d+1} \in \mathbb{R}^d$ and determines the missing $y_{d+2}, ..., y_k$, with $k = 2d$ if $p \leq 1$, and $k = 2d+1$ if $p > 1$ such that $\{y_1, ..., y_k\} = \{c_x(0,Y) = A(Y)\}$ for some $A \in \mathrm{Aff}_d$, see Theorem 4.2.22. (As we chose randomly these vectors, we are in a non-degenerate case with probability 1). Theorem 4.2.28 only covers the case in which $p < 2 - \frac{2}{5}$ or $p > 2 + \frac{2}{3}$, however the numerical experiment seems to show that the result of this theorem still holds for all $2 \neq p > 1$. Figures 4.2, 4.3, 4.4, 4.5, and 4.6 show configurations (S_0, on the left) for $p = 1.9$ and $p = 2.1$ in which the result of the theorem holds, and the graphs of $\frac{1}{p-2} \log\left(\frac{\lambda}{p}\right)$ compared to $\log\left(\mathbf{y}(-p\lambda^{p-2})\right)$ as functions of $\log(\lambda)$ (on the right). The intersections are in bijection with the points in S_0 because of the non-degeneracy by Theorem 4.2.20 with the change of variable $t = -p\lambda^{p-2}$. The color of the points need to be interpreted as follows: $d+1$ blue points are chosen at random so that 0 belongs to their convex hull. Then the new d points given by Theorem 4.2.20 are colored in red. Finally the point corresponding to the first intersection of the curves on the right is colored in yellow because this special intersection differentiates the case $p \leq 1$ and the case $p > 1$. We begin with Figures 4.2 and 4.3, in two dimensions.

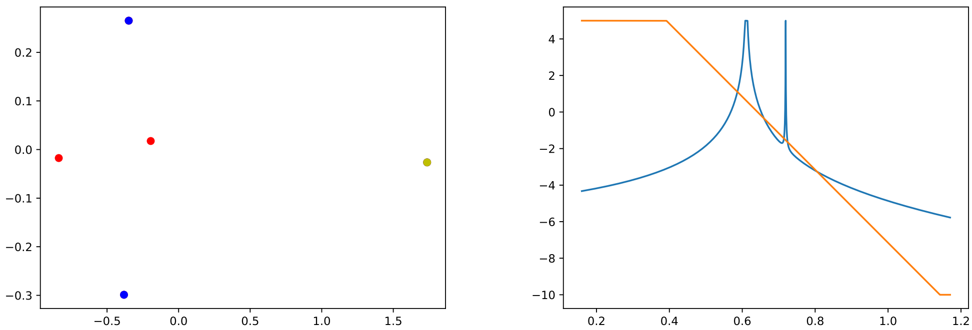


Fig. 4.2 S_0 for $d = 2$ and $p = 1.9$.

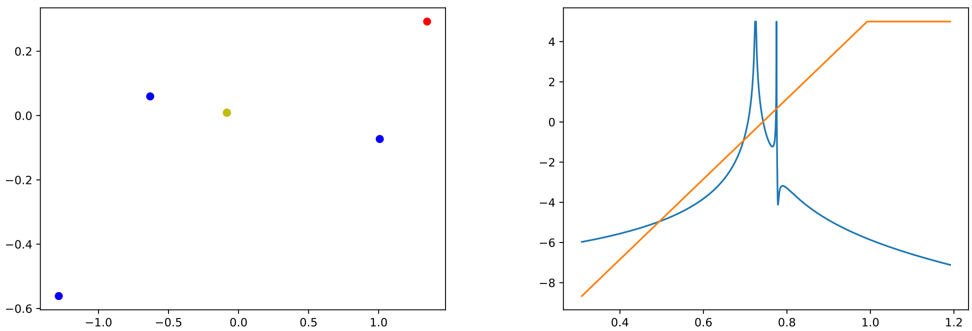


Fig. 4.3 S_0 for $d = 2$ and $p = 2.1$.

Now Figures 4.4 and 4.5, in three dimensions.

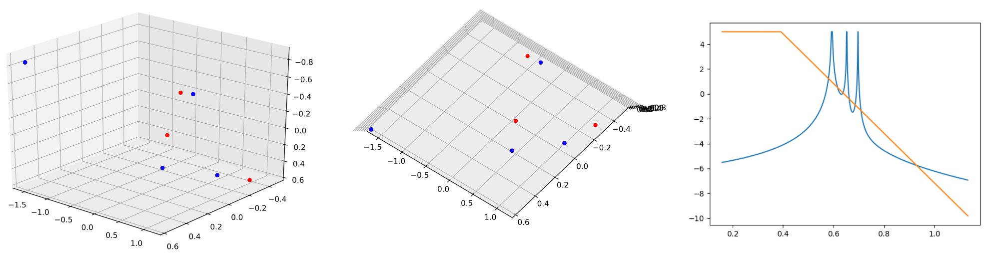


Fig. 4.4 S_0 for $d = 3$ and $p = 1.9$.

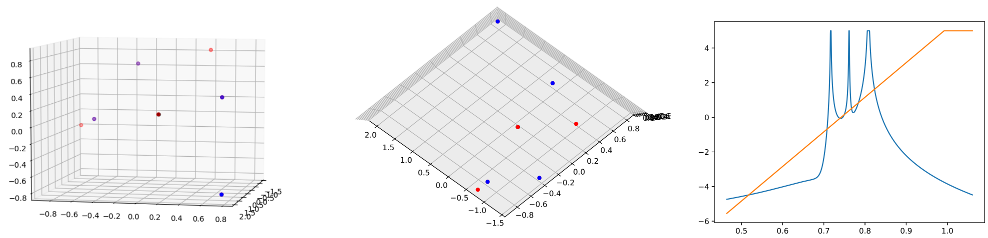


Fig. 4.5 S_0 for $d = 3$ and $p = 2.1$.

Finally, Figure 4.6 shows two experiments in which $|S_0|$ contains exactly 17 elements for $d = 8$.

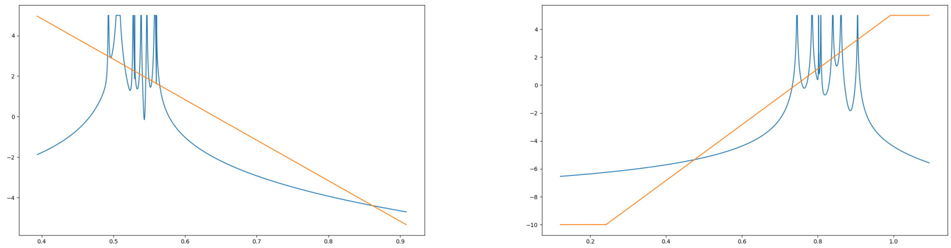


Fig. 4.6 S_0 for $d = 8$, $p = 1.9$ on the left and $p = 2.1$ on the right.

Chapter 5

Entropic approximation for multi-dimensional martingale optimal transport

We study the existing algorithms that solve the multidimensional martingale optimal transport. Then we provide a new algorithm based on entropic regularization and Newton's method. Then we provide theoretical convergence rate results and we check that this algorithm performs better through numerical experiments. We also give a simple way to deal with the absence of convex ordering among the marginals. Furthermore, we provide a new universal bound on the error linked to entropy.

Key words. Martingale optimal transport, entropic approximation, numerics, Newton.

5.1 Introduction

The problem of martingale optimal transport was introduced as the dual of the problem of robust (model-free) superhedging of exotic derivatives in financial mathematics, see Beiglböck, Henry-Labordère & Penkner [18] in discrete time, and Galichon, Henry-Labordère & Touzi [73] in continuous-time. This robust superhedging problem was introduced by Hobson [94], and was addressing specific examples of exotic derivatives by means of corresponding solutions of the Skorokhod embedding problem, see [51, 92, 93], and the survey [91].

Given two probability measures μ, ν on $\mathbb{R}^d$, with finite first order moment, martingale optimal transport differs from standard optimal transport in that the set of all interpolating probability measures $\mathcal{P}(\mu,\nu)$ on the product space is reduced to the subset $\mathcal{M}(\mu,\nu)$ restricted by the martingale condition. We recall from Strassen [146] that $\mathcal{M}(\mu,\nu) \neq \emptyset$ if and only if $\mu \preceq \nu$ in the convex order, i.e. $\mu(f) \leq \nu(f)$ for all convex functions f. Notice that the inequality $\mu(f) \leq \nu(f)$ is a direct consequence of the Jensen inequality, the reverse implication follows from the Hahn-Banach theorem.

This paper focuses on giving numerical aspects of martingale optimal transport for finite marginals. Henry-Labordère [83] used dual linear programming techniques to solve this problem, chosing well the cost functions so that the dual constraints were much easier to check. Alfonsi, Corbetta & Jourdain noticed the difficulty, when going to higher dimension to get a discrete approximation of continuous marginals in convex order, that are still in convex order in higher dimension. So they mainly solve this problem, and then do several optimal transport resolutions with primal linear programming. Guo & Obłój [77] provide convergence results in the one dimensional setting of the discrete problem converges to the continuous problem, and they provide a Bregman projection scheme for solving the martingale optimal transport problem in the one dimensional setting. We also mention Tan & Touzi [148] who used a dynamic programming approach to solve a continuous-time version of martingale optimal transport.

The idea of using Bregman projection comes from classical optimal transport. Christian Leonard [110] was the first to have the idea of introducing an entropic penalization in an optimal transport problem. The entropic penalization makes this problem smooth and strictly convex and gives a Gibbs structure to the optimal probability, which has an explicit formula as a function of the dual optimizer. The unanimous adoption of entropic methods for solving optimal transport problems came from Marco Cuturi [52] who noticed that finding the dual solution of the entropic problem was equivalent to finding two diagonal matrices that made a full matrix bistochastic, therefore allowing to use the celebrated Sinkhorn algorithm.

Historically in classical optimal transport, the practitioners used linear programming algorithm to solve it, such as the Hungarian method [107], the auction algorithm [29], the network simplex [3], we may also mention [76]. However, this method was so costly that only small problems could be treated because of the polynomial cost of linear programming algorithms. Later, Benhamou & Brenier [25] found another way of solving numerically the optimal transport problem by making it a dynamic programming problem with a final penalization on the mismatch of the final marginal of the dynamic process with the target marginal. For particular cases, it was also possible to use the Monge-Ampere equation. In the case of the square distance cost, Brenier [39] proved that the optimal coupling is concentrated on a deterministic map, which was the gradient of a "potential" convex function u. When furthermore the marginals have densities with respect to the Lebesgue measure, we may prove that u is a solution of the Monge-Ampere equation $\det D^2 u = \frac{g \circ c_x(X,\cdot)^{-1} \circ \nabla u}{f}$, where f is the density of μ and g is the density of ν. This equation satisfies a maximum principle, allowing to solve it in practice, see [28] and [27]. We also mention a smart strategy by Merigot [118], using semi-discrete transport. Levy [111] introduced a Newton method to solve the semi-discrete problem very fast.

For the entropic resolution, Leonard [110] proved that the value of the entropic penalized optimal transport converged to the one of the unpenalized problem, while the optimal transports converged as well to a solution of the optimal transport. See [42] and [48] for more precise studies of this convergence in particular cases. It have been observed by [106] that the entropic formulation was particularly useful for numerical resolution, as it allowed to use the celebrated Sinkhorn algorithm [140]. The power of this technique has been rediscovered by [52], and widely adopted by the community, see [142], [131], or [151]. This method has already been adapted to different transport problems, such as Wasserstein barycenters [2] and multi-marginal transport problems [26], gradient flows problems [128], unbalanced transport [44], and one dimensional martingale optimal transport [77].

The remarkable work by Schmitzer [137] gives very practical considerations and tricks on how to actually make the Bregman projection algorithm converge fast and stay stable in practice. Cuturi & Peyre [53] used a quasi-Newton method to solve the smooth entropic optimal transport. Their conclusion seems that the Sinkhorn algorithm is still more effective. However, [36] use an inexact Newton method (i.e. including the use of the second derivative) and manage to beat the performance of the Sinkhorn algorithm. We also mention [6] which introduces a "Greenkhorn algorithm" that outperforms the Sinkhorn algorithm according to their experiment, and similarly [150] introduces an overrelaxed version of the Sinkhorn algorithm that squares the linear convergence coefficient, and accelerates the algorithm.

Our subsequent work differs from Guo & Obłój as we explain how to deal with higher dimension, give a more effective algorithm for martingale optimal transport by inexact Newton method. We also provide a speed of convergence for the Bregman projection algorithm, and explains how to deal with the lack of convex ordering of the marginals. Finally the universal bound that we give for the error linked to the entropy term is much sharper than the previous state-of-the-art. This bound may be extended to classical optimal transport for which it does not seem to be in the literature either.

In this paper we introduce several existing algorithms for solving martingale optimal transport such as linear programming, non-smooth semi-dual optimization, and Bregman projections. We introduce the smooth Newton algorithms, and the Newton semi-implied algorithm. Then we give some theoretical results on the speed of convergence of these algorithms, together with solutions to stabilize them and make them work in practice, like the preconditioning for the Newton method, or how to deal with marginals that are not in convex order. We provide new convergence rates for the entropic approximation of the martingale optimal transport, that are much better than the existing ones. The known result is an error of the order $\varepsilon\Big(\ln(N)-1\Big)$, where N is the size of the discretized grid, while we prove that we can get a result of order $\varepsilon\frac{d}{2}$, where d is the dimension of the space of the problem (1 or 2 in this paper). These rates rely on very strong hypotheses that may be hard to check in practice. However we see on the numerical example that they are well verified in practice.

The paper is organized as follows. Section 5.2 gives the problem to solve, Section 5.3 give the different algorithms that we will compare. In Section 5.4, we provide practical solutions to some usual problems, Section 5.5 provides theoretical convergence rates for the algorithms, Section 5.6 gathers the proofs of the theoretical results, and finally Section 5.7 contains numerical results.

Notation We fix an integer $d \geq 1$.

In all this paper, $\mathbb{R}^d$ is endowed with the Euclidean structure, the Euclidean norm of $x \in \mathbb{R}^d$ will be denoted $|x|$. Let $A \subset \mathbb{R}^d$ we denote $|A|$ the Lebesgue volume of A. The map ι_A is the map equal to 0 on A, and ∞ otherwise. If V is a topological affine space and $A \subset V$ is a subset of V, $\mathrm{int}A$ is the interior of A, $\mathrm{cl}\,A$ is the closure of A, $\mathrm{aff}A$ is the smallest affine subspace of V containing A, $\mathrm{conv}A$ is the convex hull of A, and $\dim(A) := \dim(\mathrm{aff}A)$. Let $(u_\varepsilon)_{\varepsilon>0}, (v_\varepsilon)_{\varepsilon>0} \subset V$. We denote that $u_\varepsilon = o(v_\varepsilon)$ if $\lim_{\varepsilon\to 0}\frac{|u_\varepsilon|}{|v_\varepsilon|} = 0$. We further denote $u_\varepsilon \ll v_\varepsilon$. A classical property of $o(\cdot)$ is that

$$u_\varepsilon = v_\varepsilon + o(v_\varepsilon) \quad \text{if and only if} \quad u_\varepsilon = v_\varepsilon + o(u_\varepsilon). \tag{5.1.1}$$

Let $x_0 \in \mathbb{R}^d$, and $r > 0$, we denote $\mathrm{zoom}_r^{x_0} : x \longmapsto x_0 + rx$, $B_r(x_0)$ is the closed ball centered in x_0 with radius r, and we only write B_r when the center is 0. Let $f : \mathbb{R}^d \longrightarrow \mathbb{R}$, we denote $\|f\|_\infty := \sup_{x\in\mathbb{R}^d} f(x)$ its infinite norm, and $\|f\|_\infty^R := \sup_{x\in B_R} f(x)$ its infinite norm when restricted to the ball B_R, for $R \geq 0$. Let $a, b \in \mathbb{R}^d$, we denote $a \otimes b := ab^t = (a_i b_j)_{1\leq i,j\leq d}$, the only matrix in $\mathcal{M}_d(\mathbb{R})$ such that for all $x \in \mathbb{R}^d$, we have $(a\otimes b)x = (b\cdot x)a$. Let $1 \leq k \leq d+1$ and $u_1, ..., u_k \in \mathbb{R}^d$, we denote $\det_{\mathrm{aff}}(u_1, ..., u_k) := \left|\det\left(\left(e_j\cdot(u_i - u_k)\right)_{1\leq i,j\leq k-1}\right)\right|$, where $(e_j)_{1\leq j\leq k-1}$ is an orthonormal basis of $Vect\Big(u_1 - u_k, ..., u_{k-1} - u_k\Big)$.

Let $M \in \mathcal{M}_d(\mathbb{R})$, a real matrix of size d, we denote $\det M$ the determinant of M. We also denote $Com(M)$ the comatrix of M: for $1 \leq i, j \leq d$, $Com(M)_{i,j} = (-1)^{i+j}\det M^{i,j}$, where $M^{i,j}$ is the matrix of size $d-1$ obtained by removing the i^{th} line and the j^{th} row of M. Recall the useful comatrix formula:

$$Com(M)^t M = MCom(M)^t = (\det M)I_d. \tag{5.1.2}$$

As a consequence, whenever M is invertible, $M^{-1} = \frac{1}{\det M}Com(M)^t$.

We denote $\Omega := \mathbb{R}^d \times \mathbb{R}^d$ and define the two canonical maps

$$X : (x,y) \in \Omega \longmapsto x \in \mathbb{R}^d \quad \text{and} \quad Y : (x,y) \in \Omega \longmapsto y \in \mathbb{R}^d.$$

For $\varphi, \psi : \mathbb{R}^d \longrightarrow \bar{\mathbb{R}}$, and $h : \mathbb{R}^d \longrightarrow \mathbb{R}^d$, we denote

$$\varphi \oplus \psi := \varphi(X) + \psi(Y), \quad \text{and} \quad h^{\otimes} := h(X)\cdot(Y - X),$$

with the convention $\infty - \infty = \infty$.

For a Polish space $\mathcal{X}$, we denote by $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\big(\mathcal{X},\mathcal{B}(\mathcal{X})\big)$. Let $\mathcal{Y}$ be another Polish space, and $\mathbb{P}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$. The corresponding conditional kernel $\mathbb{P}_x$ is defined by:

$$\mathbb{P}(dx,dy)=\mathbb{P}\circ X^{-1}(dx)\mathbb{P}_x(dy).$$

We also use this notation for finite measures. For a measure m on $\mathcal{X}$, we denote $\mathbb{L}^1(\mathcal{X},m):=\{f\in\mathbb{L}^0(\mathcal{X}):m[|f|]<\infty\}$. We also denote simply $\mathbb{L}^1(m):=\mathbb{L}^1(\bar{\mathbb{R}},m)$.

5.2 Preliminaries

Throughout this paper, we consider two probability measures μ and ν on $\mathbb{R}^d$ with finite first order moment, and $\mu\preceq\nu$ in the convex order, i.e. $\nu(f)\ge\mu(f)$ for all integrable convex f. We denote by $\mathcal{M}(\mu,\nu)$ the collection of all probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals $\mathbb{P}\circ X^{-1}=\mu$ and $\mathbb{P}\circ Y^{-1}=\nu$. Notice that $\mathcal{M}(\mu,\nu)\neq\emptyset$ by Strassen [146].

For a derivative contract defined by a non-negative coupling function $c:\mathbb{R}^d\times\mathbb{R}^d\longrightarrow\mathbb{R}_+$, the martingale optimal transport problem is defined by:

$$\mathbf{S}_{\mu,\nu}(c) \;:=\; \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)}\mathbb{P}[c].$$

The corresponding robust superhedging problem is

$$\mathbf{I}_{\mu,\nu}(c) \;:=\; \inf_{(\varphi,\psi,h)\in\mathcal{D}_{\mu,\nu}(c)}\mu(\varphi)+\nu(\psi),$$

where

$$\mathcal{D}_{\mu,\nu}(c) \;:=\; \Big\{(\varphi,\psi,h)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^1(\mu,\mathbb{R}^d):\ \varphi\oplus\psi+h^{\otimes}\ge c\Big\}.$$

The following inequality is immediate:

$$\mathbf{S}_{\mu,\nu}(c)\le\mathbf{I}_{\mu,\nu}(c).$$

This inequality is the so-called weak duality. For upper semi-continuous coupling, we get from Beiglböck, Henry-Labordère, and Penckner [18], and Zaev [160] that there is strong duality, i.e. $\mathbf{S}_{\mu,\nu}(c)=\mathbf{I}_{\mu,\nu}(c)$. For any Borel coupling function bounded from below, Beiglböck, Nutz & Touzi [22] in dimension 1, and De March [57] in higher dimension proved that duality holds for a quasi-sure formulation of dual problem and proved dual attainability thanks to the structure of martingale transports evidenced in [58].

Along all this paper, we assume that μ and ν are discrete, i.e. we may find finite $\mathcal{X}$ and $\mathcal{Y}$ so that $\mu = \sum_{x\in\mathcal{X}} \mu_x \delta_x$, and $\nu = \sum_{y\in\mathcal{Y}} \nu_y \delta_y$, so that all the coordinates of μ and ν are positive. Similarly, duality clearly holds thanks to the finiteness of the support, and the dual problem becomes discretized as well: for $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(c)$, we can denote φ, ψ, and h as vectors $(\varphi(x))_{x\in\mathcal{X}}$, $(\varphi(y))_{y\in\mathcal{Y}}$, and $(h_i(x))_{x\in\mathcal{X},1\leq i\leq d}$.

To solve the martingale transport problem in practice, it seems necessary to discretize the problem. Guo & Oblój [77] prove that the martingale optimal transport problem with continuous μ, ν, and c is a limit of this kind of discrete problem in dimension one under reasonable assumptions. This paper does not focus on proving the convergence of the discretized problem towards the continuous problem, we focus on how to solve the discretized problem.

5.3 Algorithms

5.3.1 Primal and dual simplex algorithm

Primal

The natural strategy to solve this problem will be to use linear programming techniques such as simplex algorithm. One major problem with this approach is that the set $\mathcal{M}(\mu, \nu)$ may be empty, because in practice, the discretization of the marginals may break the convex ordering between then, thus making the set $\mathcal{M}(\mu, \nu)$ empty by Strassen theorem. This problem was relieved by Guo & Obłój [77], and by Alfonsi, Corbetta & Jourdain [5]. In [77], they deal with the problem by replacing the convex ordering constraint by an approximate convex ordering constraint which is more resilient to perturbating the marginals. In [5], they go beyond and gives several algorithms to find measures ν' (resp. μ') that are in convex order with μ (resp. with ν) and satisfy some optimality criteria such as minimality of $\nu - \nu'$ (resp. $\mu - \mu'$) in terms of $p-$Wasserstein distance. We also give in Subsubsection 5.4.3 a technique to avoid this issue.

Dual

One huge weakness of the Primal algorithm is that the size of the problem is $|\mathcal{X}||\mathcal{Y}|$, which is the size of $\mathcal{X} \times \mathcal{Y}$, the support of the probabilities we consider. When $|\mathcal{X}|$ and $|\mathcal{Y}|$ are big, it becomes a problem for memory storage. We notice that the number of constraints is $(d+1)|\mathcal{X}|+|\mathcal{Y}|$, which is much smaller, because the dual functions φ, and h are respectively in $\mathbb{R}^{\mathcal{X}}$ and in $(\mathbb{R}^d)^{\mathcal{X}}$, and the dual function ψ lies in $\mathbb{R}^{\mathcal{Y}}$. This is why in practice it makes sense to solve the Kuhn & Tucker dual problem instead of the primal one. We will see considerations on the speed of convergence in Subsubsection 5.5.4.

5.3.2 Semi-dual non-smooth convex optimization approach

It is well known from classical transport that solving directly the linear programming problem is too costly (see [125]) consequently, some alternative techniques have been developed like the Benamou-Brenier [25] approach, which inspired Tan & Touzi [148] for the continuous time optimal transport problem. The idea consists in solving a Hamilton-Jacobi-Bellman problem with a penalization on the distance between the final marginal and ν. Then an extension of this idea to our two-steps MOT problem gives the following resolution algorithm, suggested by Guo and Obłój [77]. We denote $\mathcal{M}(\mu):=\{\mathbb{P}\in\mathcal{P}(\Omega):\mathbb{P}\circ X^{-1}=\mu,\text{ and }\mathbb{P}[Y|X]=X,\,\mu-\text{a.s.}\}$, and get

$$\begin{aligned}
\mathbf{S}_{\mu,\nu}(c) &:= \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{P}[c]\\
&= \inf_{\psi\in\mathbb{L}^1(\nu)}\sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[c-\psi]+\nu[\psi]\\
&= \inf_{\psi\in\mathbb{L}^1(\nu)} \mu[(c(X,\cdot)-\psi)_{conc}(X)]+\nu[\psi]\\
&= \inf_{\psi\in\mathbb{L}^1(\nu)} V(\psi)
\end{aligned}$$

where $V(\psi):=\mu[(c(X,\cdot)-\psi)_{conc}(X)]+\nu[\psi]$ is a convex function in the variable ψ. Then the problem becomes a simple convex optimization problem. It seems appropriate in these conditions to solve the problem with using a classical gradient descent algorithm. It is proved in [148] that V has an explicit gradient. To give the explicit form of this gradient, we first need to introduce a notion of contact set. Let $f:\mathcal{Y}\longmapsto\mathbb{R}$, as $\mathcal{Y}$ is finite, $f_{conc}(x)=\inf_{f\leq g\text{ affine}}g(x)=\sup\{\sum_i\lambda_i f(y_i):y_1,...,y_{d+1}\in\mathcal{Y},\lambda_1,...,\lambda_{d+1}\geq 0:\sum_i\lambda_i y_i=x\}$. By finiteness of $\mathcal{Y}$, this supremum is a maximum. We denote $\operatorname{argconc}_f(x):=\operatorname{argmax}\{\sum_i\lambda_i f(y_i):y_1,...,y_{d+1}\in\mathcal{Y},\lambda_1,...,\lambda_{d+1}\geq 0:\sum_i\lambda_i y_i=x\}$. Then the subgradient of V at ψ is given by

$$\partial V(\psi)=\left\{\sum_{x\in\mathcal{X}}\mu_x\sum_i\lambda_i(x)\delta_{y_i(x)}-\nu:\Big(y(x),\lambda(x)\Big)\in\operatorname{argconc}_{c(x,\cdot)-\psi}(x),\text{ for all }x\in\mathcal{X}\right\}$$

Notice that this set is a singleton for a.e. $\psi \in \mathbb{L}^1(\mathcal{Y})$, as V is a convex function in finite dimensions. Then with high probability, on each gradient step, the function V will be differentiable on this point. In practice there is always uniqueness after the first step.

5.3.3 Entropic algorithms

The entropic problem in optimal transport

In practice, this problem is added some regularity by the addition of an entropic penalization (see Leonard [110], Cuturi [52]). Let $\varepsilon > 0$,

$$\mathbf{S}^{\varepsilon}_{\mu,\nu}(c) := \sup_{\mathbb{P}\in\mathcal{P}(\mu,\nu)} \mathbb{P}[c] - \varepsilon H(\mathbb{P}|m_0),$$

where $H(\mathbb{P}|m_0) := \int_\Omega \left(\ln\left(\frac{d\mathbb{P}}{dm_0}\right) - 1\right) \frac{d\mathbb{P}}{dm_0} m_0(d\omega)$. The measure m_0 is the "reference measure", we assume that it may be decomposed as $m_0 := m^{\mathcal{X}} \otimes m^{\mathcal{Y}}$ such that μ is dominated by $m^{\mathcal{X}} \in \mathcal{M}(\mathbb{R}^d)$, and ν is dominated by $m^{\mathcal{Y}} \in \mathcal{M}(\mathbb{R}^d)$. For this text we chose $m_0 := \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \delta_{(x,y)}$. By the finiteness of the supports of μ and ν, we know that $\mathbb{P}$ is absolutely continuous with respect to m_0. Denote $p := \frac{d\mathbb{P}}{dm_0}$, and abuse notation writting $p \in \mathcal{P}(\mu,\nu)$. Then $H(\mathbb{P}|m_0) := \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} (\ln p(x,y) - 1)\, p(x,y)$. This problem can be written with Lagrange multipliers,

$$\sup_{\mathbb{P}\in\mathcal{P}(\mu,\nu)} \mathbb{P}[c] - \varepsilon H(\mathbb{P}|m_0) = \inf_{(\varphi,\psi)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)} \sup_{\mathbb{P}\in\mathcal{P}(\mu,\nu)} \mathbb{P}[c - \varphi\oplus\psi] - \varepsilon H(\mathbb{P}|m_0) + \mu[\varphi] + \nu[\psi].$$

which leads to an explicit Gibbs form for the optimal kernel p. Then as the supports are finite we easily get the shape of the optimizer $p(x,y) = \exp\left(-\frac{\varphi(x)+\psi(y)-c(x,y)}{\varepsilon}\right)$, and the associated dual problem becomes

$$\mathbf{I}^{\varepsilon}_{\mu,\nu}(c) := \inf_{(\varphi,\psi)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)} \mu[\varphi] + \nu[\psi] + \varepsilon \sum_{x,y} \exp\left(-\frac{\varphi(x)+\psi(y)-c(x,y)}{\varepsilon}\right).$$

One important property that we need is the $\Gamma-$convergence. We say that F_ε $\Gamma-$converges to F when $\varepsilon \longrightarrow 0$ if for all sequence $\varepsilon_n \longrightarrow 0$, we have
(i) For all sequences $x_n \longrightarrow x$, we have $F(x) \geq \limsup_n F_{\varepsilon_n}(x_n)$.
(ii) There exists a sequence $x_n \longrightarrow x$ such that $F(x) \leq \liminf_n F_{\varepsilon_n}(x_n)$.

The $\Gamma-$convergence implies that $\min F_n \longrightarrow F$, when $n \longrightarrow \infty$, and that if x_n is a minimizer of F_n for all $n \geq 1$, and if $x_n \longrightarrow x$, then x is a minimizer of F. Leonard [110] proved this $\Gamma-$convergence of the penalized problem to the optimal transport problem.

The Bregman iterations algorithm

Coupled with the Sinkhorn algorithm [140] introduced by Marco Cuturi for optimal transport [52], this method allows an exponentially fast approximated resolution. Notice that the operator $V_\varepsilon(\varphi,\psi) := \mu[\varphi]+\nu[\psi]+\varepsilon\sum_{x,y}\exp\left(-\frac{\varphi(x)+\psi(y)-c(x,y)}{\varepsilon}\right)$ is smooth convex. The Euler-Lagrange equations $\partial_\varphi V_\varepsilon = 0$ (resp. $\partial_\psi V_\varepsilon = 0$) are exactly equivalent to the marginal relations $\mathbb{P}\circ X^{-1}=\mu$ (resp. $\mathbb{P}\circ Y^{-1}=\nu$). It was noticed in [52] that these partial optimizations can be obtained in closed form:

$$\varphi(x) = \varepsilon\ln\left(\frac{1}{\mu_x}\sum_y \exp\left(-\frac{\psi(y)-c(x,y)}{\varepsilon}\right)\right),$$

and

$$\psi(y) = \varepsilon\ln\left(\frac{1}{\nu_y}\sum_x \exp\left(-\frac{\varphi(x)-c(x,y)}{\varepsilon}\right)\right).$$

By iterating these partial optimization, we obtain the so-called Sinkhorn algorithm (see [140]) that is equivalent to a block optimization of the smooth function V_ε which dual is called Bregman projection [38], and converges exponentially fast, see Knight [104].

The entropic approach for the one period martingale optimal transport problem

As observed by Guo & Obłój [77] in dimension 1, the Sinkhorn algorithm can be extended to the martingale optimal transport problem. With exactly the same computations, we get

$$\begin{aligned}\mathbf{S}^\varepsilon_{\mu,\nu}(c) &:= \sup_{\mathbb{P}\in\mathcal{M}(\mu,\nu)} \mathbb{P}[c]-\varepsilon H(\mathbb{P}|m_0)\\ =\mathbf{I}^\varepsilon_{\mu,\nu}(c) &:= \inf_{(\varphi,\psi,h)\in\mathbb{L}^1(\mu)\times\mathbb{L}^1(\nu)\times\mathbb{L}^0(\mathbb{R}^d)} \mu[\varphi]+\nu[\psi]\\ &\qquad +\varepsilon\sum_{x,y}\exp\left(-\frac{\varphi(x)+\psi(y)+h^{\otimes}(x,y)-c(x,y)}{\varepsilon}\right).\end{aligned}$$

First notice that the $\Gamma-$convergence still holds in this easy finite case.

Proposition 5.3.1. *Let $F_\varepsilon : \mathbb{P}\in\mathcal{M}(\mu,\nu)\longmapsto \mathbb{P}[c]-\varepsilon H(\mathbb{P}|m_0)$. For $\varepsilon>0$, F_ε is strictly concave upper semi-continuous. Furthermore, F_ε $\Gamma-$converges to F_0 when $\varepsilon\longmapsto 0$.*

Proof. This $\Gamma-$convergence is easy by finiteness as the entropy is bounded by $\ln(|\mathcal{X}||\mathcal{Y}|)-1$ when $\mathbb{P}$ is a probability measure. $\square$

We denote $\Delta := \varphi\oplus\psi+h^{\otimes}-c$, the convex function to minimize becomes

$$V_\varepsilon(\varphi,\psi,h) := \mu[\varphi]+\nu[\psi]+\varepsilon\sum_{x,y}\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right).$$

Then the Sinkhorn algorithm is complemented by another step so as to account for the martingale relation:

$$\begin{aligned}
\varphi(x) &= \varepsilon\ln\left(\frac{1}{\mu_x}\sum_y\exp\left(-\frac{\psi(y)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon}\right)\right), \qquad (5.3.3)\\
\psi(y) &= \varepsilon\ln\left(\frac{1}{\nu_y}\sum_x\exp\left(-\frac{\varphi(x)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon}\right)\right),\\
0 &= \frac{1}{\mu_x}\sum_y\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right)(y-x)\\
&= \frac{-1}{\mu_x}\frac{\partial}{\partial h(x)}\varepsilon\sum_y\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right).
\end{aligned}$$

Notice that the martingale step is not closed form and is only implied. However, it may be computed almost as fast as φ, and ψ, thanks to the Newton algorithm applied to each smooth strongly convex function F_x of d variables given, for each $x\in\mathcal{X}$ with its derivatives by

$$\begin{aligned}
F_x(h) &:= \varepsilon\sum_y\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right), \qquad (5.3.4)\\
\nabla F_x(h) &= -\sum_y\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right)(y-x),\\
D^2F_x(h) &= \frac{1}{\varepsilon}\sum_y\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right)(y-x)\otimes(y-x).
\end{aligned}$$

Notice that the optimization of F_{x_1} and F_{x_2} are independent for $x_1\neq x_2$.

Truncated Newton method

For these problems, it may make sense to use a Newton method, as the problems are smooth, and the Newton method converges very fast. For very highly dimensional problems (here $(d+1)|\mathcal{X}|+|\mathcal{Y}|$), the inversion of the hessian is too much costly. Then it is in general preferred to use quasi-Newton. Instead of computing the Newton step $D^2 V_\varepsilon^{-1} \nabla V_\varepsilon$, we use a conjugate gradient algorithm to find by iterations a vector $p \in \mathcal{D}_{\mathcal{X},\mathcal{Y}}$ such that $|D^2 V_\varepsilon p - \nabla V_\varepsilon|$ is small enough, generally in practice this quantity is chosen to be smaller than $\min\left(\frac{1}{2}, \sqrt{|\nabla V_\varepsilon|}\right)$.

The conjugate gradient algorithm approximates the solution of the equation $Ax = b$ by solving it "direction by direction" along the most important direction, until a stopping criterion is reached. The exact algorithm may be found in [159].

Implied truncated Newton method

Some instabilities may appear from Newton steps as any term of the form $\exp(X/\varepsilon)$ can easily explode when ε is very small and $X > 0$. The dimension may also make the conjugate gradient from the quasi-Newton algorithm slow. A good way to avoid this problem and exploit the near-closed formula for the optimal φ and h when ψ is fixed, or optimal ψ when φ and h are fixed.

Instead of applying the truncated method to $V_\varepsilon(\varphi, \psi, h)$, we apply the truncated Newton method to $\widetilde{V}_\varepsilon(\psi) := \min_{\varphi,h} V_\varepsilon(\varphi, \psi, h)$. It is elementary that with these definitions we have

$$\inf_{\varphi,\psi,h} V_\varepsilon(\varphi, \psi, h) = \inf_{\psi} \widetilde{V}_\varepsilon(\psi).$$

Doing this variable implicitation is easy by the fact that we have a closed formula for φ and a quasi-closed formula for h. It brings the great advantage of having the first marginal and the martingale relationship verified, this fact will be exploited in Subsubsection 5.4.3.

Now we give a general framework that allows to use variables implicitation. The following framework should be used with $F = V_\varepsilon$, $x = \psi$, and $y = (\varphi, h)$. Proposition 5.3.2 below provides the appropriate convexity result together with the closed formulas for the two first derivatives of $\widetilde{V}_\varepsilon$ that are necessary to apply the truncated Newton algorithm. Let $\mathcal{A}$ and $\mathcal{B}$ finite dimensional spaces and $F : \mathcal{A} \times \mathcal{B} \longrightarrow \mathbb{R}$, we say that F is $\alpha-$convex if

$$\lambda F(\omega_1) + (1-\lambda) F(\omega_2) - F\Big(\lambda \omega_1 + (1-\lambda)\omega_2\Big) \geq \alpha \frac{\lambda(1-\lambda)}{2} |\omega_1 - \omega_2|^2,$$

for all $\omega_1, \omega_2 \in \mathcal{A} \times \mathcal{B}$, and $0 \leq \lambda \leq 1$. The case $\alpha = 0$ corresponds to the standard notion of convexity.

Proposition 5.3.2. *Let $F : \mathcal{A} \times \mathcal{B} \longrightarrow \mathbb{R}$ be a $\alpha-$convex function. Then the map $\tilde{F} : x \longmapsto \inf_{y \in \mathcal{B}} F(x, y)$ is $\alpha-$convex. Furthermore, if $\alpha > 0$ and F is C^2, then $y(x) :=$*

$\operatorname{argmin}_y F(x,y)$ *is unique and we have*

$$\begin{aligned}\nabla y(x) &= -\partial_y^2 F^{-1}\partial_{yx}^2 F\Big(x,y(x)\Big),\\ \nabla \tilde{F}(x) &= \partial_x F\Big(x,y(x)\Big),\\ D^2\tilde{F}(x) &= \Big(\partial_x^2 F - \partial_{xy}^2 F\partial_y^2 F^{-1}\partial_{yx}^2 F\Big)\Big(x,y(x)\Big).\end{aligned}$$

The proof of Proposition 5.3.2 is reported in Subsection 5.6.1.

Remark 5.3.3. *The matrix $\partial_{xy}^2 F\partial_y^2 F^{-1}\partial_{yx}^2 F$ is symmetric positive definite, therefore the curvature of the function $\tilde{F}$ is reduced by the implicitation process, making heuristically the minimization easier. This fact is also observed in practice.*

This method shall be used for the optimization of V_ε, but also for the optimization of F_x that gives the martingale step, see (5.3.4). Indeed the value of $\varphi(x)$ does not change the martingale optimality of F_x. We provide these important formulas.

The map $\widetilde{V}_\varepsilon$ and its derivatives: Let $\psi\in\mathbb{R}^{\mathcal{Y}}$, we denote $(\widehat{\varphi}_\psi^\varepsilon,\widehat{h}_\psi^\varepsilon):=\operatorname{argmin}_{\varphi,h} V_\varepsilon(\varphi,\psi,h)$, that are unique and may be found in quasi-closed form from (5.3.3). Now we give the formula for $\widetilde{V}_\varepsilon$ and its derivatives. We directly get from Proposition 5.3.2 that

$$\begin{aligned}\widetilde{V}_\varepsilon(\psi) &= V_\varepsilon\left(\widehat{\varphi}_\psi^\varepsilon,\psi,\widehat{h}_\psi^\varepsilon\right),\\ \nabla\widetilde{V}_\varepsilon(\psi) &= \partial_\psi V_\varepsilon\left(\widehat{\varphi}_\psi^\varepsilon,\psi,\widehat{h}_\psi^\varepsilon\right),\\ D^2\widetilde{V}_\varepsilon &= \left(\partial_\psi^2 V_\varepsilon - \partial_{\psi,\varphi}^2 V_\varepsilon(\partial_\varphi^2 V_\varepsilon)^{-1}\partial_{\varphi,\psi}^2 V_\varepsilon - \sum_{i=1}^d \partial_{\psi,h_i}^2 V_\varepsilon(\partial_{h_i}^2 V_\varepsilon)^{-1}\partial_{h_i,\psi}^2 V_\varepsilon\right)\left(\widehat{\varphi}_\psi^\varepsilon,\psi,\widehat{h}_\psi^\varepsilon\right).\end{aligned}$$

The last additive decomposition of $\partial_{(\varphi,h)^2}V_\varepsilon^{-1}$ stems from the fact that $\partial_{(\varphi,h)^2}V_\varepsilon\left(\widehat{\varphi}_\psi^\varepsilon,\psi,\widehat{h}_\psi^\varepsilon\right)$ is diagonal. Indeed, V_ε is a sum of functions of $(\varphi(x),h(x))$ for $x\in\mathcal{X}$, and the crossed derivative $\partial_{\varphi(x),h(x)}V_\varepsilon = \sum_y (y-x)\exp\left(-\frac{\Delta(x,y)}{\varepsilon}\right)$ cancels at $\left(\widehat{\varphi}_\psi^\varepsilon(x),\widehat{h}_\psi^\varepsilon(x)\right)$ by the martingale property induced by the optimality in $h(x)$. The same holds for $\partial_{h_i(x),h_j(x)}V_\varepsilon$

for $i \neq j$. We denote $\Delta_\psi^\varepsilon := \widehat{\varphi}_\psi^\varepsilon \oplus \psi + \left(\widehat{h}_\psi^\varepsilon\right)^\otimes - c$, and we have

$$\begin{aligned}
\widetilde{V}_\varepsilon(\psi) &= \mu\left[\widehat{\varphi}_\psi^\varepsilon\right] + \nu[\psi] + \varepsilon \sum_{x,y} \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right),\\
\nabla \widetilde{V}_\varepsilon(\psi) &= \left(\nu_y - \sum_x \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)_{y\in\mathcal{Y}},\\
D^2\widetilde{V}_\varepsilon &= \varepsilon^{-1} diag\left(\sum_x \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)\\
&\quad -\varepsilon^{-1}\sum_x \left(\sum_y \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)^{-1}\\
&\quad\times\left(\exp\left(-\frac{\Delta_\psi^\varepsilon(x,y_1)}{\varepsilon}\right)\exp\left(-\frac{\Delta_\psi^\varepsilon(x,y_2)}{\varepsilon}\right)\right)_{y_1,y_2\in\mathcal{Y}}\\
&\quad -\varepsilon^{-1}\sum_x\sum_{i=1}^d \left(\sum_y (y_i-x_i)^2 \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)^{-1}\\
&\quad\times\left((y_1-x)_i \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y_1)}{\varepsilon}\right)(y_2-x)_i\exp\left(-\frac{\Delta_\psi^\varepsilon(x,y_2)}{\varepsilon}\right)\right)_{y_1,y_2\in\mathcal{Y}}.
\end{aligned}$$

Notice that for the conjugate gradient algorithm, we only need to be able to compute the product $\left(D^2\widetilde{V}_\varepsilon\right)p$ for $p\in\mathbb{R}^\mathcal{Y}$. Then if the RAM is not sufficient to store the whole matrix $D^2\widetilde{V}_\varepsilon$, it may be convenient to only store $D_\psi := diag\left(\sum_x \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)$, $D_\varphi^{-1} := diag\left(\sum_y \exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)^{-1}$, and $D_{h_i}^{-1} := diag\left(\sum_y (y_i-x_i)^2\exp\left(-\frac{\Delta_\psi^\varepsilon(x,y)}{\varepsilon}\right)\right)^{-1}$ for all $1 \le i \le d$. Then we compute $\left(D^2\widetilde{V}_\varepsilon\right)p$ in the following way:

$$
\begin{aligned}
p_\psi &:= \left(D_\psi\right)p\\
p_\varphi &:= \left(\sum_y \exp\left(-\frac{\Delta^\varepsilon_\psi(x,y)}{\varepsilon}\right)p_y\right)_{x\in\mathcal{X}},\\
p_{h_i} &:= \left(\sum_y (y_i-x_i)\exp\left(-\frac{\Delta^\varepsilon_\psi(x,y)}{\varepsilon}\right)p_y\right)_{x\in\mathcal{X}},\\
p'_\varphi &:= \left(D_\phi^{-1}\right)p_\varphi,\\
p'_{h_i} &:= \left(D_{h_i}^{-1}\right)p_{h_i},\\
p''_\varphi &:= \left(\sum_x \exp\left(-\frac{\Delta^\varepsilon_\psi(x,y)}{\varepsilon}\right)(p''_\varphi)_x\right)_{y\in\mathcal{Y}},\\
p''_{h_i} &:= \left(\sum_x (y_i-x_i)\exp\left(-\frac{\Delta^\varepsilon_\psi(x,y)}{\varepsilon}\right)(p''_{h_i})_x\right)_{y\in\mathcal{Y}},\\
\left(D^2\widetilde{V}_\varepsilon\right)p &= \varepsilon^{-1}\left(p_\psi - p''_\varphi - \sum_{i=1}^d p''_{h_i}\right).
\end{aligned}
$$

The map $\widetilde{F}_x$ and its derivatives: In this paragraph we fix $\psi\in\mathbb{R}^{\mathcal{Y}}$ and $\varepsilon>0$. Recall the map F_x from (5.3.4). This map may be seen as a function of $\big(\varphi(x),h(x)\big)$. Then we set

$$\widetilde{F}_x(h) := \min_{\varphi(x)\in\mathbb{R}} \mu_x\varphi(x) + \varepsilon\sum_y \exp\left(-\frac{\varphi(x)+\psi+h\cdot(y-x)-c(x,y)}{\varepsilon}\right).$$

The optimizer is given by (5.3.3), hence by the closed formula

$$\widehat{\varphi}_h(x) := \operatorname*{argmin}_{\varphi(x)\in\mathbb{R}} \mu_x\varphi(x) + F_x\big(\varphi(x),h\big) = \varepsilon\ln\left(\frac{1}{\mu_x}\sum_y \exp\left(-\frac{\psi(y)+h\cdot(y-x)-c(x,y)}{\varepsilon}\right)\right).$$

A direct application of Proposition 5.3.2 gives

$$
\begin{aligned}
\widetilde{F}_x(h) &= \min_{\varphi(x)\in\mathbb{R}} \mu_x\widehat{\varphi}_h(x) + \varepsilon\sum_y \exp\left(-\frac{\widehat{\Delta}_h(x,y)}{\varepsilon}\right),\\
\nabla\widetilde{F}_x(h) &= -\sum_y (y-x)\exp\left(-\frac{\widehat{\Delta}_h(x,y)}{\varepsilon}\right),\\
D^2\widetilde{F}_x(h) &= \varepsilon^{-1}\left(\sum_y (y-x)^{2\otimes}\exp\left(-\frac{\widehat{\Delta}_h(x,y)}{\varepsilon}\right) - \mu_x^{-1}\left(\sum_y (y-x)\exp\left(-\frac{\widehat{\Delta}_h(x,y)}{\varepsilon}\right)\right)^{2\otimes}\right),
\end{aligned}
$$

where we denote $\widehat{\Delta}_h(x,y) := \widehat{\varphi}_h(x) + \psi(y) + h\cdot(y-x) - c(x,y)$, and $u^{2\otimes} := u\otimes u$ for $u\in\mathbb{R}^d$.

5.4 Solutions to practical problems

5.4.1 Preventing numerical explosion of the exponentials

As we want to make ε go to 0, all the terms like $\exp\left(\frac{\cdot}{\varepsilon}\right)$ tend to explode numerically. Here are the different risks that we have to deal with, and how we deal with them.

First the Newton algorithm is very local, and nothing guarantees that after one iteration, the value function will not explode. From our practical experience, the algorithm tends to explode for $\varepsilon < 10^{-3}$. Notice that the numerical experiment given by [36] does not go beyond 10^{-3}, we may imagine that this is because they do not use the variable implicitation technique. Furthermore, we notice from our numerical experimenting that this variable implicitation, additionaly to the stabilizing the numerical scheme, makes the convergence of the Newton algorithm much faster. Moreover, impliciting in φ and h is much more effective than impliciting in ψ, even though we have to do the implicitation in h which is much more costly than the implicitation in ψ.

For the computation of the implicitations (5.3.3), the computation of the formula $\varphi(x) = \varepsilon\ln\left(\frac{1}{\mu_x}\sum_y\exp\left(-\frac{\psi(y)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon}\right)\right)$ should be done as follows to prevent numerical explosion. First we compute $M_x := \max_{y\in\mathcal{Y}}\left\{-\frac{\psi(y)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon}\right\}$, and then the computation that we do effectively is

$$\varphi(x) \;=\; M_x + \varepsilon\ln\left(\sum_y \exp\left(-\frac{\psi(y)+h(x)\cdot(y-x)-c(x,y)}{\varepsilon} - M_x\right)\right) - \varepsilon\ln\mu_x \qquad (5.4.5)$$

In (5.4.5), the exponential arguments are always smaller than 1, and one of them is equal to 1, then any explosion makes the exponential be totally negligible when compared to $\exp(0) = 1$, this computation rule makes it very stable. Notice also the separation of $\ln\mu_x$ that allows to treat the case when the value of μ_x is extremely low (like for exemple when you discretise a Gaussian measure on a grid) even if in this case, it may be smarter to just remove the value from the grid.

Notice that the variable implicitation should also be used during each partial optimisation in $h(x)$ for $x\in\mathcal{X}$, as this Newton algorithm is highly susceptible to explode as well. The implicitation simply consists in minimizing in $\varphi(x)$ the maps F_x from (5.3.4), and has a closed form.

Another thing to take care of about h is the initial value taken for the next partial optimization of V_ε in h. On a first hand, chosing the last value for h helps diminishing the number of steps for the optimization. Also, when ε is very small, even with the implicitation, the Newton optimization may get hard if the initial value is too far from the optimum.

5.4.2 Customization of the Newton method

Preconditioning

The conjugate gradient algorithm used to compute the search direction for the Newton algorithm has a convergence rate given by $|x_k - x^*|_A \leq 2\left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^k |x_0 - x^*|_A$, where x_k is the $k-$th iterate, x^* is the solution of the problem, $|x|_A := x^t Ax$ is the Euclidean norm associated to the scalar product A, and $\kappa(A) := \|A\|\|A^{-1}\|$ is the conditioning of A. This conditioning is the fundamental parameter for this convergence speed. When ε is getting small, the conditioning raises. We also observe on the numerics that is happens when the marginals have a thin tail (e.g. Gaussian distributions). The simplest way of dealing with this conditioning problem consists in applying a "preconditioning" algorithm. We find a matrix P that is easy to invert (for example a diagonal matrix) and we use the fact that solving $Ax = b$ is equivalent with solving $P^t APx' = P^t b$, where $x' := P^{-1}x$. We use the most classical and simple preconditioning which consists in taking $P := \sqrt{diag(A)}^{-1}$. See [159] for the precise algorithm.

Line search

An important advantage of the Bregman projection algorithm over the primitive Newton algorithm is that V_ε is a Lyapunov function as the steps only consist of block minimizations of this function, whereas the Newton step may get very wrong and lose the optimal region if we are not close enough to the minimum. However in practice, some ingredients need to be added to the Newton step. Indeed, once the direction of search is decided by the conjugate gradient algorithm, in practice it is necessary to make a line search algorithm, i.e. to find a point on the line on which the value function V_ε is strictly smaller, and so does the directional gradient absolute value $|\nabla V_\varepsilon \cdot p|$, where p is the descent direction. This "descent" condition is called the Wolfe condition. A very good line search algorithm that is commonly used in practice is detailed in [159].

Remark 5.4.1. *Notice that if a value is rejected by the line search, it is important to throw away the value of h given by this wrong point, and to come back to the last value of h corresponding to a point that was not rejected by the line search.*

5.4.3 Penalization

The penalized problem

The dual solution may not be unique, which may lead to numerical unstabilities. As an example we may add any constant to φ while subtracting it to ψ without changing the value of V. A straightforward solution is to add a penalization to the minimization problem. I.e. we have the new problem

$$\min_{\psi\in\mathbb{R}^{\mathcal{Y}}} \widetilde{V}_\varepsilon(\psi)+\alpha f(\psi) \tag{5.4.6}$$

where f is a strictly convex superlinear function, so that there is a unique minimum by the fact that the gradient of $\widetilde{V}_\varepsilon$ is a difference of probabilities, which proves that this convex function is Lipschitz, whence the strict convexity and super-linearity of $\widetilde{V}_\varepsilon(\psi)+\alpha f(\psi)$. In practice we take $f(\psi):=\frac{1}{2}\sum_{y\in\mathcal{Y}}a_y\psi(y)^2$, for some $a\in\mathbb{R}^{\mathcal{Y}}$, so that $\nabla f(\psi)=\sum_{y\in\mathcal{Y}}a_y\psi(y)e_y$, where $(e_y)_{y\in\mathcal{Y}}$ is the canonical basis, and $Df(\psi)=\mathrm{diag}(a)$ have these easy closed expressions. In practice we take $a=(1)$, $a=\nu$, $a=\nu^2$, or $a=\nu/\psi_0$, where φ_0 is a fixed estimate of ψ from the last step of $\varepsilon-$scaling (see Subsection 5.4.4).

Marginals not in convex order

Problem 5.4.6 allows to solve the problem of mismatch in the convex ordering thanks to the following theorem that allows for probability measures μ,ν not in convex order to find another probability measure $\tilde{\nu}$ in convex order with μ that satisfies some optimality criterion, for example in terms of distance from ν.

Theorem 5.4.2. *Let $(\mu,\nu)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ not in convex order. Let $\nu_\alpha:=\mathbb{P}_\alpha\circ Y^{-1}$, where $\mathbb{P}_\alpha$ is the optimal probability for Problem* (5.4.6)*, where f is a super-linear, differentiable, strictly convex, and $p-$homogeneous function $\mathbb{R}^{\mathcal{Y}}\longrightarrow\mathbb{R}$ for some $p>1$. Then $\nu_\alpha\longrightarrow\nu_l$ when $\alpha\longrightarrow 0$, for some $\nu_l\succeq_c\mu$ satisfying*

$$f^*(\nu_l-\nu)=\min_{\tilde{\nu}\succeq_c\mu} f^*(\tilde{\nu}-\nu).$$

The proof of Theorem 5.4.2 is reported in Subsection 5.6.2. Notice that for $f(\psi) := \frac{1}{2}\sum_{y\in\mathcal{Y}} a_y\psi(y)^2$, we have $f^*(\gamma) = \frac{1}{2}\sum_{y\in\mathcal{Y}} a_y^{-1}\gamma(y)^2$, whence the idea of taking $a = \nu^2$.

Conjugate gradient improvement and stabilization

Adding a penalization also allows to accelerate the conjugate gradient algorithm, indeed it reduces the conditioning of the Hessian matrix by killing the small eigenvalues, and therefore accelerates the conjugate gradient algorithm's convergence. It also stabilizes this algorithm, indeed when ε is small we observe that without penalization, the numerical error may cause instabilities by returning a non positive definite Hessian. Adding the positive definite Hessian of the penalization function bypasses this instability.

5.4.4 Epsilon scaling

For all entropic algorithms, we observe that when ε is small, the algorithm may be very slow to find the region of optimality. For the Bregman projection, the formula for the speed of convergence in Subsection 5.5.4 suggests to have a strategy of $\varepsilon-$scaling: i.e. we solve the problem for $\varepsilon = 1$, so that the function to optimize is very smooth. Then solve the problem for $\varepsilon' < \varepsilon$, with the previous optimum as an initial point. We continue this algorithm until we reach the desired value for ε. In practice we divide ε by 2 at each step.

5.4.5 Grid size adaptation

It may be a huge loss of time to run the algorithm on full resolution since the beginning of $\varepsilon-$scaling. To prevent this waste of time, Schmitzer [137] suggests to raise the size of the grid at the same time than shrinking ε. In practice we give to each new point of the grid for φ, ψ, and h the value of the closest point in the previous grid. We use heuristic criteria to decide when to doble the size of the grid, avoiding for example to doble is when ε is too small as is seriously challenges the stability of the resolution scheme.

5.4.6 Kernel truncation

While ε shrinks to 0, we observe that the optimal transport tends to concentrate on graphs, as suggested in [56]. Because of the exponential, the value of the optimal probability far enough to these graphs tends to become completely negligible. For this reason, Schmitzer [137] suggests to truncate the grid in order to do much less calculation. In dimensions higher than 1, the gain in term of number of operation may quickly reach a factor 100 for small ε. In practice we removed the points in the grid when their probability were smaller than $10^{-7}\mu_x$ (resp. $10^{-7}\nu_y$) for all $x \in \mathcal{X}$ (resp. for all $y \in \mathcal{Y}$).

5.4.7 Computing the concave hull of functions

We were not able to find algorithms that compute the concave hull of a function in the literature, so we provide here the one we used. Let $f : \mathcal{Y} \longrightarrow \mathbb{R}$.

In dimension 1 the algorithm is linear in $|\mathcal{Y}|$, we use the McCallum & Avis [117] algorithm to find the points of the convex hull of the upper graph of f in a linear time and then we go through these points until we find the two consecutive points $y_1, y_2 \in \mathcal{Y}$ around the convex hull such that $y_1 < x \leq y_2$. Then $f_{conc}(x) = \frac{y_2 - x}{y_2 - y_1} f(y_1) + \frac{x - y_1}{y_2 - y_1} f(y_2)$.

In higher dimension we use Algorithm 1 in order to compute the convex hull of a function. We do not know if a better algorithm exists, but this one should be the fastest when the active points of the convex hull are already close to the maximum, this will be useful to compute $(c(x,\cdot) - \psi)_{conc}(x)$ from Theorem 5.5.5 below, so as the field "gradient" of the result that allows to find the right h. We believe that the complexity of this algorithm is quadradic in the (not so improbable) worst case of a concave function, $O\Big(n\ln(n)\Big)$ on average for a "random" function, and linear when the guess of the gradient is good. These assesments are formal and based on the observation of numerics, we do not prove anything about Algorithm 1, not even the fact that it cannot go on infinite loops. We provide it for the reader who would like to reproduce the numerical experiments without having to search for an algorithm by himself.

5.5 Convergence rates

5.5.1 Discretization error

Proposition 5.5.1. *Let $\mu \preceq \nu$ in convex order in $\mathcal{P}(\mathbb{R}^d)$ having a dual optimizer $(\varphi, \psi, h) \in \mathcal{D}_{\mu,\nu}(c)$ such that φ is $L_\varphi-$Lipschitz, and ψ is $L_\psi-$Lipschitz. Then for all $\mu' \preceq \nu'$ in convex order in $\mathcal{P}(\mathbb{R}^d)$ having a dual optimizer $(\varphi', \psi', h') \in \mathcal{D}_{\mu,\nu}(c)$ such that φ' is $L_{\varphi'}-$Lipschitz, and ψ' is $L_{\psi'}-$Lipschitz, we have*

$$\Big|\mathbf{S}_{\mu,\nu}(c) - \mathbf{S}_{\mu',\nu'}(c)\Big| \leq \max\Big(L_\varphi, L_{\varphi'}\Big) W_1(\mu', \mu) + \max\Big(L_\psi, L_{\psi'}\Big) W_1(\nu', \nu)$$

Algorithm 1 Concave hull of f.

```
procedure ConcaveHull(f, x, grid, gradientGuess)
    if gradientGuess is None then
        grad ← vector of zeros with the same size than x
        gridF ← f(grid)
    else
        grad ← gradientGuess
        gridF ← f(grid) − grad · grid
    y ← argmax gridF
    support ← [y]
    gridF ← gridF − gridF[y_0]
    while True do
        if x ∈ aff support then
            bary ← barycentric coefficients of x in the basis support
            if bary are all > 0 then
                value ← sum(bary × f(support))
                return {"value" : value; "support" : support;
                    "barycentric coefficients" : bary; "gradient" : grad}
            else
                i ← argmin bary
                remove entry i in support
                remove entry i in bary
        else
            projx ← orthogonal projection of x on aff support
            p = x − projx
            scalar ← p · (grid − x)
            if scalar are all ≤ 0 then
                Fail with error "x not in the convex hull of grid."
            y ← argmax{gridF/scalar such that scalar > 0}
            add y to support
            a ← −gridF[y]/scalar[y]
            gridF ← gridF + a × scalar
            grad ← grad − a × p
```

The proof of Proposition 5.5.1 is reported in Subsection 5.6.3.

Remark 5.5.2. *Guo & Obłój [77] provide a very similar result in Proposition 2.2. It is however very different as they need to introduce an approximately martingale optimal transport problem, and our result makes hypotheses on the Lipschitz property of the dual optimizers, which are unknown, and even their existence in unknown. In dimension* 1*, thanks to the work by Beiglböck, Lim & Obłój [21], we may prove the existence of these Lipschitz dual, thanks to some regularity assumptions on* c*. In higher dimension, there are ongoing investigations about the existence of similar results. However, by Example 4.1 in [57], it will be necessary to make assumptions on* μ,ν *as well, as the smoothness of* c *cannot guarantee the existence of a dual optimizer. Proposition 5.5.1 is entitled to be a proposition of practical use, we may formally assume that the partial dual functions that we get converge to the continuous dual and assume that their Lipschitz constant converges to the Lipschitz constant of the limit.*

We refer to Subsection 2.2 in [77] for a study of the discrete $\mathcal{W}_1-$approximation of the continuous marginals. In dimensions higher than 3, it is necessary to use a Monte-Carlo type approximation of μ and ν to avoid the curse of dimensionality linked to a grid type approximation. However, Proposition 5.5.1 is not well-adapted to estimate the error, as we know from [71] that the Wasserstein distance between a measure and its Monte-Carlo estimate is of order $n^{\frac{-1}{d}}$. Next proposition deals with this issue. For two sequences $(u_N)_{N\geq 0}$ and $(v_N)_{N\geq 0}$, we denote $u_N \approx v_N$ when $N \longrightarrow \infty$ if u_N/v_N converges to 1 in probability, when $N \longrightarrow \infty$.

Proposition 5.5.3. *Let* $\mu \preceq \nu$ *in convex order in* $\mathcal{P}(\mathbb{R}^d)$ *having a dual optimizer* $(\varphi,\psi,h) \in \mathcal{D}_{\mu,\nu}(c)$*, and* μ_N *and* ν_M*, independent Monte-Carlo estimates of* μ *and* ν *with* N *and* M *samples. If furthermore* $\mu'_N \preceq \nu'_M$ *is in convex order in* $\mathcal{P}(\mathbb{R}^d)$ *having a dual optimizer* $(\varphi_N,\psi_M,h') \in \mathcal{D}_{\mu'_N,\nu'_M}(c)$ *such that when* $N,M \longrightarrow \infty$*, we have*
(i) $(\mu_N-\mu)[\varphi] \approx (\mu_N-\mu)[\varphi_N]$*,* (ii) $(\nu_M-\nu)[\psi] \approx (\nu_M-\nu)[\psi_M]$*,*
(iii) $(\mu_N-\mu'_N)[\varphi] \approx (\mu_N-\mu'_N)[\varphi_N]$*,* (iv) $(\nu_M-\nu'_M)[\psi] \approx (\nu_M-\nu'_M)[\psi_M]$.

Then we have

$$\left|\mathbf{S}_{\mu,\nu}(c)-\mathbf{S}_{\mu',\nu'}(c)\right| \leq \alpha\sqrt{\frac{\mathrm{Var}_\mu[\varphi]}{N}+\frac{\mathrm{Var}_\nu[\psi]}{M}}+\left|(\mu_N-\mu'_N)[\varphi_N]\right|+\left|(\nu_M-\nu'_M)[\psi_M]\right|,$$

with probability converging to $1-2\int_\alpha^\infty \exp\left(-x^2/2\right)dx$*, when* $N,M \longrightarrow \infty$*.*

The proof of Proposition 5.5.3 is reported in Subsection 5.6.3.

Remark 5.5.4. *In Proposition 5.5.3, we introduce μ'_N, ν'_M because the Monte-Carlo approximation will not conserve the convex order for μ_N, ν_N in general. Then we obtain μ'_N, ν'_M from "convex ordering processes" such as the one suggested in Subsection 5.4.3, or the one suggested in [5]. In both cases, the quantity $\left|(\mu_N - \mu'_N)[\varphi_N]\right| + \left|(\nu_M - \nu'_M)[\psi_M]\right|$ may be computed exactly numerically.*

5.5.2 Entropy error

In this subsection, m_ε is a generic finite measure and no assumptions are made on μ_ε and ν_ε.

Theorem 5.5.5. *Let $\mu_\varepsilon \preceq \nu_\varepsilon$ in convex order in $\mathcal{P}(\mathbb{R}^d)$ with dual optimizer $(\varphi_\varepsilon, \psi_\varepsilon, h_\varepsilon) \in \mathcal{D}_{\mu_\varepsilon,\nu_\varepsilon}(c)$ to the $\varepsilon-$entropic dual problem with reference measure m_ε, such that we may find $\gamma, \eta, \beta > 0$, sets $(D_\varepsilon^{\mathcal{X}}, D_\varepsilon^{\mathcal{Y}})_{\varepsilon>0} \subset \mathcal{B}(\mathbb{R}^d)$, and parameters $(\alpha_\varepsilon, A_\varepsilon)_{\varepsilon>0} \subset (0,\infty)$, such that if we denote $r_\varepsilon := \varepsilon^{\frac{1}{2}-\eta}$, $m_\varepsilon^{\mathcal{X}} := m_\varepsilon \circ X^{-1}$, and $\Delta_\varepsilon := \varphi \oplus \psi + h^\otimes - c$, for $\varepsilon > 0$ small enough we have:*
(i) *$\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}(x) \le \varepsilon^{-\gamma}$, $\mu_\varepsilon-$a.e., $m_\varepsilon[\Omega] \le \varepsilon^{-\gamma}$, and $A_\varepsilon \ll \varepsilon^{-q}$ for all $q > 0$.*
(ii) *For $m_\varepsilon-$a.e. $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$, we have $(m_\varepsilon)_x [B_{\alpha_\varepsilon}(y)] \ge \varepsilon^\gamma$, and for $(m_\varepsilon)_x-$a.e. $y' \in B_{\alpha_\varepsilon}(y)$, we have $|\Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,y')| \le \gamma\varepsilon\ln(\varepsilon^{-1})$.*
(iii) *$\mu_\varepsilon\left[(D_\varepsilon^{\mathcal{X}})^c\right] \ll 1/\ln(\varepsilon^{-1})$ and for all $x \in D_\varepsilon^{\mathcal{X}}$ we may find $k_x^\varepsilon \in \mathbb{N}$, $S_x^\varepsilon \in (B_{A_\varepsilon})^{k_x^\varepsilon}$, and $\lambda_x^\varepsilon \in [0,1]^{k_x^\varepsilon}$ with $\det_{\mathrm{aff}}(S_x^\varepsilon) \ge A_\varepsilon^{-1}$, $\min \lambda_x^\varepsilon \ge A_\varepsilon^{-1}$, and $\sum_{i=1}^{k_x^\varepsilon} \lambda_{x,i}^\varepsilon S_{x,i}^\varepsilon = x$, convex combination.*
(iv) *On $B_{r_\varepsilon}(S_x^\varepsilon)$, $\Delta_\varepsilon(x,\cdot)$ is C^2, $A_\varepsilon^{-1} I_d \le \partial_y^2 \Delta_\varepsilon(x,\cdot) \le A_\varepsilon I_d$, and for all $y, y' \in B_{r_\varepsilon}(S_x^\varepsilon)$, we have $|\partial_y^2\Delta_\varepsilon(x,y) - \partial_y^2\Delta_\varepsilon(x,y')| \le \varepsilon^\eta$.*
(v) *For $x \in D_\varepsilon^{\mathcal{X}}$ and $y \notin B_{r_\varepsilon}(S_x^\varepsilon)$, we have that $\Delta_\varepsilon(x,y) \ge \sqrt{\varepsilon}\,\mathrm{dist}(y, S_x^\varepsilon)$.*
(vi) *$\nu_\varepsilon\left[(D_\varepsilon^{\mathcal{Y}})^c\right] \ll \frac{1}{A_\varepsilon \ln(\varepsilon^{-1})}$ and for all $y_0 \in D_\varepsilon^{\mathcal{Y}}$, $R, L \ge 1$, and $f : \mathbb{R}^d \longrightarrow \mathbb{R}_+$ such that $\frac{f}{\|f\|_\infty^R}$ is $L-$Lipschitz, we have*

$$\left| \int_{B_R} f(y) \left[\frac{d(m_\varepsilon)_x \circ \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_0}}{(m_\varepsilon)_x[B_{R\sqrt{\varepsilon}}(y_0)]} - \frac{dy}{|B_R|} \right] \right| \le [R+L]^\gamma \varepsilon^\beta \int_{B_R} f(y) \frac{dy}{|B_R|}.$$

Then if we denote $\mathbb{P}_\varepsilon := e^{-\frac{\Delta_\varepsilon}{\varepsilon}} m_\varepsilon$, we have

$$\lim_{\varepsilon \to 0} \frac{\mu_\varepsilon[\bar{\varphi}_\varepsilon] + \nu_\varepsilon[\psi_\varepsilon] - \mathbb{P}_\varepsilon[c]}{\varepsilon} = \frac{d}{2}, \quad \textit{where} \quad \bar{\varphi}_\varepsilon := \Big(c(X,\cdot) - \psi_\varepsilon\Big)_{conc}(X).$$

The proof of Theorem 5.5.5 is reported to Subsection 5.6.4.

Corollary 5.5.6. *Under the assumptions of 5.5.5, we have that* $\mathbb{P}_\varepsilon[c] \geq \mathbf{S}_{\mu,\nu}(c) - \frac{d}{2}\varepsilon + o(\varepsilon)$*, when* $\varepsilon \longrightarrow 0$.

Proof. We fix $h(x) \in \partial(c(x,\cdot)-\psi_\varepsilon)_{conc}(x)$ for all $x \in \mathcal{X}$, then

$$\Big((c(X,\cdot)-\psi_\varepsilon)_{conc}(X), \psi_\varepsilon, h\Big) \in \mathcal{D}_{\mu_\varepsilon,\nu_\varepsilon}(c),$$

and therefore $\mu_\varepsilon\Big[(c(X,\cdot)-\psi_\varepsilon)_{conc}(X)\Big] + \nu_\varepsilon[\psi_\varepsilon] \geq \mathbf{I}_{\mu_\varepsilon,\nu_\varepsilon}(c) \geq \mathbf{S}_{\mu_\varepsilon,\nu_\varepsilon}(c) \geq \mathbb{P}_\varepsilon[c]$. Theorem 5.5.5 concludes the proof. □

Figure 5.1 gives numerical examples of the convergence of the duality gap when ε converges to 0. In these graphs, the blue curve gives the ratio of the dominator of the duality gap $\mu_\varepsilon\Big[\Big(c(X,\cdot)-\psi_\varepsilon\Big)_{conc}(X)\Big] + \nu_\varepsilon[\psi_\varepsilon] - \mathbb{P}_\varepsilon[c]$ with respect to ε. It is meant to be compared to the green flat curve which is its theoretical limit $\frac{d}{2}$ according to Theorem 5.5.5. Finally the orange curve provides the ratio of the weaker dominator $\mu_\varepsilon\Big[\sup\Big(c(X,\cdot)-\psi_\varepsilon\Big)\Big] + \nu_\varepsilon[\psi_\varepsilon] - \mathbb{P}_\varepsilon[c]$ of the duality gap with respect to ε. The interest of this last weaker dominator is that it avoids computing the concave envelop which may be a complicated issue, while having a reasonable comparable performance in practice than the concave hull dominator as showed by the graphs and by Remark 5.5.13 below.

Figure 5.1a provides these curves for the one-dimensional cost function $c := XY^2$, μ uniform on $[-1,1]$, and $\nu = |Y|^{1.5}\mu$. The grid size adaptation method is used and the size of the grid goes from 10 when $\varepsilon = 1$ to 10000 when $\varepsilon = 10^{-5}$. Figure 5.1b provides these curves for the two-dimensional cost function $c : (x,y) \in \mathbb{R}^2 \times \mathbb{R}^2 \longmapsto x_1(y_1^2 + 2y_2^2) + x_2(2y_1^2 + y_2^2)$, μ uniform on $[-1,1]^2$, and $\nu = (|Y_1|^{1.5} + |Y_2|^{1.5})\mu$. The grid size adaptation method is used again and the size of the grid goes from 10×10 when $\varepsilon = 1$ to 160×160 when $\varepsilon = 10^{-4}$.

Remark 5.5.7. *The hypotheses of regularity are impossible to check at the current state of the art. One element that could argue in this direction is the fact that the limit of* Δ_ε *when* $\varepsilon \longrightarrow 0$*, that we shall denote* Δ_0*, should satisfy the Monge-Ampère equation* $\det\Big(\partial_y^2\Delta_0\Big) = f$*, similar to classical transport [153] that may provide some regularity, but less than the one needed to satisfy (ii) of Theorem 5.5.5, see Chapter 5 of [80]. It also justifies, in the case where* $f > 0$*, that* $\partial_y^2\Delta_\varepsilon \in GL_d(\mathbb{R})$.

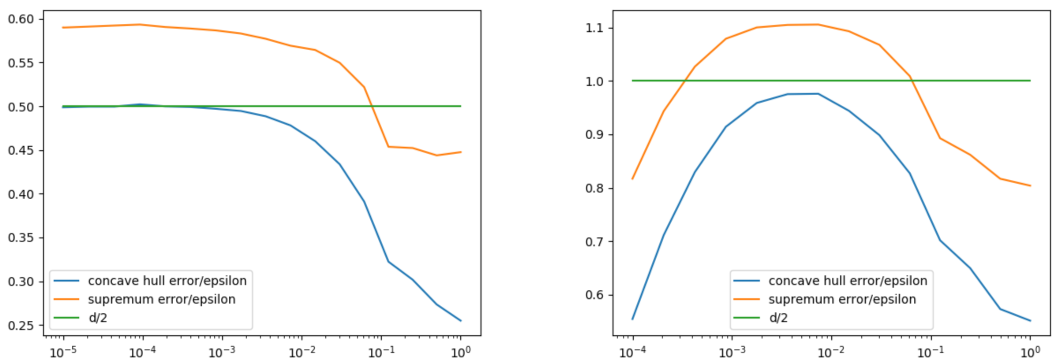
Remark 5.5.8. *Assumption (vi) on the local convergence of the reference measures is justified is we take a regular grid for* $\mathcal{Y}$ *that becomes fine fast enough. If the grid does not become fine fast enough, then the local decrease of* $\Delta_\varepsilon - \Delta_\varepsilon(x) - \nabla\Delta_\varepsilon(x)\cdot(Y-x)$

is exponential because of the shape of the kernel. We observe on the experiments that in this case the duality gap becomes indeed smaller, however as a downside, the convergence of the scheme becomes much less effective because the marginal error stays high (see Proposition 5.5.1). See Figure 5.1b.

Remark 5.5.9. *Assumption (ii) from the theorem seems to hold in one dimension, but it seems to be wrong in two dimensions, as shown in the example of Figure 5.4. However, we may still find a formula similar to* (5.5.7) *below. Therefore, we may reasonably assume that the error is still of order ε, as confirmed in a the numerical examples by Figure 5.1b.*

Remark 5.5.10. *In the case when $\mathcal{X}$ is obtained from Monte-Carlo methods, (vi) is verified with a constant that depends on the point, and is probabilistic. the exponent α may be taken taken equal to $\frac{1}{d}$, see [71].*

Remark 5.5.11. *Despite the difficulty to check the assumptions, this result is inspired and satisfied by observation on the numerics, we have tried with several cost functions, differentiable or not, and the result seems to be always satisfied, probably with the help of its universality. See figure 5.1.*



(a) Dimension 1. (b) Dimension 2.

Fig. 5.1 Duality gap for the supremum, and the concave hull dual approximation vs ε.

Remark 5.5.12. *An easier version of Theorem 5.5.5 could be obtained by the same strategy for classical optimal transport under the twist condition $c_x(x,\cdot)$ injective, replacing $\left(c(X,\cdot)-\psi_\varepsilon\right)_{conc}(X)$ by $\sup\left(c(X,\cdot)-\psi_\varepsilon\right)$.*

Remark 5.5.13. *By similar reasoning, we may prove that even for martingale optimal transport, replacing $\left(c(X,\cdot)-\psi_\varepsilon\right)_{conc}(X)$ by $\sup\left(c(X,\cdot)-\psi_\varepsilon\right)$ would still give a good* *result (see 5.1). We may formally estimate the new limit of* $\frac{\text{duality gap}}{\varepsilon}$*: for all* $y_i^\varepsilon(x)$ *that are not the optimizer (say* $y_0^\varepsilon(x)$*), the weight added to* $\frac{d}{2}$ *is* $\lambda_i^\varepsilon(x)\Big(\Delta_\varepsilon(x,y_i^\varepsilon(x))-\Delta_\varepsilon(x,y_0^\varepsilon(x))\Big)$*. By using the tools of the proof of Theorem 5.5.5, we get the formal formula, if we denote* λ_i *for the limit of* $\lambda_i^\varepsilon(X)$*, and* y_i *for the limit of* $y_i^\varepsilon(x)$*, we have*

$$\mu_\varepsilon\Big[\sup\{c(X,\cdot)-\psi_\varepsilon\}(X)\Big]+\nu_\varepsilon[\psi_\varepsilon]-\mathbb{P}_\varepsilon[c]\approx\alpha\varepsilon, \tag{5.5.7}$$

with $\alpha=\frac{d}{2}+\int_{\mathbb{R}^d}\sum_{i>0}\lambda_i\ln\left(\frac{dm^{\mathcal{Y}}\big(y_i\big)}{dm^{\mathcal{Y}}\big(y_0\big)}\frac{\lambda_0\det\big(\partial_y^2\Delta_0(x,y_0)\big)}{\lambda_i\det\big(\partial_y^2\Delta_0(x,y_i)\big)}\right)d\mu$*. Then we could reasonably make the assumption the the second term in* α *does not explode, and then the limit is still of the order of* ε*, as we may see on the numerical experiments of Figure 5.1. This result also generalises to optimal transport when there are several transport maps.*

5.5.3 Penalization error

Proposition 5.5.14. *Let* $(\mu,\nu)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ *in convex order. Let* $\nu_\alpha:=\mathbb{P}_\alpha\circ Y^{-1}$*, where* $\mathbb{P}_\alpha$ *is the optimal probability for the entropic dual implied problem with an additional penalization* αf*, where* f *is a super-linear, strictly convex, and differentiable function* $\mathbb{R}^{\mathcal{Y}}\longrightarrow\mathbb{R}$*. Then let* ψ_0 *be the only optimizer for the entropic dual implied problem with minimal* $f(\psi)$*, we have*

$$\frac{\nu_\alpha-\nu}{\alpha}\longrightarrow\nabla f(\psi_0),\quad\textit{when}\quad\alpha\longrightarrow 0.$$

The proof of Proposition 5.5.14 is reported to Subsection 5.6.5.

5.5.4 Convergence rates of the algorithms

Convergence rate for the simplex algorithm

Precise results on the convergence rate of the simplex algorithm is an open problem. Roos [133] gave an example in which the convergence takes an exponential time in the number of parameters. However the simplex algorithm is much more efficient in practice, Smale [141] proved that in average, the number of necessary steps in polynomial in the number of entries, and Spielman & Teng [144] refined this analysis by including the number of constraints in the polynomial.

However, none of these papers provide the real time of convergence of this algorithm. Schmitzer [137] reports that this algorithm is not very useful in practice as it only allows to solve a discretized problem with no more than hundreds of points.

Convergence rate for the semi-dual algorithm

We notice that any subgradient of this function is a difference of probabilities, and then the gradient is bounded. Furthermore the function V is a supremum of a finite number of affine functions, and therefore it does not have a smooth second derivative. In this condition the best theoretical way to optimize this function is by a gradient descent with a step size of order $O(1/\sqrt{n})$ at the $n-$th step, see Ben-Tal & Nemirovski [24]. Then by Theorem 5.3.1 of [24], the rate of convergence is $O(1/\sqrt{n})$ as well, which is quite slow. Furthermore, the time of computation of one step needs to compute one convex hull which has the average complexity $O(|\mathcal{Y}|\ln(|\mathcal{Y}|))$ for each $x \in \mathcal{X}$, and $O(|\mathcal{Y}|)$ in dimension 1, see Subsection 5.4.7. However, we give in Subsection 5.4.7) an algorithm that computes the concave hull in a linear time if the relying points of the concave hull do not change too much. Then let us be optimistic and assume that the computation of one concave hull is on average $O(|\mathcal{Y}|)$, then we have that the complexity is $O(|\mathcal{X}||\mathcal{Y}|)$ operations for each step. Although this algorithm is highly parallelizable, its complexity imposes to the grid to be very coarse. Indeed, to get a precision of 10^{-2}, we need an order of 10^4 operations. We shall see that the entropic algorithms are much more performing for this low precision.

Notice that even though the best theoretical algorithm is the last gradient descent, Lewis & Overton [112] showed that in most case, quasi-Newton methods converge faster, even when the convex function is non-smooth, however they find a particular case in which quasi-Newton fails at being better. The L-BFGS method is a quasi-Newton method that is adapted to high-dimensions problems. The Hessian (even though it does not exist) is approximated by a low-dimensional estimate, and the classical Newton step method is applied. See [159] for the exact algorithm. We see on simulations that this algorithm is indeed much more efficient.

Even if the quasi-Newton algorithm gives better results, the smooth entropic algorithms are much more effective in practice.

Convergence rate for the Sinkhorn algorithm

In practice, if we want to observe the transport maps from [56] to have a good precision on the estimation of the support of the optimal transport, we need to set $\varepsilon = 10^{-4}$. The rate of convergence of the Sinkhorn algorithm is given by κ^{2n} after n iterations for some $0 < \kappa < 1$, see [104]. This result is extended by [77] to the one-dimensional martingale Sinkhorn algorithm. However, we have $\kappa := \frac{\sqrt{\theta}-1}{\sqrt{\theta}+1}$, and

$$\theta := \max_{(x_1,y_1),(x_2,y_2)\in\Omega} \exp\left(\frac{c(x_1,y_1)+c(x_2,y_2)-c(x_2,y_1)-c(x_1,y_2)}{\varepsilon}\right),$$

in the case of classical transport. Then θ is of the order of $\exp(K(c)/\varepsilon)$ for some map $K(c)$ bounded from below. For $\varepsilon = 10^{-5}$, this θ is so big that κ^2 is so close to 1 that κ^{2n}, with n the number of iterations will remain approximately equal to 1.

We also see in practice for the martingale Sinkhorn algorithm that the rate of convergence is not exponential for small values of ε, see Figure 5.2 in the numerical experiment part, as the graph is logarithmic in the error, an exponential convergence rate would be characterized by a straight line. However we observe that for the Bregman projection algorithm we do not have a straight line during the first part of the iteration for the one-dimensional case, and it never happens in the two-dimensional case.

In this regime of ε small, another convergence theory looks to have a better fit with this algorithm. The Sinkhorn algorithm may be interpreted as a block coordinates descent for the optimization of the map $V_\varepsilon(\varphi,\psi)$. We optimize alternatively in φ, and in ψ. We know from Beck & Tetruashvili [15] that this optimization problem has a speed of convergence given by $\frac{LR(x_0)^2}{n}$, where $R(x_0)$ is a quantity that is of the order of $|x_0 - x^*|$ in practice, where x^* is the closest optimizer of V_ε and L is the Lipschitz constant of the gradient. This speed is more comparable to the convergence observed in practice. More precisely, L is of the order of $1/\varepsilon$. This formula shows that in order to minimize the problem for a very small ε, we first need to make $R(x_0)$ small to compensate L. This can be done by minimizing the problem for larger ε. In practice, we divide ε by 2 until we reach a sufficiently small ε. Then we make the grid finer as ε becomes small, and exploit the sparsity in the problem that appears when ε gets small. See Schmitzer [137].

We may apply the same theory for the martingale V_ε and its block optimization in (φ,h) and in ψ. Let $\mathcal{D}_{\mathcal{X},\mathcal{Y}} := \{(\varphi,\psi,h) \in \mathbb{R}^{\mathcal{X}} \times \mathbb{R}^{\mathcal{Y}} \times (\mathbb{R}^d)^{\mathcal{X}}\} \approx \mathbb{R}^{(d+1)|\mathcal{X}|+|\mathcal{Y}|}$, and for $x := (\varphi,\psi,h) \in \mathcal{D}_{\mathcal{X},\mathcal{Y}}$, let $\Delta(x) := (\varphi \oplus \psi + h^{\otimes})_{\mathcal{X}\times\mathcal{Y}}$.

Theorem 5.5.15. *Let $x_0 = (\varphi_0, \psi_0, h_0) \in \mathcal{D}_{\mathcal{X},\mathcal{Y}}$ such that $\left(e^{-\frac{\varphi_0\oplus\psi_0+h_0^{\otimes}-c}{\varepsilon}}\right)_{\mathcal{X}\times\mathcal{Y}}$ sums to 1, and for $n \geq 0$, let the n^{th} iteration of the martingale Sinkhorn algorithm:*

$$\begin{aligned} x_{n+1/2} &:= \left(\varphi_n, \psi_{n+1} := \operatorname*{argmin}_{\psi} V_\varepsilon(\phi_n, \cdot, h_n), h_n\right), \\ x_{n+1} &:= \left(\varphi_{n+1} := \operatorname*{argmin}_{\varphi} V_\varepsilon(\cdot, \psi_{n+1}, \cdot), \psi_{n+1}, h_{n+1} := \operatorname*{argmin}_{h} V_\varepsilon(\cdot, \psi_{n+1}, \cdot)\right). \end{aligned}$$

Furthermore let $\mathbb{P}_0 \in \mathcal{M}(\mu,\nu)$ and let $\mathcal{X}^*$ be the minimizing affine space of V_ε and let V_ε^* be its minimum, then we have

$$\begin{aligned} V_\varepsilon(x_n) - V_\varepsilon^* &\leq \frac{\beta R(x_0)^2 \varepsilon^{-1}}{n}, & (5.5.8) \\ V_\varepsilon(x_n) - V_\varepsilon^* &\leq \left(1 - \frac{2|\mathcal{X}|}{\beta\lambda_2} e^{-\frac{D(x_0)}{\varepsilon}}\right)^n \left(V_\varepsilon(x_0) - V_\varepsilon^*\right), & (5.5.9) \\ \textit{and} \quad \operatorname{dist}(x_n, \mathcal{X}^*) &\leq \sqrt{\frac{\lambda_2\varepsilon}{|\mathcal{X}|} e^{\frac{D(x_0)}{\varepsilon}}} \left(V_\varepsilon(x_n) - V_\varepsilon^*\right)^{\frac{1}{2}}, \end{aligned}$$

where $\beta := \left(2\ln(2) - 1\right)^{-1} \approx 2{,}6$, $\lambda_p := |\mathcal{X}|^{-1} \inf_{x\in\mathcal{D}_{\mathcal{X},\mathcal{Y}}:\Delta(x)\neq 0} \sup_{\Delta(\widetilde{x})=\Delta(x)} \frac{|\Delta(x)|_p^p}{|\widetilde{x}|_p^p}$, for $p = 1, 2$,
$D(x_0) := \lambda_1 \max(1, \|\mathcal{Y} - \mathcal{X}\|_\infty) \frac{V_\varepsilon(x_0) - \mathbb{P}_0[c] - \varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}}$,
$(\mathbb{P}_0)_{\min} := \min_{x\in\mathcal{X}\times\mathcal{Y}} \mathbb{P}_0[\{x\}]$, $\|\mathcal{Y} - \mathcal{X}\|_\infty := \sup_{x\in\mathcal{Y}-\mathcal{X}} |y - x|_\infty$,

and we may choose $R(x_0)$ among $R(x_0) := \left\{ \begin{array}{c} \sup_{V_\varepsilon(x)\leq V_\varepsilon(x_0)} \operatorname{dist}(x, \mathcal{X}^*) \\ \sqrt{\frac{\lambda_2\varepsilon}{|\mathcal{X}|} e^{\frac{D(x_0)}{\varepsilon}}} \left(V_\varepsilon(x_0) - V_\varepsilon^*\right)^{\frac{1}{2}} \\ 2\lambda_1 \frac{V_\varepsilon(x_0) - \mathbb{P}_0[c] - \varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}} \\ \sup_{k\geq 0} \operatorname{dist}(x_k, \mathcal{X}^*) \end{array} \right.$.

The proof of Theorem 5.5.15 is reported in Subsection 5.6.6.

Remark 5.5.16. *By the same arguments, we may prove a similar theorem for the case of optimal transport with $\lambda_1 = \min\left(1, \frac{|\mathcal{Y}|}{|\mathcal{X}|}\right)$, and $\lambda_2 = \frac{1+\left(\frac{|\mathcal{Y}|}{|\mathcal{X}|}\right)^2}{1+\frac{|\mathcal{Y}|}{|\mathcal{X}|}}$.*

Remark 5.5.17. *The theoretical rate of convergence given by Theorem 5.5.15 becomes pretty bad when $\varepsilon \longrightarrow 0$. We observe it in practise when we apply this algorithm with a small ε and a starting point $x_0 = 0$. This emphasizes the need of using the epsilon scaling trick, of Subsection 5.4.4.*

Remark 5.5.18. *Depending on the experiment, in some cases we observe a linear convergence like* (5.5.8) *(see Figure 5.2a), however in other cases, we observe a convergence speed that looks more like* (5.5.9) *(see Figure 5.2b). However, the convergence rates that we provide here are generic, if we wanted to have convergence rates that look more like the one observed, we would need to look for the asymptotic convergence rates like it was suggested by Peyré in [129] for the case of classical transport.*

Remark 5.5.19. *The positive probability* $\mathbb{P}_0 \in \mathcal{M}(\mu,\nu)$ *is necessary. We know from [58] that for some (possibly elementary)* $\mu \preceq \nu$ *in convex order, we way find* $(x_0,y_0) \in \mathcal{X}\times\mathcal{Y}$ *such that* $\mathbb{P}[\{(x_0,y_0)\}] = 0$ *for all* $\mathbb{P}\in\mathcal{M}(\mu,\nu)$ *even thought* $\mu[\{x_0\}] > 0$ *and* $\nu[\{y_0\}] > 0$ *(see Example 2.2 in [58]). Therefore, in this situation there is no optimal* $x^* \in \mathcal{D}_{\mathcal{X}\times\mathcal{Y}}$*, as this would mean that* $\Delta(x^*)_{x_0,y_0} = \infty$*.*

Convergence rate for the Newton algorithm

When the current point gets close enough from the optimum, the convergence rate of the Newton algorithm is quadratic if the hessian is Lipschitz, i.e. $|x_k - x^*|$ and $|\nabla V_\varepsilon(x_k)|$ both converge quadratically to 0, see Theorem 3.5 in [159]. The truncated Newton is a bit slower, but still has a superlinear convergence rate, see Theorem 7.2 in [159].

Convergence rate for the implied Newton algorithm

The important parameter for Newton algorithm is the Lipschitz constant of the Hessian of the objective function. However in the case of variable implicitation, the presence of $\partial_y^2 F^{-1}$ in the Hessian, and the addition of the variation of $y(x)$ in the Lipschitz analysis may kill the Lipschitz property of the Hessian of $\tilde{F}$. The following proposition solves this problem.

Proposition 5.5.20. *Let* $(\tilde{x}_n)_{n\geq 0}$ *the Newton iterations applied to* $\tilde{F}$ *starting from* $\tilde{x}_0 := x_0 \in \mathcal{X}$*. Now let* $(x_n)_{n\geq 0}$ *the sequence defined by recurrence by* $x_0 := x_0$*, then for all* $n \geq 0$*,* $y_n := y(x_n)$*, and let* (x,y) *be the result of a Newton step from* (x_n,y_n)*, and we set* $x_{n+1} := x$*.*

Then $(\tilde{x}_n)_{n\geq 0} = (x_n)_{n\geq 0}$*.*

The proof of Proposition 5.5.20 is reported to Subsection 5.6.7. This proposition implies that the theoretical convergence of the Newton algorithm on F can be extended to the Newton algorithm applied to $\tilde{F}$, indeed the partial minimization in y only decreases the distance from the current point to the minimum around the minimum.

In practice we observe that the convergence for this implied algorithm is much faster and much more stable than the non-implied Newton algorithm.

5.6 Proofs of the results

5.6.1 Minimized convex function

Proof of Proposition 5.3.2 Let $x_1, x_2 \in \mathcal{A}$, $y_1, y_2 \in \mathcal{B}$, and $0 \le \lambda \le 1$. We have

$$\begin{aligned}\lambda F(x_1,y_1)+(1-\lambda)F(x_2,y_2) &\ge F\Big(\lambda(x_1,y_1)+(1-\lambda)(x_2,y_2)\Big)\\ &\quad +\alpha\frac{\lambda(1-\lambda)}{2}|(x_1,y_1)-(x_2-y_2)|^2\\ &\ge \tilde{F}(\lambda x_1+(1-\lambda)x_2)+\alpha\frac{\lambda(1-\lambda)}{2}|x_1-x_2|^2.\end{aligned}$$

By minimizing over y_1 and y_2, we get

$$\lambda\tilde{F}(x_1,y_1)+(1-\lambda)\tilde{F}(x_2,y_2)\ge\tilde{F}(\lambda x_1+(1-\lambda)x_2)+\alpha\frac{\lambda(1-\lambda)}{2}|x_1-x_2|^2,$$

which establishes the $\alpha-$convexity of $\tilde{F}$.

Now if we further assume that $\alpha > 0$ and F is C^2, $y \longmapsto F(x,y)$ is $\alpha-$convex, and therefore strictly convex and super-linear. Hence, there is a unique minimizer $y(x)$. Using the first order derivative condition of this optimum, we have $\partial_y F\Big(x,y(x)\Big)=0$. By the fact that $\partial_y^2 F$ is positive definite (bigger than αId by $\alpha-$convexity), we may apply the local inversion theorem, which proves that $y(x)$ is C^1 in the neighborhood of x. We also obtain $\partial_{yx}^2 F\Big(x,y(x)\Big)+\partial_y^2 F\Big(x,y(x)\Big)\nabla y(x)=0$, which gives the following expression of ∇y:

$$\nabla y(x) = -\partial_y^2 F^{-1}\partial_{yx}^2 F\Big(x,y(x)\Big).$$

Now we may compute the derivatives of $\tilde{F}$. By definition, we have $\tilde{F}(x)=F\Big(x,y(x)\Big)$, then just differentiating this expression, we get

$$\begin{aligned}\nabla\tilde{F}(x) &= \partial_x F\Big(x,y(x)\Big)+\partial_y F\Big(x,y(x)\Big)\nabla y(x)\\ &= \partial_x F\Big(x,y(x)\Big),\end{aligned}$$

where the second equality comes from the fact that $\partial_y F\Big(x,y(x)\Big)=0$ because $y(x)$ is a minimizer. Finally we get the Hessian by deriving again this expression and injecting the value of $\nabla y(x)$. □

5.6.2 Limit marginal

Proof of Theorem 5.4.2 Let $\alpha > 0$, we are considering the following minimization problem:

$$\begin{aligned}
\inf_{\psi} \widetilde{V}_\varepsilon(\psi) + \alpha f(\psi) &= \inf_{\varphi,\psi,h} \mu[\varphi] + \nu[\psi] + \varepsilon \int e^{\frac{-\Delta}{\varepsilon}} dm_0 + \alpha f(\psi) \\
&= \inf_{\psi} \sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[c-\psi] - \varepsilon H(\mathbb{P}|m_0) + \nu[\psi] + \alpha f(\psi) \qquad (5.6.10) \\
&= \sup_{\mathbb{P}\in\mathcal{M}(\mu)} \inf_{\psi} \mathbb{P}[c] - \varepsilon H(\mathbb{P}|m_0) + (\nu - \mathbb{P}\circ Y^{-1})[\psi] + \alpha f(\psi) \\
&= \sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[c] - \varepsilon H(\mathbb{P}|m_0) - (\alpha f)^*(\mathbb{P}\circ Y^{-1} - \nu) \\
&= \sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[c] - \varepsilon H(\mathbb{P}|m_0) - \alpha^{-\frac{1}{p-1}} f^*(\mathbb{P}\circ Y^{-1} - \nu), \quad (5.6.11)
\end{aligned}$$

where the first equality comes from a mutualisation of the infima, the second comes from a partial dualisation of the infimum in φ, h in a supremum over $\mathbb{P} \in \mathcal{M}(\mu)$, we obtain the third equality by applying the minimax theorem and reordering the terms, the fourth equality the definition of the Fenchel-Legendre transform, and the fifth and final equality is just a consequence of the transformation of a multiplyer of a $p-$homogeneous function by the Fenchel-Legendre conjugate. Let $(\alpha_n)_{n\geq 1}$ converging to 0. As $\mathcal{Y}$ is finite, the set $\mathcal{P}(\mathcal{Y})$ is compact. Then we may assume up to extracting a subsequence that ν_{α_n} converges to some limit ν_l. The first order optimality equation for all $y \in \mathcal{Y}$ gives that $\nu - \nu_{\alpha_n} + \alpha_n \nabla f(\psi_n)$, where ψ_n is the unique optimizer of $\widetilde{V}_\varepsilon + \alpha f$. By the $p-$homogeneity of f, the gradient ∇f is $(p-1)-$homogeneous. Then we have the convergence $\widetilde{\psi}_n := \frac{\psi_n}{\alpha_n^{\frac{1}{p-1}}} \underset{n\to\infty}{\longrightarrow} \psi_l := \nabla f^{-1}(\nu_l - \nu)$. As we have by (5.6.10)

$$\inf_{\psi} \widetilde{V}_\varepsilon(\psi) + \alpha_n f(\psi) = \inf_{\psi} \sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[c-\psi] - \varepsilon H(\mathbb{P}|m_0) + \nu[\psi] + \alpha_n f(\psi)$$

By dividing this equation by $\alpha_n^{\frac{1}{p-1}}$, we have that ψ_l is the minimizer of the strictly convex function $\sup_{\mathbb{P}\in\mathcal{M}(\mu)} \mathbb{P}[-\psi] + \nu[\psi] + f(\psi)$, it is therefore unique. Then ν_l is unique as well. By (5.6.10), $\mathbb{P}_{\alpha_n}$ tends to minimize $f^*(\mathbb{P}\circ Y^{-1} - \nu)$, by the fact that $\nu_l = \lim_{\alpha\to 0} \mathbb{P}_\alpha \circ Y^{-1}$, which concludes the proof. □

5.6.3 Discretization error

Proof of Proposition 5.5.1 We have that (φ, ψ, h) is a dual optimizer for (μ, ν). Then $\varphi \oplus \psi + h^{\otimes} \geq c$, and if $\mathbb{P}' \in \mathcal{M}(\mu', \nu')$, then $\mathbb{P}'[c] \leq \mu'[\varphi] + \nu'[\psi] \leq \mu[\varphi] + \nu[\psi] +$

$L_\varphi W_1(\mu',\mu)+L_\psi W_1(\nu',\nu)$. If we take the supremum in $\mathbb{P}'$, we get that

$$\begin{aligned}\mathbf{S}_{\mu',\nu'}(c) &\leq \mu[\varphi]+\nu[\psi]+L_\varphi W_1(\mu',\mu)+L_\psi W_1(\nu',\nu)\\ &= \mathbf{I}_{\mu,\nu}(c)+L_\varphi W_1(\mu',\mu)+L_\psi W_1(\nu',\nu)\\ &= \mathbf{S}_{\mu,\nu}(c)+L_\varphi W_1(\mu',\mu)+L_\psi W_1(\nu',\nu).\end{aligned}$$

As the reasoning may be symmetrical in $\Big((\mu,\nu),(\mu',\nu')\Big)$, we get the result. □

Proof of Proposition 5.5.3 Similar to the proof of Proposition 5.5.1, we have that

$$\mathbf{S}_{\mu'_N,\nu'_M}(c)-\mathbf{S}_{\mu,\nu}(c)\leq(\mu'_N-\mu)[\varphi]+(\nu'_M-\nu)[\psi],$$

and

$$\mathbf{S}_{\mu,\nu}(c)-\mathbf{S}_{\mu'_N,\nu'_M}(c)\leq(\mu-\mu'_N)[\varphi_N]+(\nu-\nu'_M)[\psi_M].$$

The first inequality gives

$$\mathbf{S}_{\mu'_N,\nu'_M}(c)-\mathbf{S}_{\mu,\nu}(c)\leq(\mu_N-\mu)[\varphi]+(\nu_M-\nu)[\psi]+(\mu'_N-\mu_N)[\varphi]+(\nu'_M-\nu_M)[\psi].$$

The two first terms are independent and their sum $(\mu_N-\mu)[\varphi]+(\nu_M-\nu)[\psi]$ is equivalent in law to $\sqrt{\frac{\mathrm{Var}_\mu[\varphi]}{N}+\frac{\mathrm{Var}_\nu[\psi]}{M}}\mathcal{N}(0,1)$ when N,M go to infinity. Then doing the same work on the symmetric inequality and using the Assumptions (i) to (iv), we get the result. □

5.6.4 Entropy error

Lemma 5.6.1. *Let $r,a>0$, $F:\mathbb{R}^d\longrightarrow\mathbb{R}^d$ and $x_0\in\mathbb{R}^d$ such that $\|(\nabla F)^{-1}(x_0)\|\leq a^{-1}$, and on $B_r(x_0)$, we have that F is C^1, that ∇F is invertible, and that $\|\nabla F-\nabla F(x_0)\|<a/2$. Then F is a C^1-diffeomorphism on $B_r(x_0)$.*

Proof. We claim that F is injective on $B_r(x_0)$, we also have that ∇F is invertible on this set. Then by the global inversion theorem, F is a C^1-diffeomorphism on $B_r(x_0)$.

Now we prove the claim that F is injective on $B_r(x_0)$. Let $x,y\in B_r(x_0)$,

$$\begin{aligned}F(y)-F(x) &= \int_0^1\nabla F(tx+(1-t)y)(y-x)dt\\ &= \nabla F(x)(y-x)+\int_0^1[\nabla F(tx+(1-t)y)-\nabla F(x)](y-x)dt\\ &= \nabla F(x)\left(y-x+\nabla F(x)^{-1}\int_0^1[\nabla F(tx+(1-t)y)-\nabla F(x)](y-x)dt\right).\end{aligned}$$

Then we assume that $F(y) = F(x)$, and we suppose for contradiction that $x \neq y$. Therefore by the fact that ∇F is invertible, we have

$$\begin{aligned} |y-x| &= \left|\nabla F(x)^{-1}\int_0^1 \left[\nabla F(tx+(1-t)y) - \nabla F(x_0) + \nabla F(x_0) - \nabla F(x)\right](y-x)dt\right| \\ &< \|\nabla F(x)^{-1}\| 2\frac{a}{2}|y-x| \\ &\leq |y-x|, \end{aligned}$$

Then we get the contradiction $|y-x| < |y-x|$. The injectivity is proved. □

In order to prove Lemma 5.6.3, we first need the following technical lemma.

Lemma 5.6.2. *Let an integer $d \geq 1$, $k \leq d$, $r > 0$, and $(y_i)_{1\leq i\leq k+1}$, $k+1$ differentiable maps $B_r \longrightarrow \mathbb{R}^d$ such that $|\nabla y_i| \leq A$, $|y_i| \leq A$ and $\det_{\text{aff}}(y_1,...,y_k) \geq A^{-1}$. Let $(u_i)_{k+1\leq i\leq d}$ an orthonormal basis of $\Big((y_i(0) - y_{k+1}(0))_{1\leq i\leq k}\Big)^\perp$, $(e_i)_{1\leq i\leq d}$ be the orthonormal basis obtained from $\Big((y_i - y_{k+1})_{1\leq i\leq d}, (u_i)_{k+1\leq i\leq d}\Big)$ from the Gram-Schmidt process, p the orthogonal projection of 0 on $\text{aff}(y_1,...,y_{k+1})$, and $(\lambda_i)_{1\leq i\leq k+1}$ the unique coefficients such that $p = \sum_{i=1}^{k+1}\lambda_i y_i$, barycentric combination. Then the maps e_i, λ_i, and p are differentiable on B_r and we may find $C, q > 0$, only depending on d such that if $r \leq C^{-1}A^{-q}$, then we have that $|\nabla e_i| \leq CA^q$, $|\nabla p| \leq CA^q$, $|\nabla \lambda_i| \leq CA^q$, and $\nabla \det_{\text{aff}}(y_1,...,y_{k+1}) \leq CA^q$.*

Proof. The determinant is a polynomial expression of the coefficients, therefore if these coefficients are bounded by A. Then we may find $C_{\det}, q_{\det} > 0$ (only depending on d) such that $|\nabla \det| \leq C_{\det}A^{q_{\det}}$.

Let $(v_i)_{1\leq i\leq d} := \Big((y_i - y_{k+1})_{1\leq i\leq d}, (u_i)_{k+1\leq i\leq d}\Big)$. Notice that for all i, we have $|\nabla v_i| \leq 2A$, and by the fact that

$$\begin{aligned} \det_{\text{aff}}\Big(y_1(0),...,y_{k+1}(0)\Big) &:= \Big|\det\Big(y_1(0) - y_{k+1}(0),...,y_k(0) - y_{k+1}(0), u_{k+1},...,u_d\Big)\Big| \\ &\geq A^{-1}, \end{aligned}$$

we have that $|\det(v_1,...,v_d)| \geq \frac{1}{2}A^{-1}$ on B_r for $r \leq C_{\det}^{-1}A^{-q_{\det}}\frac{1}{2}A^{-1}$.

Recall that $e_1 := \frac{v_1}{|v_1|}$. By the fact that $|v_1|...|v_d| \geq |\det(v_1,...,v_d)|$, we have that $|v_1| \geq A^{-d}$. Then we may find $C_1, q_1 > 0$ such that $\nabla e_1 \leq C_1A^{q_1}$. Now for $1 \leq i \leq k$, we have that $e_i := \frac{v_i - \sum_{j<i}(e_j\cdot v_i)e_j}{\left|v_i - \sum_{j<i}(e_j\cdot v_i)e_j\right|}$. Notice that

$$\begin{aligned}\left|\det\left(v_1,...,v_{i-1},v_i-\sum_{j<i}(e_j\cdot v_i)e_j,v_{i+1},...,v_d\right)\right| &= |\det(v_1,...,v_{i-1},v_i,v_{i+1},...,v_d)| \\ &\geq A^{-1},\end{aligned}$$

and therefore we have that $\left|v_i - \sum_{j<i}(e_j\cdot v_i)e_j\right| \geq A^{-d}$. Therefore, by induction, we may find C_i, q_i such that $|\nabla e_i| \leq C_i A^{q_i}$.

Now notice that $p := y_{k+1} + \sum_{i=1}^k e_i\cdot(0-y_{k+1})e_i$. Then we may find $C_0, q_0 > 0$ such that $|\nabla p| \leq C_0 A^{q_0}$.

Finally let $\lambda := (\lambda_1,...,\lambda_{k+1})$, $M' := \Big[e_i\cdot(y_j - y_{k+1})\Big]_{1\leq i\leq k, 1\leq j\leq k+1}$, $M := \left[\frac{1...1}{M'}\right]$, and $P := \left[\frac{1}{p-y_{k+1}}\right]$. We have that $M\lambda = P$, and therefore $\lambda = M^{-1}P$. Recall that $M^{-1} = \det(M)^{-1}Com(M)^t$ (see (5.1.2)), therefore we may find $C', q' > 0$ such that $|M^{-1}| \leq C'A^{q'}$, and $|\nabla(M^{-1})| \leq C'A^{q'}$. Then we may find $C'', q'' > 0$ such that $|\nabla\lambda_i| \leq C''A^{q''}$ for all i.

Finally, by the fact that

$$\det_{\text{aff}}(y_1,...,y_{k+1}) = |\det(y_1-y_{k+1},...,y_k-y_{k+1},e_{k+1},...,e_d)|,$$

we may find C''', q''' such that $\det_{\text{aff}}(y_1,...,y_{k+1}) \leq C'''A^{q'''}$.

The lemma is proved for

$$C := \max(C_0,...,C_d,C',C'',C''',2C_{\det}),\ \text{and for}\ q := \max(q_0,...,q_d,q',q'',q''',q_{\det}+1).$$

□

Lemma 5.6.3. *Let $A,r,e,\delta,h,H > 0$ and $F:\mathbb{R}^d \longrightarrow \mathbb{R}$ such that we may find $k\in\mathbb{N}$ and $S\in(B_A)^k$ such that for all $y\in S$, we have on $B_r(y)$ that F is C^2, $A^{-1}I_d \leq D^2F \leq AI_d$, and $|D^2F - D^2F(y)| \leq e$. Furthermore, $\det_{\text{aff}} S \geq A^{-1}$, $\nabla F(S) = \{0\}$, and $\left|\sum_{i=1}^k \lambda_i S_i\right| \leq H$, convex combination with $\min\lambda \geq A^{-1}$. Furthermore, assume that $F(S)\subset[-h,h]$, and $F \geq \delta\text{dist}(Y,S)$ on $B_r(S)^c$. Then we may find $C,q>0$ such that if $\delta, r \geq CA^qH$, $e,H \leq C^{-1}A^{-q}$, and $h\leq rH$, then we have that $\Big(0,F_{conv}(0)\Big) = \sum_{i=1}^k \overline{\lambda}_i\Big(\bar{y}_i, F(\bar{y}_i)\Big)$, convex combination with $|\bar{y}_i - S_i| \leq CA^qH$ for all i.*

Proof.

Step 1: For all i, the map $y \longmapsto \nabla F(y)$ is a C^1-diffeomorphism on $B_r(y_i)$ by Lemma 5.6.1. Then we define the map $z_i(a) := \nabla F^{-1}(a)$ which is defined on $B_{rA^{-1}}$. Notice that its gradient is given by $\nabla z_i(a) := D^2 F^{-1}(a)$. Now we define the map $\Phi : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ as follows: $\Phi_i(a) := a \cdot \Big(z_i(a) - z_{k+1}(a)\Big) - \Big(F\big(z_i(a)\big) - F\big(z_{k+1}(a)\big)\Big)$, for $1 \le i \le k$, and $\Phi_i(a) := e_i(a) \cdot p(a)$ for $k+1 \le i \le d$, where $p(a)$ is the orthogonal projection of 0 on $\mathrm{aff}\big(z_1(a), ..., z_k(a)\big)$, and $\big(e_{k+1}(a), ..., e_d(a)\big)$ is the orthonormal basis of $\Big(\mathrm{aff}\big(z_1(a), ..., z_k(a)\big)\Big)^\perp$ defined as the Gram-Schmidt basis obtained from $(z_1(a) - z_{k+1}(a), ..., z_k(a) - z_{k+1}(a), u_{k+1}, ..., u_d)$, where $(u_{k+1}, ..., u_d)$ is a fixed basis of $\Big(z_1(0) - z_{k+1}(0), ..., z_k(0) - z_{k+1}(0)\Big)^\perp$.

Step 2: Now we prove that the convex hull $\Big(F\Big)_{conv}(0)$ is determined by the equation $\Phi(a) = 0$ for a small enough. Let $|a| \le rA^{-1}$ such that $\Phi(a) = 0$. Then $a\Big(z_i(a) - z_{k+1}(a)\Big) - \Big(F\big(z_1(a)\big) - F\big(z_{k+1}(a)\big)\Big) = 0$, and therefore let $b := F\big(z_1(a)\big) - az_1(a) = ... = F\big(z_k(a)\big) - az_k(a)$. Then the map $y \mapsto ay + b$ is tangent to F at all z_i. Furthermore, $p(a)$ is orthogonal to $\Big(\mathrm{aff}\big(z_1(a), ..., z_k(a)\big)\Big)^\perp$, implying that $p(a) = 0$. Then $0 \in \mathrm{aff}\big(z_1(a), ..., z_k(a)\big)$. By Lemma 5.6.2, we may find $C_1, q_1 > 0$ (only depending on d) such that $|\nabla \det_{\mathrm{aff}}(z_1, ..., z_k)| \le C_1 A^{q_1}$. Therefore, if $a \le \frac{1}{2} A^{-2} C_1^{-1} A^{-q_1}$, we have that $|\det_{\mathrm{aff}}(z_1, ..., z_k)| \ge \frac{1}{2} A^{-1}$ and we may find $(\overline{\lambda}_i)_{1 \le i \le k+1}$ such that $p(a) = \sum_{i=1}^{k+1} \overline{\lambda}_i z_i(a)$, and $\overline{\lambda}_i \ge \frac{1}{2} A^{-1}$ by Lemma 5.6.2 together with the fact that $\min \lambda \ge A^{-1}$. Now we prove that $F \ge aY + b$. This holds on each $B_r\big(z_i(a)\big)$ by convexity of F on these balls, together with the fact that $aY + b$ is tangent to F. Now out of these balls, $F \ge \delta \mathrm{dist}(Y, S)$ by assumption. Furthermore, $|z_i(a) - z_i(0)| \le A|a|$, and $\Big|\nabla F\big(z_i(a)\big)\Big| \le A^2|a|$, while similar, we have $\Big|F\big(z_i(a)\big)\Big| \le h + \frac{1}{2} A^3 |a|^2$. Notice that as it is tangent, $aY + b = \nabla F\big(z_i(a)\big)\big(Y - z_i(a)\big) + F\big(z_i(a)\big) = \nabla F\big(z_i(a)\big)\big(Y - S_i\big) + \nabla F\big(z_i(a)\big)\big(S_i - z_i(a)\big) + F\big(z_i(a)\big)$, for all i. Then,

$$
\begin{aligned}
|aY + b| &\le \Big|\nabla F\big(z_i(a)\big)\Big|\,|Y - S_i| + \Big|\nabla F\big(z_i(a)\big)\big(S_i - z_i(a)\big) + F\big(z_i(a)\big)\Big| \\
&\le A^2|a|\,|Y - S_i| + \frac{3}{2} A^3 |a|^2 + h
\end{aligned}
$$

Therefore, if $\delta \ge A^2|a| + \Big(\frac{3}{2} A^3|a|^2 + h\Big) r^{-1}$, then $F \ge \delta \mathrm{dist}(Y, S) \ge aY + b$. This holds in particular if $r \ge h/H$, implying that $A^2|a| + \Big(\frac{3}{2} A^3 |a|^2 + h\Big) r^{-1} \le A^2|a| + \frac{3}{2} A^3 |a|^2 r^{-1} + hH/h \le A^2|a| + \frac{3}{2} A^3|a| + H$. Finally, the following domination is sufficient:

$$
\delta \ge A^2|a| + \frac{3}{2} A^3|a| + H. \tag{5.6.12}
$$

Step 3: Now we prove that Φ may be locally inverted. If $1 \leq i \leq k$, we have $\nabla\Phi_i(a) = z_i(a) - z_{k+1}(a)$. If $k+1 \leq i \leq d$, we have $\nabla\Phi_i(a) = \nabla p(a)e_i(a) + \nabla e_i(a)p(a)$. We may rewrite the previous expression by introducing the locally smooth maps $\lambda_j(a)$ such that $p(a) =: \sum_{j=1}^k \lambda_j(a) z_j(a)$, convex combination. Then $\nabla p(a) = \sum_{j=1}^k \lambda_j(a)\nabla z_j(a) + \sum_{j=1}^k \nabla\lambda_j(a) z_j(a)$. Notice that by the relationship $\sum_{j=1}^k \lambda_j(a) = 1$, we have that $\sum_{j=1}^k \nabla\lambda_j(a) = 0$, therefore $\sum_{j=1}^k \nabla\lambda_j(a) z_j(a) = \sum_{j=1}^k \nabla\lambda_j(a)\Big(z_j(a) - z_{k+1}(a)\Big)$ and $\sum_{j=1}^k \nabla\lambda_j(a)\Big(z_j(a) - z_{k+1}(a)\Big)e_i = 0$ as $z_j - z_{k+1} \perp e_i$. Finally, we have $\nabla p(a)e_i(a) = \sum_{j=1}^k \lambda_j(a) D^2F^{-1}\Big(z_j(a)\Big)e_i(a)$.

Step 4: Now we provide a bound for $\nabla e_i(a)p(a)$. We have the control $|p(a)| \leq |p(0)| + \sup_{B_a}|\nabla p| r \leq H + C_2A^{q_2}a$ for $a \in B_r$ by Lemma 5.6.2 for some $C_2, q_2 > 0$. Therefore by Lemma 5.6.2, we may find $C_3, q_3 > 0$ so that if $H \leq C_3^{-1}A^{-q_3}$, we have the control $|\nabla e_i(a)| \leq C_3A^{q_3}$, whence the inequality

$$|\nabla e_i(a)p(a)| \leq C_3A^{q_3}(H + C_2A^{q_2}a). \tag{5.6.13}$$

Step 5: Now we provide a lower bound to $\det\nabla\Phi$. Notice that $\nabla\Phi = P_0 + P'$ with $P_0 := (z_i - z_{k+1} : i \leq k, De_i : i \geq k+1)$, and $P' := (\nabla e_i p : i \geq k+1)$, where $D := \sum_{j=1}^k \lambda_j D^2F^{-1}(z_j)$. Let $M_{basis} := Mat(z_i - z_{k+1} : i \leq k, e_i : i \geq k+1)$, then $P_0M_{basis}^{-1}$ may be written as a block matrix as follows: $P_0M_{basis}^{-1} = \left[\begin{array}{c|c} I_k & \cdot \\ \hline 0 & D_{basis}\end{array}\right]$, with $D_{basis} := (e_i^tDe_j)_{k+1\leq i,j\leq d}$. Then $\det(P_0M_{basis}^{-1}) = \det(D_{basis}) \geq A^{-(d-k)}$, $\det(M_{basis}) = \det(z_i - z_{k+1} : i \leq k) \geq A^{-1}$, and therefore $\det P_0 \geq A^{-(d-k+1)}$. Then by Lemma 5.6.2, as k lines of P_0 are dominated by $2A$ and $d-k$ are dominated by A, we have $\det\nabla\Phi(a) \geq C_4^{-1}A^{-q_4}$ if $a, H \leq C_4^{-1}A^{-q_4}$, and $a \leq r$ for some $C_4, q_4 > 0$.

Step 6: Finally $|\Phi| \leq A + A = 2A$. In order to apply Lemma 5.6.1, we need to control $|\nabla\Phi(a) - \nabla\Phi(a')|$ for $a, a' \in B_r$. For $i \leq k$, $|\nabla\Phi_i(a) - \nabla\Phi_i(a')| = |D^2F\Big(z_i(a)\Big)^{-1} - D^2F\Big(z_{k+1}(a')\Big)^{-1}| \leq 2A^2e$. For $i \geq k+1$,

$$\begin{aligned}|\nabla\Phi_i(a) - \nabla\Phi_i(a')| &= |\nabla\Big(p(a)\cdot e_i(a)\Big) - \nabla\Big(p(a')\cdot e_i(a')\Big)| \\ &\leq |\nabla p(a)e_i(a) - \nabla p(a')e_i(a')| + |\nabla e_i(a)p(a) - \nabla e_i(a')p(a')| \\ &\leq |\nabla p(a)e_i(a) - \nabla p(a')e_i(a')| + 2\Big(H + C_2A^{q_2}(|a| + |a'|)\Big)C_1A^{q_1}.\end{aligned}$$

We consider the first term:

$$\begin{aligned}
|\nabla p(a)e_i(a)-\nabla p(a')e_i(a')| &\le \left|\sum_{j=1}^k \lambda_j(a')\Big(D^2F^{-1}(z_j(a))-D^2F^{-1}(z_j(a'))\Big)\right| \\
&\quad +\left|\sum_{j=1}^k \Big(\lambda_j(a)-\lambda_j(a')\Big)D^2F^{-1}\Big(z_j(a)\Big)\right| \\
&\le A^2e+C_1A^{q_1}|a-a'|A,
\end{aligned}$$

and therefore we may find $C_5,q_5>0$ such that if $|a|,e\le C_5^{-1}A^{-q_5}$, then we have that $|\nabla\Phi(a)-\nabla\Phi(a')|\le\frac12|det\nabla\Phi(0)|^{-1}\|Com\Big(\nabla\Phi(0)\Big)^t\|=\|\nabla\Phi(0)\|$. Then we may apply Lemma 5.6.1: Φ is a C^1-diffeomorphism on B_r, we may find $C_6,q_6>0$ such that $C_6^{-1}A^{-q_6}\le|\nabla\Phi|\le C_6A^{q_6}$. By assumption, we have $|\Phi(0)|\le dH$. Furthermore, $B_{rC_6^{-1}A^{-q_6}}\Big(\Phi(0)\Big)\subset\Phi(B_r)$. Therefore, if $H\le rC_6^{-1}d^{-1}A^{-q_6}$, then we may find $a_0\in B_r$ such that $\Phi(a_0)=0$. We have

$$|a_0|=|\Phi^{-1}(0)-\Phi^{-1}(\Phi(0))|\le C_6A^{q_6}|\Phi(0)|\le C_6dA^{q_6}H.$$

By Step 2, $z_1(a_0),...,z_k(a_0)$ have the required property. Moreover,

$$|z_i(a_0)-S_i|=|z_i(a_0)-z_i(0)|\le C_6dA^{q_6+1}H.$$

Finally, (5.6.12) is satisfied if $\delta\ge C_7A^{q_7}H$, with $C_7:=C_6+\frac52$, and $q_7:=\max(3,q_6)$. The lemma is proved for $C:=\max(3,C_1,...,C_7,C_6d)$ and $q:=\max(3,q_1,...,q_7,q_6+1)$. □

Proof of Theorem 5.5.5

Step 1: We claim that we may find $C_1>0$ such that for ε small enough, we have $\Delta_\varepsilon\ge -C_1\varepsilon\ln(\varepsilon^{-1})$, $m_\varepsilon-$a.e. Indeed, by the fact that $(\varphi_\varepsilon,\psi_\varepsilon,h_\varepsilon)$ is an optimum, we have that $e^{-\frac{\Delta_\varepsilon}{\varepsilon}}m_\varepsilon$ is a probability distribution. Therefore, $e^{-\frac{\Delta_\varepsilon(X,\cdot)}{\varepsilon}}(m_\varepsilon)_X/\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal X}}$ is a probability measure, $\mu_\varepsilon-$a.s. Then by (i), $m_\varepsilon-$a.s., we have that

$$\begin{aligned}
1\ge\int_{B_{\alpha_\varepsilon}(Y)}e^{-\frac{\Delta_\varepsilon(X,y)}{\varepsilon}}(m_\varepsilon)_X(dy)/\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal X}} &\ge \varepsilon^\gamma\int_{B_{\alpha_\varepsilon}(Y)}e^{-\frac{\Delta_\varepsilon(X,y)-\gamma\varepsilon\ln(\varepsilon^{-1})}{\varepsilon}}(m_\varepsilon)_X(dy) \\
&\ge \varepsilon^{3\gamma}e^{-\frac{\Delta_\varepsilon(X,y)}{\varepsilon}}.
\end{aligned}$$

Hence $\Delta_\varepsilon(X,Y)\ge-3\gamma\varepsilon\ln(\varepsilon^{-1})$, $m_\varepsilon-$a.s. The claim is proved for ε small enough.

Step 2: We claim that we may find $C_2 > 0$ such that $\int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \ll \varepsilon$, and $\int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \ll 1$. Indeed, let $C_2 > 0$. By (i), we have

$$\int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \leq m_\varepsilon[\Omega] C_2\varepsilon\ln(\varepsilon^{-1})\varepsilon^{C_2} \leq C_2\varepsilon\ln(\varepsilon^{-1})\varepsilon^{C_2+\gamma}.$$

Similar, we have that $\int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \leq C_2\varepsilon^{C_2+\gamma}$. Therefore, up to choosing C_2 large enough, the claim holds.

Step 3: Let $\bar{\Delta}_\varepsilon := \Delta_\varepsilon - \Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X) - \nabla\Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X)\cdot(Y-X)$. We claim that $\int_{x\notin D_\varepsilon^{\mathcal{X}}} \bar{\Delta}_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \ll \varepsilon$. Indeed by (iii) we have that $\mu_\varepsilon[D_\varepsilon^{\mathcal{X}}] \ll 1/\ln(\varepsilon^{-1})$, therefore we have

$$\begin{aligned}
\int_{x\notin D_\varepsilon^{\mathcal{X}}} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon &\leq \int_{x\notin D_\varepsilon^{\mathcal{X}}} C_2\varepsilon\ln(\varepsilon^{-1}) e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon + \int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \\
&\leq C_2\varepsilon\ln(\varepsilon^{-1})\mu_\varepsilon[(D_\varepsilon^{\mathcal{X}})^c] + \int_{\Delta_\varepsilon \geq C_2\varepsilon\ln(\varepsilon^{-1})} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \\
&\ll \varepsilon.
\end{aligned}$$

Finally, by the martingale property of $e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon$, we have

$$\begin{aligned}
\int_{x\notin D_\varepsilon^{\mathcal{X}}} \bar{\Delta}_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon &= \int_{x\notin D_\varepsilon^{\mathcal{X}}} \Big(\Delta_\varepsilon - \Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X) \\
&\quad -\nabla\Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X)\cdot(Y-X)\Big) e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \\
&= \int_{x\notin D_\varepsilon^{\mathcal{X}}} \Big(\Delta_\varepsilon - \Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X)\Big) e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon \\
&\leq \int_{x\notin D_\varepsilon^{\mathcal{X}}} \Delta_\varepsilon e^{-\frac{\Delta_\varepsilon}{\varepsilon}} dm_\varepsilon + \mu_\varepsilon[D_\varepsilon^{\mathcal{X}}] C_1\varepsilon\ln(\varepsilon^{-1}) \\
&\ll \varepsilon.
\end{aligned}$$

Step 4: Let $x \notin D_\varepsilon^{\mathcal{X}}$, we denote $S_x^\varepsilon = \{s_1, ..., s_k\}$, where $k := k_x^\varepsilon$. We claim that for all $S' := (s_1', ..., s_k') \in \mathbb{R}^d$ such that $s_i' \in B_{r_\varepsilon}(s_i)$ for all i, and $\sum \lambda_i' s_i' = x' \in B_{r_\varepsilon}(x)$ we have for $\varepsilon > 0$ small enough that $S' \in (B_{2A_\varepsilon})^k$, $|\det_{\text{aff}} S'| \geq \frac{1}{2}A_\varepsilon^{-1}$, $\min \lambda_x^\varepsilon \geq \frac{1}{2}A_\varepsilon^{-1}$.

Indeed, $\sum \lambda_i'(s_i' + x - x') = x$, with $s_i' + x - x' \in B_{2r_\varepsilon}(s_i)$. By Lemma 5.6.2, we may find C_3, q_3 such that $|\lambda_i' - \lambda_i| \leq C_3 A_\varepsilon^{q_3} r_\varepsilon$, $|\det_{\text{aff}} S' - \det_{\text{aff}} S| \leq C_3 A_\varepsilon^{q_3} r_\varepsilon$, and $|s_i'| \leq A_\varepsilon + r_\varepsilon$. Now by the fact that $r_\varepsilon \ll A_\varepsilon^{-q_3}$, the claim is proved.

Step 5: We claim that up to shrinking $D_\varepsilon^{\mathcal{X}}$, we may assume that for $x \notin D_\varepsilon^{\mathcal{X}}$, we have $\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}(x) \geq \varepsilon^{\gamma+1}$. Indeed, $\mu_\varepsilon^{\mathcal{X}}\left[\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}} \leq \varepsilon^{\gamma+1}\right] \leq \varepsilon^{\gamma+1} m_\varepsilon^{\mathcal{X}}[\mathbb{R}^d]$. Therefore,

$$\mu_\varepsilon^{\mathcal{X}}\left[\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}} \leq \varepsilon^{\gamma+1}\right] \leq \varepsilon^{\gamma+1} m_\varepsilon^{\mathcal{X}}[\mathbb{R}^d] = \varepsilon^{\gamma+1} m_\varepsilon[\Omega] \leq \varepsilon \ll 1/\ln(\varepsilon^{-1}).$$

Therefore we may shrink $D_\varepsilon^{\mathcal{X}}$ by removing $\left\{\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}} > \varepsilon^{\gamma+1}\right\}$ from it.

Step 6: We claim that for $\varepsilon > 0$ small enough, we may find unique $y_i \in B_{r_\varepsilon}(s_i)$ such that $\nabla\Delta(x, y_i) = 0$ for all i, $B_{\overline{r}_\varepsilon}(y_i) \subset B_{r_\varepsilon}(s_i)$, with $\overline{r}_\varepsilon := \varepsilon^{\frac{1}{2}-\overline{\eta}}$, where $0 < \overline{\eta} < \frac{\eta}{2}$, and finally $\Delta_\varepsilon(x, y) \geq \frac{1}{2}\sqrt{\varepsilon}\,\mathrm{dist}\big(y, (y_1, ..., y_k)\big)$ for $y \notin \cup_{i=1}^k B_{\overline{r}_\varepsilon}(y_i)$. Indeed $\Delta_\varepsilon(x, \cdot)$ is strictly convex on $B_{r_\varepsilon}(s_i)$ for all i. Furthermore, let

$$m_i := \frac{1}{\bar\lambda_i}\int_{B_{r_\varepsilon}(s_i)} y e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} dm_\varepsilon^{\mathcal{Y}}\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1},$$

with $\bar\lambda_i := \int_{B_{r_\varepsilon}(s_i)} e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} dm_\varepsilon^{\mathcal{Y}}\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}$. By the martingale property of $e^{-\frac{\Delta_\varepsilon}{\varepsilon}} m_\varepsilon$, we have

$$\sum_i \bar\lambda_i m_i = x - \int_{(\cup_i B_{r_\varepsilon}(s_i))^c} y e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} dm_\varepsilon^{\mathcal{Y}}\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1},$$

therefore by Step 5,

$$\begin{aligned}
\left|\sum_{i=1}^k \bar\lambda_i m_i - x\right| &= \left|\int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c} y e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} m_\varepsilon^{\mathcal{Y}}(dy)\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}\right| \\
&\leq \int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c} |y| e^{-\varepsilon^{-\frac{1}{2}}\mathrm{dist}(y, S_x^\varepsilon)} m_\varepsilon^{\mathcal{Y}}(dy)\varepsilon^{-\gamma-1} \\
&\leq \int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c, |y|\geq 2A_\varepsilon} |y| e^{-\varepsilon^{-\frac{1}{2}}\mathrm{dist}(y, S_x^\varepsilon)} m_\varepsilon^{\mathcal{Y}}(dy)\varepsilon^{-\gamma-1} \\
&\quad + \int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c, |y|< 2A_\varepsilon} 2A_\varepsilon e^{-\varepsilon^{-\frac{1}{2}}\mathrm{dist}(y, S_x^\varepsilon)} m_\varepsilon^{\mathcal{Y}}(dy)\varepsilon^{-\gamma-1}.
\end{aligned}$$

Observe that as $y_i \leq A_\varepsilon$, we have $\mathrm{dist}(y, S_x^\varepsilon) \geq |y| - A_\varepsilon$. Furthermore, if $A_\varepsilon \geq 2$, and $|y| \geq 2A_\varepsilon$, we have $|y| \leq e^{|y|-A_\varepsilon}$. Then we have

$$\begin{aligned}\left|\sum_{i=1}^k \bar{\lambda}_i m_i - x\right| &\leq \int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c, |y|\geq 2A_\varepsilon} e^{-\left(\varepsilon^{-\frac{1}{2}}-1\right)(|y|-A_\varepsilon)} m_\varepsilon^{\mathcal{Y}}(dy)\varepsilon^{-\gamma-1} \\ &\quad + \int_{\left(\cup_{i=1}^k B_{r_\varepsilon}(s_i)\right)^c, |y|< 2A_\varepsilon} 2A_\varepsilon e^{-\varepsilon^{-\frac{1}{2}} r_\varepsilon} m_\varepsilon^{\mathcal{Y}}(dy)\varepsilon^{-\gamma-1} \\ &\leq e^{-\left(\varepsilon^{-\frac{1}{2}}-1\right)A_\varepsilon} m_\varepsilon^{\mathcal{Y}}[\mathbb{R}^d]\varepsilon^{-\gamma-1} + 2A_\varepsilon e^{-\varepsilon^{-\eta}} m_\varepsilon^{\mathcal{Y}}[\mathbb{R}^d]\varepsilon^{-\gamma-1} \\ &\ll \varepsilon. \end{aligned} \tag{5.6.14}$$

Similar, we have $\sum_i \bar{\lambda}_i = 1 + o(\varepsilon)$, with uniform convergence of $o(\varepsilon)$ in x. By Step 4, we have that $\bar{\lambda}_i \geq \frac{1}{2}A_\varepsilon^{-1}$ for $\varepsilon > 0$ small enough, as $\varepsilon \ll r_\varepsilon$. Therefore, we may find $y \in B_{r_\varepsilon}(s_i)$ such that $\Delta_\varepsilon(x,y) < \varepsilon^{1-\frac{\eta}{2}}$, as otherwise, similar to (5.6.14), we would have $\bar{\lambda}_i \ll \varepsilon$. Notice that for y in the boundary of $B_{r_\varepsilon}(s_i)$, by (v), we have $\Delta_\varepsilon(x,y) \geq \varepsilon^{\frac{1}{2}} r_\varepsilon = \varepsilon^{1-\eta} > \varepsilon^{1-\frac{\eta}{2}}$, then as $\Delta_\varepsilon(x,\cdot)$ is strictly convex on $B_{r_\varepsilon}(s_i)$, we may find a unique minimizer $y_i \in B_{r_\varepsilon}(s_i)$. Now let $l := \mathrm{dist}\big(y_i, \partial B_{r_\varepsilon}(s_i)\big)$, we have that $\Delta_\varepsilon(x,y_i) + \frac{1}{2}A_\varepsilon l^2 \geq \varepsilon^{1-\eta}$. Then by the inequality $\Delta_\varepsilon(x,y_i) \leq \varepsilon^{1-\frac{\eta}{2}}$, we have that $l \geq \sqrt{2A_\varepsilon^{-1}}\varepsilon^{\frac{1}{2}-\frac{\eta}{2}}\sqrt{1-\varepsilon^{\frac{\eta}{2}}}$. Now, let $0 < \overline{\eta} < \frac{\eta}{2}$, we have $l \leq \varepsilon^{\frac{1}{2}-\overline{\eta}}$ for $\varepsilon > 0$ small enough. Finally, let $y \notin \cup_{i=1}^k B_{\bar{r}_\varepsilon}(y_i)$, we treat two cases. Case 1: $y \in B_{r_\varepsilon}(y_i) \setminus B_{\bar{r}_\varepsilon}(y_i)$ for some i. Then by (iv) we have $\Delta_\varepsilon(x,y) \geq \Delta_\varepsilon(x,y_i) + \frac{1}{2}A_\varepsilon^{-1}\bar{r}_\varepsilon^2 \geq -C_1\varepsilon\ln(\varepsilon^{-1}) + \sqrt{\varepsilon}\bar{r}_\varepsilon \geq \frac{1}{2}\sqrt{\varepsilon}|y-y_i| \geq \frac{1}{2}\sqrt{\varepsilon}\,\mathrm{dist}\big(y,(y_1,...,y_k)\big)$, for $\varepsilon > 0$ small enough.

Case 2: $y \notin B_{r_\varepsilon}(y_i)$ for all i. Then let $1 \leq i \leq k$, we have $|y - s_i| \leq r_\varepsilon$, and recall that $|y_i - s_i| \leq r_\varepsilon$. Then $|y-y_i| \leq |y-s_i| + r_\varepsilon \leq 2|y-y_i|$, and therefore $\mathrm{dist}\big(y,(y_1,...,y_k)\big) \leq 2\mathrm{dist}(y,S_x^\varepsilon)$. By (v) we have $\Delta_\varepsilon(x,y) \geq \sqrt{\varepsilon}\,\mathrm{dist}(y,S_x^\varepsilon) \geq \frac{1}{2}\sqrt{\varepsilon}\,\mathrm{dist}\big(y,(y_1,...,y_k)\big)$, for $\varepsilon > 0$ small enough.

The claim is proved. Now, up to changing η to $\overline{\eta}$, the properties (i) to (vi) are still satisfied, and the properties of (iv) and (v) also hold if we replace S_x^ε by $(y_1,...,y_k)$.

Step 7: Let $D_\varepsilon^{\mathcal{X}\to\mathcal{Y}} := \left\{x \in D_\varepsilon^{\mathcal{X}} : B_{r_\varepsilon}(y) \setminus D_\varepsilon^{\mathcal{Y}} = \emptyset, \text{ for some } y \in S_x^\varepsilon\right\}$. We claim that we have $\mu_\varepsilon\left[D_\varepsilon^{\mathcal{X}\to\mathcal{Y}}\right] \ll 1/\ln(\varepsilon^{-1})$. Indeed, for $x \in D_\varepsilon^{\mathcal{X}}$, and for all $y \in S_x^\varepsilon$, we have $(\mathbb{P}_\varepsilon)_x[B_{r_\varepsilon}(y)] \geq A_\varepsilon^{-1}$ by Step 6. Therefore, if for some such y, we have that $B_{r_\varepsilon}(y) \subset (D_\varepsilon^{\mathcal{Y}})^c$, then $A_\varepsilon^{-1} \leq (\mathbb{P}_\varepsilon)_x[B_{r_\varepsilon}(y)] \leq (\mathbb{P}_\varepsilon)_x[(D_\varepsilon^{\mathcal{Y}})^c]$. Then, if we integrate along μ_ε on $D_\varepsilon^{\mathcal{X}\to\mathcal{Y}}$, together with (vi) we get that

$$A_\varepsilon^{-1}\mu_\varepsilon[D_\varepsilon^{\mathcal{X}\to\mathcal{Y}}] \leq \mathbb{P}_\varepsilon[Y \notin D_\varepsilon^{\mathcal{Y}}] = \nu_\varepsilon[(D_\varepsilon^{\mathcal{Y}})^c] \ll A_\varepsilon^{-1}/\ln(\varepsilon^{-1}).$$

The claim is proved. Now up to shrinking $D_\varepsilon^{\mathcal{X}}$, we may assume that $D_\varepsilon^{\mathcal{X}} \cap D_\varepsilon^{\mathcal{X}\to\mathcal{Y}} = \emptyset$.

Step 8: We claim that up to raising γ, if $\frac{\left|f_{|B_{2R}\setminus B_R}\right|}{\|f\|_\infty^{2R}} \le \frac{1}{4}\varepsilon^\beta$ and $R \ge r_\varepsilon/\sqrt{\varepsilon}$, then (vi) may be applied to y_i (even if $y_i \notin D_\varepsilon^{\mathcal{Y}}$), up to replacing $(m_\varepsilon)_x[B_{R\sqrt{\varepsilon}}(y_i)]$ by $(m_\varepsilon)_x[B_{2R\sqrt{\varepsilon}}(y_i')]/2^d$ for some $y_i' \in B_{r_\varepsilon}(y_i)$. Indeed, by Step 7 we may find $y_i' \in B_{r_\varepsilon}(y_i) \cap D_\varepsilon^{\mathcal{Y}}$. Let f have such property. Now let $\widetilde{f}$ defined by $f = \widetilde{f}$ on $B_{R/2}$, and $\widetilde{f}(y) := \left(1 - \frac{2|y|}{R}\right) f\left(\frac{Ry}{2|y|}\right)$. Let $L, R \ge 1$ such that f is $L-$Lipschitz, then $\widetilde{f}$ is $L-$Lipschitz. Therefore, We have $\left|\int_{B_{2R}} \widetilde{f}(y)\left[\frac{d(m_\varepsilon)_x \circ \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_i'}}{(m_\varepsilon)_x[B_{2R}(y_i')]} - \frac{dy}{|B_{2R}|}\right]\right| \le [2R+L]^\gamma \varepsilon^\beta \int_{B_{2R}} \widetilde{f}(y)\frac{dy}{|B_{2R}|}$ from (vi). Now, as $R \ge r_\varepsilon/\sqrt{\varepsilon}$, we have that $B_{R\sqrt{\varepsilon}}(y_i) \subset B_{2R\sqrt{\varepsilon}}(y_i')$. Then

$$\left|\int_{B_R} f(y)\left[\frac{d(m_\varepsilon)_x \circ \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_i}}{(m_\varepsilon)_x[B_{2R}(y_i')]} - \frac{dy}{|B_{2R}|}\right]\right| \le [2R+L]^\gamma \varepsilon^\beta \int_{B_R} f(y)\frac{dy}{|B_{2R}|} + |\widetilde{f} - f|_\infty.$$

As we may find $y^* \in B_{R/2}$ such that $f(y^*) = \|f\|_\infty^R$, we have $f(y) \ge \|f\|_\infty^R (1 - L|y - y^*|)$, and $\int_{B_R} f(y)\frac{dy}{|B_{2R}|} \ge |B_1|L^{-d}$. Therefore, as $|B_{2R}| = 2^d|B_R|$, we have

$$\begin{aligned}
\left|\int_{B_R} f(y)\left[\frac{d(m_\varepsilon)_x \circ \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_i}}{(m_\varepsilon)_x[B_{2R}(y_i')]/2^d} - \frac{dy}{|B_R|}\right]\right| &\le \left([2R+L]^\gamma \varepsilon^\beta + |B_1|^{-1}L^d\varepsilon^\beta\right)\int_{B_R} f(y)\frac{dy}{|B_R|} \\
&\le [R+L]^{2\gamma+2}\varepsilon^\beta \int_{B_R} f(y)\frac{dy}{|B_R|}. \qquad (5.6.15)
\end{aligned}$$

From now we replace γ by $\gamma' := 2\gamma + 2$.

Step 9: As a preparation for this step, we observe that (i) to (vi) are preserved if we replace η by any $0 \le \eta' \le \eta$. Then, up to lowering η, we may assume without loss of generality that $\eta < \beta/\gamma$. Therefore

$$\beta - \eta\gamma > 0. \qquad (5.6.16)$$

We claim that $\sum_{i=1}^{k_x} \int_{B_{r_\varepsilon}(y_i)} (\Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,y_i))(\mathbb{P}_\varepsilon)_x(dy) = \frac{d}{2}\varepsilon + o(\varepsilon)$, where the convergence of $o(\varepsilon)$ is uniform in x. Indeed, consider

$$\begin{aligned}
\lambda_i &:= \int_{B_{r_\varepsilon}(y_i)} \exp\left(-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}\right) m^{\mathcal{Y}}(dy)\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1} \qquad (5.6.17) \\
&= \frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}\int_{B_{\varepsilon^{-\eta}}} \exp\left(-\frac{\Delta_\varepsilon\left(x, \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_i}(y)\right)}{\varepsilon}\right) m^{\mathcal{Y}} \circ \mathrm{zoom}_{\sqrt{\varepsilon}}^{y_i}(dy).
\end{aligned}$$

We want to compare λ_i to

$$\lambda_i' := \frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}\int_{B_{\varepsilon^{-\eta}}} \exp\left(-\frac{\Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)}{\varepsilon}\right) dy \frac{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d}{|B_{\varepsilon^{-\eta}}|}.$$

Notice that the map $F : y \longrightarrow \frac{\Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)}{\varepsilon}$ may be differentiated in

$$\nabla F = \frac{\partial_y \Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)}{\sqrt{\varepsilon}},$$

which is bounded by $A_\varepsilon \varepsilon^{-\eta}$, by the fact that $\nabla F(0) = 0$ and $D^2 F \leq A_\varepsilon I_d$ by (iv). Then, we observe that the map $f : y \longrightarrow \exp\left(-\frac{\Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)}{\varepsilon}\right)$ satisfies that $\frac{f}{\left\|f\right\|^{\varepsilon^{-\eta}}_\infty}$ is $A_\varepsilon \varepsilon^{-\eta}-$Lipschitz. We may apply (5.6.15) to f, with $R = r_\varepsilon/\sqrt{\varepsilon}$, as for $y \in B_{2R}\setminus B_R$, we have

$$|F(y)| \leq e^{-\frac{\Delta_\varepsilon(x,y_i)}{\varepsilon} - A_\varepsilon^{-1}|y|^2} \leq |F|^{2R}_\infty e^{-A_\varepsilon \varepsilon^{-2\eta}} \leq \frac{1}{4}\varepsilon^\beta,$$

for $\varepsilon > 0$ small enough, and get that

$$\begin{aligned} |\lambda_i - \lambda_i'| &\leq \left|\int_{B_{\varepsilon^{-\eta}}} f(y)\left[\frac{(m_\varepsilon)_x \circ \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(dy)}{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d} - \frac{dy}{|B_{\varepsilon^{-\eta}}|}\right]\right|(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d \frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1} \\ &\leq \left[\varepsilon^{-\eta}(A_\varepsilon + 1)\right]^\gamma \varepsilon^\beta \int_{B_{\varepsilon^{-\eta}}} f(y)\frac{dy}{|B_{\varepsilon^{-\eta}}|}(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d \frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1} \\ &= (A_\varepsilon + 1)^\gamma \varepsilon^{\beta - \eta\gamma}\lambda_i'. \end{aligned} \tag{5.6.18}$$

Similar, we claim that the map g, defined by

$$g : y \longrightarrow \left(\Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right) - \Delta_\varepsilon\left(x, y_i\right)\right)\exp\left(-\frac{\Delta_\varepsilon\left(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)}{\varepsilon}\right),$$

satisfies that $\frac{g}{\left\|g\right\|^{\varepsilon^{-\eta}}_\infty}$ is $e^{-1}A_\varepsilon \varepsilon^{-\eta}-$Lipschitz. Now, we want to compare

$$D_i := \int_{B_{\varepsilon^{-\eta}}} g(y) d(m_\varepsilon)_x \circ \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(dy)\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1},$$

with

$$D_i' := \int_{B_{\varepsilon^-\eta}} g(y)dy \frac{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i)]/2^d}{|B_{\varepsilon^\eta}|}\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}.$$

Hence similar, by (5.6.15), we have

$$\begin{aligned}
|D_i - D_i'| &= \left|\int_{B_{\varepsilon^{-\eta}}} g(y)\left[\frac{d(m_\varepsilon)_x\circ \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(dy)}{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d} - \frac{dy}{|B_{\varepsilon^{-\eta}}|}\right]\right|(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}\\
&\le \left[\varepsilon^{-\eta}(e^1A_\varepsilon+1)\right]^\gamma\varepsilon^\beta\int_{B_{\varepsilon^{-\eta}}} g(x)\frac{dx}{|B_{\varepsilon^{-\eta}}|}(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}\\
&= (e^{-1}A_\varepsilon+1)^\gamma\varepsilon^{\beta-\eta\gamma}D_i'. \qquad (5.6.19)
\end{aligned}$$

Now we denote $K_i := \frac{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d}{\left|B_{\varepsilon^{-\eta}}\right|}\frac{d\mu_\varepsilon}{dm^{\mathcal{X}}}(x)^{-1}\exp\left(-\frac{\Delta_\varepsilon\left(x,y_i\right)}{\varepsilon}\right)$, so that

$$\lambda_i' = K_i\int_{B_{\varepsilon^{-\eta}}}\exp\left(-\frac{\Delta_\varepsilon\left(x,\mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)-\Delta_\varepsilon\left(x,y_i\right)}{\varepsilon}\right)dy.$$

We now compare λ_i' with $\lambda_i'' := K_i\int_{\mathbb{R}^d} e^{-\partial_y^2\Delta_\varepsilon(x,y_i)y^2}dy$. By the formula of the Gaussian integral, we have $\lambda_i'' = K_i\sqrt{2\Pi}^d\sqrt{\det\partial_y^2\Delta_\varepsilon\left(x,y_i\right)}$. Similar to (5.6.14), the part of the integral out of $B_{\varepsilon^{-\eta}}$ is uniformly negligible in front of ε. We assume that $\varepsilon > 0$ is small enough so that this integral is uniformly smaller than ε. By (iv), we have that

$$\begin{aligned}
\left(\partial_y^2\Delta_\varepsilon\left(x,y_i\right)-\varepsilon^\eta\right)(y-y_i)^2 &\le \Delta_\varepsilon\left(x,\mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\right)-\Delta_\varepsilon\left(x,y_i\right)\\
&\le \left(\partial_y^2\Delta_\varepsilon\left(x,y_i\right)+\varepsilon^\eta\right)(y-y_i)^2.
\end{aligned}$$

Therefore, we have

$$K_i\sqrt{2\Pi}^d\sqrt{\det\left[\partial_y^2\Delta_\varepsilon\left(x,y_i\right)-\varepsilon^\eta I_d\right]}-\varepsilon \le \lambda_i' \le K_i\sqrt{2\Pi}^d\sqrt{\det\left[\partial_y^2\Delta_\varepsilon\left(x,y_i\right)-\varepsilon^\eta I_d\right]}+\varepsilon.$$

By the fact that $\varepsilon I_d \ll \varepsilon^\eta I_d \ll A_\varepsilon^{-1}I_d \le \partial_y^2\Delta_\varepsilon\left(x,y_i\right)$, we may find $C_4, q_4 > 0$ such that

$$|\lambda_i' - \lambda_i''| \le C_4A_\varepsilon^{q_4}\varepsilon^\eta\lambda_i''. \qquad (5.6.20)$$

Similar, we get that the integral of g may be approximated by the integral of

$$\widetilde{g}(y) := \varepsilon\partial_y^2\Delta_\varepsilon\left(x,y_i\right)y^2\exp\left(-\partial_y^2\Delta_\varepsilon\left(x,y_i\right)y^2\right).$$

Let $D_i'' := K_i \int_{\mathbb{R}^d} \widetilde{g}(y)dy$. Similar than the previous computation, up to raising C_4 and q_4, we have

$$|D_i' - D_i''| \leq C_4 A_\varepsilon^{q_4} \varepsilon^\eta D_i''. \tag{5.6.21}$$

Now we compute the value of D_i''. By change of variables $z = \sqrt{\partial_y^2 \Delta_\varepsilon\big(x, y_i\big)}y$, where $\sqrt{A}$ applied to a symmetrical positive definite matrix denotes the only symmetrical positive definite square root of the matrix A, we get that $D_i'' = \varepsilon K_i \frac{d}{2}\sqrt{\det \partial_y^2 \Delta_\varepsilon\big(x, y_i\big)}$.

We observe that from (5.6.16) and (i), together with (5.6.18), (5.6.19), (5.6.20), and (5.6.21), we have for all i that $\lambda_i = \lambda_i' + o(\lambda_i')$, $\lambda_i' = \lambda_i'' + o(\lambda_i'')$, $D_i = D_i' + o(D_i')$, and $D_i' = D_i'' + o(D_i'')$. Finally, using (5.1.1) and the fact that we can sum up positive $o(\cdot)$, we get

$$\begin{aligned}
\sum_{i=1}^{k_x} \int_{B_{r_\varepsilon}(y_i)} \left(\Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,y_i)\right)(\mathbb{P}_\varepsilon)_x(dy) &= \sum_{i=1}^{k} D_i \\
&= \sum_{i=1}^{k} D_i'' + o\left(\sum_{i=1}^{k} D_i' + D_i''\right) \\
&= \sum_{i=1}^{k} \varepsilon K_i \frac{d}{2}\sqrt{\det \partial_y^2 \Delta_\varepsilon\big(x, y_i\big)} + o\left(\sum_{i=1}^{k} D_i\right) \\
&= \varepsilon \frac{d}{2} \sum_{i=1}^{k} \lambda_i'' + o\left(\sum_{i=1}^{k} D_i\right) \\
&= \varepsilon \frac{d}{2} \sum_{i=1}^{k} \lambda_i + o\left(\sum_{i=1}^{k} D_i + \varepsilon(\lambda_i' + \lambda_i'')\right) \\
&= \varepsilon \frac{d}{2} + o\left(\varepsilon + \sum_{i=1}^{k} D_i\right) \\
&= \varepsilon \frac{d}{2} + o(\varepsilon),
\end{aligned}$$

where all the $o(\cdot)$ are uniform in x, thanks to all the controls established in this step. The claim is proved.

Step 10: We claim that for $\varepsilon > 0$ small enough, we have

$$\text{dist}\Big(x, \text{aff}(y_1, ..., y_k)\Big) \leq C_5 A_\varepsilon^{q_5} \varepsilon^{\frac{1}{2}+\beta-\eta\gamma} + C_6 A_\varepsilon^{q_6} \varepsilon^{\frac{1}{2}+\eta} + \varepsilon. \tag{5.6.22}$$

Indeed, let $z_i := \lambda_i^{-1}\int_{B_{r_\varepsilon}(y_i)} ye^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} m_\varepsilon^{\mathcal{Y}}(dy)\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}$, where recall that

$$\lambda_i = \int_{B_{r_\varepsilon}(y_i)} e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} m_\varepsilon^{\mathcal{Y}}(dy)\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}.$$

We have that $x = \int_{\mathbb{R}^d} ye^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} m_\varepsilon^{\mathcal{Y}}(dy)\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}$, by the martingale property of $e^{-\frac{\Delta_\varepsilon}{\varepsilon}}m_\varepsilon$. Similar to (5.6.14), we have $\left|\sum_{i=1}^k \lambda_i z_i - x\right| \ll \varepsilon$. Similar, we also have $\sum_{i=1}^k \lambda_i = 1 + o(\varepsilon)$. Therefore, $\mathrm{dist}\Big(x, \mathrm{aff}(z_1,...,z_k)\Big) \ll \varepsilon$. Now let

$$z_i' := \frac{1}{\lambda_i'}\int_{B_{r_\varepsilon}(y_i)} ye^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} dy \frac{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d}{|B_{r_\varepsilon}|}\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1},$$

where recall that $\lambda_i' = \int_{B_{r_\varepsilon}(y_i)} e^{-\frac{\Delta_\varepsilon(x,y)}{\varepsilon}} dy \frac{(m_\varepsilon)_x[B_{r_\varepsilon}(y_i)]}{|B_{r_\varepsilon}|}\left(\frac{d\mu_\varepsilon}{dm_\varepsilon^{\mathcal{X}}}\right)^{-1}$. We claim that we may find universal $C_5, q_5 > 0$, such that $|\lambda_i(z_i - y_i) - \lambda_i'(z_i' - y_i)| \leq C_5 A_\varepsilon^{q_5}\varepsilon^{\frac{1}{2}+\beta-\eta\gamma}$. Indeed, if u is a unit vector, we have

$$h : y \longrightarrow \sqrt{\varepsilon}|y\cdot u|\exp\left(-\frac{\Delta_\varepsilon\Big(x, \mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(y)\Big) - \Delta_\varepsilon(x,y_i)}{\varepsilon}\right).$$

We claim that $\frac{h}{\|h\|_\infty^{\varepsilon^{-\eta}}}$ is $\varepsilon^{-\eta}A_\varepsilon\left(1+\frac{\sqrt{A_\varepsilon}}{2y_0^3}\right)$−Lipschitz, where $y_0 > 0$ is the unique positive real satisfying $2y_0^2 = e^{-y_0^2}$, furthermore the function is small enough on $B_{2R}(y_i)\setminus B_R(y_i)$ so that we may apply (5.6.15):

$$\begin{aligned}\Big|u\cdot\Big(\lambda_i(z_i-y_i)-\lambda_i'(z_i'-y_i)\Big)\Big| &= \left|\int_{B_{\varepsilon^{-\eta}}(y_i)} h(y)\left[\frac{d(m_\varepsilon)_x\circ\mathrm{zoom}^{y_i}_{\sqrt{\varepsilon}}(dy)}{(m_\varepsilon)_x[B_{2r_\varepsilon}(y_i')]/2^d} - \frac{dy}{|B_{\varepsilon^{-\eta}}|}\right]\right| K_i|B_{\varepsilon^{-\eta}}| \\ &\leq \left[\varepsilon^{-\eta}A_\varepsilon\left(1+\frac{\sqrt{A_\varepsilon}}{2y_0^3}\right)\right]^\gamma \varepsilon^\beta\int_{B_{\varepsilon^{-\eta}}(y_i)} h(y)dyK_i, \\ &\leq \left[A_\varepsilon\left(1+\frac{\sqrt{A_\varepsilon}}{2y_0^3}\right)\right]^\gamma \varepsilon^{\frac{1}{2}+\beta-\gamma\eta} IA_\varepsilon K_i, \end{aligned} \quad (5.6.23)$$

where $I := \int_{\mathbb{R}}^d |y \cdot u| e^{-|y|^2} dy$, as recall that by Step 9, $K_i = \frac{\lambda_i''}{\sqrt{2\Pi}^d \sqrt{\det \partial_y^2 \Delta_\varepsilon(x,y_i)}} \leq A_\varepsilon^{d/2}$. Now we consider again a unit vector u, and

$$\begin{aligned}
(z_i' - y_i) \cdot u &= \frac{\sqrt{\varepsilon}}{\lambda_i'} \int_{B_{\varepsilon^{-\eta}}} \Big((y \cdot u)_+ - (y \cdot u)_- \Big) e^{-\frac{\Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,y_i)}{\varepsilon}} dy K_i \\
&\leq \frac{\sqrt{\varepsilon}}{\lambda_i'} \int_{B_{\varepsilon^{-\eta}}} (y \cdot u)_+ e^{-(\partial_y^2 \Delta_\varepsilon(x,y_i) - \varepsilon^\eta I_d) y^2} dy K_i \\
&\quad - \frac{\sqrt{\varepsilon}}{\lambda_i'} \int_{B_{\varepsilon^{-\eta}}} (y \cdot u)_- e^{-(\partial_y^2 \Delta_\varepsilon(x,y_i) + \varepsilon^\eta I_d) y^2} dy K_i \\
&\leq \frac{K_i \sqrt{\varepsilon}}{\lambda_i'} (I/2) \left(\sqrt{\partial_y^2 \Delta_\varepsilon(x,y_i) - \varepsilon^\eta I_d}^{-1} u^2 - \sqrt{\partial_y^2 \Delta_\varepsilon(x,y_i) + \varepsilon^\eta I_d}^{-1} u^2 \right) \\
&\leq C_6 A_\varepsilon^{q_6} \varepsilon^{\frac{1}{2} + \eta},
\end{aligned}$$

for some $C_6, q_6 > 0$, independent of x and u, as $\partial_y^2 \Delta_\varepsilon(x,y_i) \geq A_\varepsilon^{-1} I_d \gg \varepsilon^\eta I_d$, and $\lambda_i' \geq \frac{1}{2} A_\varepsilon^{-1}$. Then $|z_i' - y_i| \leq C_6 A_\varepsilon^{q_6} \varepsilon^{\frac{1}{2} + \eta}$.

Finally, with (5.6.18) and (5.6.20), up to raising C_5, C_6, q_5, q_6, we get the estimate $|z_i - y_i| \leq C_5 A_\varepsilon^{q_5} \varepsilon^{\frac{1}{2} + \beta - \gamma\eta} + C_6 A_\varepsilon^{q_6} \varepsilon^{\frac{1}{2} + \eta}$. We finally get the desired estimate from the fact that $\text{dist}\Big(x, \text{aff}(y_1, ..., y_k)\Big)$.

Step 11: We now claim that

$$\int_{B_{r_\varepsilon}(y_i)} \bar{\Delta}_\varepsilon(x,\cdot) d(\mathbb{P}_\varepsilon)_x = \int_{B_{r_\varepsilon}(y_i)} \Big(\Delta_\varepsilon(x,\cdot) - \Delta_\varepsilon(x,y_i) \Big) d(\mathbb{P}_\varepsilon)_x + o(\varepsilon), \qquad (5.6.24)$$

where the convergence speed of $o(\varepsilon)$ is independent of the choice of x and i. Indeed, by (5.6.22) and the fact that

$$|\Delta_\varepsilon(x,y_i)| \leq \max(C_1, C_2) \varepsilon \ln(\varepsilon^{-1}) \leq \varepsilon^{1 - \frac{\eta}{2}} \leq r_\varepsilon H_\varepsilon,$$

with $H_\varepsilon := \varepsilon^{\frac{1}{2} + \frac{1}{2} \min(\eta, \beta - \eta\gamma)}$ for ε small enough. Furthermore, by (5.6.22), we have that for $\varepsilon > 0$ small enough, $\text{dist}\Big(x, \text{aff}(y_1, ..., y_k)\Big) \leq H_\varepsilon$. Then, we may apply Lemma 5.6.3: we may find $C_7, q_7 > 0$ such that we have $\bar{\Delta}_\varepsilon(x,y) = \Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,\bar{y}_i) - \nabla \Delta_\varepsilon(x,\bar{y}_i) \cdot (y - \bar{y}_i)$, with $|y_i - \bar{y}_i| \leq C_7 A_\varepsilon^{q_7} H_\varepsilon \leq C_7 A_\varepsilon^{q_7} \varepsilon^{\frac{1}{2}}$, as by Step 6, we have that for $y \notin \cup_{i=1}^k B_{r_\varepsilon}(y_i)$, we have $\Delta_\varepsilon(x,y) \geq \frac{1}{2} \sqrt{\varepsilon}\, \text{dist}\Big(y, (y_1, ..., y_k)\Big)$, where $\frac{1}{2}\sqrt{\varepsilon} \gg C_7 A_\varepsilon^{q_7} H_\varepsilon$, and $r_\varepsilon \gg C_7 A_\varepsilon^{q_7} H_\varepsilon$ as well. Notice that therefore, we have $0 \leq \Delta_\varepsilon(x,\bar{y}_i) - \Delta_\varepsilon(x,y_i) \leq \frac{1}{2} A_\varepsilon C_7^2 A_\varepsilon^{2q_7} H_\varepsilon^2 \ll \varepsilon$, and similar, we have $|\nabla \Delta_\varepsilon(x,\bar{y}_i) \cdot (y_i - \bar{y}_i)| \leq A_\varepsilon C_7^2 A_\varepsilon^{2q_7} H_\varepsilon^2 \ll \varepsilon$. Finally, $\int_{B_{r_\varepsilon}(y_i)} \nabla \Delta_\varepsilon(x,\bar{y}_i) \cdot (y - y_i) d(\mathbb{P}_\varepsilon)_x = \nabla \Delta_\varepsilon(x,\bar{y}_i) \cdot \int_{B_{r_\varepsilon}(y_i)} (y - y_i) d(\mathbb{P}_\varepsilon)_x$. We have from the computations in Step 10 that $\left| \int_{B_{r_\varepsilon}(y_i)} (y - y_i) d(\mathbb{P}_\varepsilon)_x \right| \leq C_3 A_\varepsilon^{q_3} \varepsilon^{\frac{1}{2}}$, and $|\nabla \Delta_\varepsilon(x,\bar{y}_i) \cdot \int_{B_{r_\varepsilon}(y_i)} (y - y_i) d(\mathbb{P}_\varepsilon)_x| \leq C_3 C_7 A_\varepsilon^{q_3 + 2q_7} \varepsilon^{\frac{1}{2}} H_\varepsilon \ll \varepsilon$, whence the result.

Step 12: Now using Step 3 and Step 9, integrating against μ_ε, with the uniform error estimate (5.6.24), together with controls that are independent of x, similar to (5.6.14), to deal with $\left(\cup_{i=1}^k B_{r_\varepsilon}(y_i)\right)^c$, we get

$$\begin{aligned}\int \bar{\Delta}_\varepsilon d\mathbb{P}_\varepsilon &= \int_{D_\varepsilon^{\mathcal{X}}} \sum_{i=1}^{k_x} \int_{B_{r_\varepsilon}(y_i)} \left(\Delta_\varepsilon(x,y) - \Delta_\varepsilon(x,y_i)\right)(\mathbb{P}_\varepsilon)_x(dy)\mu_\varepsilon(dx) + o(\varepsilon)\\ &= \varepsilon\frac{d}{2} + o(\varepsilon).\end{aligned}$$

Finally, notice that

$$\begin{aligned}\bar{\Delta}_\varepsilon &= \Delta_\varepsilon - \Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X) - \nabla\Big(\Delta_\varepsilon(X,\cdot)\Big)_{conv}(X)\cdot(Y-X)\\ &= \psi_\varepsilon - c - \Big(\psi_\varepsilon - c(X,\cdot)\Big)_{conv}(X) - \nabla\Big(\psi_\varepsilon - c(X,\cdot)\Big)_{conv}(X)\cdot(Y-X)\\ &= \bar{\varphi}_\varepsilon + \psi_\varepsilon - c - \nabla\Big(\psi_\varepsilon - c(X,\cdot)\Big)_{conv}(X)\cdot(Y-X).\end{aligned}$$

Therefore we have $\mathbb{P}_\varepsilon[\bar{\Delta}_\varepsilon] = \mu_\varepsilon[\bar{\varphi}_\varepsilon] + \nu_\varepsilon[\psi_\varepsilon] - \mathbb{P}_\varepsilon[c]$. Whence the result of the theorem. □

5.6.5 Asymptotic penalization error

Proof of Proposition 5.5.14 As $\mu \preceq_c \nu$, $\widetilde{V}_\varepsilon$ is convex with a finite global minimum. Then the minimum of $\widetilde{V}_\varepsilon + \alpha f$ converges to a minimum of $\widetilde{V}_\varepsilon$. More precisely, let ψ_α be the only global minimizer of $\widetilde{V}_\varepsilon + \alpha f$, then ψ_α is also the minimizer of the map $\frac{1}{\alpha}\left(\widetilde{V}_\varepsilon - \widetilde{V}_\varepsilon(\psi_0)\right) + f$, which $\Gamma-$converges to $\iota_{\left\{\psi:\widetilde{V}_\varepsilon(\psi)\neq\widetilde{V}_\varepsilon(\psi_0)\right\}} + f$, whose unique global minimizer is ψ_0. Therefore $\psi_\alpha \longrightarrow \psi_0$ when $\alpha \longrightarrow 0$. Now the first order condition gives that $\frac{\nu_\alpha - \nu}{\alpha} = \nabla f(\psi_\alpha) \longrightarrow \nabla f(\psi_0)$, when $\alpha \longrightarrow 0$, by convexity and differentiability of f, guaranteeing that $\psi \longmapsto \nabla f(\psi)$ is continuous. □

5.6.6 Convergence of the martingale Sinkhorn algorithm

Proof of Theorem 5.5.15 This result stems from an indirect application of Theorem 5.2 in [15]. By a direct application of this theorem we get that

$$V_\varepsilon(x_k) - V_\varepsilon^* \le \tfrac{2\min(L_1,L_2)R^2(x_0)}{n-1}, \tag{5.6.25}$$

$$\text{and that}\quad V_\varepsilon(x_k) - V_\varepsilon^* \le \left(1 - \tfrac{\sigma}{\min(L_1,L_2)}\right)^{n-1}\left(V_\varepsilon(x_0) - V_\varepsilon^*\right), \tag{5.6.26}$$

with $R(x_0) := \sup_{V_\varepsilon(x)\leq V_\varepsilon(x_0)} \mathrm{dist}(x,\mathcal{X}^*)$, L_1 (resp. L_2) is the Lipschitz constant of the $\psi-$gradient (resp. $(\varphi,h)-$gradient) of V_ε, and σ is the strong convexity parameter of V_ε. Furthermore, the strong convexity gives that

$$\mathrm{dist}(x_k,\mathcal{X}^*) \leq \sqrt{\frac{2}{\sigma}}(V_\varepsilon(x_k) - V_\varepsilon^*)^{\frac{1}{2}}. \tag{5.6.27}$$

However the gradient ∇V_ε is locally but not globally Lipschitz, nor V_ε strongly convex. Therefore we need to refine the theorem by looking carefully at where these constants are used in its proof.

Step 1: The constant L_1 is used for Lemma 5.1 in [15]. We need for all $k\geq 0$ to have $V_\varepsilon(x_k) - V_\varepsilon(x_{k+1/2}) \geq \frac{1}{2L_1}|\nabla V_\varepsilon(x_k)|^2$. We may start from x_1, this way all the x_k are such that $(e^{-\frac{\Delta(x_k)_{i,j}}{\varepsilon}})_{i\in\mathcal{X},j\in\mathcal{Y}}$ is a probability distribution. Then $|\partial^2_\psi V_\varepsilon(x_k)| \leq \varepsilon^{-1}$. Let $u\in\mathbb{R}^{\mathcal{Y}}$, then $|\partial^2_\psi V_\varepsilon(\varphi_k,\psi_k+u,h_k)| \leq \varepsilon^{-1}e^{\frac{|u|_\infty}{\varepsilon}}$. We want to find $C,L>0$ such that $V_\varepsilon(x_k) - V_\varepsilon(\varphi_k,\psi_k - C\partial_\psi V_\varepsilon(x_k),h_k) \geq \frac{1}{2L}|\partial_\psi V_\varepsilon(x_k)|$, then L may be use to replace L_1 in the final step of the proof of Lemma 5.1 in [15]. Recall that $|\partial_\psi V_\varepsilon(x_k)|_\infty \leq 1$, as it is the difference of two probability vectors. We have

$$\begin{aligned}
&V_\varepsilon(x_k) - V_\varepsilon\Big(\varphi_k,\psi_k - C\partial_\psi V_\varepsilon(x_k),h_k\Big)\\
&= -\partial_\psi V_\varepsilon(x_k)\cdot\Big(-C\partial_\psi V_\varepsilon(x_k)\Big)\\
-C^2\int_0^1 &(1-t)\partial_\psi V_\varepsilon(x_k)^t\partial^2_\psi V_\varepsilon\Big(\varphi_k,\psi_k - tC\partial_\psi V_\varepsilon(x_k),h_k\Big)\partial_\psi V_\varepsilon(x_k)dt\\
&\leq C|\partial_\psi V_\varepsilon(x_k)|^2 - C^2|\partial_\psi V_\varepsilon(x_k)|^2\int_0^1\varepsilon^{-1}(1-t)e^{\frac{tC}{\varepsilon}}dt\\
&= C|\partial_\psi V_\varepsilon(x_k)|^2 - C^2|\partial_\psi V_\varepsilon(x_k)|^2\varepsilon^{-1}\frac{e^{\frac{C}{\varepsilon}}-1-\frac{C}{\varepsilon}}{\frac{C^2}{\varepsilon^2}}\\
&= \left(C - \varepsilon\left(e^{\frac{C}{\varepsilon}} - 1 - \frac{C}{\varepsilon}\right)\right)|\partial_\psi V_\varepsilon(x_k)|^2.
\end{aligned}$$

Deriving with respect to C gives the equation $C = \varepsilon\ln(2)$. We get

$$V_\varepsilon(x_k) - V_\varepsilon\Big(\varphi_k,\psi_k - C\partial_\psi V_\varepsilon(x_k),h_k\Big) \geq \varepsilon\Big(2\ln(2)-1\Big)|\partial_\psi V_\varepsilon(x_k)|^2.$$

Therefore we may use $L := \varepsilon^{-1}\Big(4\ln(2)-2\Big)^{-1}$.

Step 2: The constant σ is used to get the result from (3.21) in [15]. Then we just need the inequality

$$V_\varepsilon(y) \geq V_\varepsilon(x) + \nabla V_\varepsilon(x)\cdot(y-x) + \frac{\sigma}{2}|y-x|^2, \tag{5.6.28}$$

to hold for some $y\in\mathcal{X}^*$ and $x=x_k$ for all $k\geq 0$. Now we give a lower bound for σ. Notice that $V_\varepsilon=\mu[\varphi]+\nu[\psi]+\varepsilon\sum_{x\in\mathcal{X},y\in\mathcal{Y}}exp\left(-\frac{\cdot}{\varepsilon}\right)\circ\Delta_{x,y}$. Then for $x_0,u\in\mathcal{D}_{\mathcal{X},\mathcal{Y}}$, we have

$$\begin{aligned} u^tD^2V_\varepsilon(x_0)u &= \varepsilon^{-1}\sum_{x\in\mathcal{X},y\in\mathcal{Y}}exp\left(-\frac{\cdot}{\varepsilon}\right)\circ\Delta_{x,y}(x_0)\Delta_{x,y}(u)^2 \\ &\geq \varepsilon^{-1}exp\left(-\frac{|\Delta(x)|_\infty}{\varepsilon}\right)|\Delta(u)|^2. \end{aligned}$$

Then, by definition of λ_2, we may find $\widetilde{u}$ such that $\Delta(u)=\Delta(\widetilde{u})$, and

$$u^tD^2V_\varepsilon(x_0)u\geq\frac{|\mathcal{X}|}{\lambda_2\varepsilon}exp\left(-\frac{|\Delta(x)|_\infty}{\varepsilon}\right)|\widetilde{u}|^2. \tag{5.6.29}$$

Now, we claim that $|\Delta(x)|_\infty\leq D(x_0)$. Then let $x^*\in\mathcal{X}^*$ and consider (5.6.29) for $u=x^*-x$. Then we have that $x+\widetilde{u}\in\mathcal{X}^*$, and therefore, we may take $y=x+\widetilde{u}$ for (5.6.28), and therefore use

$$\sigma:=\frac{|\mathcal{X}|}{\lambda_2\varepsilon}exp\left(-\frac{D(x_0)}{\varepsilon}\right), \tag{5.6.30}$$

in this equation.
Step 3: Now we prove our claim that $|\Delta(x)|_\infty\leq D(x_0)$. Indeed

$$V_\varepsilon(x_0)\geq V_\varepsilon(x)=\mu[\varphi]+\nu[\psi]+\varepsilon=\mathbb{P}_0[\varphi\oplus\psi+h^\otimes]+\varepsilon=\mathbb{P}_0[\Delta(x)]+\mathbb{P}_0[c]+\varepsilon.$$

Therefore we have $\mathbb{P}_0[\Delta(x)]\leq V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon$, and finally

$$(\mathbb{P}_0)_{min}|\Delta(x)|_1\leq V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon. \tag{5.6.31}$$

Then $|\Delta(x)|_\infty\leq D(x_0)$ stems from the definition of λ_1.
Step 4: Now we provide the bound on $R(x_0)$. From the proof of Theorem 5.2 in [15], what is needed to make the proof work is $R(x_0)=\sup_{k\geq0}\mathrm{dist}(x_k,\mathcal{X}^*)$, which is smaller than $\sup_{V_\varepsilon(x)\leq V_\varepsilon(x_0)}\mathrm{dist}(x,\mathcal{X}^*)$. Furthermore, from (5.6.27) together with (5.6.30), we get that the supremum $\sup_{k\geq0}\mathrm{dist}(x_k,\mathcal{X}^*)$ is also smaller than $\sqrt{\frac{\lambda_2\varepsilon}{|\mathcal{X}|}e^{\frac{D(x_0)}{\varepsilon}}}\big(V_\varepsilon(x_0)-V_\varepsilon^*\big)^{\frac12}$. Finally, from (5.6.31) together with the definition of λ_1, we may find that $\widetilde{x}^*,\widetilde{x}_k\in\mathcal{D}_{\mathcal{X},\mathcal{Y}}$ such that $\Delta(x_k)=\Delta(\widetilde{x}_k)$, $\widetilde{x}^*\in\mathcal{X}^*$, $|\widetilde{x}_k|_1\leq\lambda_1\frac{V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}}$, and $|\widetilde{x}^*|_1\leq\lambda_1\frac{V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}}$. Then $|\widetilde{x}_k-\widetilde{x}^*|\leq|\widetilde{x}_k-\widetilde{x}^*|_1\leq2\lambda_1\frac{V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}}$ by the fact that $|\cdot|\leq$

$|\cdot|_1$. Finally, as $\widetilde{x}^*+x_k-\widetilde{x}_k\in\mathcal{X}^*$, we have $\mathrm{dist}(x_k,\mathcal{X}^*)\le 2\lambda_1\frac{V_\varepsilon(x_0)-\mathbb{P}_0[c]-\varepsilon}{|\mathcal{X}|(\mathbb{P}_0)_{\min}}$. Therefore the same bound holds for $\sup_{k\ge 0}\mathrm{dist}(x_k,\mathcal{X}^*)$.
Step 5: Finally, as we focus on the L_1 optimization phase, we may replace $n-1$ by n in the convergence formula (5.6.25) and (5.6.26), see the proof of Theorem 5.2 in [15]. The result is proved. □

5.6.7 Implied Newton equivalence

Proof of Proposition 5.5.20 We apply the Newton step in the algorithm to $\big(x,y(x)\big)$. we are looking for p such that $D^2Fp=\nabla F$. First $\nabla F\big(x,y(x)\big)=\partial_xF\big(x,y(x)\big)=\nabla\tilde{F}(x)$, then if we decompose $p=p_x\oplus p_y\in\mathcal{X}\oplus\mathcal{Y}$, the equation becomes

$$\partial_x^2Fp_x+\partial_{xy}^2Fp_y=\nabla\tilde{F}(x),\quad\text{and}\quad\partial_{yx}^2Fp_x+\partial_y^2Fp_y=0.$$

The solution to this equation system is given by $p_y=-\partial_y^2F^{-1}\partial_{yx}^2Fp_x$, and

$$\Big(\partial_x^2Fp_x-\partial_{xy}^2F\partial_y^2F^{-1}\partial_{yx}^2F\Big)p_x=\nabla\tilde{F}(x)\tag{5.6.32}$$

The conclusion follows from the fact that (5.6.32) is the step for the Newton algorithm applied to $\tilde{F}$. The Newton step on y does not matter, as y will be immediately thrown away and replaced by $y(x)$. □

5.7 Numerical experiment

5.7.1 An hybrid algorithm

The steps of the Newton algorithm are theoretically very performing if the current point is close enough to the optimum. What is really time-consuming is the computation of the descent direction with the conjugate gradient algorithm. The idea of preferring the Newton method to the Bregman projection method in the case of martingale optimal transport comes from the fact that, unlike in the case of classical transport, projecting on the martingale constraint is more costly than projecting on the marginal constraints, as we use a Newton algorithm instead of a closed formula. From the experiment, we would say that in dimension 1 it takes 7 times more time, and 20 times more in dimension 2. The implied Newton algorithm performs this projection only for the Newton step, whereas it is not necessary for the conjugate gradient algorithm.

We Notice that the Bregman projection algorithm is more effective at the beginning, to find the optimal region, and then it converges slower. In contrast, the Newton algorithm is slow at the beginning when it is searching the neighborhood of the optimum, but when its finds this neighborhood, the convergence gets very fast. Then it makes sense to apply an hybrid algorithm that starts with Bregman projections, and concludes with the Newton method. We call this dual-method algorithm the hybrid algorithm. We see on the simulations that it generally out-performs the two other algorithms.

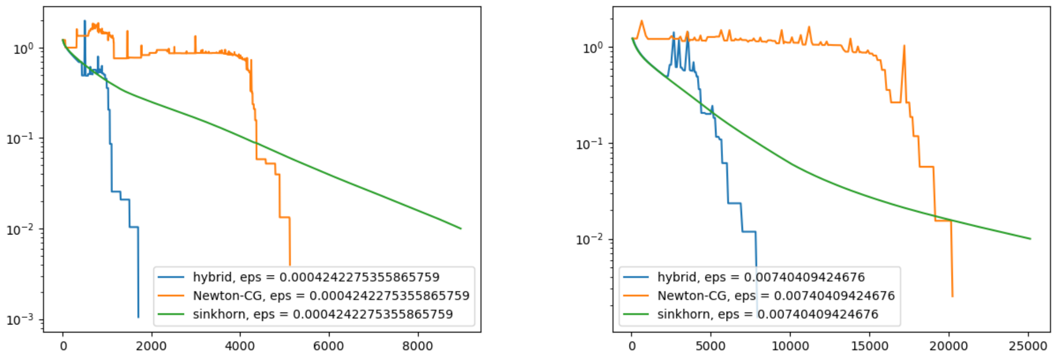
Figure 5.2 compares the evolution of the gradient error in dimension 1 and 2 of the longest step of the three algorithms in terms of computation time. What we call here the gradient error is the norm 1 of the gradient of the function $\widetilde{V}_\varepsilon$ that we are minimizing, and which is also equal to the difference between the target measure ν and the current measure. In the case of Newton algorithms, the penalization gradient is also included, then we use a coefficient in front of this penalization so that it does not interfere too much with the equation between the current and the target measure. We use the $\varepsilon-$scaling technique. For each value of ε, we iterate the minimization algorithm until the error is smaller than 10^{-2}. Then at the final iteration we lower the target error to the one we want.

The green line corresponds to the Bregman projections algorithm. The orange line corresponds to the implied truncated Newton algorithm. All the techniques evocated in Section 5.4 are applied. We use the diagonal of the Hessian to precondition the conjugate gradient algorithm. The coefficient in front of the quadratic penalization, which is normalized by ν^2, is set to 10^{-2}. Finally the blue line corresponds to the "hybrid algorithm", which consists in doing some Bregman projection steps before switching to the implied truncated Newton algorithm. The moment of switching is chosen by very empirical criteria: we do it after having the initial error divided by 2 or after 100 iteration, or if the initial error is divided by 1.1 if the initial error is smaller than 0.1.

Figure 5.2a gives the computation times of these three entropic algorithms, for a grid size going from 10 to 2500 while ε goes from 1 to 10^{-4}, with the cost function $c := XY^2$, μ uniform on $[-1,1]$, and $\nu := \frac{1}{K}|Y|^{1.5}\mu$, where $K := (|Y|^{1.5}\mu)[\mathbb{R}]$. By [85] the optimal coupling that we get is the "left curtain" coupling studied in [20]. We show the curves for the value of ε that takes the largest amount of time, the one for which the time of computation is the most important for $\varepsilon = 4.2 \times 10^{-4}$.

We conduct the same experiment on a two dimensional problem. The difference of efficiency between the algorithms should be even bigger, as the computing of the optimal h becomes more costly, as the optimization of a convex function of two variables.

Let $d = 2$, $c : (x,y) \in \mathbb{R}^2 \times \mathbb{R}^2 \longmapsto x_1(y_1^2 + 2y_2^2) + x_2(2y_1^2 + y_2^2)$, μ uniform on $[-1,1]^2$, and $\nu = \frac{1}{K}(|Y_1|^{1.5} + |Y_2|^{1.5})\mu$ where $K := \Big((|Y_1|^{1.5} + |Y_2|^{1.5})\mu\Big)[\mathbb{R}^2]$. We start with a 10×10 grid and scale it to a 160×160 one while ε scales from 1 to 10^{-4}. Figure 5.2b gives the computation times of the three entropic algorithms. Once again we show the curves for the value of ε that takes the largest amount of time, the one for which the time of computation is the most important for $\varepsilon = 7.4 \times 10^{-3}$.

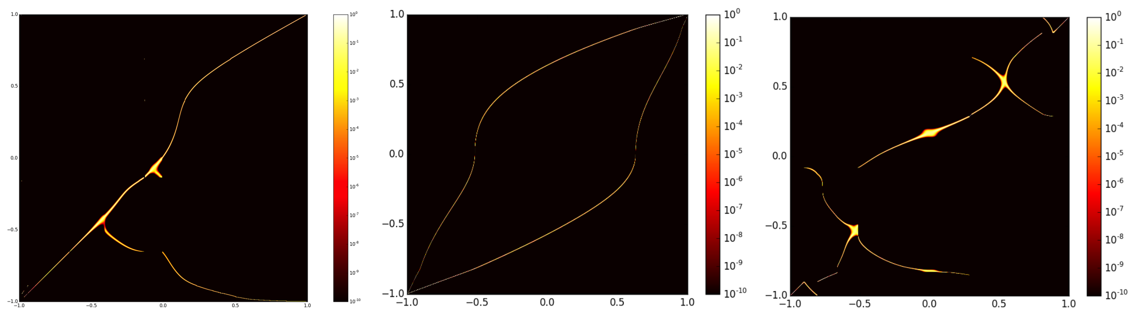


(a) Dimension 1. (b) Dimension 2.

Fig. 5.2 Log plot of the size of the gradient VS time for the Bregman projection algorithm, the Newton algorithm, and the Hybrid algorithm.

5.7.2 Results for some simple cost functions

Examples in one dimension



(a) $c = XY^2$. (b) $c = |X - Y|$. (c) $c = \sin(8XY)$.

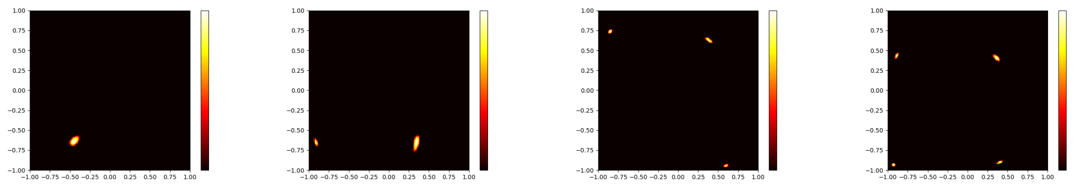
Fig. 5.3 Optimal coupling for different costs in dimension one.

Figure 5.3 give the solution for three different costs for $\varepsilon = 10^{-5}$ with $\mu := (\mu_1 + \mu_2)/2$ and $\nu := (\nu_1 + \nu_2)/2$ with mu_1 uniform on $[-1,1]$, $\nu_1 = \frac{1}{K}|Y|^{1.5}\mu_1$ with $K =$

$(\frac{1}{K}|Y|^{1.5}\mu_1)[\mathbb{R}]$, μ_2 is the law of $\exp(\mathcal{N}(-\frac{1}{2}\sigma_1^2,\sigma_1^2))-1$ with $\sigma_1=0.1$, and ν_2 is the law of $\exp(\mathcal{N}(-\frac{1}{2}\sigma_2^2,\sigma_2^2))-1$ with $\sigma_2=0.2$. The scale indicates the mass in each point of the grid, the mass of the entropic approximation of the optimal coupling is the yellow zone. Notice that in all the cases the optimal coupling is supported on at most two maps. We saw this in all our experiment, we conjecture that for almost all μ,ν this is the case.

Figure 5.3a shows well the left curtain coupling from [20] and [85]. Figure 5.3b shows the optimal coupling for the distance cost. This coupling has been studied by Hobson & Neuberger [93]. They predict that this coupling is concentrated on two graphs. Finally, Figure 5.3c shows how we may find solutions for any kind of cost function.

Example in two dimensions



(a) $X=-(0.45,0.65)$ (b) $X=(0.3,-0.66)$ (c) $X=(0.2,0.44)$ (d) $X=(0.13,0.16)$

Fig. 5.4 Optimal coupling conditioned to several values of X.

In dimension 2, it has been proved in [56] that for the cost function $c:(x,y)\in\mathbb{R}^2\times\mathbb{R}^2\longmapsto x_1(y_1^2+2y_2^2)+x_2(2y_1^2+y_2^2)$, the kernel of optimal probabilities are concentrated on the intersection of two ellipses with fixed characteristics, except for their position and their scale. Figure 5.4 is meant to test this theoretical result. We do an entropic approximation with a grid 160×160, and $\varepsilon=10^{-4}$. Then we selected 4 points $x_1:=(-0.45,-0.65)$, $x_2:=(0.3,-0.66)$, $x_3:=(0.2,0.44)$, and $x_4:=(0.13,0.16)$ and draw the kernels of the approximated optimal transport $\mathbb{P}^*$ conditioned to $X=x_i$ for $i=1,2,3,4$. We see on this figure that for $i=1,2,3,4$, $\mathbb{P}^*(\cdot|X=x_i)$ is concentrated on exactly i points, showing that all the numbers between 1 and 4 are reached. It seems that no trivial result can be proved on the number of maps that we may reach.

References

[1] Acciaio, B., Beiglböck, M., Penkner, F., and Schachermayer, W. (2016). A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance*, 26(2):233–251.

[2] Agueh, M. and Carlier, G. (2011). Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924.

[3] Ahuja, R. K., Magnanti, T. L., and Orlin, J. B. (1988). Network flows.

[4] Aksamit, A., Deng, S., Ob, J., Tan, X., et al. (2017). Robust pricing–hedging duality for american options in discrete time financial markets. Technical report.

[5] Alfonsi, A., Corbetta, J., and Jourdain, B. (2017). Sampling of probability measures in the convex order and approximation of martingale optimal transport problems.

[6] Altschuler, J., Weed, J., and Rigollet, P. (2017). Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In *Advances in Neural Information Processing Systems*, pages 1961–1971.

[7] Ambrosio, L. (2003). Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces*, pages 1–52. Springer.

[8] Ambrosio, L. and Gigli, N. (2013). A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, pages 1–155. Springer.

[9] Ambrosio, L., Gigli, N., and Savaré, G. (2008). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

[10] Ambrosio, L., Kirchheim, B., Pratelli, A., et al. (2004). Existence of optimal transport maps for crystalline norms. *Duke Mathematical Journal*, 125(2):207–241.

[11] Ambrosio, L. and Pratelli, A. (2003). Existence and stability results in the l1 theory of optimal transportation. In *Optimal transportation and applications*, pages 123–160. Springer.

[12] Bachelier, L. (1900). *Théorie de la spéculation*. Gauthier-Villars.

[13] Backhoff, J., Beiglboeck, M., Huesmann, M., and Källblad, S. (2017). Martingale benamou–brenier: A probabilistic perspective. *preprint*.

[14] Bartl, D. and Kupper, M. (2017). A pointwise bipolar theorem. *arXiv preprint arXiv:1702.02490*.

[15] Beck, A. and Tetruashvili, L. (2013). On the convergence of block coordinate descent type methods. *SIAM journal on Optimization*, 23(4):2037–2060.

[16] Beer, G. (1991). A polish topology for the closed subsets of a polish space. *Proceedings of the American Mathematical Society*, 113(4):1123–1133.

[17] Beiglböck, M., Cox, A. M., and Huesmann, M. (2014). Optimal transport and skorokhod embedding. *Inventiones mathematicae*, pages 1–74.

[18] Beiglböck, M., Henry-Labordère, P., and Penkner, F. (2013). Model-independent bounds for option prices: a mass transport approach. *Finance and Stochastics*, 17(3):477–501.

[19] Beiglböck, M., Henry-Labordere, P., and Touzi, N. (2015a). Monotone martingale transport plans and skorohod embedding. *preprint*.

[20] Beiglböck, M. and Juillet, N. (2016). On a problem of optimal transport under marginal martingale constraints. *The Annals of Probability*, 44(1):42–106.

[21] Beiglböck, M., Lim, T., and Oblój, J. (2017). Dual attainment for the martingale transport problem. *arXiv preprint arXiv:1705.04273*.

[22] Beiglböck, M., Nutz, M., and Touzi, N. (2015b). Complete duality for martingale optimal transport on the line. *arXiv preprint arXiv:1507.00671*.

[23] Beiglböck, M. and Pratelli, A. (2012). Duality for rectified cost functions. *Calculus of Variations and Partial Differential Equations*, 45(1-2):27–41.

[24] Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. Siam.

[25] Benamou, J.-D. and Brenier, Y. (2000). A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393.

[26] Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015). Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138.

[27] Benamou, J.-D., Collino, F., and Mirebeau, J.-M. (2016). Monotone and consistent discretization of the monge-ampere operator. *Mathematics of computation*, 85(302):2743–2775.

[28] Benamou, J.-D., Froese, B. D., and Oberman, A. M. (2014). Numerical solution of the optimal transportation problem using the monge–ampère equation. *Journal of Computational Physics*, 260:107–126.

[29] Bertsekas, D. P. (1988). The auction algorithm: A distributed relaxation method for the assignment problem. *Annals of operations research*, 14(1):105–123.

[30] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic optimal control: The discrete time case*, volume 23. Academic Press New York.

[31] Biagini, S., Bouchard, B., Kardaras, C., and Nutz, M. (2017). Robust fundamental theorem for continuous processes. *Mathematical Finance*, 27(4):963–987.

[32] Bianchini, S. and Caravenna, L. (2009). On the extremality, uniqueness and optimality of transference plans.

[33] Bianchini, S. and Cavalletti, F. (2013). The monge problem for distance cost in geodesic spaces. *Communications in Mathematical Physics*, 318(3):615–673.

[34] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.

[35] Bouchard, B., Nutz, M., et al. (2015). Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, 25(2):823–859.

[36] Brauer, C., Clason, C., Lorenz, D., and Wirth, B. (2017). A sinkhorn-newton method for entropic optimal transport. *arXiv preprint arXiv:1710.06635*.

[37] Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

[38] Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217.

[39] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417.

[40] Burzoni, M., Frittelli, M., Maggis, M., et al. (2017). Model-free superhedging duality. *The Annals of Applied Probability*, 27(3):1452–1477.

[41] Caffarelli, L., Feldman, M., and McCann, R. (2002). Constructing optimal maps for monge's transport problem as a limit of strictly convex costs. *Journal of the American Mathematical Society*, 15(1):1–26.

[42] Carlier, G., Duval, V., Peyré, G., and Schmitzer, B. (2017). Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418.

[43] Carr, P. and Madan, D. (1999). Option valuation using the fast fourier transform. *Journal of computational finance*, 2(4):61–73.

[44] Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2016). Scaling algorithms for unbalanced transport problems. *arXiv preprint arXiv:1607.05816*.

[45] Choquet, G. (1959a). Ensembles k-analytiques et k-sousliniens. cas général et cas métrique. *Ann. Inst. Fourier. Grenoble*, 9:75–81.

[46] Choquet, G. (1959b). Forme abstraite du théorème de capacitabilité. In *Annales de l'institut Fourier*, volume 9, pages 83–89.

[47] Claisse, J., Guo, G., and Henry-Labordere, P. (2015). Robust hedging of options on local time.

[48] Cominetti, R. and San Martín, J. (1994). Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1-3):169–187.

[49] Cox, A. (2008). Extending chacon-walsh: minimality and generalised starting distributions. In *Séminaire de probabilités XLI*, pages 233–264. Springer.

[50] Cox, A. M. and Källblad, S. (2017). Model-independent bounds for asian options: a dynamic programming approach. *SIAM Journal on Control and Optimization*, 55(6):3409–3436.

[51] Cox, A. M. and Obłój, J. (2011). Robust pricing and hedging of double no-touch options. *Finance and Stochastics*, 15(3):573–605.

[52] Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300.

[53] Cuturi, M. and Peyré, G. (2016). A smoothed dual approach for variational wasserstein problems. *SIAM Journal on Imaging Sciences*, 9(1):320–343.

[54] Davies, R. O. (1952). On accessibility of plane sets and differentiation of functions of two real variables. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pages 215–232. Cambridge University Press.

[55] De March, H. (2018a). Entropic resolution for multi-dimensional optimal transport (work in progress).

[56] De March, H. (2018b). Local structure of the optimizer of multi-dimensional martingale optimal transport. *arXiv preprint arXiv:1805.09469*.

[57] De March, H. (2018c). Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv preprint arXiv:1805.01757*.

[58] De March, H. and Touzi, N. (2017). Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *arXiv preprint arXiv:1702.08298*.

[59] Delbaen, F. and Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520.

[60] Deng, S. and Tan, X. (2016). Duality in nondominated discrete-time models for americain options. *arXiv preprint arXiv:1604.05517*.

[61] Denis, L. and Martini, C. (2006). A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *The Annals of Applied Probability*, pages 827–852.

[62] Dobrushin, R. L. (1996). Perturbation methods of the theory of gibbsian fields. In *Lectures on probability theory and statistics*, pages 1–66. Springer.

[63] Dolinsky, Y. and Soner, H. M. (2014). Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1-2):391–427.

[64] Dupire, B. (1997). Pricing and hedging with smiles. *Mathematics of derivative securities*, 1(1):103–111.

[65] Ekren, I. and Soner, H. M. (2016). Constrained optimal transport. *arXiv preprint arXiv:1610.02940*.

[66] El Karoui, N. and Quenez, M.-C. (1995). Dynamic programming and pricing of contingent claims in an incomplete market. *SIAM journal on Control and Optimization*, 33(1):29–66.

[67] Evans, L. C. and Gangbo, W. (1999). *Differential equations methods for the Monge-Kantorovich mass transfer problem*, volume 653. American Mathematical Soc.

[68] Evans, L. C. and Gariepy, R. F. (2015). *Measure theory and fine properties of functions*. CRC press.

[69] Fahim, A. and Huang, Y.-J. (2016). Model-independent superhedging under portfolio constraints. *Finance and Stochastics*, 20(1):51–81.

[70] Föllmer, H. and Schied, A. (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

[71] Fournier, N. and Guillin, A. (2015). On the rate of convergence in wasserstein distance of the empirical measure. *Probability Theory and Related Fields*, 162(3-4):707–738.

[72] Fremlin, D. H. (2000). *Measure theory*, volume 4. Torres Fremlin.

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

[74] Ghoussoub, N., Kim, Y.-H., and Lim, T. (2015). Structure of optimal martingale transport plans in general dimensions. *arXiv preprint arXiv:1508.01806*.

[75] Godel, K. (1947). What is cantor's continuum problem? *The American Mathematical Monthly*, 54(9):515–525.

[76] Goldberg, A. V. and Tarjan, R. E. (1989). Finding minimum-cost circulations by canceling negative cycles. *Journal of the ACM (JACM)*, 36(4):873–886.

[77] Guo, G. and Obloj, J. (2017). Computational methods for martingale optimal transport problems. *arXiv preprint arXiv:1710.07911*.

[78] Guo, G., Tan, X., and Touzi, N. (2016a). Optimal skorokhod embedding under finitely many marginal constraints. *SIAM Journal on Control and Optimization*, 54(4):2174–2201.

[79] Guo, G., Tan, X., and Touzi, N. (2016b). Tightness and duality of martingale transport on the skorokhod space. *Stochastic Processes and their Applications*.

[80] Gutiérrez, C. E. (2001). *The Monge-Ampere equation*, volume 42. Springer.

[81] Hagan, P. S., Kumar, D., Lesniewski, A. S., and Woodward, D. E. (2002). Managing smile risk. *The Best of Wilmott*, 1:249–296.

[82] Hartshorne, R. (2013). *Algebraic geometry*, volume 52. Springer Science & Business Media.

[83] Henry-Labordere, P. (2017). *Model-free Hedging: A Martingale Optimal Transport Viewpoint*. CRC Press.

[84] Henry-Labordere, P., Obłój, J., Spoida, P., Touzi, N., et al. (2016). The maximum maximum of a martingale with given **n** marginals. *The Annals of Applied Probability*, 26(1):1–44.

[85] Henry-Labordère, P. and Touzi, N. (2016). An explicit martingale version of the one-dimensional brenier theorem. *Finance and Stochastics*, 20(3):635–668.

[86] Hess, C. (1986). *Contribution à l'étude de la mesurabilité, de la loi de probabilité et de la convergence des multifonctions*. PhD thesis.

[87] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343.

[88] Himmelberg, C. J. (1975). Measurable relations. *Fundamenta mathematicae*, 87(1):53–72.

[89] Hiriart-Urruty, J.-B. and Lemaréchal, C. (2013). *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer science & business media.

[90] Hirsch, F., Profeta, C., Roynette, B., and Yor, M. (2011). *Peacocks and associated martingales, with explicit constructions*. Springer Science & Business Media.

[91] Hobson, D. (2011). The skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 267–318. Springer.

[92] Hobson, D. and Klimmek, M. (2015). Robust price bounds for the forward starting straddle. *Finance and Stochastics*, 19(1):189–214.

[93] Hobson, D. and Neuberger, A. (2012). Robust bounds for forward start options. *Mathematical Finance*, 22(1):31–56.

[94] Hobson, D. G. (1998). Robust hedging of the lookback option. *Finance and Stochastics*, 2(4):329–347.

[95] Hou, Z. and Obloj, J. (2015). On robust pricing-hedging duality in continuous time. *arXiv preprint arXiv:1503.02822*.

[96] Jameson, G. (2006). Counting zeros of generalised polynomials: Descartes' rule of signs and laguerre's extensions. *The Mathematical Gazette*, 90(518):223–234.

[97] Johansen, S. (1974). The extremal convex functions. *Mathematica Scandinavica*, 34(1):61–68.

[98] Källblad, S., Tan, X., Touzi, N., et al. (2017). Optimal skorokhod embedding given full marginals and azéma–yor peacocks. *The Annals of Applied Probability*, 27(2):686–719.

[99] Kantorovich, L. and Akilov, G. (1959). Functional analysis in normed spaces [funktsional'nyi analiz v normirovannykh prostranstvakh]. *Fizmatgiz, Moscow*.

[100] Kantorovich, L. V. (2006). On a problem of monge. *Journal of Mathematical Sciences*, 133(4):1383–1383.

[101] Kantorovich, L. V. and Rubinstein, G. S. (1958). On a space of completely additive functions. *Vestnik Leningrad. Univ*, 13(7):52–59.

[102] Kantorovitch, L. (1958). On the translocation of masses. *Management Science*, 5(1):1–4.

[103] Kellerer, H. G. (1984). Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 67(4):399–432.

[104] Knight, P. A. (2008). The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275.

[105] Komiya, H. (1988). Elementary proof for sion's minimax theorem. *Kodai mathematical journal*, 11(1):5–7.

[106] Kosowsky, J. and Yuille, A. L. (1994). The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural networks*, 7(3):477–490.

[107] Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval Research Logistics (NRL)*, 2(1-2):83–97.

[108] Kuhn, H. W. and Tucker, A. W. (2014). Nonlinear programming. In *Traces and emergence of nonlinear programming*, pages 247–258. Springer.

[109] Larson, P. B. (2009). The filter dichotomy and medial limits. *Journal of Mathematical Logic*, 9(02):159–165.

[110] Léonard, C. (2012). From the schrödinger problem to the monge–kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920.

[111] Lévy, B. (2015). A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693–1715.

[112] Lewis, A. S. and Overton, M. L. (2013). Nonsmooth optimization via quasi-newton methods. *Mathematical Programming*, 141(1-2):135–163.

[113] Lim, T. (2014). Optimal martingale transport between radially symmetric marginals in general dimensions. *arXiv preprint arXiv:1412.3530*.

[114] Lim, T. (2016). Multi-martingale optimal transport. *arXiv preprint arXiv:1611.01496*.

[115] Mandelbrot, B. B. and Hudson, R. L. (2010). *The (mis) behaviour of markets: a fractal view of risk, ruin and reward*. Profile books.

[116] Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437.

[117] McCallum, D. and Avis, D. (1979). A linear algorithm for finding the convex hull of a simple polygon. *Information Processing Letters*, 9(5):201–206.

[118] Mérigot, Q. (2011). A multiscale approach to optimal transport. In *Computer Graphics Forum*, volume 30, pages 1583–1592. Wiley Online Library.

[119] Meyer, P.-A. (1973). Limites médiales d'après mokobodzki. *Séminaire de Probabilités de Strasbourg*, 7:198–204.

[120] Mokobodzki, G. (1967). Ultrafiltres rapides sur $\mathbb{N}$. construction d'une densité relative de deux potentiels comparables. *Séminaire Brelot-Choquet-Deny. Théorie du potentiel*, 12:1–22.

[121] Monge, G. (1781). *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale.

[122] Monroe, I. (1972). On embedding right continuous martingales in brownian motion. *The Annals of Mathematical Statistics*, pages 1293–1311.

[123] Neufeld, A., Nutz, M., et al. (2013). Superreplication under volatility uncertainty for measurable claims. *Electronic journal of probability*, 18.

[124] Nutz, M., Stebegg, F., and Tan, X. (2017). Multiperiod martingale transport. *arXiv preprint arXiv:1703.10588*.

[125] Oberman, A. M. and Ruan, Y. (2015). An efficient linear programming method for optimal transportation. *arXiv preprint arXiv:1509.03668*.

[126] Obłój, J. et al. (2004). The skorokhod embedding problem and its offspring. *Probability Surveys*, 1:321–392.

[127] Oblój, J. and Siorpaes, P. (2017). Structure of martingale transports in finite dimensions.

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

[129] Peyré, G., Cuturi, M., et al. (2017). Computational optimal transport. Technical report.

[130] Possamaï, D., Royer, G., Touzi, N., et al. (2013). On the robust superhedging of measurable claims. *Electronic Communications in Probability*, 18.

[131] Rabin, J. and Papadakis, N. (2015). Convex color image segmentation with optimal transport distances. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 256–269. Springer.

[132] Rockafellar, R. T. (2015). *Convex analysis*. Princeton university press.

[133] Roos, C. (1990). An exponential example for terlaky's pivoting rule for the criss-cross simplex method. *Mathematical Programming*, 46(1-3):79–84.

[134] Rüschendorf, L. (1991). Fréchet-bounds and their applications. *Advances in probability distributions with given marginals*, pages 151–187.

[135] Rüschendorf, L. and Rachev, S. T. (1990). A characterization of random variables with minimum l2-distance. *Journal of Multivariate Analysis*, 32(1):48–54.

[136] Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, pages 99–102.

[137] Schmitzer, B. (2016). Stabilized sparse scaling algorithms for entropy regularized transport problems. *arXiv preprint arXiv:1610.06519*.

[138] Schoutens, W. (2003). Book series.

[139] Shafarevich, I. R. (2013). *Basic Algebraic Geometry 1: Varieties in Projective Space*. Springer Science & Business Media.

[140] Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348.

[141] Smale, S. (1983). On the average number of steps of the simplex method of linear programming. *Mathematical programming*, 27(3):241–262.

[142] Solomon, J., De Goes, F., Peyré, G., Cuturi, M., Butscher, A., Nguyen, A., Du, T., and Guibas, L. (2015). Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 34(4):66.

[143] Soner, H. M., Touzi, N., Zhang, J., et al. (2013). Dual formulation of second order target problems. *The Annals of Applied Probability*, 23(1):308–347.

[144] Spielman, D. A. and Teng, S.-H. (2004). Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463.

[145] Stebegg, F. (2014). Model-independent pricing of asian options via optimal martingale transport. *arXiv preprint arXiv:1412.1429*.

[146] Strassen, V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439.

[147] Sudakov, V. N. (1979). *Geometric problems in the theory of infinite-dimensional probability distributions*. Number 141. American Mathematical Soc.

[148] Tan, X., Touzi, N., et al. (2013). Optimal transportation under controlled stochastic dynamics. *The annals of probability*, 41(5):3201–3240.

[149] t.b. (https://math.stackexchange.com/users/5363/t b). Medial limit of mokobodzki (case of banach limit). Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/54562 (version: 2017-04-13).

[150] Thibault, A., Chizat, L., Dossal, C., and Papadakis, N. (2017). Overrelaxed sinkhorn-knopp algorithm for regularized optimal transport. *arXiv preprint arXiv:1711.01851*.

[151] Thorpe, M., Park, S., Kolouri, S., Rohde, G. K., and Slepčev, D. (2017). A transportation l^ p lp distance for signal analysis. *Journal of Mathematical Imaging and Vision*, 59(2):187–210.

[152] Trudinger, N. S. and Wang, X.-J. (2001). On the monge mass transfer problem. *Calculus of Variations and Partial Differential Equations*, 13(1):19–31.

[153] Trudinger, N. S. and Wang, X.-J. (2006). On the second boundary value problem for monge-ampere type equations and optimal transportation. *arXiv preprint math/0601086*.

[154] Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical review*, 36(5):823.

[155] Vaserstein, L. N. (1969). Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72.

[156] Villani, C. (2003). *Topics in optimal transportation*. Number 58. American Mathematical Soc.

[157] Villani, C. (2008). *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

[158] Wagner, D. H. (1977). Survey of measurable selection theorems. *SIAM Journal on Control and Optimization*, 15(5):859–903.

[159] Wright, S. and Nocedal, J. (1999). Numerical optimization. *Springer Science*, 35(67-68):7.

[160] Zaev, D. A. (2015). On the monge–kantorovich problem with additional linear constraints. *Mathematical Notes*, 98(5-6):725–741.

université PARIS-SACLAY | École doctorale de mathématiques Hadamard (EDMH)

Titre : Transport optimal de martingale multidimensionnel

Mots clés : transport optimal martingale, composantes irréductibles, dualité, structure locale, numérique, finance robuste, risque de modèle, régularisation entropique.

Résumé : Nous étudions dans cette thèse divers aspects du transport optimal martingale en dimension plus grande que un, de la dualité à la structure locale, puis nous proposons finalement des méthodes d'approximation numérique.
On prouve d'abord l'existence de composantes irréductibles intrinsèques aux transports martingales entre deux mesures données, ainsi que la canonicité de ces composantes. Nous avons ensuite prouvé un résultat de dualité pour le transport optimal martingale en dimension quelconque, la dualité point par point n'est plus vraie mais une forme de dualité quasi-sûre est démontrée. Cette dualité permet de démontrer la possibilité de décomposer le transport optimal quasi-sûre en une série de sous-problèmes de transports optimaux point par point sur chaque composante irréductible. On utilise enfin cette dualité pour démontrer un principe de monotonie martingale, analogue au célèbre principe de monotonie du transport optimal classique. Nous étudions ensuite la structure locale des transports optimaux, déduite de considérations différentielles. On obtient ainsi une caractérisation de cette structure en utilisant des outils de géométrie algébrique réelle. On en déduit la structure des transports optimaux martingales dans le cas des coûts puissances de la norme euclidienne, ce qui permet de résoudre une conjecture qui date de 2015. Finalement, nous avons comparé les méthodes numériques existantes et proposé une nouvelle méthode qui s'avère plus efficace et permet de traiter un problème intrinsèque de la contrainte martingale qu'est le défaut d'ordre convexe. On donne également des techniques pour gérer en pratique les problèmes numériques.

Title : Multi-dimensional martingale optimal transport

Keywords : martingale optimal transport, irreducible components, duality, local structure, numerics, robust finance, model risk, entropic regularization.

Abstract : In this thesis, we study various aspects of martingale optimal transport in dimension greater than one, from duality to local structure, and finally we propose numerical approximation methods.
We first prove the existence of irreducible intrinsic components to martingal transport between two given measurements, as well as the canonicity of these components. We have then proved a duality result for optimal martingale transport in any dimension, point-by-point duality is no longer true but a form of quasi-safe duality is demonstrated. This duality makes it possible to demonstrate the possibility of decomposing the quasi-safe optimal transport into a series of optimal transport subproblems point by point on each irreducible component. Finally, this duality is used to demonstrate a principle of martingale monotony, analogous to the famous monotonic principle of classical optimal transport. We then study the local structure of optimal transport, deduced from differential considerations. We thus obtain a characterization of this structure using tools of real algebraic geometry. We deduce the optimal martingal transport structure in the case of the power costs of the Euclidean norm, which makes it possible to solve a conjecture that dates from 2015. Finally, we compared the existing numerical methods and proposed a new method which proves more efficient and allows to treat an intrinsic problem of the martingale constraint which is the defect of convex order. Techniques are also provided to manage digital problems in practice.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France