On the Stability of the Martingale Optimal Transport Problem

William Margheriti

UNIVERSITÉ PARIS-EST

École doctorale : MATHÉMATIQUES ET SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

Thèse de doctorat

Spécialité : Mathématiques

présentée par

William MARGHERITI

Sur la stabilité du problème de transport optimal martingale

Thèse dirigée par Benjamin JOURDAIN
et préparée au CERMICS, École des Ponts ParisTech

Soutenue le 17 décembre 2020 devant le jury composé de :

Jean-François DELMAS	École des Ponts ParisTech	*Président*
Nicolas JUILLET	Université de Strasbourg	*Rapporteur*
Nizar TOUZI	École Polytechnique	*Rapporteur*
Virginie EHRLACHER	École des Ponts ParisTech	*Examinatrice*
Nathael GOZLAN	Université de Paris	*Examinateur*
Sébastien ROLAND	Société Générale	*Examinateur*
Benjamin JOURDAIN	École des Ponts ParisTech	*Directeur de thèse*

Remerciements (Acknowledgements)

Le doctorat en mathématiques que j'ai l'honneur de recevoir m'offre un privilège exceptionnel, celui de pouvoir faire de ma passion mon métier. À qui dois-je une telle chance ? Ne vous fiez pas à la page de garde. Elle prend un malin plaisir à vous induire en erreur. Elle présente mon nom en grand et en gras pour faire croire au lecteur que je ne dois l'accomplissement de cette thèse qu'à moi-même. Les quelques paragraphes qui suivent sont l'occasion de rectifier la vérité. Ce travail que j'ai réalisé n'aurait pas été possible sans l'aide et le soutien de nombreuses personnes, que je me dois de reconnaître et remercier.

Bien sûr mes premiers remerciements vont à mon directeur de thèse Benjamin Jourdain. Il m'a d'abord fait découvrir la théorie des probabilités en première année à l'École des Ponts, puis initié au calcul stochastique l'année suivante, avec une pédagogie extraordinaire. Je le connaissais comme excellent professeur, puis j'ai eu l'occasion de le découvrir comme éminent chercheur lorsqu'il a encadré ma thèse de doctorat. Au cours de ces trois années, il a toujours été disponible lorsque j'en avais besoin, m'a fait confiance et a fait preuve de beaucoup de patience. Il m'a très souvent impressionné par ses talents de mathématicien, et n'a jamais hésité à partager ses découvertes et son mérite avec moi. Merci pour ces trois années exceptionnellement riches en enseignement.

Je tiens aussi à remercier sincèrement Nizar Touzi et Nicolas Juillet d'avoir accepté d'être rapporteurs et donc d'avoir pris le temps de relire en détail mon mémoire. Je suis très heureux que mes travaux les aient intéressés, et les remercie pour leurs retours. Je remercie Jean-François Delmas d'avoir accepté de présider le jury de ma thèse, mais aussi de m'avoir initié aux processus stochastiques en deuxième année à l'École des Ponts, et pour les discussions fructueuses que nous avons eues. Merci à Virginie Ehrlacher d'avoir accépté d'être membre du jury, et de m'avoir enseigné les fondamentaux de la théorie de l'intégration de Lebesgue en première année à l'École des Ponts. Enfin, je remercie Nathael Gozlan et Sébastien Roland de m'avoir fait l'honneur de leur présence dans le jury.

Je suis très reconnaissant envers la “Chaire Risques Financiers” de la Fondation du Risque pour avoir financé mes trois années de thèse. Merci aussi à l'Inria de m'avoir accueilli dans l'équipe Mathrisk, et bien sûr au CERMICS de m'avoir accueilli dans son laboratoire.

Ich danke Mathias Beiglböck herzlich für die Einladung an die Universität Wien zu einer Arbeitssitzung mit seinem Doktoranden Gudmund Pammer. Es war der Beginn einer schönen Zusammenarbeit, bei der ich viel gelernt habe. Mathias' exzellenter Ruf ist unbestritten. Was Gudmund betrifft, so habe ich jemanden kennengelernt, der ebenso intelligent wie freundlich ist. Er hat mich während dieser Zusammenarbeit sehr beeindruckt, und ich sage ihm eine glänzende Forschungskarriere voraus. Ich möchte ihm herzlich danken und wünsche ihm alles Gute bei seiner Forschung sowie persönlich.

Je remercie bien sûr tous les membres de ma famille pour leur soutien indéfectible depuis toujours. Ils n'ont jamais cessé un seul instant de croire en moi. Étant essentiellement des personnes de lettres, ils peuvent croire que leur apport dans cette thèse de mathématiques est plutôt réduit. Ils ont pourtant bien plus contribué que ce qu'ils n'imaginent. Évidemment je ne peux pas citer tout ce qu'ils ont fait pour moi depuis ma naissance, donc je n'évoquerai pour chacun d'eux qu'un seul élément ayant eu un impact significatif sur mon parcours. Merci d'abord à mon père d'avoir éveillé ma curiosité scientifique lorsque j'étais petit en me parlant notamment d'astronomie et d'étymologie. Merci à ma mère d'avoir toujours veillé à ce que je m'exprime en bon français. Merci à ma soeur de m'avoir appris à lire (rien que ça !). Enfin, merci à Nuria d'être une partenaire si fiable et fidèle, toujours à mes côtés pour aller de l'avant.

Évidemment je remercie aussi l'ensemble de mes professeurs de prépa du collège Stanislas, qui ont eu un rôle fondamental dans mon parcours. Il y a d'abord Raphaël Shabetai, qui m'a énormément appris lors de mes deux années en Maths spé (oui, deux...), et qui m'a fait confiance plusieurs années de suite pour être interrogateur de khôlles dans sa classe. Sa bienveillance dont il a toujours fait preuve à mon égard a été d'un grand soutien. Merci à Christine Leygnac pour son enseignement en physique-chimie et toutes les discussions intéressantes que nous avons eues, ainsi que Charlotte Audiard, dont l'enseignement a très fortement contribué à la diminution des fautes d'anglais dans ce mémoire. Merci à Alain Camanes, professeur en Maths sup, qui m'a lui aussi fait confiance pour être interrogateur de khôlles dans sa classe, ainsi que Louis Candelon pour son enseignement en physique-chimie.

Je remercie aussi mes professeurs du lycée Gerson, qui m'ont énormément apporté : M. Ramassamy, Mme Roblot, M. Guégan, Mme Tomalino, M. Schmidt-Preuss, Mme Picot, M. Herault, Mme Chenitzer, ainsi que M. Person. C'est à cette période que j'ai effectué mon année de Préparation Militaire Marine (PMM), sous le commandement du Capitaine de Corvette (R) Guillaume Wattinne. Je remercie tous les instructeurs de la PMM Richelieu pour cette période unique et exceptionnellement formatrice, lors de laquelle j'ai appris tellement de choses. Je les remercie en particulier de m'avoir fait l'immense honneur de me permettre de défiler le 14 juillet 2010 sur les Champs-Élysées en tant que matelot de réserve.

Je remercie l'ensemble des chercheurs du CERMICS pour l'ambiance agréable dont j'ai profité ces trois dernières années, en particulier Aurélien Alfonsi et Benjamin Jourdain pour m'avoir respectivement confié le cours de probabilité de première année à l'École des Ponts ainsi que les TD de calcul stochastique appliqué à la finance de deuxième année à l'École des Ponts. Bien sûr une mention spéciale à Isabelle Simunic, secrétaire générale du CERMICS, pour son travail exceptionnel, et sans qui le laboratoire serait bien vide. Merci aussi à Vincent Lefieux de m'avoir confié le cours de statistiques de deuxième année à l'École des Ponts.

Enfin, merci à l'ensemble de mes amis (par ordre alphabétique pour ne pas faire de jaloux) : l'ensemble des doctorants du CERMICS, Oumaima Bencheikh, Thomas Bittar, Adel Cherchali, Victor Cohen, Rafaël Coyaud, Guillaume Dalle, Sébastien Deschamps, Dylan Dronnier, Maël Forcier, Clément Henin, Ezéchiel Kahn, Sofiane Martel, Thomas Martin, Sophian Mehalla, Mouad Ramil, Inass Sekkat, Adrien Touboul, Benoît Tran, Cyrille Vessaire, Alexandre Zhou ; le meilleur stagiaire du CERMICS, Mohamed Detsouli ; mes amis du Master 2 Probabilités et modèles aléatoires, Quentin Cormier, Marie Garin, Angelo Rosello, Clare Wallace ; ma collègue stagiaire de Natixis, Élise Pouille ; mes amis de l'École des Ponts, Samy Benbelkacem, Mohamed Boubakour, Thomas Daniel, Geoffrey Daniel, Ousmane Galokho, Maxime Heinrich, Cefax Ouamara, Eïda Ouchtar, Martin Troussard, Youcef Zemmouri ; mes amis de l'Imperical College London, Astrid Delaporte-Sprengers, Tamanna Saghir, Émilie Tarouilly ; mes amis de prépa, Juliette Caucheteux, Quentin Desrousseaux ; mes amis du lycée, Balthazar Dadvisard, Abel Quérou, Henri Mirande ; mon ami du collège Rémi Langlois, mon ami d'école primaire Guillaume Huard, mon ami Fabien Zapata, et enfin mon amie de crèche (!), Margaux Schmit.

Titre : Sur la stabilité du problème de transport optimal martingale

Résumé : Cette thèse est motivée par l'étude de la stabilité du problème de transport optimal martingale, et s'articule naturellement autour de deux parties. Dans la première partie, nous exhibons une nouvelle famille de couplages martingale entre deux mesures de probabilités unidimensionnelles μ et ν comparables dans l'ordre convexe. Cette famille contient en particulier le couplage martingale transformée inverse, qui est explicite en termes des fonctions quantiles des marginales. L'intégrale $\mathcal{M}_1(\mu,\nu)$ de $|x-y|$ contre chacun de ces couplages est majorée par le double de la distance de Wasserstein $\mathcal{W}_1(\mu,\nu)$ entre μ et ν. Nous montrons une inégalité similaire lorsque $|x-y|$ et $\mathcal{W}_1$ sont respectivement remplacés par $|x-y|^\rho$ et le produit de $\mathcal{W}_\rho$ par le moment centré d'ordre ρ de la seconde marginale élevé à l'exposant $\rho-1$, pour $\rho\in[1,+\infty[$ quelconque. Nous étudions ensuite la généralisation de cette nouvelle inégalité de stabilité à la dimension supérieure. Enfin, nous établissons une forte connexion entre notre nouvelle famille de couplages martingale et la projection d'un couplage entre deux marginales données comparables dans l'ordre convexe sur l'ensemble des couplages martingale entre ces mêmes marginales. Cette dernière projection est prise par rapport à la distance de Wasserstein adaptée, qui majore la distance de Wasserstein usuelle et induit donc une topologie plus fine et mieux adaptée pour la modélisation financière, puisqu'elle prend en compte la structure temporelle des martingales. Dans la seconde partie, nous prouvons que tout couplage martingale dont les marginales sont approchées par des mesures de probabilité comparables dans l'ordre convexe peut être lui-même approché par des couplages martingale au sens de la distance de Wasserstein adaptée. Nous traitons ensuite d'applications variées de ce résultat. En particulier, nous renforçons un résultat de stabilité portant sur le problème de transport optimal faible et établissons un résultat de stabilité pour le problème de transport optimal martingale faible. Nous en déduisons la stabilité par rapport aux marginales du prix de sur-réplication de contrats à termes sur le VIX.

Mots-clefs : Transport optimal martingale, Couplages martingale, Distance de Wasserstein, Ordre convexe, Distance de Wasserstein adaptée, Finance robuste, Transport faible, Stabilité

Title: On the Stability of the Martingale Optimal Transport Problem

Abstract: This thesis is motivated by the study of the stability of the Martingale Optimal Transport problem, and is naturally structured around two parts. In the first part, we exhibit a new family of martingale couplings between two one-dimensional probability measures μ and ν in the convex order. This family contains in particular the inverse transform martingale coupling which is explicit in terms of the quantile functions of these marginal densities. The integral $\mathcal{M}_1(\mu,\nu)$ of $|x-y|$ with respect to each of these couplings is smaller than or equal to twice the Wasserstein distance $\mathcal{W}_1(\mu,\nu)$ between μ and ν. We show that a similar inequality holds when replacing $|x-y|$ and $\mathcal{W}_1$ respectively with $|x-y|^\rho$ and the product of $\mathcal{W}_\rho$ times the centred ρ-th moment of the second marginal to the power $\rho-1$, for any $\rho\in[1,+\infty)$. We then study the generalisation of this new stability inequality to higher dimensions. Last, we establish a strong connection between our new family of martingale couplings and the projection of a coupling between two given marginals in the convex order onto the set of martingale couplings between the same marginals. The latter projection is taken with respect to the adapted Wasserstein distance, which is greater than or equal to the usual Wasserstein distance and therefore induces a finer topology, which is more suitable to financial modelisation since it takes into account the temporal structure of martingales. In the second part, we prove that any martingale coupling whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance. We then discuss various applications of this result. In particular, we strengthen a stability result on the Weak Optimal Transport problem and establish a stability result on the Weak Martingale Optimal Transport problem. We deduce the stability with respect to the marginals of the superreplication price of VIX futures.

Keywords: Martingale Optimal Transport, Martingale couplings, Wasserstein distance, Convex order, Adapted Wasserstein distance, Robust finance, Weak transport, Stability

Contents

Chapter 1

Introduction (in French)

L'apport de la théorie mathématique, et en particulier le développement de la théorie des probabilités, a permis depuis plusieurs décennies de développer l'industrie des produits dérivés financiers, et donc les marchés financiers, de manière considérable. Certes, il faut reconnaître que l'histoire de la finance de marché remonte bien avant celle des mathématiques : des historiens rapportent l'existence de contrats à terme en Égypte sur les matières agricoles en -1400 av. J.-C., sous le règne du pharaon Akhénaton, ou encore l'existence de contrats d'option d'achat sur les olives de Grèce en -700 av. J.-C., dont l'émetteur serait Thalès lui-même. Bien que ce dernier ait employé ses connaissances en mathématiques et astronomie pour prévoir les récoltes des années suivantes, on ne peut bien sûr pas affirmer qu'il s'agisse déjà là du point de départ des mathématiques financières. Il faudra attendre le tout début du XX^e^ siècle pour que les grands esprits, Louis Bachelier en tête dans sa thèse intitulée « théorie de la spéculation » [13], reconnaissent le besoin fondamental du développement d'outils mathématiques sophistiqués afin d'avoir une emprise sur les produits dérivés. Il emprunte ainsi au botaniste Robert Brown le célèbre mouvement brownien $(B_t)_{t\in\mathbb{R}_+}$, ce processus fondamental et incontournable, dont la structure particulièrement erratique restitue les allures des cours boursiers, qui est à l'origine de tous les modèles connus utilisés par l'ensemble des acteurs financiers de la planète.

Supposons qu'une banque vende un produit dérivé, typiquement une option de vente, dite *put*, ou une option d'achat, dite *call*. Ce produit, proposé à la date 0, rapporte à celui qui l'achète un profit P, dit *payoff*, à la date T, dite *maturité*. Ce payoff dépend de la valeur S_T en T d'un actif risqué dit *sous-jacent* dont on ne connaît la valeur qu'en 0. La banque est alors confrontée à deux problématiques fondamentales : l'*évaluation* (pricing) et la *couverture* (hedging). Autrement dit, à quel prix c doit-elle vendre son produit, et comment doit-elle investir le montant c sur le marché pour répliquer le payoff, c'est-à-dire obtenir un montant P à la date T afin de pouvoir honorer son contrat avec le client. En 1973, Black, Scholes et Merton [40, 138] révolutionnent le domaine et marquent le point de départ des mathématiques financières modernes en développant un modèle quantifié qui répond précisément à ces deux problématiques. Ils démontrent que si la loi de l'actif risqué est celle d'un mouvement brownien géométrique, ou de manière équivalente s'il suit une dynamique $dS_t = \mu S_t\, dt + \sigma S_t\, dB_t$, alors il est possible, en achetant ou vendant à chaque instant compris entre 0 et T une quantité bien définie de sous-jacent et d'actif sans risque, de répliquer le payoff P. Cette stratégie de réplication, ou de couverture, nécessite cependant un apport initial c en 0. Le prix du produit dérivé est alors défini comme la quantité c à détenir en 0 pour amorcer la stratégie de réplication. Cette approche aboutit à la fameuse formule de Black-Scholes qui fournit une expression explicite du prix d'un call ou d'un put. Elle repose notamment sur l'hypothèse d'*absence d'opportunité d'arbitrage*, selon laquelle il est impossible de faire un profit sans prendre de risques. Il s'agit en pratique d'une hypothèse très réaliste, dans la mesure où il existe des arbitragistes dont le métier consiste à repérer de telles opportunités et à les saisir rapidement. Ainsi, tout écart entre le prix théorique d'un produit et son cours côté sur le marché se dissipe rapidement.

Cependant, le modèle de Black-Scholes-Merton s'est avéré insatisfaisant pour rendre compte de ce que l'on observe sur les marchés. La distribution gaussienne des log-rendements, la constance du taux d'intérêt et celle de la volatilité sont autant d'hypothèses simplificatrices qui font de ce modèle surtout une source d'inspiration sur laquelle vont se fonder des modèles plus réalistes. Parmi les alternatives les plus connues, on peut citer le modèle d'Ornstein-Uhlenbeck [188], les modèles à volatilité stochastique [107], à volatilité locale [74, 75], ou encore le modèle de Heston avec taux d'intérêt stochastique [94]. On notera que pour modéliser les sauts que peuvent parfois présenter les cours des actifs à la suite par exemple de défauts de paiement, faillites ou crises majeures, le mouvement brownien dont les trajectoires sont continues devra parfois céder sa place à des processus de Lévy plus généraux [125]. Pour une étude détaillée du sujet, on pourra consulter [171, 172, 128, 79, 155].

Nombre de modèles ont donc émergé dans un contexte de dérégulation, de déréglementation, d'ouverture des marchés financiers et de grandes volatilités des actifs échangés. Une des conséquences de la variabilité des modèles est naturellement la variabilité des conclusions. En effet, deux modèles différents impliquent des prévisions d'évolution du marché différentes, et aboutissent donc à des calculs de prix différents pour un même produit dérivé. Or dans notre contexte actuel d'après crise de 2007-2008, nous nous devons d'accorder une importance particulière au calcul du risque. Mais la notion même de risque n'a pas toujours une définition claire. Une approche naturelle serait par exemple de considérer une famille de modèles $(\mathcal{M}_\theta)_{\theta\in\Theta}$ paramétrée par un ensemble Θ, et de calculer l'erreur type induite par le passage du modèle $\mathcal{M}_\theta$ au modèle $\mathcal{M}_{\theta'}$ pour $\theta, \theta' \in \Theta$ [70, 181, 143, 154]. Cette méthode peut être jugée insatisfaisante, puisque le risque calculé dépend de la famille de modèles choisie, et dépend donc d'une forme de modélisation. Très récemment, une nouvelle approche a été développée pour quantifier le risque de manière absolument indépendante du modèle. Plus précisément, il s'agit de déterminer les bornes inférieure et supérieure de l'intervalle des possibilités de tout prix issu d'un modèle quelconque sous l'hypothèse d'absence d'opportunité d'arbitrage. De telles bornes s'expriment comme un problème dit de *transport optimal martingale*. La résolution d'un tel problème représente donc un enjeu important en mathématiques financières, et motivera tout ce qui suit.

Nous allons donc résumer la littérature sur la théorie du transport optimal martingale, après avoir abordé celle dont elle est originaire, à savoir le transport optimal classique. Nous expliquerons ensuite l'apport de cette thèse, qui se distingue naturellement en deux parties. La première partie, motivée par la preuve de la stabilité du problème de transport optimal martingale par rapport à ses marginales, voit l'apparition d'une nouvelle famille de couplages martingale en dimension 1, qui satisfont à une inégalité de stabilité. On explore ensuite la généralisation de cette inégalité à la dimension supérieure. Dans la seconde partie, nous prouvons qu'en dimension 1, tout couplage martingale entre des limites de suites de marginales peut être approché par des couplages martingale entre ces marginales, au sens d'une distance Wasserstein plus adaptée à la structure des martingales que la distance de Wasserstein classique. La démonstration de ce résultat est divisée en plusieurs étapes, dont l'une repose sur l'inégalité de stabilité montrée dans la première partie. De ce théorème, nous déduisons des applications sur la stabilité de problèmes de transport optimal faible et transport optimal martingale faible.

1.1 Le transport optimal

1.1.1 Définition

En 1781, le mathématicien français Gaspard Monge publie un mémoire sur un problème d'ingénierie de son époque [140, 10, 11, 87]. Il s'agit de transporter un volume de terre (le *déblai*) de son lieu initial vers un espace qu'il doit occuper (le *remblai*), le tout pour un coût minimal. Supposons pour commencer que le déblai se situe en des points $x_1, \cdots, x_n \in \mathbb{R}$ et le remblai en des points $y_1, \cdots, y_m \in \mathbb{R}$, où $n, m \in \mathbb{N}^*$. Pour tout $i \in \{1, \cdots, n\}$, le point x_i contient une fraction $\mu_i \in]0, 1]$ du volume totale initial, et pour tout $j \in \{1, \cdots, m\}$, il nous faut transporter en y_j une fraction $\nu_j \in]0, 1]$ du volume total de terre. De plus, le transport d'une unité de masse du point x_i vers le point y_j a un coût $c(x_i, y_j) \geq 0$. Le problème de Monge consiste alors à déterminer l'application $T : \{x_1, \cdots, x_n\} \to \{y_1, \cdots, y_m\}$ pour laquelle le coût de transport total

$$\sum_{i=1}^{n} c(x_i, T(x_i))\mu_i \tag{1.1.1}$$

est minimal, en respectant la contrainte que l'image du déblai par l'application T corresponde bien au remblai imposé, i.e.

$$\forall j \in \{1, \cdots, m\}, \quad \nu_j = \sum_{i=1}^{n} \mu_i \mathbb{1}_{\{T(x_i)=y_j\}}. \tag{1.1.2}$$

En s'autorisant la considération de molécules de terre de taille infinitésimale, Monge propose une formulation plus générale de ce problème. Supposons que sur l'intervalle infinitésimal $[x, x + dx]$ se trouve une fraction de terre $\mu(x)\, dx$, et que l'espace à remplir en $[y, y + dy]$ doive contenir une fraction $\nu(y)\, dy$ du volume. De plus, envoyer la masse de terre comprise entre x et $x + dx$ via une application mesurable $T : \mathbb{R} \to \mathbb{R}$ a un coût infinitésimal $c(x, T(x))\mu(x)\, dx$. Dans ce cas, (1.1.1) s'écrit $\int_{\mathbb{R}} c(x, T(x))\mu(x)\, dx$, et (1.1.2) s'écrit $T_\sharp\mu = \nu$, où $\sharp$ désigne l'opération de mesure image. En considérant des espaces mesurés quelconques, on peut ainsi formuler le problème de Monge dans sa généralité : étant données deux mesures de probabilité μ et ν définies respectivement sur E et F, et une application de coût $c : E \times F \to [0, +\infty]$, déterminer

$$C(\mu, \nu) = \inf_{\substack{T:E\to F \text{ mesurable} \\ T_\sharp\mu=\nu}} \int_E c(x, T(x))\, \mu(dx). \tag{1.1.3}$$

Ce problème d'optimisation est la première formulation historique d'un problème de *transport optimal*. Son succès s'explique notamment par le fait que la portée de ce problème dépasse largement le cadre spécifique du transport de terre, et s'adapte en fait à un champ extrêmement vaste d'applications. En effet, avec les notations de (1.1.1), on peut par exemple imaginer au point x_i la présence d'un entrepôt contenant une proportion μ_i de masques, que l'hôpital situé en y_j doit recevoir une proportion ν_j de masques, et qu'il y a un coût donné $c(x_i, y_j)$ par masque pour transporter les masques de l'entrepôt en x_i à l'hôpital en y_j. La résolution du problème de Monge dans ce cadre permet alors de déterminer à quel hôpital un entrepôt doit envoyer son stock de masques pour répondre exactement à la demande tout en faisant un maximum d'économies.

On notera cependant que le problème de Monge n'autorise pas un entrepôt à fournir des masques à plusieurs hôpitaux : l'entrepôt en x_i ne peut founir que l'hôpital qui se trouve en $T(x_i)$. De manière générale, l'optimum du problème de Monge peut être insatisfaisant lorsque l'on cherche à résoudre un problème où cela a du sens de transporter une masse en x non pas nécessairement en un seul point y, mais potentiellement diffusée à plusieurs endroits. Le problème de Monge peut aussi ne pas avoir de solution. Par exemple dans le cas $E = F = \mathbb{R}$, il est impossible de trouver une application $T : \mathbb{R} \to \mathbb{R}$ dont l'image de $\mu = \delta_0$ soit la mesure $\nu = \frac{1}{2}(\delta_0 + \delta_1)$. Plus généralement, l'impossibilité demeure si $n < m$ dans (1.1.2). C'est ainsi qu'en 1942, le mathématicien russe Leonid Kantorovich redécouvre et modernise le problème de transport optimal en l'interprétant comme un problème de couplage probabiliste optimal [114], et fera le lien avec le problème de Monge 6 ans plus tard [115]. Cela donne naissance au problème dit de Monge-Kantorovich : étant données deux mesures de probabilité μ et ν définies respectivement sur E et F, et une application de coût $c : E \times F \to [0, +\infty]$, déterminer

$$C'(\mu, \nu) = \inf_{X\sim\mu,\ Y\sim\nu} \mathbb{E}[c(X, Y)]. \tag{1.1.4}$$

Notons que l'infimum de (1.1.4) se fait toujours sur un ensemble non vide puisqu'il est toujours possible de considérer deux variables aléatoires indépendantes suivant respectivement μ et ν. De plus, il est facile de montrer que le minimum existe sous des hypothèses peu exigentes de régularité sur l'application de coût c, typiquement de la semi-continuité inférieure. Pour toute application mesurable $T : E \to F$ telle que $T_\sharp\mu = \nu$ et pour toute variable aléatoire X de loi μ, la variable aléatoire $Y = T(X)$ a pour loi ν, ce qui montre que $C'(\mu, \nu) \leq C(\mu, \nu)$. Réciproquement, si le couple optimal (X, Y) qui minimise (1.1.4) est tel qu'il existe une application mesurable $T : E \to F$ satisfaisant $Y = T(X)$, alors $C(\mu, \nu) \leq C'(\mu, \nu)$, et ainsi les problèmes de Monge et de Monge-Kantorovich coïncident. Cette dernière condition, d'apparence restrictive, est en fait régulièrement satisfaite. On verra par exemple dans la section 1.1.4 ci-dessous que dans le cas $E = F = \mathbb{R}$, il est suffisant que μ n'ait pas d'atome.

Le mathématicien bulgare Svetlozar Rachev [160], élève de Kantorovich, donne un bon aperçu de la conscience des nombreux domaines d'application du problème de Monge-Kantorovich qu'avait déjà la communauté mathématique en 1985 : géométrie différentielle [192, 184], analyse fonctionnelle [116, 132], programmation linéaire en dimension infinie [189, 166, 159, 119], théorie des probabilités [80, 81, 72, 197], statistiques [196, 111], théorie de l'information [193, 91, 92], physique statistique [71, 52], théorie des système dynamiques [147] et théorie matricielle [100, 135, 146]. Motivé par l'étude de la mécanique des fluides, le mathématicien français Yann Brenier formule à la fin des années 1980 son célèbre théorème de décomposition polaire [47, 48], à savoir que sous des hypothèses assez générales, l'unique solution du problème de Monge-Kantorovich dans $\mathbb{R}^d$ pour le coût quadratique $c : (x, y) \mapsto |y - x|^2$ est de la forme $(X, \nabla\varphi(X))$, où φ est une fonction convexe déterminée de manière unique. D'après le raisonnement du paragraphe précédent, cette solution coïncide donc avec celle du problème de Monge. De plus, la fonction φ est alors la solution d'une équation de type Monge-Ampère, ce qui donne une nouvelle impulsion au transport optimal en attirant les mathématiciens intéressés par la résolution d'équations aux dérivées partielles. Depuis, nombre de chercheurs de domaines en apparence très éloignés du transport optimal cherchent à prouver une connexion avec ce dernier, comme les météorologues pour la résolution d'équations dites semi-géostrophiques [61, 58, 60, 59, 57]. Récemment, le transport optimal est devenu un outil incontournable de certaines branches du machine learning, comme l'analyse de formes [1, 158, 179, 86], l'adaptation de domaine [56], la classification multi-labels [83] ou encore les réseaux adverses génératifs qui produisent artificiellement des images au réalisme troublant [118, 12].

Tous les exemples cités font appel au problème de Monge-Kantorovich sous sa forme originale, c'est-à-dire celle donnée par (1.1.4). Le contexte qui nous intéresse ici est celui des mathématiques financières. Nous expliquons dans la section 1.2 ci-dessous en quoi le problème de transport optimal nous est utile, sous réserve de lui ajouter une contrainte dite de martingalité, conséquence de l'absence d'opportunité d'arbitrage. La résolution de ce problème modifié, appelé problème de *transport optimal martingale*, et son application en mathématiques financières, représentent notre motivation principale. Naturellement, beaucoup de résultats de la théorie du transport optimal martingale s'inspirent de la théorie du transport optimal classique. Pour une étude en profondeur de cette dernière, on pourra se référer à [161, 162, 7, 190, 9, 191, 8, 168]. Avant d'aborder le transport optimal martingale, nous résumons les idées essentielles de la théorie classique qui serviront de source d'inspiration pour l'extension au cas martingale.

1.1.2 Monotonie cyclique et dualité

La solution optimale à un problème de transport optimal ne peut avoir n'importe quelle structure. Pour le voir, reprenons l'exemple des entrepôts devant fournir des masques à des hôpitaux. Supposons que la solution optimale implique, entre autres, d'envoyer au moins un masque de l'entrepôt situé en x_i vers l'hôpital situé en y_i, pour $i \in \{1, \cdots, N\}$ où

$N \in \mathbb{N}\backslash\{0,1\}$. Redirigeons alors un masque venant de l'entrepôt situé en x_1 vers l'hôpital situé en y_2, puis un masque venant de x_2 vers y_3, etc., puis un masque de x_{N-1} vers y_N, et enfin un masque de x_N vers y_1. Pour tout $i \in \{1, \cdots, N\}$, l'opération de redirection du masque en x_i destiné en y_i et détourné en y_{i+1} induit une modification du coût de $c(x_i, y_{i+1}) - c(x_i, y_i)$, avec pour convention $y_{N+1} = y_1$. Puisque l'on a modifié la solution optimale, le coût global qui en résulte est nécessairement supérieur ou égal au coût optimal, ce qui implique que la modification globale $\sum_{i=1}^{N}(c(x_i, y_{i+1}) - c(x_i, y_i))$ est positive, soit

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1}).$$

Cette observation motive la définition suivante : on dira qu'un ensemble mesurable $\Gamma \subset E \times F$ est c-cycliquement monotone si

$$\forall N \in \mathbb{N}\backslash\{0,1\}, \quad \forall (x_1, y_1), \cdots, (x_N, y_N) \in \Gamma, \quad \sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N-1} c(x_i, y_{i+1}) + c(x_N, y_1),$$

et la loi d'un couple de variables aléatoires (X, Y) à valeurs dans $E \times F$ est dite c-cycliquement monotone s'il existe un ensemble c-cycliquement monotone Γ tel que $\mathbb{P}((X, Y) \in \Gamma) = 1$. Intuitivement, une distribution est c-cycliquement monotone si on ne peut trouver de cycle de redirection tel que décrit ci-dessus menant à un coût inférieur, et donc si le coût de transport associé à cette distribution ne peut être amélioré. Conformément à cette idée, on obtient le théorème suivant, dont une démonstration peut être trouvée dans [191]. On rappelle qu'un espace topologique est dit polonais si et seulement s'il est séparable et sa topologie est induite par une distance pour laquelle il est complet. Évidemment un tel espace est implicitement muni de la tribu borélienne.

Théorème 1.1.1. *Soient μ et ν deux mesures de probabilité définies respectivement sur les espaces polonais E et F et une application de coût $c : E \times F \to \mathbb{R}_+$ semi-continue inférieure telle que la quantité $C'(\mu, \nu)$ définie par* (1.1.4) *soit finie. Soient X et Y deux variables aléatoires de lois respectives μ et ν. Alors les deux assertions suivantes sont équivalentes :*

- *(i) (X, Y) est optimal pour le problème* (1.1.4) *;*
- *(ii) La loi de (X, Y) est c-cycliquement monotone.*

Ce théorème a pour implication fondamentale la stabilité du transport optimal par rapport à ses marginales sous des hypothèses peu restrictives. Pour simplifier, supposons l'application de coût $c : E \times F \to \mathbb{R}$ continue bornée. Pour tout $n \in \mathbb{N}$, soit μ_n une mesure de probabilité sur E convergeant étroitement vers μ, ν_n une mesure de probabilité sur F convergeant étroitement vers ν, et π_n une distribution optimale pour $C'(\mu_n, \nu_n)$. Alors à extraction d'une sous-suite près, $(\pi_n)_{n\in\mathbb{N}}$ converge étroitement vers une distribution π. D'une part, d'après le théorème de Portmanteau, $C'(\mu_n, \nu_n) = \int_{X\times Y} c(x, y)\, \pi_n(dx, dy)$ converge vers $\int_{X\times Y} c(x, y)\, \pi(dx, dy)$. D'autre part, on montre facilement la c-monotonie cyclique de π, et donc son optimalité pour $C'(\mu, \nu)$ en vertu du théorème 1.1.1. On en déduit que

$C'(\mu_n, \nu_n)$ converge vers $C'(\mu, \nu)$. On reporte à la section 1.1.5 les conséquences de cette implication sur la résolution numérique du problème de transport optimal.

Classiquement, le théorème 1.1.1 se prouve de pair avec un principe de dualité. Pour illustrer ce principe avec l'exemple des entrepôts devant fournir des masques à des hôpitaux, on adopte le point de vue d'un prestataire qui propose à l'ensemble hospitalier, composé des hôpitaux et de leurs entrepôts, de lui déléguer sa problématique de transport. Plus précisément, le prestataire propose à l'entrepôt en x_i d'acheter le masque à un prix unitaire $\psi(x_i)$, et de vendre à l'hôpital situé en y_j un masque au prix unitaire $\varphi(y_j)$. Ainsi, pour chaque masque transporté de x_i à y_j, le prestataire reçoit $\varphi(y_j) - \psi(x_i)$, payé par l'ensemble hospitalier, ce qui n'est bien sûr intéressant pour ce dernier que si cette valeur est inférieure au cout $c(x_i, y_j)$ qu'il aurait dépensé en le transportant lui-même. Naturellement, l'objectif du prestataire est de maximiser son gain global, à savoir $\sum_{j=1}^{p} \varphi(y_j)\nu_j - \sum_{i=1}^{n} \psi(x_i)\mu_i$, tout en proposant une prestation admissible, c'est-à-dire $\varphi(y_j) - \psi(x_i) \leq c(x_i, y_j)$ pour tout $(i,j) \in \{1, \cdots, n\} \times \{1, \cdots, p\}$. La problématique du prestataire est qualifiée de duale, étant donné qu'elle associe au problème de minimisation originel un problème de maximisation. En adaptant de manière évidente les quantités évoquées au cas de mesures de probabilité quelconques, on aboutit au *problème dual de Kantorovich* : étant données deux mesures de probabilité μ et ν définies respectivement sur E et F, et une application de coût $c : E \times F \to [0, +\infty]$, déterminer

$$D'(\mu, \nu) = \sup_{\substack{(\psi,\varphi) \in L^1(\mu) \times L^1(\nu), \\ \forall (x,y) \in E \times F,\ \varphi(y) - \psi(x) \leq c(x,y)}} \left(\int_F \varphi(y)\, \nu(dy) - \int_E \psi(x)\, \mu(dx) \right), \tag{1.1.5}$$

où pour toute mesure de probabilité η définie sur un espace mesurable H, $L^1(\eta)$ désigne l'ensemble des applications $\phi : H \to \mathbb{R}$ mesurables et η-intégrables, i.e. telles que $\int_H |\phi(z)|\, \eta(dz) < +\infty$. On peut alors montrer que le saut de dualité est nul.

Théorème 1.1.2 (Dualité de Kantorovich)**.** *Soient μ et ν deux mesures de probabilité définies respectivement sur les espaces polonais E et F et une application de coût $c : E \times F \to \mathbb{R}_+^*$ semi-continue inférieure. Alors*

$$C'(\mu, \nu) = D'(\mu, \nu).$$

Les théorèmes 1.1.1 et 1.1.2 se relient notamment par le fait suivant : si (X, Y) est optimal pour $C'(\mu, \nu)$ et (φ, ψ) est optimal pour $D'(\mu, \nu)$, alors $\Gamma = \{(x, y) \in E \times F \mid \varphi(y) - \psi(x) = c(x, y)\}$ est c-cyliquement monotone et $\mathbb{P}((X, Y) \in \Gamma) = 1$.

1.1.3 Les distances de Wasserstein

Dans le problème de Monge-Kantorovich (1.1.4) où E est un espace polonais égal à F, la quantité $C'(\mu, \nu)$ peut être vue comme une manière de quantifier l'écart entre les mesures de probabilité μ et ν. On remarque que pour tout $(x, y) \in E^2$, $C'(\delta_x, \delta_y) = c(x, y)$. Ainsi, pour que l'application C' vérifie effectivement les axiomes d'une distance sur l'ensemble des mesures de probabilité sur E, il est nécessaire que c soit elle-même une distance sur E.

Puisqu'une distance se doit d'être finie, on ne peut espérer que C' puisse mesurer autre chose que l'écart entre des mesures de probabilité μ sur E satisfaisant $C'(\mu, \delta_{x_0}) = \mathbb{E}[c(X, x_0)] < +\infty$, où x_0 est un élément quelconque de E. Dobrushin montre en 1970 [71] que ces conditions nécessaires sont suffisantes à faire de C' une distance. Grâce à l'inégalité de Minkowski, on obtient en fait une famille de distances paramétrée par un réel $\rho \geq 1$, de sorte que la finesse de la topologie induite croisse avec ρ. Plus formellement, soient d une distance sur un espace polonais E, $\mathcal{P}(E)$ l'ensemble des mesures de probabilité sur E, $\rho \in [1, +\infty[$ et $\mathcal{P}_\rho(E) \subset \mathcal{P}(E)$ l'ensemble des mesures de moment d'ordre ρ fini, i.e.

$$\begin{aligned}\mathcal{P}_\rho(E) &= \left\{\mu \in \mathcal{P}(E) \mid \exists x_0 \in E,\ \int_E d^\rho(x, x_0)\, \mu(dx) < +\infty\right\} \\ &= \left\{\mu \in \mathcal{P}(E) \mid \forall x_0 \in E,\ \int_E d^\rho(x, x_0)\, \mu(dx) < +\infty\right\}.\end{aligned}$$

Alors l'application

$$\mathcal{W}_\rho : \mathcal{P}_\rho(E) \times \mathcal{P}_\rho(E) \ni (\mu, \nu) \mapsto \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[d^\rho(X, Y)]^{\frac{1}{\rho}} \tag{1.1.6}$$

définit une distance sur $\mathcal{P}_\rho(E)$, qui porte différents noms selon les auteurs. À l'instar de Dobrushin, nous l'appelerons dorénavant *distance de Wasserstein* d'ordre ρ, d'après le nom du mathématicien russo-américain l'ayant introduite l'année d'avant [193].

Une simple application de l'inégalité de Hölder montre que pour tous $\rho, \rho' \in [1, +\infty[$ tels que $\rho \leq \rho'$, $\mathcal{P}_{\rho'}(E) \subset \mathcal{P}_\rho(E)$, et les distances de Wasserstein vérifient $\mathcal{W}_\rho \leq \mathcal{W}_{\rho'}$. De plus, toute convergence de mesures de probabilité au sens de $\mathcal{W}_\rho$ implique une convergence au sens de $\mathcal{W}_1$, qui est donc la plus faible. La convergence $\mathcal{W}_1$ demeure toutefois plus forte que la convergence étroite. Pour comparer ces topologies, on dispose d'un résultat qui accroît considérablement le confort d'utilisation de la convergence $\mathcal{W}_\rho$.

Théorème 1.1.3. *Soient (E, d) un espace métrique séparable complet, $x_0 \in E$, $\rho \in [1, +\infty[$, $\mu \in \mathcal{P}_\rho(E)$ et $(\mu_n)_{n\in\mathbb{N}} \in (\mathcal{P}_\rho(E))^{\mathbb{N}}$. Alors les assertions suivantes sont équivalentes :*

(i) $\mathcal{W}_\rho(\mu_n, \mu) \underset{n\to+\infty}{\longrightarrow} 0$;

(ii) $\mu_n \underset{n\to+\infty}{\longrightarrow} \mu$ *étroitement et* $\int_E d^\rho(x, x_0)\, \mu_n(dx) \underset{n\to+\infty}{\longrightarrow} \int_E d^\rho(x, x_0)\, \mu(dx)$;

(iii) Pour toute application continue $f : E \to \mathbb{R}$ à croissance au plus en exposant ρ, i.e. telle que $|f(x)| \leq K(1 + d^\rho(x, x_0))$ pour tout $x \in E$ et un certain $K \in \mathbb{R}_+$, on a

$$\int_E f(x)\, \mu_n(dx) \underset{n\to+\infty}{\longrightarrow} \int_E f(x)\, \mu(dx).$$

En plus de sa maniabilité, le théorème 1.1.3 permet par exemple de voir que dans le cas général, la notion de convergence $\mathcal{W}_\rho$ est impactée par le changement de distance sur E. En effet, la convergence $\mathcal{W}_\rho$ est souvent strictement plus forte que la convergence étroite. Or si l'on remplace d par la distance $\delta = \frac{d}{1+d}$ (ou $\min(1, d)$) qui lui est uniformément équivalente (i.e. l'application identité est uniformément continue de (E,d) dans (E,δ), et de (E,δ) dans (E,d)), le caractère borné de cette dernière et le point (iii) du théorème 1.1.3 montrent que la convergence $\mathcal{W}_\rho$ pour E muni de δ est équivalente à la convergence étroite. Notons quand même que si l'on remplace d par une distance d' qui lui est Lipschitz-équivalente (i.e. il existe $L, L' \in \mathbb{R}_+^*$ tels que $Ld \leq d' \leq L'd$), il est clair que la notion de convergence $\mathcal{W}_\rho$ ne change pas.

Du point de vue de la structure topologique, il est connu que pour (E,d) un espace métrique séparable complet, l'espace $(\mathcal{P}(E), d_P)$ est lui aussi séparable complet [38], où d_P désigne la distance de Prokhorov [156], qui induit la topologie de la convergence étroite. Cette propriété s'étend au contexte de la distance de Wasserstein : si (E,d) est séparable et complet, alors pour tout $\rho \in [1,+\infty[$, $(\mathcal{P}_\rho(E), \mathcal{W}_\rho)$ l'est aussi.

Enfin, puisque la distance de Wasserstein $\mathcal{W}_1$ est un cas particulier de problème de Monge-Kantorovich, on dispose d'après le théorème 1.1.2 d'une formulation duale. Nous avons vu que si (φ,ψ) est optimal pour $D'(\mu,\nu)$, alors tout couple (X,Y) optimal pour $\mathcal{W}_1(\mu,\nu)$ est concentré sur l'ensemble des $(x,y) \in E^2$ satisfaisant $\varphi(y) - \psi(x) = d(x,y)$. Pour $x = y$, on obtient $\varphi(x) = \psi(x)$. De plus, les applications $x \mapsto d(x,y)$ à y fixé et $y \mapsto d(x,y)$ à x fixé étant 1-Lipschitziennes, on en déduit qu'il est suffisant de chercher un couple optimal (φ,ψ) pour $D'(\mu,\nu)$ parmi les couples (f,f) où $f : E \to \mathbb{R}$ est une application 1-Lipschitzienne. Il en résulte la formule de *dualité de Kantorovich-Rubinstein* pour la distance $\mathcal{W}_1$:

$$\mathcal{W}_1(\mu,\nu) = \sup_{f:E\to\mathbb{R} \text{ 1-Lipschitz}} \left(\int_E f(y)\,\nu(dy) - \int_E f(x)\,\mu(dx) \right). \tag{1.1.7}$$

Pour une étude approfondie des distances de Wasserstein, on renvoie aux ouvrages cités plus haut sur la théorie classique du transport optimal. Pour avoir un aperçu de l'abondance des applications des distances de Wasserstein, on pourra par exemple consulter [45, 163, 167, 165, 195, 43, 185, 53, 136].

1.1.4 Le couplage comonotone de Hoeffding-Fréchet

Jusqu'à maintenant nous avons considéré le problème de Monge-Kantorovich dans un espace polonais quelconque. Sans surprise, les secrets du transport optimal se dévoilent plus facilement lorsque l'on travaille dans $\mathbb{R}$. Il existe en effet des outils unidimensionnels bien commodes qui aident considérablement à la compréhension de ce qu'il se passe sur la droite réelle. Pour une mesure de probabilité μ sur $\mathbb{R}$, on note $F_\mu : x \mapsto \mu((-\infty,x])$ sa fonction de répartition, et F_μ^{-1} sa fonction quantile, c'est-à-dire l'inverse généralisé à gauche de F_μ, définie pour tout $u \in (0,1)$ par

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

La fonction de répartition F_μ est croissante et continue à droite, tandis que la fonction quantile F_μ^{-1} est croissante et continue à gauche. Si F_μ est bijective, alors F_μ^{-1} est effectivement son inverse au sens classique. Dans le cas général, si X est distribuée selon μ, alors $F_\mu^{-1}(F_\mu(X)) = X$ presque partout. Réciproquement, si U est uniforme sur $]0,1[$, l'égalité

$F_\mu(F_\mu^{-1}(U)) = U$ ne peut être vérifiée si $]0,1[\nsubseteq F_\mu(\mathbb{R})$, ce qui se produit lorsque F_μ présente des sauts de discontinuité. Pour tout $x \in \mathbb{R}$, on note $F_\mu(x-) = \lim_{y\to x, y<x} F_\mu(y)$. On a alors $F_\mu(x) - F_\mu(x-) = \mu(\{x\})$, et une manière de combler les sauts de discontinuité de F_μ est d'ajouter une loi uniforme sur le saut $]F_\mu(x-), F_\mu(x)[$ pour pouvoir explorer presque sûrement tout l'intervalle $]0,1[$. Selon ce principe, on obtient que si V est uniformément distribuée sur $]0,1[$ et indépendante de U, alors la variable aléatoire $W = F_\mu(F_\mu^{-1}(U)-) + V\mu(\{F_\mu^{-1}(U)\})$ suit la loi uniforme sur $]0,1[$. Cette observation a une conséquente importante sur la manière de représenter les lois unidimensionnelles. D'une part, si U est uniforme sur $]0,1[$, alors on montre facilement que $F_\mu^{-1}(U)$ suit la loi μ. Cette propriété est appelée *méthode de la transformée inverse* ou *méthode de simulation par inversion de la fonction de répartition*. Réciproquement, si X suit la loi μ et V est uniforme sur $]0,1[$ et indépendante de X, alors d'après ce qui précède la variable aléatoire $W = F_\mu(X-) + V\mu(\{X\})$ suit la loi uniforme sur $]0,1[$. Par définition de l'inverse généralisé, on montre facilement que W satisfait $F_\mu^{-1}(W) = X$ presque sûrement. Cela montre que la méthode de la transformée inverse n'induit aucune perte de généralité : toute variable aléatoire réelle peut être représentée comme l'image d'une variable aléatoire uniforme sur $]0,1[$ par sa fonction quantile, quitte à enrichir l'espace de probabilité.

Soient alors deux variables aléatoires réelles X et Y, suivant respectivement μ et ν. D'après ce qui précède, on peut trouver deux variables aléatoires U et V uniformes sur $]0,1[$ telles que presque sûrement, $(X,Y) = (F_\mu^{-1}(U), F_\nu^{-1}(V))$. La donnée de la loi jointe de (X,Y) est donc équivalente à la donnée de la loi jointe de (U,V), ou encore à la donnée d'une copule bidimensionnelle, c'est-à-dire la fonction de répartition d'un vecteur bidimensionnel dont les marginales suivent des uniformes sur $]0,1[$: c'est le théorème de Sklar [175]. On appelle couplage *comonotone* ou de *Hoeffding-Fréchet* la loi du vecteur $(F_\mu^{-1}(U), F_\nu^{-1}(U))$, où U est uniforme sur $]0,1[$. On définit aussi son antagoniste le couplage *antimonotone*, à savoir la loi du vecteur $(F_\mu^{-1}(U), F_\nu^{-1}(1-U))$. Ces couplages sont d'une importance capitale en théorie du transport optimal, puisqu'ils réalisent les extrema du problème de Monge-Kantorovich pour une classe très générale de coûts [50, 161].

Théorème 1.1.4. *Soient μ et ν deux mesures de probabilité sur $\mathbb{R}$ et $c : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ une application satisfaisant la condition de Monge, i.e. continue à droite (pour l'ordre partiel naturel sur $\mathbb{R}^2$) et telle que*

$$\forall (x,y), (x',y') \in \mathbb{R}^2, \quad (x,y) \le (x',y') \implies c(x',y') - c(x,y') - c(x',y) + c(x,y) \le 0.$$

Supposons de plus que pour toutes variables aléatoires $X \sim \mu$ et $Y \sim \nu$, $c(X,Y)$ est intégrable. Alors pour $X \sim \mu$, $Y \sim \nu$ et U uniforme sur $]0,1[$,

$$\mathbb{E}[c(F_\mu^{-1}(U), F_\nu^{-1}(U))] \le \mathbb{E}[c(X,Y)] \le \mathbb{E}[c(F_\mu^{-1}(U), F_\nu^{-1}(1-U))].$$

On notera que pour toute application convexe $f : \mathbb{R} \to \mathbb{R}$, l'application $(x,y) \mapsto f(|y-x|)$ satisfait à la condition de Monge. Pour tout $\rho \in [1,+\infty[$, l'application $x \mapsto |x|^\rho$ est bien sûr convexe, et le théorème 1.1.4 a donc pour conséquence remarquable que le couplage de Hoeffding-Fréchet est optimal pour $\mathcal{W}_\rho(\mu,\nu)$, indépendamment de ρ :

$$\forall \rho \in [1,+\infty[, \quad \mathcal{W}_\rho(\mu,\nu) = \left(\int_0^1 |F_\nu^{-1}(u) - F_\mu^{-1}(u)|^\rho \, du\right)^{\frac{1}{\rho}}. \tag{1.1.8}$$

Lorsque $\rho > 1$, la stricte convexité de l'application $x \mapsto |x|^\rho$ a pour conséquence que le couplage de Hoeffding-Fréchet est même l'unique couplage optimal pour $\mathcal{W}_\rho(\mu, \nu)$. Dans le cas $\rho = 1$, l'unicité disparaît. On pourra s'en convaincre en observant par exemple que s'il existe $a \in \mathbb{R}$ tel que $\mu(]-\infty, a]) = 1 = \nu([a, +\infty[)$, alors tout couplage entre μ et ν est optimal pour $\mathcal{W}_1(\mu, \nu)$. Pour $\rho = 1$, l'égalité (1.1.8) indique que la distance $\mathcal{W}_1(\mu, \nu)$ est simplement l'aire qui sépare les courbes de F_μ^{-1} et F_ν^{-1}. En échangeant l'axe des abscisses et des ordonnées, on voit qu'il s'agit aussi de l'aire qui sépare les courbes de F_μ et F_ν, d'où

$$\mathcal{W}_1(\mu, \nu) = \int_{\mathbb{R}} |F_\nu(t) - F_\mu(t)| \, dt. \tag{1.1.9}$$

Dans la théorie du transport optimal martingale, nous chercherons des couplages martingale qui exprimeront une proximité au sens de la distance $\mathcal{W}_\rho$. Le couplage de Hoeffding-Fréchet nous servira alors de point de départ. Bien sûr, ce dernier n'est en général pas admissible dans le contexte martingale, et il faudra donc le modifier. D'une part, la modification doit être suffisante pour le rendre martingale. D'autre part, la modification doit être la plus faible possible, en un sens à préciser, pour conserver une faible distance $\mathcal{W}_\rho$. Il nous faut donc analyser plus en détails le couplage de Hoeffding-Fréchet, et plus particulièrement la loi conditionnelle de $F_\nu^{-1}(U)$ sachant $F_\mu^{-1}(U)$. Soient $f, g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ continues bornées et $W = F_\mu(F_\mu^{-1}(U)-) + V\mu(\{F_\mu^{-1}(U)\})$ où V est uniforme sur $]0,1[$ et indépendante de U. D'après ce qui précède, W suit la loi uniforme sur $]0,1[$, et $F_\mu^{-1}(W) = F_\mu^{-1}(U)$ presque sûrement. En posant $(X, Y) = (F_\mu^{-1}(U), F_\nu^{-1}(U))$, on a donc

$$\begin{aligned}\mathbb{E}[f(X)g(Y)] &= \mathbb{E}[f(F_\mu^{-1}(U))g(F_\nu^{-1}(U))] = \mathbb{E}[f(F_\mu^{-1}(W))g(F_\nu^{-1}(W))] \\ &= \mathbb{E}[f(F_\mu^{-1}(U))g(F_\nu^{-1}(W))] = \mathbb{E}[f(X)g(F_\nu^{-1}(F_\mu(X-) + V\mu(\{X\})))].\end{aligned}$$

On en déduit que $\mathbb{E}[g(Y)|X] = \varphi(X)$, où $\varphi : x \mapsto \mathbb{E}[g(F_\nu^{-1}(F_\mu(x-) + V\mu(\{x\})))]$. Pour tout $x \in \mathbb{R}$, $F_\mu(x-) + V\mu(\{x\})$ suit une loi uniforme sur $[F_\mu(x-), F_\mu(x)]$ si $\mu(\{x\}) > 0$, ou bien est constante égale à $F_\mu(x)$ si $\mu(\{x\}) = 0$. En posant la convention que la loi uniforme sur un singleton $\{a\}$ est δ_a, on en déduit que si (X, Y) est comonotone, alors la loi conditionnelle de Y sachant X est l'image par F_ν^{-1} de la loi uniforme sur $[F_\mu(X-), F_\mu(X)]$. On retrouve alors un fait bien connu : le couplage de Hoeffding-Fréchet est la loi d'un couple $(X, T(X))$ si et seulement si F_ν^{-1} est constante sur les sauts de F_μ, c'est-à-dire sur les intervalles de la forme $](F_\mu(x-), F_\mu(x)]$ (on peut indifféremment ouvrir ou fermer l'intervalle à droite par continuité à gauche de F_ν^{-1}). Dans ce cas, l'application T est égale à $F_\nu^{-1} \circ F_\mu$ et appelée *transport de Monge*. L'hypothèse de l'existence d'un transport de Monge est évidemment satisfaite lorsque F_μ ne saute pas, ou de manière équivalente lorsque μ n'a pas d'atome. Cela donne des conditions très générales sous lesquelles la solution du problème de Monge-Kantorovich dans $\mathbb{R}$ coïncide avec celle du problème de Monge.

1.1.5 Résolution numérique

Étant donné le nombre considérable d'applications du transport optimal, sa résolution numérique est un enjeu majeur. Il a été mentionné précédemment que la résolution du problème de

Monge-Kantorovich est stable par rapport à ses marginales, conséquence du théorème 1.1.1. Puisque l'on sait approcher des mesures quelconques μ et ν par des mesures empiriques $\hat{\mu}$ et $\hat{\nu}$, l'approche historique fut naturellement de chercher à résoudre $C'(\hat{\mu}, \hat{\nu})$ pour des mesures à support fini, puis de passer à la limite pour approcher $C'(\mu, \nu)$. Soient $\hat{\mu}$ et $\hat{\nu}$ à support fini de la forme $\hat{\mu} = \sum_{i=1}^{n} p_i \delta_{x_i}$ et $\hat{\nu} = \sum_{j=1}^{m} q_j \delta_{y_j}$, où $p_1, \cdots, p_n, q_1, \cdots, q_m \in [0, 1]$ et $\sum_{i=1}^{n} p_i = \sum_{j=1}^{m} q_j = 1$. Toute loi jointe π de marginales $\hat{\mu}$ et $\hat{\nu}$ s'écrit sous la forme $\pi = \sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} \delta_{(x_i, y_j)}$. La mesure π est bien une mesure de probabilité si et seulement si pour tout $(i, j) \in \{1, \cdots, n\} \times \{1, \cdots, m\}$, $r_{i,j} \geq 0$ et $\sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} = 1$. De plus, π a pour première marginale $\hat{\mu}$ si et seulement si pour tout $i \in \{1, \cdots, n\}$, $\sum_{j=1}^{m} r_{i,j} = p_i$, et a pour seconde marginale $\hat{\nu}$ si et seulement si pour tout $j \in \{1, \cdots, m\}$, $\sum_{i=1}^{n} r_{i,j} = q_j$. Le problème de Monge-Kantorovich revient donc à minimiser

$$\inf_{(r_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}} \sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} c(x_i, y_j), \tag{1.1.10}$$

sur l'ensemble des $(r_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ satisfaisant aux contraintes linéaires mentionnées. Il s'agit donc d'un problème de programmation linéaire. Kantorovich et son élève Gavurin développent à cet effet en 1939 une méthode de potentiels pour le résoudre [117] (l'écart de 10 ans avec la publication s'expliquant par la censure soviétique). On peut aussi faire appel aux algorithmes classiques de la recherche opérationnelle : l'algorithme du simplexe [63], l'algorithme hongrois [126], l'algorithme des points intérieurs [122] ou encore l'algorithme des enchères [36, 37]. Ces algorithmes sont cependant trop coûteux en temps de calcul et en mémoire pour pouvoir être utilisés en pratique par certains domaines, comme l'analyse d'images qui traite des pixels et donc des variables par millions ou plus [178, 182].

Une possibilité est d'ajouter un terme de régularisation entropique et donc de chercher à minimiser

$$\inf_{(r_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}} \left(\sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} c(x_i, y_j) + \lambda \sum_{i=1}^{n} \sum_{j=1}^{m} r_{i,j} \ln(r_{i,j} - 1) \right), \tag{1.1.11}$$

où $(r_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}$ est soumis aux mêmes contraintes linéaires, et λ est un réel strictement positif. Étant prouvé sous certaines conditions que le coût du problème de transport optimal avec pénalisation entropique converge bien vers le problème d'origine [55, 131, 51], nombre de chercheurs ont profité de cette méthode qui permet de recourir à l'algorithme de Sinkhorn [173, 174] pour obtenir des performances de résolution très élevées [85, 62, 32, 179, 157, 187].

Des performances supérieures se font en général au prix d'un sacrifice sur la généralité de la fonction de coût. Des méthodes ont notamment été développées lorsque c est le carré d'une distance, autrement dit pour calculer une distance $\mathcal{W}_2$. Lorsque les marginales μ et ν ont des supports respectifs S_μ et S_ν, des densités respectives f_μ et f_ν par rapport à la mesure de Lebesgue et que l'optimum pour le problème dual $D'(\mu, \nu)$ est suffisament régulier et fortement convexe, ce dernier peut être associé à l'équation de Monge-Ampère [98]

$$f_\mu(\nabla u) \det(D^2 u) = f_\nu(y), \quad \nabla u(S_\nu) \subset S_\mu. \tag{1.1.12}$$

Il est alors possible de résoudre ce problème par différences finies [82, 35, 33, 34]. Il est aussi possible de résoudre l'équation (1.1.12) par une approche semi-discrète [61], c'est-à-dire lorsque μ est à densité et ν est à support fini. À l'origine développée pour gérer les conditions aux limites de Dirichlet [145, 139, 142] de l'équation de Monge-Ampère, cette méthode a pu être utilisée pour la résolution du transport optimal [137, 65, 95, 123, 133, 124].

Enfin, on pourra citer la méthode de résolution numérique fondée sur la formulation de Benamou-Brenier [28, 29, 30, 49, 148, 134, 54], à savoir l'identification de la distance $\mathcal{W}_2^2$ avec le problème de minimisation

$$\inf \int_{\mathbb{R}^d} \int_0^1 \rho(t,x)|v(t,x)|^2 \, dx \, dt,$$

où le minimum porte sur l'ensemble des densités $\rho : [0,1] \times \mathbb{R}^d \to \mathbb{R}_+$ et champs de vitesse $v : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfaisant

$$\rho(0,\cdot) = f_\mu, \quad \rho(1,\cdot) = f_\nu \quad \text{et} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0.$$

Lorsque c est simplement la distance, c'est-à-dire lorsque l'on veut calculer une distance $\mathcal{W}_1$, on pourra consulter la méthode d'utilisation de coefficients d'ondelettes [170], de Lagrangien augmenté [31] ou d'éléments finis [180].

1.2 Le transport optimal martingale

1.2.1 Motivation

Pour comprendre comment le transport optimal, et plus exactement sa variante martingale, nous sont utiles dans le calcul de bornes de prix de produits dérivés, il nous faut d'abord présenter les bases de la théorie des mathématiques financières. Pour un traitement bien plus complet, on consultera les ouvrages de référence [171, 172, 128, 79, 155].

On considère un marché composé de $d+1$ actifs financiers, où $d \in \mathbb{N}^*$, qui peuvent s'échanger à des dates $t \in \{0, \cdots, T\}$, où $T \in \mathbb{N}^*$. Soient $i \in \{0, \cdots, d+1\}$ et $t \in \{0, \cdots, T\}$. On note S_t^i le cours de l'actif i à la date t. Plus formellement, S_t^i est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On considèrera que l'actif d'indice $i = 0$ est dit *sans-risque*. Par convention, on pose $S_0^0 = 1$. En notant $r \in \mathbb{R}$ le taux d'intérêt supposé constant, 1€ placé en t rapporte $(1+r)$€ à la date $t+1$, d'où $S_t^0 = (1+r)^t$. Par souci de clarté, on supposera $r = 0$, ce qui à certaines renormalisations près n'a aucune incidence sur la suite. En particulier, $S_t^0 = 1$. On décide alors d'investir sur ce marché. On note ϕ_0^i la quantité d'actif i détenue en 0, et pour $t > 0$, ϕ_t^i la quantité d'actif i détenue sur l'intervalle $]t-1, t]$, de sorte que la valeur de notre portefeuille à la date t soit donnée par

$$V_t = \sum_{i=0}^{d} \phi_t^i S_t^i. \tag{1.2.1}$$

Notons qu'il est possible d'avoir $\phi_t^i < 0$, ce qui correspond à une dette en actif i. Cependant, une stratégie de portefeuille est admissible seulement si l'on impose à l'investisseur d'être toujours en mesure de rembourser ses dettes, soit $V_t \geq 0$ à tout instant. De plus, on impose une condition d'auto-financement, c'est-à-dire que l'on interdit un quelconque financement du portefeuille extérieur au marché considéré. En particulier, les variations de la valeur du portefeuille entre les dates t et $t+1$ ne proviennent que des variations des cours des actifs risqués, soit

$$V_{t+1} - V_t = \sum_{i=1}^{d} \phi_{t+1}^i (S_{t+1}^i - S_t^i). \tag{1.2.2}$$

Une stratégie d'arbitrage est alors par définition une stratégie de portefeuille admissible telle que $V_0 = 0$ et $\mathbb{P}(V_T > 0) \neq 0$. Nous l'avons évoqué, on ne considère que des marchés *viables*, c'est-à-dire sans opportunité d'arbitrage. On dispose alors du théorème suivant, connu sous le nom de premier théorème fondamental de l'évaluation d'actifs.

Théorème 1.2.1 (Premier théorème fondamental de l'évaluation d'actifs)**.** *Le marché ne présente pas de stratégie d'arbitrage si et seulement s'il existe une probabilité* $\mathbb{P}^*$ *équivalente à* $\mathbb{P}$*, appelée* probabilité risque-neutre*, sous laquelle les actifs actualisés* $\left(\frac{S_t^1}{S_t^0}, \cdots, \frac{S_t^d}{S_t^0}\right)_{t\in\{0,\cdots,T\}}$ *sont des martingales.*

Puisque nous avons supposé le taux d'intérêt nul, nous avons donc que sous $\mathbb{P}^*$, le panier d'actifs $(S_t = (S_t^1, \cdots, S_t^d))_{t\in\{0,\cdots,T\}}$ est une martingale. On montre alors facilement que sous $\mathbb{P}^*$, le processus $(V_t)_{t\in\{0,\cdots,T\}}$ est lui aussi une martingale (avec des considérations sur la prévisibilité de $(\phi_t^i)_{t\in\{0,\cdots,T\}}$). Soit alors une option européenne de maturité T dont le payoff dépend de la valeur des actifs financiers aux dates $t_1, t_2 \in \{0, \cdots, T\}$ telles que $t_1 < t_2$, c'est-à-dire un contrat financier qui garantit à son détenteur une valeur $c(S_{t_1}, S_{t_2})$ à l'instant T, où $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$. Supposons alors que notre stratégie de portefeuille réplique l'option, c'est-à-dire $V_T = c(S_{t_1}, S_{t_2})$. Puisque $(V_t)_{t\in\{0,\cdots,T\}}$ est une martingale, son espérance est constante égale à sa valeur d'origine V_0 (considérée déterministe sous l'hypothèse d'une information à l'instant 0 modélisée par la tribu grossière $\mathcal{F}_0 = \{\emptyset, \Omega\}$), soit

$$V_0 = \mathbb{E}^*[V_T] = \mathbb{E}^*[c(S_{t_1}, S_{t_2})],$$

où $\mathbb{E}^*$ désigne l'espérance sous $\mathbb{P}^*$. Ainsi, toute stratégie de couverture de l'option de payoff $c(S_{t_1}, S_{t_2})$ a pour valeur initiale $\mathbb{E}[c(S_{t_1}, S_{t_2})]$. Par absence d'opportunité d'arbitrage, $\mathbb{E}[c(S_{t_1}, S_{t_2})]$ est donc le prix de cette option. Pour $d = 1$, Breeden et Litzenberger [46] montrent que les lois respectives μ et ν de S_{t_1} et S_{t_2} sont déterminées de manière unique dès que l'on dispose des cours des calls et des puts pour tout prix d'exercice aux dates t_1 et t_2. De manière analogue lorsque $d > 1$, la connaissance des lois respectives μ et ν de S_{t_1} et S_{t_2} se déduit des cours d'options de payoff $(\lambda \cdot S_{t_1} - K)^+$ et $(\lambda \cdot S_{t_2} - K)^+$ pour tout prix d'exercice K et $\lambda \in \mathbb{R}^d$ de composantes positives de somme 1. De plus, le théorème 1.2.1 implique que (S_{t_1}, S_{t_2}) est une martingale, soit $\mathbb{E}[S_{t_2}|S_{t_1}] = S_{t_1}$ presque sûrement. On en déduit que le prix P de l'option de payoff $c(S_{t_1}, S_{t_2})$ vérifie

$$\inf_{\substack{X\sim\mu, Y\sim\nu \\ \mathbb{E}[Y|X]=X \text{ p.s.}}} \mathbb{E}[c(X,Y)] \leq P \leq \sup_{\substack{X\sim\mu, Y\sim\nu \\ \mathbb{E}[Y|X]=X \text{ p.s.}}} \mathbb{E}[c(X,Y)]. \tag{1.2.3}$$

On reconnaît bien sûr dans les termes de gauche et de droite le problème de Monge-Kantorovich (1.1.4), auquel on a ajouté une contrainte martingale. Insistons sur le fait que l'encadrement (1.2.3) n'a introduit aucune hypothèse de modèle, à l'exception de l'absence d'opportunité d'arbitrage, d'où la robustesse de ces bornes de prix.

1.2.2 Définition

Comme on a pu l'apercevoir en (1.2.3), on appellera donc problème de transport optimal martingale le problème suivant introduit en 2013 par Beiglböck, Henry-Labordère et Penkner dans le cas discret [23] : étant données deux mesures de probabilité μ et ν définies respectivement sur $\mathbb{R}^d$ et de premier moment fini, où $d \in \mathbb{N}^*$, et une application de coût $c : \mathbb{R}^d \times \mathbb{R}^d \to [0, +\infty]$, déterminer

$$C_M(\mu, \nu) = \inf_{\substack{X \sim \mu, Y \sim \nu \\ \mathbb{E}[Y|X] = X \text{ p.s.}}} \mathbb{E}[c(X, Y)]. \tag{1.2.4}$$

Nous avons vu que le problème de Monge-Kantorovich réalise toujours l'infimum sur un ensemble non vide. Cette particularité sympathique n'est malheureusement pas partagée par le problème (1.2.4). En effet, on se convainc rapidement qu'il n'existe par exemple pas de couple (X, Y) tel que $X \sim \delta_0$, $Y \sim \delta_1$, et $\mathbb{E}[Y|X] = X$ presque sûrement. Plus généralement, si $\mathbb{E}[Y \mid X] = X$ presque sûrement et $f : \mathbb{R}^d \to \mathbb{R}$ est une application convexe, alors l'inégalité de Jensen pour les espérances conditionnelles implique que

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\mathbb{E}[Y|X])] \leq \mathbb{E}[\mathbb{E}[f(Y)|X]] = \mathbb{E}[f(Y)].$$

On dit alors que X est dominée par Y dans l'ordre convexe, ce que l'on note $X \leq_{cx} Y$. Si X et Y sont de lois respectives μ et ν, alors on vient de montrer que $X \leq_{cx} Y$ est une condition nécessaire pour que $C_M(\mu, \nu)$ réalise un infimum sur un ensemble non vide. D'après le célèbre théorème de Strassen, il y a en fait équivalence [183]. Plus précisément, si X et Y sont des variables aléatoires intégrables, alors $C_M(\mu, \nu)$ réalise l'infimum sur un ensemble non vide si et seulement si $X \leq_{cx} Y$. Cette équivalence est illustrée en dimension $d = 1$ par l'existence des couplages dits left-curtain et right-curtain introduits par Beiglböck et Juillet [25].

Le problème de transport optimal martingale a aussi été introduit en temps continu en 2014 par Galichon, Henry-Labordère et Touzi [84] pour résoudre un problème de sur-couverture, classiquement résolu à travers un problème de plongement de Skorokhod. Ce dernier, formulé par Skorokhod en 1961 [176, 177], consistait à déterminer pour une mesure de probabilité μ centrée de moment d'ordre 2 fini et un mouvement brownien B, un temps d'arrêt intégrable τ tel que B_τ suive la loi μ. On pourra consulter le rapport d'Obłój qui donne une bonne vue d'ensemble de l'évolution jusqu'en 2004 de ce problème [144]. Le lien entre les deux problèmes est explicité par [24, 22, 97, 17].

Pour une étude détaillée de l'application du problème de transport optimal martingale en finance, on pourra consulter le livre de Pierre Henry-Labordère [103].

1.2.3 Dualité

Le problème (1.2.4) est qualifié de primal, et concerne l'évaluation d'une option. Il est aussi possible d'adopter une approche duale, dans l'esprit du problème (1.1.5), associée à un problème de couverture, ou plus exactement de sous-couverture. Reprenons comme précédemment l'exemple d'un marché viable composé de $d+1$ actifs financiers avec un taux d'intérêt nul. On se place sous la probabilité risque-neutre $\mathbb{P}^*$. On cherche à déterminer une stratégie de sous-couverture de l'option de payoff $c(S_{t_1}, S_{t_2})$, en s'autorisant l'achat de toute option de payoff ne dépendant que de S_{t_1}, ou de S_{t_2}, ainsi que les Δ-hedges, c'est-à-dire les options de payoff $H(S_{t_1}) \cdot (S_{t_2} - S_{t_1})$ où $H : \mathbb{R}^d \to \mathbb{R}^d$ est continue bornée. On dispose donc d'une option de payoff de la forme $\psi(S_{t_1}) + \varphi(S_{t_2}) + H(S_{t_1}) \cdot (S_{t_2} - S_{t_1})$, qui est bien une sous-couverture si ce payoff demeure inférieur à $c(S_{t_1}, S_{t_2})$. D'après le théorème 1.2.1, (S_{t_1}, S_{t_2}) est une martingale sous $\mathbb{P}^*$, donc les δ-hedges ont un prix nul. En revanche, le prix des options de payoffs $\psi(S_{t_1})$ et $\varphi(S_{t_2})$ sont respectivement égaux à $\mathbb{E}^*[\psi(S_{t_1})]$ et $\mathbb{E}^*[\varphi(S_{t_2})]$, d'après ce qui précède. La détermination d'une telle stratégie qui donne le prix le plus élevé mène naturellement à la formulation duale du problème de transport optimal martingale : étant données deux mesures de probabilité μ et ν définies sur $\mathbb{R}^d$, où $d \in \mathbb{N}^*$, et une application de coût $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, déterminer

$$D_M(\mu, \nu) = \sup_{\substack{(\psi,\varphi,H) \in L^1(\mu) \times L^1(\nu) \times C_b^0(\mathbb{R}^d, \mathbb{R}^d), \\ \forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d,\ \psi(x) + \varphi(y) + H(x) \cdot (y-x) \le c(x,y)}} \left(\int_{\mathbb{R}^d} \psi(x)\, \mu(dx) + \int_{\mathbb{R}^d} \varphi(y)\, \nu(dy) \right), \tag{1.2.5}$$

où $C_b^0(\mathbb{R}^d, \mathbb{R}^d)$ désigne l'ensemble des applications continues bornées de $\mathbb{R}^d$ dans lui-même. Précisons que l'affectation d'un signe $+$ devant la fonction ψ par rapport au signe $-$ du problème (1.1.5) n'a aucune autre justification que celle d'être ici une convention plus naturelle par rapport au contexte de la stratégie de sous-couverture. Notons que quitte à remplacer le payoff c par son opposé, la résolution de $D_M(\mu, \nu)$ permet de déterminer le prix le moins élevé d'une stratégie de sur-couverture, ce qui en pratique est souvent plus naturel. Beiglböck, Henry-Labordère et Penkner montrent alors que le saut de dualité en dimension 1 est nul [23]. On dit que deux mesures de probabilité μ et ν sont dans l'ordre convexe, ce que l'on note $\mu \le_{cx} \nu$, si toutes variables aléatoires X et Y telles que $X \sim \mu$ et $Y \sim \nu$ vérifient $X \le_{cx} Y$.

Théorème 1.2.2. *Soient μ et ν deux mesures de probabilité sur $\mathbb{R}_+$ telles que $\mu \le_{cx} \nu$ et $c : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ une application continue telle qu'il existe une constante $K \in \mathbb{R}$ satisfaisant*

$$\forall (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+, \quad \max(c(x,y), 0) \le K(1 + x + y).$$

Alors $C_M(\mu, \nu) = D_M(\mu, \nu)$.

Dans le cadre du théorème 1.2.2, l'optimum de $D_M(\mu, \nu)$ n'est pas nécessairement atteint, voir le contre-exemple de Beiglböck, Henry-Labordère et Penkner [23]. Il est cependant atteint lorsque l'on s'autorise la formulation quasi-sûre de Beiglböck, Nutz et Touzi [27].

1.2.4 Stabilité et résolution numérique

Avec les mêmes notations et raisonnement que pour (1.1.10), on voit que la résolution du problème de transport optimal martingale entre deux mesures à support fini $\mu = \sum_{i=1}^n p_i \delta_{x_i}$ et $\nu = \sum_{j=1}^m q_j \delta_{y_j}$ revient à la résolution du problème de programmation linéaire

$$\inf_{(r_{i,j})_{1 \leq i \leq n, 1 \leq j \leq m}} \sum_{i=1}^n \sum_{j=1}^m r_{i,j} c(x_i, y_j), \tag{1.2.6}$$

sous les contraintes

$$r_{i,j} \geq 0, \quad \sum_{j=1}^m r_{i,j} = p_i, \quad \sum_{i=1}^n r_{i,j} = q_j \quad \text{et} \quad \sum_{j=1}^m r_{i,j} y_j = p_i x_i, \tag{1.2.7}$$

où la dernière contrainte est celle de martingalité. On peut alors s'en remettre aux algorithmes classiques de la recherche opérationnelle déjà évoqués [63, 126, 122, 36, 37], à l'instar de Pierre Henry-Labordère qui discrétise le problème [102]. Cependant, cette méthode nous confronte à deux problématiques importantes. La première est l'échantillonage dans l'ordre convexe, c'ést-à-dire la possibilité d'approcher les marginales par des lois à support fini et dans l'ordre convexe, pour que le problème approché ne soit pas un infimum sur un ensemble vide. En effet, les mesures empiriques d'échantillons indépendants et identiquement distribués suivant deux distributions dans l'ordre convexe n'ont aucune raison d'avoir en général le même barycentre, et donc ne sont en général pas dans l'ordre convexe. Alfonsi, Corbetta et Jourdain ont apporté une solution à ce problème en dimensions 1 [3] et supérieure [4]. Ils posent néanmoins la seconde problématique, et non des moindres, de la stabilité du problème de transport optimal martingale par rapport aux marginales. La réponse à cette question est cruciale pour déterminer si le problème discrétisé converge effectivement vers le problème d'origine. De plus, il a été mentionné précédemment que les marginales sont en général connues en vertu du résultat de Breeden et Litzenberger [46], qui nécessite la connaissance des cours de calls et de puts pour une infinité de prix d'exercices, ce dont le marché ne dispose pas. Les marginales sont donc en pratique extrapolées à partir des données disponibles. Par conséquent, il existe un bruit intrinsèque à la donnée de ces marginales. Dans ce contexte, la résolution numérique du problème de transport optimal martingale n'a de sens que si ce dernier est effectivement stable par rapport à ses marginales.

La question de la stabilité du problème de transport optimal martingale est restée ouverte quelques années. Des résultats partiels ont été donnés par Guo et Obłój [96], et la stabilité des couplages left-curtain de Nicolas Juillet [112] a aussi représenté un progrès important pour les coûts satisfaisant la condition dite de Spence-Mirrlees [106]. En 2019, la stabilité en dimension 1 est finalement confirmée sous des conditions assez générales par Backhoff-Veraguas et Pammer [20], suivis de près par un travail indépendant de Johannes Wiesel [194]. Pour le prouver, les premiers ont exploité l'équivalence entre la notion de monotonie martingale et optimalité, une sorte d'analogue martingale du théorème 1.1.1, prouvée notamment par Beiglböck et Juillet [25]. En exploitant la stabilité de cette notion, ils démontrent la stabilité du problème de transport optimal martingale pour les coûts continus à croissance au plus linéaire. Le second s'est quant à lui directement confronté à la formulation primale du problème, en développant la notion de réarrangement martingale de couplage. La question de la stabilité en dimension supérieure est en revanche toujours ouverte et reste un domaine actif de recherche.

Comme pour le transport optimal classique, on peut accroître les performances en ajoutant un terme de régularisation entropique, et donc résoudre

$$\inf_{(r_{i,j})_{1\leq i\leq n, 1\leq j\leq m}} \left(\sum_{i=1}^{n}\sum_{j=1}^{m} r_{i,j} c(x_i, y_j) + \lambda \sum_{i=1}^{n}\sum_{j=1}^{m} r_{i,j} \ln(r_{i,j} - 1) \right), \tag{1.2.8}$$

sous les contraintes (1.2.7), où λ est un réel strictement positif. Par rapport au problème équivalent du transport optimal classique, seules des contraintes linéaires égalité ont été ajoutées, ce qui permet d'utiliser la méthode itérative des projections de Bregman décrite par Benamou, Carlier, Cuturi, Nenna et Peyré [32]. Le cas des projections sur les contraintes des marginales sont explicitées par ce même article. Pour la projection sur les contraintes de martingale, on peut employer par exemple la méthode itérative de changement d'échelle généralisée par Darroch et Ratcliff [64]. Guo et Obłój complètent cette approche avec une relaxation de la contrainte martingale, et montrent la convergence du problème aux contraintes relâchées vers le problème de départ [96].

Hadrien De March fait une étude comparative de ces algorithmes et en propose un nouveau, fondé sur la méthode de Newton et la régularisation entropique [66]. Ce nouvel algorithme a le double avantage d'accroître significativement la performance, et de gérer le problème d'échantillonage dans l'ordre convexe soulevé par Alfonsi, Corbetta et Jourdain [3, 4]. Plus récemment, Eckstein et Kupper ont proposé une méthode unifiée de résolution par réseaux de neurones pour résoudre une classe assez générale de problèmes comprenant notamment les problèmes de transport optimal et transport optimal martingale [76]. Leur algorithme peut cependant être instable pour des paramètres de régularisation trop grands et ne pas converger vers la vraie solution, comme illustré par Pierre Henry-Labordère [104], qui propose lui aussi une méthode par réseaux de neurones, dans le cadre d'un algorithme primal-dual pour des fonctions de coût satisfaisant la condition de Spence-Mirrlees. On pourra aussi consulter Tan et Touzi [186] pour la résolution d'une version semimartingale continue du problème de transport optimal.

1.3 Partie 1 : Nouvelle famille de couplages martingale et inégalités de stabilité

1.3.1 Chapitre 2 : Une nouvelle famille de couplages martingale en dimension 1

Soient deux mesures de probabilité distinctes μ et ν définies sur $\mathbb{R}$, de premier moment fini et dans l'ordre convexe. Nous avons vu que pour $\rho \in [1, +\infty[$, la distance de Wasserstein

$\mathcal{W}_\rho(\mu,\nu)$ se définit comme le problème de Monge-Kantorovich pour l'application de coût $(x,y) \mapsto |y-x|^\rho$. Nous définissons son analogue martingale $\mathcal{M}_\rho(\mu,\nu)$ par

$$\mathcal{M}_\rho(\mu,\nu) = \inf_{\substack{X\sim\mu,\ Y\sim\nu \\ \mathbb{E}[Y|X]=X \text{ p.s.}}} \mathbb{E}[|Y-X|^\rho]^{\frac{1}{\rho}}. \tag{1.3.1}$$

Motivés par la recherche d'une preuve de la stabilité du problème de transport optimal martingale en dimension 1, à une époque où les résultats plus généraux de Backhoff-Veraguas et Pammer [20] et Wiesel [194] n'étaient pas encore connus, nous prouvons une inégalité de stabilité qui montre que lorsque μ et ν sont dans l'ordre convexe et proches pour la distance $\mathcal{W}_1$, il existe alors un couplage martingale qui exprime cette proximité. Ce résultat fait l'objet du théorème principal du chapitre 2.

Théorème 1.3.1. *Soient $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ telles que $\mu \le_{cx} \nu$. Alors*

$$\mathcal{M}_1(\mu,\nu) \le 2\mathcal{W}_1(\mu,\nu).$$

Le théorème 1.3.1 se prouve en construisant une famille $((X^Q,Y^Q))_{Q\in\mathcal{Q}}$ de couplages martingale entre μ et ν, paramétrée par un ensemble $\mathcal{Q}$ de mesures de probabilité sur $]0,1[^2$, satisfaisant

$$\forall Q \in \mathcal{Q}, \quad \mathbb{E}[|Y^Q - X^Q|] \le 2\mathcal{W}_1(\mu,\nu).$$

Pour $0 < a < b$, $\mu = \frac{1}{2}(\delta_{-a}+\delta_a)$ et $\nu = \frac{1}{2}(\delta_{-b}+\delta_b)$, on calcule facilement $\mathcal{M}_1(\mu,\nu) = \left(1+\frac{a}{b}\right)\mathcal{W}_1(\mu,\nu)$, ce qui montre pour $b \to a$ que la constante 2 du théorème 1.3.1 ne peut être améliorée. Nous montrons aussi que pour tout $\rho > 1$, il n'existe pas de constante $C \in \mathbb{R}_+^*$ indépendante de (μ,ν) telle que $\mathcal{M}_\rho(\mu,\nu) \le C\mathcal{W}_\rho(\mu,\nu)$.

Hobson et Klimmek [109] caractérisent le couplage martingale optimal pour $\mathcal{M}_1(\mu,\nu)$ sous une certaine condition de structure, à savoir l'existence d'un intervalle borné tel que $\mu(I) = 1 = \nu(I^\complement)$. Lorsque le support de ν est de cardinal inférieur ou égal à 2, on voit très facilement qu'il ne peut exister qu'un seul couplage martingale entre μ et ν. Dans ce cas, tous nos couplages martingales (X^Q,Y^Q) sont identiques, et coïncident avec celui de Hobson et Klimmek. Ce dernier ne fait en revanche jamais partie de notre famille lorsque le cardinal de ν est supérieur ou égal à 3. Un autre couplage martingale de référence et déjà évoqué est le left-curtain introduit par Beiglböck et Juillet [25], optimal pour le problème de transport optimal martingale dès que le coût c satisfait la condition de Spence-Mirrlees, à savoir $\frac{\partial c^3}{\partial x \partial y^2} \le 0$. Henry-Larbordère et Touzi [106] explicitent ce couplage sous la condition que ν n'a pas d'atome et que l'ensemble des maxima locaux de $F_\nu - F_\mu$ est fini, à travers la résolution de deux équations différentielles ordinaires couplées démarrant de chaque maximiseur local le plus à droite. Nous démontrons qu'en général, le couplage left-curtain ne fait pas partie de notre famille.

Nous avons vu dans la section 1.1.4 que le couplage de Hoeffding-Fréchet est optimal pour $\mathcal{W}_1(\mu,\nu)$. On s'inspire donc de ce dernier pour construire nos couplages martingale. Ainsi pour U uniformément distribuée sur $]0,1[$ et V à valeurs dans $]0,1[$ et de loi conditionnelle Q_U

sachant U, on pose $X^Q = F_\mu^{-1}(U)$, $Y^Q = F_\nu^{-1}(U)$ avec probabilité $P(U,V)$ et $Y^Q = F_\nu^{-1}(V)$ avec probabilité $1 - P(U,V)$. De l'égalité

$$F_\mu^{-1}(U) = F_\nu^{-1}(U)\frac{F_\nu^{-1}(V) - F_\mu^{-1}(U)}{F_\nu^{-1}(V) - F_\nu^{-1}(U)} + F_\nu^{-1}(V)\frac{F_\mu^{-1}(U) - F_\nu^{-1}(U)}{F_\nu^{-1}(V) - F_\nu^{-1}(U)},$$

on déduit que (X^Q, Y^Q) est un couplage martingale si l'on pose $P(U,V) = \frac{F_\nu^{-1}(V) - F_\mu^{-1}(U)}{F_\nu^{-1}(V) - F_\nu^{-1}(U)}$, sous réserve que cette quantité soit bien presque sûrement dans l'intervalle $[0,1]$. Par la méthode de la transformée inverse, X est bien distribuée suivant Y. En revanche, pour assurer que Y soit bien distribuée selon ν, il faut préciser la loi conditionnelle Q_U de V sachant U. On définit alors une famille $\mathcal{Q}$ de mesures de probabilité sur $]0,1[^2$ telle que tout $Q \in \mathcal{Q}$ s'écrive sous la forme

$$Q(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,Q_u(dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,Q_v(du),$$

où γ est un facteur de normalisation, et $Q(\{(u,v) \in]0,1[^2 \mid u < v\}) = 1$. En définissant alors U uniformément distribuée sur $]0,1[$, V de loi conditionnelle sachant U égale à Q_U, et W uniformément distribuée sur $]0,1[$ indépendante de (U,V), on montre que $P(U,V) \in [0,1]$ presque sûrement, et les variables aléatoires

$$X^Q = F_\mu^{-1}(U) \quad \text{et} \quad Y^Q = F_\nu^{-1}(U)\mathbb{1}_{\{W \le P(U,V)\}} + F_\nu^{-1}(V)\mathbb{1}_{\{W > P(U,V)\}} \tag{1.3.2}$$

sont telles que Y^Q suive bien la loi ν, et donc (X^Q, Y^Q) est un couplage martingale entre μ et ν d'après ce qui précède. On montre que ces couplages sont proches du couplage de Hoeffding-Fréchet. Pour simplifier, supposons qu'il existe un transport de Monge T (voir section 1.1.4). Alors tout couplage de notre famille minimise $\mathbb{E}[|Y - T(X)|]$ sur l'ensemble des couplages martingales entre μ et ν :

$$\forall Q \in \mathcal{Q}, \quad \mathbb{E}[|Y^Q - T(X^Q)| = \inf_{\substack{X\sim\mu,\ Y\sim\nu \\ \mathbb{E}[Y|X]=X \text{ p.s.}}} \mathbb{E}[|Y - T(X)|] = \mathcal{W}_1(\mu,\nu).$$

Par inégalité triangulaire, on a donc

$$\mathcal{M}_1(\mu,\nu) \le \mathbb{E}[|Y^Q - X^Q|] \le \mathbb{E}[|Y^Q - T(X^Q)|] + \mathbb{E}[|T(X^Q) - X^Q|] = 2\mathcal{W}_1(\mu,\nu),$$

ce qui prouve le théorème 1.3.1.

Parmi notre famille $((X^Q, Y^Q))_{Q\in\mathcal{Q}}$ de couplages martingale figure un couplage particulier. Pour le définir, on pose $\Psi_+ : u \mapsto \int_0^u (F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$, $\Psi_- : v \mapsto \int_0^v (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv$, et l'application φ définie pour tout $u \in]0,1[$ par $\varphi(u) = \Psi_-^{-1}(\Psi_+(u))$ si $F_\mu^{-1}(u) > F_\nu^{-1}(u)$, $\varphi(u) = \Psi_+^{-1}(\Psi_-(u))$ si $F_\mu^{-1}(u) < F_\nu^{-1}(u)$ et $\varphi(u) = u$ si $F_\mu^{-1}(u) = F_\nu^{-1}(u)$. Alors la mesure $Q^{IT} = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\delta_{\varphi(u)}(dv)$ est un élément de $\mathcal{Q}$. Le couplage (X^{IT}, Y^{IT}) défini par (1.3.2) pour $Q = Q^{IT}$ est alors un couplage martingale, complètement explicite en termes de la différence des fonctions quantiles de μ et ν, baptisé le *couplage martingale transformée inverse*. Ce couplage se particularise dans la famille par son caractère explicite, mais aussi par ses propriétés d'optimisation par rapport au transport de Monge. En effet, supposons à nouveau qu'il existe un transport de Monge T. Alors

$$\forall \rho \in]-\infty, 1] \cup [2, +\infty[, \quad \mathbb{E}[|Y^{IT} - T(X^{IT})|^\rho \mathbb{1}_{\{Y^{IT} \neq T(X^{IT})\}}] = \inf_{Q \in \mathcal{Q}} \mathbb{E}[|Y^Q - X^Q|^\rho \mathbb{1}_{\{Y^Q \neq T(X^Q)\}}]$$
$$\forall \rho \in [1, 2], \quad \mathbb{E}[|Y^{IT} - T(X^{IT})|^\rho] = \sup_{Q \in \mathcal{Q}} \mathbb{E}[|Y^Q - X^Q|^\rho],$$

la présence des indicatrices ne servant qu'à assurer une bonne définition pour les exposants négatifs. En particulier pour $\rho = 0$, le couplage martingale transformée inverse maximise $\mathbb{P}(Y^Q = T(X^Q))$ parmi $Q \in \mathcal{Q}$.

Par convexité de $\mathcal{Q}$ et exhibition de deux éléments distincts, on montre que la famille $\mathcal{Q}$ est de cardinal infini indénombrable. Or lorsque ν est réduit à deux atomes, il n'existe qu'un unique couplage martingale, donc $\{(X^Q, Y^Q) \mid Q \in \mathcal{Q}\}$ est réduit à un singleton. Cependant, on montre que sous des conditions assez générales, notamment lorsque μ et ν sont à densité par rapport à la mesure de Lebesgue, que l'application qui à $Q \in \mathcal{Q}$ associe la loi de (X^Q, Y^Q) est injective, ce qui démontre que notre famille contient un nombre infini indénombrable de couplages martingales.

La construction est ensuite généralisée au cas des sous-martingales et surmartingales.

1.3.2 Chapitre 3 : Inégalité Wasserstein martingale pour des mesures de probabilité dans l'ordre convexe

Soient deux mesures de probabilité distinctes $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ dans l'ordre convexe, où $\rho \in [1, +\infty[$. On cherche dans un premier temps à étendre l'inégalité de stabilité $\mathcal{M}_1(\mu, \nu) \leq 2\mathcal{W}_1(\mu, \nu)$, prouvée dans le chapitre 2, à un indice ρ quelconque. On rappelle que pour tout $\rho > 1$, il n'existe pas de constante $C \in \mathbb{R}_+^*$ indépendante de (μ, ν) telle que $\mathcal{M}_\rho(\mu, \nu) \leq C\mathcal{W}_\rho(\mu, \nu)$. Dans le chapitre 3, nous montrons que le contrôle de la quantité $\mathcal{M}_\rho(\mu, \nu)$ par $\mathcal{W}_\rho(\mu, \nu)$ peut en fait faire intervenir le moment centré $\sigma_\rho(\nu)$ d'ordre ρ de ν, défini par

$$\sigma_\rho(\nu) = \inf_{c \in \mathbb{R}} \left(\int_{\mathbb{R}} |y - c|^\rho \, \nu(dy) \right)^{\frac{1}{\rho}}.$$

En effet, dans le cas $\rho = 2$, la condition de martingalité implique la propriété remarquable que $\mathcal{M}_2(\mu, \nu)$ ne dépend que des marginales, à savoir

$$\mathcal{M}_2^2(\mu, \nu) = \int_{\mathbb{R}} |y|^2 \, \nu(dy) - \int_{\mathbb{R}} |x|^2 \, \mu(dx). \tag{1.3.3}$$

Une simple utilisation de l'inégalité de Cauchy-Schwarz, l'inégalité de Jensen et la définition de l'ordre convexe montre alors que $\mathcal{M}_2^2(\mu, \nu) \leq 2\mathcal{W}_2(\mu, \nu)\sigma_2(\nu)$. Plus généralement, on étend le théorème 1.3.1 en dimension 1 à un indice ρ quelconque.

Proposition 1.3.2. *Soient $\rho \in [1, +\infty[$ et $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ tels que $\mu \leq_{cx} \nu$ et $\mu \neq \nu$. Pour tout $Q \in \mathcal{Q}$, le couplage martingale (X^Q, Y^Q) défini par (1.3.2) vérifie*

$$\mathbb{E}[|Y^Q - X^Q|^\rho] \leq K_\rho \mathcal{W}_\rho(\mu, \nu) \sigma_\rho^{\rho-1}(\nu), \tag{1.3.4}$$

où

$$K_\rho = \inf\left\{2^{\rho-1}\gamma_1 + 2(2^{\rho-2}\vee 1)\gamma_2 \mid (\gamma_1,\gamma_2)\in\mathbb{R}_+^2 \ \textit{et}\ \forall x\in\mathbb{R}_+,\ \frac{x+x^\rho}{1+x}\le \gamma_1+\gamma_2 x^{\rho-1}\right\}. \tag{1.3.5}$$

Pour formuler un énoncer plus précis, on introduit la constante C_ρ définie par

$$C_\rho = \inf\left\{C>0 \mid \forall \mu,\nu\in\mathcal{P}_\rho(\mathbb{R}) \text{ tels que } \mu\le_{cx}\nu,\ \mathcal{M}_\rho^\rho(\mu,\nu)\le C\mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu)\right\}. \tag{1.3.6}$$

On déduit immédiatement de l'inégalité (1.3.4) que $C_\rho \le K_\rho$. De plus, $C_2 = 2$ d'après ce qui précède, et $C_1 = 2$ d'après le chapitre 2. En calculant les quantités pour $0<a<b$, $\mu = \frac{1}{2}(\delta_{-a}+\delta_a)$ et $\nu=\frac{1}{2}(\delta_{-b}+\delta_b)$, l'exemple qui nous permettait de montrer que la constante 2 était optimale pour le théorème 1.3.1, on peut aussi déduire une borne inférieure sur C_ρ.

Proposition 1.3.3. *Soit $\rho\in[1,+\infty[$. Alors la constante C_ρ définie par* (1.3.6) *vérifie*

$$2^{\rho-1}\sup_{x\in(1,+\infty)}\frac{x+x^\rho}{(1+x)^\rho}\le C_\rho\le K_\rho\le \begin{cases}\left(1+\sup_{x\in[0,1]}\frac{x-x^{\rho-1}}{1+x}\right)2^{\rho-1} & \textit{pour } \rho>2;\\ 2+\left(2\sup_{x\in[1,+\infty)}\frac{x-x^{\rho-1}}{x^{\rho-1}(1+x)}\right)\wedge 2^{\rho-1} & \textit{pour } 1<\rho<2,\end{cases} \tag{1.3.7}$$

et $K_1 = K_2 = 2$.

Pour $\rho=1$ et $\rho=2$, on retrouve $C_1=C_2=2$. Pour $\rho>2$, une simple étude de fonctions montre que $2^{\rho-1}\le C_\rho\le \frac{3}{2}2^{\rho-1}$. Enfin, on montre que les exposants de l'inégalité de stabilité (1.3.4) sont les bons, au sens où

$$\forall\rho\in]1,+\infty[,\quad \forall s\in]1,\rho],\qquad \sup_{\substack{\mu,\nu\in\mathcal{P}_\rho(\mathbb{R})\\ \mu\le_{cx}\nu}}\frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho^s(\mu,\nu)\sigma_\rho^{\rho-s}(\nu)}=+\infty. \tag{1.3.8}$$

On explore ensuite la généralisation de l'inégalité de stabilité à la dimension supérieure. On considère donc cette fois $\mu,\nu\in\mathcal{P}_\rho(\mathbb{R}^d)$ où $d\in\mathbb{N}^*$, toujours dans l'ordre convexe, et on généralise de manière évidente les quantités $\mathcal{M}_\rho(\mu,\nu)$ et $\sigma_\rho(\nu)$, pour une norme donnée $|\cdot|$ sur $\mathbb{R}^d$. On étudie donc la constante

$$C_{\rho,d} = \inf\left\{C>0 \mid \forall \mu,\nu\in\mathcal{P}_\rho(\mathbb{R}^d) \text{ tels que } \mu\le_{cx}\nu,\ \mathcal{M}_\rho^\rho(\mu,\nu)\le C\mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu)\right\}. \tag{1.3.9}$$

Un résultat global de stabilité du problème de transport optimal martingale en dimension supérieure n'est pas encore connu à l'heure actuelle. La détermination de $C_{\rho,d}$, et notamment la démonstration de sa finitude, représenterait un pas en avant vers la preuve de la stabilité en dimension d. En effet, on déduit de cette finitude que si la fonction de coût $c:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ est à croissance au plus en exposant ρ, i.e. $|c(x,y)|\le K(1+|x|^\rho+|y|^\rho)$ pour tout $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$ et un certain $K\in\mathbb{R}_+$, et $(\nu_n)_{n\in\mathbb{N}}\in\mathcal{P}_\rho(\mathbb{R}^d)^{\mathbb{N}}$ est une suite de mesures de probabilité dominant ν dans l'ordre convexe et convergeant vers celui-ci en distance $\mathcal{W}_\rho$, alors $C_M(\mu,\nu_n)$ converge vers $C_M(\mu,\nu)$ lorsque $n\to+\infty$.

Lorsque $\mathbb{R}^d$ est munie de la norme L^ρ, l'inégalité de stabilité en dimension 1 se tensorise, c'est-à-dire que nous conservons $\mathcal{M}_\rho^\rho(\mu,\nu) \leq C_\rho \mathcal{W}_\rho^\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu)$ dès que μ et ν sont des lois de vecteurs aléatoires aux composantes indépendantes. Cette inégalité reste vraie aussi dans le cas où μ et ν sont reliées par une relation de changement d'échelle, c'est-à-dire qu'il existe $\lambda \in \mathbb{R}_+$ tel que pour X distribuée suivant μ, $X + \lambda(X - \mathbb{E}[X])$ est distribuée suivant ν, sous une condition supplémentaire sur la loi conditionnelle de X. Lorsque l'on relâche cette dernière contrainte, le cas du changement d'échelle nous renseigne sur les valeurs minimales des constantes recherchées.

Proposition 1.3.4. *Soient $d \in \mathbb{N}^*\backslash\{1\}$ et $\rho \in [1,+\infty[$. Indépendamment du choix de la norme sur $\mathbb{R}^d$, nous avons*

$$C_{1,d} \geq 3 \quad \textit{et} \quad \forall \rho \in]1,+\infty[, \quad C_{\rho,d} \geq 2 \vee \left(2^{\rho-1} \sup_{x\in]1,+\infty[} \frac{x+x^\rho}{(1+x)^\rho}\right).$$

1.3.3 Chapitre 4 : Réarrangements martingale de couplages en dimension 1

Soient deux mesures de probabilité distinctes μ et ν définies sur $\mathbb{R}$ de premier moment fini. On note $\Pi(\mu,\nu)$ l'ensemble des couplages entre μ et ν, c'est-à-dire l'ensemble des lois jointes de tout couple (X,Y) tel que X et Y soient distribuées respectivement selon μ et ν. Si de plus $\mu \leq_{cx} \nu$, on note $\Pi^M(\mu,\nu)$ l'ensemble des couplages martingale entre μ et ν, c'est-à-dire l'ensemble des lois jointes de tout couple (X,Y) tel que X et Y soient distribuées respectivement selon μ et ν, et $\mathbb{E}[Y|X] = X$ presque sûrement.

Pour prouver la stabilité du problème de transport optimal martingale en dimension 1, Wiesel [194] a abordé la formulation primale du problème. Pour ce faire, il développe la notion de projection de $\Pi(\mu,\nu)$ sur $\Pi^M(\mu,\nu)$, c'est-à-dire qu'il associe à un couplage P le couplage martingale M de mêmes marginales qui lui est le plus proche. Pour mesurer la distance entre deux couplages, on peut naturellement penser à la distance de Wasserstein. Mais la topologie induite par cette dernière ne convient pas à toutes les situations, surtout en mathématiques financières. En effet, la symétrie de cette distance ne prend pas en compte la structure temporelle des martingales. On peut facilement se convaincre que deux processus stochastiques de lois très proches en distance de Wasserstein peuvent tout de même induire des informations très différentes. Pour reprendre l'exemple particulièrement instructif de Backhoff-Veraguas, Bartl, Beiglböck et Eder [14], on peut considérer deux processus (X,Y) et (X',Y') qui vérifient, pour un certain $\varepsilon > 0$,

$$\begin{aligned} &\mathbb{P}(X=0,Y=-1) = \mathbb{P}(X=0,Y=1) = \frac{1}{2}, \\ \text{et} \quad &\mathbb{P}(X'=-\varepsilon,Y'=-1) = \mathbb{P}(X'=\varepsilon,Y'=1) = \frac{1}{2}\cdot \end{aligned}$$

Au sens de la distance de Wasserstein, ces lois sont très proches, ce que l'on peut confirmer très rapidement par le théorème 1.1.3 (iii) pour $\varepsilon \to 0$. Pourtant, l'observation de

X n'apporte aucune information sur la valeur de Y, alors que celle de X' permet de déterminer celle de Y' avec certitude. Pour rendre compte de cette différence d'information, il est nécessaire de renforcer ou d'adapter la topologie usuelle. Plusieurs approches ont alors été développées, comme la topologie faible adaptée, la topologie de l'information de Hellwig [101], la topologie faible étendue de Aldous [2], ou encore la topologie d'arrêt optimal [15]. De manière remarquable, toutes ces topologies sont en fait égales, du moins en temps discret [15]. On choisit donc dans le chapitre 4 de se focaliser sur la distance dite de Wasserstein adaptée d'ordre $\rho \in [1, +\infty[$, notée $\mathcal{AW}_\rho$. Soient $\mu, \nu, \mu', \nu' \in \mathcal{P}_\rho(\mathbb{R})$ et $\pi \in \Pi(\mu, \nu)$, $\pi' \in \Pi(\mu', \nu')$ des couplages de la forme $\pi(dx, dy) = \mu(dx)\, \pi_x(dy)$ et $\pi'(dx', dy') = \mu'(dx')\, \pi'_{x'}(dy')$. On définit alors la distance $\mathcal{AW}_\rho$ entre π et π' par

$$\mathcal{AW}_\rho(\pi, \pi') = \inf_{X\sim\mu,\ X'\sim\mu'} \mathbb{E}\left[|X - X'|^\rho + \mathcal{W}_\rho^\rho(\pi_X, \pi'_{X'})\right]^{\frac{1}{\rho}}. \tag{1.3.10}$$

On vérifie facilement que $\mathcal{AW}_\rho$ domine $\mathcal{W}_\rho$ et induit donc une topologie plus fine. Pour une étude approfondie de cette distance, on pourra consulter [149, 150, 151, 152, 129, 39]. On notera que Wiesel [194] travaille avec une distance dite de Wasserstein imbriquée, notée $\mathcal{W}_\rho^{nd}$, définie par le problème de Monge-Kantorovich restreint aux couplages bicausaux, comme cela est fait par Backhoff-Veraguas, Bartl, Beiglböck et Wiesel [16]. Par un lemme très simple, on montre l'égalité $\mathcal{AW}_\rho = \mathcal{W}_\rho^{nd}$.

Wiesel [194] définit donc un réarrangement martingale d'un couplage $P \in \Pi(\mu, \nu)$ comme un couplage martingale $M \in \Pi^M(\mu, \nu)$ qui minimise la distance $\mathcal{W}_1^{nd}(P, M)$. Toutefois il montre avec un contre-exemple qu'un tel réarrangement n'existe pas toujours. Pour garantir son existence, Wiesel définit une hypothèse dite de dispersion barycentrique, sous laquelle il montre qu'un couplage admet bien un réarrangement martingale. En particulier, le couplage de Hoeffding-Fréchet satisfait à cette hypothèse, et nous montrons que lorsqu'il existe un transport de Monge, le couplage martingale transformée inverse est un réarrangement martingale du couplage de Hoeffding-Fréchet. Dans le cas général, on montre que notre famille de couplages martingale du chapitre 2 contient toujours un réarrangement martingale du couplage de Hoeffding-Fréchet.

Proposition 1.3.5. *Soient $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ telles que $\mu \leq_{cx} \nu$, et P^{HF} le couplage de Hoeffding-Fréchet entre μ et ν. Alors il existe $Q \in \mathcal{Q}$ tel que la loi M^Q du couple (X^Q, Y^Q) défini par* (1.3.2) *soit un réarrangement martingale de P^{HF} :*

$$\mathcal{AW}_1(P^{HF}, M^Q) = \inf_{M\in\Pi^M(\mu,\nu)} \mathcal{AW}_1(P^{HF}, M).$$

Dans le cas d'un couplage quelconque $P \in \Pi(\mu, \nu)$ satisfaisant à l'hypothèse de dispersion barycentrique de Wiesel [194], nous construisons un couplage martingale dans le même esprit que le couplage martingale transformée inverse, et prouvons qu'il s'agit d'un réarrangement martingale de P. Wiesel proposait déjà une construction dans [194], mais qui n'est vraiment explicite que lorsque P est à support fini, et repose sur un passage à la limite sinon. La compréhension de notre construction, quant à elle, ne dépend pas de la finitude du support de P.

1.4 Partie 2 : Approximation de couplages martingale pour la distance de Wasserstein adaptée et applications

Cette partie est le fruit d'une collaboration avec Mathias Beiglböck, Benjamin Jourdain et Gudmund Pammer.

1.4.1 Chapitre 5 : Approximation de couplages martingale sur la droite réelle pour la distance de Wasserstein adaptée

Soient $\rho \in [1, +\infty[$, $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ et $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ convergeant respectivement vers μ et ν en $\mathcal{W}_\rho$. Il est alors bien connu que pour tout $P \in \Pi(\mu, \nu)$, il existe une suite de couplages $(P_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi(\mu_n, \nu_n)$ convergeant en $\mathcal{W}_\rho$ vers π. On dispose même d'un contrôle :

$$\inf_{P_n\in\Pi(\mu_n,\nu_n)} \mathcal{W}_\rho^\rho(P, P_n) \leq \mathcal{W}_\rho^\rho(\mu, \mu_n) + \mathcal{W}_\rho^\rho(\nu, \nu_n) \underset{n\to+\infty}{\longrightarrow} 0. \tag{1.4.1}$$

Cependant, pour des raisons évoquées au chapitre 4, la distance de Wasserstein n'est pas toujours la plus adaptée pour mesurer la distance entre des couplages, en particulier dans le cas martingale. L'enjeu du chapitre 5 est alors de montrer que tout couplage martingale $M \in \Pi^M(\mu, \nu)$ peut se faire approcher par une suite de couplages martingale $(M_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi^M(\mu_n, \nu_n)$ pour la distance $\mathcal{AW}_\rho$, sous réserve bien sûr que toutes les marginales considérées soient dans l'ordre convexe. Pour arriver à ce résultat, on montre dans un premier temps que l'approximation en $\mathcal{AW}_\rho$ est possible si l'on n'impose ni martingalité des couplages, ni ordre convexe entre les marginales.

Proposition 1.4.1. *Soient $\rho \in [1, +\infty[$ et $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ convergeant respectivement vers μ et ν en $\mathcal{W}_\rho$. Soit $P \in \Pi(\mu, \nu)$. Alors il existe une suite de couplages $(P_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi(\mu_n, \nu_n)$ qui converge vers P en $\mathcal{AW}_\rho$:*

$$\inf_{P_n\in\Pi(\mu_n,\nu_n)} \mathcal{AW}_\rho(P, P_n) \underset{n\to+\infty}{\longrightarrow} 0.$$

De plus, si pour tout $n \in \mathbb{N}$ et $x \in \mathbb{R}$ tel que $\mu_n(\{x\}) > 0$, il existe $x' \in \mathbb{R}$ tel que

$$F_\mu(x'-) \leq F_{\mu_n}(x-) < F_{\mu_n}(x) \leq F_\mu(x'),$$

ce qui est notamment toujours vérifié lorsque μ_n n'a pas d'atome, alors

$$\inf_{P_n\in\Pi(\mu_n,\nu_n)} \mathcal{AW}_\rho^\rho(P, P_n) \leq \mathcal{W}_\rho^\rho(\mu, \mu_n) + \mathcal{W}_\rho^\rho(\nu, \nu_n). \tag{1.4.2}$$

Supposons avec les notations de la proposition 1.4.1 que P soit un couplage martingale. Soient $n \in \mathbb{N}$ et X, Y, X_n, Y_n des variables aléatoires telles que $(X, Y) \sim P$ et $(X_n, Y_n) \sim P_n$.

En notant P_X et $(P_n)_X$ les lois conditionnelles respectives de Y sachant X et de Y_n sachant X_n, l'inégalité triangulaire et la martingalité de (X, Y) impliquent

$$\begin{aligned}\mathbb{E}[|X_n - \mathbb{E}[Y_n|X_n]|] &\leq \mathbb{E}[|X_n - X| + |X - \mathbb{E}[Y_n|X_n]|] = \mathbb{E}[|X_n - X| + |\mathbb{E}[Y|X] - \mathbb{E}[Y_n|X_n]|] \\ &\leq \mathbb{E}[|X_n - X| + \mathcal{W}_1((P_n)_{X_n}, P_X)].\end{aligned}$$

En choisissant (X, X_n) optimal pour $\mathcal{AW}_1(P, P_n)$ vu comme le problème de Monge-Kantorovich (1.3.10), on obtient ainsi

$$\mathbb{E}[|X_n - \mathbb{E}[Y_n|X_n]|] \leq \mathcal{AW}_1(P, P_n) \underset{n\to+\infty}{\longrightarrow} 0.$$

En ce sens, le couplage (X_n, Y_n) est presque martingale. Ainsi, sous l'hypothèse d'ordre convexe entre les marginales, il est naturel de s'attendre à ce que l'on puisse étendre ce résultat au cas martingale. Mais cela requiert en fait un travail considérable, qui fait précisément l'objet du théorème principal. Avant d'énoncer ce dernier, nous mentionnons que la proposition 1.4.1 se généralise au cas où les marginales sont définies sur des espaces polonais quelconques. Cette généralité se fait malheureusement au prix de la perte d'un contrôle de type (1.4.2). La définition de la distance $\mathcal{AW}_\rho$ donnée par (1.3.10) se généralise de manière évidente au cas de couplages définis sur un produit d'espaces polonais.

Proposition 1.4.2. *Soient $\rho \in [1, +\infty[$, E et F deux espaces polonais quelconques et $(\mu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(E)^\mathbb{N}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(F)^\mathbb{N}$ convergeant respectivement vers μ et ν en $\mathcal{W}_\rho$. Soit $P \in \Pi(\mu, \nu)$. Alors il existe une suite de couplages $(P_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi(\mu_n, \nu_n)$ qui converge vers P en $\mathcal{AW}_\rho$:*

$$\inf_{P_n\in\Pi(\mu_n,\nu_n)} \mathcal{AW}_\rho(P, P_n) \underset{n\to+\infty}{\longrightarrow} 0.$$

De retour à la dimension 1, notre résultat principal est le suivant.

Théorème 1.4.3. *Soient $\rho \in [1, +\infty[$, $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^\mathbb{N}$ convergeant respectivement vers μ et ν en $\mathcal{W}_\rho$, et telles que pour tout $n \in \mathbb{N}$, $\mu_n \leq_{cx} \nu_n$. Soit $M \in \Pi^M(\mu, \nu)$. Alors il existe une suite de couplages martingale $(M_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi^M(\mu_n, \nu_n)$ qui converge vers M en $\mathcal{AW}_\rho$:*

$$\inf_{M_n\in\Pi^M(\mu_n,\nu_n)} \mathcal{AW}_\rho(M, M_n) \underset{n\to+\infty}{\longrightarrow} 0.$$

La preuve de ce théorème est particulièrement technique. Pour la simplifier, on peut supposer sans perte de généralité que $\rho = 1$, par un raisonnement similaire à l'équivalence des points (i) et (ii) du théorème 1.1.3. On peut de plus supposer sans perte de généralité que (μ, ν) est irréductible. On rappelle que la fonction potentiel u_η d'une mesure de probabilité $\eta \in \mathcal{P}_1(\mathbb{R})$ est définie par $u_\eta : y \mapsto \int_\mathbb{R} |y - x|\, \eta(dx)$, et que (μ, ν) est dite irréductible si l'ensemble $I = \{x \in \mathbb{R} \mid u_\mu(x) < u_\nu(x)\}$ est un intervalle, et $\mu(I) = \nu(\overline{I}) = 1$. Lorsque $\mu \leq_{cx} \nu$, Beiglböck et Juillet [25] prouvent l'existence d'une décomposition en composantes irréductibles : il existe une famille $(I_n)_{n\in N}$ au plus dénombrable d'intervalles ouverts disjoints telle que pour tout couplage martingale (X, Y) entre μ et ν, on ait $\mathbb{P}(Y \in \overline{I_n}|X \in I_n) = 1$ pour tout $n \in N$.

Malgré ces simplications, la preuve du théorème 1.4.3 reste très technique. Pour le sentir, on donne sans plus d'explication la figure 1.1 ci-dessous, qui illustre une partie des découpages de l'intervalle $I = \{x \in \mathbb{R} \mid u_\mu(x) < u_\nu(x)\}$ impliqués dans notre démonstration.

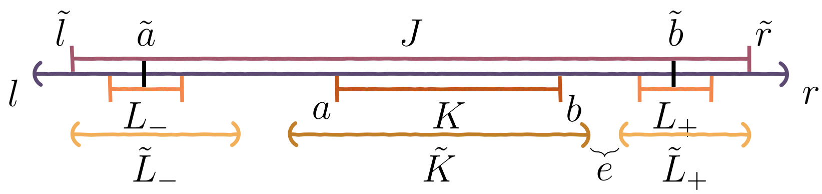

Figure 1.1: Points et intervalles impliqués dans la démonstration du théorème 1.4.3. Les extrémités des intervalles fermés sont repérées par des barres verticales, et celles des intervalles ouverts par des parenthèses.

Notons que la dernière étape de la démonstration de ce théorème repose sur l'inégalité de stabilité du théorème 1.3.1. En effet, on aboutit au cours de la démonstration à l'existence d'un couplage martingale $\mathring{M}_n$ proche de M en distance $\mathcal{AW}_1$ entre μ_n et une mesure de probabilité ν'_n dominée par ν_n dans l'ordre convexe et proche de celle-ci en distance $\mathcal{W}_1$. Il suffit alors de composer $\mathring{M}_n$ par le couplage martingale transformée inverse présenté dans le chapitre 2 pour aboutir à un couplage martingale proche de M entre μ_n et ν_n.

1.4.2 Chapitre 6 : Stabilité des problèmes de transport optimal faible et transport optimal martingale faible

Ce dernier chapitre est consacré aux applications du chapitre 5. Parmi elles, on retrouve la stabilité du problème de transport optimal pour des espaces polonais quelconques, ainsi que la stabilité du problème de transport optimal martingale en dimension 1. Plus précisément, considérons $\rho \in [1, +\infty[$, E et F deux espaces polonais quelconques, $(\mu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(E)^{\mathbb{N}}$ et $(\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(F)^{\mathbb{N}}$ deux suites de mesures de probabilité convergeant respectivement vers μ et ν en $\mathcal{W}_\rho$, et $c : E \times F \to \mathbb{R}$ une application continue telle que $|c(x, y)| \le K(1+|x|^\rho+|y|^\rho)$ pour tout $(x, y) \in E \times F$ et un certain $K \in \mathbb{R}_+$. On retrouve alors grâce à la proposition 1.4.2 la stabilité du problème de Monge-Kantorovich, à savoir

$$C'(\mu_n, \nu_n) \underset{n\to+\infty}{\to} C'(\mu, \nu).$$

Lorsque $E = F = \mathbb{R}$ et que pour tout $n \in \mathbb{N}$, $\mu_n \le_{cx} \nu_n$, nous retrouvons grâce au théorème 1.4.3 la stabilité du problème de transport optimal martingale en dimension 1 :

$$C_M(\mu_n, \nu_n) \underset{n\to+\infty}{\to} C_M(\mu, \nu).$$

En réalité, on prouve même la stabilité de problèmes plus généraux, à savoir de transport optimal faible pour des espaces polonais quelconques et de transport optimal martingale faible en dimension 1. On rappelle la formulation du problème de transport optimal faible, introduit par Gozlan, Roberto, Samson et Tetali [90] et étudié par ces mêmes auteurs avec Shu [89] : étant données deux mesures de probabilité μ et ν définies respectivement sur les espaces polonais E et F, et une application $c : E \times P(F) \to [0, +\infty]$ mesurable, déterminer

$$V(\mu,\nu) = \inf_{X\sim\mu,\ Y\sim\nu} \mathbb{E}[c(X, \mathcal{L}(Y|X))], \tag{1.4.3}$$

où $\mathcal{L}(Y|X)$ désigne la loi conditionnelle de Y sachant X. Lorsque $E = F = \mathbb{R}^d$ où $d \in \mathbb{N}^*$, et $\mu \leq_{cx} \nu$, le problème de transport optimal martingale faible consiste à déterminer

$$V_M(\mu,\nu) = \inf_{\substack{X\sim\mu,\ Y\sim\nu \\ \mathbb{E}[Y|X]=X \text{ p.s.}}} \mathbb{E}[c(X, \mathcal{L}(Y|X))]. \tag{1.4.4}$$

Depuis les travaux de Backhoff-Veraguas, Beiglböck et Pammer [18], et Backhoff-Veraguas et Pammer [20], la stabilité du problème (1.4.3) pour des espaces polonais quelconques en $\mathcal{AW}_\rho$ est connue. Nous retrouvons ce résultat sans le recours à l'outil de monotonie martingale. De plus, nous établissons un résultat de stabilité du problème 1.4.4 en dimension 1 en $\mathcal{AW}_\rho$.

Théorème 1.4.4. *Soient $\rho \in [1, +\infty[$ et $c : \mathbb{R} \times \mathcal{P}_\rho(\mathbb{R}) \to \mathbb{R}$ continue telle que*

$$\exists K > 0, \quad \forall (x,p) \in \mathbb{R} \times \mathcal{P}_\rho(\mathbb{R}), \quad |c(x,p)| \leq K \left(1 + |x|^\rho + \int_{\mathbb{R}} |y|^\rho \, p(dy)\right).$$

Soient $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ convergeant respectivement vers $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ en $\mathcal{W}_\rho$, et telles que $\mu_n \leq_{cx} \nu_n$ pour tout $n \in \mathbb{N}$. Alors

$$V_M(\mu_n, \nu_n) \underset{n\to+\infty}{\to} V_M(\mu,\nu).$$

Si de plus c est strictement convexe en son second argument, alors l'unique loi d'un minimiseur de $V_M(\mu_n,\nu)$ converge en $\mathcal{AW}_\rho$ vers l'unique loi d'un minimiseur de $V_M(\mu,\nu)$ lorque $n \to +\infty$.

Le théorème 1.4.4 est donc formulé en dimension 1. Sa démonstration s'adapte de manière immédiate à la dimension supérieure, sous réserve que le théorème 1.4.3 soit lui-même généralisé à la dimension supérieure. Pour retrouver la stabilité du problème de transport optimal martingale avec une application de coût $c : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, il suffit d'appliquer le théorème 1.4.4 avec le coût $(x,p) \mapsto \int_{\mathbb{R}} c(x,y) \, p(dy)$. Le théorème 1.4.4 a aussi pour conséquence la stabilité d'un problème de transport optimal martingale avec une formulation de type Benamou-Brenier suggérée par Backhoff-Veraguas, Beiglböck, Huesmann et Källblad [17]. La solution à ce problème est appelée mouvement brownien étiré, dont nous prouvons la stabilité trajectorielle. Nous utilisons aussi le théorème 1.4.4 pour montrer que la monotonie martingale est une condition suffisante à l'optimalité du problème (1.4.4) sous des conditions assez générales.

Citons enfin une dernière application du théorème 1.4.4, à savoir la stabilité par rapport aux marginales du prix de sur-réplication de contrats à termes sur le VIX. Le VIX, abréviation de *Volatility Index*, est une mesure couramment utilisée pour déterminer le sentiment que les investisseurs ont d'un marché. Par définition, le VIX est la volatilité implicite du S&P 500 calculée sur un horizon de 30 jours. Typiquement, le VIX croît lorsque les investisseurs s'attendent à des variations brutales du marché, et tendent alors à acheter plus d'options, ce qui fait augmenter la demande et les rend ainsi plus chères. Le marché est alors dit volatile. Réciproquement, le VIX décroît lorsque la demande en options est plus faible, ce qui traduit un marché perçu comme calme.

Considérons un marché financier composé de deux actifs financiers : l'actif sans risque, et le S&P 500 $(S_t)_{t\in\{T_1,T_2\}}$, qui peuvent s'échanger à des dates T_1 et $T_2 = T_1 + 30$ jours. On suppose connus les cours des calls pour tout prix d'exercices $K \geq 0$, ce qui en vertu de la formule de Breeden-Litzenberger [46] permet de connaître les lois respectives μ et ν de S_{T_1} et S_{T_2}. On autorise l'échange à la date 0 d'options vanille de maturités T_1 et T_2, et l'échange à la date T_1 de S&P 500 et de contrats à terme de payoff $\frac{-2}{T_2-T_1}\ln\frac{S_{T_2}}{S_{T_1}}$ en T_2. Dans ce cadre, Guyon, Menegaux and Nutz [99] expriment la borne supérieure des prix de contrats à terme sur VIX expirant en T_1 issus d'un modèle quelconque sans opportunité d'arbitrage, comme le plus petit prix de sur-réplication au temps 0:

$$P_{\text{super}}(\mu,\nu) = \inf\left(\mathbb{E}[u_1(S_{T_1})] + \mathbb{E}[u_2(S_{T_2})]\right), \tag{1.4.5}$$

où l'infimum porte sur tous les coupls $(u_1,u_2) \in L^1(\mu)\times L^1(\nu)$ et les applications mesurables Δ^S, Δ^L tels que pour tout $(x,y,v) \in]0,+\infty[^2\times[0,+\infty[$,

$$u_1(x) + u_2(y) + \Delta^S(x,v)(y-x) + \Delta^L(x,v)\left(-\frac{2}{T_2-T_1}\ln\frac{y}{x} - v^2\right) \geq v. \tag{1.4.6}$$

Notons que le problème primal $P_{\text{super}}(\mu,\nu)$ implique dans (1.4.6) trois variables x, y, s, qui représentent respectivement le S&P 500 au temps T_1, le S&P 500 au temps T_2, et le VIX au temps T_1. On s'attendrait alors naturellement à ce que la formulation duale implique trois marginales. De manière remarquable, elle prend en fait la forme d'un problème de transport optimal martingale faible à deux marginales seulement, grâce à la concavité de la racine carrée, comme le montrent Guyon, Menegaux et Nutz [99, Proposition 4.10]. On a alors

$$P_{\text{super}}(\mu,\nu) = V_M(\mu,\nu),$$

où $V_M(\mu,\nu)$ est la solution du problème de transport optimal martingale faible (1.4.4) associé au coût

$$c :]0,+\infty[\times\mathcal{P}_1(]0,+\infty[) \ni (x,p) \mapsto \sqrt{-\frac{2}{T_2-T_1}\int_{]0,+\infty[}\ln\left(\frac{y}{x}\right)\, p(dy)}.$$

On en déduit la stabilité du prix de sur-réplication de contrats à termes sur le VIX.

Corollaire 1.4.5. *Soient $\rho \in [1,+\infty[$ et $(\mu_n)_{n\in\mathbb{N}}, (\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ convergeant respectivement vers $\mu,\nu \in \mathcal{P}_\rho(\mathbb{R})$ en $\mathcal{W}_\rho$, et telles que $\mu_n \leq_{cx} \nu_n$ pour tout $n \in \mathbb{N}$. Alors*

$$P_{super}(\mu_n,\nu_n) \underset{n\to+\infty}{\longrightarrow} P_{super}(\mu,\nu).$$

Part I
A new family of martingale couplings and stability inequalities

Chapter 2

A new family of one dimensional martingale couplings

Abstract

In this paper, we exhibit a new family of martingale couplings between two one-dimensional probability measures μ and ν in the convex order. This family is parametrised by two dimensional probability measures on the unit square with respective marginal densities proportional to the positive and negative parts of the difference between the quantile functions of μ and ν. It contains the inverse transform martingale coupling which is explicit in terms of the quantile functions of these marginal densities. The integral of $|x-y|$ with respect to each of these couplings is smaller than twice the $\mathcal{W}_1$ distance between μ and ν. When the comonotonous coupling between μ and ν is given by a map T, the elements of the family minimise $\int_{\mathbb{R}} |y-T(x)|\, M(dx,dy)$ among all martingale couplings between μ and ν. When μ and ν are in the decreasing (resp. increasing) convex order, the construction is generalised to exhibit super (resp. sub) martingale couplings.

Keywords: Convex order, Martingale Optimal Transport, Wasserstein distance, Martingale couplings.

AMS MSC 2010: 60G42, 60E15, 91G80.

2.1 Introduction

For all $d \in \mathbb{N}^*$, $\rho \geq 1$ and μ, ν in the set $\mathcal{P}_\rho(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with finite ρ-th moment, we define the Wasserstein distance with index ρ by $\mathcal{W}_\rho(\mu,\nu) = (\inf_{P \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\rho \, P(dx,dy))^{1/\rho}$, where $\Pi(\mu,\nu)$ denotes the set of couplings between μ and ν, that is $\Pi(\mu,\nu) = \{P \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) \mid \forall A \in \mathcal{B}(\mathbb{R}^d),\ P(A \times \mathbb{R}^d) = \mu(A) \text{ and } P(\mathbb{R}^d \times A) = \nu(A)\}$. Let $\Pi^{\mathrm{M}}(\mu,\nu)$ be the set of martingale couplings between μ and ν, that is $\Pi^{\mathrm{M}}(\mu,\nu) = \{M \in \Pi(\mu,\nu) \mid \mu(dx)\text{-a.e.}, \int_{\mathbb{R}^d} |y| \, m(x,dy) < +\infty \text{ and } \int_{\mathbb{R}^d} y \, m(x,dy) = x\}$. The celebrated Strassen theorem [183] ensures that if $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, then $\Pi^{\mathrm{M}}(\mu,\nu) \neq \emptyset$ iff μ and ν are in the convex order. We recall that two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ are in the convex order, and denote $\mu \leq_{cx} \nu$, if $\int_{\mathbb{R}^d} f(x) \, \mu(dx) \leq \int_{\mathbb{R}^d} f(x) \, \nu(dx)$ for any convex function $f : \mathbb{R}^d \to \mathbb{R}$. We denote $\mu <_{cx} \nu$ if $\mu \leq_{cx} \nu$ and $\mu \neq \nu$. For all $\rho \geq 1$ and $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$, we define $\mathcal{M}_\rho(\mu,\nu)$ by

$$\mathcal{M}_\rho(\mu,\nu) = \left(\inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\rho \, M(dx,dy) \right)^{1/\rho}.$$

Our main result is the following stability inequality which shows that if μ and ν are in the convex order and close to each other, then there exists a martingale coupling which expresses this proximity:

$$\forall \mu, \nu \in \mathcal{P}_1(\mathbb{R}) \text{ such that } \mu \leq_{cx} \nu, \quad \mathcal{M}_1(\mu,\nu) \leq 2\mathcal{W}_1(\mu,\nu). \tag{2.1.1}$$

It is well known (see for instance Remark 2.19 (ii) Chapter 2 [190]) that for all $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$,

$$\mathcal{W}_\rho(\mu,\nu) = \left(\int_0^1 \left| F_\mu^{-1}(u) - F_\nu^{-1}(u) \right|^\rho du \right)^{1/\rho}, \tag{2.1.2}$$

where we denote by $F_\eta(x) = \eta((-\infty,x])$ and $F_\eta^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_\eta(x) \geq u\}, u \in (0,1)$, the cumulative distribution function and the quantile function of a probability measure η on $\mathbb{R}$. We prove the inequality (2.1.1) by exhibiting a new family of martingale couplings M such that $\int_{\mathbb{R}\times\mathbb{R}} |x-y| \, M(dx,dy) \leq 2\mathcal{W}_1(\mu,\nu)$. We will show (see the proof of Theorem 2.2.12) that the constant 2 is sharp in (2.1.1). We will also see that (2.1.1) cannot be generalised with $\mathcal{M}_1(\mu,\nu)$ and $\mathcal{W}_1(\mu,\nu)$ replaced with $\mathcal{M}_\rho(\mu,\nu)$ and $\mathcal{W}_\rho(\mu,\nu)$ for $\rho > 1$. The case $\rho = 2$ is easy, since for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ and $M \in \Pi^{\mathrm{M}}(\mu,\nu)$, $\int_{\mathbb{R}\times\mathbb{R}} |x-y|^2 \, M(dx,dy) = \int_{\mathbb{R}} y^2 \, \nu(dy) - \int_{\mathbb{R}} x^2 \, \mu(dx)$, which is independent from M. For all $n \in \mathbb{N}^*$, let μ_n be the centred Gaussian distribution with variance n^2. Then we get that $\mathcal{M}_2(\mu_n,\mu_{n+1}) = \sqrt{2n+1} \underset{n\to+\infty}{\longrightarrow} +\infty$, whereas $\mathcal{W}_\rho(\mu_n,\mu_{n+1}) = (\int_0^1 |nF_{\mu_1}^{-1}(u) - (n+1)F_{\mu_1}^{-1}(u)|^\rho \, du)^{1/\rho} = \mathbb{E}[|G|^\rho]^{1/\rho} < +\infty$ for $G \sim \mathcal{N}_1(0,1)$, which makes the equivalent of (2.1.1) impossible to hold. Extension to the case $\rho > 2$ is immediate with the same example thanks to Jensen's inequality which provides $\mathcal{M}_\rho(\mu_n,\mu_{n+1}) \geq \mathcal{M}_2(\mu_n,\mu_{n+1}) = \sqrt{2n+1}$, whereas $\mathcal{W}_\rho(\mu_n,\mu_{n+1})$ is still bounded.

This problem is motivated by the resolution of the Martingale Optimal Transport (MOT) problem introduced by Beiglböck, Henry-Labordère and Penkner [23] in a discrete time setting, and Galichon, Henry-Labordère and Touzi [84] in a continuous time setting. For adaptations of celebrated results on classical optimal transport theory to the MOT problem, we refer to Henry-Labordère, Tan and Touzi [105] and Henry-Labordère and Touzi [106]. To tackle numerically the MOT problem, we refer to Alfonsi, Corbetta and Jourdain [3], Alfonsi, Corbetta and Jourdain [4], De March [66] and Guo and Obłój [96]. On duality, we refer to Beiglböck, Nutz and Touzi [27], Beiglböck, Lim and Obłój [26] and De March [68]. We also refer to De March [67] and De March and Touzi [69] for the multi-dimensional case. Once the martingale optimal transport problem is discretised by approximating μ and ν by probability measures with finite support and in the convex order, one can raise the question of the convergence of the discrete optimal cost towards the continuous one. The present paper is a step forward in proving the stability of the martingale optimal transport problem with respect to the marginals.

We develop in Section 2.2 an abstract construction of a new family of martingale couplings between two probability measures μ and ν on the real line with finite first moments and comparable in the convex order. This family is parametrised by two dimensional probability measures on the unit square with respective marginal densities proportional to the positive and the negative parts of the difference $F_\mu^{-1} - F_\nu^{-1}$ between the quantile functions of μ and ν. Moreover, each martingale coupling in the family is obtained as the image of $\mathbb{1}_{(0,1)}(u)\,du\,\widetilde{m}^Q(u,dy)$ by $(u,y)\mapsto(F_\mu^{-1}(u),y)$ where $\widetilde{m}^Q$ is a Markov kernel on $(0,1)\times\mathbb{R}$ such that $\int_{(0,1)}\widetilde{m}^Q(u,\{y\in\mathbb{R}\mid |y-F_\nu^{-1}(u)|=(y-F_\nu^{-1}(u))\mathrm{sg}(F_\mu^{-1}(u)-F_\nu^{-1}(u))\})du=1$, where $\mathrm{sg}(x)=\mathbb{1}_{\{x>0\}}-\mathbb{1}_{\{x<0\}}$ for $x\in\mathbb{R}$. Therefore, for (U,Y) distributed according to $\mathbb{1}_{(0,1)}(u)\,du\,\widetilde{m}^Q(u,dy)$, $(F_\mu^{-1}(U),Y)$ is a martingale coupling and

$$\begin{aligned}\mathbb{E}[|Y-F_\nu^{-1}(U)|] &= \mathbb{E}[\mathrm{sg}(F_\mu^{-1}(U)-F_\nu^{-1}(U))\mathbb{E}[Y-F_\nu^{-1}(U)|U]] \\ &= \mathbb{E}[|F_\mu^{-1}(U)-F_\nu^{-1}(U)|]=\mathcal{W}_1(\mu,\nu).\end{aligned} \tag{2.1.3}$$

When the comonotonous coupling between μ and ν, that is the law of $(F_\mu^{-1}(U),F_\nu^{-1}(U))$, is given by a map T, the elements of the family minimise $\int_{\mathbb{R}}|y-T(x)|\,M(dx,dy)$ among all martingale couplings between μ and ν. We deduce from (2.1.3) that $\mathbb{E}[|Y-F_\mu^{-1}(U)|]\le\mathbb{E}[|Y-F_\nu^{-1}(U)|]+\mathbb{E}[|F_\nu^{-1}(U)-F_\mu^{-1}(U)|]=2\mathcal{W}_1(\mu,\nu)$ which implies (2.1.1) as soon as the parameter set of probability measures on the unit square is non empty.

In Section 2.3, we give an explicit example of such a probability measure on the unit square. We call the associated martingale coupling the inverse transform martingale coupling. This coupling is explicit in terms of the cumulative distribution functions of the above-mentioned densities and their left-continuous generalised inverses. It is therefore more explicit than the left-curtain (and right-curtain) coupling introduced by Beiglböck and Juillet [25] which under the condition that ν has no atoms and the set of local maximal values of $F_\nu-F_\mu$ is finite can be made explicit according to Henry-Labordère and Touzi [106] by solving two coupled ordinary differential equations starting from each right-most local maximiser. We also check that the inverse transform martingale coupling is stable with respect to its marginals μ and ν for the Wasserstein distance. The building brick of the inverse transform martingale coupling is a martingale coupling between $\mu_{u,v}=p\delta_{F_\mu^{-1}(u)}+(1-p)\delta_{F_\mu^{-1}(v)}$ and $\nu_{u,v}=p\delta_{F_\nu^{-1}(u)}+(1-p)\delta_{F_\nu^{-1}(v)}$ with $0<u<v<1$ such that

$$F_\nu^{-1}(u)<F_\mu^{-1}(u)<F_\mu^{-1}(v)<F_\nu^{-1}(v), \tag{2.1.4}$$

where we choose a common weight p (resp. $1-p$) for $F_\mu^{-1}(u)$ and $F_\nu^{-1}(u)$ (resp. $F_\mu^{-1}(v)$ and $F_\nu^{-1}(v)$) to help ensuring that the second marginal is equal to ν when the first is equal to μ. Then p is given by the equality of the means which in view of the condition (2.1.4) on the supports is equivalent to the convex order between $\mu_{u,v}$ and $\nu_{u,v}$: $\frac{1-p}{p} = \frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\mu^{-1}(v)}$. We rely on the necessary condition of Theorem 3.A.5 Chapter 3 [169]: $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ are such that $\mu \leq_{cx} \nu$ iff for all $u \in [0,1]$, $\int_0^u F_\mu^{-1}(v)\,dv \geq \int_0^u F_\nu^{-1}(v)dv$ with equality for $u=1$. This implies that for all $u \in [0,1]$, $\Psi_+(u) := \int_0^u (F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv \geq \int_0^u (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv := \Psi_-(u)$ where $x^+ := \max(x,0)$ and $x^- := \max(-x,0)$ respectively denote the positive and negative parts of a real number x. We now choose $v = \Psi_-^{-1}(\Psi_+(u))$ where Ψ_-^{-1} is the left-continuous generalised inverse of Ψ_-. Then $d\Psi_+(u)$ a.e. $u < v$ (consequence of $\Psi_- \leq \Psi_+$) and $F_\nu^{-1}(u) < F_\mu^{-1}(u) < F_\mu^{-1}(v) < F_\nu^{-1}(v)$ (consequence of the definitions of Ψ_+ and Ψ_-, see Section 2.3.1). Moreover the key equality $\frac{dv}{du} = \frac{(F_\mu^{-1}-F_\nu^{-1})^+(u)}{(F_\mu^{-1}-F_\nu^{-1})^-(v)} = \frac{1-p}{p}$ explains why the construction succeeds. More details are given in Section 2.3.

The cardinality of this new family of martingale couplings between μ and ν is discussed in Section 2.4. This family is shown to be convex and is therefore either a singleton like when ν only weighs two points, or uncountably infinite like when $\mu(\{x\}) = \nu(\{x\}) = 0$ for all $x \in \mathbb{R}$.

The construction is finally generalised in Section 2.5 to exhibit super (resp. sub) martingale couplings as soon as μ is smaller than ν in the decreasing (resp. increasing) convex order. We recall that two probability measures $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ are in the decreasing (resp. increasing) convex order and denote $\mu \leq_{dcx} \nu$ (resp. $\mu \leq_{icx} \nu$) if $\int_\mathbb{R} f(x)\,\mu(dx) \leq \int_\mathbb{R} f(x)\,\nu(dx)$ for any decreasing (resp. increasing) convex function $f : \mathbb{R} \to \mathbb{R}$. In particular, we generalise the stability inequality to the super (resp. sub) martingale case.

Throughout the present article, a capital letter M which denotes a coupling between μ and ν is associated to its small letter m which denotes the regular conditional probability distribution of M with respect to μ, that is the (μ-a.e.) unique Markov kernel such that $M(dx,dy) = \mu(dx)\,m(x,dy)$.

Acknowledgements: We thank Jean-François Delmas (CERMICS) for numerous and fruitful discussions.

2.2 A new family of martingale couplings

2.2.1 A simple example

Let us construct a coupling in dimension 1 which shows that (2.1.1) holds true in a simple case. We say that a centred probability measure $\mu \in \mathcal{P}_1(\mathbb{R})$ is symmetric if $\mu = \bar{\mu}$, where $\bar{\mu}$ denotes the image of μ by $x \mapsto -x$. Let then μ and ν be centred and symmetric probability measures on $\mathbb{R}$ such that $F_\mu^{-1}(u) \geq F_\nu^{-1}(u)$ for all $u \in (0,1/2]$ and $F_\mu^{-1}(u) \leq F_\nu^{-1}(u)$ for all $u \in (1/2,1)$. Let U be a random variable uniformly distributed on $(0,1)$. According to the inverse transform sampling, the probability distributions of $F_\mu^{-1}(U)$ and $F_\nu^{-1}(U)$ are respectively μ and ν. Let Y be the random variable defined by

$$Y = F_\nu^{-1}(U)\mathbb{1}_{\{F_\nu^{-1}(U)\neq 0, V\leq \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)}{2F_\nu^{-1}(U)}\}} - F_\nu^{-1}(U)\mathbb{1}_{\{F_\nu^{-1}(U)\neq 0, V> \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)}{2F_\nu^{-1}(U)}\}}, \tag{2.2.1}$$

where V is a random variable uniformly distributed on $(0,1)$ independent from U. It is clear by symmetry of μ that $F_\mu(0)\geq 1/2$, so $F_\mu^{-1}(1/2)\leq 0$. Moreover, for all $x\in\mathbb{R}$ and $u>1/2$, $F_\mu(x)\geq u$ implies $x\geq 0$, so $F_\mu^{-1}(u)\geq 0$. Therefore, we have

$$\forall u\in(0,1/2],\quad F_\nu^{-1}(u)\leq F_\mu^{-1}(u)\leq 0 \quad\text{and}\quad \forall u\in(1/2,1),\quad 0\leq F_\mu^{-1}(u)\leq F_\nu^{-1}(u). \tag{2.2.2}$$

In particular, when $F_\nu^{-1}(U)=0$, then $F_\mu^{-1}(U)=0$ and $Y=0$. Let us check that Y is distributed according to ν. Using that $(F_\mu^{-1}(U),F_\nu^{-1}(U))$ and $(-F_\mu^{-1}(U),-F_\nu^{-1}(U))$ are identically distributed (see Lemma 2.6.5 below in Section 2.6), we have for all measurable and bounded functions $h:\mathbb{R}\to\mathbb{R}$,

$$\begin{aligned}\mathbb{E}[h(Y)] &= \mathbb{E}\left[h(F_\nu^{-1}(U))\mathbb{1}_{\{F_\nu^{-1}(U)\neq 0, V\leq \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)}{2F_\nu^{-1}(U)}\}}\right]\\ &\quad + \mathbb{E}\left[h(F_\nu^{-1}(U))\mathbb{1}_{\{F_\nu^{-1}(U)\neq 0, V> \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)}{2F_\nu^{-1}(U)}\}}\right] + h(0)\mathbb{P}(F_\nu^{-1}(U)=0)\\ &= \mathbb{E}[h(F_\nu^{-1}(U))].\end{aligned}$$

Moreover, according to (2.2.2), we have $\frac{F_\mu^{-1}(u)+F_\nu^{-1}(u)}{2F_\nu^{-1}(u)}\in[0,1]$ for all $u\in(0,1)$ such that $F_\nu^{-1}(u)\neq 0$. In addition to that, we have $F_\nu^{-1}(u)\frac{F_\mu^{-1}(u)+F_\nu^{-1}(u)}{2F_\nu^{-1}(u)} - F_\nu^{-1}(u)\left(1-\frac{F_\mu^{-1}(u)+F_\nu^{-1}(u)}{2F_\nu^{-1}(u)}\right) = F_\mu^{-1}(u)$ for all $u\in(0,1)$ such that $F_\nu^{-1}(u)\neq 0$. So $\mathbb{E}[Y|U] = F_\mu^{-1}(U)\mathbb{1}_{\{F_\nu^{-1}(U)\neq 0\}} = F_\mu^{-1}(U)$ since $F_\nu^{-1}(U)=0$ implies $F_\mu^{-1}(U)=0$. So we deduce that $\mathbb{E}[Y|F_\mu^{-1}(U)]=F_\mu^{-1}(U)$. Therefore, the law of $(F_\mu^{-1}(U),Y)$ is an explicit martingale coupling between μ and ν.

Furthermore, remarking that $|Y-F_\nu^{-1}(U)| = (Y-F_\nu^{-1}(U))\mathrm{sg}(F_\mu^{-1}(U)-F_\nu^{-1}(U))$, we deduce from the equality (2.1.3) that $\mathbb{E}[|Y-F_\mu^{-1}(U)|]\leq \mathbb{E}[|Y-F_\nu^{-1}(U)|]+\mathbb{E}[|F_\nu^{-1}(U)-F_\mu^{-1}(U)|] = 2\mathcal{W}_1(\mu,\nu)$, so (2.1.1) holds.

2.2.2 Definition

Let μ and ν be two probability measures on $\mathbb{R}$ with finite first moment such that $\int_\mathbb{R} x\,\mu(dx) = \int_\mathbb{R} y\,\nu(dy)$ and $\mu\neq\nu$. We recall that Ψ_+ and Ψ_- are defined for all $u\in[0,1]$ by $\Psi_+(u) = \int_0^u (F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv$ and $\Psi_-(u) = \int_0^u (F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv$. Let $\mathcal{U}_+$, $\mathcal{U}_-$ and $\mathcal{U}_0$ be defined by

$$\mathcal{U}_+ = \{u\in(0,1)\mid F_\mu^{-1}(u)>F_\nu^{-1}(u)\},\quad \mathcal{U}_- = \{u\in(0,1)\mid F_\mu^{-1}(u)<F_\nu^{-1}(u)\}$$
$$\text{and}\quad \mathcal{U}_0 = \{u\in(0,1)\mid F_\mu^{-1}(u)=F_\nu^{-1}(u)\}. \tag{2.2.3}$$

Notice that $d\Psi_+(u)$-a.e. (resp. $d\Psi_-(u)$-a.e.), we have $u\in\mathcal{U}_+$ (resp. $u\in\mathcal{U}_-$). Since μ and ν have equal means, we can set $\gamma = \int_0^1 (F_\mu^{-1}-F_\nu^{-1})^+(u)\,du = \int_0^1 (F_\mu^{-1}-F_\nu^{-1})^-(u)\,du\in(0,+\infty)$. We note $\mathcal{Q}$ the set of probability measures $Q(du,dv)$ on $(0,1)^2$ such that

(i) Q has first marginal $\frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du = \frac{1}{\gamma}\,d\Psi_+(u)$;

(ii) Q has second marginal $\frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv = \frac{1}{\gamma}\,d\Psi_-(v)$;

(iii) $Q\left(\{(u,v) \in (0,1)^2 \mid u < v\}\right) = 1$.

Example 2.2.1. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Suppose that the difference of the quantile functions changes sign only once, that is there exists $p \in (0,1)$ such that $u \mapsto \int_0^u (F_\mu^{-1}(v) - F_\nu^{-1}(v))\,dv$ is nondecreasing on $[0,p]$ and nonincreasing on $[p,1]$. Then one can easily see that any probability measure Q defined on $(0,1)$ satisfying properties (i) and (ii) of the definition of $\mathcal{Q}$ is concentrated on $(0,p) \times (p,1)$ and therefore satisfies (iii). In particular, the probability measure Q_1 defined on $(0,1)^2$ by

$$Q_1(du,dv) = \frac{1}{\gamma^2}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv \tag{2.2.4}$$

is an element of $\mathcal{Q}$.

In view of (i) and (ii), one could rewrite (iii) as $Q\left(\{(u,v) \in \mathcal{U}_+ \times \mathcal{U}_- \mid u < v\}\right) = 1$. A characterisation of the support of Q in terms of the irreducible components of μ and ν is given by Proposition 2.2.8 below. In the general case, the construction of a probability measure $Q \in \mathcal{Q}$ is not straightforward, but a direct consequence of Proposition 2.3.1 below is that $\mathcal{Q}$ is non-empty as long as $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are such that $\mu <_{cx} \nu$. Moreover, the convexity of $\mathcal{Q}$ is clear.

Proposition 2.2.2. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Then $\mathcal{Q}$ is a non-empty convex set.*

Let Q be an element of $\mathcal{Q}$. Let π_-^Q and π_+^Q be two sub-Markov kernels on $(0,1)$ such that for du-almost all $u \in \mathcal{U}_+$ and dv-almost all $v \in \mathcal{U}_-$, $\pi_+^Q(u,(0,1)) = 1$, $\pi_-^Q(v,(0,1)) = 1$ and

$$Q(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi_+^Q(u,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv\,\pi_-^Q(v,du).$$

Let $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ be the Markov kernel defined by

$$\begin{cases} \displaystyle\int_{v\in(0,1)} \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\,\delta_{F_\nu^{-1}(v)}(dy)\,\pi_+^Q(u,dv) + \int_{v\in(0,1)} \frac{F_\nu^{-1}(v) - F_\mu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\,\pi_+^Q(u,dv)\,\delta_{F_\nu^{-1}(u)}(dy) \\ \qquad\qquad \text{for } u \in \mathcal{U}_+ \text{ such that } \pi_+^Q(u,\{v \in (0,1) \mid F_\nu^{-1}(v) > F_\mu^{-1}(u)\}) = 1; \\ \displaystyle\int_{v\in(0,1)} \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\,\delta_{F_\nu^{-1}(v)}(dy)\,\pi_-^Q(u,dv) + \int_{v\in(0,1)} \frac{F_\nu^{-1}(v) - F_\mu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\,\pi_-^Q(u,dv)\,\delta_{F_\nu^{-1}(u)}(dy) \\ \qquad\qquad \text{for } u \in \mathcal{U}_- \text{ such that } \pi_-^Q(u,\{v \in (0,1) \mid F_\nu^{-1}(v) < F_\mu^{-1}(u)\}) = 1; \\ \delta_{F_\nu^{-1}(u)}(dy) \qquad \text{otherwise.} \end{cases} \tag{2.2.5}$$

For any Markov kernel $(\widetilde{m}(u,dy))_{u\in(0,1)}$, we denote by $(m(x,dy))_{x\in\mathbb{R}}$ the Markov kernel defined by

$$\begin{cases} \delta_x(dy) & \text{if } F_\mu(x)=0 \text{ or } F_\mu(x_-)=1; \\ \dfrac{1}{\mu(\{x\})}\displaystyle\int_{u=F_\mu(x_-)}^{F_\mu(x)} \widetilde{m}(u,dy)\,du & \text{if } \mu(\{x\})>0; \\ \widetilde{m}(F_\mu(x),dy) & \text{otherwise.} \end{cases} \tag{2.2.6}$$

For all $x\in\mathbb{R}$ such that $F_\mu(x)>0$ and $F_\mu(x_-)<1$, $m(x,dy)$ can be rewritten as

$$m(x,dy)=\int_{v=0}^{1}\widetilde{m}(F_\mu(x_-)+v(F_\mu(x)-F_\mu(x_-)),dy)\,dv. \tag{2.2.7}$$

Conversely, let $(p(x,dy))_{x\in\mathbb{R}}$ be a Markov kernel. Let then $(\widetilde{m}(u,dy))_{u\in(0,1)}$ be the Markov kernel defined for all $u\in(0,1)$ by $\widetilde{m}(u,dy)=p(F_\mu^{-1}(u),dy)$. Let $(m(x,dy))_{x\in\mathbb{R}}$ be the Markov kernel defined by (2.2.6). Let $x\in\mathbb{R}$ be such that $F_\mu(x_-)>0$ and $F_\mu(x)<1$. If $\mu(\{x\})>0$, then for all $u\in(F_\mu(x_-),F_\mu(x)]$, $F_\mu^{-1}(u)=x$. Hence $m(x,dy)=\frac{1}{\mu(\{x\})}\int_{u=F_\mu(x_-)}^{F_\mu(x)}\widetilde{m}(u,dy)\,du=\frac{1}{\mu(\{x\})}\int_{u=F_\mu(x_-)}^{F_\mu(x)}p(x,dy)\,du=p(x,dy)$. By Lemma 2.6.3 below, $F_\mu^{-1}(F_\mu(x))=x$, $\mu(dx)$-almost everywhere. So for $\mu(dx)$-almost all $x\in\mathbb{R}$ such that $F_\mu(x_-)>0$, $F_\mu(x)<1$ and $\mu(\{x\})=0$, $m(x,dy)=p(F_\mu^{-1}(F_\mu(x)),dy)=p(x,dy)$. Therefore, for $\mu(dx)$-almost all $x\in\mathbb{R}$, $p(x,dy)=m(x,dy)$.

Throughout the present article, for any $Q\in\mathcal{Q}$, $(m^Q(x,dy))_{x\in\mathbb{R}}$ and M^Q will respectively denote the Markov kernel given by (2.2.6) when $(\widetilde{m}(u,dy))_{u\in(0,1)}=(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ and the probability measure on $\mathbb{R}^2$ defined by $M^Q(dx,dy)=\mu(dx)\,m^Q(x,dy)$.

Proposition 2.2.3. *Let μ and ν be two distinct probability measures on $\mathbb{R}$ with finite first moment and equal means such that $\mathcal{Q}$ is non-empty. Let $Q\in\mathcal{Q}$. Then the probability measure M^Q is a martingale coupling between μ and ν.*

One can easily check thanks to Jensen's inequality that the existence of a martingale coupling between μ and ν implies that $\mu\le_{cx}\nu$ (see Remark 2.3.2 for a proof). A direct consequence of the latter fact and the last two propositions is an easy characterisation of the emptiness of $\mathcal{Q}$.

Corollary 2.2.4. *Let μ and ν be two distinct probability measures on $\mathbb{R}$ with finite first moment and equal means. Then $\mathcal{Q}\neq\emptyset$ iff $\mu\le_{cx}\nu$.*

The proof of Proposition 2.2.3 relies on the two following lemmas.

Lemma 2.2.5. *Let $Q\in\mathcal{Q}$. For du-almost all $u\in(0,1)$,*

$$\begin{cases} u\in\mathcal{U}_+ \implies F_\nu^{-1}(v)>F_\mu^{-1}(u),\ \pi_+^Q(u,dv)\text{-a.e;} \\ u\in\mathcal{U}_- \implies F_\nu^{-1}(v)<F_\mu^{-1}(u),\ \pi_-^Q(u,dv)\text{-a.e.} \end{cases}$$

Proof of Lemma 2.2.5. We have

$$\int_{(0,1)} \left(\int_{(0,1)} \mathbb{1}_{\{F_\nu^{-1}(v) \le F_\mu^{-1}(u)\}} \pi_+^Q(u, dv) \right) (F_\mu^{-1} - F_\nu^{-1})^+(u)\, du = \gamma \int_{(0,1)^2} \mathbb{1}_{\{F_\nu^{-1}(v) \le F_\mu^{-1}(u)\}}\, Q(du, dv)$$
$$\le \gamma \int_{(0,1)^2} \mathbb{1}_{\{F_\nu^{-1}(v) \le F_\mu^{-1}(v)\}}\, Q(du, dv) = \int_{(0,1)^2} \mathbb{1}_{\{F_\mu^{-1}(v) - F_\nu^{-1}(v) \ge 0\}} (F_\mu^{-1} - F_\nu^{-1})^-(v)\, dv\, \pi_-^Q(v, du)$$
$$= 0,$$

where we used for the inequality that $u < v$, $Q(du, dv)$-almost everywhere and that F_μ^{-1} is nondecreasing. So for du-almost all $u \in \mathcal{U}_+$, $\pi_+^Q(u, dv)$-a.e., $F_\nu^{-1}(v) > F_\mu^{-1}(u)$. With a symmetric reasoning, we obtain that for du-almost all $u \in \mathcal{U}_-$, $\pi_-^Q(u, dv)$-a.e., $F_\nu^{-1}(v) < F_\mu^{-1}(u)$. □

Lemma 2.2.6. *Let $(\widetilde{m}(u, dy))_{u \in (0,1)}$ be a Markov kernel and let $(m(x, dy))_{x \in \mathbb{R}}$ be given by (2.2.6). Then*

$$\mu(dx)\, m(x, dy) = (F_\mu^{-1}(u), y)_\sharp \left(\mathbb{1}_{(0,1)}(u)\, du\, \widetilde{m}(u, dy) \right),$$

where $\sharp$ denotes the pushforward operation.

Proof of Lemma 2.2.6. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a measurable and nonnegative function. By Lemma 2.6.4 below, $F_\mu(x) > 0$ and $F_\mu(x_-) < 1$, $\mu(dx)$-almost everywhere. So using (2.2.7), we have

$$\int_{\mathbb{R} \times \mathbb{R}} h(x, y)\, \mu(dx)\, m(x, dy)$$
$$= \int_{\mathbb{R} \times \mathbb{R} \times (0,1)} h(x, y) \mathbb{1}_{\{0 < F_\mu(x), F_\mu(x_-) < 1\}}\, \mu(dx)\, \widetilde{m}(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)), dy)\, dv.$$

Let $\theta : (x, v) \mapsto F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-))$. By Lemma 2.6.6 below, $x = F_\mu^{-1}(\theta(x, v))$, $\mu(dx) \otimes dv$-almost everywhere on $\mathbb{R} \times (0,1)$ and $\theta(x, v)_\sharp(\mu(dx) \otimes \mathbb{1}_{(0,1)}(v)\, dv) = \mathbb{1}_{(0,1)}(u)\, du$. So

$$\int_{\mathbb{R} \times \mathbb{R}} h(x, y)\, \mu(dx)\, m(x, dy)$$
$$= \int_{\mathbb{R} \times \mathbb{R} \times (0,1)} h(F_\mu^{-1}(\theta(x, v)), y) \mathbb{1}_{\{0 < F_\mu(F_\mu^{-1}(\theta(x,v))), F_\mu(F_\mu^{-1}(\theta(x,v)_-)) < 1\}}\, \mu(dx)\, \widetilde{m}(\theta(x, v), dy)\, dv$$
$$= \int_{\mathbb{R} \times (0,1)} h(F_\mu^{-1}(u), y) \mathbb{1}_{\{0 < F_\mu(F_\mu^{-1}(u)), F_\mu(F_\mu^{-1}(u)_-) < 1\}}\, \widetilde{m}(u, dy)\, du.$$

By Lemma 2.6.4 below and the inverse transform sampling, $F_\mu(F_\mu^{-1}(u)) > 0$ and $F_\mu(F_\mu^{-1}(u)_-) < 1$, du-almost everywhere on $(0, 1)$, hence

$$\int_{\mathbb{R} \times \mathbb{R}} h(x, y)\, \mu(dx)\, m(x, dy) = \int_{\mathbb{R} \times (0,1)} h(F_\mu^{-1}(u), y)\, \widetilde{m}(u, dy)\, du.$$

□

Proof of Proposition 2.2.3. Let us show that M^Q defines a coupling between μ and ν. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable and nonnegative (or bounded) function. We want to show that

$$\int_{\mathbb{R}\times\mathbb{R}} h(y)\,\mu(dx)m^Q(x,dy) = \int_{\mathbb{R}} h(y)\,\nu(dy),$$

which by Lemma 2.2.6 and the inverse transform sampling is equivalent to

$$\int_0^1 \int_{\mathbb{R}} h(y)\,\widetilde{m}^Q(u,dy)\,du = \int_0^1 h(F_\nu^{-1}(u))\,du. \tag{2.2.8}$$

Thanks to Lemma 2.2.5, we get for du-almost all $u \in (0,1)$,

$$\begin{aligned}
&\int_{\mathbb{R}} h(y)\,\widetilde{m}^Q(u,dy) \\
&= \int_{(0,1)} \left(1 - \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\right) h(F_\nu^{-1}(u))(\pi_+^Q(u,dv)\mathbb{1}_{\{F_\mu^{-1}(u)>F_\nu^{-1}(u)\}} + \pi_-^Q(u,dv)\mathbb{1}_{\{F_\mu^{-1}(u)<F_\nu^{-1}(u)\}}) \\
&\quad + \int_{(0,1)} \left(\frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\right) h(F_\nu^{-1}(v))(\pi_+^Q(u,dv)\mathbb{1}_{\{F_\mu^{-1}(u)>F_\nu^{-1}(u)\}} + \pi_-^Q(u,dv)\mathbb{1}_{\{F_\mu^{-1}(u)<F_\nu^{-1}(u)\}}) \\
&\quad + h(F_\nu^{-1}(u))\mathbb{1}_{\{F_\mu^{-1}(u)=F_\nu^{-1}(u)\}} \\
&= h(F_\nu^{-1}(u)) + \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^+(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}(h(F_\nu^{-1}(v)) - h(F_\nu^{-1}(u)))\,\pi_+^Q(u,dv) \\
&\quad + \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^-(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(v)}(h(F_\nu^{-1}(v)) - h(F_\nu^{-1}(u)))\,\pi_-^Q(u,dv).
\end{aligned} \tag{2.2.9}$$

Since

$$\begin{aligned}
&\int_{(0,1)^2} \frac{(F_\mu^{-1} - F_\nu^{-1})^+(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}(h(F_\nu^{-1}(v) - h(F_\nu^{-1}(u)))\,\pi_+^Q(u,dv)\,du \\
&= \gamma \int_{(0,1)^2} \frac{h(F_\nu^{-1}(v)) - h(F_\nu^{-1}(u))}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} Q(du,dv) \\
&= \int_{(0,1)^2} \frac{h(F_\nu^{-1}(v)) - h(F_\nu^{-1}(u))}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} (F_\mu^{-1} - F_\nu^{-1})^-(v)\,\pi_-^Q(v,du)\,dv \\
&= -\int_{(0,1)^2} \frac{(F_\mu^{-1} - F_\nu^{-1})^-(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(v)}(h(F_\nu^{-1}(v)) - h(F_\nu^{-1}(u)))\,\pi_-^Q(u,dv)\,du,
\end{aligned}$$

we deduce that $\int_0^1 \int_{\mathbb{R}} h(y)\,\widetilde{m}^Q(u,dy)\,du = \int_0^1 h(F_\nu^{-1}(u))\,du$. We conclude that M^Q is a coupling between μ and ν. In particular for $h : y \mapsto |y|$, using the inverse transform sampling, we have

$$\int_0^1 \int_{\mathbb{R}} |y|\,\widetilde{m}^Q(u,dy)\,du = \int_0^1 |F_\nu^{-1}(u)|\,du = \int_{\mathbb{R}} |y|\,\nu(dy) < +\infty.$$

So $\int_{\mathbb{R}} y\,\widetilde{m}^Q(u,dy)$ is well defined du-almost everywhere on $(0,1)$.

Let us show now that M^Q defines a martingale coupling between μ and ν. By Lemma 2.2.5, for du-almost all $u \in \mathcal{U}_+$,

$$\begin{aligned}\int_{\mathbb{R}} y\,\widetilde{m}^Q(u,dy) &= \int_{(0,1)} \left(1 - \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\right) F_\nu^{-1}(u)\,\pi_+^Q(u,dv) \\ &\quad + \int_{(0,1)} \left(\frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\right) F_\nu^{-1}(v)\,\pi_+^Q(u,dv) \\ &= \int_{(0,1)} (F_\nu^{-1}(u) + F_\mu^{-1}(u) - F_\nu^{-1}(u))\,\pi_+^Q(u,dv) \\ &= F_\mu^{-1}(u). \end{aligned} \tag{2.2.10}$$

In the same way, for du-almost all $u \in \mathcal{U}_-$,

$$\int_{\mathbb{R}} y\,\widetilde{m}^Q(u,dy) = F_\mu^{-1}(u). \tag{2.2.11}$$

Else if $u \in \mathcal{U}_0$, then by definition of $\widetilde{m}^Q(u,dy)$,

$$\int_{\mathbb{R}} y\,\widetilde{m}^Q(u,dy) = F_\nu^{-1}(u) = F_\mu^{-1}(u),$$

so for du-almost all $u \in (0,1)$, $\int_{\mathbb{R}} y\,\widetilde{m}^Q(u,dy) = F_\mu^{-1}(u)$.

Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable and bounded function. By Lemma 2.2.6,

$$\int_{\mathbb{R}\times\mathbb{R}} h(x)(y-x)\,\mu(dx)\,m^Q(x,dy) = \int_0^1 h(F_\mu^{-1}(u)) \left(\int_{\mathbb{R}} (y - F_\mu^{-1}(u))\,\widetilde{m}^Q(u,dy)\right) du = 0.$$

So $\mu(dx)\,m^Q(x,dy)$ is a martingale coupling between μ and ν. $\square$

Let $H : \mathbb{R}^2 \to \mathbb{R}$ be a measurable and bounded function. Lemma 2.2.6 and (2.2.9) written with $h : y \mapsto H(F_\mu^{-1}(u), y)$ yield the following formula, which illustrates well how the martingale coupling M^Q differs from the comonotous coupling between μ and ν:

$$\begin{aligned}&\int_{\mathbb{R}\times\mathbb{R}} H(x,y)\,M^Q(dx,dy) - \int_0^1 H(F_\mu^{-1}(u), F_\nu^{-1}(u))\,du \\ &= \gamma \int_{(0,1)^2} \frac{H(F_\mu^{-1}(u), F_\nu^{-1}(v)) - H(F_\mu^{-1}(u), F_\nu^{-1}(u)) + H(F_\mu^{-1}(v), F_\nu^{-1}(u)) - H(F_\mu^{-1}(v), F_\nu^{-1}(v))}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\,Q(du,dv). \end{aligned} \tag{2.2.12}$$

Notice that the last integral is well defined since $Q(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi_+^Q(u,dv)$ and according to Lemma 2.2.5, $Q(du,dv)$-almost everywhere, $F_\nu^{-1}(v) > F_\mu^{-1}(u) > F_\nu^{-1}(u)$. Moreover, the fact that μ and ν have finite first moment along with the inverse transform sampling show that (2.2.12) also holds for any measurable map $H : \mathbb{R}^2 \to \mathbb{R}$ with at most linear growth. As shown in the next proposition, we can easily deduce from this formula that the map $\mathcal{Q} \ni Q \mapsto M^Q$ is one-to-one as soon as F_μ and F_ν are continuous.

Proposition 2.2.7. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. If F_μ and F_ν are continuous, then the map $\mathcal{Q} \ni Q \mapsto M^Q$ is one-to-one.*

Proof. Let $Q, Q' \in \mathcal{Q}$ be such that $Q \neq Q'$. Then there exists a borel set $A \subset (0,1)^2$ such that $Q(A) \neq Q'(A)$. Let $H : (x,y) \mapsto (y - F_\nu^{-1}(F_\mu(x)))^+ \mathbb{1}_{\{F_\mu(x) \in (0,1)\}} \mathbb{1}_A(F_\mu(x), F_\nu(y)) \mathbb{1}_{\{F_\mu(x) < F_\mu(y)\}}$. Since F_μ and F_ν are continuous, for all $u, v \in (0,1)$, we have $F_\mu(F_\mu^{-1}(u)) = u$ and $F_\nu(F_\nu^{-1}(v)) = v$, so $H(F_\mu^{-1}(u), F_\nu^{-1}(v)) = (F_\nu^{-1}(v) - F_\nu^{-1}(u))^+ \mathbb{1}_A(u,v) \mathbb{1}_{\{u<v\}}$. We deduce that for all $u, v \in (0,1)$, $H(F_\mu^{-1}(u), F_\nu^{-1}(u)) = H(F_\mu^{-1}(v), F_\nu^{-1}(v)) = 0$ and since $(Q+Q')(du,dv)$-almost everywhere on $(0,1)^2$, $u < v$, we have that $(Q+Q')(du,dv)$-almost everywhere on $(0,1)^2$, $H(F_\mu^{-1}(v), F_\nu^{-1}(u)) = 0$. Since $H(x,y)$ grows at most linearly in $F_\nu^{-1}(F_\mu(x))$ and y, one can easily deduce from the integrability of μ and ν and the inverse transform sampling that (2.2.12) holds. Using that $(Q+Q')(du,dv)$ almost everywhere on $(0,1)^2$, $F_\nu^{-1}(u) < F_\mu^{-1}(u) < F_\nu^{-1}(v)$, which is a consequence of Lemma 2.2.5, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathbb{R}} H(x,y)\, M^Q(dx,dy) - \int_{\mathbb{R}\times\mathbb{R}} H(x,y)\, M^{Q'}(dx,dy) \\
=&\gamma \int_{(0,1)^2} \frac{H(F_\mu^{-1}(u), F_\nu^{-1}(v))}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\, Q(du,dv) - \gamma \int_{(0,1)^2} \frac{H(F_\mu^{-1}(u), F_\nu^{-1}(v))}{F_\nu^{-1}(v) - F_\nu^{-1}(u)}\, Q'(du,dv) \\
=&\gamma(Q(A) - Q'(A)) \neq 0,
\end{aligned}$$

hence $M^Q \neq M^{Q'}$ and the map $\mathcal{Q} \ni Q \mapsto M^Q$ is one-to-one. □

According to Theorem A.4 [25], there exist $N \in \mathbb{N}^* \cup \{+\infty\}$ and a sequence of disjoint open intervals $((\underline{t}_n, \overline{t}_n))_{1 \le n \le N}$ such that

$$\left\{ t \in \mathbb{R} \mid \int_{-\infty}^t F_\mu(x)\,dx < \int_{-\infty}^t F_\nu(x)\,dx \right\} = \bigcup_{n=1}^N (\underline{t}_n, \overline{t}_n). \tag{2.2.13}$$

These intervals are called the irreducible components of the pair (μ, ν). Moreover, there exists a unique decomposition of probability measures $(\mu_n, \nu_n)_{1\le n\le N}$, such that the choice of any martingale coupling M between μ and ν reduces to the choice of a sequence of martingale couplings $(M_n)_{1\le n\le N}$. More precisely, for all $1 \le n \le N$,

$$F_\mu(\underline{t}_n) \le F_\nu(\underline{t}_n) \le F_\nu((\overline{t}_n)_-) \le F_\mu((\overline{t}_n)_-), \quad F_\mu(\underline{t}_n) < F_\mu((\overline{t}_n)_-), \tag{2.2.14}$$

and μ_n and ν_n are given by

$$\left\{ \begin{array}{rcl} \mu_n(dx) &=& \frac{1}{F_\mu((\overline{t}_n)_-) - F_\mu(\underline{t}_n)} \mathbb{1}_{(\underline{t}_n, \overline{t}_n)}(x)\, \mu(dx); \\ \nu_n(dy) &=& \frac{1}{F_\mu((\overline{t}_n)_-) - F_\mu(\underline{t}_n)} \Big(\mathbb{1}_{(\underline{t}_n, \overline{t}_n)}(y)\, \nu(dy) + (F_\nu(\underline{t}_n) - F_\mu(\underline{t}_n))\, \delta_{\underline{t}_n}(dy) \\ && \qquad + (F_\mu((\overline{t}_n)_-) - F_\nu((\overline{t}_n)_-))\, \delta_{\overline{t}_n}(dy) \Big). \end{array} \right. \tag{2.2.15}$$

Then a probability measure M on $\mathbb{R}^2$ is a martingale coupling between μ and ν if and only if there exists a sequence $(M_n)_{1\le n\le N}$ such that for all $1 \le n \le N$, M_n is a martingale coupling between μ_n and ν_n and

$$M(dx,dy) = \mathbb{1}_{\mathbb{R}\setminus \bigcup_{n=1}^N (\underline{t}_n, \overline{t}_n)}(x)\, \mu(dx)\, \delta_x(dy) + \sum_{n=1}^N \mu((\underline{t}_n, \overline{t}_n))\, M_n(dx,dy).$$

We can establish a strong connection between the support of any probability measure $Q \in \mathcal{Q}$ and the irreducible components of (μ, ν).

Proposition 2.2.8. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Let $(\underline{t}_n, \overline{t}_n)_{1 \le n \le N}$ denote the irreducible components of (μ, ν). Then for all $Q \in \mathcal{Q}$, we have*

$$Q\left(\bigcup_{1 \le n \le N} (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))^2\right) = 1.$$

Proof. Let $Q \in \mathcal{Q}$. By Lemma A.8 [4], we have

$$\mathcal{W} := \bigcup_{n=1}^{N} (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-)) = \left\{u \in (0,1) \mid \int_0^u F_\mu^{-1}(v)\,dv > \int_0^u F_\nu^{-1}(v)\,dv\right\}.$$

Let $u \in (0,1)$ be such that $F_\mu^{-1}(u) > F_\nu^{-1}(u)$, that is $u \in \mathcal{U}_+$. Since $\mu \le_{cx} \nu$, according to the necessary condition of Theorem 3.A.5 Chapter 3 [169] (see also Remark 2.3.2 for a proof), for all $q \in [0,1]$, $\int_0^q F_\mu^{-1}(v)\,dv \ge \int_0^q F_\nu^{-1}(v)\,dv$. By left-continuity of F_μ^{-1} and F_ν^{-1}, we deduce that $\int_0^u F_\mu^{-1}(v)\,dv > \int_0^u F_\nu^{-1}(v)\,dv$, that is $u \in \mathcal{W}$. So $\mathcal{U}_+ \subset \mathcal{W}$.

Let $1 \le n \le N$. Then M^Q transports $(\underline{t}_n, \overline{t}_n)$ to $[\underline{t}_n, \overline{t}_n]$, namely for $\mu(dx)$-almost all $x \in (\underline{t}_n, \overline{t}_n)$, $m^Q(x, [\underline{t}_n, \overline{t}_n]) = 1$. So using Lemma 2.2.6 for the last equality, we have

$$\begin{aligned}
\int_{F_\mu(\underline{t}_n)}^{F_\mu((\overline{t}_n)_-)} du = \mu((\underline{t}_n, \overline{t}_n)) &= \int_{\mathbb{R}} \mathbb{1}_{\{\underline{t}_n < x < \overline{t}_n\}} \mu(dx) \\
&= \int_{\mathbb{R}^2} \mathbb{1}_{\{\underline{t}_n < x < \overline{t}_n\}} \mathbb{1}_{\{\underline{t}_n \le y \le \overline{t}_n\}}\, \mu(dx)\, m^Q(x, dy) \\
&= \int_{(0,1)\times\mathbb{R}} \mathbb{1}_{\{\underline{t}_n < F_\mu^{-1}(u) < \overline{t}_n\}} \mathbb{1}_{\{\underline{t}_n \le y \le \overline{t}_n\}}\, du\, \widetilde{m}^Q(u, dy).
\end{aligned}$$

Using Lemma 2.6.3 below, one can easily see that for all $u \in (0,1)$, $\mathbb{1}_{\{F_\mu(\underline{t}_n) < u < F_\mu((\overline{t}_n)_-)\}} \le \mathbb{1}_{\{\underline{t}_n < F_\mu^{-1}(u) < \overline{t}_n\}} \le \mathbb{1}_{\{F_\mu(\underline{t}_n) < u \le F_\mu((\overline{t}_n)_-)\}}$. So

$$\begin{aligned}
\int_{F_\mu(\underline{t}_n)}^{F_\mu((\overline{t}_n)_-)} du &= \int_{(0,1)\times\mathbb{R}} \mathbb{1}_{\{F_\mu(\underline{t}_n) < u < F_\mu((\overline{t}_n)_-)\}} \mathbb{1}_{\{\underline{t}_n \le y \le \overline{t}_n\}}\, du\, \widetilde{m}^Q(u, dy) \\
&= \int_{F_\mu(\underline{t}_n)}^{F_\mu((\overline{t}_n)_-)} \widetilde{m}^Q(u, [\underline{t}_n, \overline{t}_n])\, du.
\end{aligned}$$

So for du-almost all $u \in (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))$, $\widetilde{m}^Q(u, [\underline{t}_n, \overline{t}_n]) = 1$. By Lemma 2.2.5, $d\Psi_+(u)$-almost everywhere on $(F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))$,

$$\begin{aligned}
1 &= \pi_+^Q(u, \{v \in (0,1) \mid F_\nu^{-1}(v) \in [\underline{t}_n, \overline{t}_n]\}) \\
&= \pi_+^Q(u, \mathcal{U}_- \cap (u,1) \cap \{v \in (0,1) \mid F_\nu^{-1}(v) \in [\underline{t}_n, \overline{t}_n]\}),
\end{aligned}$$

where the last equality derives from conditions (ii) and (iii) satisfied by Q. Let $u \in (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))$. Let us check that

$$\mathcal{U}_- \cap (u,1) \cap \{v \in (0,1) \mid F_\nu^{-1}(v) \in [\underline{t}_n, \overline{t}_n]\} \subset \mathcal{U}_- \cap (u,1) \cap (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-)]. \quad (2.2.16)$$

Let $v \in (0,1)$ be such that $F_\nu^{-1}(v) \in [\underline{t}_n, \overline{t}_n]$. First of all, if $v > u$ then $v > F_\mu(\underline{t}_n)$. Second, if $v > F_\mu((\overline{t}_n)_-)$, then according to (2.2.14) and Lemma 2.6.3 below, we have $F_\nu((\overline{t}_n)_-) \le F_\mu((\overline{t}_n)_-) < v \le F_\nu(\overline{t}_n)$. In that case, if $v \le F_\mu(\overline{t}_n)$, then $v \in (F_\nu((\overline{t}_n)_-), F_\nu(\overline{t}_n)] \cap (F_\mu((\overline{t}_n)_-), F_\mu(\overline{t}_n)]$, so $F_\nu^{-1}(v) = F_\mu^{-1}(v) = \overline{t}_n$ and $v \in \mathcal{U}_0$. Else if $v > F_\mu(\overline{t}_n)$, then $v \in (F_\mu(\overline{t}_n), F_\nu(\overline{t}_n)]$ so $F_\nu^{-1}(v) \le \overline{t}_n < F_\mu^{-1}(v)$ and $v \in \mathcal{U}_+$. This proves (2.2.16).

Using conditions (ii) and (iii) satisfied by Q again and the fact that the second marginal of Q has a density, we get that $d\Psi_+(u)$-almost everywhere on $(F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))$,

$$\begin{aligned}
1 &= \pi_+^Q(u, \mathcal{U}_- \cap (u,1) \cap \{v \in (0,1) \mid F_\nu^{-1}(v) \in [\underline{t}_n, \overline{t}_n]\}) \\
&\le \pi_+^Q(u, \mathcal{U}_- \cap (u,1) \cap (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-)]) \\
&= \pi_+^Q(u, (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))).
\end{aligned}$$

We deduce that

$$\begin{aligned}
Q\left(\bigcup_{1\le n\le N} (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))^2\right) &= \sum_{n=1}^N Q\left((F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))^2\right) \\
&= \frac{1}{\gamma}\sum_{n=1}^N \int_{F_\mu(\underline{t}_n)}^{F_\mu((\overline{t}_n)_-)} d\Psi_+(u)\, \pi_+^Q(u, (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))) \\
&= \frac{1}{\gamma}\sum_{n=1}^N \int_{F_\mu(\underline{t}_n)}^{F_\mu((\overline{t}_n)_-)} d\Psi_+(u) = \frac{1}{\gamma}\, d\Psi_+(\mathcal{W}) \\
&\ge \frac{1}{\gamma}\, d\Psi_+(\mathcal{U}_+) \\
&= 1,
\end{aligned}$$

where we used the fact that $\mathcal{U}_+ \subset \mathcal{W}$ for the inequality. □

The next proposition clarifies the structure of the set of martingale couplings deriving from $\mathcal{Q}$ and states a linearity property of the map $Q \in \mathcal{Q} \mapsto M^Q$. In particular, it ensures that the set of martingale couplings deriving from $\mathcal{Q}$ is either a singleton, or uncountably infinite.

Proposition 2.2.9. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Then for all $Q, Q' \in \mathcal{Q}$ and $\lambda \in [0,1]$,*

$$M^{\lambda Q + (1-\lambda)Q'} = \lambda M^Q + (1-\lambda) M^{Q'}.$$

In particular, the set $\{M^Q \mid Q \in \mathcal{Q}\}$ is convex.

Proof. Let $Q, Q' \in \mathcal{Q}$ and let $\lambda \in [0,1]$. It is straightforward that for du-almost all $u \in \mathcal{U}_+$ and dv-almost all $v \in \mathcal{U}_-$,

$$\begin{aligned}
\pi_+^{\lambda Q + (1-\lambda)Q'}(u, dy) &= \lambda \pi_+^Q(u, dy) + (1-\lambda)\pi_+^{Q'}(u, dy); \\
\pi_-^{\lambda Q + (1-\lambda)Q'}(v, dy) &= \lambda \pi_-^Q(v, dy) + (1-\lambda)\pi_-^{Q'}(v, dy).
\end{aligned}$$

Using Lemma 2.2.5, we get that for du-almost all $u \in (0,1)$,

$$\widetilde{m}^{\lambda Q+(1-\lambda)Q'}(u,dy) = \lambda \widetilde{m}^{Q}(u,dy) + (1-\lambda)\widetilde{m}^{Q'}(u,dy).$$

Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a measurable and bounded function. By Lemma 2.2.6,

$$\begin{aligned}
&\int_{\mathbb{R}\times\mathbb{R}} h(x,y)\, M^{\lambda Q+(1-\lambda)Q'}(dx,dy) \\
&= \int_{\mathbb{R}\times\mathbb{R}} h(x,y)\,\mu(dx)\, m^{\lambda Q+(1-\lambda)Q'}(x,dy) = \int_0^1 \left(\int_{\mathbb{R}} h(F_\mu^{-1}(u),y)\, \widetilde{m}^{\lambda Q+(1-\lambda)Q'}(u,dy)\right) du \\
&= \lambda \int_0^1 \left(\int_{\mathbb{R}} h(F_\mu^{-1}(u),y)\, \widetilde{m}^{Q}(u,dy)\right) du + (1-\lambda)\int_0^1 \left(\int_{\mathbb{R}} h(F_\mu^{-1}(u),y)\, \widetilde{m}^{Q'}(u,dy)\right) du \\
&= \lambda \int_{\mathbb{R}\times\mathbb{R}} h(x,y)\,\mu(dx)\, m^{Q}(x,dy) + (1-\lambda)\int_{\mathbb{R}\times\mathbb{R}} h(x,y)\,\mu(dx)\, m^{Q'}(x,dy) \\
&= \int_{\mathbb{R}\times\mathbb{R}} h(x,y)\,(\lambda M^Q + (1-\lambda)M^{Q'})(dx,dy).
\end{aligned}$$

So $M^{\lambda Q+(1-\lambda)Q'} = \lambda M^Q + (1-\lambda)M^{Q'}$. $\square$

We deduce that if $Q, Q' \in \mathcal{Q}$ are such that $M^Q \neq M^{Q'}$, then there exists a whole segment of martingale couplings between μ and ν, all parametrised by $\mathcal{Q}$. More details are given in Section 2.4. Let us complete this section by revisiting the example given in Section 2.2.1.

Example 2.2.10. Suppose now that $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are symmetric with common mean $\alpha \in \mathbb{R}$, that is $(x-\alpha)_\sharp\mu(dx) = (\alpha-x)_\sharp\mu(dx)$ and $(y-\alpha)_\sharp\nu(dy) = (\alpha-y)_\sharp\nu(dy)$ where $\sharp$ denotes the pushforward operation. Suppose in addition that their respective quantile functions satisfy $F_\mu^{-1} \geq F_\nu^{-1}$ on $(0,1/2]$ and $F_\mu^{-1} \leq F_\nu^{-1}$ on $(1/2,1)$. We saw in Section 2.2.1 that when U is a random variable uniformly distributed on $[0,1]$ and Z is given by (2.2.1), $(F_\mu^{-1}(U), Z)$ is an explicit coupling between μ and ν in the case $\alpha = 0$. Let us show here that this coupling is in fact associated to a particular element of $\mathcal{Q}$. According to Lemma 2.6.5 below, we have $F_\mu^{-1}(u) = 2\alpha - F_\mu^{-1}(1-u)$ and $F_\nu^{-1}(u) = 2\alpha - F_\nu^{-1}(1-u)$ for du-almost all $u \in (0,1)$, which is helpful in order to see that the probability measure Q_2 defined on $(0,1)^2$ by

$$Q_2(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\, du\, \delta_{1-u}(dv) \tag{2.2.17}$$

is an element of $\mathcal{Q}$ (in particular to check that it satisfies (ii)). For that element Q_2, using (2.2.5), Lemma 2.2.5 and Lemma 2.6.5 below, we have for du-almost all $u \in \mathcal{U}_+ \cup \mathcal{U}_-$,

$$\widetilde{m}^{Q_2}(u,dy) = \frac{F_\mu^{-1}(u) + F_\nu^{-1}(u) - 2\alpha}{2(F_\nu^{-1}(u) - \alpha)}\delta_{F_\nu^{-1}(u)}(dy) + \frac{F_\nu^{-1}(u) - F_\mu^{-1}(u)}{2(F_\nu^{-1}(u) - \alpha)}\delta_{2\alpha - F_\nu^{-1}(u)}(dy), \tag{2.2.18}$$

and $\widetilde{m}^{Q_2}(u,dy) = \delta_{F_\nu^{-1}(u)}(dy)$ if $u \in \mathcal{U}_0$. Let $u \in (0,1)$. If $F_\mu^{-1}(u) = F_\nu^{-1}(u) \neq \alpha$, then $\delta_{F_\nu^{-1}(u)}(dy)$ coincides with the right-hand side of (2.2.18). Furthermore if $F_\nu^{-1}(u) = \alpha$, since $\alpha \geq F_\mu^{-1}(u) \geq F_\nu^{-1}(u)$ or $\alpha \leq F_\mu^{-1}(u) \leq F_\nu^{-1}(u)$ by an easy generalisation of (2.2.2), then

$F_\mu^{-1}(u) = \alpha$. Therefore (2.2.18) holds for du-almost all $u \in (0,1)$ such that $F_\nu^{-1}(u) \neq \alpha$ and $\widetilde{m}^{Q_2}(u, dy) = \delta_{F_\nu^{-1}(u)}(dy)$ for du-almost all $u \in (0,1)$ such that $F_\nu^{-1}(u) = \alpha$.

Let U and V be two independent random variables uniformly distributed on $(0,1)$ and let Y be defined as in (2.2.1) but with the mean α taken into account, that is

$$Y = F_\nu^{-1}(U)\mathbb{1}_{\left\{F_\nu^{-1}(U)\neq\alpha, V \leq \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)-2\alpha}{2(F_\nu^{-1}(U)-\alpha)}\right\}} + (2\alpha - F_\nu^{-1}(U))\mathbb{1}_{\left\{F_\nu^{-1}(U)\neq\alpha, V > \frac{F_\mu^{-1}(U)+F_\nu^{-1}(U)-2\alpha}{2(F_\nu^{-1}(U)-\alpha)}\right\}}$$
$$+ \alpha\mathbb{1}_{\{F_\nu^{-1}(U)=\alpha\}}.$$

Then (U, Y) is distributed according to $\mathbb{1}_{(0,1)}(u)\, du\, \widetilde{m}^{Q_2}(u, dy)$. By Lemma 2.2.6, $(F_\mu^{-1}(U), Y)$ is distributed according to $\mu(dx)\, m(x, dy)$.

2.2.3 Optimality property

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. It is well known that F_ν^{-1} is constant on the jumps of F_μ, that is F_ν^{-1} is constant on the intervals of the form $(F_\mu(x_-), F_\mu(x)]$, iff the comonotonous coupling between μ and ν is concentrated on the graph of a map $T : \mathbb{R} \to \mathbb{R}$, and then

$$T = F_\nu^{-1} \circ F_\mu. \tag{2.2.19}$$

We will refer to T as the Monge transport map.

Proposition 2.2.11. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Suppose in addition that F_ν^{-1} is constant on the intervals of the form $(F_\mu(x_-), F_\mu(x)]$. Let T be the Monge transport map. Let $Q \in \mathcal{Q}$. Then*

$$\inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M(dx, dy) = \int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M^Q(dx, dy) = \mathcal{W}_1(\mu, \nu).$$

Proof. This is a particular case of Proposition 2.2.18 below. Indeed, let $M(dx, dy) = \mu(dx)\, m(x, dy)$ be a martingale coupling between μ and ν. Let $(\widetilde{m}(u, dy))_{u\in(0,1)}$ be the kernel defined for all $u \in (0,1)$ by $\widetilde{m}(u, dy) = m(F_\mu^{-1}(u), dy)$. Using the inverse transform sampling, we have

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, \mu(dx)\, m(x, dy) &= \int_{(0,1)\times\mathbb{R}} |y - T(F_\mu^{-1}(u))|\, du\, m(F_\mu^{-1}(u), dy) \\ &= \int_{(0,1)\times\mathbb{R}} |y - F_\nu^{-1}(F_\mu((F_\mu^{-1}(u))))|\, du\, \widetilde{m}(u, dy),\end{aligned}$$

where we used for the last equality that $T = F_\nu^{-1} \circ F_\mu$. Let $u \in (0,1)$. If there exists $x \in \mathbb{R}$ such that $u = F_\mu(x)$, then $F_\mu(F_\mu^{-1}(u)) = F_\mu(F_\mu^{-1}(F_\mu(x))) = F_\mu(x) = u$, so $F_\nu^{-1}(F_\mu(F_\mu^{-1}(u))) = F_\nu^{-1}(u)$. Else there exists x in the set of discontinuities of F_μ such that $F_\mu(x_-) \leq u < F_\mu(x)$. In that case, if $u > F_\mu(x_-)$ then $x = F_\mu^{-1}(u)$, so $F_\nu^{-1}(F_\mu(F_\mu^{-1}(u))) = F_\nu^{-1}(F_\mu(x)) = F_\nu^{-1}(u)$ since F_ν^{-1} is constant on the jumps of F_μ. Hence

$$du\text{-a.e. on } (0,1), \quad F_\nu^{-1}(F_\mu(F_\mu^{-1}(u))) = F_\nu^{-1}(u). \tag{2.2.20}$$

We deduce that

$$\int_{(0,1)\times\mathbb{R}} |y - F_\nu^{-1}(F_\mu(F_\mu^{-1}(u)))|\, du\, \widetilde{m}(u,dy) = \int_{(0,1)\times\mathbb{R}} |y - F_\nu^{-1}(u)|\, du\, \widetilde{m}(u,dy).$$

With a similar reasoning, we have

$$\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, \mu(dx)\, m^Q(x,dy) = \int_{(0,1)\times\mathbb{R}} |y - F_\nu^{-1}(u)|\, du\, \widetilde{m}^Q(u,dy).$$

Therefore, using Proposition 2.2.18 combined with Remark 2.2.19 below, we get that $\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|\, M(dx,dy)$ is minimised when $M = M^Q$, for which we have $\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|\, M^Q(dx,dy) = \mathcal{W}_1(\mu,\nu)$. □

2.2.4 Stability inequality

We can now state our main result. In the minimisation of the cost function $(x,y) \mapsto |x-y|$ with respect to the couplings between μ and ν, the addition of the martingale constraint does not cost more than a factor 2.

Theorem 2.2.12. *For all $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu <_{cx} \nu$ and for all Q in the non-empty set $\mathcal{Q}$,*

$$\int_{\mathbb{R}\times\mathbb{R}} |x-y|\, M^Q(dx,dy) \le 2\mathcal{W}_1(\mu,\nu). \tag{2.2.21}$$

Consequently,

$$\mathcal{M}_1(\mu,\nu) \le 2\mathcal{W}_1(\mu,\nu). \tag{2.2.22}$$

Moreover, the constant 2 is sharp.

The proof of Theorem 2.2.12 relies on Proposition 2.2.18 below. Note that since $\Pi^{\mathrm{M}}(\mu,\nu) \subset \Pi(\mu,\nu)$, we always have $\mathcal{W}_1(\mu,\nu) \le \mathcal{M}_1(\mu,\nu)$. Moreover, the stability inequality (2.2.22) can be tensorised: it holds in greater dimension when the marginals are independent, as the next corollary states.

Corollary 2.2.13. *Let $d \in \mathbb{N}^*$ and $\mu_1,\cdots,\mu_d,\nu_1,\cdots,\nu_d \in \mathcal{P}_1(\mathbb{R})$ be such that for all $1 \le i \le d$, $\mu_i \le_{cx} \nu_i$. Let $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ and $\nu = \nu_1 \otimes \cdots \otimes \nu_d$. Then $\mu \le_{cx} \nu$ and*

$$\mathcal{M}_1(\mu,\nu) \le 2\mathcal{W}_1(\mu,\nu),$$

when $\mathbb{R}^d$ is endowed with the L^1-norm.

Proof of Corollary 2.2.13. For all $1 \le i \le d$, since $\mu_i \le_{cx} \nu_i$, Strassen's theorem or Proposition 2.2.3 and Corollary 2.2.4 ensure the existence of a martingale coupling $M_i(dx_i,dy_i) = \mu_i(dx_i)\, m_i(x_i,dy_i)$ between μ_i and ν_i. Let then M be the probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ defined by $M(dx,dy) = \mu(dx)\, m_1(x_1,dy_1)\cdots m_d(x_d,dy_d)$. Then it is clear that M is a martingale coupling between μ and ν, which shows that $\mu \le_{cx} \nu$, and

$$\mathcal{M}_1(\mu,\nu) \le \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|\, M(dx,dy) = \sum_{i=1}^d \int_{\mathbb{R}^d\times\mathbb{R}^d} |x_i-y_i|\, M(dx,dy) = \sum_{i=1}^d \int_{\mathbb{R}\times\mathbb{R}} |x_i-y_i|\, M_i(dx_i,dy_i).$$

For all $1 \leq i \leq d$, let $\mathcal{Q}_i$ denote the set $\mathcal{Q}$ with respect to $\mu = \mu_i$ and $\nu = \nu_i$ and let $Q_i \in \mathcal{Q}_i$. Then for $M_1 = M^{Q_1}, \cdots, M_d = M^{Q_d}$, we deduce from Theorem 2.2.12 that

$$\mathcal{M}_1(\mu,\nu) \leq \sum_{i=1}^d \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i| \, M^{Q_i}(dx_i, dy_i) \leq 2 \sum_{i=1}^d \mathcal{W}_1(\mu_i,\nu_i).$$

Let $P \in \Pi(\mu,\nu)$ be a coupling between μ and ν. For all $1 \leq i \leq d$, let P_i be the marginals of P with respect to the coordinates i and $i+d$, so that P_i is a coupling between μ_i and ν_i. Then

$$\begin{aligned}\sum_{i=1}^d \mathcal{W}_1(\mu_i,\nu_i) &\leq \sum_{i=1}^d \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i| \, P_i(dx_i, dy_i) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \sum_{i=1}^d |x_i - y_i| \, P(dx,dy) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y| \, P(dx,dy).\end{aligned}$$

Since the inequality above is true for any coupling P between μ and ν, we deduce that $\sum_{i=1}^d \mathcal{W}_1(\mu_i,\nu_i) \leq \mathcal{W}_1(\mu,\nu)$, which proves the assertion. ☐

In the following remarks, we first look in which case the minimiser of (2.2.22), studied by Hobson and Klimmek [109], derives from $\mathcal{Q}$. Second, we see that the left-curtain martingale coupling introduced by Beiglböck and Juillet [25] does not always satisfy (2.2.22).

Remark 2.2.14. The optimal martingale coupling $M \in \Pi^{\mathrm{M}}(\mu,\nu)$ which minimises $\int_{\mathbb{R}\times\mathbb{R}} |x-y| \, M(dx,dy)$ was actually characterised by Hobson and Klimmek [109] under the dispersion assumption that there exists a bounded interval E of positive length such that $(\mu-\nu)^+(E^\complement) = (\nu-\mu)^+(E) = 0$. They show that the optimal coupling M^{HK} is unique. Moreover, in the simpler case where $\mu \wedge \nu = 0$, if $a < b$ denote the endpoints of E, then there exist two nonincreasing functions $R : (0,1) \to (-\infty, a]$ and $S : (0,1) \to [b, +\infty)$ such that for all $u \in (0,1)$, denoting $\widetilde{m}^{HK}(u,dy) = m^{HK}(F_\mu^{-1}(u), dy)$ where $m^{HK}(x,dy)\,\mu(dx) = M^{HK}(dx,dy)$, one has

$$\widetilde{m}^{HK}(u,dy) = \frac{S(u) - F_\mu^{-1}(u)}{S(u) - R(u)} \delta_{R(u)}(dy) + \frac{F_\mu^{-1}(u) - R(u)}{S(u) - R(u)} \delta_{S(u)}(dy).$$

We can discuss in which case M^{HK} derives from $\mathcal{Q}$. Suppose first that F_ν^{-1} takes at least three different values, that is there exist $u, v, w \in (0,1)$ such that $F_\nu^{-1}(u) < F_\nu^{-1}(v) < F_\nu^{-1}(w)$. By left-continuity of F_ν^{-1}, there exists $\varepsilon > 0$ such that $F_\nu^{-1}(u) < F_\nu^{-1}(v-\varepsilon)$ and $F_\nu^{-1}(v) < F_\nu^{-1}(w-\varepsilon)$. Let $I_1 = (0,u]$, $I_2 = (v-\varepsilon, v]$ and $I_3 = (w-\varepsilon, 1]$. Those three intervals are such that for all $s \in I_1$ (resp. $s \in I_2$) and $t \in I_2$ (resp. $t \in I_3$), we have $F_\nu^{-1}(s) < F_\nu^{-1}(t)$. Since R is nonincreasing, if the graph of R meets the graph of F_ν^{-1} on one of those three intervals, then they cannot meet on the two others. We can assert the same with the graph of S since S is nonincreasing as well. Therefore, there exists $k \in \{1,2,3\}$ such that the intersection of $F_\nu^{-1}(I_k)$ and $R(I_k) \cup S(I_k)$ is empty. In particular, for all $t \in I_k$, $\widetilde{m}^{HK}(t, \{F_\nu^{-1}(t)\}) = 0$. However, thanks to Lemma 2.2.5, we can see that for all $Q \in \mathcal{Q}$, the Markov kernel $\widetilde{m}^Q$ is such that $\widetilde{m}^Q(u, \{F_\nu^{-1}(u)\}) > 0$ for du-almost all $u \in (0,1)$. Therefore, M^{HK} does not derive from $\mathcal{Q}$.

If F_ν^{-1} does not take more than two different values, that is if ν is reduced to two atoms at most, then there exists a unique martingale coupling between μ and ν, so M^{HK} derives of course from $\mathcal{Q}$.

Note that the maximisation problem $\sup_{M\in\Pi^{\mathrm{M}}(\mu,\nu)}\int_{\mathbb{R}\times\mathbb{R}}|x-y|\,M(dx,dy)$ is discussed by Hobson and Neuberger [110].

Example 2.2.15. For instance, if $\mu(dx)=\frac{1}{2}\mathbb{1}_{[-1,1]}(x)\,dx$ and $\nu(dx)=\frac{1}{2}(\mathbb{1}_{[-2,-1)}+\mathbb{1}_{(1,2]})(x)\,dx$, then (see Example 6.1 [109] for an equivalent calculation)

$$m^{HK}(x,dy)=\left(\frac{1}{2}-\frac{3x}{2\sqrt{12-3x^2}}\right)\delta_{-\frac{1}{2}(x+\sqrt{12-3x^2})}(dy)+\left(\frac{1}{2}+\frac{3x}{2\sqrt{12-3x^2}}\right)\delta_{\frac{1}{2}(-x+\sqrt{12-3x^2})}(dy),$$

which satisfies $m^{HK}(x,\{F_\nu^{-1}(F_\mu(x))\})>0$ iff $x\in\{(3-\sqrt{33})/6,(\sqrt{33}-3)/6\}$. On the other hand, for all $Q\in\mathcal{Q}$, the Markov kernel m^Q is such that $m^Q(x,\{F_\nu^{-1}(F_\mu(x))\})>0$ for dx-almost all $x\in(-1,1)$.

Remark 2.2.16. We investigate an example where the left-curtain martingale coupling introduced by Beiglböck and Juillet [25] does not satisfy (2.2.21). Let $\mu\in\mathcal{P}_1(\mathbb{R})$ be with density f_μ and let $u>1$ and $d>0$. Let M^{LC} be defined by

$$M^{LC}(dx,dy)=\mu(dx)\,\left(\mathbb{1}_{\{x>0\}}\,(q\,\delta_{ux}(dy)+(1-q)\,\delta_{-dx}(dy))+\mathbb{1}_{\{x\le 0\}}\delta_x(dy)\right),$$

where $q=\frac{1+d}{u+d}$. Let ν denote the second marginal of M^{LC}. So ν has density f_ν defined by $f_\nu(x)=\frac{q}{u}f_\mu(\frac{x}{u})$ for all $x>0$ and $f_\nu(x)=f_\mu(x)+\frac{1-q}{d}f_\mu(-\frac{x}{d})$ for all $x\le 0$. Then M^{LC} is the left-curtain martingale coupling between μ and ν. One can easily compute $\int_{\mathbb{R}^d}|y-x|\,M^{LC}(dx,dy)=2\frac{(u-1)(1+d)}{u+d}\int_{\mathbb{R}_+}xf_\mu(x)\,dx$. On the other hand, $\mathcal{W}_1(\mu,\nu)=\int_{\mathbb{R}}|F_\mu(t)-F_\nu(t)|\,dt$ (see for instance Remark 2.19 (iii) Chapter 2 [190]). From the relation between f_ν and f_μ, one can deduce that for all $x\ge 0$, $F_\nu(x)=1-q+qF_\mu(x/u)$, and for all $x\le 0$, $F_\nu(x)=F_\mu(x)+(1-q)\overline{F}_\mu(-x/d)$, where $\overline{F}_\mu:x\mapsto\mu((x,+\infty))=1-F_\mu(x)$. Using $|x|=x+2x^-$, we have

$$\begin{aligned}
\mathcal{W}_1(\mu,\nu)&=\int_{\mathbb{R}_-}(1-q)\overline{F}_\mu(-x/d)\,dx+\int_{\mathbb{R}_+}|\overline{F}_\mu(x)-q\overline{F}_\mu(x/u)|\,dx\\
&=\int_{\mathbb{R}_-}(1-q)\overline{F}_\mu(-x/d)\,dx+\int_{\mathbb{R}_+}(\overline{F}_\mu(x)-q\overline{F}_\mu(x/u))\,dx\\
&\quad+2\int_{\mathbb{R}_+}(\overline{F}_\mu(x)-q\overline{F}_\mu(x/u))^-\,dx\\
&=d(1-q)\int_{\mathbb{R}+}xf_\mu(x)\,dx+(1-qu)\int_{\mathbb{R}_+}xf_\mu(x)\,dx+2\int_{\mathbb{R}_+}(\overline{F}_\mu(x)-q\overline{F}_\mu(x/u))^-\,dx\\
&=2\int_{\mathbb{R}_+}(\overline{F}_\mu(x)-q\overline{F}_\mu(x/u))^-\,dx.
\end{aligned}$$

Then M^{LC} satisfies (2.2.21) iff

$$\frac{(u-1)(1+d)}{u+d}\int_{\mathbb{R}_+}xf_\mu(x)\,dx\le 2\int_{\mathbb{R}_+}\left(\bar{F}_\mu(x)-q\bar{F}_\mu(x/u)\right)^-\,dx. \tag{2.2.23}$$

The next example illustrates a contradiction of (2.2.23) and therefore (2.2.21) for M^{LC}.

Example 2.2.17. Let $\mu(dx) = \lambda \exp(-\lambda x) \mathbb{1}_{\{x>0\}}\, dx$, where $\lambda > 0$, and let ν be the probability distribution with density f_ν given by $f_\nu(x) = \frac{q}{u} f_\mu(x/u)$ for $x > 0$ and $f_\nu(x) = \frac{1-q}{d} f_\mu(-x/d)$ for $x \leq 0$. Then for all $x \in \mathbb{R}$, $\overline{F}_\mu(x) = \exp(-\lambda x)$, and (2.2.23) is equivalent to

$$\begin{aligned}
\frac{(u-1)(1+d)}{u+d} \times \frac{1}{\lambda} &> 2\int_{\mathbb{R}_+} \left(\exp(-\lambda x) - q\exp(-\lambda x/u)\right)^-\, dx \\
&= 2\int_{\frac{\ln q}{\lambda(\frac{1}{u}-1)}}^{+\infty} \left(q\exp(-\lambda x/u) - \exp(-\lambda x)\right) dx \\
\iff \frac{(u-1)q}{\lambda} &> 2\left(\frac{qu}{\lambda}\exp\left(-\frac{\ln q}{1-u}\right) - \frac{1}{\lambda}\exp\left(-\frac{\ln q}{\frac{1}{u}-1}\right)\right) = 2\frac{q}{\lambda}(u-1)q^{-1/(1-u)} \\
\iff 2^{1-u} &> q = \frac{1+d}{u+d},
\end{aligned}$$

which can be satisfied for example with $u = \frac{5}{4}$ and $d = \frac{1}{4}$. Note that this condition does not depend on the value of λ. Therefore, the left-curtain martingale coupling

$$M^{LC}(dx, dy) = \lambda \exp(-\lambda x)\mathbb{1}_{\{x>0\}}\, dx \left(\frac{5}{6}\,\delta_{\frac{5x}{4}}(dy) + \frac{1}{6}\,\delta_{-\frac{x}{4}}(dy)\right)$$

does not satisfy (2.2.21), for any $\lambda > 0$.

Proposition 2.2.18. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Let $Q \in \mathcal{Q}$. Then the Markov kernel $(\widetilde{m}^Q(u, dy))_{u\in(0,1)}$ minimises*

$$\int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\, \widetilde{m}(u, dy)\, du$$

among all Markov kernels $(\widetilde{m}(u, dy))_{u\in(0,1)}$ such that

$$\left\{\begin{array}{rcl} \int_{u\in(0,1)} \widetilde{m}(u, dy)\, du &=& \nu(dy) \\ \int_{\mathbb{R}} |y|\, \widetilde{m}(u, dy) < +\infty \quad \text{and} \quad \int_{\mathbb{R}} y\, \widetilde{m}(u, dy) &=& F_\mu^{-1}(u),\ du\text{-almost everywhere on } (0,1) \end{array}\right. \tag{2.2.24}$$

Moreover, $\int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\, \widetilde{m}^Q(u, dy)\, du = \mathcal{W}_1(\mu, \nu)$.

Remark 2.2.19. If $(\widetilde{m}(u, dy))_{u\in(0,1)}$ is a Markov kernel satisfying (2.2.24), then using Lemma 2.2.6, we get that $\mu(dx)\, m(x, dy)$ with $(m(x, dy))_{x\in\mathbb{R}}$ denoting the Markov kernel given by (2.2.6) is a martingale coupling between μ and ν.

Conversely, if $\mu(dx)\, m(x, dy)$ is a martingale coupling between μ and ν, then using the inverse transform sampling, we get that the Markov kernel $(m(F_\mu^{-1}(u), dy))_{u\in(0,1)}$ satisfies (2.2.24).

Remark 2.2.20. The martingale couplings parametrised by $Q \in \mathcal{Q}$ are not the only ones to minimise $\int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\, \widetilde{m}(u, dy)\, du$ among all Markov kernels $(\widetilde{m}(u, dy))_{u\in(0,1)}$ which satisfy (2.2.24). Indeed, let $\mu = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, $\nu = \frac{1}{8}\delta_{-8} + \frac{1}{4}\delta_{-6} + \frac{5}{8}\delta_4$ and

$$M = \frac{1}{8}\left(2\delta_{(-1,-6)} + 2\delta_{(-1,4)} + \delta_{(1,-8)} + 3\delta_{(1,4)}\right).$$

For $m(-1,dy)=\frac{1}{2}\delta_{-6}+\frac{1}{2}\delta_4$ and $m(1,dy)=\frac{1}{4}\delta_{-8}+\frac{3}{4}\delta_4$, we have $M(dx,dy)=\mu(dx)\,m(x,dy)$. Let $(\widetilde{m}(u,dy))_{u\in(0,1)}$ be defined for all $u\in(0,1)$ by $\widetilde{m}(u,dy)=m(F_\mu^{-1}(u),dy)$. It is easy to see that M is a martingale coupling between μ and ν, so $(\widetilde{m}(u,dy))_{u\in(0,1)}$ satisfies (2.2.24). For all $u\in(0,1)$, we have $F_\mu^{-1}(u)=\mathbb{1}_{\{u\le 1/2\}}(-1)+\mathbb{1}_{\{u>1/2\}}$ and $F_\nu^{-1}(u)=\mathbb{1}_{\{u\le 1/8\}}(-8)+\mathbb{1}_{\{1/8<u\le 3/8\}}(-6)+\mathbb{1}_{\{u>3/8\}}\times 4$. So for all $u\in(0,1)$, we have

$$\widetilde{m}(u,dy)=\mathbb{1}_{\{u\le\frac{1}{2}\}}\left(\frac{1}{2}\delta_{-6}+\frac{1}{2}\delta_4\right)+\mathbb{1}_{\{u>\frac{1}{2}\}}\left(\frac{1}{4}\delta_{-8}+\frac{3}{4}\delta_4\right).$$

We can compute $\int_0^1\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}(u,dy)\,du=\frac{17}{4}=\int_0^1|F_\mu^{-1}(u)-F_\nu^{-1}(u)|\,du=\mathcal{W}_1(\mu,\nu)$, so $(\widetilde{m}(u,dy))_{u\in(0,1)}$ is optimal.

Thanks to Lemma 2.2.5, we can see that for all $Q\in\mathcal{Q}$, the Markov kernel $\widetilde{m}^Q$ is such that $\widetilde{m}^Q(u,\{F_\nu^{-1}(u)\})>0$ for du-almost all $u\in(0,1)$. However for all $u\in(0,1/8]$, we have $\widetilde{m}(u,\{F_\nu^{-1}(u)\})=\widetilde{m}(u,\{-8\})=0$. Therefore, $\widetilde{m}$ does not derive from $\mathcal{Q}$.

Proof of Proposition 2.2.18. Let $\widetilde{m}$ be a Markov kernel satisfying (2.2.24). By Jensen's inequality, for du-almost every $u\in(0,1)$,

$$|F_\nu^{-1}(u)-F_\mu^{-1}(u)|=\left|\int_{\mathbb{R}}(F_\nu^{-1}(u)-y)\,\widetilde{m}(u,dy)\right|\le\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}(u,dy).$$

So $\int_0^1|F_\nu^{-1}(u)-F_\mu^{-1}(u)|\,du\le\int_0^1\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}(u,dy)\,du$.

Therefore, to conclude, it is sufficient to prove that $\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}^Q(u,dy)=|F_\nu^{-1}(u)-F_\mu^{-1}(u)|$, du-almost everywhere on $(0,1)$.

Applying (2.2.9) to the measurable and nonnegative function $h:y\mapsto|F_\nu^{-1}(u)-y|$ yields for du-almost all $u\in(0,1)$

$$\begin{aligned}\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}^Q(u,dy)&=\int_{(0,1)}\frac{(F_\mu^{-1}-F_\nu^{-1})^+(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}|F_\nu^{-1}(u)-F_\nu^{-1}(v)|\pi_+^Q(u,dv)\\&\quad+\int_{(0,1)}\frac{(F_\mu^{-1}-F_\nu^{-1})^-(u)}{F_\nu^{-1}(u)-F_\nu^{-1}(v)}|F_\nu^{-1}(u)-F_\nu^{-1}(v)|\pi_-^Q(u,dv).\end{aligned}$$

Using Lemma 2.2.5, we deduce that for du-almost all $u\in(0,1)$

$$\begin{aligned}\int_{\mathbb{R}}|F_\nu^{-1}(u)-y|\,\widetilde{m}^Q(u,dy)&=\int_{(0,1)}(F_\mu^{-1}-F_\nu^{-1})^+(u)\pi_+^Q(u,dv)+\int_{(0,1)}(F_\mu^{-1}-F_\nu^{-1})^-(u)\pi_-^Q(u,dv)\\&=|F_\nu^{-1}(u)-F_\mu^{-1}(u)|.\end{aligned}$$

□

Proof of Theorem 2.2.12. Let $Q\in\mathcal{Q}$ and let $\widetilde{m}^Q$ be the Markov kernel defined by (2.2.5). By Lemma 2.2.6 and Proposition 2.2.18,

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}}|y-x|\,\mu(dx)\,m^Q(x,dy)&=\int_0^1\int_{\mathbb{R}}|y-F_\mu^{-1}(u)|\,\widetilde{m}^Q(u,dy)\,du\\&\le\int_0^1\int_{\mathbb{R}}|y-F_\nu^{-1}(u)|\,\widetilde{m}^Q(u,dy)\,du\end{aligned}$$

$$
\begin{aligned}
&+\int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - F_\mu^{-1}(u)| \, \widetilde{m}^Q(u, dy)\, du \\
&= 2\mathcal{W}_1(\mu, \nu).
\end{aligned}
$$

Since $M^Q(dx, dy) = \mu(dx)\, m^Q(x, dy)$ is a martingale coupling between μ and ν (Proposition 2.2.3), we get (2.2.22).

Let us show now that the constant 2 is sharp, that is

$$
\sup_{\substack{\mu,\nu \in \mathcal{P}_1(\mathbb{R}) \\ \mu <_{cx} \nu}} \frac{\mathcal{M}_1(\mu, \nu)}{\mathcal{W}_1(\mu, \nu)} = 2.
$$

Let $a, b \in \mathbb{R}$ be such that $0 < a < b$. Let $\mu = \frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$ and $\nu = \frac{1}{2}\delta_{-b} + \frac{1}{2}\delta_b$. Since μ and ν are two probability measures with equal means such that μ is concentrated on $[-a, a]$ and ν on $\mathbb{R}\backslash[-a, a]$, then $\mu <_{cx} \nu$. Any coupling H between μ and ν is of the form

$$
H = r\delta_{(-a,-b)} + r'\delta_{(-a,b)} + p\delta_{(a,b)} + p'\delta_{(a,-b)},
$$

where $r, r', p, p' \geq 0$ and $p + p' = r + r' = p + r' = p' + r = 1/2$. One can easily see that H is a martingale coupling iff $b(p - p') = a/2$ and $b(r' - r) = -a/2$, that is

$$
H = \frac{(b + a)}{4b}\delta_{(-a,-b)} + \frac{(b - a)}{4b}\delta_{(-a,b)} + \frac{(b + a)}{4b}\delta_{(a,b)} + \frac{(b - a)}{4b}\delta_{(a,-b)}. \tag{2.2.25}
$$

Since there is only one martingale coupling, we trivially have

$$
\mathcal{M}_1(\mu, \nu) = \int_{\mathbb{R}\times\mathbb{R}} |x - y| \, H(dx, dy) = \frac{b^2 - a^2}{b}\cdot
$$

On the other hand, since $\mathcal{W}_1(\mu, \nu) = \int_{\mathbb{R}} |F_\mu(t) - F_\nu(t)|\, dt$ (see for instance Remark 2.19 (iii) Chapter 2 [190]),

$$
\mathcal{W}_1(\mu, \nu) = \int_{-\infty}^{-b} 0\, dt + \int_{-b}^{-a} \frac{1}{2}\, dt + \int_{-a}^{a} 0\, dt + \int_a^b \frac{1}{2}\, dt + \int_b^{+\infty} 0\, dt = b - a.
$$

So, we have

$$
\frac{\mathcal{M}_1(\mu, \nu)}{\mathcal{W}_1(\mu, \nu)} = 1 + \frac{a}{b},
$$

which tends to 2 as b tends to a. $\square$

Also, the stability inequality (2.2.22) does not generalise with $\mathcal{M}_1(\mu, \nu)$ and $\mathcal{W}_1(\mu, \nu)$ replaced with $\mathcal{M}_\rho(\mu, \nu)$ and $\mathcal{W}_\rho(\mu, \nu)$ for $\rho > 1$, as shown in the next proposition in general dimension.

Proposition 2.2.21. *Let $d \geq 1$ and $\rho > 1$. Then*

$$
\sup_{\substack{\mu,\nu \in \mathcal{P}_\rho(\mathbb{R}^d) \\ \mu <_{cx} \nu}} \frac{\mathcal{M}_\rho(\mu, \nu)}{\mathcal{W}_\rho(\mu, \nu)} = +\infty.
$$

The proof of Proposition 2.2.21 will use the following lemma for the case $1 < \rho < 2$.

Lemma 2.2.22. *Let $d \geq 1$ and $\rho \in (1,2)$. Let $|\cdot|$ denote the Euclidean norm on $\mathbb{R}^d$. Then there exists $C_\rho > 0$ such that*

$$\forall (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, \quad |x-y|^\rho \geq C_\rho \left(|x|^\rho - \frac{\rho}{\rho-1} |x|^{\rho-2} \langle x, y \rangle_{\mathbb{R}^d} + \frac{1}{\rho-1} |y|^\rho \right), \qquad (2.2.26)$$

where, by convention, for all $y \in \mathbb{R}^d$ and for $x = 0$ we choose $|x|^{\rho-2}\langle x, y\rangle_{\mathbb{R}^d}$ equal to its limit 0 as $x \to 0$.

When $\rho = 2$, both sides of the inequality are equal with $C_2 = 1$.

Proof of Lemma 2.2.22. If $x = 0$, any $C_\rho \leq \rho - 1$ suits. Else, dividing by $|x|^\rho$ and using that $y/|x|$ explores $\mathbb{R}^d$ when y explores $\mathbb{R}^d$, we see that the statement reduces to show that for all $x, y \in \mathbb{R}^d$ such that $|x| = 1$,

$$|x-y|^\rho \geq C_\rho \left(1 - \frac{\rho}{\rho-1} \langle x, y \rangle_{\mathbb{R}^d} + \frac{1}{\rho-1} |y|^\rho \right).$$

For all $x, y \in \mathbb{R}^d$ such that $|x| = 1$, there exist $y_1, y_2 \in \mathbb{R}$ such that $y = y_1 x + y_2 x^\perp$, where $x^\perp$ is an element of $\operatorname{span}(x)^\perp$ such that $|x^\perp| = 1$. The inequality to prove becomes

$$\forall (y_1, y_2) \in \mathbb{R}^2, \quad \left((1-y_1)^2 + y_2^2 \right)^{\rho/2} \geq C_\rho \left(1 - \frac{\rho}{\rho-1} y_1 + \frac{1}{\rho-1} (y_1^2 + y_2^2)^{\rho/2} \right). \qquad (2.2.27)$$

Let $L : (y_1, y_2) \mapsto ((1-y_1)^2 + y_2^2)^{\rho/2}$ and $R : (y_1, y_2) \mapsto 1 - \frac{\rho}{\rho-1} y_1 + \frac{1}{\rho-1}(y_1^2 + y_2^2)^{\rho/2}$. When $(y_1, y_2) \to (1, 0)$, we have

$$\begin{aligned}
R(y_1, y_2) &= \frac{1}{\rho-1} \left(\rho - 1 - \rho(y_1 - 1 + 1) + (1 + 2(y_1 - 1) + (y_1-1)^2 + y_2^2)^{\rho/2} \right) \\
&= \frac{1}{\rho-1} \Big(-1 - \rho(y_1 - 1) + 1 + \rho(y_1 - 1) + \frac{\rho}{2}(y_1-1)^2 + \frac{\rho}{2} y_2^2 \\
&\qquad + \rho(\frac{\rho}{2} - 1)(y_1 - 1)^2 + o((y_1-1)^2 + y_2^2) \Big) \\
&= \frac{1}{\rho-1} \left(\frac{\rho}{2}(y_1-1)^2 + \frac{\rho}{2} y_2^2 - \rho(1 - \frac{\rho}{2})(y_1-1)^2 + o((y_1-1)^2 + y_2^2) \right).
\end{aligned}$$

Since $\rho < 2$, $L(y_1, y_2) \geq (1-y_1)^2 + y_2^2$ for any (y_1, y_2) in the ball centred at $(1, 0)$ with radius 1. So

$$\limsup_{\substack{(y_1,y_2) \to (1,0) \\ (y_1,y_2) \neq (1,0)}} \frac{R(y_1, y_2)}{L(y_1, y_2)} \leq \frac{\rho}{2(\rho-1)},$$

On the other hand, when $y_1^2 + y_2^2 \to +\infty$,

$$\frac{R(y_1, y_2)}{L(y_1, y_2)} \sim \frac{(y_1^2 + y_2^2)^{\rho/2}}{(\rho-1)(y_1^2 + y_2^2)^{\rho/2}} = \frac{1}{\rho-1}.$$

So $(y_1, y_2) \mapsto R(y_1, y_2)/L(y_1, y_2)$ is defined and continuous on $(\mathbb{R}^d \times \mathbb{R}^d) \backslash \{(1,0)\}$, bounded from above in the ball centred at $(1,0)$ with radius 1 and has a finite limit when the norm of (y_1, y_2) tends to $+\infty$. Therefore this function is bounded from above on $(\mathbb{R}^d \times \mathbb{R}^d) \backslash \{(1,0)\}$ by a certain constant $K \geq \frac{1}{\rho-1}$. Since both sides of (2.2.27) vanish for $(y_1, y_2) = (1,0)$, we conclude that this inequality holds with constant $C_\rho = \frac{1}{K}$ and (2.2.26) with constant $C_\rho = \frac{1}{K}$. □

Proof of Proposition 2.2.21. Since all norms on $\mathbb{R}^d$ are equivalent, we can suppose that $\mathbb{R}^d$ is endowed with the Euclidean norm. The case $\rho \geq 2$ was addressed in the introduction in the one dimensional case. Its extension to dimension d is immediate. Indeed, for all $n \in \mathbb{N}^*$, let $\mu_n = \mathcal{N}_1(0, n^2)$ and $\mu'_n(dx_1, \cdots, dx_d) = (x_1, 0, \cdots, 0)_\sharp \mu_n(dx_1)$ where $\sharp$ denotes the pushforward operation. By reduction to the one dimensional case, we have

$$\frac{\mathcal{M}_\rho(\mu'_n, \mu'_{n+1})}{\mathcal{W}_\rho(\mu'_n, \mu'_{n+1})} = \frac{\mathcal{M}_\rho(\mu_n, \mu_{n+1})}{\mathcal{W}_\rho(\mu_n, \mu_{n+1})} \underset{n\to+\infty}{\longrightarrow} +\infty.$$

We now consider the case $1 < \rho < 2$. Let $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$ be such that $\mu <_{cx} \nu$, and let M be a martingale coupling between μ and ν, which exists according to Strassen's theorem or Proposition 2.2.3 and Corollary 2.2.4. Thanks to Lemma 2.2.22, there exists $C_\rho > 0$ such that

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^\rho \, M(dx,dy) \geq C_\rho \Bigg(&\int_{\mathbb{R}^d} |x|^\rho \, \mu(dx) - \frac{\rho}{\rho-1} \int_{\mathbb{R}^d\times\mathbb{R}^d} |x|^{\rho-2} \langle x, y\rangle_{\mathbb{R}^d} \, M(dx,dy) \\ &+ \frac{1}{\rho-1} \int_{\mathbb{R}^d} |y|^\rho \, \nu(dx) \Bigg).\end{aligned}$$

Since $M(dx,dy) = \mu(dx)\, m(x,dy)$ is a martingale coupling, we have for $\mu(dx)$-almost all $x \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} |x|^{\rho-2}\langle x, y\rangle_{\mathbb{R}^d} \, m(x,dy) = |x|^\rho$, where both sides are equal to 0 when $x = 0$. So we get

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^\rho \, M(dx,dy) \geq \frac{C_\rho}{\rho-1} \left(\int_{\mathbb{R}^d} |y|^\rho \, \nu(dx) - \int_{\mathbb{R}^d} |x|^\rho \, \mu(dx) \right).$$

For all $n \in \mathbb{N}$, let $\mu_n = \mathcal{N}_d(0, n^2 I_d)$. Let $G \sim \mathcal{N}_d(0, I_d)$. Then for all $n \in \mathbb{N}$, $\mathcal{W}^\rho_\rho(\mu_n, \mu_{n+1}) \leq \mathbb{E}[|G|^\rho]$ and

$$\begin{aligned}\frac{\mathcal{M}^\rho_\rho(\mu_n, \mu_{n+1})}{\mathcal{W}^\rho_\rho(\mu_n, \mu_{n+1})} &\geq \frac{C_\rho}{\rho-1} \frac{(\mathbb{E}[|(n+1)G|^\rho] - \mathbb{E}[|nG|^\rho])}{\mathbb{E}[|G|^\rho]} \\ &= \frac{((n+1)^\rho - n^\rho) C_\rho}{\rho-1} \\ &\sim_{n\to+\infty} \frac{\rho C_\rho}{\rho-1} n^{\rho-1} \underset{n\to+\infty}{\longrightarrow} +\infty.\end{aligned}$$

□

2.3 The inverse transform martingale coupling

2.3.1 Definition and stability of the inverse transform martingale coupling

Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be such that $\mu<_{cx}\nu$. We recall that Ψ_+ and Ψ_- are defined for all $u\in[0,1]$ by $\Psi_+(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv$ and $\Psi_-(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv$. Let Ψ_-^{-1} (resp. Ψ_+^{-1}) denote the left continuous generalised inverse of Ψ_- (resp. Ψ_+). Let $\varphi:[0,1]\to[0,1]$ and $\widetilde{\varphi}:[0,1]\to[0,1]$ be defined for all $u\in[0,1]$ by

$$\begin{aligned}\varphi(u)&=\Psi_-^{-1}(\Psi_+(u))=\inf\{r\in[0,1]\mid\Psi_-(r)\geq\Psi_+(u)\};\\ \widetilde{\varphi}(u)&=\Psi_+^{-1}(\Psi_-(u))=\inf\{r\in[0,1]\mid\Psi_+(r)\geq\Psi_-(u)\},\end{aligned}$$

which are well defined thanks to the equality $\Psi_-(1)=\Psi_+(1)$, consequence of the equality of the means.

Let Q^{IT} be the measure defined on $(0,1)^2$ by

$$Q^{IT}(du,dv)=\frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du\,\pi_+^{IT}(u,dv)\quad\text{where }\pi_+^{IT}(u,dv)=\mathbb{1}_{\{0<\varphi(u)<1\}}\,\delta_{\varphi(u)}(dv),\tag{2.3.1}$$

with $\gamma=\Psi_-(1)=\Psi_+(1)$. According to the next proposition, this measure belongs to $\mathcal{Q}$.

Proposition 2.3.1. *Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be such that $\mu<_{cx}\nu$. The measure Q^{IT} is an element of $\mathcal{Q}$ as defined in Section 2.2. Moreover,*

$$Q^{IT}(du,dv)=\frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv\,\pi_-^{IT}(v,du)\quad\textit{where }\pi_-^{IT}(v,du)=\mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,\delta_{\widetilde{\varphi}(v)}(du).$$

Let us then write $(\widetilde{m}^{IT}(u,dy))_{u\in(0,1)}$ instead of $(\widetilde{m}^{Q^{IT}}(u,dy))_{u\in(0,1)}$ and $(m^{IT}(x,dy))_{x\in\mathbb{R}}$ instead of $(m^{Q^{IT}}(x,dy))_{x\in\mathbb{R}}$. By Proposition 2.2.3, the probability measure $M^{IT}(dx,dy)=\mu(dx)\,m^{IT}(x,dy)$ is a martingale coupling between μ and ν, which we call the inverse transform martingale coupling.

We deduce from the expression of π_-^{IT} given in Proposition 2.3.1 that the definition of $(\widetilde{m}^{IT}(u,dy))_{u\in(0,1)}$ reduces to

$$\begin{cases}\dfrac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u))-F_\nu^{-1}(u)}\delta_{F_\nu^{-1}(\varphi(u))}(dy)+\left(1-\dfrac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u))-F_\nu^{-1}(u)}\right)\delta_{F_\nu^{-1}(u)}(dy)\\ \qquad\qquad\text{if }F_\nu^{-1}(\varphi(u))>F_\mu^{-1}(u)>F_\nu^{-1}(u)\text{ and }\varphi(u)<1;\\ \dfrac{F_\nu^{-1}(u)-F_\mu^{-1}(u)}{F_\nu^{-1}(u)-F_\nu^{-1}(\widetilde{\varphi}(u))}\delta_{F_\nu^{-1}(\widetilde{\varphi}(u))}(dy)+\left(1-\dfrac{F_\nu^{-1}(u)-F_\mu^{-1}(u)}{F_\nu^{-1}(u)-F_\nu^{-1}(\widetilde{\varphi}(u))}\right)\delta_{F_\nu^{-1}(u)}(dy)\\ \qquad\qquad\text{if }F_\nu^{-1}(\widetilde{\varphi}(u))<F_\mu^{-1}(u)<F_\nu^{-1}(u)\text{ and }\widetilde{\varphi}(u)<1;\\ \delta_{F_\nu^{-1}(u)}(dy)\qquad\text{otherwise.}\end{cases}\tag{2.3.2}$$

Note that if $F_\mu^{-1}(u) > F_\nu^{-1}(u)$, then by left-continuity of F_μ^{-1} and F_ν^{-1}, $\Psi_+(u) > 0$, which implies $\varphi(u) > 0$. Therefore $F_\mu^{-1}(u) > F_\nu^{-1}(u)$ implies $\varphi(u) > 0$ so that with the condition $\varphi(u) < 1$, $F_\nu^{-1}(\varphi(u))$ makes sense. For similar reasons, if $F_\mu^{-1}(u) < F_\nu^{-1}(u)$ and $\widetilde{\varphi}(u) < 1$ then $F_\nu^{-1}(\widetilde{\varphi}(u))$ makes sense.

Remark 2.3.2. We recall the celebrated Strassen theorem: if $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, then $\mu \le_{cx} \nu$ iff there exists a martingale coupling between μ and ν. The sufficient condition is a straightforward consequence of Jensen's inequality. Indeed, if $M(dx, dy) = \mu(dx)\, m(x, dy)$ is a martingale coupling between $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, then for all convex functions $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}} f(x)\,\mu(dx) = \int_{\mathbb{R}} f\left(\int_{\mathbb{R}} y\, m(x, dy)\right)\mu(dx) \le \int_{\mathbb{R}^2} f(y)\, m(x, dy)\,\mu(dx) = \int_{\mathbb{R}} f(y)\,\nu(dy).$$

Conversely, suppose that $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are such that $\mu \le_{cx} \nu$. For $t \in \mathbb{R}$, $\int_{\mathbb{R}}(t - x)^+\,\mu(dx) \le \int_{\mathbb{R}}(t - x)^+\,\nu(dx)$ by convexity of $x \in \mathbb{R} \mapsto (t - x)^+$. By the Fubini-Tonelli theorem, $\int_{\mathbb{R}}(t - x)^+\,\mu(dx) = \int_{-\infty}^t F_\mu(x)\,dx$. Hence $\varphi_\mu(t) = \int_{-\infty}^t F_\mu(x)\,dx \le \varphi_\nu(t) = \int_{-\infty}^t F_\nu(x)\,dx$ for all $t \in \mathbb{R}$. Hence the respective Fenchel-Legendre transforms φ_μ^* and φ_ν^* of φ_μ and φ_ν satisfy $\varphi_\mu^* \ge \varphi_\nu^*$. For all $u \in (0, 1)$ and for all $t \in \mathbb{R}$, $F_\mu^{-1}(u) \le t \iff u \le F_\mu(t)$, so

$$\sup_{q \in [0,1]}\left(qt - \int_0^q F_\mu^{-1}(u)\,du\right) = \int_0^{F_\mu(t)}(t - F_\mu^{-1}(u))\,du = \int_0^1 (t - F_\mu^{-1}(u))^+\,du = \varphi_\mu(t).$$

Since $q \mapsto (\int_0^q F_\mu^{-1}(u)\,du)$ is convex on $[0, 1]$, we get the well known fact (see for instance Lemma A.22 [79]) that for all $q \in \mathbb{R}$, $\varphi_\mu^*(q) = (\int_0^q F_\mu^{-1}(u)\,du)\mathbb{1}_{[0,1]}(q) + (+\infty)\mathbb{1}_{[0,1]^\complement}(q)$. Hence

$$\int_0^q F_\mu^{-1}(u)\,du \ge \int_0^q F_\nu^{-1}(u)\,du \quad \text{for all } q \in [0, 1], \text{ with equality for } q = 1. \tag{2.3.3}$$

We will see in the proof of Proposition 2.3.1 that if $\mu \ne \nu$, then (2.3.3) implies that Q^{IT} belongs to $\mathcal{Q}$, which ensures that the inverse transform martingale coupling M^{IT} exists. If $\mu = \nu$, the existence of a martingale coupling is straightforward. Therefore, the construction of the inverse transform martingale coupling gives a constructive proof of the necessary condition in Strassen's theorem in dimension 1.

Proof of Proposition 2.3.1. By Lemma 2.6.1 below,

$$Q^{IT}((0, 1)^2) = \frac{1}{\gamma}\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\,du = \frac{1}{\gamma}\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^-(u)\mathbb{1}_{\{0<u<1\}}\,du = 1,$$

so Q^{IT} is a probability measure on $(0, 1)^2$. Let $h : (0, 1)^2 \to \mathbb{R}$ be a measurable and bounded function. We have

$$\int_{(0,1)^2} h(u, v)\, Q^{IT}(du, dv) = \frac{1}{\gamma}\int_{(0,1)} h(u, \varphi(u))(F_\mu^{-1} - F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\,du. \tag{2.3.4}$$

Since Ψ_- is continuous, one has $\Psi_-(\Psi_-^{-1}(u)) = u$ for all $u \in (0, 1)$. By Lemma 2.6.3 below, we deduce that $\widetilde{\varphi}(\varphi(u)) = u$, $(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$-almost everywhere on $(0, 1)$. Therefore, by Lemma 2.6.1 below,

$$\begin{aligned}
&\int_{(0,1)} h(u,\varphi(u))(F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\,du \\
&= \int_{(0,1)} h(\widetilde{\varphi}(\varphi(u)),\varphi(u))(F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\mathbb{1}_{\{0<\widetilde{\varphi}(\varphi(u))<1\}}\,du \\
&= \int_{(0,1)} h(\widetilde{\varphi}(v),v)(F_\mu^{-1}-F_\nu^{-1})^-(v)\mathbb{1}_{\{0<v<1\}}\mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,dv \\
&= \int_{(0,1)^2} h(u,v)(F_\mu^{-1}-F_\nu^{-1})^-(v)\mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,\delta_{\widetilde{\varphi}(v)}(du)\,dv.
\end{aligned} \tag{2.3.5}$$

So

$$\int_{(0,1)^2} h(u,v)\,Q^{IT}(du,dv) = \frac{1}{\gamma}\int_{(0,1)^2} h(u,v)(F_\mu^{-1}-F_\nu^{-1})^-(v)\mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,\delta_{\widetilde{\varphi}(v)}(du)\,dv.$$

Hence $Q^{IT}(du,dv) = \frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv\,\pi_-^{IT}(v,du)$, where $\pi_-^{IT}(v,du) = \mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,\delta_{\widetilde{\varphi}(v)}(du)$. Moreover, since Q^{IT} is a probability measure on $(0,1)^2$, it proves that

$$d\Psi_+(u)\text{-a.e. (resp. } d\Psi_-(u)\text{-a.e.)},\quad 0<\varphi(u)<1 \text{ (resp. } 0<\widetilde{\varphi}(u)<1). \tag{2.3.6}$$

Therefore, it is clear that Q^{IT} has first marginal $\frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du$ and second marginal $\frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv$. For $h:(u,v)\mapsto \mathbb{1}_{\{u<v\}}$, (2.3.4) writes

$$Q^{IT}\left(\{(u,v)\in(0,1)^2 \mid u<v\}\right) = \frac{1}{\gamma}\int_0^1 \mathbb{1}_{\{u<\varphi(u)\}}(F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\,du.$$

Let us show that $u<\varphi(u)$, $(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du$-almost everywhere on $(0,1)$. By the definition of φ and Lemma 2.6.3 below, for all $u\in(0,1)$, $\varphi(u)\le u \iff \Psi_-^{-1}(\Psi_+(u))\le u \iff \Psi_+(u)\le\Psi_-(u)$. Recall that since $\mu\le_{cx}\nu$, according (2.3.3), for all $u\in(0,1)$, $\int_0^u F_\mu^{-1}(v)\,dv \ge \int_0^u F_\nu^{-1}(v)\,dv$, so $\Psi_+(u)\ge\Psi_-(u)$. Therefore, we get that

$$\forall u\in(0,1),\quad \varphi(u)\le u \iff \Psi_+(u)=\Psi_-(u). \tag{2.3.7}$$

Suppose $F_\mu^{-1}(u)>F_\nu^{-1}(u)$. Since F_μ^{-1} and F_ν^{-1} are left continuous, this implies $F_\mu^{-1}(u-\varepsilon)>F_\nu^{-1}(u-\varepsilon)$ for $\varepsilon>0$ small enough. So, for $\varepsilon>0$ small enough, $\Psi_-(u)=\Psi_-(u-\varepsilon)\le\Psi_+(u-\varepsilon)<\Psi_+(u)$, which implies

$$u<\varphi(u),\quad (F_\mu^{-1}-F_\nu^{-1})^+(u)\,du\text{-almost everywhere on } (0,1). \tag{2.3.8}$$

So

$$Q^{IT}\left(\{(u,v)\in(0,1)^2 \mid u<v\}\right) = \frac{1}{\gamma}\int_0^1 (F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\varphi(u)<1\}}\,du = Q^{IT}((0,1)^2)=1,$$

since Q^{IT} is a probability measure on $(0,1)^2$. $\square$

We end this section with the stability of the inverse transform martingale coupling with respect to its marginals μ and ν for the Wasserstein distance topology. The following proposition is a direct consequence of Proposition 2.5.10, whose proof is given in the supermartingale setting. For the sake of generality, the only martingale coupling between a probability measure $\mu \in \mathcal{P}_1(\mathbb{R})$ and itself, namely $\mu(dx)\,\delta_x(dy)$, is still called inverse transform martingale coupling.

Proposition 2.3.3. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$. Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ be two sequences of probability measures on $\mathbb{R}$ with finite first moments such that for all $n \in \mathbb{N}$, $\mu_n \leq_{cx} \nu_n$. For all $n \in \mathbb{N}$, let M_n^{IT} (resp. M^{IT}) be the inverse transform martingale coupling between μ_n and ν_n (resp. between μ and ν).*

If $\mathcal{W}_1(\mu_n, \mu) \underset{n\to+\infty}{\longrightarrow} 0$ and $\mathcal{W}_1(\nu_n, \nu) \underset{n\to+\infty}{\longrightarrow} 0$, then

$$\mathcal{W}_1(M_n^{IT}, M^{IT}) \underset{n\to+\infty}{\longrightarrow} 0.$$

2.3.2 Optimality properties

Let us now suppose that $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are such that $\mu <_{cx} \nu$ and there exists $p \in (0,1)$ such that $u \mapsto \int_0^u (F_\mu^{-1}(v) - F_\nu^{-1}(v))\,dv$ is nondecreasing on $[0,p]$ and nonincreasing on $[p,1]$. We saw in Example 2.2.1 a concrete example of an element $Q_1 \in \mathcal{Q}$. Any probability measure Q defined on $(0,1)$ satisfying properties (i) and (ii) of the definition of $\mathcal{Q}$ is concentrated on $(0,p) \times (p,1)$ and therefore satisfies (iii). The probability measure Q_1 is a simple example that comes to mind. The inverse transform martingale coupling presented in this section is a valid example as well and inspires another coupling which is sort of the nonincreasing twin of the inverse transform martingale coupling.

Let $\chi_- : u \in [0,1] \mapsto \int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv = \int_0^u (F_\mu^{-1} - F_\nu^{-1})^-(1-v)\,dv$, $\chi_+ : u \in [0,1] \mapsto \int_0^u (F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv$ and $\Gamma = \chi_-^{-1} \circ \chi_+$ where χ_-^{-1} denotes the left continuous generalised inverse of χ_-, that is

$$\Gamma : u \in [0,1] \mapsto \inf\{r \in [0,1] \mid \chi_-(r) \geq \chi_+(u)\},$$

which is well defined since $\chi_+(1) = \chi_+(p) = \chi_-(1-p) = \gamma$, consequence of the equality of the means. Let Q^{NIT} be the probability measure defined on $(0,1)^2$ by

$$Q^{NIT}(du, dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi_+^{NIT}(u, dv) \quad \text{where } \pi_+^{NIT}(u, dv) = \mathbb{1}_{\{\Gamma(u)>0\}}\,\delta_{1-\Gamma(u)}(dv). \tag{2.3.9}$$

Proposition 2.3.4. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Assume that there exists $p \in (0,1)$ such that $u \mapsto \int_0^u (F_\mu^{-1}(v) - F_\nu^{-1}(v))\,dv$ is nondecreasing on $[0,p]$ and nonincreasing on $[p,1]$. Then $Q^{NIT} \in \mathcal{Q}$.*

In the symmetric case, that is when μ and ν are symmetric and $p = 1/2$, we have $\Gamma(u) = u$ and therefore $Q^{NIT} = Q_2$ (see (2.2.17)). Hence Q^{NIT} is a generalisation of the symmetric coupling.

Proof of Proposition 2.3.4. Note that $\Gamma(1) \leq 1-p$, hence $\Gamma(u) < 1$ for all $u \in (0,1)$. It is clear that Q^{NIT} satisfies property (i) of the definition of $\mathcal{Q}$. By Lemma 2.6.1 below applied with the functions $f_1 : u \in (0,1) \mapsto (F_\mu^{-1} - F_\nu^{-1})^+(u)$ and $f_2 : u \in (0,1) \mapsto (F_\mu^{-1} - F_\nu^{-1})^-(1-u)$, we have

$$\begin{aligned}\frac{1}{\gamma}\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^+(u)h(1-\Gamma(u))\mathbb{1}_{\{\Gamma(u)>0\}}\,du &= \frac{1}{\gamma}\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^-(1-v)h(1-v)\,dv \\ &= \frac{1}{\gamma}\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^{-1}(v)h(v)\,dv,\end{aligned}$$

for any measurable and bounded function $h : (0,1) \to \mathbb{R}$. So Q^{NIT} satisfies (ii) as well, and therefore (iii). $\square$

We saw with Proposition 2.2.18 that for all $Q \in \mathcal{Q}$, $\int_0^1 \int_\mathbb{R} |F_\nu^{-1}(u) - y|\,\widetilde{m}^Q(u,dy)\,du = \mathcal{W}_1(\mu,\nu)$. The next proposition shows that the inverse transform martingale coupling and its nonincreasing twin, when it exists, play particular roles among the martingale couplings which derive from $\mathcal{Q}$ when $|F_\nu^{-1}(u) - y|$ is replaced with $|F_\nu^{-1}(u) - y|^\rho$ with $\rho \in \mathbb{R}$.

Proposition 2.3.5. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. For all $\rho \in \mathbb{R}$ and for any Markov kernel $(\widetilde{m}(u,dy))_{u\in(0,1)}$, let $\mathcal{C}_\rho(\widetilde{m})$ be defined by*

$$\mathcal{C}_\rho(\widetilde{m}) = \int_{\mathbb{R}\times(0,1)} |F_\nu^{-1}(u) - y|^\rho \mathbb{1}_{\{y \neq F_\nu^{-1}(u)\}}\,\widetilde{m}(u,dy)\,du. \tag{2.3.10}$$

Then, for all $Q \in \mathcal{Q}$,

$$\begin{aligned}\forall \rho \in (-\infty,1]\cup[2,+\infty), &\quad \mathcal{C}_\rho(\widetilde{m}^{IT}) \leq \mathcal{C}_\rho(\widetilde{m}^Q);\\ \forall \rho \in [1,2], &\quad \mathcal{C}_\rho(\widetilde{m}^Q) \leq \mathcal{C}_\rho(\widetilde{m}^{IT});\\ \forall \rho \in \{1,2\}, &\quad \mathcal{C}_\rho(\widetilde{m}^{IT}) = \mathcal{C}_\rho(\widetilde{m}^Q).\end{aligned} \tag{2.3.11}$$

Let us now assume that there exists $p \in (0,1)$ such that $u \mapsto \int_0^u (F_\mu^{-1}(v) - F_\nu^{-1}(v))\,dv$ is nondecreasing on $[0,p]$ and nonincreasing on $[p,1]$ and denote $(\widetilde{m}^{NIT}(u,dy))_{u\in(0,1)}$ for $(\widetilde{m}^{Q^{NIT}}(u,dy))_{u\in(0,1)}$. Then, for all $Q \in \mathcal{Q}$,

$$\begin{aligned}\forall \rho \in (-\infty,1]\cup[2,+\infty), &\quad \mathcal{C}_\rho(\widetilde{m}^Q) \leq \mathcal{C}_\rho(\widetilde{m}^{NIT});\\ \forall \rho \in [1,2], &\quad \mathcal{C}_\rho(\widetilde{m}^{NIT}) \leq \mathcal{C}_\rho(\widetilde{m}^Q);\\ \forall \rho \in \{1,2\}, &\quad \mathcal{C}_\rho(\widetilde{m}^{NIT}) = \mathcal{C}_\rho(\widetilde{m}^Q).\end{aligned} \tag{2.3.12}$$

Remark 2.3.6. Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. By Proposition 2.3.5 for $\rho = 0$, we deduce that

$$\sup_{Q\in\mathcal{Q}} \{\mathbb{P}(Y = F_\nu^{-1}(U)) \mid (U,Y) \sim \mathbb{1}_{(0,1)}(u)\,du\,\widetilde{m}^Q(u,dy)\}$$

is attained for the inverse transform martingale coupling.

Suppose in addition that F_ν^{-1} is constant on the intervals of the form $(F_\mu(x_-), F_\mu(x)]$. Let $M(dx,dy) = \mu(dx)\,m(x,dy)$ be a martingale coupling between μ and ν. Let $(\widetilde{m}(u,dy))_{u\in(0,1)}$

be the kernel defined for all $u \in (0,1)$ by $\widetilde{m}(u, dy) = m(F_\mu^{-1}(u), dy)$. Let T be the Monge transport map. According to (2.2.20), $F_\nu^{-1}(u) = F_\nu^{-1}(F_\mu(F_\mu^{-1}(u)))$ for du-almost all $u \in (0,1)$. So by Lemma 2.2.6, for all $\rho \in \mathbb{R}$,

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|^\rho \mathbb{1}_{\{y \neq T(x)\}}\, \mu(dx)\, m(x, dy) &= \int_0^1 \int_{\mathbb{R}} |y - T(F_\mu^{-1}(u))|^\rho \mathbb{1}_{\{y \neq T(F_\mu^{-1}(u))\}}\, \widetilde{m}(u, dy)\, du \\ &= \int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|^\rho \mathbb{1}_{\{y \neq F_\nu^{-1}(u)\}}\, \widetilde{m}(u, dy)\, du.\end{aligned}$$

We deduce that the supremum of $\mathbb{P}(Y = T(X))$ among all random variables X and Y such that $(X, Y) \sim M^Q$ for $Q \in \mathcal{Q}$ is attained for the inverse transform martingale coupling.

Proof of Proposition 2.3.5. Let $\rho \in \mathbb{R}$ and $Q \in \mathcal{Q}$. Let $\varepsilon > 0$ and $f_\varepsilon : \mathbb{R} \to \mathbb{R}$ be defined for all $x \in \mathbb{R}$ by

$$f_\varepsilon(x) = \varepsilon^{\rho-2}\left((\rho-1)x + (2-\rho)\varepsilon\right)\mathbb{1}_{\{x \leq \varepsilon\}} + x^{\rho-1}\mathbb{1}_{\{x > \varepsilon\}}.$$

It is clear that f_ε is convex for $\rho \in (-\infty, 1] \cup [2, +\infty)$ and concave for $\rho \in [1, 2]$. Let $c_\varepsilon : (0,1)^2 \to \mathbb{R}$ be the right-continuous function defined for all $(u, v) \in (0,1)^2$ by $c_\varepsilon(u, v) = f_\varepsilon(|F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|)$.

If $\rho \in (-\infty, 1] \cup [2, +\infty)$, then c_ε satisfies the Monge condition, that is for all $u, u', v, v' \in (0,1)$ such that $u \leq u'$ and $v \leq v'$,

$$c_\varepsilon(u', v') - c_\varepsilon(u, v') - c_\varepsilon(u', v) + c_\varepsilon(u, v) \leq 0,$$

which follows from the monotonicity of F_ν^{-1} and the fact that $(x, y) \mapsto f_\varepsilon(|x - y|)$ is convex and therefore satisfies the Monge condition. Since Q has marginals $d\Psi_+/\gamma$ and $d\Psi_-/\gamma$, by Theorem 3.1.2 Chapter 3 [161], we have

$$\int_0^1 c_\varepsilon(\Psi_+^{-1}(\gamma u), \Psi_-^{-1}(\gamma u))\, du \leq \int_{(0,1)^2} c_\varepsilon(u, v)\, Q(du, dv) \leq \int_0^1 c_\varepsilon(\Psi_+^{-1}(\gamma u), \Psi_-^{-1}(\gamma(1-u)))\, du.$$

It is easy to check that for all $u, v \in (0,1)$, the map $(0,1) \ni \varepsilon \mapsto c_\varepsilon(u, v)$ is nonincreasing, bounded from below by $2 - \rho$ and converges to $|F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|^{\rho-1}$ when $\varepsilon \to 0$ where by convention, we choose $0^0 = 1$ and for all $\alpha < 0$ and $x = 0$, we choose x^α equal to its limit $+\infty$ as $x \to 0_+$. Therefore, by the monotone convergence theorem for $\varepsilon \to 0$, we have

$$\begin{aligned}\forall \rho \in (-\infty, 1] \cup [2, +\infty), \quad &\int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\, du \\ &\leq \int_{(0,1)^2} |F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|^{\rho-1}\, Q(du, dv) \\ &\leq \int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma(1-u))_+)|^{\rho-1}\, du.\end{aligned} \tag{2.3.13}$$

If $1 \leq \rho \leq 2$, then $(x, y) \mapsto f_\varepsilon(|x - y|)$ is concave so $-c_\varepsilon$ satisfies the Monge condition and a symmetric reasoning shows that

$$\int_0^1 c_\varepsilon(\Psi_+^{-1}(\gamma u), \Psi_-^{-1}(\gamma(1-u)))\, du \leq \int_{(0,1)^2} c_\varepsilon(u, v)\, Q(du, dv) \leq \int_0^1 c_\varepsilon(\Psi_+^{-1}(\gamma u), \Psi_-^{-1}(\gamma u))\, du. \tag{2.3.14}$$

It is easy to check that for all $u, v \in (0,1)$, the map $(0,1) \ni \varepsilon \mapsto c_\varepsilon(u,v)$ is bounded from above by $1+|F_\nu^{-1}(u_+)-F_\nu^{-1}(v_+)|^{\rho-1}$ and converges to its lower bound $|F_\nu^{-1}(u_+)-F_\nu^{-1}(v_+)|^{\rho-1}$ when $\varepsilon \to 0$. Consider one of the three integrals in (2.3.14). If the pointwise limit for $\varepsilon \to 0$ of its integrand is integrable, then we can apply the dominated convergence theorem. Otherwise, the integral is infinite for all $\varepsilon \in (0,1)$. Therefore, for $\varepsilon \to 0$, we have

$$\begin{aligned}\forall 1 \le \rho \le 2, \quad & \int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma(1-u))_+)|^{\rho-1}\,du \\ &\le \int_{(0,1)^2} |F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|^{\rho-1}\,Q(du,dv) \\ &\le \int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\,du. \quad (2.3.15)\end{aligned}$$

For all $\rho \in \mathbb{R}$, applying (2.2.9) to the measurable and nonnegative function $h : y \mapsto |F_\nu^{-1}(u) - y|^\rho \mathbb{1}_{\{y \neq F_\nu^{-1}(u)\}}$ yields du-almost everywhere on $(0,1)$,

$$\begin{aligned}&\int_{\mathbb{R}} |F_\nu^{-1}(u) - y|^\rho \mathbb{1}_{\{y\neq F_\nu^{-1}(u)\}}\,\widetilde{m}^Q(u,dy) \\ &= \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^+(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} |F_\nu^{-1}(u) - F_\nu^{-1}(v)|^\rho \mathbb{1}_{\{F_\nu^{-1}(v)\neq F_\nu^{-1}(u)\}}\,\pi_+^Q(u,dv) \\ &\quad + \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^-(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(v)} |F_\nu^{-1}(u) - F_\nu^{-1}(v)|^\rho \mathbb{1}_{\{F_\nu^{-1}(v)\neq F_\nu^{-1}(u)\}}\,\pi_-^Q(u,dv),\end{aligned}$$

where according to Lemma 2.2.5, for $(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$-almost all $u \in (0,1)$, $\pi_+^Q(u,dv)$-a.e., $F_\nu^{-1}(v) > F_\nu^{-1}(u)$ and for $(F_\mu^{-1} - F_\nu^{-1})^-(u)\,du$-almost all $u \in (0,1)$, $\pi_-^Q(u,dv)$-a.e, $F_\nu^{-1}(v) < F_\nu^{-1}(u)$. We deduce that

$$\begin{aligned}C_\rho(\widetilde{m}^Q) &= \int_{(0,1)^2} (F_\mu^{-1} - F_\nu^{-1})^+(u)|F_\nu^{-1}(u) - F_\nu^{-1}(v)|^{\rho-1}\,du\,\pi_+^Q(u,dv) \\ &\quad + \int_{(0,1)^2} (F_\mu^{-1} - F_\nu^{-1})^-(u)|F_\nu^{-1}(u) - F_\nu^{-1}(v)|^{\rho-1}\,du\,\pi_-^Q(u,dv) \qquad (2.3.16) \\ &= 2\gamma \int_{(0,1)^2} |F_\nu^{-1}(u) - F_\nu^{-1}(v)|^{\rho-1}\,Q(du,dv).\end{aligned}$$

Since the set of discontinuities of F_ν^{-1} is at most countable and since the marginals of Q have densities, we have

$$C_\rho(\widetilde{m}^Q) = 2\gamma \int_{(0,1)^2} |F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|^{\rho-1}\,Q(du,dv). \qquad (2.3.17)$$

Let us show that

$$\mathcal{C}_\rho(\widetilde{m}^{IT}) = 2\gamma \int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\,du. \qquad (2.3.18)$$

By Lemma 2.6.3 below, $\Psi_+^{-1}(\Psi_+(u)) = u$, $d\Psi_+(u)$-almost everywhere on $(0,1)$, so using (2.3.16), Proposition 2.6.2 below and the fact that $0 < \Psi_\pm^{-1}(u) < 1$ for all $u \in (0,\gamma)$, we have

$$\begin{aligned}
\mathcal{C}_\rho(\widetilde{m}^{IT}) &= 2\int_0^1 (F_\mu^{-1} - F_\nu^{-1})^+(u)|F_\nu^{-1}(u) - F_\nu^{-1}(\varphi(u))|^{\rho-1}\mathbb{1}_{\{0<\varphi(u)<1\}}\,du \\
&= 2\int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\Psi_+(u))) - F_\nu^{-1}(\Psi_-^{-1}(\Psi_+(u)))|^{\rho-1}\mathbb{1}_{\{0<\Psi_-^{-1}(\Psi_+(u))<1\}}\,d\Psi_+(u) \\
&= 2\int_0^\gamma |F_\nu^{-1}(\Psi_+^{-1}(u)) - F_\nu^{-1}(\Psi_-^{-1}(u))|^{\rho-1}\mathbb{1}_{\{0<\Psi_-^{-1}(u)<1\}}\,du \\
&= 2\gamma\int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)) - F_\nu^{-1}(\Psi_-^{-1}(\gamma u))|^{\rho-1}\,du.
\end{aligned}$$

Since the set of discontinuities of Ψ_+^{-1}, Ψ_-^{-1}, $(\Psi_+\circ F_\nu)^{-1} = F_\nu^{-1}\circ\Psi_+^{-1}$ and $(\Psi_-\circ F_\nu)^{-1} = F_\nu^{-1}\circ\Psi_-^{-1}$ are at most countable, we get that for du-almost all $u\in(0,1)$, $F_\nu^{-1}(\Psi_+^{-1}(\gamma u)) = F_\nu^{-1}\circ\Psi_+^{-1}(\gamma u_+) = F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+)$ and $F_\nu^{-1}(\Psi_-^{-1}(\gamma u)) = F_\nu^{-1}(\Psi^{-1}(\gamma u)_+)$, which proves (2.3.18). Then (2.3.11) is deduced from (2.3.13), (2.3.15), (2.3.17) and (2.3.18).

Assume now that there exists $p\in(0,1)$ such that $u\mapsto \int_0^u (F_\mu^{-1}(v) - F_\nu^{-1}(v))\,dv$ is nondecreasing on $[0,p]$ and nonincreasing on $[p,1]$. For all $u\in(0,1)$, $\chi_+(u) = \Psi_+(u)$ and $\chi_-(u) = \gamma - \Psi_-(1-u)$. If U is a random variable uniformly distributed on $(0,1)$, one can easily check that $1-\Psi_-^{-1}(\gamma(1-U))$ has distribution $d\chi_-/\gamma$. Since $u\mapsto 1-\Psi_-^{-1}(\gamma(1-u))$ is nondecreasing, it is shown in Lemma A.3 [4] that $1-\Psi_-^{-1}(\gamma(1-u)) = \chi_-^{-1}(\gamma u)$, du-almost everywhere on $(0,1)$. So we show with similar arguments as above that

$$\mathcal{C}_\rho(\widetilde{m}^{NIT}) = 2\gamma\int_0^1 |F_\nu^{-1}(\Psi_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma(1-u))_+)|^{\rho-1}\,du. \tag{2.3.19}$$

Then (2.3.12) is deduced from (2.3.13), (2.3.15), (2.3.17) and (2.3.19). □

2.4 On the uniqueness of martingale couplings parametrised by $\mathcal{Q}$

Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. A direct consequence of Proposition 2.2.9 is that the set of martingale couplings between μ and ν parametrised by $\mathcal{Q}$ is either a singleton, or uncountably infinite. Since $\mathcal{Q}$ is convex, we deduce from Proposition 2.4.2 below that $\mathcal{Q}$ is infinite as soon as $\mu <_{cx} \nu$. When μ and ν are such that F_μ and F_ν are continuous, Corollary 2.4.5 below ensures that there exist uncountably many martingale couplings between μ and ν parametrised by $\mathcal{Q}$. However this does not necessarily hold in the general case. We saw that when ν is reduced to two atoms only, there exists a unique martingale coupling between μ and ν. Suppose now that the comonotonous coupling is a martingale coupling between μ and ν, and $\mu,\nu\in\mathcal{P}_2(\mathbb{R})$. For any martingale coupling $M\in\Pi^{\mathrm{M}}(\mu,\nu)$, we have $\int_{\mathbb{R}\times\mathbb{R}} |x-y|^2\,M(dx,dy) = \int_{\mathbb{R}} y^2\,\nu(dy) - \int_{\mathbb{R}} x^2\,\mu(dx)$. So all the martingale couplings between μ and ν yield the same quadratic cost. In particular, they yield the same quadratic cost as the comonotonous coupling, which is the only minimiser of the quadratic cost among $\Pi(\mu,\nu)$.

So the comonotonous coupling is the only martingale coupling between μ and ν. The next proposition states that this conclusion still holds when μ and ν only have finite first order moments.

Proposition 2.4.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. If the comonotonous coupling between μ and ν is a martingale coupling, that is for U a random variable uniformly distributed on $(0,1)$,*

$$\mathbb{E}[F_\nu^{-1}(U)|F_\mu^{-1}(U)] = F_\mu^{-1}(U) \quad \textit{almost surely},$$

then it is the only martingale coupling between μ and ν.

Proof. Let U be a random variable uniformly distributed on $(0,1)$. The couple $(U, F_\nu^{-1}(U))$ is distributed according to $\mathbb{1}_{(0,1)}(u)\,du\,\delta_{F_\nu^{-1}(u)}(dy)$. By Lemma 2.2.6 applied with the Markov kernel $(\widetilde{m}(u,dy))_{u\in(0,1)} = (\delta_{F_\nu^{-1}(u)}(dy))_{u\in(0,1)}$, we get that $(F_\mu^{-1}(U), F_\nu^{-1}(U))$ is distributed according to $\mu(dx)\,m(x,dy)$ where $(m(x,dy))_{x\in\mathbb{R}}$ is given by (2.2.6). By Lemma 2.6.4 below combined with the inverse transform sampling and (2.2.7), we get that $(F_\mu^{-1}(U), F_\nu^{-1}(U))$ is distributed according to $\mu(dx)\int_{v=0}^1 \delta_{F_\nu^{-1}(F_\mu(x_-)+v(F_\mu(x)-F_\mu(x_-)))}(dy)\,dv$. So almost surely,

$$\begin{aligned}
F_\mu^{-1}(U) &= \mathbb{E}\left[F_\nu^{-1}(U)|F_\mu^{-1}(U)\right] \\
&= \int_{v=0}^1 \left(\int_{y\in\mathbb{R}} y\,\delta_{F_\nu^{-1}(F_\mu(F_\mu^{-1}(U)_-)+v(F_\mu(F_\mu^{-1}(U))-F_\mu(F_\mu^{-1}(U)_-)))}(dy)\right) dv \\
&= \int_0^1 F_\nu^{-1}(F_\mu(F_\mu^{-1}(U)_-) + v(F_\mu(F_\mu^{-1}(U)) - F_\mu(F_\mu^{-1}(U)_-)))\,dv.
\end{aligned}$$

By the inverse transform sampling, we deduce that for $\mu(dx)$-almost all $x\in\mathbb{R}$,

$$\int_0^1 F_\nu^{-1}(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)))\,dv = x. \tag{2.4.1}$$

Let $(\underline{t}_n, \overline{t}_n)_{1\le n\le N}$ denote the irreducible components of (μ,ν), whose definition is given by (2.2.13). We recall that the choice of any martingale coupling M between μ and ν reduces to the choice of a sequence of martingale couplings $(M_n)_{1\le n\le N}$ such that for all $1\le n\le N$, M_n is a martingale coupling between the probability measures μ_n and ν_n defined by (2.2.15). If for each n, μ_n reduces to a single atom, then we necessarily have $M_n(dx,dy) = \mu_n(dx)\,\nu_n(dy)$, so there is a unique choice of the sequence $(M_n)_{1\le n\le N}$ and therefore M, which is the comonotonous coupling. Let us then prove that μ_n reduces to a single atom.

Let I be the at most countable set of $x\in\mathbb{R}$ such that $\mu(\{x\})>0$ and F_ν^{-1} is nonconstant on $(F_\mu(x_-), F_\mu(x)]$. Let us show that

$$\bigcup_{x\in I}(F_\mu(x_-), F_\mu(x)) = \bigcup_{n=1}^N (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-))\,. \tag{2.4.2}$$

By Lemma A.8 [4], we have

$$\bigcup_{n=1}^N (F_\mu(\underline{t}_n), F_\mu((\overline{t}_n)_-)) = \left\{u\in(0,1) \mid \int_0^u F_\mu^{-1}(v)\,dv > \int_0^u F_\nu^{-1}(v)\,dv\right\}. \tag{2.4.3}$$

Let $u \in (0,1)$. Suppose first that there exists $t \in \mathbb{R}$ such that $u = F_\mu(t)$. We recall that $(F_\mu^{-1}(U), F_\nu^{-1}(U))$ is distributed according to $\mu(dx) \int_{v=0}^1 \delta_{F_\nu^{-1}(F_\mu(x_-)+v(F_\mu(x)-F_\mu(x_-)))}(dy)\,dv$. So

$$\begin{aligned}
\int_0^{F_\mu(t)} F_\nu^{-1}(v)\,dv &= \int_0^1 \mathbb{1}_{\{v \leq F_\mu(t)\}} F_\nu^{-1}(v)\,dv = \int_0^1 \mathbb{1}_{\{F_\mu^{-1}(v) \leq t\}} F_\nu^{-1}(v)\,dv \\
&= \int_{x\in\mathbb{R}} \left(\int_{v=0}^1 \mathbb{1}_{\{x \leq t\}} \left(\int_{y\in\mathbb{R}} y\,\delta_{F_\nu^{-1}(F_\mu(x_-)+v(F_\mu(x)-F_\mu(x_-)))}(dy) \right) dv \right) \mu(dx) \\
&= \int_{x\in\mathbb{R}} \mathbb{1}_{\{x \leq t\}} \left(\int_{v=0}^1 F_\nu^{-1}(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)))\,dv \right) \mu(dx) \\
&= \int_{x\in\mathbb{R}} \mathbb{1}_{\{x \leq t\}} x\,\mu(dx) = \int_0^1 \mathbb{1}_{\{F_\mu^{-1}(v) \leq t\}} F_\mu^{-1}(v)\,dv = \int_0^{F_\mu(t)} F_\mu^{-1}(v)\,dv,
\end{aligned}$$

where we used (2.4.1) for the fifth equality and the inverse transform sampling for the sixth equality. By continuity, we also deduce that for all $t \in \mathbb{R}$, $\int_0^{F_\mu(t_-)} F_\nu^{-1}(v)\,dv = \int_0^{F_\mu(t_-)} F_\mu^{-1}(v)\,dv$.

Suppose now that there exists $x \in \mathbb{R}$ in the set of discontinuities of F_μ such that $F_\mu(x_-) < u < F_\mu(x)$. According to (2.4.1), we have $\int_{F_\mu(x_-)}^{F_\mu(x)} F_\nu^{-1}(v)\,dv = \mu(\{x\})x = \int_{F_\mu(x_-)}^{F_\mu(x)} x\,dv = \int_{F_\mu(x_-)}^{F_\mu(x)} F_\mu^{-1}(v)\,dv$.

If F_ν^{-1} is constant on $(F_\mu(x_-), F_\mu(x)]$, then for all $v \in (F_\mu(x_-), F_\mu(x)]$, $F_\nu^{-1}(v) = x = F_\mu^{-1}(v)$, so

$$\begin{aligned}
\int_0^u F_\nu^{-1}(v)\,dv &= \int_0^{F_\mu(x_-)} F_\nu^{-1}(v)\,dv + \int_{F_\mu(x_-)}^u F_\nu^{-1}(v)\,dv = \int_0^{F_\mu(x_-)} F_\mu^{-1}(v)\,dv + \int_{F_\mu(x_-)}^u F_\mu^{-1}(v)\,dv \\
&= \int_0^u F_\mu^{-1}(v)\,dv.
\end{aligned}$$

If F_ν^{-1} is nonconstant on $(F_\mu(x_-), F_\mu(x)]$, then using the monotonicity of F_ν^{-1}, one can easily show that for all $u \in (F_\mu(x_-), F_\mu(x))$,

$$\frac{1}{u - F_\mu(x_-)} \int_{F_\mu(x_-)}^u F_\nu^{-1}(v)\,dv < \frac{1}{F_\mu(x) - F_\mu(x_-)} \int_{F_\mu(x_-)}^{F_\mu(x)} F_\nu^{-1}(v)\,dv.$$

We deduce that for all $u \in (F_\mu(x_-), F_\mu(x))$,

$$\int_{F_\mu(x_-)}^u F_\nu^{-1}(v)\,dv < \frac{u - F_\mu(x_-)}{F_\mu(x) - F_\mu(x_-)} x\mu(\{x\}) = (u - F_\mu(x_-))x = \int_{F_\mu(x_-)}^u x\,dv = \int_{F_\mu(x_-)}^u F_\mu^{-1}(v)\,dv,$$

and $\int_0^u F_\mu^{-1}(v)\,dv > \int_0^u F_\nu^{-1}(v)\,dv$. With (2.4.3), we deduce (2.4.2). Since the intervals $((\underline{t}_n, \bar{t}_n))_{1\leq n\leq N}$ are disjoint, the intervals $((F_\mu(\underline{t}_n), F_\mu((\bar{t}_n)_-)))_{1\leq n\leq N}$ are disjoint as well. By equality of unions of disjoint intervals, we proved that for all $1 \leq n \leq N$, there exists $x \in I$ such that $(F_\mu(\underline{t}_n), F_\mu((\bar{t}_n)_-)) = (F_\mu(x_-), F_\mu(x))$. So $x \in (\underline{t}_n, \bar{t}_n)$ and $\mu((\underline{t}_n, \bar{t}_n)) = F_\mu((\bar{t}_n)_-) - F_\mu(\underline{t}_n) = F_\mu(x) - F_\mu(x_-) = \mu(\{x\})$. So $\mu_n = \delta_x$, and the discussion above concludes that there exists only one martingale coupling between μ and ν, namely the comonotonous coupling. □

We saw in Section 2.3.1 that we can build a nonincreasing twin of the inverse transform martingale coupling (see (2.3.9)) as soon as the two marginals satisfy the assumption in Proposition 2.3.4. This corresponds to a general inversion of the monoticity of φ on $(0,1)$. In the general case, such an inversion is not possible on $(0,1)$, but can be made locally.

Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Since $\mu \neq \nu$, there exists $u \in (0,1)$ such that $\Psi_+(u) > \Psi_-(u)$. Let $v = \Psi_+^{-1}(\Psi_+(u))$. Then $\Psi_+(v) = \Psi_+(\Psi_+^{-1}(\Psi_+(u)) = \Psi_+(u)$ so that $v > 0$ and $\Psi_+(v) > \Psi_-(u) \geq \Psi_-(v)$. By left-continuity of Ψ_- and Ψ_+, there exists $\eta \in (0,v)$ such that $\Psi_+(w) > \Psi_-(w)$ for all $w \in [v-\eta, v]$. By definition of v, we have $\Psi_+(v-\eta) < \Psi_+(v)$, so there exists $u_0 \in (v-\eta, v)$ such that $(F_\mu^{-1} - F_\nu^{-1})^+(u_0) > 0$. Since $u_0 \in (v-\eta,v)$, we have $\Psi_+(v) > \Psi_+(u_0) > \Psi_-(u_0)$ so $1 > \varphi(u_0) > u_0$ according to (2.3.7). By left-continuity of F_μ^{-1}, F_ν^{-1} and φ, there exists $\varepsilon \in (0,u_0)$ such that

$$\forall u \in [u_0-\varepsilon, u_0], \quad 1 > \varphi(u) > u_0 \quad \text{and} \quad F_\mu^{-1}(u) > F_\nu^{-1}(u). \tag{2.4.4}$$

Since $(u_0-\varepsilon, u_0] \subset \mathcal{U}_+$, Ψ_+ is increasing and is therefore one-to-one onto from $(u_0-\varepsilon,u_0]$ to $(\Psi_+(u_0-\varepsilon), \Psi_+(u_0)]$. Since the set of discontinuities of Ψ_-^{-1} is at most countable, up to choosing ε smaller, we may also suppose that in addition to (2.4.4), ε is such that Ψ_-^{-1} is continuous at $\Psi_+(u_0-\varepsilon)$. Let then $\zeta : [0,1] \to [0,1]$ and $\widetilde{\zeta} : [0,1] \to [0,1]$ be defined for all $u \in (0,1)$ by

$$\zeta(u) = \Psi_-^{-1}(G(\Psi_+(u))) \quad \text{and} \quad \widetilde{\zeta}(u) = \Psi_+^{-1}(G(\Psi_-(u))), \tag{2.4.5}$$

where $G : u \mapsto u\mathbb{1}_{(u_0-\varepsilon,u_0]^\complement}(\Psi_+^{-1}(u)) + (\Psi_+(u_0) - u + \Psi_+(u_0-\varepsilon))\,\mathbb{1}_{(u_0-\varepsilon,u_0]}(\Psi_+^{-1}(u))$. Let Q^ζ be the measure defined on $(0,1)^2$ by

$$Q^\zeta(du,dv) = \frac{1}{\gamma}\left(F_\mu^{-1} - F_\nu^{-1}\right)^+(u)\,du\,\pi_+^\zeta(u,dv) \quad \text{where } \pi_+^\zeta(u,dv) = \mathbb{1}_{\{0<\zeta(u)<1\}}\,\delta_{\zeta(u)}(dv), \tag{2.4.6}$$

with $\gamma = \Psi_-(1) = \Psi_+(1)$.

Proposition 2.4.2. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. The measure Q^ζ defined by (2.4.6) is an element of $\mathcal{Q}$. Moreover,*

$$Q^\zeta(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv\,\pi_-^\zeta(v,du) \quad \textit{where } \pi_-^\zeta(v,du) = \mathbb{1}_{\{0<\widetilde{\zeta}(v)<1\}}\,\delta_{\widetilde{\zeta}(v)}(du).$$

As said above, Ψ_+ is one-to-one onto from $(u_0-\varepsilon,u_0]$ to $(\Psi_+(u_0-\varepsilon),\Psi_+(u_0)]$. So, for all $u \in (u_0-\varepsilon,u_0]$, $\Psi_+^{-1}(\Psi_+(u)) = u$ and $G(\Psi_+(u)) = \Psi_+(u_0) - \Psi_+(u) + \Psi_+(u_0-\varepsilon)$. So

$$\forall u \in (u_0-\varepsilon,u_0], \quad \zeta(u) = \Psi_-^{-1}(\Psi_+(u_0) - \Psi_+(u) + \Psi_+(u_0-\varepsilon)). \tag{2.4.7}$$

Since Ψ_- is continuous, Ψ_-^{-1} is one-to-one. Moreover, Ψ_+ is increasing on $(u_0-\varepsilon,u_0]$, so for all $u \in (u_0-\varepsilon,u_0]\backslash\{\Psi_+^{-1}(\frac{\Psi_+(u_0)+\Psi_+(u_0-\varepsilon)}{2})\}$, $\zeta(u) \neq \varphi(u)$. Since $(u_0-\varepsilon,u_0] \subset \mathcal{U}_+$, considering the first marginal of Q^ζ and Q^{IT}, we deduce that $Q^\zeta \neq Q^{IT}$. As a direct consequence of the convexity of $\mathcal{Q}$, we deduce that $\mathcal{Q}$ is uncountably infinite.

Corollary 2.4.3. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Then $\mathcal{Q}$ is uncountably infinite.*

Proof of Proposition 2.4.2. Let $h : (0,1)^2 \to \mathbb{R}$ be a measurable and bounded function. We have

$$\begin{aligned}\int_{(0,1)^2} h(u,v)\, Q^\zeta(du,dv) &= \frac{1}{\gamma}\int_{(0,1)^2} h(u,v)(F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\{0<\zeta(u)<1\}}\,\delta_{\zeta(u)}(dv)\,du \\ &= \frac{1}{\gamma}\int_0^1 h(u,\zeta(u))\mathbb{1}_{\{0<\zeta(u)<1\}}\,d\Psi_+(u) \\ &= \frac{1}{\gamma}\int_0^1 h(\Psi_+^{-1}(\Psi_+(u)),\zeta(u))\mathbb{1}_{\{0<\zeta(u)<1\}}\mathbb{1}_{\{0<u<1\}}\,d\Psi_+(u),\end{aligned} \tag{2.4.8}$$

where the last equality is a consequence of Lemma 2.6.3 below. By Proposition 2.6.2 below,

$$\int_{(0,1)^2} h(u,v)\, Q^\zeta(du,dv) = \frac{1}{\gamma}\int_0^{\Psi_+(1)} h(\Psi_+^{-1}(u),\Psi_-^{-1}(G(u)))\mathbb{1}_{\{0<\Psi_-^{-1}(G(u))<1\}}\mathbb{1}_{\{0<\Psi_+^{-1}(u)<1\}}\,du.$$

By Lemma 2.6.3 below, for all $u \in (0,\Psi_+(1))$, $u_0-\varepsilon < \Psi_+^{-1}(u) \le u_0 \iff \Psi_+(u_0-\varepsilon) < u \le \Psi_+(u_0)$. Hence G is a piecewise affine function which satisfies $G(G(u)) = u$ for all $u \in (0,\Psi_+(1))\backslash\{\Psi_+(u_0)\}$ and $G(G(\Psi_+(u_0))) = \Psi_+(u_0-\varepsilon)$. So by the change of variables $w = G(u)$, we have

$$\int_{(0,1)^2} h(u,v)\, Q^\zeta(du,dv) = \frac{1}{\gamma}\int_0^{\Psi_+(1)} h(\Psi_+^{-1}(G(w)),\Psi_-^{-1}(w))\mathbb{1}_{\{0<\Psi_-^{-1}(w)<1\}}\mathbb{1}_{\{0<\Psi_+^{-1}(G(w))<1\}}\,dw. \tag{2.4.9}$$

By continuity of Ψ_- and Proposition 2.6.2 below, using that $\Psi_+(1) = \Psi_-(1)$, we have

$$\begin{aligned}&\int_{(0,1)^2} h(u,v)\, Q^\zeta(du,dv) \\ &= \frac{1}{\gamma}\int_0^1 h(\Psi_+^{-1}(G(\Psi_-(u))),\Psi_-^{-1}(\Psi_-(u)))\mathbb{1}_{\{0<\Psi_-^{-1}(\Psi_-(u))<1\}}\mathbb{1}_{\{0<\Psi_+^{-1}(G(\Psi_-(u)))<1\}}\,d\Psi_-(u) \\ &= \frac{1}{\gamma}\int_0^1 h(\tilde\zeta(u),u)\mathbb{1}_{\{0<\tilde\zeta(u)<1\}}\,d\Psi_-(u),\end{aligned}$$

where we used for the last equality that $\Psi_-^{-1}(\Psi_-(u)) = u$, $d\Psi_-(u)$-almost everywhere on $(0,1)$ according to Lemma 2.6.3 below.

Hence $Q(du,dv) = \frac{1}{\gamma}(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv\,\pi_-^\zeta(v,du)$ where $\pi_-^\zeta(v,du) = \mathbb{1}_{\{0<\tilde\zeta(v)<1\}}\,\delta_{\tilde\zeta(v)}(du)$.

For $h : (u,v) \mapsto 1$, (2.4.9) writes

$$Q^\zeta((0,1)^2) = \frac{1}{\gamma}\int_0^{\Psi_+(1)} \mathbb{1}_{\{0<\Psi_-^{-1}(w)<1\}}\mathbb{1}_{\{0<\Psi_+^{-1}(G(w))<1\}}\,dw.$$

By continuity of Ψ_-, Proposition 2.6.2 and Lemma 2.6.3 below,

$$\int_0^{\Psi_+(1)} \mathbb{1}_{\{0<\Psi_-^{-1}(w)<1\}}\,dw = \int_0^1 \mathbb{1}_{\{0<\Psi_-^{-1}(\Psi_-(w))<1\}}\,d\Psi_-(w) = \int_0^1 d\,\Psi_-(u) = \Psi_-(1) = \Psi_+(1).$$

So $0 < \Psi_-^{-1}(w) < 1$, dw-almost everywhere on $(0,\Psi_+(1))$. By a similar reasoning, $0 < \Psi_+^{-1}(w) < 1$ for dw-almost all $w \in (0,\Psi_+(1))$. Since G is piecewise affine and bijective from $(0, \Psi_+(1))\backslash\{\Psi_+(u_0)\}$ to itself, $0 < \Psi_+^{-1}(G(w)) < 1$ for dw-almost all $w \in (0, \Psi_+(1))$. Hence

$$Q^\zeta((0,1)^2) = \frac{1}{\gamma}\int_0^{\Psi_+(1)} dw = 1,$$

so Q^ζ is a probability measure, with first marginal $\frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$ and second marginal $\frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv$.

We have

$$Q^\zeta\left(\{(u,v) \in (0,1)^2 \mid u < v\}\right) = \frac{1}{\gamma}\int_0^1 \mathbb{1}_{\{u<\zeta(u)\}}\mathbb{1}_{\{0<\zeta(u)<1\}}\,d\Psi_+(u).$$

According to (2.3.8), $u < \varphi(u)$, $d\Psi_+(u)$-almost everywhere on $(0,1)$. According to (2.4.7) and (2.4.4), for all $u \in (u_0 - \varepsilon, u_0]$, $\zeta(u) \geq \zeta(u_0)$ and

$$\zeta(u_0) = \varphi(u_0 - \varepsilon) > u_0. \tag{2.4.10}$$

So for all $u \in (u_0 - \varepsilon, u_0]$, $\zeta(u) > u_0 \geq u$. Moreover, by Lemma 2.6.3 below, $\Psi_+^{-1}(\Psi_+(u)) = u$, $d\Psi_+(u)$-almost everywhere on $(0,1)$. So ζ coincides with φ, $d\Psi_+$-almost everywhere on $(u_0 - \varepsilon, u_0]^\complement$, hence $u < \zeta(u)$, $d\Psi_+(u)$-almost everywhere on $(0,1)$. So using (2.4.8) for $h = 1$, we get that

$$Q^\zeta\left(\{(u,v) \in (0,1)^2 \mid u < v\}\right) = \frac{1}{\gamma}\int_0^1 \mathbb{1}_{\{0<\zeta(u)<1\}}\,d\Psi_+(u) = Q^\zeta((0,1)^2),$$

which is equal to 1 since Q^ζ is a probability measure on $(0,1)^2$. □

Corollary 2.4.4. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$. Let $u_0 \in (0,1)$ and $\varepsilon \in (0, u_0)$ be such that (2.4.4) is satisfied and Ψ_-^{-1} is continuous at $\Psi_+(u_0 - \varepsilon)$. If F_ν^{-1} is nonconstant on $(\varphi(u_0 - \varepsilon), \varphi(u_0)]$ and if F_μ^{-1} is such that for all $\varepsilon' \in (0, \varepsilon)$, the set $\{u \in (0,1) \mid F_\mu^{-1}(u_0 - \varepsilon') < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)\}$ has positive Lebesgue measure, then there exist uncountably many martingale couplings parametrised by $\mathcal{Q}$ between μ and ν.*

Notice that by left-continuity, the condition on F_μ^{-1} in the statement of Corollary 2.4.4 is satisfied if for all $\varepsilon' \in (0, \varepsilon)$, F_μ^{-1} takes at least three different values on $[u_0 - \varepsilon', u_0]$. A direct consequence of Corollary 2.4.4 is the infinite amount of martingale couplings between μ and ν when F_μ^{-1} and F_ν^{-1} are increasing, or equivalently when F_μ and F_ν are continuous.

Corollary 2.4.5. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu <_{cx} \nu$ and $\mu(\{x\}) = \nu(\{x\}) = 0$ for all $x \in \mathbb{R}$. Then there exist uncountably many martingale couplings parametrised by $\mathcal{Q}$ between μ and ν.*

Remark 2.4.6. Corollary 2.4.5 is also a consequence of Corollary 2.4.3 together with Proposition 2.2.7.

Proof of Corollary 2.4.4. Let ζ be defined by (2.4.5) and let Q^ζ be the probability measure defined by (2.4.6). By Proposition 2.3.1, Proposition 2.4.2 and Lemma 2.2.5, for du-almost all $u \in (u_0 - \varepsilon, u_0)$, $F_\nu^{-1}(\zeta(u)) > F_\mu^{-1}(u) > F_\nu^{-1}(u)$ and $F_\nu^{-1}(\varphi(u)) > F_\mu^{-1}(u) > F_\nu^{-1}(u)$.

Since F_ν^{-1} is left-continuous and nonconstant on $(\varphi(u_0 - \varepsilon), \varphi(u_0)]$, F_ν^{-1} is nonconstant on $(\varphi(u_0 - \varepsilon), \varphi(u_0))$. So there exist $a, b \in (\varphi(u_0 - \varepsilon), \varphi(u_0))$ such that $F_\nu^{-1}(a) < F_\nu^{-1}(b)$. Let then $c = \inf\{u \in (a, b) \mid F_\nu^{-1}(u) = F_\nu^{-1}(b)\}$. Let $u \in [a, b]$. If $F_\nu^{-1}(u) = F_\nu^{-1}(b)$, then $c \le u$. Else if $F_\nu^{-1}(u) < F_\nu^{-1}(b)$, then $c \ge u$. We deduce that $a \le c \le b$ and for all $u, v \in (\varphi(u_0 - \varepsilon), \varphi(u_0))$ such that $u < c < v$, we have $F_\nu^{-1}(u) < F_\nu^{-1}(b) \le F_\nu^{-1}(v)$.

Using (2.4.10), we have $F_\nu^{-1}(\varphi(u_0)) > F_\nu^{-1}(\varphi(u_0 - \varepsilon)) = F_\nu^{-1}(\zeta(u_0))$. The map φ is left-continuous, and since ε is such that Ψ_-^{-1} is continuous at $\Psi_+(u_0 - \varepsilon)$, ζ is left-continuous at u_0. So there exists $\tau \in (0, \varepsilon)$ such that for all $u \in [u_0 - \tau, u_0]$, $\zeta(u) < c < \varphi(u)$. We deduce that for all $u \in [u_0 - \tau, u_0]$, $F_\nu^{-1}(\varphi(u)) > F_\nu^{-1}(\zeta(u))$. So for du-almost all $u \in [u_0 - \tau, u_0]$, we have

$$F_\nu^{-1}(\varphi(u)) > F_\nu^{-1}(\zeta(u)) > F_\mu^{-1}(u) > F_\nu^{-1}(u). \tag{2.4.11}$$

Let $a, b, c, d \in \mathbb{R}$ be such that $a > b > c > d$. Then

$$\begin{aligned}\left(\frac{c-d}{a-d}\right) a^2 + \left(\frac{a-c}{a-d}\right) d^2 &= \frac{ca^2 - da^2 + ad^2 - cd^2}{a-d} = \frac{(a-d)(ac - ad + dc)}{(a-d)} = a(c-d) + dc \\ &> b(c-d) + dc \\ &= \left(\frac{c-d}{b-d}\right) b^2 + \left(\frac{b-c}{b-d}\right) d^2.\end{aligned}$$

Thanks to (2.4.11) and this inequality applied with

$$(a, b, c, d) = (F_\nu^{-1}(\varphi(u)), F_\nu^{-1}(\zeta(u)), F_\mu^{-1}(u), F_\nu^{-1}(u)),$$

we deduce that for du-almost all $u \in [u_0 - \tau, u_0]$,

$$\int_{\mathbb{R}} y^2\, \widetilde{m}^{IT}(u, dy) > \int_{\mathbb{R}} y^2\, \widetilde{m}^{Q^\zeta}(u, dy). \tag{2.4.12}$$

By Lemma 2.2.6, we have

$$\int_{\mathbb{R}^2} \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau) < x < F_\mu^{-1}(u_0)\}} y^2\, M^{IT}(dx, dy) = \int_{u=0}^1 \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau) < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)\}} \int_{y\in\mathbb{R}} y^2\, \widetilde{m}^{IT}(u, dy)\, du.$$

For all $u \in (0, 1)$ such that $F_\mu^{-1}(u_0 - \tau) < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)$, we have $u \in [u_0 - \tau, u_0]$. So by (2.4.11), for $Q \in \{Q^{IT}, Q^\zeta\}$ and for du-almost all $u \in (0, 1)$ such that $F_\mu^{-1}(u_0 - \tau) < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)$, $y^2 \le \max(F_\nu^{-1}(\varphi(u_0))^2, F_\nu^{-1}(u_0 - \varepsilon)^2)$, $\widetilde{m}^Q(u, dy)$-almost everywhere. Therefore, for $Q \in \{Q^{IT}, Q^\zeta\}$, we have

$$\int_{u=0}^1 \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau) < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)\}} \int_{y\in\mathbb{R}} y^2\, \widetilde{m}^Q(u, dy)\, du \le \max(F_\nu^{-1}(\varphi(u_0))^2, F_\nu^{-1}(u_0 - \varepsilon)^2) < \infty.$$

Since by assumption the Lebesgue measure of $\{u \in (0,1) \mid F_\mu^{-1}(u_0-\tau) < F_\mu^{-1}(u) < F_\mu^{-1}(u_0)\}$ is positive, according to (2.4.12), we get that

$$\int_{\mathbb{R}^2} \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau)<x<F_\mu^{-1}(u_0)\}} y^2\, M^{IT}(dx,dy) > \int_{u=0}^1 \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau)<F_\mu^{-1}(u)<F_\mu^{-1}(u_0)\}} \int_{y\in\mathbb{R}} y^2\, \widetilde{m}^{Q^\zeta}(u,dy)\,du$$
$$= \int_{\mathbb{R}^2} \mathbb{1}_{\{F_\mu^{-1}(u_0-\tau)<x<F_\mu^{-1}(u_0)\}} y^2\, M^{Q^\zeta}(dx,dy).$$

So $M^{IT} \neq M^{Q^\zeta}$. By Proposition 2.2.9, we deduce that $(M^{\lambda Q^{IT}+(1-\lambda)Q^\zeta})_{\lambda\in[0,1]}$ is a family of distinct martingale couplings between μ and ν. □

2.5 Corresponding super and submartingale couplings

We recall that two probability measures $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ are in the decreasing (resp. increasing) convex order and denote $\mu \le_{dcx} \nu$ (resp. $\mu \le_{icx} \nu$) if $\int_\mathbb{R} f(x)\,\mu(dx) \le \int_\mathbb{R} f(y)\,\nu(dy)$ for any decreasing (resp. increasing) convex function $f : \mathbb{R} \to \mathbb{R}$. For two probability measures $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \le_{dcx} \nu$ (resp. $\mu \le_{icx} \nu$), let

$$\mathcal{S}_1(\mu,\nu) = \inf \int_{\mathbb{R}\times\mathbb{R}} |x-y|\, M(dx,dy),$$

where the infimum is taken over all supermartingale (resp. submartingale) couplings M between μ and ν. Our main result, namely Theorem 2.2.12, can be generalised for the decreasing and increasing convex orders. We use the definitions of $\mathcal{U}_+$, $\mathcal{U}_-$, $\mathcal{U}_0$ given by (2.2.3) and the definitions of Ψ_+ and Ψ_- given at the beginning of Section 2.3.1.

Theorem 2.5.1. *For all $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \le_{dcx} \nu$,*

$$\mathcal{S}_1(\mu,\nu) \le 2\Psi_-(1) + \mathcal{W}_1(\mu,\nu). \tag{2.5.1}$$

For all $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \le_{icx} \nu$,

$$\mathcal{S}_1(\mu,\nu) \le 2\Psi_+(1) + \mathcal{W}_1(\mu,\nu). \tag{2.5.2}$$

Remark 2.5.2. In the martingale case, that is $\mu \le_{cx} \nu$, we have that $2\Psi_-(1) = 2\Psi_+(1) = \mathcal{W}_1(\mu,\nu)$, consequence of the equality of the means and (2.1.2), so that we find Theorem 2.2.12 again.

The statement (2.5.2) for the increasing convex order can easily be deduced from (2.5.1) for the decreasing convex order. Indeed, let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \le_{icx} \nu$. For any probability measure τ on $\mathbb{R}$ or $\mathbb{R}^2$, let $\overline{\tau}$ denote the image of τ by $x \mapsto -x$, so that $\overline{\mu} \le_{dcx} \overline{\nu}$. By the inverse transform sampling, $-F_\mu^{-1}(1-U)$ is distributing according to $\overline{\mu}$ for U a random variable uniformly distributed on $(0,1)$. Since $u \mapsto -F_\mu^{-1}(1-u)$ is nondecreasing, we have $F_{\overline{\mu}}^{-1}(u) = -F_\mu^{-1}(1-u)$, du-almost everywhere on $(0,1)$ (see for instance Lemma A.3 [4] and Lemma 2.6.5 below for an idea of the proof). The fact that quantile functions are left-continuous and have at most countable sets of discontinuities then yields $F_{\overline{\mu}}^{-1}(u) = -F_{\mu}^{-1}((1-u)_+)$ for all $u \in (0,1)$. Since the map $M \mapsto \overline{M}$ is a one-to-one correspondence between the set of supermartingale couplings between $\overline{\mu}$ and $\overline{\nu}$ and the set of submartingale couplings between μ and ν, we have $\mathcal{S}_1(\overline{\mu},\overline{\nu}) = \mathcal{S}_1(\mu,\nu)$. So if (2.5.1) is true for $\overline{\mu} \leq_{dcx} \overline{\nu}$, then

$$\begin{aligned}\mathcal{S}_1(\mu,\nu) = \mathcal{S}_1(\overline{\mu},\overline{\nu}) &\leq 2\int_0^1 (F_{\overline{\mu}}^{-1} - F_{\overline{\nu}}^{-1})^-(u)\,du + \mathcal{W}_1(\overline{\mu},\overline{\nu})\\ &= 2\int_0^1 (F_{\nu}^{-1} - F_{\mu}^{-1})^-(u)\,du + \mathcal{W}_1(\mu,\nu)\\ &= 2\int_0^1 (F_{\mu}^{-1} - F_{\nu}^{-1})^+(u)\,du + \mathcal{W}_1(\mu,\nu) = 2\Psi_+(1) + \mathcal{W}_1(\mu,\nu),\end{aligned}$$

hence (2.5.2) holds.

Frow now on, we suppose $\mu \leq_{dcx} \nu$. We recall that two probability measures $\eta,\tau \in \mathcal{P}(\mathbb{R})$ are in the stochastic order, denoted $\eta \leq_{st} \tau$, iff for all $u \in (0,1)$, $F_\eta^{-1}(u) \leq F_\tau^{-1}(u)$, and in that case $\tau \leq_{dcx} \eta$. If $\nu \leq_{st} \mu$, then for U a random variable uniformly distributed on $(0,1)$, by the inverse transform sampling, $(F_\mu^{-1}(U), F_\nu^{-1}(U))$ is a supermartingale coupling between μ and ν, that is $\mathbb{E}[F_\nu^{-1}(U) | F_\mu^{-1}(U)] \leq F_\mu^{-1}(U)$ almost surely. In that case,

$$\mathcal{S}_1(\mu,\nu) \leq \mathbb{E}[F_\nu^{-1}(U) - F_\mu^{-1}(U)] = \mathcal{W}_1(\mu,\nu), \tag{2.5.3}$$

so (2.5.1) is satisfied as soon as $\nu \leq_{st} \mu$, which is equivalent to $\Psi_-(1) = 0$. If $\nu \not\leq_{st} \mu$, then Inequality (2.5.1) is a direct consequence of Proposition 2.5.5 and Proposition 2.5.7 below. As mentioned above, this concludes the proof of Theorem 2.5.1.

Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{dcx} \nu$ and $\nu \not\leq_{st} \mu$, so that $\Psi_-(1) > 0$. According to Theorem 4.A.3 Chapter 4 [169], for all $u \in [0,1]$, $\int_0^u F_\mu^{-1}(v)\,dv \geq \int_0^u F_\nu^{-1}(v)\,dv$. This implies that for all $u \in [0,1]$, $\Psi_+(u) \geq \Psi_-(u)$. Let $\widetilde{\mathcal{U}}_+$ be a measurable subset of $\mathcal{U}_+$ which satisfies

$$\forall u \in (0,1), \quad \int_0^u \mathbb{1}_{\widetilde{\mathcal{U}}_+(v)}(F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv \geq \int_0^u (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv, \quad \text{with equality for } u = 1. \tag{2.5.4}$$

Let $u_d = \Psi_+^{-1}(\Psi_-(1))$. Since Ψ_+ is continuous, $\Psi_+(u_d) = \Psi_-(1)$. If $u_d = 0$, then $\Psi_-(1) = 0$, which implies that for all $u \in (0,1)$, $F_\mu^{-1}(u) \geq F_\nu^{-1}(u)$. We deduce that the condition $\nu \not\leq_{st} \mu$ is equivalent to $u_d > 0$. One readily sees that (2.5.4) is satisfied for $\widetilde{\mathcal{U}}_+ = (0, u_d)$. Let

$$\overline{u} = \sup\{u \in [0,1] \mid \Psi_+(u) = \Psi_-(u)\}. \tag{2.5.5}$$

We deduce from the definition of $\overline{u}$ and (2.5.4) that $\Psi_+(\overline{u}) = \int_0^{\overline{u}} \mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$, so for du-almost all $u \in \mathcal{U}_+ \cap [0,\overline{u}]$, $u \in \widetilde{\mathcal{U}}_+$. Therefore, the only room for manoeuvre of $\widetilde{\mathcal{U}}_+$ is $[\overline{u}, 1]$.

Let $\gamma = \int_0^1 (F_\mu^{-1} - F_\nu^{-1})^-(u)\,du \in (0,+\infty)$. We note $\mathcal{Q}$ the set of probability measures Q on $(0,1)^2$ such that there exists a measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ which satisfies (2.5.4) and

(i) Q has first marginal $\frac{1}{\gamma}\mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$;

(ii) Q has second marginal $\frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv$;

(iii) $Q\left(\{(u,v) \in (0,1)^2 \mid u < v\}\right) = 1$.

For $\widetilde{\mathcal{U}}_+$ a measurable subset of $\mathcal{U}_+$ satisfying (2.5.4), let $\widetilde{\Psi}_+ : [0,1] \to \mathbb{R}_+$ be defined for all $u \in [0,1]$ by $\widetilde{\Psi}_+(u) = \int_0^u \mathbb{1}_{\widetilde{\mathcal{U}}_+}(v)(F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv$. Let $\varphi : [0,1] \to [0,1]$ and $\widetilde{\varphi} : [0,1] \to [0,1]$ be defined for all $u \in [0,1]$ by

$$\begin{aligned}
\varphi(u) &= \Psi_-^{-1}(\widetilde{\Psi}_+(u)) = \inf\{r \in [0,1] \mid \Psi_-(r) \geq \widetilde{\Psi}_+(u)\};\\
\widetilde{\varphi}(u) &= \widetilde{\Psi}_+^{-1}(\Psi_-(v)) = \inf\{r \in [0,1] \mid \widetilde{\Psi}_+(r) \geq \Psi_-(u)\},
\end{aligned}$$

which are well defined thanks to the equality $\Psi_-(1) = \widetilde{\Psi}_+(1)$, consequence of (2.5.4). Let then $Q^{IT}_{\widetilde{\mathcal{U}}_+}$ be the measure defined on $(0,1)^2$ by

$$Q^{IT}_{\widetilde{\mathcal{U}}_+}(du,dv) = \frac{1}{\gamma}\mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du\,\pi_+^Q(u,dv) \quad \text{where } \pi_+^Q(u,dv) = \mathbb{1}_{\{0<\varphi(u)<1\}}\,\delta_{\varphi(u)}(dv). \tag{2.5.6}$$

Proposition 2.5.3. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{dcx} \nu$ and $\nu \not\leq_{st} \mu$. Let $\widetilde{\mathcal{U}}_+$ be a measurable subset of $\mathcal{U}_+$ such that (2.5.4) holds. The measure $Q^{IT}_{\widetilde{\mathcal{U}}_+}$ defined by (2.5.6) is an element of $\mathcal{Q}$. Moreover,*

$$Q^{IT}_{\widetilde{\mathcal{U}}_+}(du,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv\,\pi_-^Q(v,du) \quad \textit{where } \pi_-^Q(v,du) = \mathbb{1}_{\{0<\widetilde{\varphi}(v)<1\}}\,\delta_{\widetilde{\varphi}(v)}(du). \tag{2.5.7}$$

Proof of Proposition 2.5.3. A mild adaptation of the proof of Proposition 2.3.1 is conclusive. In particular, (2.3.5) is replaced with

$$\int_0^1 h(u,\varphi(u))\,d\widetilde{\Psi}_+(u) = \int_0^1 h(\widetilde{\varphi}(v),v)\,d\Psi_-(v),$$

for any measurable and bounded function $h : [0,1]^2 \to \mathbb{R}$, consequence of Lemma 2.6.1 below with $f_1 : u \mapsto \mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u)$, $f_2 : v \mapsto (F_\mu^{-1} - F_\nu^{-1})^-(v)$ and $u_0 = 1$, which gives the key property to show that $Q^{IT}_{\widetilde{\mathcal{U}}_+} \in \mathcal{Q}$. □

The existence of the inverse transform supermartingale coupling introduced below for the choice $\widetilde{\mathcal{U}}_+ = (0,u_d)$ implies that $\mathcal{Q}$ is non-empty. More generally, for any measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ satisfying (2.5.4), there exists $Q \in \mathcal{Q}$ with first marginal $\frac{1}{\gamma}\mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du$. For Q an element of $\mathcal{Q}$, let π_+^Q and π_-^Q be two sub-Markov kernels such that

$$Q(du,dv) = \frac{1}{\gamma}\mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1}(u) - F_\nu^{-1})^+(u)\,du\,\pi_+^Q(u,dv) = \frac{1}{\gamma}(F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv\,\pi_-^Q(v,du).$$

Let $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ be the Markov kernel defined by

$$\begin{cases} \displaystyle\int_{(0,1)} \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(v)}(dy) \pi_+^Q(u, dv) + \int_{(0,1)} \frac{F_\nu^{-1}(v) - F_\mu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \pi_+^Q(u, dv) \delta_{F_\nu^{-1}(u)}(dy) \\ \qquad \text{for } u \in \widetilde{\mathcal{U}}_+ \text{ such that } \pi_+^Q(u, \{v \in (0,1) \mid F_\nu^{-1}(v) > F_\mu^{-1}(u)\}) = 1; \\ \displaystyle\int_{\widetilde{\mathcal{U}}_+} \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(v)}(dy) \pi_-^Q(u, dv) + \int_{\widetilde{\mathcal{U}}_+} \frac{F_\nu^{-1}(v) - F_\mu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \pi_-^Q(u, dv) \delta_{F_\nu^{-1}(u)}(dy) \\ \qquad \text{for } u \in \mathcal{U}_- \text{ such that } \pi_-^Q(u, \{v \in (0,1) \mid F_\nu^{-1}(v) < F_\mu^{-1}(u)\}) = 1; \\ \delta_{F_\nu^{-1}(u)}(dy) \qquad \text{otherwise.} \end{cases} \tag{2.5.8}$$

The idea of this construction is as follows: for $u \in \mathcal{U}_-$, we can associate to $F_\mu^{-1}(u)$ a martingale contribution with $F_\nu^{-1}(u)$ and $F_\nu^{-1}(v)$ as in Section 2.2. If $F_\mu^{-1}(u) = F_\nu^{-1}(u)$, we associate $F_\nu^{-1}(u)$ to $F_\mu^{-1}(u)$. For $u \in \mathcal{U}_+ \backslash \widetilde{\mathcal{U}}_+$, we only associate $F_\nu^{-1}(u) < F_\mu^{-1}(u)$ to $F_\mu^{-1}(u)$ since there is no partner $v \in \mathcal{U}_- \cap (u, 1)$ available to construct a martingale contribution: all such possible partners have already been associated to values in $\widetilde{\mathcal{U}}_+$. Since du-almost all u in $\mathcal{U}_+ \cap [0, \overline{u}]$ belong to $\widetilde{\mathcal{U}}_+$, our construction is such that we associate to $F_\mu^{-1}(u)$ a martingale contribution at least for du-almost all $u \in \mathcal{U}_+ \cap [0, \overline{u}]$, which is actually not a particularity of our construction but a common property satisfied by all supermartingale couplings, as shown in the next proposition.

Proposition 2.5.4. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{dcx} \nu$, $M(dx, dy) = \mu(dx)\, m(x, dy)$ be a supermartingale coupling between μ and ν and $\overline{u}$ be defined by* (2.5.5)*. Then $\int_\mathbb{R} y\, m(x, dy) = x$ for $\mu(dx)$-almost all $x \leq F_\mu^{-1}(\overline{u})$, or equivalently,*

$$\int_\mathbb{R} \left(x - \int_\mathbb{R} y\, m(x, dy) \right) \mathbb{1}_{\{x \leq F_\mu^{-1}(\overline{u})\}}\, \mu(dx) = 0. \tag{2.5.9}$$

Proof. Let $(\widetilde{m}(u, dy))_{u \in (0,1)} = (m(F_\mu^{-1}(u), dy))_{u \in (0,1)}$ and η be the image of μ by the map $x \mapsto \int_\mathbb{R} y\, m(x, dy)$. Since M is a supermartingale coupling, we deduce by the inverse transform sampling that $F_\mu^{-1}(u) \geq \int_\mathbb{R} y\, \widetilde{m}(u, dy)$ for $du-$almost all $u \in (0,1)$. Therefore,

$$\int_0^{\overline{u}} F_\mu^{-1}(u)\, du \geq \int_0^{\overline{u}} \int_\mathbb{R} y\, \widetilde{m}(u, dy)\, du = \int_0^{\overline{u}} \int_\mathbb{R} y\, m(F_\mu^{-1}(u), dy)\, du. \tag{2.5.10}$$

Let U and V be two independent random variables uniformly distributed on $(0,1)$. By the inverse transform sampling, $\int_\mathbb{R} y\, m(F_\mu^{-1}(U), dy)$ is distributed according to η, so by Lemma 2.6.6 below, the map $f : (0,1)^2 \to \mathbb{R}$ defined for all $u, v \in (0,1)$ by

$$f(u, v) = F_\eta \left(\left(\int_\mathbb{R} y\, m(F_\mu^{-1}(u), dy) \right)_- \right) + v\eta \left(\left\{ \int_\mathbb{R} y\, m(F_\mu^{-1}(u), dy) \right\} \right)$$

is such that $f(U, V)$ is uniformly distributed on $(0,1)$ and $F_\eta^{-1}(f(U, V)) = \int_\mathbb{R} y\, m(F_\mu^{-1}(U), dy)$ almost surely. For $d \in \{1, 2\}$, let λ_d denote the Lebesgue measure on $[0,1]^d$ and $A = f((0, \overline{u}) \times$

$(0,1))$. Then $\overline{u} = \lambda_2((0,\overline{u}) \times (0,1)) \le \lambda_2(f^{-1}(f((0,\overline{u}) \times (0,1)))) = \lambda_1(A)$. We deduce that $\lambda_1(A \cap (0,\overline{u})^\complement) = \lambda_1(A) - \lambda_1(A \cap (0,\overline{u})) \ge \lambda_1((0,\overline{u})) - \lambda_1(A \cap (0,\overline{u})) = \lambda_1(A^\complement \cap (0,\overline{u}))$ and

$$\begin{aligned}\int_0^{\overline{u}} \int_{\mathbb{R}} y\, m(F_\mu^{-1}(u), dy)\, du &= \int_{(0,\overline{u})\times(0,1)} F_\eta^{-1}(f(u,v))\, du\, dv = \int_{A\cap(0,\overline{u})^\complement} F_\eta^{-1}(u)\, du + \int_{A\cap(0,\overline{u})} F_\eta^{-1}(u)\, du \\ &\ge \lambda_1(A \cap (0,\overline{u})^\complement) F_\eta^{-1}(\overline{u}) + \int_{A\cap(0,\overline{u})} F_\eta^{-1}(u)\, du \\ &\ge \lambda_1(A^\complement \cap (0,\overline{u})) F_\eta^{-1}(\overline{u}) + \int_{A\cap(0,\overline{u})} F_\eta^{-1}(u)\, du \\ &\ge \int_{A^\complement\cap(0,\overline{u})} F_\eta^{-1}(u)\, du + \int_{A\cap(0,\overline{u})} F_\eta^{-1}(u)\, du = \int_0^{\overline{u}} F_\eta^{-1}(u)\, du.\end{aligned}$$

For any convex function $f : \mathbb{R} \to \mathbb{R}$, Jensen's inequality yields

$$\int_{\mathbb{R}} f(y)\, \eta(dy) = \int_{\mathbb{R}} f\left(\int_{\mathbb{R}} y\, m(x, dy)\right) \mu(dx) \le \int_{\mathbb{R}} \int_{\mathbb{R}} f(y)\, m(x, dy)\, \mu(dx) = \int_{\mathbb{R}} f(y)\, \nu(dy),$$

hence $\eta \le_{cx} \nu$. We then deduce from (2.3.3) that $\int_0^{\overline{u}} F_\eta^{-1}(u)\, du \ge \int_0^{\overline{u}} F_\nu^{-1}(v)\, dv$. Finally, we showed that

$$0 = \Psi_+(\overline{u}) - \Psi_-(\overline{u}) = \int_0^{\overline{u}} F_\mu^{-1}(u)\, du - \int_0^{\overline{u}} F_\nu^{-1}(u)\, du \ge \int_0^{\overline{u}} F_\mu^{-1}(u)\, du - \int_0^{\overline{u}} \int_{\mathbb{R}} y\, m(F_\mu^{-1}(u), dy)\, du \ge 0,$$

where the last inequality comes from (2.5.10). Therefore, $\int_0^{\overline{u}} (F_\mu^{-1}(u) - \int_{\mathbb{R}} y\, m(F_\mu^{-1}(u), dy))\, du = 0$. Let $u \in (0,1)$ be such that $\overline{u} \le u$ and $F_\mu^{-1}(\overline{u}) = F_\mu^{-1}(u)$. Then $\Psi_+(u) = \Psi_+(\overline{u}) = \Psi_-(\overline{u}) \le \Psi_-(u) \le \Psi_+(u)$, so these inequalities are equalities and $u = \overline{u}$ by definition of $\overline{u}$. Therefore, $u \le \overline{u} \iff F_\mu^{-1}(u) \le F_\mu^{-1}(\overline{u})$ and by the inverse transform sampling,

$$\begin{aligned} 0 &= \int_0^1 \left(F_\mu^{-1}(u) - \int_{\mathbb{R}} y\, m(F_\mu^{-1}(u), dy)\right) \mathbb{1}_{F_\mu^{-1}(u) \le F_\mu^{-1}(\overline{u})}\, du \\ &= \int_0^1 \left(x - \int_{\mathbb{R}} y\, m(x, dy)\right) \mathbb{1}_{x \le F_\mu^{-1}(\overline{u})}\, \mu(dx), \end{aligned}$$

which proves (2.5.9). $\square$

Let $(m^Q(x,dy))_{x\in\mathbb{R}}$ be the Markov kernel defined as in (2.2.6) with $(\widetilde{m}(u,dy))_{u\in(0,1)}$ replaced with $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$. Then $\mu(dx)\, m^Q(x,dy)$ is expected to be a supermartingale coupling between μ and ν, as the next proposition states.

Proposition 2.5.5. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \le_{dcx} \nu$ and $\nu \not\le_{st} \mu$. Then for all Q in the non-empty set $\mathcal{Q}$, the probability measure $M^Q(dx,dy) = \mu(dx)\, m^Q(x,dy)$ is a supermartingale coupling between μ and ν.*

Notice that M^Q is a martingale coupling between μ and ν iff μ and ν have equal means, which is equivalent to $\Psi_+(1) = \Psi_-(1)$.

Proof of Proposition 2.5.5. With the very same arguments as in Section 2.2, we show that $M^Q(dx,dy)$ is a coupling between μ and ν (see Proposition 2.2.3). The same calculation as (2.2.10) for du-almost all $u \in \widetilde{\mathcal{U}}_+ \cup \mathcal{U}_-$ and the definition of $\widetilde{m}^Q$ for $u \in \mathcal{U}_0$ and $u \in \mathcal{U}_+ \backslash \widetilde{\mathcal{U}}_+$ yield $\int_\mathbb{R} |y| \widetilde{m}^Q(u,dy) < +\infty$ and

$$\int_\mathbb{R} y\, \widetilde{m}^Q(u,dy) = \begin{cases} F_\mu^{-1}(u) & \text{for } du\text{-almost all } u \in \widetilde{\mathcal{U}}_+ \cup \mathcal{U}_-; \\ F_\nu^{-1}(u) = F_\mu^{-1}(u) & \text{for } u \in \mathcal{U}_0; \\ F_\nu^{-1}(u) < F_\mu^{-1}(u) & \text{for } u \in \mathcal{U}_+ \backslash \widetilde{\mathcal{U}}_+. \end{cases} \tag{2.5.11}$$

Therefore, for du-almost all $u \in (0,1)$, $\int_\mathbb{R} y\, \widetilde{m}^Q(u,dy) \le F_\mu^{-1}(u)$. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable nonnegative and bounded function. By Lemma 2.2.6,

$$\int_{\mathbb{R}\times\mathbb{R}} h(x)(y-x)\, \mu(dx)\, m^Q(x,dy) = \int_0^1 h(F_\mu^{-1}(u)) \left(\int_\mathbb{R} (y - F_\mu^{-1}(u))\, \widetilde{m}^Q(u,dy) \right) du \le 0. \tag{2.5.12}$$

Therefore, for all $Q \in \mathcal{Q}$, $M^Q(dx,dy)$ is a supermartingale coupling between μ and ν. □

For $\widetilde{\mathcal{U}}_+$ a measurable subset of $\mathcal{U}_+$ which satifies (2.5.4), let us write $(m^{IT}_{\widetilde{\mathcal{U}}_+}(x,dy))_{x\in\mathbb{R}}$ instead of $(m^{Q^{IT}_{\widetilde{\mathcal{U}}_+}}(x,dy))_{x\in\mathbb{R}}$ and $(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+}(u,dy))_{u\in(0,1)}$ instead of $(\widetilde{m}^{Q^{IT}_{\widetilde{\mathcal{U}}_+}}(u,dy))_{u\in(0,1)}$, whose definition, given by (2.5.8), reduces to

$$\begin{cases} \dfrac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(\varphi(u))}(dy) + \left(1 - \dfrac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)}\right) \delta_{F_\nu^{-1}(u)}(dy) \\ \qquad \text{if } u \in \widetilde{\mathcal{U}}_+,\ F_\nu^{-1}(\varphi(u)) > F_\mu^{-1}(u) > F_\nu^{-1}(u) \text{ and } \varphi(u) < 1; \\ \dfrac{F_\nu^{-1}(u) - F_\mu^{-1}(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(\widetilde{\varphi}(u))} \delta_{F_\nu^{-1}(\widetilde{\varphi}(u))}(dy) + \left(1 - \dfrac{F_\nu^{-1}(u) - F_\mu^{-1}(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(\widetilde{\varphi}(u))}\right) \delta_{F_\nu^{-1}(u)}(dy) \\ \qquad \text{if } F_\nu^{-1}(\widetilde{\varphi}(u)) < F_\mu^{-1}(u) < F_\nu^{-1}(u) \text{ and } \widetilde{\varphi}(u) < 1; \\ \delta_{F_\nu^{-1}(u)}(dy) \qquad \text{otherwise.} \end{cases} \tag{2.5.13}$$

Then $M^{IT}_{\widetilde{\mathcal{U}}_+}(dx,dy) = \mu(dx)\, m^{IT}_{\widetilde{\mathcal{U}}_+}(x,dy)$ is a supermartingale coupling. Let $Q^{ITS} = Q^{IT}_{(0,u_d)}$, that is the element of $\mathcal{Q}$ defined by (2.5.6) for $\widetilde{\mathcal{U}}_+ = (0,u_d)$. Denoting $(\widetilde{m}^{ITS}(u,dy))_{u\in(0,1)}$ and $(m^{ITS}(x,dy))_{x\in\mathbb{R}}$ respectively for $(\widetilde{m}^{Q^{ITS}}(u,dy))_{u\in(0,1)} = (\widetilde{m}^{IT}_{(0,u_d)}(u,dy))_{u\in(0,1)}$ and $(m^{Q^{ITS}}(x,dy))_{x\in\mathbb{R}} = (m^{IT}_{(0,u_d)}(x,dy))_{x\in\mathbb{R}}$, we call inverse transform supermartingale coupling the probability measure $M^{ITS}(dx,dy) = \mu(dx)\, m^{ITS}(x,dy)$.

The next statement generalises Proposition 2.2.18.

Proposition 2.5.6. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \le_{dcx} \nu$ and $\nu \not\le_{st} \mu$. Let $Q \in \mathcal{Q}$. Then the Markov kernel $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ minimises*

$$\int_0^1 \int_\mathbb{R} |F_\nu^{-1}(u) - y|\, \widetilde{m}(u,dy)\, du$$

among all Markov kernels $(\widetilde{m}(u,dy))_{u\in(0,1)}$ *such that*

$$\begin{cases} & \int_{u\in(0,1)} \widetilde{m}(u,dy)\,du \quad = \quad \nu(dy) \\ \int_{\mathbb{R}} |y|\,\widetilde{m}(u,dy) < +\infty \quad \text{and} & \int_{\mathbb{R}} y\,\widetilde{m}(u,dy) \quad \leq \quad F_\mu^{-1}(u), \ du\text{-almost everywhere on } (0,1) \end{cases} \tag{2.5.14}$$

Moreover, $\int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\,\widetilde{m}^Q(u,dy)\,du = 2\Psi_-(1)$.

Proof. Let $\widetilde{m}$ be a Markov kernel satisfying (2.5.14). By monotonicity of the negative part and Jensen's inequality, for du-almost every $u \in (0,1)$,

$$(F_\mu^{-1} - F_\nu^{-1})^-(u) \leq \left(\int_{\mathbb{R}} (y - F_\nu^{-1}(u))\,\widetilde{m}(u,dy)\right)^- \leq \int_{\mathbb{R}} (y - F_\nu^{-1}(u))^-\,\widetilde{m}(u,dy).$$

Using the equality $2x^- = |x| - x$ valid for $x \in \mathbb{R}$ and the inverse transform sampling, we deduce that

$$\begin{aligned} &2\Psi_-(1) \\ &\leq 2\int_0^1 \int_{\mathbb{R}} (y - F_\nu^{-1}(u))^-\,\widetilde{m}(u,dy)\,du \\ &= \int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\,\widetilde{m}(u,dy)\,du - \int_0^1 \int_{\mathbb{R}} y\,\widetilde{m}(u,dy)\,du + \int_0^1 F_\nu^{-1}(u)\,du \\ &= \int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\,\widetilde{m}(u,dy)\,du - \int_{\mathbb{R}} y\,\nu(dy) + \int_{\mathbb{R}} y\,\nu(dy) = \int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\,\widetilde{m}(u,dy)\,du. \end{aligned}$$

According to Proposition 2.5.5, $\mu(dx)\,m^Q(x,dy)$ is a coupling between μ and ν, so by Lemma 2.2.6, $\int_{u\in(0,1)} \widetilde{m}^Q(u,dy)\,dy = \nu(dy)$. Moreover, we deduce from (2.5.11) that $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ satisfies (2.5.14). Therefore, to conclude, it is sufficient to prove that $\int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\,\widetilde{m}^Q(u,dy)\,du = 2\Psi_-(1)$.

Using the definition (2.5.8) of $\widetilde{m}^Q$, we get for du-almost all $u \in (0,1)$

$$\begin{aligned} \int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\,\widetilde{m}^Q(u,dy) = &\int_{(0,1)} \mathbb{1}_{\widetilde{\mathcal{U}}_+}(u) \frac{(F_\mu^{-1} - F_\nu^{-1})^+(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} |F_\nu^{-1}(u) - F_\nu^{-1}(v)| \pi_+^Q(u,dv) \\ &+ \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^-(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(v)} |F_\nu^{-1}(u) - F_\nu^{-1}(v)| \pi_-^Q(u,dv). \end{aligned}$$

A mild adaptation of the proof of Lemma 2.2.5 yields for du-almost all $u \in (0,1)$,

$$\begin{cases} u \in \widetilde{\mathcal{U}}_+ & \Longrightarrow \quad F_\nu^{-1}(v) > F_\mu^{-1}(u), \ \pi_+^Q(u,dv)\text{-a.e;} \\ u \in \mathcal{U}_- & \Longrightarrow \quad F_\nu^{-1}(v) < F_\mu^{-1}(u), \ \pi_-^Q(u,dv)\text{-a.e.} \end{cases} \tag{2.5.15}$$

We deduce that $\int_{\mathbb{R}} |F_\nu^{-1}(u) - y|\,\widetilde{m}^Q(u,dy) \leq \mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u) + (F_\nu^{-1} - F_\mu^{-1})^-(u)$ for du-almost all $u \in (0,1)$. Using (2.5.4) for $u = 1$, we conclude that

$$\int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\,\widetilde{m}^Q(u,dy)\,du \leq 2\Psi_-(1).$$

□

The next statement generalises the first statement in Theorem 2.2.12.

Proposition 2.5.7. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \le_{dcx} \nu$ and $\nu \not\le_{st} \mu$. For all Q in the non-empty set $\mathcal{Q}$,*

$$\int_{\mathbb{R}\times\mathbb{R}} |x-y|\, M^Q(dx,dy) \le 2\Psi_-(1) + \mathcal{W}_1(\mu,\nu). \tag{2.5.16}$$

Proof. Let $Q \in \mathcal{Q}$ and let $\widetilde{m}^Q$ be the Markov kernel defined by (2.5.8). By Lemma 2.2.6 and Proposition 2.5.6,

$$\begin{aligned}
\int_{\mathbb{R}\times\mathbb{R}} |y-x|\, M^Q(dx,dy) &= \int_{\mathbb{R}\times\mathbb{R}} |y-x|\, \mu(dx)\, m^Q(x,dy) = \int_0^1 \int_{\mathbb{R}} |y - F_\mu^{-1}(u)|\, \widetilde{m}^Q(u,dy)\, du \\
&\le \int_0^1 \int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\, \widetilde{m}^Q(u,dy)\, du \\
&\quad + \int_0^1 \int_{\mathbb{R}} |F_\nu^{-1}(u) - F_\mu^{-1}(u)|\, \widetilde{m}^Q(u,dy)\, du \\
&= 2\Psi_-(1) + \mathcal{W}_1(\mu,\nu).
\end{aligned}$$

□

Among all the measurable subsets $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ which satisfy (2.5.4), $(0,u_d)$ is the leftmost one. This is one of the reasons for which the inverse transform supermartingale coupling plays a particular role among the supermartingale couplings which derive from $\mathcal{Q}$, as stated in the next Proposition. It is also natural to investigate the rightmost measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ which satisfies (2.5.4), that is such that $\widetilde{\Psi}_+$ is as small as possible. Notice that a measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ satisfies (2.5.4) iff it satisfies

$$\forall u \in (0,1),\ \int_{1-u}^1 \mathbb{1}_{\widetilde{\mathcal{U}}_+(v)} (F_\mu^{-1} - F_\nu^{-1})^+(v)\, dv \le \int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})^-(v)\, dv, \text{ with equality for } u=1. \tag{2.5.17}$$

Therefore, we look for a measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ such that for $u\in[0,1]$, $\int_{1-u}^1 \mathbb{1}_{\widetilde{\mathcal{U}}_+}(v)\, (F_\mu^{-1} - F_\nu^{-1})^+(v)\, dv$ is as big as possible while still being smaller than $\int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})^-(v)\, dv$ with equality for $u=1$. This is equivalent to have

$$\begin{aligned}
&\int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})(v)\, dv - \int_{1-u}^1 \mathbb{1}_{\mathcal{U}_+\setminus\widetilde{\mathcal{U}}_+}(v) (F_\mu^{-1} - F_\nu^{-1})(v)\, dv \\
&= \int_{1-u}^1 \mathbb{1}_{\widetilde{\mathcal{U}}_+}(v)\, (F_\mu^{-1} - F_\nu^{-1})^+(v)\, dv - \int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})^-(v)\, dv \le 0,
\end{aligned}$$

with equality for $u=1$. Therefore, we look for a measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ such that $\int_{1-u}^1 \mathbb{1}_{\mathcal{U}_+\setminus\widetilde{\mathcal{U}}_+}(v)(F_\mu^{-1} - F_\nu^{-1})(v)\, dv$ is as small as possible while still being greater than $\int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})(v)\, dv$. Let then $R : [0,1] \to \mathbb{R}$ be defined for all $u\in[0,1]$ by

$$R(u) = \sup_{v\in[0,u]} \int_{1-v}^1 (F_\mu^{-1} - F_\nu^{-1})(w)\, dw, \tag{2.5.18}$$

which can easily be proved to be the minimum of the set of nonnegative and nondecreasing functions $f : [0,1] \to \mathbb{R}$ which satisfy $f(u) \geq \int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})(v)\,dv$ for all $u \in [0,1]$. The following proposition makes the connection between R and the rightmost measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ which satisfies (2.5.4).

Proposition 2.5.8. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{dcx} \nu$, R be defined by (2.5.18) and $B : [0,1] \ni u \mapsto \int_{1-u}^1 (F_\mu^{-1} - F_\nu^{-1})(v)\,dv$. Let*

$$\widetilde{\mathcal{U}}_+^R = \{u \in \mathcal{U}_+ \mid R(1-u) > B(1-u)\} \quad \text{and} \quad \widetilde{\Psi}_+^R : u \mapsto \int_0^u \mathbb{1}_{\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1} - F_\nu^{-1})(v)\,dv. \tag{2.5.19}$$

Then $\widetilde{\mathcal{U}}_+^R$ is a measurable subset of $\mathcal{U}_+$ which satisfies (2.5.4) and for any measurable subset $\widetilde{\mathcal{U}}_+$ of $\mathcal{U}_+$ satisfying (2.5.4), we have that

$$\forall u \in [0,1], \quad \widetilde{\Psi}_+^R(u) \leq \widetilde{\Psi}_+(u).$$

Proof. For $\varepsilon > 0$, let $\varphi_\varepsilon : \mathbb{R} \to \mathbb{R}$ be a continuously differentiable map such that $\varphi_\varepsilon(x) = 0$ for $x \leq -\varepsilon$, $\varphi_\varepsilon(x) = x$ for $x \geq \varepsilon$, $\varphi_\varepsilon'(x) \in [0,1]$ for $x \in \mathbb{R}$ and $\varphi_\varepsilon'(0) = 1$. One could choose for instance

$$\varphi_\varepsilon : x \mapsto \left(\frac{\varepsilon}{2} + x + \frac{1}{2\varepsilon}x^2\right)\mathbb{1}_{\{-\varepsilon < x \leq 0\}} + \left(\frac{\varepsilon}{2} + x - \frac{3}{2\varepsilon}x^2 + \frac{1}{\varepsilon^2}x^3\right)\mathbb{1}_{\{0 < x \leq \varepsilon\}} + x\mathbb{1}_{\{x > \varepsilon\}}.$$

Since φ_ε is continuously differentiable, the chain rule formula (see for instance Proposition 4.6 Chapter 0 [164]) yields for all $0 \leq u < v \leq 1$,

$$\varphi_\varepsilon((B-R)(v)) - \varphi_\varepsilon((B-R)(u)) = \int_{(u,v]} \varphi_\varepsilon'((B-R)(w))\,d(B-R)(w).$$

We deduce from the dominated convergence theorem for $\varepsilon \to 0$ that for all $0 \leq u < v \leq 1$,

$$(B-R)^+(v) - (B-R)^+(u) = \int_{(u,v]} \mathbb{1}_{\{(B-R)(w) \geq 0\}}\,d(B-R)(w).$$

Since $R(u) \geq B(u)$ for all $u \in [0,1]$, we get that

$$0 = d(B-R)^+(u) = \mathbb{1}_{\{R(u)=B(u)\}}\,dB(u) - \mathbb{1}_{\{R(u)=B(u)\}}\,dR(u). \tag{2.5.20}$$

According to Theorem 1.1.1 [153], the map R solves a Skorokhod problem and may increase only at points $u \in (0,1)$ such that $R(u) = B(u)$, that is $dR(u) = \mathbb{1}_{\{R(u)=B(u)\}}\,dR(u)$. With (2.5.20), we deduce that

$$dR(u) = \mathbb{1}_{\{R(u)=B(u)\}}(F_\mu^{-1} - F_\nu^{-1})(1-u)\,du.$$

By monotonicity of R, we have

$$0 \leq \int_{(0,1)} \mathbb{1}_{\{F_\mu^{-1}(1-u) \leq F_\nu^{-1}(1-u)\}}\,dR(u)$$

$$= \int_{(0,1)} \mathbb{1}_{\{R(u)=B(u)\}} \mathbb{1}_{\{F_\mu^{-1}(1-u) \le F_\nu^{-1}(1-u)\}} (F_\mu^{-1} - F_\nu^{-1})(1-u)\, du \le 0,$$

so those inequalities are equalities and for $dR(u)$-almost all $u \in (0,1)$, $1-u \in \mathcal{U}_+$. Therefore, $dR(u) = \mathbb{1}_{\{R(u)=B(u)\}} \mathbb{1}_{\{(1-u)\in\mathcal{U}_+\}} (F_\mu^{-1} - F_\nu^{-1})(1-u)\, du$, so that the set $\widetilde{\mathcal{U}}_+^R := \{u \in \mathcal{U}_+ \mid R(1-u) > B(1-u)\}$ is such that for all $u \in [0,1]$, $R(u) = \int_{1-u}^1 \mathbb{1}_{\mathcal{U}_+ \setminus \widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1} - F_\nu^{-1})(v)\, dv$.

Let us now prove that $\widetilde{\mathcal{U}}_+^R$ satisfies (2.5.4), which will end the proof. Let $\widetilde{\Psi}_+^R : u \mapsto \int_0^u \mathbb{1}_{\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1} - F_\nu^{-1})(v)\, dv$. On the one hand, using that $\Psi_+(u) \ge \Psi_-(u)$ for all $u \in [0,1]$, we have

$$\begin{aligned} B(1) \le R(1) = \sup_{v\in[0,1]} B(v) &= \sup_{v\in[0,1]} \left(\Psi_+(1) - \Psi_-(1) - \Psi_+(1-v) + \Psi_-(1-v)\right) \\ &\le \Psi_+(1) - \Psi_-(1) = B(1), \end{aligned}$$

so those inequalities are equalities and $R(1) = \Psi_+(1) - \Psi_-(1)$. We deduce that $\widetilde{\Psi}_+^R(1) = \Psi_+(1) - R(1) = \Psi_-(1)$. On the other hand, for $u \in [0,1]$,

$$\widetilde{\Psi}_+^R(u) = \widetilde{\Psi}_+^R(1) + R(1-u) - \Psi_+(1) + \Psi_+(u) \ge \Psi_-(1) + B(1-u) - \Psi_+(1) + \Psi_+(u) = \Psi_-(u),$$

so $\widetilde{\mathcal{U}}_+^R$ satisfies (2.5.4). $\square$

Proposition 2.5.9. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \le_{dcx} \nu$ and $\nu \not\le_{st} \mu$. For all $\rho \in \mathbb{R}$ and for any Markov kernel $(\widetilde{m}(u,dy))_{u\in(0,1)}$, let $\mathcal{C}_\rho(\widetilde{m})$ be defined by*

$$\mathcal{C}_\rho(\widetilde{m}) = \int_{\mathbb{R}\times(0,1)} |F_\nu^{-1}(u) - y|^\rho \mathbb{1}_{\{y \ne F_\nu^{-1}(u)\}}\, \widetilde{m}(u,dy)\, du. \tag{2.5.21}$$

Let $(\widetilde{m}^R(u,dy))_{u\in(0,1)} = (\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+^R}(u,dy))_{u\in(0,1)}$, where $\widetilde{\mathcal{U}}_+^R$ is defined by (2.5.19)*. Then, for all $Q \in \mathcal{Q}$,*

$$\begin{aligned} \forall \rho \in (-\infty, 1], &\quad \mathcal{C}_\rho(\widetilde{m}^{ITS}) \le \mathcal{C}_\rho(\widetilde{m}^Q); \\ \forall \rho \in [1,2], &\quad \mathcal{C}_\rho(\widetilde{m}^Q) \le \mathcal{C}_\rho(\widetilde{m}^{ITS}); \\ \forall \rho \in [2,+\infty), &\quad \mathcal{C}_\rho(\widetilde{m}^R) \le \mathcal{C}_\rho(\widetilde{m}^Q). \end{aligned} \tag{2.5.22}$$

Proof. Let $\widetilde{\mathcal{U}}_+$ be a subset of $\mathcal{U}_+$ which satisfies (2.5.4). Let Q be any element of $\mathcal{Q}$ with first marginal $\frac{1}{\gamma}\mathbb{1}_{\widetilde{\mathcal{U}}_+}(u)(F_\mu^{-1} - F_\nu^{-1})^+(u)\, du$, and $Q^{IT}_{\widetilde{\mathcal{U}}_+}$ be defined by (2.5.6). Reasoning like in the derivation of (2.3.13) and (2.3.15), we obtain

$$\begin{aligned} \forall \rho \in (-\infty,1] \cup [2,+\infty), \quad \int_0^1 &|F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\gamma u)_+) - F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\, du \\ &\le \int_{(0,1)^2} |F_\nu^{-1}(u_+) - F_\nu^{-1}(v_+)|^{\rho-1}\, Q(du,dv), \end{aligned} \tag{2.5.23}$$

and

$$\forall 1\le\rho\le 2,\quad \int_{(0,1)^2}|F_\nu^{-1}(u_+)-F_\nu^{-1}(v_+)|^{\rho-1}\,Q(du,dv)$$
$$\le\int_0^1|F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\gamma u)_+)-F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\,du.\quad(2.5.24)$$

Moreover, (2.3.17) and (2.3.18) generalise into

$$C_\rho(\widetilde{m}^Q)=2\gamma\int_{(0,1)^2}|F_\nu^{-1}(u_+)-F_\nu^{-1}(v_+)|^{\rho-1}\,Q(du,dv)$$

and

$$\mathcal{C}_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+})=2\gamma\int_0^1|F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\gamma u)_+)-F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+)|^{\rho-1}\,du.$$

We deduce that

$$\forall\rho\in(-\infty,1]\cup[2,+\infty),\quad \mathcal{C}_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+})\le\mathcal{C}_\rho(\widetilde{m}^Q)\quad\text{and}\quad\forall\rho\in[1,2],\quad\mathcal{C}_\rho(\widetilde{m}^Q)\le\mathcal{C}_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+}).\quad(2.5.25)$$

Notice that since $\Psi_-(u)\le\widetilde{\Psi}_+(u)$ for $u\in[0,1]$, we have $\Psi_-^{-1}(v)\ge\widetilde{\Psi}_+^{-1}(v)$ for $v\in(0,\gamma)$, so by monotonicity of F_ν^{-1}, we also have

$$C_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+})=2\gamma\int_0^1\left(F_\nu^{-1}(\Psi_-^{-1}(\gamma u)_+-F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\gamma u)_+))\right)^{\rho-1}\,du.\quad(2.5.26)$$

Let $\widetilde{\Psi}_+^{ITS}:[0,1]\to\mathbb{R}$ and $\widetilde{\Psi}_+^R:[0,1]\to\mathbb{R}$ be defined for all $u\in[0,1]$ by $\widetilde{\Psi}_+^{ITS}(u)=\int_0^{u\wedge u_d}(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv$ and $\widetilde{\Psi}_+^R(u)=\int_0^u\mathbb{1}_{\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv$. For all $u\in[0,u_d]$, $\widetilde{\Psi}_+(u)=\int_0^u\mathbb{1}_{\widetilde{\mathcal{U}}_+}(v)(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\le\int_0^u(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\le\widetilde{\Psi}_+^{ITS}(u)$ and for all $u\in[u_d,1]$, $\widetilde{\Psi}_+(u)\le\widetilde{\Psi}_+(1)=\Psi_-(1)=\widetilde{\Psi}^{ITS}(u)$. Moreover, let $\alpha:u\mapsto\int_{1-u}^1\mathbb{1}_{\mathcal{U}_+\backslash\widetilde{\mathcal{U}}_+}(v)(F_\mu^{-1}-F_\nu^{-1})(v)$. The map α is nonnegative, nondecreasing and satisfies

$$\begin{aligned}\alpha(u)&=\int_{1-u}^1\mathbb{1}_{\mathcal{U}_+}(v)(F_\mu^{-1}-F_\nu^{-1})(v)\,dv-\widetilde{\Psi}_+(1)+\widetilde{\Psi}_+(1-u)\\&\ge\int_{1-u}^1(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv-\Psi_-(1)+\Psi_-(1-u)\\&=\int_{1-u}^1(F_\mu^{-1}-F_\nu^{-1})(v)\,dv,\end{aligned}$$

where we used (2.5.4) for the inequality. By definition of R, we deduce that for all $u\in[0,1]$, $\alpha(u)\ge R(u)$, hence

$$\begin{aligned}\widetilde{\Psi}_+^R(u)&=\widetilde{\Psi}_+^R(1)-\int_u^1\mathbb{1}_{\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv=\widetilde{\Psi}_+(1)-\int_u^1\mathbb{1}_{\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\\&=\widetilde{\Psi}_+(1)-\Psi_+(1)+\Psi_+(u)+\int_u^1\mathbb{1}_{\mathcal{U}_+\backslash\widetilde{\mathcal{U}}_+^R}(v)(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\\&=\widetilde{\Psi}_+(1)-\Psi_+(1)+\Psi_+(u)+R(1-u)\\&\le\widetilde{\Psi}_+(1)-\Psi_+(1)+\Psi_+(u)+\alpha(1-u)=\widetilde{\Psi}_+(u).\end{aligned}$$

Since $\widetilde{\Psi}_+^R(u) \leq \widetilde{\Psi}_+(u) \leq \widetilde{\Psi}_+^{ITS}(u)$ for all $u \in [0,1]$, we deduce that

$$\forall u \in (0,\gamma), \quad (\widetilde{\Psi}_+^{ITS})^{-1}(u) \leq \widetilde{\Psi}_+^{-1}(u) \leq (\widetilde{\Psi}_+^R)^{-1}(u). \tag{2.5.27}$$

By (2.5.26), (2.5.27) and monotonicity of the maps $\mathbb{R}_+ \ni x \mapsto x^{\rho-1}$ and F_ν^{-1}, we have

$$\begin{aligned} &\forall \rho \in (-\infty,1], \quad C_\rho(\widetilde{m}^{ITS}) \leq C_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+}) \leq C_\rho(\widetilde{m}^R), \\ \text{and} \quad &\forall \rho \in [1,+\infty), \quad C_\rho(\widetilde{m}^R) \leq C_\rho(\widetilde{m}^{IT}_{\widetilde{\mathcal{U}}_+}) \leq C_\rho(\widetilde{m}^{ITS}). \end{aligned} \tag{2.5.28}$$

Then (2.5.22) is deduced from (2.5.25) and (2.5.28). □

We now show the stability of the inverse transform supermartingale coupling with respect to its marginals for the Wasserstein distance topology. So far, the inverse transform supermartingale coupling has been defined just before Proposition 2.5.6 for all $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ such that $\mu \leq_{dcx} \nu$ and $\nu \not\leq_{st} \mu$. When $\nu \leq_{st} \mu$, we simply define the inverse transform supermartingale coupling as the comonotonous coupling between μ and ν.

Proposition 2.5.10. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{dcx} \nu$. Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ be two sequences of probability measures on $\mathbb{R}$ with finite first moments such that for all $n \in \mathbb{N}$, $\mu_n \leq_{dcx} \nu_n$. For all $n \in \mathbb{N}$, let M_n^{ITS} (resp. M^{ITS}) be the inverse transform supermartingale coupling between μ_n and ν_n (resp. μ and ν).*

If $\mathcal{W}_1(\mu_n,\mu) \underset{n\to+\infty}{\longrightarrow} 0$ and $\mathcal{W}_1(\nu_n,\nu) \underset{n\to+\infty}{\longrightarrow} 0$, then

$$\mathcal{W}_1(M_n^{ITS}, M^{ITS}) \underset{n\to+\infty}{\longrightarrow} 0.$$

Proof. For all $n \in \mathbb{N}$, let $\Psi_{n+} : u \in [0,1] \mapsto \int_0^u (F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^+(v)\,dv$, $\Psi_{n-} : u \in [0,1] \mapsto \int_0^u (F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^-(v)\,dv$, $(u_d)_n = \Psi_{n+}^{-1}(\Psi_{n-}(1))$ if $\nu_n \not\leq_{st} \mu_n$ and $(u_d)_n = 0$ otherwise, $\widetilde{\mathcal{U}}_{n+} = (0,(u_d)_n)$, $\widetilde{\Psi}_{n+} : u \in [0,1] \mapsto \int_0^u \mathbb{1}_{\widetilde{\mathcal{U}}_{n+}}(v)(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^+(v)\,dv$, $\varphi_n = \Psi_{n-}^{-1} \circ \widetilde{\Psi}_{n+}$ and $\widetilde{\varphi}_n = \widetilde{\Psi}_{n+}^{-1} \circ \Psi_{n-}$. Let $H : \mathbb{R}^2 \to \mathbb{R}$ be a bounded and continuous function such that h is Lipschitz continuous with respect to its second variable. One can easily prove that (2.2.12) still holds in the supermartingale case, which writes, for $Q = Q^{IT}_{\widetilde{\mathcal{U}}_{n+}}$,

$$\begin{aligned} &\int_{\mathbb{R}^2} H(x,y)\,M_n^{ITS}(dx,dy) \\ &= \int_{(0,1)} H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u))\,du \\ &\quad + \int_{(0,1)} \mathbb{1}_{\widetilde{\mathcal{U}}_{n+}}(u) \frac{(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^+(u)}{F_{\nu_n}^{-1}(\varphi_n(u)) - F_{\nu_n}^{-1}(u)} (H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(\varphi_n(u))) - H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)))\,du \\ &\quad + \int_{(0,1)} \frac{(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^-(u)}{F_{\nu_n}^{-1}(u) - F_{\nu_n}^{-1}(\widetilde{\varphi}_n(u))} (H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(\widetilde{\varphi}_n(u))) - H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)))\,du, \end{aligned} \tag{2.5.29}$$

where the last two integrands are zero when $\nu_n \leq_{st} \mu_n$. Since μ_n converges weakly towards μ, then $F_{\mu_n}^{-1}(u)$ (resp. $F_{\nu_n}^{-1}(u)$) converges towards $F_\mu^{-1}(u)$ (resp. $F_\nu^{-1}(u)$) du-almost everywhere on $(0,1)$. Since H is continuous and bounded, by the dominated convergence theorem,

$$\int_{(0,1)} H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u))\,du \underset{n\to+\infty}{\longrightarrow} \int_{(0,1)} H(F_\mu^{-1}(u), F_\nu^{-1}(u))\,du. \tag{2.5.30}$$

Since for all $u \in [0,1]$, $x \mapsto x^+$ is Lipschitz continuous with constant 1,

$$\begin{aligned}
|\Psi_{n-}(u) - \Psi_-(u)| &\leq \int_0^u |(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^-(v) - (F_\mu^{-1} - F_\nu^{-1})^-(v)|\,dv \\
&\leq \int_0^u |F_{\mu_n}^{-1}(v) - F_\mu^{-1}(v)|\,dv + \int_0^u |F_{\nu_n}^{-1}(v) - F_\nu^{-1}(v)|\,dv \\
&\leq \mathcal{W}_1(\mu_n, \mu) + \mathcal{W}_1(\nu_n, \nu),
\end{aligned}$$

so Ψ_{n-} converges uniformly to Ψ_- on $[0,1]$. We deduce with the same reasoning that Ψ_{n+} converges uniformly to Ψ_+ on $[0,1]$. Since $\widetilde{\mathcal{U}}_{n+} = (0,(u_d)_n)$, we deduce from the definition of $(u_d)_n$ that for all $u \in [0,1]$, $\widetilde{\Psi}_{n+}(u) = \Psi_{n+}(u \wedge (u_d)_n) = \Psi_{n+}(u) \wedge \Psi_{n-}(1)$. Let $(a,b,c,d) \in \mathbb{R}^4$. Then $((a-b)^+ - (c-d)^+)((b-a)^+ - (d-c)^+) = -(a-b)^+(d-c)^+ - (c-d)^+(b-a)^+ \leq 0$, so $(a-b)^+ - (c-d)^+$ and $(b-a)^+ - (d-c)^+$ have opposite signs. Therefore, we can apply the inequality $|x| \leq |x+\alpha| \vee |x+\beta|$ valid for $(x,\alpha,\beta) \in \mathbb{R}^3$ such that α and β have opposite signs with $(x,\alpha,\beta) = (a\wedge b - c\wedge d, (a-b)^+ - (c-d)^+, (b-a)^+ - (d-c)^+)$, which yields

$$\begin{aligned}
|a\wedge b - c\wedge d| &\leq |a\wedge b - c\wedge d + (a-b)^+ - (c-d)^+| \vee |a\wedge b - c\wedge d + (b-a)^+ - (d-c)^+| \\
&= |a-c| \vee |b-d|.
\end{aligned} \tag{2.5.31}$$

Using (2.5.31) with $(a,b,c,d) = (\Psi_{n+}(u), \Psi_{n-}(1), \Psi_+(u), \Psi_-(1))$, we deduce that

$$|\widetilde{\Psi}_{n+}(u) - \widetilde{\Psi}_+(u)| = |\Psi_{n+}(u)\wedge\Psi_{n-}(1) - \Psi_+(u)\wedge\Psi_-(1)| \leq |\Psi_{n+}(u) - \Psi_+(u)| \vee |\Psi_{n-}(1) - \Psi_-(1)|,$$

hence $\widetilde{\Psi}_{n+}$ converges uniformly to $\widetilde{\Psi}_+$ on $[0,1]$. If $\nu \leq_{st} \mu$, then we deduce from the Lipschitz continuity of H with respect to its second variable, (2.5.29) and (2.5.30) that there exists $K \in \mathbb{R}_+$ such that

$$\begin{aligned}
&\left|\int_{\mathbb{R}\times\mathbb{R}} H(x,y)\,M_n^{ITS}(dx,dy) - \int_{\mathbb{R}\times\mathbb{R}} H(x,y)\,M^{ITS}(dx,dy)\right| \\
&\leq \left|\int_{(0,1)} H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u))\,du - \int_{(0,1)} H(F_\mu^{-1}(u), F_\nu^{-1}(u))\,du\right| + K(\widetilde{\Psi}_{n+}(1) + \Psi_{n-}(1)) \\
&\underset{n\to+\infty}{\longrightarrow} K(\widetilde{\Psi}_+(1) + \Psi_-(1)) = 0.
\end{aligned}$$

We conclude that $M_n^{ITS} \underset{n\to+\infty}{\longrightarrow} M^{ITS}$ for the weak convergence topology as soon as $\nu \leq_{st} \mu$. From now on, we suppose $\nu \not\leq_{st} \mu$. Since $\Psi_{n-}(1) \underset{n\to+\infty}{\longrightarrow} \Psi_-(1) > 0$, $\nu_n \not\leq_{st} \mu_n$ for n large enough, so we can suppose without loss of generality that $\nu_n \not\leq_{st} \mu_n$ for all $n \in \mathbb{N}$.

Using Lemma 2.6.3 below for the first equality, then Proposition 2.6.2 below for the second equality and the change of variables $u = \widetilde{\Psi}_{n+}(1)v$ with the equality $\widetilde{\Psi}_{n+}(1) = \Psi_{n-}(1)$ for the last equality, we have

$$\begin{aligned}
&\int_{(0,1)} \frac{(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^+(u)}{F_{\nu_n}^{-1}(\varphi_n(u)) - F_{\nu_n}^{-1}(u)} (H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(\varphi_n(u))) - H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)))\, du \\
&= \int_{(0,1)} \frac{H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(u))), F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\widetilde{\Psi}_{n+}(u)))) - H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(u))), F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(u))))}{F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\widetilde{\Psi}_{n+}(u))) - F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(u)))}\, d\widetilde{\Psi}_{n+}(u) \\
&= \int_{(0,\widetilde{\Psi}_{n+}(1))} \frac{H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(u)), F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(u))) - H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(u)), F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(u)))}{F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(u)) - F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(u))}\, du \\
&= \widetilde{\Psi}_{n+}(1) \int_{(0,1)} \frac{H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)v)), F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n-}(1)v))) - H(F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)v)), F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)v)))}{(F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n-}(1)v)) - F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)v)))}\, dv.
\end{aligned}$$

Since H is Lipschitz continuous with respect to its second variable, then the integrand above is bounded. Moreover, for all $x \in \mathbb{R}$, $|\widetilde{\Psi}_{n+}(F_{\mu_n}(x)) - \widetilde{\Psi}_+(F_\mu(x))| \le \sup_{[0,1]} |\widetilde{\Psi}_{n+} - \widetilde{\Psi}_+| + |\widetilde{\Psi}_+(F_{\mu_n}(x)) - \widetilde{\Psi}_+(F_\mu(x))|$, so $\widetilde{\Psi}_{n+}(F_{\mu_n}(x))/\widetilde{\Psi}_{n+}(1) \underset{n\to+\infty}{\to} \widetilde{\Psi}_+(F_\mu(x))/\widetilde{\Psi}_+(1)$ for all $x \in \mathbb{R}$ outside the at most countable set of discontinuities of F_μ. This implies that $d(\widetilde{\Psi}_{n+}(F_{\mu_n}(x))/\widetilde{\Psi}_{n+}(1))$ converges to $d(\widetilde{\Psi}_+(F_\mu(x))/\widetilde{\Psi}_+(1))$ for the weak convergence topology. We deduce the pointwise convergence of the left continuous pseudo-inverses du-almost everywhere on $(0,1)$, that is $F_{\mu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)u)) \underset{n\to+\infty}{\longrightarrow} F_\mu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)u))$ for du-almost all $u \in (0,1)$. In the same way, $F_{\nu_n}^{-1}(\widetilde{\Psi}_{n+}^{-1}(\widetilde{\Psi}_{n+}(1)u)) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)u))$ and $F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n-}(1)u)) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}(\Psi_-^{-1}(\Psi_-(1)u))$ for du-almost all $u \in (0,1)$. Therefore, by the dominated convergence theorem,

$$\begin{aligned}
&\int_{(0,1)} \frac{(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^+(u)}{F_{\nu_n}^{-1}(\varphi_n(u)) - F_{\nu_n}^{-1}(u)} (H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(\varphi_n(u))) - H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)))\, du \\
&\underset{n\to+\infty}{\longrightarrow} \widetilde{\Psi}_+(1) \int_{(0,1)} \frac{H(F_\mu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)v)), F_\nu^{-1}(\Psi_-^{-1}(\Psi_-(1)v))) - H(F_\mu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)v)), F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)v)))}{(F_\nu^{-1}(\Psi_-^{-1}(\Psi_-(1)v)) - F_\nu^{-1}(\widetilde{\Psi}_+^{-1}(\widetilde{\Psi}_+(1)v)))}\, dv \\
&= \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^+(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)} (H(F_\mu^{-1}(u), F_\nu^{-1}(\varphi(u))) - H(F_\mu^{-1}(u), F_\nu^{-1}(u)))\, du.
\end{aligned}$$

We can show in the same way that

$$\begin{aligned}
&\int_{(0,1)} \frac{(F_{\mu_n}^{-1} - F_{\nu_n}^{-1})^-(u)}{F_{\nu_n}^{-1}(u) - F_{\nu_n}^{-1}(\widetilde{\varphi}_n(u))} (H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(\widetilde{\varphi}_n(u))) - H(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)))\, du \\
&\underset{n\to+\infty}{\longrightarrow} \int_{(0,1)} \frac{(F_\mu^{-1} - F_\nu^{-1})^-(u)}{F_\nu^{-1}(u) - F_\nu^{-1}(\widetilde{\varphi}(u))} (H(F_\mu^{-1}(u), F_\nu^{-1}(\widetilde{\varphi}(u))) - H(F_\mu^{-1}(u), F_\nu^{-1}(u)))\, du.
\end{aligned}$$

Finally, we showed that

$$\int_{\mathbb{R}\times\mathbb{R}} H(x,y)\, M_n^{ITS}(x,dy) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}\times\mathbb{R}} H(x,y)\, M^{ITS}(x,dy),$$

for any bounded and continuous function $H : \mathbb{R}^2 \to \mathbb{R}$ which is Lipschitz continuous with respect to its second variable, that is $M_n^{ITS} \underset{n\to+\infty}{\longrightarrow} M^{ITS}$ for the weak convergence topology.

Since the convergence for the Wasserstein distance topology is equivalent to the convergence for the weak convergence topology and the convergence of the first order moments (see for instance Theorem 6.9 Chapter 6 [191]), $\int_{\mathbb{R}} |x| \, \mu_n(dx) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}} |x| \, \mu(dx)$ and $\int_{\mathbb{R}} |y| \, \nu_n(dy) \underset{n\to+\infty}{\longrightarrow} \int_{\mathbb{R}} |y| \, \nu(dy)$. Therefore, $\mathcal{W}_1(M_n^{ITS}, M^{ITS}) \underset{n\to+\infty}{\longrightarrow} 0$ when $\mathbb{R}^2$ is endowed with the L^1-norm. Since all norms on $\mathbb{R}^2$ are equivalent, $\mathcal{W}_1(M_n^{ITS}, M^{ITS}) \underset{n\to+\infty}{\longrightarrow} 0$ when $\mathbb{R}^2$ is endowed with any norm. □

2.6 Appendix

We begin with a key result for the construction of the inverse transform martingale coupling.

Lemma 2.6.1. *Let $f_1, f_2 : (0,1) \to \mathbb{R}$ be two measurable nonnegative and integrable functions and $u_0 \in [0,1]$ be such that $\int_0^{u_0} f_1(u) \, du = \int_0^1 f_2(u) \, du$. Let $\Psi_1 : [0,1] \ni u \mapsto \int_0^u f_1(v) \, dv$, $\Psi_2 : [0,1] \ni u \mapsto \int_0^u f_2(v) \, dv$ and $\Gamma = \Psi_2^{-1} \circ \Psi_1$ where Ψ_2^{-1} denotes the càg pseudo-inverse of Ψ_2. Then Γ is well defined on $[0, u_0]$ and for any measurable and bounded function $h : [0,1] \to \mathbb{R}$,*

$$\int_0^{u_0} h(\Gamma(u)) f_1(u) \, du = \int_0^1 h(v) f_2(v) \, dv.$$

The proof of Lemma 2.6.1 relies on the next proposition, which is a well known result of integration by continuous and nondecreasing substitution, whose proof can be found for instance in Proposition 4.10 Chapter 0 [164].

Proposition 2.6.2. *Let $a, b \in \mathbb{R}$ be such that $a < b$. Let $\Psi : [a,b] \to \mathbb{R}$ be a continuous and nondecreasing function. Then for any Borel function $f : [\Psi(a), \Psi(b)] \to \mathbb{R}$,*

$$\int_a^b f(\Psi(s)) \, d\Psi(s) = \int_{\Psi(a)}^{\Psi(b)} f(t) \, dt.$$

Proof of Lemma 2.6.1. Let $h : [0,1] \to \mathbb{R}$ be a measurable and bounded function. Since Ψ_1 is nondecreasing and continuous, using Proposition 2.6.2, we have

$$\int_0^{u_0} h(\Gamma(u)) f_1(u) \, du = \int_0^{u_0} h(\Psi_2^{-1}(\Psi_1(u))) \, d\Psi_1(u) = \int_0^{\Psi_1(u_0)} h(\Psi_2^{-1}(w)) \, dw.$$

Since $\int_0^{u_0} f_1(u) \, du = \int_0^1 f_2(u) \, du$, we have $\Psi_1(u_0) = \Psi_2(1)$, and since Ψ_2 is nondecreasing and continuous, using once again Proposition 2.6.2, we have

$$\int_0^{\Psi_1(u_0)} h(\Psi_2^{-1}(w)) \, dw = \int_0^{\Psi_2(1)} h(\Psi_2^{-1}(w)) \, dw = \int_0^1 h(\Psi_2^{-1}(\Psi_2(v))) \, d\Psi_2(v).$$

Since by Lemma 2.6.3 below, $\Psi_2^{-1}(\Psi_2(v)) = v$, $d\Psi_2(v)$-almost everywhere on $(0,1)$, we conclude that

$$\int_0^1 h(\Psi_2^{-1}(\Psi_2(v))) \, d\Psi_2(v) = \int_0^1 h(v) f_2(v) \, dv.$$

□

We complete this section with standard lemmas with their proofs, so that the present article is self-contained.

Lemma 2.6.3. *Let $I \subset \mathbb{R}$ be an interval, $F : I \to \mathbb{R}$ be a bounded and nondecreasing càdlàg function, $F(I)$ be the image of I by F and F^{-1} be the left continuous pseudo-inverse of F, that is*

$$F^{-1} : u \in F(I) \mapsto \inf\{r \in I \mid F(r) \geq u\}$$

Then for all $(x, u) \in I \times F(I)$, $F(x) \geq u \iff x \geq F^{-1}(u)$. Moreover, $F^{-1}(F(x)) = x$, $dF(x)$-almost everywhere on I.

Proof. Let $(x, u) \in I \times F(I)$. If $F(x) \geq u$, then by definition of the infimum, $x \geq F^{-1}(u)$. Conversely, if $x \geq F^{-1}(u)$, then let $(r_n)_{n\in\mathbb{N}} \in I^{\mathbb{N}}$ be a decreasing sequence converging to $F^{-1}(u)$. For all $n \in \mathbb{N}$, $F(r_n) \geq u$. By right-continuity of F, we get $F(F^{-1}(u)) \geq u$ for $n \to +\infty$. By monotonicity of F, we have $F(x) \geq F(F^{-1}(u)) \geq u$.

Let us now prove the second statement. Let $a = \inf F(I)$ and $b = \sup F(I)$. If $a = b$, then $dF(x)$ is the trivial measure on I so the statement is straightforward. Else, let $G : I \to [0, 1]$ be defined for all $x \in I$ by $G(x) = (F(x) - a)/(b - a)$ and let G^{-1} be its left-continuous pseudo-inverse. It is well known that for all $u \in (0, 1)$, $G^{-1}(G(G^{-1}(u))) = G^{-1}(u)$. So $G^{-1}(G(G^{-1}(U))) = G^{-1}(U)$, where U is a random variable uniformly distributed on $[0, 1]$. By the inverse transform sampling, it implies that $G^{-1}(G(x)) = x$, $dG(x)$-almost everywhere on I. For all $u \in F(I)$, we have $F^{-1}(u) = G^{-1}((u - a)/(b - a))$, hence $F^{-1}(F(x)) = G^{-1}(G(x)) = x$, $dG(x)$-almost everywhere on I. Since $dG(x) = \frac{1}{b-a} dF(x)$, $dG(x)$ and $dF(x)$ are equivalent, so $F^{-1}(F(x)) = x$, $dF(x)$-almost everywhere on I.

∎

Lemma 2.6.4. *Let $\mu \in \mathcal{P}(\mathbb{R})$. Then $F_\mu(x) > 0$ and $F_\mu(x_-) < 1$, $\mu(dx)$-almost everywhere on $\mathbb{R}$.*

Proof. If $\{x \in \mathbb{R} \mid F_\mu(x) = 0\}$ is nonempty, then it is an interval of the form $(-\infty, a]$ or $(-\infty, a)$, depending on whether $F_\mu(a) = 0$ or not. If $F_\mu(a) = 0$, then $\mu(\{x \in \mathbb{R} \mid F_\mu(x) = 0\}) = \mu((-\infty, a]) = F_\mu(a) = 0$. Else, since for all $x < a$, $F_\mu(x) = 0$, then $\mu(\{x \in \mathbb{R} \mid F_\mu(x) = 0\}) = \mu((-\infty, a)) = F_\mu(a_-) = 0$.

If $\{x \in \mathbb{R} \mid F_\mu(x_-) = 1\}$ is nonempty, then it is an interval of the form $[a, +\infty)$ or $(a, +\infty)$, depending whether $F_\mu(a_-) = 1$ or not. If $F_\mu(a_-) = 1$, then $\mu(\{x \in \mathbb{R} \mid F_\mu(x_-) = 1\}) = \mu([a, +\infty)) = 1 - F_\mu(a_-) = 0$. Else, since for all $x > a$, $F_\mu(x_-) = 1$, then $\mu(\{x \in \mathbb{R} \mid F_\mu(x_-) = 1\}) = \mu((a, +\infty)) = 1 - F_\mu(a) = 1 - \lim_{x\to a, x>a} F_\mu(x_-) = 0$, by right continuity of F_μ. ∎

Lemma 2.6.5. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$. Then μ is symmetric with mean $\alpha \in \mathbb{R}$, that is $(x - \alpha)_\sharp\mu(dx) = (\alpha - x)_\sharp\mu(dx)$ where $\sharp$ denotes the pushforward operation, iff*

$$F_\mu^{-1}(u_+) = 2\alpha - F_\mu^{-1}(1 - u),$$

for all $u \in (0, 1)$. In that case, $F_\mu^{-1}(u) = 2\alpha - F_\mu^{-1}(1 - u)$ for $u \in (0, 1)$ up to the at most countable set of discontinuities of F_μ^{-1}.

Proof. Let U be a random variable uniformly distributed on $[0,1]$. Then, by the inverse transform sampling, $F_\mu^{-1}(1-U) \sim \mu$, so $2\alpha - F_\mu^{-1}(1-U) \sim \mu$ since μ is symmetric with mean α. Since $u \mapsto 2\alpha - F_\mu^{-1}(1-u)$ is nondecreasing, then one can show that $2\alpha - F_\mu^{-1}(1-u) = F_\mu^{-1}(u)$, du-almost everywhere on $(0,1)$. Indeed, as shown in Lemma A.3 [4], for all $u, q \in (0,1)$ such that $q < u$, if $F_\mu^{-1}(u) < 2\alpha - F_\mu^{-1}(1-q)$, then

$$\begin{aligned}\mathbb{P}(2\alpha - F_\mu^{-1}(1-U) \leq F_\mu^{-1}(u)) &\leq \mathbb{P}(2\alpha - F_\mu^{-1}(1-U) < 2\alpha - F_\mu^{-1}(1-q)) \\ &\leq q < u \leq \mathbb{P}(F_\mu^{-1}(U) \leq F_\mu^{-1}(u)) = \mathbb{P}(2\alpha - F_\mu^{-1}(1-U) \\ &\leq F_\mu^{-1}(u)),\end{aligned}$$

which is contradictory, so $F_\mu^{-1}(u) \geq \sup_{q\in(0,u)}(2\alpha - F_\mu^{-1}(1-q))$. By symmetry, $2\alpha - F_\mu^{-1}(1-u) \geq \sup_{q\in(0,u)} F_\mu^{-1}(q) = F_\mu^{-1}(u)$ by left-continuity and monotonicity of F_μ^{-1}. Since F_μ^{-1} has an at most countable set of discontinuities, then for du-almost all $u \in (0,1)$, $2\alpha - F_\mu^{-1}(1-u) = \sup_{q\in(0,u)}(2\alpha - F_\mu^{-1}(1-q)) \leq F_\mu^{-1}(u) \leq 2\alpha - F_\mu^{-1}(1-u)$. Therefore, $2\alpha - F_\mu^{-1}(1-u) = F_\mu^{-1}(u_+)$, du-almost everywhere on $(0,1)$ and even everywhere on $(0,1)$ since both sides are right-continuous.

□

Lemma 2.6.6. *Let $\mu \in \mathcal{P}(\mathbb{R})$, let $X : \Omega \to \mathbb{R}$ be a random variable with distribution μ and let V be a random variable independent from X and uniformly distributed on $(0,1)$. Let $W : \Omega \to \mathbb{R}$ be the random variable defined by*

$$W = F_\mu(X_-) + V(F_\mu(X) - F_\mu(X_-)).$$

Then W is uniformly distributed on $(0,1)$, and $F_\mu^{-1}(W) = X$ almost surely.

Proof. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable and bounded function. Then

$$\begin{aligned}\mathbb{E}[h(W)] &= \mathbb{E}[h(F_\mu(X_-) + V(F_\mu(X) - F_\mu(X_-))] \\ &= \int_0^1 \int_{\mathbb{R}} h(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)))\, \mu(dx)\, dv \\ &= \int_0^1 \int_{\mathbb{R}} \mathbb{1}_{\{\mu(\{x\})=0\}} h(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)))\, \mu(dx)\, dv \\ &\quad + \int_0^1 \int_{\mathbb{R}} \mathbb{1}_{\{\mu(\{x\})>0\}} h(F_\mu(x_-) + v(F_\mu(x) - F_\mu(x_-)))\, \mu(dx)\, dv \\ &= \int_{\mathbb{R}} \mathbb{1}_{\{\mu(\{x\})=0\}} h(F_\mu(x))\, \mu(dx) + \sum_{x\in\mathbb{R}:\mu(\{x\})>0} \int_{F_\mu(x_-)}^{F_\mu(x)} h(v)\, dv \\ &= \int_0^1 \mathbb{1}_{\{\mu(\{F_\mu^{-1}(u)\})=0\}} h(F_\mu(F_\mu^{-1}(u)))\, du + \sum_{x\in\mathbb{R}:\mu(\{x\})>0} \int_{F_\mu(x_-)}^{F_\mu(x)} h(v)\, dv \\ &= \int_0^1 \mathbb{1}_{\{\mu(\{F_\mu^{-1}(u)\})=0\}} h(u)\, du + \sum_{x\in\mathbb{R}:\mu(\{x\})>0} \int_{F_\mu(x_-)}^{F_\mu(x)} h(v)\, dv,\end{aligned}$$

where we used for the last but one equality the inverse transform sampling, and for the last equality the fact that $F_\mu(F_\mu^{-1}(u)) = u$ if F_μ is continuous at $F_\mu^{-1}(u)$. One can easily see that for all $x \in \mathbb{R}$ and $u \in (0,1)$,

$$F_\mu(x_-) < u \leq F_\mu(x) \implies x = F_\mu^{-1}(u) \implies F_\mu(x_-) \leq u \leq F_\mu(x),$$

which implies

$$\bigcup_{x\in\mathbb{R}:\mu(\{x\})>0} (F_\mu(x_-), F_\mu(x)] \subset \{u \in (0,1) \mid \mu(\{F_\mu^{-1}(u)\}) > 0\} \subset \bigcup_{x\in\mathbb{R}:\mu(\{x\})>0} [F_\mu(x_-), F_\mu(x)],$$

so

$$\sum_{x\in\mathbb{R}:\mu(\{x\})>0} \int_{F_\mu(x_-)}^{F_\mu(x)} h(v)\,dv = \int_0^1 \mathbb{1}_{\{\mu(\{F_\mu^{-1}(u)\})>0\}} h(u)\,du.$$

Therefore, $\mathbb{E}[h(W)] = \int_0^1 \mathbb{1}_{\{\mu(\{F_\mu^{-1}(u)\})=0\}} h(u)\,du + \int_0^1 \mathbb{1}_{\{\mu(\{F_\mu^{-1}(u)\})>0\}} h(u)\,du = \int_0^1 h(u)\,du$. So W is uniformly distributed on $(0,1)$. Moreover, on $\{F_\mu(X_-) = F_\mu(X)\}$, $W = F_\mu(X)$ and by Lemma 2.6.3, $F_\mu^{-1}(W) = X$ almost surely. Since $V > 0$ a.s., on $\{F_\mu(X_-) < F_\mu(X)\}$, a.s., $F_\mu(X_-) < W \leq F_\mu(X)$ so $F_\mu^{-1}(W) = X$. $\square$

Chapter 3

Martingale Wasserstein inequality for probability measures in the convex order

Abstract

It is known since Chapter 2 that two one-dimensional probability measures in the convex order admit a martingale coupling with respect to which the integral of $|x-y|$ is smaller than twice their $\mathcal{W}_1$-distance (Wasserstein distance with index 1). We showed in 2 that replacing $|x-y|$ and $\mathcal{W}_1$ respectively with $|x-y|^\rho$ and $\mathcal{W}_\rho^\rho$ does not lead to a finite multiplicative constant. We show here that a finite constant is recovered when replacing $\mathcal{W}_\rho^\rho$ with the product of $\mathcal{W}_\rho$ times the centred ρ-th moment of the second marginal to the power $\rho-1$. Then we study the generalisation of this new stability inequality to higher dimension.

Keywords: Convex order, Martingale Optimal Transport, Wasserstein distance, Martingale couplings.

3.1 Introduction

For all $d \in \mathbb{N}^*$, $\rho \geq 1$ and μ, ν in the set $\mathcal{P}_\rho(\mathbb{R}^d)$ of probability measures on $\mathbb{R}^d$ with finite ρ-th moment, we define the Wasserstein distance with index ρ by

$$\mathcal{W}_\rho(\mu,\nu) = \left(\inf_{P \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\rho \, P(dx,dy) \right)^{1/\rho},$$

where $\Pi(\mu,\nu)$ denotes the set of couplings between μ and ν, that is

$$\Pi(\mu,\nu) = \{P \in \mathcal{P}_1(\mathbb{R}^d \times \mathbb{R}^d) \mid \forall A \in \mathcal{B}(\mathbb{R}^d),\ P(A \times \mathbb{R}^d) = \mu(A) \text{ and } P(\mathbb{R}^d \times A) = \nu(A)\}.$$

Let $\Pi^{\mathrm{M}}(\mu,\nu)$ be the set of martingale couplings between μ and ν, that is

$$\Pi^{\mathrm{M}}(\mu,\nu) = \left\{ M \in \Pi(\mu,\nu) \mid \mu(dx)\text{-a.e.},\ \int_{\mathbb{R}^d} y\, m(x,dy) = x \right\},$$

where for all $M \in \Pi(\mu,\nu)$, $(m(x,dy))_{x\in\mathbb{R}}$ denotes a regular conditional probability distribution of M with respect to μ. The celebrated Strassen theorem [183] ensures that if $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$, then $\Pi^{\mathrm{M}}(\mu,\nu) \neq \emptyset$ iff μ and ν are in the convex order. We recall that two probability measures $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$ are in the convex order, and denote $\mu \leq_{cx} \nu$, if

$$\int_{\mathbb{R}^d} f(x)\,\mu(dx) \leq \int_{\mathbb{R}^d} f(y)\,\nu(dy), \tag{3.1.1}$$

for any convex function $f : \mathbb{R}^d \to \mathbb{R}$. For all $\rho \geq 1$ and $\mu,\nu \in \mathcal{P}_\rho(\mathbb{R}^d)$, we define $\mathcal{M}_\rho(\mu,\nu)$ by

$$\mathcal{M}_\rho(\mu,\nu) = \left(\inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^\rho \, M(dx,dy) \right)^{1/\rho}.$$

Notice that when $\mathbb{R}^d$ is endowed with the Euclidean norm, the martingale property $\int_{\mathbb{R}^d} \langle x,y\rangle\, M(dx,dy) = \int_{\mathbb{R}^d} |x|^2\,\mu(dx)$ valid for any martingale coupling $M \in \Pi^{\mathrm{M}}(\mu,\nu)$ yields the remarkable property that $\mathcal{M}_2(\mu,\nu)$ depends only on the marginals, namely

$$\mathcal{M}_2^2(\mu,\nu) = \int_{\mathbb{R}^d} |y|^2\,\nu(dy) - \int_{\mathbb{R}^d} |x|^2\,\mu(dx). \tag{3.1.2}$$

It was shown in Chapter 2 that if μ and ν are in the convex order and close to each other, then there exists a martingale coupling which expresses this proximity:

$$\forall \mu,\nu \in \mathcal{P}_1(\mathbb{R}) \text{ such that } \mu \leq_{cx} \nu, \quad \mathcal{M}_1(\mu,\nu) \leq 2\mathcal{W}_1(\mu,\nu), \tag{3.1.3}$$

where the constant 2 is sharp. This stability inequality was proved by exhibiting for all $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ in the convex order a subset $\mathcal{Q}$ of two dimensional probability measures on the unit square and a family $(M^Q)_{Q\in\mathcal{Q}}$ of martingale couplings between μ and ν such that for all $Q \in \mathcal{Q}$, $\int_{\mathbb{R}\times\mathbb{R}} |y-x|\,M^Q(dx,dy) \leq 2\mathcal{W}_1(\mu,\nu)$. A particular martingale coupling stands out from the latter family: the so called inverse transform martingale coupling. This coupling is explicit in terms of the cumulative distribution functions of the marginal distributions and their left-continuous generalised inverses. It is therefore more explicit than the left-curtain (and right-curtain) coupling introduced by Beiglböck and Juillet [25] and which under the condition that ν has no atoms and the set of local maximal values of $F_\nu - F_\mu$ is finite can be explicited according to Henry-Labordère and Touzi [106] by solving two coupled ordinary differential equations starting from each right-most local maximiser. Many properties of the inverse transform martingale coupling and the family from which it derives are discussed in Chapter 2. In this paper, we prove a more general stability inequality:

$$\forall \rho \geq 1, \quad \exists C \in \mathbb{R}_+^*, \quad \forall \mu, \nu \in \mathcal{P}_\rho(\mathbb{R}) \text{ such that } \mu \leq_{cx} \nu, \quad \mathcal{M}_\rho^\rho(\mu, \nu) \leq C\mathcal{W}_\rho(\mu, \nu)\sigma_\rho^{\rho-1}(\nu), \tag{3.1.4}$$

where the centred moment $\sigma_\rho(\eta)$ of order ρ of $\eta \in \mathcal{P}_\rho(\mathbb{R})$ is defined by

$$\sigma_\rho(\eta) = \inf_{c \in \mathbb{R}} \left(\int_{\mathbb{R}} |y - c|^\rho \, \eta(dy) \right)^{1/\rho}.$$

For all $\rho \geq 1$, let C_ρ denote the optimal constant C in (3.1.4), that is

$$C_\rho = \inf \left\{ C > 0 \mid \forall \mu, \nu \in \mathcal{P}_\rho(\mathbb{R}) \text{ such that } \mu \leq_{cx} \nu, \ \mathcal{M}_\rho^\rho(\mu, \nu) \leq C\mathcal{W}_\rho(\mu, \nu)\sigma_\rho^{\rho-1}(\nu) \right\}. \tag{3.1.5}$$

One readily notices that (3.1.3) is a particular case of (3.1.4) for $\rho = 1$ and $C = 2$. Moreover, since 2 is sharp for (3.1.3), we have $C_1 = 2$. One can also obtain that $C_2 = 2$ when $\mathbb{R}^d$ is endowed with the Euclidean norm with simple arguments which hold in general dimension. Indeed, let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ be such that $\mu \leq_{cx} \nu$ and $\pi \in \Pi(\mu, \nu)$ be optimal for $\mathcal{W}_2(\mu, \nu)$. Then by (3.1.2), the martingale property, the Cauchy-Schwarz inequality, Jensen's inequality and the definition of the convex order, for all $c \in \mathbb{R}$, we have

$$\begin{aligned}
\mathcal{M}_2^2(\mu, \nu) &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(|y|^2 - |x|^2 \right) \pi(dx, dy) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle y - x, y + x \rangle \, \pi(dx, dy) - 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle y - x, c \rangle \, \pi(dx, dy) \\
&= \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle y - x, y - c + x - c \rangle \, \pi(dx, dy) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x||y - c + x - c| \, \pi(dx, dy) \\
&\leq \sqrt{\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - x|^2 \, \pi(dx, dy)} \sqrt{\int_{\mathbb{R}^d \times \mathbb{R}^d} |y - c + x - c|^2 \, \pi(dx, dy)} \\
&\leq \mathcal{W}_2(\mu, \nu) \sqrt{2 \int_{\mathbb{R}^d \times \mathbb{R}^d} (|y - c|^2 + |x - c|^2) \, \pi(dx, dy)} \\
&\leq \mathcal{W}_2(\mu, \nu) \sqrt{4 \int_{\mathbb{R}^d \times \mathbb{R}^d} |y - c|^2 \, \pi(dx, dy)} = 2\mathcal{W}_2(\mu, \nu) \sqrt{\int_{\mathbb{R}^d} |y - c|^2 \, \nu(dy)}.
\end{aligned} \tag{3.1.6}$$

By bias-variance decomposition, one can see that $\sigma_2(\nu)$ is the standard deviation of ν, so for c equal to the mean of ν, (3.1.6) writes $\mathcal{M}_2^2(\mu, \nu) \leq 2\mathcal{W}_2(\mu, \nu)\sigma_2(\nu)$, hence $C_2 \leq 2$. On the other hand, for all $n \in \mathbb{N}^*$, let μ_n be the centred Gaussian distribution with variance n^2.

Then we get that $\mathcal{M}_2(\mu_n, \mu_{n+1}) = \sqrt{2n+1}$. It is well known (see for instance Remark 2.19 (ii) Chapter 2 [190]) that for all $\rho \geq 1$ and $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$,

$$\mathcal{W}_\rho(\mu, \nu) = \left(\int_0^1 \left| F_\mu^{-1}(u) - F_\nu^{-1}(u) \right|^\rho du \right)^{1/\rho}, \tag{3.1.7}$$

where we denote by $F_\eta(x) = \eta((-\infty, x]), x \in \mathbb{R}$ and $F_\eta^{-1}(u) = \inf\{x \in \mathbb{R} \mid F_\eta(x) \geq u\}, u \in (0,1)$, the cumulative distribution function and the quantile function of a probability measure η on $\mathbb{R}$. Therefore, for $G \sim \mathcal{N}_1(0,1)$, $\mathcal{W}_2(\mu_n, \mu_{n+1}) = (\int_0^1 |nF_{\mu_1}^{-1}(u) - (n+1)F_{\mu_1}^{-1}(u)|^2\, du)^{1/2} = \mathbb{E}[|G|^2]^{1/2} = 1$. We deduce that for all $n \in \mathbb{N}^*$, $2n+1 \leq C_2\sqrt{(n+1)^2}$, which implies for $n \to +\infty$ that $C_2 \geq 2$. Hence $C_2 = 2$.

The generalisation of the stability inequality (3.1.3) is motivated by the resolution of the Martingale Optimal Transport (MOT) problem introduced by Beiglböck, Henry-Labordère and Penkner [23] in a discrete time setting, and Galichon, Henry-Labordère and Touzi [84] in a continuous time setting. For adaptations of celebrated results on classical optimal transport theory to the MOT problem, we refer to Beiglböck and Juillet [25], Henry-Labordère, Tan and Touzi [105] and Henry- Labordère and Touzi [84]. On duality, we refer to Beiglböck, Nutz and Touzi [27], Beiglböck, Lim and Obłój [26] and De March [68]. We also refer to De March [67] and De March and Touzi [69] for the multi-dimensional case.

About the numerical resolution of the MOT problem, one can look at Alfonsi, Corbetta and Jourdain [3, 4], De March [66], Guo and Obłój [96] and Henry-Labordère [104]. When μ and ν are finitely supported, then the MOT problem amounts to linear programming. In the general case, once the MOT problem is discretised by approximating μ and ν by probability measures with finite support and in the convex order, Alfonsi, Corbetta and Jourdain raised the question of the convergence of the discrete optimal cost towards the continuous one. Partial results were first brought by Guo and Obłój [96] and the stability of left-curtain couplings obtained by Juillet [112]. More recently, Backhoff-Veraguas and Pammer [20] and Wiesel [194] independently gave a positive answer in dimension one under mild regularity assumption. The generalisation of our stability inequality to higher dimension is a step forward in giving a positive answer of the stability of the MOT problem in higher dimension. Indeed, it yields the stability for continuous costs which satisfy a growth constraint when the second marginal is increased in the convex order. More precisely, let $\rho \geq 1$, $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$ be such that $\mu \leq_{cx} \nu$ and $(\nu_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R}^d)^{\mathbb{N}}$ be such that $\nu \leq_{cx} \nu_n$ for all $n \in \mathbb{N}$ and ν_n converges to ν in $\mathcal{W}_\rho$ as $n \to +\infty$. Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be continuous and growing at most as the ρ-th power of its variables, i.e. $|c(x,y)| \leq K(1 + |x|^\rho + |y|^\rho)$ for all $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d$ and a certain $K \in \mathbb{R}_+$. It is well known that any sequence $(\pi_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi^{\mathrm{M}}(\mu, \nu_n)$ is tight and has all its accumulation points with respect to the weak convergence topology in $\Pi^{\mathrm{M}}(\mu, \nu)$. Then one can readily derive the first inequality

$$V(\mu, \nu) := \inf_{\pi \in \Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,y)\, \pi(dx, dy) \leq \liminf_{n\to+\infty} V(\mu, \nu_n).$$

On the other hand, for any martingale coupling $\pi_n \in \Pi^M(\mu, \nu_n)$ we have

$$\limsup_{n\to+\infty} V(\mu, \nu_n) \leq \limsup_{n\to+\infty} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x,z)\, \pi_n(dx, dz). \tag{3.1.8}$$

Recall that a sequence $(\tau_n)_{n\in\mathbb{N}}$ of probability measures on $\mathbb{R}^d$ converges to τ in $\mathcal{W}_\rho$ iff $(\int_{\mathbb{R}^d} f(x)\,\tau_n(dx))_{n\in\mathbb{N}}$ converges to $\int_{\mathbb{R}^d} f(x)\,\tau(dx)$ for any real-valued continuous map $f:\mathbb{R}^d\to\mathbb{R}$ which grows at most as the ρ-th power of its variables. Hence it suffices to find π_n such that $\mathcal{W}_\rho(\pi,\pi_n)$ vanishes as $n\to+\infty$ to conclude that the right-hand side in (3.1.8) and therefore $V(\mu,\nu_n)$ converges to $V(\mu,\nu)$. By Lemma 3.4.1 below there exist for all $n\in\mathbb{N}$ a martingale coupling $M_n\in\Pi^{\mathrm{M}}(\nu,\nu_n)$ optimal for $\mathcal{M}_\rho(\nu,\nu_n)$ and a martingale coupling $\pi\in\Pi^M(\mu,\nu)$ optimal for $V(\mu,\nu)$. Let $(m_n(y,dy'))_{y\in\mathbb{R}^d}$ be a regular conditional probability distribution of M_n with respect to ν and $\pi_n(dx',dy')=\int_{y\in\mathbb{R}^d} m_n(y,dy')\,\pi(dx',dy)\in\Pi^{\mathrm{M}}(\mu,\nu_n)$. Since $\pi(dx,dy)\,\delta_x(dx')\,m_n(y,dy')$ is a coupling between π and π_n, we have

$$\begin{aligned}\mathcal{W}_\rho^\rho(\pi,\pi_n)&\le\int_{\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R}^d}|y-y'|^\rho\,\pi(dx,dy)\,m_n(y,dy')=\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-y'|^\rho\,M_n(dy,dy')\\&=\mathcal{M}_\rho^\rho(\nu,\nu_n)\le C_\rho\mathcal{W}_\rho(\nu,\nu_n)\sigma_\rho^{\rho-1}(\nu_n).\end{aligned}$$

If C_ρ is finite, then by convergence of $(\nu_n)_{n\in\mathbb{N}}$ in $\mathcal{W}_\rho$, the sequences $(\mathcal{W}_\rho(\nu,\nu_n))_{n\in\mathbb{N}}$ and $(\sigma_\rho^{\rho-1}(\nu_n))_{n\in\mathbb{N}}$ are bounded, hence $(\pi_n)_{n\in\mathbb{N}}$ converges to π in $\mathcal{W}_\rho$.

We present our main result in Section 3.2, namely the new one-dimensional stability inequality which extends the previous one, see Chapter 2, to any index $\rho\ge1$. Then Section 3.3 extends the latter inequality in higher dimension as soon as the marginals satisfy specific conditions. Finally Section 3.4 is devoted to the proof of some required lemmas.

Acknowledgements We thank Nicolas Juillet for his remarks on a preliminary version of this paper.

3.2 A new stability inequality

We come back a moment on the family $(M^Q)_{Q\in\mathcal{Q}}$ parametrised by $\mathcal{Q}$ mentioned in the introduction since it will have particular significance in the present section. We briefly recall the construction and main properties, see Chapter 2 for an extensive study. Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be such that $\mu\le_{cx}\nu$ and $\mu\ne\nu$. For $u\in[0,1]$ we define

$$\Psi_+(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\quad\text{and}\quad\Psi_-(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv,\tag{3.2.1}$$

with respective left continuous generalised inverses Ψ_+^{-1} and Ψ_-^{-1}. We then define $\mathcal{Q}$ as the set of probability measures on $(0,1)^2$ with first marginal $\frac{1}{\Psi_+(1)}d\Psi_+$, second marginal $\frac{1}{\Psi_+(1)}d\Psi_-$ and such that $u<v$ for $Q(du,dv)$-almost every $(u,v)\in(0,1)^2$. Since $d\,\Psi_+$ and $d\,\Psi_-$ are concentrated on two disjoint Borel sets, there exists for each $Q\in\mathcal{Q}$ a probability kernel $(\pi^Q(u,dv))_{u\in(0,1)}$ such that

$$Q(du,dv)=\frac{1}{\Psi_+(1)}d\Psi_+(u)\,\pi^Q(u,dv)=\frac{1}{\Psi_+(1)}d\Psi_-(v)\,\pi^Q(v,du),$$

and we exhibit a probability kernel $(\widetilde{m}^Q(u,dy))_{u\in(0,1)}$ which satisfies for du-almost all $u \in (0,1)$ such that $F_\mu^{-1}(u) \neq F_\nu^{-1}(u)$

$$\widetilde{m}^Q(u,dy) = \int_{v\in(0,1)} \left(\frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(v)}(dy) + \frac{F_\nu^{-1}(v) - F_\mu^{-1}(u)}{F_\nu^{-1}(v) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(u)}(dy) \right) \pi^Q(u,dv), \tag{3.2.2}$$

and $\widetilde{m}^Q(u,dy) = \delta_{F_\nu^{-1}(u)}(dy)$ for all $u \in (0,1)$ such that $F_\mu^{-1}(u) = F_\nu^{-1}(u)$. Then the measure

$$M^Q(dx,dy) = \int_0^1 \delta_{F_\mu^{-1}(u)}(dx)\, \widetilde{m}^Q(u,dy)\, du \tag{3.2.3}$$

is a martingale coupling between μ and ν which satisfies $\int_{\mathbb{R}\times\mathbb{R}} |y-x|\, M^Q(dx,dy) \leq 2\mathcal{W}_1(\mu,\nu)$. The latter fact is based on the property that for du-almost all $u \in (0,1)$,

$$\int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\, \widetilde{m}^Q(u,dy) = |F_\mu^{-1}(u) - F_\nu^{-1}(u)|, \tag{3.2.4}$$

as showed by Proposition 2.2.18 and its proof.

We also recall some standard results about cumulative distribution functions and quantile functions since they will prove very handy one-dimensional tools. Proofs can be found for instance in Section 2.6. For any probability measure η on $\mathbb{R}$:

(1) F_η, resp. F_η^{-1}, is right continuous, resp. left continuous, and nondecreasing;

(2) For all $(x,u) \in \mathbb{R} \times (0,1)$,

$$F_\eta^{-1}(u) \leq x \iff u \leq F_\eta(x), \tag{3.2.5}$$

which implies

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{3.2.6}$$

$$\text{and} \quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{3.2.7}$$

(3) For $\mu(dx)$-almost every $x \in \mathbb{R}$,

$$0 < F_\eta(x), \quad F_\eta(x-) < 1 \quad \text{and} \quad F_\eta^{-1}(F_\eta(x)) = x; \tag{3.2.8}$$

(4) Denoting by $\lambda_{(0,1)}$, resp. $\lambda_{(0,1)^2}$, the Lebesgue measure on $(0,1)$, resp. $(0,1)^2$, we have

$$\Big((u,v) \mapsto F_\mu(F_\mu^{-1}(u)-) + v\mu(\{F_\mu^{-1}(u)\})\Big)_\sharp \lambda_{(0,1)^2} = \lambda_{(0,1)}, \tag{3.2.9}$$

where $\sharp$ denotes the pushforward operation.

(5) The image of the Lebesgue measure on $(0,1)$ by F_η^{-1} is η.

The property (5) is referred to as inverse transform sampling.

We can now state and prove our main result. For all $\rho \geq 1$ and $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ in the convex order, we provide an estimate of the martingale Wasserstein function $\mathcal{M}_\rho(\mu,\nu)$ in terms of the Wasserstein distance $\mathcal{W}_p(\mu,\nu)$ and the centred ρ-th moment of ν.

Proposition 3.2.1. *Let $\rho \geq 1$ and $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$. Then*

(i) For all $Q \in \mathcal{Q}$, the martingale coupling $M^Q \in \Pi^{\mathrm{M}}(\mu,\nu)$ defined by (3.2.3) *satisfies*

$$\mathcal{M}_\rho^\rho(\mu,\nu) \leq K_\rho \mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu), \tag{3.2.10}$$

where

$$K_\rho = \inf\left\{2^{\rho-1}\gamma_1 + 2(2^{\rho-2} \vee 1)\gamma_2 \mid (\gamma_1,\gamma_2) \in \mathbb{R}_+^2 \text{ and } \forall x \in \mathbb{R}_+,\ \frac{x+x^\rho}{1+x} \leq \gamma_1 + \gamma_2 x^{\rho-1}\right\}. \tag{3.2.11}$$

(ii) The constant C_ρ defined by (3.1.5) *satisfies*

$$2^{\rho-1} \sup_{x\in(1,+\infty)} \frac{x+x^\rho}{(1+x)^\rho} \leq C_\rho \leq K_\rho \leq \begin{cases} \left(1+\sup_{x\in[0,1]} \frac{x-x^{\rho-1}}{1+x}\right) 2^{\rho-1} & \text{for } \rho > 2; \\ 2 + \left(2\sup_{x\in[1,+\infty)} \frac{x-x^{\rho-1}}{x^{\rho-1}(1+x)}\right) \wedge 2^{\rho-1} & \text{for } 1 < \rho < 2, \end{cases} \tag{3.2.12}$$

and $K_1 = K_2 = 2$.

(iii) $\mathcal{W}_\rho(\mu,\nu)$ and $\sigma_\rho(\nu)$ have the right exponent in (3.1.4) *in the following sense:*

$$\forall \rho > 1, \quad \forall s \in (1,\rho], \qquad \sup_{\substack{\mu,\nu\in\mathcal{P}_\rho(\mathbb{R})\\ \mu\leq_{cx}\nu}} \frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho^s(\mu,\nu)\sigma_\rho^{\rho-s}(\nu)} = +\infty. \tag{3.2.13}$$

Remark 3.2.2. (i) We easily check that $\sup_{x\in(1,+\infty)} \frac{x+x^1}{(1+x)^1} = 2$. Moreover, for all $a > 0$ one can easily derive the inequalities $1+\rho a \leq (1+a)^\rho \leq 1+\rho a(1+a)^{\rho-1}$. Equivalently, for all $x > 0$ we have $x^\rho + \rho x^{\rho-1} \leq (1+x)^\rho \leq x^\rho + \rho(1+x)^{\rho-1}$. Then for $\rho \geq 2$, the inequality $x^\rho + x \leq x^\rho + \rho x^{\rho-1}$ valid for all $x > 1$ implies that $\sup_{x\in(1,+\infty)} \frac{x+x^\rho}{(1+x)^\rho} = 1$. Moreover, for $\rho \in (1,2)$ we find for $x > 1$ large enough, namely such that $x > \rho(1+x)^{\rho-1}$, that $x^\rho + \rho(1+x)^{\rho-1} < x^\rho + x$, hence $\sup_{x\in(1,+\infty)} \frac{x+x^\rho}{(1+x)^\rho} > 1$.

By (3.2.12) applied with $\rho = 1$ and $\rho = 2$ we then recover $C_1 = C_2 = 2$. Moreover, we also easily check that $\sup_{x\in[0,1]} \frac{x-x^{\rho-1}}{1+x} \leq \frac{1}{2}$, hence by (3.2.12) again we find for all $\rho > 2$ that $2^{\rho-1} \leq C_\rho \leq \frac{3}{2}2^{\rho-1}$.

(ii) We show (see (3.2.20) below) that $\mathcal{W}_\rho(\mu,\nu) \leq 2\sigma_\rho(\nu)$, so that for $s \in [0,1]$,

$$\mathcal{M}_\rho^\rho(\mu,\nu) \leq 2^{1-s} C_\rho \mathcal{W}_\rho^s(\mu,\nu)\sigma_\rho^{\rho-s}(\nu).$$

Proof of Proposition 3.2.1. Let us prove (iii) first. One can readily show (see for instance (2.2.25)) that for $a, b \in \mathbb{R}$ such that $0 < a < b$,

$$H = \frac{(b+a)}{4b}\delta_{(-a,-b)} + \frac{(b-a)}{4b}\delta_{(-a,b)} + \frac{(b+a)}{4b}\delta_{(a,b)} + \frac{(b-a)}{4b}\delta_{(a,-b)} \tag{3.2.14}$$

is the only martingale coupling between $\mu = \frac{1}{2}\delta_{-a} + \frac{1}{2}\delta_a$ and $\nu = \frac{1}{2}\delta_{-b} + \frac{1}{2}\delta_b$. Consequenly, for $\rho \geq 1$ we trivially have

$$\mathcal{M}_\rho^\rho(\mu,\nu) = \int_{\mathbb{R}\times\mathbb{R}} |x-y|^\rho \, H(dx,dy) = \frac{1}{2b}\left((a+b)(b-a)^\rho + (b-a)(a+b)^\rho\right).$$

On the other hand, since $\mathcal{W}_\rho(\mu,\nu) = \left(\int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)| \, du\right)^{1/\rho}$ (see for instance Remark 2.19 (ii) Chapter 2 [190]),

$$\mathcal{W}_\rho(\mu,\nu) = \left(\int_0^{1/2} |-a-(-b)|^\rho \, du + \int_{1/2}^1 |a-b|^\rho \, du\right)^{1/\rho} = b-a.$$

Moreover, for all $c \in \mathbb{R}$, $\int_{\mathbb{R}} |y-c|^\rho \, \nu(dy) = \frac{1}{2}(|b-c|^\rho + |b+c|^\rho)$, which attains its infimum for $c = 0$, hence $\sigma_\rho(\nu) = b$. So for all $s \in [1,\rho]$, we have

$$\frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho^s(\mu,\nu)\sigma_\rho(\nu)^{\rho-s}} = \frac{1}{2b^{\rho+1-s}}\left((a+b)(b-a)^{\rho-s} + (a+b)^\rho(b-a)^{1-s}\right) \geq \frac{(a+b)^\rho(b-a)^{1-s}}{2b^{\rho+1-s}}, \tag{3.2.15}$$

which tends to $+\infty$ as b tends to a as soon as $\rho > 1$ and $s \in (1,\rho]$, which proves (iii). Furthermore, (3.2.15) applied with $s = 1$, $a = 1$ and $b > 1$ yields

$$\frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho(\mu,\nu)\sigma_\rho(\nu)^{\rho-1}} = \frac{(1+b)(b-1)^{\rho-1} + (1+b)^\rho}{2b^\rho}.$$

In particular for $b = \frac{x+1}{x-1}$ where x denotes any real number in $(1,+\infty)$, the latter equality writes

$$\frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho(\mu,\nu)\sigma_\rho(\nu)^{\rho-1}} = 2^{\rho-1}\frac{x+x^\rho}{(1+x)^\rho},$$

which proves the lower bound in (3.2.12) by taking the supremum over all $x \in (1,+\infty)$ in the right-hand side. Note that considering more general measures μ and ν in the convex order, each concentrated on two atoms, does not yield a greater lower bound.

We now show (i). Let $Q \in \mathcal{Q}$. Since the probability measure M^Q defined by (3.2.3) belongs to $\Pi^{\mathrm{M}}(\mu,\nu)$, we have by definition of $\mathcal{M}_\rho(\mu,\nu)$ and the definition (3.2.2) of $\widetilde{m}^Q$ that

$$\begin{aligned}
\mathcal{M}_\rho^\rho(\mu,\nu) &\leq \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho \, M^Q(dx,dy) = \int_{(0,1)\times\mathbb{R}} |y - F_\mu^{-1}(u)|^\rho \, du \, \widetilde{m}^Q(u,dy) \\
&= \int_{(0,1)^2} |F_\nu^{-1}(v) - F_\mu^{-1}(u)|^\rho \frac{|F_\mu^{-1}(u) - F_\nu^{-1}(u)|}{|F_\nu^{-1}(v) - F_\nu^{-1}(u)|} \, du \, \pi^Q(u,dv) \\
&\quad + \int_{(0,1)^2} |F_\nu^{-1}(u) - F_\mu^{-1}(u)|^\rho \frac{|F_\nu^{-1}(v) - F_\mu^{-1}(u)|}{|F_\nu^{-1}(v) - F_\nu^{-1}(u)|} \, du \, \pi^Q(u,dv).
\end{aligned} \tag{3.2.16}$$

By Lemma 2.2.5 we have $du\,\pi^Q(u,dv)$-almost everywhere

$$|F_\nu^{-1}(v)-F_\mu^{-1}(u)|+|F_\mu^{-1}(u)-F_\nu^{-1}(u)|=|F_\nu^{-1}(v)-F_\nu^{-1}(u)| \quad \text{and} \quad |F_\nu^{-1}(v)-F_\mu^{-1}(u)|\neq 0. \tag{3.2.17}$$

Let $\gamma_1,\gamma_2\geq 0$ be such that for all $x\in\mathbb{R}_+$, $\frac{x+x^\rho}{1+x}\leq\gamma_1+\gamma_2x^{\rho-1}$. Therefore, for all $(a,b)\in\mathbb{R}_+\times\mathbb{R}_+^*$ we have

$$\frac{a^\rho b+ab^\rho}{a+b}\leq\gamma_1a^\rho+\gamma_2ab^{\rho-1}. \tag{3.2.18}$$

By (3.2.16), (3.2.17) and (3.2.18) with $(a,b)=(|F_\nu^{-1}(u)-F_\mu^{-1}(u)|,|F_\nu^{-1}(v)-F_\mu^{-1}(u)|)$ for $du\,\pi^Q(u,dv)$-almost all $(u,v)\in(0,1)$, we get

$$\begin{aligned}
&\mathcal{M}_\rho^\rho(\mu,\nu)\\
&\leq\gamma_1\int_{(0,1)}|F_\nu^{-1}(u)-F_\mu^{-1}(u)|^\rho\,du+\gamma_2\int_{(0,1)}|F_\nu^{-1}(v)-F_\nu^{-1}(u)|^{\rho-1}|F_\mu^{-1}(u)-F_\nu^{-1}(u)|\,du\,\pi^Q(u,dv)\\
&=\gamma_1\mathcal{W}_\rho^\rho(\mu,\nu)+\gamma_2\int_{(0,1)}|F_\nu^{-1}(v)-F_\nu^{-1}(u)|^{\rho-1}|F_\mu^{-1}(u)-F_\nu^{-1}(u)|\,du\,\pi^Q(u,dv).
\end{aligned} \tag{3.2.19}$$

Let $c\in\mathbb{R}$. On the one hand, the inverse transform sampling and the definition (3.1.1) of the convex order applied with $x\mapsto|x-c|^\rho$ yield

$$\begin{aligned}
&\mathcal{W}_\rho^\rho(\mu,\nu)\\
&=\int_{(0,1)}|F_\nu^{-1}(u)-F_\mu^{-1}(u)|^\rho\,du\leq 2^{\rho-1}\left(\int_{(0,1)}|F_\nu^{-1}(u)-c|^\rho\,du+\int_{(0,1)}|F_\mu^{-1}(u)-c|^\rho\,du\right)\\
&=2^{\rho-1}\left(\int_\mathbb{R}|y-c|^\rho\,\nu(dy)+\int_\mathbb{R}|x-c|^\rho\,\mu(dx)\right)\leq 2^\rho\int_\mathbb{R}|y-c|^\rho\,\nu(dy).
\end{aligned} \tag{3.2.20}$$

We deduce that

$$\mathcal{W}_\rho^\rho(\mu,\nu)=\mathcal{W}_\rho(\mu,\nu)\mathcal{W}_\rho^{\rho-1}(\mu,\nu)\leq\mathcal{W}_\rho(\mu,\nu)\times 2^{\rho-1}\left(\int_\mathbb{R}|y-c|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}. \tag{3.2.21}$$

On the other hand we have

$$\begin{aligned}
&\int_{(0,1)}|F_\nu^{-1}(v)-F_\nu^{-1}(u)|^{\rho-1}|F_\mu^{-1}(u)-F_\nu^{-1}(u)|\,du\,\pi^Q(u,dv)\\
&=\int_{(0,1)^2}|F_\nu^{-1}(v)-F_\nu^{-1}(u)|^{\rho-1}(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv)\\
&\quad+\int_{(0,1)^2}|F_\nu^{-1}(v)-F_\nu^{-1}(u)|^{\rho-1}(F_\mu^{-1}-F_\nu^{-1})^-(u)\,du\,\pi^Q(u,dv).
\end{aligned} \tag{3.2.22}$$

Using the inequality $|x-y|^{\rho-1}\leq(2^{\rho-2}\vee 1)(|x|^{\rho-1}+|y|^{\rho-1})$ valid for all $(x,y)\in\mathbb{R}$ and the fact that $(F_\mu^{-1}-F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv)=Q(du,dv)=(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv\,\pi^Q(v,du)$, we get

$$
\begin{aligned}
&\int_{(0,1)^2} |F_\nu^{-1}(v) - F_\nu^{-1}(u)|^{\rho-1}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv) \\
&\le (2^{\rho-2} \vee 1)\left(\int_{(0,1)^2} |F_\nu^{-1}(v) - c|^{\rho-1}\,(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv)\right. \\
&\qquad\qquad \left. + \int_{(0,1)^2} |F_\nu^{-1}(u) - c|^{\rho-1}\,(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv)\right) \\
&= (2^{\rho-2} \vee 1)\left(\int_{(0,1)^2} |F_\nu^{-1}(u) - c|^{\rho-1}(F_\mu^{-1} - F_\nu^{-1})^-(u)\,du\,\pi^Q(u,dv)\right. \\
&\qquad\qquad \left. + \int_{(0,1)^2} |F_\nu^{-1}(u) - c|^{\rho-1}(F_\mu^{-1} - F_\nu^{-1})^+(u)\,du\,\pi^Q(u,dv)\right) \\
&= (2^{\rho-2} \vee 1)\int_{(0,1)} |F_\nu^{-1}(u) - c|^{\rho-1}|F_\mu^{-1}(u) - F_\nu^{-1}(u)|\,du.
\end{aligned}
\tag{3.2.23}
$$

Similarly, we have

$$
\begin{aligned}
&\int_{(0,1)^2} |F_\nu^{-1}(v) - F_\nu^{-1}(u)|^{\rho-1}(F_\mu^{-1} - F_\nu^{-1})^-(u)\,du\,\pi^Q(u,dv) \\
&\le (2^{\rho-2} \vee 1)\int_{(0,1)} |F_\nu^{-1}(u) - c|^{\rho-1}|F_\mu^{-1}(u) - F_\nu^{-1}(u)|\,du.
\end{aligned}
\tag{3.2.24}
$$

Plugging (3.2.23) and (3.2.24) in (3.2.22) for the first inequality, using Hölder's inequality for the second inequality and the inverse transform sampling for the equality, we have

$$
\begin{aligned}
&\int_{(0,1)} |F_\nu^{-1}(v) - F_\nu^{-1}(u)|^{\rho-1}|F_\mu^{-1}(u) - F_\nu^{-1}(u)|\,du\,\pi^Q(u,dv) \\
&\le 2(2^{\rho-2} \vee 1)\int_{(0,1)} |F_\nu^{-1}(u) - c|^{\rho-1}|F_\mu^{-1}(u) - F_\nu^{-1}(u)|\,du \\
&\le 2(2^{\rho-2} \vee 1)\left(\int_{(0,1)} |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^\rho\,du\right)^{1/\rho}\left(\int_{(0,1)} |F_\nu^{-1}(u) - c|^\rho\,du\right)^{(\rho-1)/\rho} \\
&= 2(2^{\rho-2} \vee 1)\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}} |y-c|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}.
\end{aligned}
$$

The latter inequality and (3.2.21) plugged in (3.2.19) then yields

$$
\mathcal{M}_\rho^\rho(\mu,\nu) \le (2^{\rho-1}\gamma_1 + 2(2^{\rho-2} \vee 1)\gamma_2)\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}} |y-c|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}.
$$

By taking in the right-hand side the infimum over all $(\gamma_1,\gamma_2) \in \mathbb{R}_+ \times \mathbb{R}_+$ such that for all $x \in \mathbb{R}_+$, $\frac{x+x^\rho}{1+x} \le \gamma_1 + \gamma_2 x^{\rho-1}$ and over all $c \in \mathbb{R}$, we deduce that

$$
\mathcal{M}_\rho^\rho(\mu,\nu) \le K_\rho \mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu).
$$

To complete the proof, it remains to prove the estimate of K_ρ in (ii). For all $x \in \mathbb{R}_+$ we have $\frac{x+x^1}{1+x} \leq 2 + 0x^0$ and $\frac{x+x^2}{1+x} = 0 + x^1$, hence $K_1 \leq 2$ and $K_2 \leq 2$. Since $2 = C_1 \leq K_1$ and $2 = C_2 \leq K_2$, this proves the equality for $\rho \in \{1, 2\}$. Let $\rho > 2$ and $\gamma_1 = \sup_{x \in [0,1]} \frac{x - x^{\rho-1}}{1+x}$. For all $x > 1$, $x - x^{\rho-1} < 0$, so $\gamma_1 = \sup_{x \in \mathbb{R}_+} \frac{x - x^{\rho-1}}{1+x}$. Then for all $x \in \mathbb{R}_+$,

$$\frac{x + x^\rho}{1+x} = \frac{x - x^{\rho-1}}{1+x} + x^{\rho-1} \leq \gamma_1 + x^{\rho-1},$$

hence $K_\rho \leq 2^{\rho-1}\gamma_1 + 2^{\rho-1} \times 1 = \left(1 + \sup_{x \in [0,1]} \frac{x - x^{\rho-1}}{1+x}\right) 2^{\rho-1}$. Let now $\rho \in (1, 2)$. For all $x \in \mathbb{R}_+$, $\frac{x+x^\rho}{1+x} \leq 1 + x^{\rho-1}$, hence $K_\rho \leq 2^{\rho-1} + 2$. Moreover, let $\gamma_2 = 1 + \sup_{x \in [1,+\infty)} \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}$. For all $x \in (0, 1)$, $x - x^{\rho-1} < 0$, so $\gamma_2 = 1 + \sup_{x \in \mathbb{R}_+^*} \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}$. Then for all $x \in \mathbb{R}_+^*$ we have

$$\frac{x + x^\rho}{1+x} = \left(1 + \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}\right) x^{\rho-1} \leq \gamma_2 x^{\rho-1},$$

hence $K_\rho \leq 2\gamma_2 = 2 + 2\sup_{x \in [1,+\infty)} \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}$. We deduce that

$$K_\rho \leq \left(2^{\rho-1} + 2\right) \wedge \left(2 + 2 \sup_{x \in [1,+\infty)} \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}\right) = 2 + \left(2 \sup_{x \in [1,+\infty)} \frac{x - x^{\rho-1}}{x^{\rho-1}(1+x)}\right) \wedge 2^{\rho-1}.$$

□

Remark 3.2.3. For $\rho = 2$, by (3.1.2) for the first inequality and the fact that $\sigma_2(\nu)$ is the standard deviation of ν, consequence of the bias-variance decomposition, for the last equality, we have

$$\begin{aligned}
\mathcal{W}_2^2(\mu, \nu) &\leq \mathcal{M}_2^2(\mu, \nu) \\
&= \int_\mathbb{R} y^2 \, \nu(dy) - \int_\mathbb{R} x^2 \, \mu(dx) \\
&\leq \int_\mathbb{R} y^2 \, \nu(dy) - \left(\int_\mathbb{R} x \, \mu(dx)\right)^2 = \int_\mathbb{R} y^2 \, \nu(dy) - \left(\int_\mathbb{R} y \, \nu(dy)\right)^2 \\
&= \sigma_2(\nu),
\end{aligned}$$

where the inequalities are equalities as soon as μ is reduced to an atom. Therefore we can improve the constant 2^ρ in (3.2.20) at least in the case $\rho = 2$. We can then naturally wonder whether we can also improve this constant for any $\rho > 1$. The constant

$$C'_\rho = \sup_{\nu \in \mathcal{P}_\rho(\mathbb{R})} \frac{\int_\mathbb{R} \left|y - \int_\mathbb{R} z \, \nu(dz)\right|^\rho \, \nu(dy)}{\int_\mathbb{R} |y|^\rho \, \nu(dy)}$$

is studied in [141]. For all $\nu \in \mathcal{P}_\rho(\mathbb{R})$, let $c \in \mathbb{R}$ be such that $\sigma_\rho^\rho(\nu) = \int_\mathbb{R} |y - c|^\rho \, \nu(dy)$ and ν_c be the image of ν by $y \mapsto y - c$. Then we have

$$\int_\mathbb{R} |z|^\rho \, \nu_c(dz) = \sigma_\rho^\rho(\nu) \quad \text{and} \quad \int_\mathbb{R} \left|y - \int_\mathbb{R} z \, \nu_c(dz)\right|^\rho \nu_c(dy) = \int_\mathbb{R} \left|y - \int_\mathbb{R} z \, \nu(dz)\right|^\rho \nu(dy) = \mathcal{W}_\rho^\rho(\mu_\nu, \nu),$$

where we denote $\mu_\nu = \delta_{\int_{\mathbb{R}} y\,\nu(dy)}$, which is dominated by ν in the convex order. We deduce that

$$C'_\rho = \sup_{\nu \in \mathcal{P}_\rho(\mathbb{R})} \frac{\mathcal{W}_\rho^\rho(\mu_\nu, \nu)}{\sigma_\rho^\rho(\nu)}.$$

Yet by [141, Theorem 2.3] we have $C'_\rho \sim_{\rho\to+\infty} \frac{2^{\rho-1}}{\sqrt{2e\rho}}$, which shows that we cannot lower the constant 2^ρ in (3.2.20) by a factor more than $2\sqrt{2e\rho}$ asymptotically for $\rho \to +\infty$.

3.3 Towards a multidimensional generalisation

3.3.1 Extension of the one dimensional inequality

One might legitimately ask whether the new stability inequality (3.1.4) holds in higher dimension $d \in \mathbb{N}^*$, that is if for all $\rho \geq 1$, there exists $C \in \mathbb{R}_+^*$ such that for all $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$ satisfying $\mu \leq_{cx} \nu$,

$$\mathcal{M}_\rho^\rho(\mu, \nu) \leq C \mathcal{W}_\rho(\mu, \nu) \sigma_\rho^{\rho-1}(\nu). \tag{3.3.1}$$

For all $d \in \mathbb{N}^*$ and $\rho \geq 1$, we define $C_{\rho,d}$ by

$$C_{\rho,d} = \inf \left\{ C > 0 \mid \forall \mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d) \text{ such that } \mu \leq_{cx} \nu,\ \mathcal{M}_\rho^\rho(\mu, \nu) \leq C \mathcal{W}_\rho(\mu, \nu) \sigma_\rho^{\rho-1}(\nu) \right\}.$$

The constant $C_{\rho,d}$ is well defined but is potentially infinite. Of course, for $d = 1$, we get $C_{\rho,d} = C_\rho$. Moreover, $C_{\rho,d}$ depends a priori on the choice of the norm in $\mathbb{R}^d$, but since all norms on $\mathbb{R}^d$ are equivalent, $C_{\rho,d}$ is finite for one specific norm iff it is finite for any norm.

We first look at extensions of the one dimensional inequality which give the same optimal constant. We begin with the fact that the stability inequality (3.1.4) can be tensorised: it holds in greater dimension when the marginals are independent.

Proposition 3.3.1. *Let $d \in \mathbb{N}^*$ and $\mu_1, \nu_1 \cdots, \mu_d, \nu_d \in \mathcal{P}_1(\mathbb{R})$ be such that for all $1 \leq i \leq d$, $\mu_i \leq_{cx} \nu_i$. Let $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ and $\nu = \nu_1 \otimes \cdots \otimes \nu_d$. Then $\mu \leq_{cx} \nu$ and*

$$\mathcal{M}_\rho^\rho(\mu, \nu) \leq C_\rho \mathcal{W}_\rho(\mu, \nu) \sigma_\rho^{\rho-1}(\nu), \tag{3.3.2}$$

where $\mathbb{R}^d$ is endowed with the L^ρ-norm.

Proof. For all $1 \leq i \leq d$, there exists by Lemma 3.4.1 below a martingale coupling $M_i \in \Pi^{\mathrm{M}}(\mu_i, \nu_i)$ between μ_i and ν_i, optimal for $\mathcal{M}_\rho(\mu_i, \nu_i)$. Let then M be the probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ defined by

$$M(dx, dy) = \mu(dx)\, m_1(x_1, dy_1) \cdots m_d(x_d, dy_d) = M_1(dx_1, dy_1) \otimes \cdots \otimes M_d(dx_d, dy_d).$$

It is clear that M is a martingale coupling between μ and ν, which shows that $\mu \leq_{cx} \nu$, and

$$\mathcal{M}_\rho^\rho(\mu, \nu) \leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\rho \, M(dx, dy) = \sum_{i=1}^d \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_i - y_i|^\rho \, M(dx, dy)$$

$$= \sum_{i=1}^{d} \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i|^\rho \, M_i(dx_i, dy_i).$$

Then for all $c = (c_1, \cdots, c_d) \in \mathbb{R}^d$ we have

$$\begin{aligned}
\mathcal{M}_\rho^\rho(\mu,\nu) &\le \sum_{i=1}^{d} \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i|^\rho \, M_i(dx_i, dy_i) = \sum_{i=1}^{d} \mathcal{M}_\rho^\rho(\mu_i,\nu_i) \\
&\le C_\rho \sum_{i=1}^{d} \mathcal{W}_\rho(\mu_i,\nu_i) \left(\int_{\mathbb{R}} |y_i - c_i|^\rho \, \nu_i(dy_i) \right)^{(\rho-1)/\rho} \\
&\le C_\rho \left(\sum_{i=1}^{d} \mathcal{W}_\rho^\rho(\mu_i,\nu_i) \right)^{1/\rho} \left(\sum_{i=1}^{d} \int_{\mathbb{R}} |y_i - c_i|^\rho \, \nu_i(dy_i) \right)^{(\rho-1)/\rho},
\end{aligned} \tag{3.3.3}$$

where for the last inequality we applied Hölder's inequality to the sum over i. Let $P \in \Pi(\mu,\nu)$ be a coupling between μ and ν. For $1 \le i \le d$, let P_i be the marginals of P with respect to the coordinates i and $i+d$, so that P_i is a coupling between μ_i and ν_i. Then

$$\begin{aligned}
\sum_{i=1}^{d} \mathcal{W}_\rho^\rho(\mu_i,\nu_i) &\le \sum_{i=1}^{d} \int_{\mathbb{R}\times\mathbb{R}} |x_i - y_i|^\rho \, P_i(dx_i, dy_i) = \int_{\mathbb{R}^d\times\mathbb{R}^d} \sum_{i=1}^{d} |x_i - y_i|^\rho \, P(dx, dy) \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} |x - y|^\rho \, P(dx, dy).
\end{aligned}$$

Since the inequality above is true for any coupling P between μ and ν, we get

$$\sum_{i=1}^{d} \mathcal{W}_\rho^\rho(\mu_i,\nu_i) \le \mathcal{W}_\rho^\rho(\mu,\nu), \tag{3.3.4}$$

which is in fact even an equality according to [5, Proposition 1.1]. We then deduce from (3.3.3) and (3.3.4) that

$$\begin{aligned}
\mathcal{M}_\rho^\rho(\mu,\nu) &\le C_\rho \mathcal{W}_\rho(\mu,\nu) \left(\sum_{i=1}^{d} \int_{\mathbb{R}} |y_i - c_i|^\rho \, \nu_i(dy_i) \right)^{(\rho-1)/\rho} \\
&= C_\rho \mathcal{W}_\rho(\mu,\nu) \left(\int_{\mathbb{R}^d} \sum_{i=1}^{d} |y_i - c_i|^\rho \, \nu(dy) \right)^{(\rho-1)/\rho} \\
&= C_\rho \mathcal{W}_\rho(\mu,\nu) \left(\int_{\mathbb{R}^d} |y - c|^\rho \, \nu(dy) \right)^{(\rho-1)/\rho}.
\end{aligned}$$

By taking the infimum over all $c \in \mathbb{R}^d$, we get (3.3.2). □

We now look at two measures μ and ν such that for X distributed according to μ, there exists $\lambda \ge 0$ such that ν is the probability distribution of $X + \lambda(X - \mathbb{E}[X])$ and the conditional probability distribution of X given the direction of $X - \mathbb{E}[X]$ has mean $\mathbb{E}[X]$. In order to transcribe formally the latter condition, we give the following definition.

Definition 3.3.2. Let $d \in \mathbb{N}^*\backslash\{1\}$ and $H : \mathbb{R}^d \to \mathbb{R}^d$ be a measurable map. We say that H is *direction-dependent* iff $|H(x)| = 1$ for all $x \in \mathbb{R}^d$ and

$$\forall x, y \in \mathbb{R}^d\backslash\{0\}, \quad H(x) = H(y) \iff y \in \mathrm{Span}(x).$$

In dimension $d \in \mathbb{N}^*\backslash\{1\}$, a natural example of a direction-dependent map $H : \mathbb{R}^d \to \mathbb{R}^d$ is given by the one defined for all $x = (x_1, \cdots, x_d) \in \mathbb{R}^d\backslash\{0\}$ by

$$\begin{cases} H(x) &= \frac{x}{|x|} \quad \text{if} \quad x_1 > 0 \text{ or there exists } i \in \{1, \cdots, d-1\} \text{ such that } x_1 = \cdots = x_i = 0 \text{ and } x_{i+1} > 0; \\ H(x) &= -\frac{x}{|x|} \quad \text{otherwise,} \end{cases} \tag{3.3.5}$$

and $H(0)$ is any vector with norm 1.

Proposition 3.3.3. *Let $d \in \mathbb{N}^*\backslash\{1\}$, $r \in [1, +\infty]$ and $\mathbb{R}^d$ be endowed with the L^r-norm. Let $\rho \geq 1$, $\mu \in \mathcal{P}_\rho(\mathbb{R}^d)$ be with mean $\alpha \in \mathbb{R}^d$, $\lambda : \mathbb{R}^d \to \mathbb{R}_+$, $H : \mathbb{R}^d \to \mathbb{R}^d$ be a direction-dependent measurable map in the sense of Definition 3.3.2 and ν be the image of μ by the map $x \mapsto x + \lambda(H(x-\alpha))(x-\alpha) = \alpha + (1 + \lambda(H(x-\alpha)))(x-\alpha)$.*

If $\mathbb{E}[|\lambda(H(X-\alpha))(X-\alpha)|^\rho] < +\infty$ and $\mathbb{E}[X|H(X-\alpha)] = \alpha$ almost surely for X distributed according to μ, then $\mu \leq_{cx} \nu$. If moreover λ is constant, then

$$\mathcal{M}_\rho^\rho(\mu, \nu) \leq C_\rho \mathcal{W}_\rho(\mu, \nu) \sigma_\rho^{\rho-1}(\nu). \tag{3.3.6}$$

Remark 3.3.4. Suppose that $\mu \in \mathcal{P}_\rho(\mathbb{R}^d)$ is symmetric with mean $\alpha \in \mathbb{R}^d$, that is $(x - \alpha)_\sharp\mu(dx) = (\alpha - x)_\sharp\mu(dx)$. Let H be defined by (3.3.5) and X be distributed according to μ. Then $(X-\alpha, H(X-\alpha)) \overset{d}{=} (\alpha-X, H(\alpha-X)) = (\alpha-X, H(X-\alpha))$, so $\mathbb{E}[X-\alpha|H(X-\alpha)] = \mathbb{E}[\alpha - X|H(X-\alpha)]$ a.s., hence $\mathbb{E}[X|H(X-\alpha)] = \alpha$ a.s.

The proof of Proposition 3.3.3 relies on the following lemma, whose proof is deferred to Section 3.4, which explains why we can endow $\mathbb{R}^d$ with the L^r-norm for $r \in [1, \infty]$.

Lemma 3.3.5. *Let $d \in \mathbb{N}^*\backslash\{1\}$, $r \in [1, +\infty]$, $\mathbb{R}^d$ be endowed with the L^r-norm, $\mathbb{S}^{d-1} = \{a \in \mathbb{R}^d \mid |a| = 1\}$ and $\mathrm{sgn} : \mathbb{R} \to \mathbb{R}, x \mapsto \mathbb{1}_{\{x\geq 0\}} - \mathbb{1}_{\{x<0\}}$. For all $a = (a_1, \cdots, a_d) \in \mathbb{S}^{d-1}$ and $c = (c_1, \cdots, c_d) \in \mathbb{R}^d$, let c_a be defined by*

$$c_a = \begin{cases} \left(\sum_{i=1}^d c_i \,\mathrm{sgn}(a_i)|a_i|^{r-1}\right) a & \text{if} \quad r < +\infty \\ c_i \,\mathrm{sgn}(a_i) a & \text{if} \quad r = +\infty, \text{ where } i \in \{1, \cdots, d\} \text{ is such that } |a_i| = 1. \end{cases} \tag{3.3.7}$$

Then

$$\forall a \in \mathbb{S}^{d-1}, \quad \forall c \in \mathbb{R}^d, \quad \forall y \in \mathrm{Span}(a), \quad |y - c_a| \leq |y - c|.$$

Proof of Proposition 3.3.3. Up to replacing μ and ν by their respective images by the map $x \mapsto x - \alpha$, we may suppose without loss of generality that $\alpha = 0$.

Let $(p(a, dx))_{a\in H(\mathbb{R}^d)}$ be a probability kernel such that $(H_\sharp\mu)(da)\, p(a, dx)$ is the image of μ by the map $x \mapsto (H(x), x)$. For all $a \in H(\mathbb{R}^d)$, let $\widetilde{p}(a, dy)$ be the image of $p(a, dx)$ by the map $x \mapsto (1+\lambda(H(x)))x$. For all $a \in H(\mathbb{R}^d)$, let $q(a,\cdot)$, resp. $\tilde{q}(a,\cdot)$, be the image of $p(a,\cdot)$, resp. $\tilde{p}(a,\cdot)$, by the map $y \mapsto \langle y, a\rangle$. We have

$$\int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |x|^\rho \, p(a,dx)\right) (H_\sharp\mu)(da) = \int_{\mathbb{R}^d} |x|^\rho \, \mu(dx) < +\infty,$$

$$\text{and} \quad \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |y|^\rho \, \tilde{p}(a,dy)\right) (H_\sharp\mu)(da) = \int_{\mathbb{R}^d} |(1+\lambda(H(x)))x|^\rho \, \mu(dx) < +\infty,$$

so $H_\sharp\mu(da)$-almost everywhere, $p(a,\cdot)$ and $q(a,\cdot)$, and therefore $\tilde{p}(a,\cdot)$ and $\tilde{q}(a,\cdot)$, belong to $\mathcal{P}_\rho(\mathbb{R}^d)$. Moreover we see by definition of H that for $H_\sharp\mu(da)$-almost all $a \in H(\mathbb{R}^d)$, $p(a, \mathrm{Span}(a)) = 1$, so

$$p(a,dx) = \int_{\mathbb{R}} \delta_{ta}(dx)\, q(a,dt) \quad \text{and} \quad \tilde{p}(a,dy) = \int_{\mathbb{R}} \delta_{sa}(dy)\, \tilde{q}(a,ds).$$

By assumption, we have $H_\sharp\mu(da)$-almost everywhere

$$a\int_{\mathbb{R}} t\, q(a,dt) = \int_{\mathbb{R}^d} x\, p(a,dx) = 0.$$

Since $\tilde{q}(a,\cdot)$ is the image of $q(a,\cdot)$ by the map $y \mapsto (1+\lambda(H(y)))y$, or equivalently by the map $y \mapsto (1+\lambda(a))y$, by Lemma 3.4.4 below, for $H_\sharp\mu(da)$-almost all $a \in H(\mathbb{R}^d)$, $q(a,\cdot) \leq_{cx} \tilde{q}(a,\cdot)$. Up to replacing $q(a,\cdot)$ and therefore $\tilde{q}(a,\cdot)$ by δ_0 on a $H_\sharp\mu$-null set, we may suppose without loss of generality that $q(a,\cdot), \tilde{q}(a,\cdot) \in \mathcal{P}_\rho(\mathbb{R}^d)$ and $q(a,\cdot) \leq_{cx} \tilde{q}(a,\cdot)$ for all $a \in H(\mathbb{R}^d)$. By Lemma 3.4.1 below, for all $a \in H(\mathbb{R}^d)$ there exists a martingale coupling $M^a \in \Pi^{\mathrm{M}}(q(a,\cdot), \tilde{q}(a,\cdot))$ optimal for $\mathcal{M}_\rho(q(a,\cdot), \tilde{q}(a,\cdot))$.

Assume for a moment that the map $a \mapsto M^a$ is measurable, where $\mathcal{P}(\mathbb{R}\times\mathbb{R})$ is endowed with the σ-field generated by the weak convergence topology. For all $a \in H(\mathbb{R}^d)$, let $\widetilde{M}^a$ be the image of M^a by the map $(t,s) \mapsto (ta, sa)$. Then the map $a \mapsto \widetilde{M}^a$ is also measurable, which is equivalent to say (see for instance [6, Theorem 19.12]) that $(\widetilde{M}^a)_{a\in H(\mathbb{R}^d)}$ is a probability kernel from $H(\mathbb{R}^d)$ to $\mathbb{R}^d \times \mathbb{R}^d$. Hence we can define

$$\overline{M}(dx,dy) = \int_{a\in H(\mathbb{R}^d)} \widetilde{M}^a(dx,dy)\, (H_\sharp\mu)(da).$$

For all $a \in H(\mathbb{R}^d)$, M^a is a martingale coupling between $q(a,\cdot)$ and $\tilde{q}(a,\cdot)$, hence we easily see that $\widetilde{M}^a$ is a martingale coupling between the respective images of $q(a,\cdot)$ and $\tilde{q}(a,\cdot)$ by the map $t \mapsto at$, namely $p(a,\cdot)$ and $\tilde{p}(a,\cdot)$. Therefore one can also readily show that $\overline{M}$ is a martingale coupling between $\int_{a\in H(\mathbb{R}^d)} p(a,dx)\,(H_\sharp\mu)(da) = \mu(dx)$ and $\int_{a\in H(\mathbb{R}^d)} \tilde{p}(a,dy)\,(H_\sharp\mu)(da) = \nu(dy)$. Consequently, we have

$$\begin{aligned}\mathcal{M}_\rho^\rho(\mu,\nu) &\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\, \overline{M}(dx,dy) = \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\, \widetilde{M}^a(dx,dy)\right) (H_\sharp\mu)(da) \\ &= \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}\times\mathbb{R}} |s-t|^\rho\, M^a(dt,ds)\right) (H_\sharp\mu)(da) = \int_{H(\mathbb{R}^d)} \mathcal{M}_\rho^\rho(q(a,\cdot), \tilde{q}(a,\cdot))\, (H_\sharp\mu)(da).\end{aligned} \tag{3.3.8}$$

Let $c \in \mathbb{R}^d$. For all $a \in H(\mathbb{R}^d)$, let c_a be defined by (3.3.7) and $s_a \in \mathbb{R}$ be such that $c_a = s_a a$. If the map λ is constant equal to some $\lambda \in \mathbb{R}_+$ (with a slight abuse of notation), then using the definition of C_ρ for the first inequality, Lemma 3.4.2 below for the first equality, Lemma 3.3.5 for the second inequality, Hölder's inequality for the third inequality and Lemma 3.4.2 again for the last equality (there the constancy of λ plays a crucial role), we deduce that

$$\begin{aligned}
&\mathcal{M}_\rho^\rho(\mu,\nu) \\
&\le \int_{H(\mathbb{R}^d)} C_\rho \mathcal{W}_\rho(q(a,\cdot),\widetilde{q}(a,\cdot)) \left(\int_{\mathbb{R}} |s-s_a|^\rho\,\widetilde{q}(a,ds)\right)^{(\rho-1)/\rho} (H_\sharp\mu)(da) \\
&= C_\rho\lambda \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}} |t|^\rho\, q(a,dt)\right)^{1/\rho} \left(\int_{\mathbb{R}} |sa-s_a a|^\rho\,\widetilde{q}(a,ds)\right)^{(\rho-1)/\rho} (H_\sharp\mu)(da) \\
&= C_\rho\lambda \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |x|^\rho\, p(a,dx)\right)^{1/\rho} \left(\int_{\mathbb{R}^d} |y-c_a|^\rho\,\widetilde{p}(a,dy)\right)^{(\rho-1)/\rho} (H_\sharp\mu)(da) \\
&\le C_\rho\lambda \int_{H(\mathbb{R}^d)} \left(\int_{\mathbb{R}^d} |x|^\rho\, p(a,dx)\right)^{1/\rho} \left(\int_{\mathbb{R}^d} |y-c|^\rho\,\widetilde{p}(a,dy)\right)^{(\rho-1)/\rho} (H_\sharp\mu)(da) \\
&\le C_\rho\lambda \left(\int_{H(\mathbb{R}^d)}\int_{\mathbb{R}^d} |x|^\rho\, p(a,dx)\,(H_\sharp\mu)(da)\right)^{1/\rho} \left(\int_{H(\mathbb{R}^d)}\int_{\mathbb{R}^d} |y-c|^\rho\,\widetilde{p}(a,dy)\,(H_\sharp\mu)(da)\right)^{(\rho-1)/\rho} \\
&\le C_\rho\lambda \left(\int_{\mathbb{R}^d} |x|^\rho\,\mu(dx)\right)^{1/\rho} \left(\int_{\mathbb{R}^d} |(1+\lambda)x-c|^\rho\,\mu(dx)\right)^{(\rho-1)/\rho} \\
&= C_\rho\mathcal{W}_\rho(\mu,\nu) \left(\int_{\mathbb{R}^d} |y-c|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}.
\end{aligned} \tag{3.3.9}$$

By taking the infimum over all $c \in \mathbb{R}^d$, we get (3.3.6).

It remains to overcome the issue of the measurability of the map $a \mapsto M^a$. For all $q, q' \in \mathcal{P}_\rho(\mathbb{R})$ such that $q \le_{cx} q'$, there is not necessarily uniqueness of a martingale coupling between q and q' optimal for $\mathcal{M}_\rho^\rho(q,q')$, as illustrated for instance by (3.1.2) for $\rho = 2$ and $\mathbb{R}^d$ endowed with the Euclidean norm, for which any martingale coupling is optimal. A solution to overcome that hurdle is to fix $\varepsilon > 0$ and choose a martingale coupling which is optimal for an appropriate cost function V^ε instead of $\mathcal{M}_\rho$. This alternate cost function V^ε will be defined so that it has a unique minimiser which is almost optimal for $\mathcal{M}_\rho$ in a sense precised below. Let then $(g_k)_{k\in\mathbb{N}}$ be a family of 1-Lipschitz continuous functions bounded by 1 which separates $\mathcal{P}(\mathbb{R})$ (see [78, Theorem 4.5.(a)]). Define $C^\varepsilon : \mathbb{R}\times\mathcal{P}_\rho(\mathbb{R}) \to \mathbb{R}_+$ for all $\varepsilon > 0$ and $(x,\tau) \in \mathbb{R}\times\mathcal{P}_\rho(\mathbb{R})$ by

$$C^\varepsilon(x,\tau) = \int_{\mathbb{R}} |x-y|^\rho\,\tau(dy) + \varepsilon\sum_{k\in\mathbb{N}} \frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2.$$

For $q, q' \in \mathcal{P}_\rho(\mathbb{R})$ such that $q \le_{cx} q'$, let

$$V^\varepsilon(q,q') = \inf_{M\in\Pi^{\mathrm{M}}(q,q')} \int_{\mathbb{R}} C^\varepsilon(x, m(x,\cdot))\, q(dx), \tag{3.3.10}$$

where for all $M \in \Pi^{\mathrm{M}}(q, q')$, $(m(x,\cdot))_{x\in\mathbb{R}}$ denotes a regular conditional probability distribution of M with respect to μ. Note that (3.3.10) is an instance of the class of the weak optimal transport problems introduced by Gozlan, Roberto, Samson and Tetali [90]. Some key results of the classical optimal transport theory, such as existence, duality and the monotonicity principle still hold for weak costs, as showed by Backhoff-Veraguas, Beiglböck and Pammer [18]. More recently a stronger stability result was proved by the authors and Beiglböck and Pammer, see Chapter 6 below, and will be used in the present proof.

We now show that for all $\varepsilon > 0$ and $a \in H(\mathbb{R}^d)$, there exists a unique martingale coupling $M^{\varepsilon,a}$ between $q(a,\cdot)$ and $\tilde{q}(a,\cdot)$ optimal for $V^\varepsilon(q(a,\cdot),\tilde{q}(a,\cdot))$, and the map $a \mapsto M^{\varepsilon,a}$ is measurable. To do so, we show that C^ε is continuous on $\mathbb{R}\times\mathcal{P}_\rho(\mathbb{R})$ with respect to the product metric $((x,\tau),(x',\tau')) \mapsto |x-x'| + \mathcal{W}_\rho(\tau,\tau')$, strictly convex in the second argument and such that there exists a constant $K > 0$ which satisfies for all $(x,\tau) \in \mathbb{R}\times\mathcal{P}_\rho(\mathbb{R})$

$$|C^\varepsilon(x,\tau)| \le K\left(1 + |x|^\rho + \int_{\mathbb{R}} |y|^\rho\,\tau(dy)\right). \tag{3.3.11}$$

Let us show that C^ε is continuous. Let $x\in\mathbb{R}$, $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $\tau\in\mathcal{P}_\rho(\mathbb{R})$ and $(\tau_n)_{n\in\mathbb{N}} \in \mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ be such that $x_n \to x$ and $\tau_n \to \tau$ in $\mathcal{W}_\rho$ as $n \to +\infty$. We recall that the latter is equivalent to say that for any continuous function $f : \mathbb{R} \to \mathbb{R}$ which satisfies the growth constraint

$$\exists \alpha > 0, \quad \forall y \in \mathbb{R}, \quad |f(y)| \le \alpha(1 + |y|^\rho), \tag{3.3.12}$$

we have $\int_{\mathbb{R}} f(y)\,\tau_n(dy) \to \int_{\mathbb{R}} f(y)\,\tau(dy)$ as $n \to +\infty$. On the one hand,

$$\begin{aligned}\left|\int_{\mathbb{R}} |x-y|^\rho\,\tau(dy) - \int_{\mathbb{R}} |x_n-y|^\rho\,\tau_n(dy)\right| &\le \left|\int_{\mathbb{R}} |x-y|^\rho\,\tau(dy) - \int_{\mathbb{R}} |x-y|^\rho\,\tau_n(dy)\right| \\ &\quad + \int_{\mathbb{R}} \big||x-y|^\rho - |x_n-y|^\rho\big|\,\tau_n(dy).\end{aligned}$$

Since the map $y \mapsto |x-y|^\rho$ satisfies (3.3.12), the first summand of the right-hand side vanishes as n goes to $+\infty$. Moreover, one can easily show that for all $z \ge 1$, $z^\rho - 1 \le \rho z^{\rho-1}(z-1)$, from which we deduce $|b^\rho - a^\rho| \le \rho(a\vee b)^{\rho-1}|b-a|$ for all $a,b \ge 0$. Applying the latter inequality with $(a,b) = (|x-y|,|x_n-y|)$ yields

$$\begin{aligned}\int_{\mathbb{R}} \big||x-y|^\rho - |x_n-y|^\rho\big|\,\tau_n(dy) &\le \rho\int_{\mathbb{R}} (|x-y|\vee|x_n-y|)^{\rho-1}|x_n-x|\,\tau_n(dy) \\ &\le \rho\int_{\mathbb{R}} (|x-y| + |x_n-x|)^{\rho-1}|x_n-x|\,\tau_n(dy),\end{aligned}$$

where the right-hand side vanishes as n goes to $+\infty$ by the dominated convergence theorem. This proves that the map $(x,\tau) \mapsto \int_{\mathbb{R}} |x-y|^\rho\,\tau(dy)$ is continuous. On the other hand, by Kantorovich and Rubinstein's duality theorem and Jensen's inequality, we have

$$\begin{aligned}&\left|\sum_{k\in\mathbb{N}} \frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2 - \sum_{k\in\mathbb{N}} \frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau_n(dy)\right|^2\right| \\ &\le \sum_{k\in\mathbb{N}} \frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy) + \int_{\mathbb{R}} g_k(y)\,\tau_n(dy)\right|\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy) - \int_{\mathbb{R}} g_k(y)\,\tau_n(dy)\right|\end{aligned}$$

$$\le \sum_{k\in\mathbb{N}} \frac{2}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy) - \int_{\mathbb{R}} g_k(y)\,\tau_n(dy)\right|$$
$$\le \sum_{k\in\mathbb{N}} \frac{2}{2^k}\mathcal{W}_1(\tau,\tau_n) \le 4\mathcal{W}_\rho(\tau,\tau_n) \underset{n\to+\infty}{\longrightarrow} 0.$$

We deduce that $(x,\tau)\mapsto \sum_{k\in\mathbb{N}}\frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2$ and therefore C^ε is continuous. Let us now show that C^ε is strictly convex in the second argument. Let $x\in\mathbb{R}$. The map $\tau\mapsto \int_{\mathbb{R}}|x-y|^\rho\,\tau(dy)$ is clearly linear so it is sufficient to show that $\tau\mapsto \sum_{k\in\mathbb{N}}\frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2$ is strictly convex. Let $\tau,\tau'\in\mathcal{P}_\rho(\mathbb{R})$ be such that $\tau\neq\tau'$ and $\alpha\in(0,1)$. Then there exists $k_0\in\mathbb{N}$ such that $\int_{\mathbb{R}} g_{k_0}(y)\,\tau(dy)\neq\int_{\mathbb{R}} g_{k_0}(y)\,\tau'(dy)$ and the strict convexity of the square implies

$$\left|\alpha\int_{\mathbb{R}} g_k(y)\,\tau(dy)+(1-\alpha)\int_{\mathbb{R}} g_k(y)\,\tau'(dy)\right|^2 \le \alpha\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2+(1-\alpha)\left|\int_{\mathbb{R}} g_k(y)\,\tau'(dy)\right|^2$$

for all $k\in\mathbb{N}$, with the inequality strict at least for $k=k_0$, which yields strict convexity of $\tau\mapsto\sum_{k\in\mathbb{N}}\frac{1}{2^k}\left|\int_{\mathbb{R}} g_k(y)\,\tau(dy)\right|^2$. Finally, let us show that C^ε satisfies (3.3.11). Let $x\in\mathbb{R}$ and $\tau\in\mathcal{P}_\rho(\mathbb{R})$. Then

$$|C^\varepsilon(x,\tau)| \le 2^{\rho-1}\int_{\mathbb{R}}(|x|^\rho+|y|^\rho)\,\tau(dy)+\varepsilon\sum_{k\in\mathbb{N}}\frac{1}{2^k} \le K\left(1+|x|^\rho+\int_{\mathbb{R}}|y|^\rho\,\tau(dy)\right),$$

where $K=(2\varepsilon)\vee 2^{\rho-1}$. Let $\widetilde{\Pi}=\{(q,q')\in\mathcal{P}_\rho(\mathbb{R})\times\mathcal{P}_\rho(\mathbb{R})\mid q\le_{cx} q'\}$. By Theorem 6.2.6 below, for all $(q,q')\in\widetilde{\Pi}$, there exists a unique martingale coupling $M^\varepsilon_{q,q'}\in\Pi^{\mathrm{M}}(q,q')$ optimal for $V^\varepsilon(q,q')$. Moreover, if $(q_k)_{k\in\mathbb{N}},(q'_k)_{k\in\mathbb{N}}\in\mathcal{P}_\rho(\mathbb{R})^{\mathbb{N}}$ are such that q_k and q'_k respectively converge to q and q' in $\mathcal{W}_\rho$ as $k\to+\infty$, then $M^\varepsilon_{q_k,q'_k}$ converges to $M^\varepsilon_{q,q'}$ for a finer topology than the one induced by $\mathcal{W}_\rho$ as $k\to+\infty$. This implies that the map $(q,q')\mapsto M^\varepsilon_{q,q'}$ is measurable from $\widetilde{\Pi}$ to $\mathcal{P}(\mathbb{R}\times\mathbb{R})$, where $\widetilde{\Pi}$ is endowed with the trace of the Borel σ-algebra of $\mathcal{P}_\rho(\mathbb{R})\times\mathcal{P}_\rho(\mathbb{R})$ endowed with the product of the $\mathcal{W}_\rho$-distance topologies. Since $(\mathcal{P}_\rho(\mathbb{R}),\mathcal{W}_\rho)$ is separable [191, Theorem 6.18], the latter σ-algebra coincides with the product σ-algebra $\mathcal{B}_\rho\otimes\mathcal{B}_\rho$, where $\mathcal{B}_\rho$ denotes the Borel σ-algebra of $\mathcal{P}_\rho(\mathbb{R})$ endowed with the $\mathcal{W}_\rho$-distance topology [6, Theorem 4.44]. By Lemma 3.4.5 below, $\mathcal{B}_\rho\otimes\mathcal{B}_\rho=\mathcal{B}\otimes\mathcal{B}$, where $\mathcal{B}$ denotes the Borel σ-algebra of $\mathcal{P}_\rho(\mathbb{R})$ endowed with the weak convergence topology. We deduce that the map $(q,q')\mapsto M^\varepsilon_{q,q'}$ is measurable from $\widetilde{\Pi}$ to $\mathcal{P}(\mathbb{R}\times\mathbb{R})$, where $\widetilde{\Pi}$ is endowed with the trace of $\mathcal{B}\otimes\mathcal{B}$.

For all $\varepsilon>0$ and $a\in H(\mathbb{R}^d)$, let then $M^{\varepsilon,a}=M^\varepsilon_{q(a,\cdot),\widetilde{q}(a,\cdot)}$. Since $(q(a,\cdot))_{a\in H(\mathbb{R}^d)}$ and $(\widetilde{q}(a,\cdot))_{a\in H(\mathbb{R}^d)}$ are probability kernels, the maps $a\mapsto q(a,\cdot)$ and $a\mapsto\widetilde{q}(a,\cdot)$ are measurable [6, Theorem 19.12], and so is $a\mapsto(q(a,\cdot),\widetilde{q}(a,\cdot))$ where the codomain $\widetilde{\Pi}$ is endowed with the trace of $\mathcal{B}\otimes\mathcal{B}$. By composition, the map $a\mapsto M^{\varepsilon,a}$ is measurable. Let $\widetilde{M}^{\varepsilon,a}$ be the image of $M^{\varepsilon,a}$ by the map $(t,s)\mapsto(ta,sa)$ and

$$\overline{M}^\varepsilon(dx,dy)=\int_{a\in H(\mathbb{R}^d)}\widetilde{M}^{\varepsilon,a}(dx,dy)\,(H_\sharp\mu)(da).$$

We see with the very same arguments as the ones given above for the martingale coupling $\overline{M}$ that $\overline{M}^\varepsilon$ is a martingale coupling between μ and ν, which proves that $\mu \leq_{cx} \nu$. Like in (3.3.8) we get

$$\begin{aligned}\mathcal{M}_\rho^\rho(\mu,\nu) &\leq \int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\,\overline{M}^\varepsilon(dx,dy) = \int_{H(\mathbb{R}^d)}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\widetilde{M}^{\varepsilon,a}(dx,dy)\right)(H_\sharp\mu)(da)\\ &= \int_{H(\mathbb{R}^d)}\left(\int_{\mathbb{R}\times\mathbb{R}}|s-t|^\rho\,M^{\varepsilon,a}(dt,ds)\right)(H_\sharp\mu)(da)\\ &\leq \int_{H(\mathbb{R}^d)} V^\varepsilon(q(a,\cdot),q'(a,\cdot))\,(H_\sharp\mu)(da).\end{aligned}$$

For all $q,q'\in\mathcal{P}_\rho(\mathbb{R})$ such that $q\leq_{cx} q'$, we have $V^\varepsilon(q,q')\leq\mathcal{M}_\rho^\rho(q,q')+2\varepsilon$, hence $\mathcal{M}_\rho^\rho(\mu,\nu)\leq\int_{H(\mathbb{R}^d)}\mathcal{M}_\rho^\rho(q(a,\cdot),\widetilde{q}(a,\cdot))\,(H_\sharp\mu)(da)+2\varepsilon$. Assuming that λ is constant, we then derive the same estimates as in (3.3.9), except that we add $+2\varepsilon$ at the end of each line. By taking the limit $\varepsilon\to0$, the same conclusion holds. $\square$

3.3.2 The scaling case

We call scaling case the situation in which two measures μ and ν are such that for X distributed according to μ, there exists $\lambda\geq0$ such that ν is the probability distribution of $X+\lambda(X-\mathbb{E}[X])=\mathbb{E}[X]+(1+\lambda)(X-\mathbb{E}[X])$. In the previous section we already considered this case under a complementary assumption on the conditional probability distribution of X, see Proposition 3.3.3. We release here the latter constraint and study the impact on the constant C in (3.3.1). We begin with the proof that the scaling case provides a lower bound for $C_{\rho,d}$ for any $\rho\geq1$ and $d\in\mathbb{N}^*\backslash\{1\}$, the case $d=1$ having already been treated by Proposition 3.2.1.

Proposition 3.3.6. *Let $d\in\mathbb{N}^*\backslash\{1\}$ and $\rho\geq1$. Regardless of the norm $\mathbb{R}^d$ is endowed with, we have*

$$C_{1,d}\geq3,\quad\text{and for all }\rho>1,\quad C_{\rho,d}\geq2\vee\left(2^{\rho-1}\sup_{x\in(1,+\infty)}\frac{x+x^\rho}{(1+x)^\rho}\right).$$

The proof of Proposition 3.3.6 relies on the following lemma, which allows us to propagate to higher dimension the lower bound in (3.2.12) for $d=1$ and the lower bound we exhibit in the proof for $d=2$.

Lemma 3.3.7. *Let $d,d'\in\mathbb{N}^*$ be such that $d<d'$, $\rho\geq1$, $|\cdot|$ be a norm on $\mathbb{R}^d$ and $|\cdot|'$ be a norm on $\mathbb{R}^{d'}$ satisfying the following consistency condition:*

$$\forall x_1,\cdots,x_d\in\mathbb{R},\quad |(x_1,\cdots,x_d,0,\cdots,0)|'=|(x_1,\cdots,x_d)|. \tag{3.3.13}$$

Then $C_{\rho,d}\leq C_{\rho,d'}$ for $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ respectively endowed with $|\cdot|$ and $|\cdot|'$.

Proof of Lemma 3.3.7. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ be such that $\mu \leq_{cx} \nu$. Let μ' and ν' be the respective images of μ and ν by the map $\mathbb{R}^d \ni (x_1, \cdots, x_d) \mapsto (x_1, \cdots, x_d, 0, \cdots 0) \in \mathbb{R}^{d'}$. Let $c = (c_1, \cdots, c_d) \in \mathbb{R}^d$ and $c' = (c_1, \cdots, c_d, 0, \cdots, 0) \in \mathbb{R}^{d'}$. If $\mathbb{R}^d$ and $\mathbb{R}^{d'}$ are respectively endowed with $|\cdot|$ and $|\cdot|'$, then by (3.3.13) and the definition of $C_{\rho,d'}$, we have

$$\begin{aligned}\mathcal{M}_\rho^\rho(\mu,\nu) = \mathcal{M}_\rho^\rho(\mu',\nu') &\leq C_{\rho,d'} \mathcal{W}_\rho(\mu',\nu') \left(\int_{\mathbb{R}^{d'}} |y - c'|^\rho \, \nu'(dy)\right)^{(\rho-1)/\rho} \\ &= C_{\rho,d'} \mathcal{W}_\rho(\mu,\nu) \left(\int_{\mathbb{R}^d} |y - c|^\rho \, \nu(dy)\right)^{(\rho-1)/\rho}.\end{aligned}$$

By taking the infimum over all $c \in \mathbb{R}$, we get $\mathcal{M}_\rho^\rho(\mu,\nu) \leq C_{\rho,d'} W_\rho(\mu,\nu) \sigma_\rho^{\rho-1}(\nu)$, hence $C_{\rho,d} \leq C_{\rho,d'}$. □

Proof of Proposition 3.3.6. By Proposition 3.2.1, we have $C_{\rho,1} \geq 2^{\rho-1} \sup_{x \in (1,+\infty)} \frac{x+x^\rho}{(1+x)^\rho}$. Let N be a norm on $\mathbb{R}^2$ and $N' : y \mapsto N(y)/N((1,0))$, so that for all $x \in \mathbb{R}$, $|x| = N'((x,0))$. By Lemma 3.3.7, we deduce that

$$2^{\rho-1} \sup_{x \in (1,+\infty)} \frac{x + x^\rho}{(1+x)^\rho} \leq C_{\rho,1} \leq C_{\rho,2}, \tag{3.3.14}$$

for $\mathbb{R}^2$ endowed with N' and therefore with N. Assume for a moment that we also have

$$C_{1,2} \geq 3 \quad \text{and} \quad \forall \rho > 1, \quad C_{\rho,2} \geq 2. \tag{3.3.15}$$

Let us now consider $d \in \mathbb{N}^* \backslash \{1\}$ and $|\cdot|$ a norm on $\mathbb{R}^d$. Let $N : \mathbb{R}^2 \to \mathbb{R}$ be defined for all $(x,y) \in \mathbb{R}^2$ by $N((x,y)) = |(x,y,0,\cdots,0)|$. One can easily check that N is a norm on $\mathbb{R}^2$. By (3.3.14), (3.3.15) and Lemma 3.3.7, we have $3 \leq C_{1,2} \leq C_{1,d}$, and for $\rho > 1$, $2 \vee \left(2^{\rho-1} \sup_{x \in (1,+\infty)} \frac{x+x^\rho}{(1+x)^\rho}\right) \leq C_{\rho,2} \leq C_{\rho,d}$, which is conclusive.

It remains to show (3.3.15). Let $p, q, r \in [0,1]$ be such that $p + q + r = 1$, i, j and k be three non-collinear points in $\mathbb{R}^2$ and α be their weighted barycentre, namely $\alpha = pi + qj + rk$. Let $\ell \in (0,1)$ and i', j' and k' be the respective images of i, j and k after a homothety with centre α and scale factor $1 - \ell$, that is

$$i' - \alpha = (1-\ell)(i - \alpha), \quad j' - \alpha = (1-\ell)(j-\alpha), \quad \text{and} \quad k' - \alpha = (1-\ell)(k-\alpha).$$

Let then

$$\mu = p\delta_{i'} + q\delta_{j'} + r\delta_{k'} \quad \text{and} \quad \nu = p\delta_i + q\delta_j + r\delta_k.$$

Since ν is the image of μ by the map $x \mapsto x + \frac{\ell}{1-\ell}(x - \alpha)$, by Lemma 3.4.2 below applied with $\lambda = \frac{\ell}{1-\ell}$,

$$\mathcal{W}_\rho(\mu,\nu) = \frac{\ell}{1-\ell} \left(p|i' - \alpha|^\rho + q|j' - \alpha|^\rho + r|k' - \alpha|^\rho\right)^{1/\rho} = \ell(p|i-\alpha|^\rho + q|j-\alpha|^\rho + r|k-\alpha|^\rho)^{1/\rho}.$$

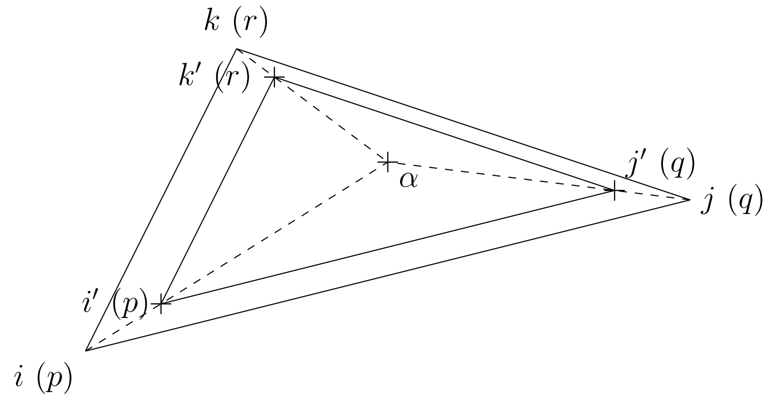

Figure 3.1: Points involved in the proof. A number in brackets is the weight associated to the point. The probability measure μ, resp. ν, is concentrated on the inner, resp. outer, triangle.

On the other hand, since i, j and k are non-collinear, each of the points i', j' and k' can be written as a unique convex combination of the points i, j and k, given by

$$\begin{aligned} i' &= (1-\ell+\ell p)i + \ell q j + \ell r k; \\ j' &= \ell p i + (1-\ell+\ell q)j + \ell r k; \\ k' &= \ell p i + \ell q j + (1-\ell+\ell r)k. \end{aligned} \tag{3.3.16}$$

We deduce from (3.3.16) the only martingale coupling M between μ and ν:

$$\begin{aligned} M(dx,dy) = \mu(dx)\Big(&\mathbb{1}_{\{x=i'\}}((1-\ell+\ell p)\delta_i + \ell q\delta_j + \ell r\delta_k)(dy) \\ &+ \mathbb{1}_{\{x=j'\}}(\ell p\delta_i + (1-\ell+\ell q)\delta_j + \ell r\delta_k)(dy) \\ &+ \mathbb{1}_{\{x=k'\}}(\ell p\delta_i + \ell q\delta_j + (1-\ell+\ell r)\delta_k)(dy)\Big). \end{aligned}$$

We can then compute

$$\begin{aligned} \mathcal{M}_\rho^\rho(\mu,\nu) = \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho\, M(dx,dy) &= p\left((1-\ell+\ell p)|i-i'|^\rho + \ell q|j-i'|^\rho + \ell r|k-i'|^\rho\right) \\ &+ q\left(\ell p|i-j'|^\rho + (1-\ell+\ell q)|j-j'|^\rho + \ell r|k-j'|^\rho\right) \\ &+ r\left(\ell p|i-k'|^\rho + \ell q|j-k'|^\rho + (1-\ell+\ell r)|k-k'|^\rho\right). \end{aligned}$$

Therefore, we get

$$\frac{\mathcal{M}_\rho^\rho(\mu,\nu)}{\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}^2}|y-\alpha|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}}$$

$$= \frac{\begin{array}{l}\Big[p\left((1-\ell+\ell p)\ell^{\rho-1}|i-\alpha|^\rho+q|j-i'|^\rho+r|k-i'|^\rho\right)\\ +q\left(p|i-j'|^\rho+(1-\ell+\ell q)\ell^{\rho-1}|j-\alpha|^\rho+r|k-j'|^\rho\right)\\ +r\left(p|i-k'|^\rho+q|j-k'|^\rho+(1-\ell+\ell r)\ell^{\rho-1}|k-\alpha|^\rho\right)\Big]\end{array}}{\left(p|i-\alpha|^\rho+q|j-\alpha|^\rho+r|k-\alpha|^\rho\right)^{1/\rho}\left(p|i-\alpha|^\rho+q|j-\alpha|^\rho+r|k-\alpha|^\rho\right)^{(\rho-1)/\rho}} \tag{3.3.17}$$

$$\underset{\ell\to0^+}{\longrightarrow}\frac{\begin{array}{l}\Big[p(\mathbb{1}_{\{\rho=1\}}|i-\alpha|^\rho+q|j-i|^\rho+r|i-k|^\rho)\\ +q(p|j-i|^\rho+\mathbb{1}_{\{\rho=1\}}|j-\alpha|^\rho+r|k-j|^\rho)\\ +r(p|i-k|^\rho+q|k-j|^\rho+\mathbb{1}_{\{\rho=1\}}|k-\alpha|^\rho)\Big]\end{array}}{p|i-\alpha|^\rho+q|j-\alpha|^\rho+r|k-\alpha|^\rho}$$

$$=\mathbb{1}_{\{\rho=1\}}+2\frac{pq|j-i|^\rho+qr|k-j|^\rho+rp|i-k|^\rho}{p|i-\alpha|^\rho+q|j-\alpha|^\rho+r|k-\alpha|^\rho}.$$

We deduce that

$$C_{\rho,2}\geq\mathbb{1}_{\{\rho=1\}}+2\frac{pq|j-i|^\rho+qr|k-j|^\rho+rp|i-k|^\rho}{p|i-\alpha|^\rho+q|j-\alpha|^\rho+r|k-\alpha|^\rho}. \tag{3.3.18}$$

Let $n\in\mathbb{N}^*$, $i=(0,0)$, $j=(1,0)$, $k=(\frac12,\frac1{n^2})$, $p=\frac1{2n}$, $q=\frac1{2n}$ and $r=1-\frac1n$. Then $|k-j|$, $|i-k|$, $|i-\alpha|$ and $|j-\alpha|$ converge to $|j-i|/2$ when $n\to+\infty$ and $|k-\alpha|=o(\frac1n)$. So

$$C_{\rho,2}=\mathbb{1}_{\{\rho=1\}}+2\frac{\frac1{4n^2}|j-i|^\rho+2\times\frac1{2n}\frac{|j-i|^\rho}{2^\rho}+o(\frac1n)}{2\times\frac1{2n}\frac{|j-i|^\rho}{2^\rho}+o(\frac1n)}\sim_{n\to+\infty}\mathbb{1}_{\{\rho=1\}}+2\frac{\frac{|j-i|^\rho}{2n}}{\frac{|j-i|^\rho}{2n}}=\mathbb{1}_{\{\rho=1\}}+2, \tag{3.3.19}$$

which shows (3.3.15) and completes the proof. □

We end the section by showing that we can estimate the constant C in (3.3.1) when we restrict to the scaling case.

Proposition 3.3.8. *Let $d\in\mathbb{N}^*$ and $\mathbb{R}^d$ be endowed with any norm. Let $\rho\geq1$, $\lambda\geq0$ and $\mu\in\mathcal{P}_\rho(\mathbb{R}^d)$ be with mean $\alpha\in\mathbb{R}^d$. Let ν be the image of μ by the map $x\mapsto x+\lambda(x-\alpha)$. Then*

$$\mathcal{M}_\rho^\rho(\mu,\nu)\leq2^{\rho-1}\frac{3+\lambda}{1+\lambda}\mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu). \tag{3.3.20}$$

Remark 3.3.9. Suppose that there exists a direction-dependent measurable map $H:\mathbb{R}^d\to\mathbb{R}^d$ in the sense of Definition 3.3.2 such that for X distributed according to μ, $\mathbb{E}[X|H(X-$

$\alpha)] = \alpha$ almost surely. Then by Proposition 3.3.3, we see that $2^{\rho-1}\frac{3+\lambda}{1+\lambda}$ could be replaced in (3.3.20) with C_ρ. In view of (3.2.12) and Remark 3.2.2, for $\rho \in (1,2)$, $C_\rho > 2^{\rho-1}$ so $2^{\rho-1}\frac{3+\lambda}{1+\lambda}$ is sharper for λ in a neighbourhood of $+\infty$. However, the smallest constant independent of λ induced by (3.3.20) is $3 \times 2^{\rho-1}$, which is greater than C_ρ, at least for $\rho \geq 2$.

Proof. For all $x \in \mathbb{R}^d$, let $m(x, dy)$ be the probability kernel defined by

$$m(x, dy) = \frac{1}{1+\lambda}\delta_{x+\lambda(x-\alpha)}(dy) + \frac{\lambda}{1+\lambda}\nu(dy).$$

For all measurable and bounded map $h : \mathbb{R}^d \to \mathbb{R}$, we have

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} h(y)\,\mu(dx)\,m(x, dy) &= \frac{1}{1+\lambda}\int_{\mathbb{R}^d} h(x+\lambda(x-\alpha))\,\mu(dx) + \frac{\lambda}{1+\lambda}\int_{\mathbb{R}^d} h(y)\,\nu(dy)\\ &= \int_{\mathbb{R}^d} h(y)\,\nu(dy).\end{aligned}$$

Moreover, for all $x \in \mathbb{R}^d$,

$$\begin{aligned}\int_{\mathbb{R}^d} y\,m(x, dy) &= \frac{1}{1+\lambda}(x+\lambda(x-\alpha)) + \frac{\lambda}{1+\lambda}\int_{\mathbb{R}^d}(x'+\lambda(x'-\alpha))\,\mu(dx')\\ &= \frac{1}{1+\lambda}(x+\lambda(x-\alpha)) + \frac{\lambda}{1+\lambda}\alpha = x.\end{aligned}$$

So $\mu(dx)\,m(x, dy)$ is a martingale coupling between μ and ν, and

$$\begin{aligned}\mathcal{M}_\rho^\rho(\mu,\nu) &\leq \int_{\mathbb{R}\times\mathbb{R}} |y-x|^\rho\,\mu(dx)\,m(x, dy)\\ &= \frac{1}{1+\lambda}\int_{\mathbb{R}^d}\lambda^\rho|x-\alpha|^\rho\,\mu(dx) + \frac{\lambda}{1+\lambda}\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy).\end{aligned}$$

On the one hand, using Lemma 3.4.2 below and the fact that $\mu(dx)\,\nu(dy)$ is a coupling between μ and ν, we have

$$\begin{aligned}\int_{\mathbb{R}^d}|x-\alpha|^\rho\,\mu(dx) &= \frac{1}{\lambda^\rho}\mathcal{W}_\rho^\rho(\mu,\nu) = \frac{1}{\lambda^\rho}\mathcal{W}_\rho(\mu,\nu)\mathcal{W}_\rho^{\rho-1}(\mu,\nu)\\ &\leq \frac{1}{\lambda^\rho}\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}.\end{aligned}$$

On the other hand, Minkowski's inequality and Lemma 3.4.2 below yield

$$\begin{aligned}&\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy)\\ &= \left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{1/\rho}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}\\ &\leq \left(\left(\int_{\mathbb{R}^d}|x-\alpha|^\rho\,\mu(dx)\right)^{1/\rho} + \left(\int_{\mathbb{R}^d}|y-\alpha|^\rho\,\nu(dy)\right)^{1/\rho}\right)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}\end{aligned}$$

$$= (2+\lambda)\left(\int_{\mathbb{R}^d} |x-\alpha|^\rho\,\mu(dx)\right)^{1/\rho}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}$$
$$= \frac{2+\lambda}{\lambda}\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}.$$

We deduce that $\mathcal{M}_\rho^\rho(\mu,\nu) \le \frac{3+\lambda}{1+\lambda}\mathcal{W}_\rho(\mu,\nu)\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho}$. Using Minkowski's inequality and the definition of convex order, for all $c\in\mathbb{R}$ we get

$$\left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |y-x|^\rho\,\mu(dx)\,\nu(dy)\right)^{(\rho-1)/\rho} \le \left(\left(\int_{\mathbb{R}^d} |x-c|^\rho\,\mu(dx)\right)^{1/\rho} + \left(\int_{\mathbb{R}^d} |y-c|^\rho\,\nu(dy)\right)^{1/\rho}\right)^{\rho-1}$$
$$\le 2^{\rho-1}\left(\int_{\mathbb{R}^d} |y-c|^\rho\,\nu(dy)\right)^{(\rho-1)/\rho}.$$

By taking the infimum over all $c\in\mathbb{R}$, we get

$$\mathcal{M}_\rho^\rho(\mu,\nu) \le 2^{\rho-1}\frac{3+\lambda}{1+\lambda}\mathcal{W}_\rho(\mu,\nu)\sigma_\rho^{\rho-1}(\nu).$$

□

3.4 Lemmas

Lemma 3.4.1. *Let $d\in\mathbb{N}^*$, $\rho\ge 1$ and $\mu,\nu\in\mathcal{P}_\rho(\mathbb{R}^d)$ be such that $\mu\le_{cx}\nu$. Let $c:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}_+$ be lower semicontinuous. Then the infimum*

$$C(\mu,\nu) = \inf_{M\in\Pi^M(\mu,\nu)}\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,M(dx,dy)$$

is attained, i.e. there exists $M\in\Pi^M(\mu,\nu)$ such that $C(\mu,\nu) = \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,M(dx,dy)$.

Proof. Let $M_n\in\Pi^{\mathrm{M}}(\mu,\nu)$, $n\in\mathbb{N}$ be a sequence of martingale couplings between μ and ν such that

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} c(x,y)\,M_n(dx,dy) \underset{n\to+\infty}{\longrightarrow} C(\mu,\nu).$$

The probability measures μ and ν are tight: for all $\varepsilon>0$ there exists a compact subset $K\subset\mathbb{R}^d$ such that $\mu(K)\ge 1-\varepsilon$ and $\nu(K)\ge 1-\varepsilon$. Therefore, for all $n\in\mathbb{N}$,

$$M_n((K\times K)^\complement) \le M_n((K^\complement\times\mathbb{R}^d)\cup(\mathbb{R}^d\times K^\complement)) \le \mu(K^\complement)+\nu(K^\complement)\le 2\varepsilon.$$

We deduce that $(M_n)_{n\in\mathbb{N}}$ is tight. By Prokhorov's theorem, there exists an increasing map $\varphi:\mathbb{N}\to\mathbb{N}$ such that $(M_{\varphi(n)})_{n\in\mathbb{N}}$ converges weakly towards $M\in\mathcal{P}(\mathbb{R}^d\times\mathbb{R}^d)$. Since the projections maps $(x,y)\mapsto x$ and $(x,y)\mapsto y$ are continuous, the respective marginals of $M_{\varphi(n)}$ converge to the respective marginals of M. Since for all $n\in\mathbb{N}$, $M_{\varphi(n)}$ has marginals μ and ν, so does M, hence $M\in\Pi(\mu,\nu)$. Moreover, for all $n\in\mathbb{N}$ let (X_n,Y_n) be a bivariate random variable distributed according to M_n and (X,Y) be distributed according to M.

Since μ and ν belong to $\mathcal{P}_1(\mathbb{R}^d)$, $((X_n, Y_n))_{n\in\mathbb{N}}$ is uniformly integrable. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous and bounded map. Since $(X_{\varphi(n)}, Y_{\varphi(n)})_{n\in\mathbb{N}}$ converges in distribution to (X, Y), is uniformly integrable and $(x, y) \mapsto f(x)(y - x)$ is continuous with at most linear growth, we have

$$\begin{aligned}0 = \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)(y-x)\, M_{\varphi(n)}(dx, dy) &= \mathbb{E}[f(X_{\varphi(n)})(Y_{\varphi(n)} - X_{\varphi(n)})] \\ \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)(Y - X)] &= \int_{\mathbb{R}^d\times\mathbb{R}^d} f(x)(y-x)\, M(dx, dy).\end{aligned}$$

We deduce that $M \in \Pi^{\mathrm{M}}(\mu, \nu)$. Then by the Portmanteau theorem, we get

$$C(\mu, \nu) \le \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y)\, M(dx, dy) \le \liminf_{n\to+\infty} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x, y)\, M_{\varphi(n)}(dx, dy) = C(\mu, \nu),$$

so M is optimal for $C(\mu, \nu)$. □

Lemma 3.4.2. *Let $d \in \mathbb{N}^*$, $\mathbb{R}^d$ be endowed with any norm, $\rho \ge 1$, $\lambda \ge 0$, $\mu \in \mathcal{P}_\rho(\mathbb{R}^d)$ and $\alpha \in \mathbb{R}^d$. Let ν be the image of μ by the map $x \mapsto x + \lambda(x - \alpha)$. Then*

$$\mathcal{W}_\rho(\mu, \nu) = \lambda \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho}. \tag{3.4.1}$$

Remark 3.4.3. Let $\eta_0, \eta_1 \in \mathcal{P}_\rho(\mathbb{R}^d)$ and $\gamma \in \Pi(\eta_0, \eta_1)$ be optimal for $\mathcal{W}_\rho(\eta_0, \eta_1)$. For all $t \in [0, 1]$, let η_t be the image of γ by $(x, y) \mapsto (1-t)x + ty$. It is well known that the curve $[0, 1] \ni t \mapsto \eta_t$ is a constant speed geodesic in $(\mathcal{P}_\rho(\mathbb{R}^d), \mathcal{W}_\rho)$ connecting η_0 to η_1 [9, Theorem 7.2.2]. Moreover, for all $0 \le s \le t \le 1$, the image of γ by $((1-s)x + sy, (1-t)x + ty)$ is an optimal transport plan between η_s and η_t for the $\mathcal{W}_\rho$-distance.

In particular for $\eta_0 = \delta_\alpha$ and $\eta_1 = \nu$, the unique coupling $\gamma(dx, dy) = \delta_\alpha(dx)\, \nu(dy)$ is optimal for $\mathcal{W}_\rho(\eta_0, \eta_1)$, and for $t = 1/(1+\lambda)$, $\eta_t = \mu$. Therefore, the image of γ by $(x, y) \mapsto ((1-t)x + ty, y)$, that is the image of μ by $x \mapsto (x, x + \lambda(x - \alpha))$, is an optimal transport plan between μ and ν for the $\mathcal{W}_\rho$-distance, which implies (3.4.1).

We add here a quick proof for the sake of completeness.

Proof of Lemma 3.4.2. We have, by the triangle inequality for the metric $\mathcal{W}_\rho$,

$$\left(\int_{\mathbb{R}^d} |y - \alpha|^\rho\, \nu(dy) \right)^{1/\rho} = \mathcal{W}_\rho(\delta_\alpha, \nu) \le \mathcal{W}_\rho(\delta_\alpha, \mu) + \mathcal{W}_\rho(\mu, \nu) = \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho} + \mathcal{W}_\rho(\mu, \nu),$$

so

$$\mathcal{W}_\rho(\mu, \nu) \ge \left(\int_{\mathbb{R}^d} |y - \alpha|^\rho\, \nu(dy) \right)^{1/\rho} - \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho} = \lambda \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho}.$$

Since $\mu(dx)\, \delta_{x+\lambda(x-\alpha)}(dy)$ is a coupling between μ and ν, we also have $\mathcal{W}_\rho(\mu, \nu) \le \lambda \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho}$, hence $\mathcal{W}_\rho(\mu, \nu) = \lambda \left(\int_{\mathbb{R}^d} |x - \alpha|^\rho\, \mu(dx) \right)^{1/\rho}$. □

Lemma 3.4.4. *Let $d \in \mathbb{N}^*$, $\lambda > 0$, $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\alpha \in \mathbb{R}^d$. Let ν be the image of μ by the map $x \mapsto x + \lambda(x - \alpha)$. Then $\mu \leq_{cx} \nu$ iff α is the mean of μ.*

Proof. If $\mu \leq_{cx} \nu$, then μ and ν have the same mean, so

$$\int_{\mathbb{R}^d} x\,\mu(dx) = \int_{\mathbb{R}^d} y\,\nu(dy) = \int_{\mathbb{R}^d} x\,\mu(dx) + \lambda\left(\int_{\mathbb{R}^d} x\,\mu(dx) - \alpha\right),$$

which implies that $\alpha = \int_{\mathbb{R}^d} x\,\mu(dx)$.

Conversely, suppose that $\alpha = \int_{\mathbb{R}^d} x\,\mu(dx)$. Then $\alpha = \int_{\mathbb{R}^d} y\,\nu(dy)$ and for all convex function $f : \mathbb{R}^d \to \mathbb{R}$,

$$\begin{aligned}\int_{\mathbb{R}^d} f(x)\,\mu(dx) &= \int_{\mathbb{R}^d} f\left(\frac{\lambda}{1+\lambda}\alpha + \frac{1}{1+\lambda}(x + \lambda(x-\alpha))\right)\mu(dx)\\ &\leq \int_{\mathbb{R}^d}\left(\frac{\lambda}{1+\lambda}f(\alpha) + \frac{1}{1+\lambda}f(x+\lambda(x-\alpha))\right)\mu(dx)\\ &= \frac{\lambda}{1+\lambda}f\left(\int_{\mathbb{R}^d} y\,\nu(dy)\right) + \frac{1}{1+\lambda}\int_{\mathbb{R}^d} f(y)\,\nu(dy)\\ &\leq \frac{\lambda}{1+\lambda}\int_{\mathbb{R}^d} f(y)\,\nu(dy) + \frac{1}{1+\lambda}\int_{\mathbb{R}^d} f(y)\,\nu(dy)\\ &= \int_{\mathbb{R}^d} f(y)\,\nu(dy),\end{aligned}$$

where we used Jensen's inequality in the last inequality. We deduce that $\mu \leq_{cx} \nu$. □

Lemma 3.4.5. *Let (E, d_E) be a Polish space, $\rho \geq 1$ and $\mathcal{P}_\rho(E)$ be the set of probability measures on E with finite ρ-th moment. Let $\mathcal{B}$, resp. $\mathcal{B}_\rho$ be the Borel σ-algebra on $\mathcal{P}_\rho(E)$ with respect to the weak convergence topology, resp. the $\mathcal{W}_\rho$-distance topology. Then $\mathcal{B} = \mathcal{B}_\rho$.*

Proof. Since the $\mathcal{W}_\rho$-distance topology is finer than the weak convergence topology, we clearly have $\mathcal{B} \subset \mathcal{B}_\rho$. Therefore it remains to prove that $\mathcal{B}_\rho \subset \mathcal{B}$.

Let $x_0 \in E$ and $\Phi_\rho(E)$ be the set of all real-valued continuous functions f on E which satisfy the growth constraint

$$\exists \alpha > 0, \quad \forall x \in E, \quad |f(x)| \leq \alpha(1 + d_E^\rho(x, x_0)).$$

For all $f \in \Phi_\rho(E)$, let $\widetilde{f} : \mathcal{P}_\rho(E) \to \mathbb{R}$ be the map defined for all $p \in \mathcal{P}_\rho(E)$ by $\widetilde{f}(p) = \int_E f(x)\,p(dx)$. The $\mathcal{W}_\rho$-distance topology is then the weak topology on $\mathcal{P}_\rho(E)$ induced by the family $(\widetilde{f})_{f \in \Phi_\rho(E)}$, that is the coarsest topology on $\mathcal{P}_\rho(E)$ for which $\widetilde{f}$ is continuous for all $f \in \Phi_\rho(E)$. Any open set for this topology is a union of finitely many intersections of sets of the form $\widetilde{f}^{-1}(U)$ where $f \in \Phi_\rho(E)$ and U is an open subset of $\mathbb{R}$. On the one hand, $(\mathcal{P}_\rho(E), \mathcal{W}_\rho)$ is Polish [191, Theorem 6.18] and therefore strongly Lindelöf, hence the latter union can be assumed at most countable. On the other hand, any open subset of $\mathbb{R}$ is an at most countable union of open intervals of $\mathbb{R}$. We deduce that any open set for the $\mathcal{W}_\rho$-distance topology is an at most countable union of finitely many intersections of at most countable unions of sets of the form $\widetilde{f}^{-1}((a,b))$ where $f \in \Phi_\rho(E)$ and $(a,b) \subset \mathbb{R}$. Since $\mathcal{B}$ is closed under countable unions and intersections, it suffices to show that every set of the form $\widetilde{f}^{-1}((a,b))$ belongs to $\mathcal{B}$ to conclude that any open set of the $\mathcal{W}_\rho$-distance topology belongs to $\mathcal{B}$ and therefore $\mathcal{B}_\rho \subset \mathcal{B}$.

Let then $f \in \Phi_\rho(E)$ and $a, b \in \mathbb{R}$ be such that $a < b$ and let us show that $\widetilde{f}^{-1}((a,b)) \in \mathcal{B}$, which will end the proof. For all $n \in \mathbb{N}$, let

$$f_n : x \mapsto (f(x) \vee (-n)) \wedge n,$$

which is clearly continuous and bounded. Then for all $n \in \mathbb{N}$ and $p \in \mathcal{P}_\rho(E)$,

$$\widetilde{f}_n(p) = \int_X ((f(x) \vee (-n)) \wedge n)\, p(dx),$$

which by the dominated convergence theorem converges to $\widetilde{f}(p)$ as $n \to +\infty$, hence

$$\widetilde{f}^{-1}((a,b)) = \bigcup_{k\in\mathbb{N}^*} \bigcup_{N\in\mathbb{N}} \bigcap_{n\geq N} \widetilde{f}_n^{-1}\left(\left(a + \frac{1}{k}, b - \frac{1}{k}\right)\right).$$

Since the weak convergence topology is induced by the family of $\widetilde{g}$ for g continuous and bounded, we have that $\widetilde{f}_n^{-1}((a,b)) \in \mathcal{B}$ for all $n \in \mathbb{N}$, hence $\widetilde{f}^{-1}((a,b)) \in \mathcal{B}$. □

Proof of Lemma 3.3.5. Let $a = (a_1, \cdots, a_d) \in \mathbb{S}^{d-1}$, $c = (c_1, \cdots, c_d) \in \mathbb{R}^d$, $y \in \operatorname{Span}(a)$ and $t \in \mathbb{R}$ be such that $y = ta$. Suppose first that $r = +\infty$. Then

$$\begin{aligned} |y - c_a| &= |ta - c_i \operatorname{sgn}(a_i) a| = |t - c_i \operatorname{sgn}(a_i)| = |t|a_i| - c_i \operatorname{sgn}(a_i)| = |(ta_i - c_i)\operatorname{sgn}(a_i)| \\ &= |ta_i - c_i| \leq |ta - c| = |y - c|. \end{aligned}$$

Suppose now that $r < +\infty$. Using the fact that $|a| = 1$ for the second and third equalities, Hölder's inequality for the second inequality and the fact that $|\operatorname{sgn}(x)| = 1$ for all $x \in \mathbb{R}$ for the last but one equality, we get

$$\begin{aligned} |y - c_a| &= \left| ta - \left(\sum_{i=1}^d c_i \operatorname{sgn}(a_i)|a_i|^{r-1}\right) a \right| \\ &= \left| t\sum_{i=1}^d |a_i|^r - \sum_{i=1}^d c_i \operatorname{sgn}(a_i)|a_i|^{r-1} \right| \\ &\leq \sum_{i=1}^d |t|a_i| - c_i \operatorname{sgn}(a_i)||a_i|^{r-1} \\ &\leq \left(\sum_{i=1}^d |t|a_i| - c_i \operatorname{sgn}(a_i)|^r\right)^{1/r} \left(\sum_{i=1}^d |a_i|^r\right)^{(r-1)/r} \\ &= \left(\sum_{i=1}^d |(ta_i - c_i)\operatorname{sgn}(a_i)|^r\right)^{1/r} \\ &= |ta - c| = |y - c|. \end{aligned}$$

□

Chapter 4

One dimensional martingale rearrangement couplings

Abstract

Backhoff-Veraguas and Pammer [20] and Wiesel [194] proved independently the stability of the Martingale Optimal Transport (MOT) problem in dimension 1 under regularity assumptions on the cost function. To do so, the former showed the stability of the so called martingale C-monotonicity, which is proved sufficient for optimality, while the latter directly tackled the primal formulation of the MOT problem. More precisely, Wiesel [194] developed the notion of projection of the set of couplings between two given marginals in the convex order onto the set of martingale couplings between the same marginals. The projection is taken with respect to the nested Wasserstein distance, which is stronger than the Wasserstein distance and therefore induces a finer topology. This leads to the notion of martingale rearrangement coupling due to Wiesel [194]. The objective of the present paper is to complement the results of Wiesel [194] by providing explicit constructions of such couplings. Moreover we establish a strong connection between those couplings and the family of martingale couplings $(M^Q)_{Q\in\mathcal{Q}}$ and the inverse transform martingale coupling introduced in Chapter 2, whose integral of $(x, y) \mapsto |y - x|$ is smaller than twice the $\mathcal{W}_1$-distance of the marginals.

Keywords: Martingale couplings, Martingale Optimal Transport, Adapted Wasserstein distance, Robust finance, Convex order.

4.1 Introduction

4.1.1 Stability of the MOT problem

Let $C : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ be a nonnegative measurable function. For all $d \in \mathbb{N}^*$ denote by $\mathcal{P}(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$. For $\mu \in \mathcal{P}(\mathbb{R})$ and $\nu \in \mathcal{P}(\mathbb{R})$, the classical Optimal Transport problem consists in minimising

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} C(x,y)\, \pi(dx,dy), \tag{OT}$$

where $\Pi(\mu,\nu)$ denotes the set of probability measures in $\mathcal{P}(\mathbb{R}\times\mathbb{R})$ with first marginal μ and second marginal ν. When C is the map $(x,y) \mapsto |y-x|^\rho$ for some $\rho \geq 1$, (OT) corresponds to the well-known Wasserstein distance with index ρ to the power ρ, denoted $\mathcal{W}_\rho^\rho(\mu,\nu)$, see [8, 168, 190, 191] for a study in depth.

The OT theory is a long story: formulated by Gaspard Monge [140] in 1781 and modernised by Kantorovich [114] in 1942, it was rediscovered many times under various forms and has an impressive scope of applications until recently where it became an unmissable tool of data sciences. However, this theory under its classical form is not sufficient to solve some major problems raised by the field of mathematical finance, such as robust model-independent pricing. Indeed, Beiglböck, Henry-Labordère and Penkner [23] showed in a discrete time setting and Galichon, Henry- Labordère and Touzi [84] in a continuous time setting that one would need an additional martingale constraint to (OT) in order to get model-free bounds of an option price. This martingale constraint reflects the condition for a financial market to be arbitrage free.

This leads to the formulation of the Martingale Optimal Transport (MOT) problem: for any $\pi \in \mathcal{P}(\mathbb{R}\times\mathbb{R})$, we denote by $(\pi_x)_{x\in\mathbb{R}}$ its disintegration with respect to its first marginal μ, that is a probability kernel such that $\pi(dx,dy) = \mu(dx)\,\pi_x(dy)$. Let μ and ν be two probability distributions on the real line with finite first moment. Then the MOT problem consists in minimising

$$\inf_{\pi \in \Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} C(x,y)\, \pi(dx,dy), \tag{MOT}$$

where $\Pi^{\mathrm{M}}(\mu,\nu)$ denotes the set of martingale couplings between μ and ν, that is

$$\Pi^{\mathrm{M}}(\mu,\nu) = \left\{ \pi \in \Pi(\mu,\nu) \mid \mu(dx)\text{-almost everywhere, } \int_{\mathbb{R}} y\, \pi_x(dy) = x \right\}.$$

According to Strassen's theorem [183], the existence of a martingale coupling between two probability measures $\mu,\nu \in \mathcal{P}(\mathbb{R})$ with finite first moment is equivalent to $\mu \leq_{cx} \nu$, where $\leq_{cx}$ denotes the convex order. We recall that two finite positive measures μ,ν on $\mathbb{R}$ with finite first moment are said to be in the convex order iff we have

$$\int_{\mathbb{R}} f(x)\,\mu(dx) \leq \int_{\mathbb{R}} f(y)\,\nu(dy),$$

for every convex function $f : \mathbb{R} \to \mathbb{R}$. Note that by evaluating this inequality for the constant function equal to 1, the identity function and their opposites, we have that μ and ν have equal mass and satisfy $\int_{\mathbb{R}} x\,\mu(dx) = \int_{\mathbb{R}} y\,\nu(dy)$.

For adaptations of celebrated results on classical optimal transport theory to the MOT problem, we refer to Beiglböck and Juillet [25], Henry-Labordère, Tan and Touzi [105] and Henry- Labordère and Touzi [106]. On duality, we refer to Beiglböck, Nutz and Touzi [27], Beiglböck, Lim and Obłój [26] and De March [68]. We also refer to De March [67] and De March and Touzi [69] for the multi-dimensional case.

About the numerical resolution of the MOT problem, one can look at Alfonsi, Corbetta and Jourdain [3, 4], De March [66], Guo and Obłój [96] and Henry-Labordère [104]. When μ and ν are finitely supported, then the MOT problem amounts to linear programming. In the general case, once the MOT problem is discretised by approximating μ and ν by probability measures with finite support and in the convex order, Alfonsi, Corbetta and Jourdain raised the question of the convergence of the discrete optimal cost towards the continuous one. Partial results were first brought by Guo and Obłój [96] and the stability of left-curtain couplings obtained by Juillet [112].

More recently, Backhoff-Veraguas and Pammer [20] gave a positive answer under mild regularity assumptions by showing the stability of the so called martingale C-monotonicity, which is proved sufficient for optimality. Independently, Wiesel [194] also gave a positive answer by directly tackling the primal formulation (MOT). To set the idea down, consider a lower semi-continuous cost function $C : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ and two sequences $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ of probability measures on the real line which respectively weakly converge to μ and ν as $n \to +\infty$. It is well known that any sequence $(\pi_n)_{n\in\mathbb{N}} \in \prod_{n\in\mathbb{N}} \Pi^{\mathrm{M}}(\mu_n, \nu_n)$ is tight and has all its accumulation points with respect to the weak convergence topology in $\Pi^{\mathrm{M}}(\mu, \nu)$. Then one can readily derive the first inequality

$$\inf_{\pi\in\Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} C(x, y)\,\pi(dx, dy) \le \liminf_{n\to+\infty} \inf_{\pi\in\Pi^{\mathrm{M}}(\mu_n,\nu_n)} \int_{\mathbb{R}\times\mathbb{R}} C(x, y)\,\pi(dx, dy).$$

The other inequality which would conclude the stability of (MOT), namely

$$\inf_{\pi\in\Pi^{\mathrm{M}}(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} C(x, y)\,\pi(dx, dy) \ge \limsup_{n\to+\infty} \inf_{\pi\in\Pi^{\mathrm{M}}(\mu_n,\nu_n)} \int_{\mathbb{R}\times\mathbb{R}} C(x, y)\,\pi(dx, dy),$$

is however substantially more difficult to prove and requires a priori stronger assumptions on the regularity of C and the convergence of the marginals. In order to get the latter inequality, Wiesel [194] develops the key notion of projection of the set $\Pi(\mu, \nu)$ of couplings between μ and ν onto the set $\Pi^{\mathrm{M}}(\mu, \nu)$ of martingale couplings between μ and ν. To measure the distance between two couplings, one could naturally think of the Wasserstein distance. However this distance turns out to be inadequate in this context for reasons explained below. Hence the projection is made with respect to the nested Wasserstein distance $\mathcal{W}_1^{nd}$ which dominates $\mathcal{W}_1$ and induces therefore a finer topology.

Wiesel [194] also complements the stability inequality $\inf_{\pi\in\Pi^{\mathrm{M}}(\mu,\nu)} |y - x|\,\pi(dx, dy) \le 2\mathcal{W}_1(\mu, \nu)$ found in Chapter 2, which was at the time a step forward in proving the stability of (MOT). The latter inequality was proved by exhibiting a new family of martingale couplings $(M^Q)_{Q\in\mathcal{Q}}$ parametrised by a set $\mathcal{Q}$ of probability measures on $(0,1)^2$. A particular martingale coupling stands out from the latter family, namely the so called inverse transform martingale coupling, which was designed to be the closest martingale coupling from the Hoeffding-Fréchet coupling. In other words, under mild assumptions, the inverse transform martingale coupling is the projection of the Hoeffding-Fréchet coupling onto the set $\Pi^{\mathrm{M}}(\mu,\nu)$. We show in Section 4.2 that in the general case, such a projection can always be found in the aforementioned family $(M^Q)_{Q\in\mathcal{Q}}$.

As mentioned by Wiesel [194, Lemma 2.3], the Hoeffding-Fréchet coupling between two probability measures $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ in the convex order can be seen as a particular element of the class of couplings $\pi\in\Pi(\mu,\nu)$ which satisfy the so called barycentre dispersion assumption introduced by Wiesel [194], that is such that

$$\forall a\in\mathbb{R},\quad \int_{\mathbb{R}} \mathbb{1}_{[a,+\infty)}(x)\left(x-\int_{\mathbb{R}} y\,\pi_x(dy)\right)\mu(dx)\le 0. \tag{4.1.1}$$

The latter assumption is important in this context since it provides a sufficient condition for a coupling between μ and ν to admit a martingale rearrangement coupling [194, Proposition 2.4]. Just like we designed in Chapter 2 a family of martingale couplings around the Hoeffding-Fréchet coupling, we can design here in the same spirit a new family of martingale couplings around any coupling π satisfying the barycentre dispersion assumption and prove that they are martingale rearrangement couplings of π. For the sake of clarity we only focus in Section 4.3 on one particular element of such a family, namely the one constructed similarly to the inverse transform martingale coupling.

4.1.2 The adapted Wasserstein distance

The topology induced by the Wasserstein distance is not always well suited for any setting, especially in mathematical finance. Indeed, the symmetry of this distance does not take into account the temporal structure of martingales. One can easily get convinced that two stochastic processes very close in Wasserstein distance can yield radically unalike information, as [14, Figure 1] illustrates very well. Therefore, one needs to strengthen, or adapt this usual topology. This can be done in many different ways, such as the adapted weak topology (see below), Hellwig's information topology [101], Aldous's extended weak topology [2] or the optimal stopping topology [15]. Strikingly, all those apparently independent topologies are actually equal, at least in discrete time [15, Theorem 1.1].

Hence we may focus on the so called adapted Wasserstein distance. For an extensive background, we refer to [149, 150, 151, 152, 129, 39]. Fix $\rho\ge 1$. For all $d\in\mathbb{N}^*$ we denote by $\mathcal{P}_\rho(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$ with finite ρ-th moment. We define the metric $\mathcal{AW}_\rho$ for all $\pi,\pi'\in\mathcal{P}_\rho(\mathbb{R}\times\mathbb{R})$ by

$$\mathcal{AW}_\rho(\pi,\pi')=\inf_{\chi\in\Pi(\mu,\mu')}\left(\int_{\mathbb{R}\times\mathbb{R}}\left(|x-x'|^\rho+\mathcal{W}_\rho^\rho(\pi_x,\pi'_{x'})\right)\chi(dx,dx')\right)^{1/\rho}, \tag{4.1.2}$$

where μ, resp. μ' is the first marginal of π, resp. π'. It is easy to check that $\mathcal{W}_\rho \leq \mathcal{AW}_\rho$, so that $\mathcal{AW}_\rho$ induces a finer topology than $\mathcal{W}_\rho$.

The adapted Wasserstein distance can also be interpreted in terms of bicausal couplings as done in [16]. Let $\pi, \pi' \in \mathcal{P}_\rho(\mathbb{R}\times\mathbb{R})$. Let X, Y, X', Y' be random variables such that the distribution of (X, Y, X', Y') is a $\mathcal{W}_\rho$-optimal coupling between π and π'. In many cases, there exists a Monge transport map $T : \mathbb{R}\times\mathbb{R} \to \mathbb{R}\times\mathbb{R}$ such that $(X', Y') = T(X, Y)$. As mentioned in [14], the temporal structure of stochastic processes is then not taken into account since the present value X' is determined from the future value Y. Therefore, it is more suitable to restrict to couplings (X, Y, X', Y') between π and π' such that the conditional distribution of (X', Y') (resp. (X, Y)) given (X, Y) (resp. (X', Y')) is equal to the conditional distribution of (X', Y') (resp. (X, Y)) given X (resp. X').

Let μ and μ' denote the respective first marginal distributions of π and π' and $\eta \in \Pi(\pi, \pi')$ be a coupling between π and π'. With a slight abuse of notation, denote by $(\eta_{(x,y)}(dx', dy'))_{(x,y)\in\mathbb{R}\times\mathbb{R}}$, $(\eta_{(x',y')}(dx, dy))_{(x',y')\in\mathbb{R}\times\mathbb{R}}$, $(\eta_x(dx', dy'))_{x\in\mathbb{R}}$ and $(\eta_{x'}(dx, dy))_{x'\in\mathbb{R}}$ the probability kernels such that

$$\begin{gathered}\eta(dx, dy, dx', dy') = \pi(dx, dy)\,\eta_{(x,y)}(dx', dy') = \pi'(dx', dy')\,\eta_{(x',y')}(dx, dy),\\ \int_{y\in\mathbb{R}} \eta(dx, dy, dx', dy') = \mu(dx)\,\eta_x(dx', dy') \quad \text{and} \quad \int_{y'\in\mathbb{R}} \eta(dx, dy, dx', dy') = \mu'(dx')\,\eta_{x'}(dx, dy).\end{gathered} \tag{4.1.3}$$

Then η is called bicausal iff for $\pi(dx, dy)$-almost every $(x, y) \in \mathbb{R}\times\mathbb{R}$, resp. $\pi(dx', dy')$-almost every $(x', y') \in \mathbb{R}\times\mathbb{R}$,

$$\eta_{(x,y)}(dx', dy') = \eta_x(dx', dy'), \quad \text{resp.} \quad \eta_{(x',y')}(dx, dy) = \eta_{x'}(dx, dy). \tag{4.1.4}$$

We denote by $\Pi_{bc}(\pi, \pi')$ the set of bicausal couplings between π and π'. Another useful characterisation, see for instance Lemma 4.4.1 below for a proof, is that η is bicausal iff there exist $\chi \in \Pi(\mu, \mu')$ and $(\gamma_{(x,x')}(dy, dy'))_{(x,x')\in\mathbb{R}\times\mathbb{R}}$ such that

$$\begin{aligned} &\chi(dx, dx')\text{-almost everywhere}, \quad \gamma_{(x,x')}(dy, dy') \in \Pi(\pi_x, \pi'_{x'}),\\ \text{and} \quad &\eta(dx, dy, dx', dy') = \chi(dx, dx')\,\gamma_{(x,x')}(dy, dy'). \end{aligned} \tag{4.1.5}$$

Therefore,

$$\begin{aligned} \left(\mathcal{W}_\rho^{nd}(\pi, \pi')\right)^\rho &:= \inf_{\eta\in\Pi_{bc}(\pi,\pi')} \int_{\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}} \left(|x - x'|^\rho + |y - y'|^\rho\right)\, \eta(dx, dy, dx', dy')\\ &= \inf_{\chi\in\Pi(\mu,\mu')} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^\rho + \inf_{\gamma_{(x,x')}\in\Pi(\pi_x,\pi'_{x'})} \int_{\mathbb{R}\times\mathbb{R}} |y - y'|^\rho\, \gamma_{(x,x')}(dy, dy')\right)\, \chi(dx, dx')\\ &= \inf_{\chi\in\Pi(\mu,\mu')} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^\rho + \mathcal{W}_\rho^\rho(\pi_x, \pi'_{x'})\right)\, \chi(dx, dx') = \mathcal{AW}_\rho^\rho(\pi, \pi'). \end{aligned} \tag{4.1.6}$$

The value $\mathcal{W}_\rho^{nd}(\pi, \pi')$ is called the nested Wasserstein distance between π and π', which by (4.1.6) coincides with the adapted Wasserstein distance $\mathcal{AW}_\rho^\rho(\pi, \pi')$ between π and π'. Hence we can indifferently work with the adapted or the nested Wasserstein distance. For the remainder of the present paper we choose to work with $\mathcal{AW}_\rho$.

4.1.3 Outline

The main result of Section 4.2 is the exhibition of an element of the family $(M^Q)_{Q\in\mathcal{Q}}$ mentioned above which is a martingale rearrangement coupling of the Hoeffding-Fréchet coupling. For the sake of completeness we recall the essentials of the family $(M^Q)_{Q\in\mathcal{Q}}$ and the Hoeffing-Fréchet coupling. We also explore the stability in $\mathcal{AW}_1$ of the inverse transform martingale coupling, which implies under certain assumptions the stability in $\mathcal{AW}_1$ of the martingale rearrangement couplings.

The study is then extended in Section 4.3 to martingale rearrangement couplings of couplings which satisfy the barycentre dispersion assumption (4.1.1).

Finally we deferred to Section 4.4 the proof of the key lemma which implies equality between the adapted and the nested Wasserstein distances.

4.2 Martingale rearrangement couplings of the Hoeffding-Fréchet coupling

4.2.1 The inverse transform martingale coupling

We come back on the inverse transform martingale coupling and the family parametrised by $\mathcal{Q}$ introduced in Chapter 2 since they will have particular significance in the present paper. We briefly recall the construction and main properties, see Chapter 2 for an extensive study. Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ be such that $\mu\le_{cx}\nu$ and $\mu\neq\nu$. For $u\in[0,1]$ we define

$$\Psi_+(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^+(v)\,dv\quad\text{and}\quad\Psi_-(u)=\int_0^u(F_\mu^{-1}-F_\nu^{-1})^-(v)\,dv,\tag{4.2.1}$$

with respective left continuous generalised inverses Ψ_+^{-1} and Ψ_-^{-1}. We then define $\mathcal{Q}$ as the set of probability measures on $(0,1)^2$ with first marginal $\frac{1}{\Psi_+(1)}d\Psi_+$, second marginal $\frac{1}{\Psi_+(1)}d\Psi_-$ and such that $u<v$ for $Q(du,dv)$-almost every $(u,v)\in(0,1)^2$. Since $d\,\Psi_+$ and $d\,\Psi_-$ are concentrated on two disjoint Borel sets, there exists for each $Q\in\mathcal{Q}$ a probability kernel $(\pi_u^Q)_{u\in(0,1)}$ such that

$$Q(du,dv)=\frac{1}{\Psi_+(1)}d\Psi_+(u)\,\pi_u^Q(dv)=\frac{1}{\Psi_+(1)}d\Psi_-(v)\,\pi_v^Q(du),\tag{4.2.2}$$

and we exhibit a probability kernel $(\widetilde{m}_u^Q)_{u\in(0,1)}$ which satisfies for du-almost all $u\in(0,1)$ such that $F_\mu^{-1}(u)\neq F_\nu^{-1}(u)$

$$\widetilde{m}_u^Q(dy)=\int_{v\in(0,1)}\left(\frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\delta_{F_\nu^{-1}(v)}(dy)+\frac{F_\nu^{-1}(v)-F_\mu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\delta_{F_\nu^{-1}(u)}(dy)\right)\pi_u^Q(dv),\tag{4.2.3}$$

and $\widetilde{m}_u^Q(dy)=\delta_{F_\nu^{-1}(u)}(dy)$ for all $u\in(0,1)$ such that $F_\mu^{-1}(u)=F_\nu^{-1}(u)$. Then the measure

$$M^Q(dx,dy)=\int_0^1\delta_{F_\mu^{-1}(u)}(dx)\,\widetilde{m}_u^Q(dy)\,du\tag{4.2.4}$$

is a martingale coupling between μ and ν which satisfies $\int_{\mathbb{R}\times\mathbb{R}} |y-x|\, M^Q(dx,dy) \leq 2\mathcal{W}_1(\mu,\nu)$. The latter fact is based on the property that for du-almost all $u\in(0,1)$,

$$\int_{\mathbb{R}} |y - F_\nu^{-1}(u)|\, \widetilde{m}_u^Q(dy) = |F_\mu^{-1}(u) - F_\nu^{-1}(u)|, \tag{4.2.5}$$

as showed by Proposition 2.2.18 and its proof. Let also

$$\mathcal{U}^+ = \{u\in(0,1) \mid F_\mu^{-1}(u) > F_\nu^{-1}(u)\},\quad \mathcal{U}^- = \{u\in(0,1) \mid F_\mu^{-1}(u) < F_\nu^{-1}(u)\}, \tag{4.2.6}$$

$$\text{and}\quad \mathcal{U}^0 = \{u\in(0,1) \mid F_\mu^{-1}(u) = F_\nu^{-1}(u)\}. \tag{4.2.7}$$

Thanks to the equality $\Psi_+(1) = \Psi_-(1)$, consequence of the equality of the respective means of μ and ν, we can set for all $u\in[0,1]$

$$\varphi(u) = \begin{cases} \Psi_-^{-1}(\Psi_+(u)) & \text{if} \quad u\in\mathcal{U}^+; \\ \Psi_+^{-1}(\Psi_-(u)) & \text{if} \quad u\in\mathcal{U}^-; \\ u & \text{if} \quad u\in\mathcal{U}^0. \end{cases} \tag{4.2.8}$$

Then the measure $Q^{IT}(du,dv) = \frac{1}{\Psi_+(1)} d\Psi_+(u) \mathbb{1}_{\{0<\varphi(u)<1\}}\, \delta_{\varphi(u)}(dv)$ belongs to $\mathcal{Q}$. The martingale coupling $M^{IT} = M^{Q^{IT}}$ is the so called inverse transform martingale coupling, associated to the probability kernel $\widetilde{m}^{IT} = \widetilde{m}^{Q^{IT}}$ which satisfies for du-almost all $u\in(0,1)$

$$\widetilde{m}^{IT}(u,dy) = \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)} \delta_{F_\nu^{-1}(\varphi(u))}(dy) + \left(1 - \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)}\right) \delta_{F_\nu^{-1}(u)}(dy). \tag{4.2.9}$$

We also recall some standard results about cumulative distribution functions and quantile functions since they will prove very handy one-dimensional tools. Proofs can be found for instance in Section 2.6. For any probability measure η on $\mathbb{R}$:

(1) F_η, resp. F_η^{-1}, is right continuous, resp. left continuous, and nondecreasing;

(2) For all $(x,u)\in\mathbb{R}\times(0,1)$,

$$F_\eta^{-1}(u) \leq x \iff u \leq F_\eta(x), \tag{4.2.10}$$

which implies

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{4.2.11}$$

$$\text{and}\quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{4.2.12}$$

(3) For $\eta(dx)$-almost every $x\in\mathbb{R}$,

$$0 < F_\eta(x),\quad F_\eta(x-) < 1 \quad\text{and}\quad F_\eta^{-1}(F_\eta(x)) = x; \tag{4.2.13}$$

(4) Denoting by $\lambda_{(0,1)}$, resp. $\lambda_{(0,1)^2}$, the Lebesgue measure on $(0,1)$, resp. $(0,1)^2$, we have

$$\Big((u,v) \mapsto F_\eta(F_\eta^{-1}(u)-) + v\eta(\{F_\eta^{-1}(u)\})\Big)_\sharp \lambda_{(0,1)^2} = \lambda_{(0,1)}, \tag{4.2.14}$$

where $\sharp$ denotes the pushforward operation.

(5) The image of the Lebesgue measure on $(0,1)$ by F_η^{-1} is η.

The property (5) is referred to as inverse transform sampling.

4.2.2 The Hoeffding-Fréchet coupling

Let μ and ν be two probability measures on the real line with finite first moment. We recall that the Hoeffding-Fréchet coupling between μ and ν, denoted P^{HF}, is by definition the comonotonic coupling between μ and ν, that is the image of the Lebesgue measure on $(0,1)$ by the map $u \mapsto (F_\mu^{-1}(u), F_\nu^{-1}(u))$. Equivalently, we can write

$$P^{HF}(dx,dy) = \int_{(0,1)} \delta_{(F_\mu^{-1}(u),F_\nu^{-1}(u))}(dx,dy)\,du.$$

This coupling is of paramount importance in the classical optimal transport theory in dimension 1 since it minimises

$$\inf_{P\in\Pi(\mu,\nu)} \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\,P(dx,dy)$$

as soon as c satisfies the so called Monge condition, see [161, Theorem 3.1.2]. The latter condition being satisfied for any function $(x,y) \mapsto h(|y-x|)$ where $h : \mathbb{R} \to \mathbb{R}$ is convex, we deduce that P^{HF} is optimal for $\mathcal{W}_\rho(\mu,\nu)$ for all $\rho \geq 1$. By strict convexity, it is even the only coupling optimal for $\mathcal{W}_\rho(\mu,\nu)$ for $\rho > 1$. For all $(x,v) \in \mathbb{R}\times[0,1]$, let

$$\theta(x,v) = F_\mu(x-) + v\mu(\{x\}).$$

Using (4.2.14) for the second equality, (4.2.11) for the third one and the inverse transform sampling for the fourth one, we get

$$\begin{aligned}
\int_{\mathbb{R}\times\mathbb{R}} f(x,y)\,P^{HF}(dx,dy) &= \int_{(0,1)} f(F_\mu^{-1}(u), F_\nu^{-1}(u))\,du \\
&= \int_{(0,1)^2} f(F_\mu^{-1}(\theta(F_\mu^{-1}(u),v)), F_\nu^{-1}(\theta(F_\mu^{-1}(u),v)))\,du\,dv \\
&= \int_{(0,1)^2} f(F_\mu^{-1}(u), F_\nu^{-1}(\theta(F_\mu^{-1}(u),v)))\,du\,dv \\
&= \int_{\mathbb{R}\times(0,1)} f(x, F_\nu^{-1}(\theta(x,v)))\,\mu(dx)\,dv \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}\times(0,1)} f(x,y)\,\delta_{F_\nu^{-1}(\theta(x,v))}(dy)\,dv\right)\mu(dx).
\end{aligned}$$

We deduce that for $\mu(dx)$-almost all $x \in \mathbb{R}$,

$$P_x^{HF}(dy) = \int_{(0,1)} \delta_{F_\nu^{-1}(F_\mu(x-)+v\mu(\{x\}))\}}(dy)\,dv. \tag{4.2.15}$$

By (4.2.15) and monotonicity and left continuity of F_ν^{-1} we recover the well known fact that P^{HF} is given by a measurable map, i.e. is the image of μ by $x \mapsto (x, T(x))$ where $T : \mathbb{R} \to \mathbb{R}$ is measurable, iff for all $x \in \mathbb{R}$ such that $\mu(\{x\}) > 0$, F_ν^{-1} is constant on $(F_\mu(x-), F_\mu(x)]$. In that case, we have $T = F_\nu^{-1} \circ F_\mu$, referred to as the Monge transport map.

4.2.3 Martingale rearrangement couplings

The inverse transform martingale coupling was meant to be in a certain way the closest martingale coupling from the Hoeffding-Fréchet coupling, the latter being well known for minimising the Wasserstein distance. In order to formulate the latter assertion in a more formal way, suppose first that two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ are in the convex order and such that F_ν^{-1} is constant on the intervals of the form $(F_\mu(x-), F_\mu(x)]$ for $x \in \mathbb{R}$. As mentioned above, the Hoeffding-Fréchet coupling P^{HF} between μ and ν is then concentrated on the graph of the Monge transport map $T = F_\nu^{-1} \circ F_\mu$. We denote respectively by $M^{IT} \in \Pi^{\mathrm{M}}(\mu, \nu)$ and $M \in \Pi^{\mathrm{M}}(\mu, \nu)$ the inverse transform martingale coupling between μ and ν and any martingale coupling between μ and ν. We also denote by $\chi \in \Pi(\mu, \mu)$ a coupling between μ and μ optimal for $\mathcal{AW}_1(P^{HF}, M^{IT})$. We deduce from (4.2.5) that $\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M^{IT}(dx, dy) = \mathcal{W}_1(\mu, \nu)$. Using this equality in the second equality, we get

$$\begin{aligned}
\mathcal{AW}_1(P^{HF}, M^{IT}) &\le \int_{\mathbb{R}} \mathcal{W}_1(\delta_{T(x)}, M_x^{IT})\,\mu(dx) = \int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M^{IT}(dx, dy) = \mathcal{W}_1(\mu, \nu) \\
&= \int_{\mathbb{R}} |x - T(x)|\,\mu(dx) \le \int_{\mathbb{R}\times\mathbb{R}} (|x - x'| + |x' - T(x)|)\; \chi(dx, dx') \\
&= \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'| + \left|\int_{\mathbb{R}} y\, M_{x'}(dy) - T(x)\right|\right)\; \chi(dx, dx') \\
&\le \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'| + \mathcal{W}_1(\delta_{T(x)}, M_{x'})\right)\; \chi(dx, dx') \\
&= \mathcal{AW}_1(P^{HF}, M).
\end{aligned} \tag{4.2.16}$$

We deduce that M^{IT} is a martingale rearrangement coupling of P^{HF} in the sense of Wiesel [194], that is

$$\mathcal{AW}_1(P^{HF}, M^{IT}) = \inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \mathcal{AW}_1(P^{HF}, M). \tag{4.2.17}$$

When P^{HF} is not concentrated on the graph of a measurable map, M^{IT} is not necessarily a martingale rearrangement coupling. However we show in the next proposition that there exists a martingale rearrangement coupling in the family of martingale couplings parametrised by $\mathcal{Q}$.

Proposition 4.2.1. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$. Let P^{HF} be the Hoeffding-Fréchet coupling between μ and ν. Then there exists $Q \in \mathcal{Q}$ such that the martingale coupling M^Q defined by* (4.2.4) *is a martingale rearrangement coupling of P^{HF}:*

$$\mathcal{AW}_1(P^{HF}, M^Q) = \inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \mathcal{AW}_1(P^{HF}, M).$$

Remark 4.2.2. We show in the proof that as soon as on each interval $(F_\mu(x-), F_\mu(x)]$ for $x \in \mathbb{R}$ the sign of $u \mapsto F_\mu^{-1}(u) - F_\nu^{-1}(u)$ is constant, the inverse transform martingale coupling is a martingale rearrangement coupling of P^{HF}. Of course this includes the case where P^{HF} is concentrated on the graph of a measurable map, and in which we recover (4.2.17).

Proof. By (4.1.6) and [194, Lemma 2.1] we have

$$\inf_{M \in \Pi^{\mathrm{M}}(\mu,\nu)} \mathcal{AW}_1(P^{HF}, M) \geq \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y \, P_x^{HF}(dy) \right| \mu(dx),$$

hence it is sufficient to show that there exists $Q \in \mathcal{Q}$ such that

$$\mathcal{AW}_1(P^{HF}, M^Q) \leq \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y \, P_x^{HF}(dy) \right| \mu(dx). \tag{4.2.18}$$

If $\mu = \nu$ then the statement is straightforward, hence we suppose $\mu \neq \nu$. The proof is achieved in four steps. First we exhibit an appropriate subdivision of the intervals $(0,1)$ in order to define a measure Q on $(0,1)^2$. Second we show that Q belongs to $\mathcal{Q}$ and is therefore associated to the martingale coupling M^Q between μ and ν. Then we find for $\mu(dx)$-almost all $x \in \mathbb{R}$ a coupling $\eta_x \in \Pi(P_x^{HF}, M_x^Q)$, so that

$$\mathcal{AW}_1(P^{HF}, M^Q) \leq \int_{\mathbb{R}} \mathcal{W}_1(P_x^{HF}, M_x^Q) \, \mu(dx) \leq \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |y - y'| \, \eta_x(dy, dy') \right) \mu(dx). \tag{4.2.19}$$

Last, we show that for $\mu(dx)$-almost all $x \in \mathbb{R}$,

$$\int_{\mathbb{R} \times \mathbb{R}} |y - y'| \, \eta_x(dy, dy') = \left| x - \int_{\mathbb{R}} y \, P_x^{HF}(dy) \right|, \tag{4.2.20}$$

which implies (4.2.18) and completes the proof.

Step 1. Recall the definitions of Ψ_+, Ψ_-, $\mathcal{U}^+$ and $\mathcal{U}^-$ from (4.2.1) and (4.2.6). Let $A_\mu = \{x \in \mathbb{R} \mid \mu(\{x\}) > 0\}$. For all $x \in A_\mu$, let $\mathcal{U}_x = (F_\mu(x-), F_\mu(x)]$, $\mathcal{U}_x^+ = \mathcal{U}^+ \cap \mathcal{U}_x$, $\mathcal{U}_x^- = \mathcal{U}^- \cap \mathcal{U}_x$ and

$$\begin{aligned}
a_x = \inf \Bigg\{ a \in [F_\mu(x-), F_\mu(x)] \mid \int_{F_\mu(x-)}^{a} (x - F_\nu^{-1}(u))^+ \, du = \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^+ \, du \right) \\
\wedge \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du \right) \Bigg\},
\end{aligned}$$

$$b_x = \sup \Bigg\{ b \in [F_\mu(x-), F_\mu(x)] \mid \int_{b}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du = \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^+ \, du \right)$$

$$\wedge \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du \right) \Big\},$$

$\mathcal{V}_x^+ = (F_\mu(x-), a_x]$, $\quad \mathcal{V}_x^- = (b_x, F_\mu(x)]$, $\quad \widetilde{\mathcal{U}}^+ = \mathcal{U}^+ \backslash (\bigcup_{x \in A_\mu} \mathcal{V}_x^+)$, $\quad$ and $\quad \widetilde{\mathcal{U}}^- = \mathcal{U}^- \backslash (\bigcup_{x \in A_\mu} \mathcal{V}_x^-)$.

By monotonicity of $u \mapsto x - F_\nu^{-1}(u)$ and by definition of a_x and b_x, we have $F_\mu(x-) \leq a_x \leq b_x \leq F_\mu(x)$,

$$\begin{aligned} \int_{\mathcal{V}_x^+} (x - F_\nu^{-1}(u)) \, du &= \int_{\mathcal{V}_x^-} (x - F_\nu^{-1}(u)) \, du \\ &= \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^+ \, du \right) \wedge \left(\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du \right), \end{aligned} \tag{4.2.21}$$

$\mathcal{V}_x^+ \subset \mathcal{U}_x^+$ and $\mathcal{V}_x^- \subset \mathcal{U}_x^-$. Moreover, we have

$$\begin{aligned} &\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^+ \, du \geq \int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du \iff (F_\mu(x-), b_x] \cap \mathcal{U}_x^- = \emptyset \iff \mathcal{V}_x^- = \mathcal{U}_x^-, \\ &\int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^+ \, du \leq \int_{F_\mu(x-)}^{F_\mu(x)} (x - F_\nu^{-1}(u))^- \, du \iff (a_x, F_\mu(x)] \cap \mathcal{U}_x^+ = \emptyset \iff \mathcal{V}_x^+ = \mathcal{U}_x^+. \end{aligned} \tag{4.2.22}$$

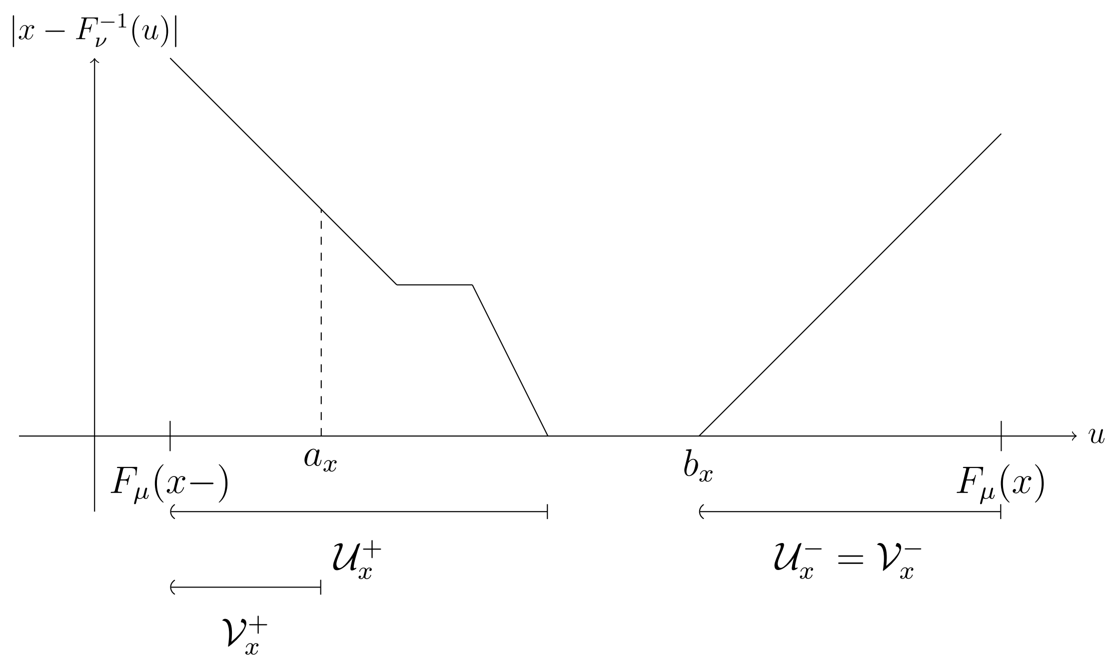

Figure 4.1: Points and intervals involved in the proof in the case where $\int_{\mathcal{U}_x} (x - F_\nu^{-1}(u))^+ \, du > \int_{\mathcal{U}_x} (x - F_\nu^{-1}(u))^- \, du$.

For $x \in A_\mu$, let Q_x be any measure on $(0,1)^2$ such that its first and second marginal are respectively

$$\frac{1}{\Psi_+(1)}\mathbb{1}_{\mathcal{V}_x^+}(u)\,(x - F_\nu^{-1}(u))^+\,du \quad \text{and} \quad \frac{1}{\Psi_+(1)}\mathbb{1}_{\mathcal{V}_x^-}(v)\,(x - F_\nu^{-1}(v))^-\,dv. \tag{4.2.23}$$

Notice that $a_x \le b_x$ implies

$$Q_x(\{(u,v) \in (0,1)^2 \mid u < v\}) = Q_x((0,1)^2). \tag{4.2.24}$$

Let now $\chi_+, \chi_- : [0,1] \to \mathbb{R}$ be defined for all $u \in [0,1]$ by

$$\chi_+(u) = \int_0^u (F_\mu^{-1} - F_\nu^{-1})^+(v)\mathbb{1}_{\widetilde{\mathcal{U}}^+}(v)\,dv \quad \text{and} \quad \chi_-(u) = \int_0^u (F_\mu^{-1} - F_\nu^{-1})^-(v)\mathbb{1}_{\widetilde{\mathcal{U}}^-}(v)\,dv.$$

For all $u \in [0,1]$, $[0,u] \cap \mathcal{U}^+$, resp. $[0,u] \cap \mathcal{U}^-$, is the disjoint union of $[0,u] \cap \widetilde{\mathcal{U}}^+$ and $[0,u] \cap \left(\bigcup_{x \in A_\mu} \mathcal{V}_x^+\right)$, resp. $[0,u] \cap \widetilde{\mathcal{U}}^-$ and $[0,u] \cap \left(\bigcup_{x \in A_\mu} \mathcal{V}_x^-\right)$. Therefore, for all $u \in [0,1]$,

$$\begin{aligned} \Psi_+(u) &= \chi_+(u) + \sum_{x \in A_\mu} \int_{[0,u] \cap \mathcal{V}_x^+} (F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv, \\ \text{and} \quad \Psi_-(u) &= \chi_-(u) + \sum_{x \in A_\mu} \int_{[0,u] \cap \mathcal{V}_x^-} (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv. \end{aligned} \tag{4.2.25}$$

Applying (4.2.25) with $u = 1$, (4.2.21) and the equality $\Psi_+(1) = \Psi_-(1)$, we get $\chi_+(1) = \chi_-(1)$, hence we can define the map $\Gamma : [0,1] \to [0,1]$ for all $u \in [0,1]$ by

$$\Gamma(u) = \begin{cases} \chi_-^{-1}(\chi_+(u)) & \text{if } u \in \widetilde{\mathcal{U}}^+; \\ \chi_+^{-1}(\chi_-(u)) & \text{if } u \in \widetilde{\mathcal{U}}^-; \\ u & \text{otherwise.} \end{cases}$$

Let then $\widetilde{Q}$ be the measure on $(0,1)^2$ defined by

$$\widetilde{Q}(du, dv) = \frac{1}{\Psi_+(1)}\,d\chi_+(u)\,\delta_{\Gamma(u)}(dv). \tag{4.2.26}$$

Step 2. We now show that the measure Q on $(0,1)^2$ defined by

$$Q = \widetilde{Q} + \sum_{x \in A_\mu} Q_x$$

is an element of $\mathcal{Q}$. Notice that if for all $x \in \mathbb{R}$, $u \mapsto F_\mu^{-1}(u) - F_\nu^{-1}(u)$ does not change sign on $\mathcal{U}_x$, then $\mathcal{V}_x^+ = \mathcal{V}_x^- = \emptyset$. In that case, $\widetilde{\mathcal{U}}^+ = \mathcal{U}^+$ and $\widetilde{\mathcal{U}}^- = \mathcal{U}^-$, hence $Q = \widetilde{Q} = Q^{IT}$, which proves the statement of Remark 4.2.2, provided that we complete the present proof.

To prove that $Q \in \mathcal{Q}$ we begin to show that

$$Q(\{(u,v) \in (0,1)^2 \mid u < v\}) = Q((0,1)^2). \tag{4.2.27}$$

In view of (4.2.24) it suffices to show that

$$\widetilde{Q}(\{(u,v) \in (0,1)^2 \mid u < v\}) = \widetilde{Q}((0,1)^2),$$

which by (4.2.26) is equivalent to

$$\Gamma(u) > u, \quad d\chi_+(u)\text{-almost everywhere.} \tag{4.2.28}$$

Let $u \in [0,1]$. Suppose first that $u \notin B_\mu := \bigcup_{x \in A_\mu} (F_\mu(x-), F_\mu(x))$. Then for all $x \in A_\mu$, we have either $u \le F_\mu(x-)$ or $F_\mu(x) \le u$, and since $\mathcal{V}_x^+, \mathcal{V}_x^- \subset \mathcal{U}_x$, either $\mathcal{V}_x^+, \mathcal{V}_x^- \subset (u,1)$ or $\mathcal{V}_x^+, \mathcal{V}_x^- \subset [0,u]$. Equivalently,

$$[0,u] \cap \mathcal{V}_x^+ = \mathcal{V}_x^+ \quad \text{and} \quad [0,u] \cap \mathcal{V}_x^- = \mathcal{V}_x^-, \quad \text{or} \quad [0,u] \cap \mathcal{V}_x^+ = [0,u] \cap \mathcal{V}_x^- = \emptyset.$$

If $[0,u] \cap \mathcal{V}_x^+ = \mathcal{V}_x^+$ and $[0,u] \cap \mathcal{V}_x^- = \mathcal{V}_x^-$, (4.2.21) yields

$$\begin{aligned}\int_{[0,u]\cap\mathcal{V}_x^+} (F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv &= \int_{\mathcal{V}_x^+} (x - F_\nu^{-1}(v))\,dv = \int_{\mathcal{V}_x^-} (x - F_\nu^{-1}(v))\,dv \\ &= \int_{[0,u]\cap\mathcal{V}_x^-} (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv.\end{aligned}$$

Else if $[0,u] \cap \mathcal{V}_x^+ = [0,u] \cap \mathcal{V}_x^- = \emptyset$, then we clearly have $\int_{[0,u]\cap\mathcal{V}_x^+} (F_\mu^{-1} - F_\nu^{-1})^+(v)\,dv = \int_{[0,u]\cap\mathcal{V}_x^-} (F_\mu^{-1} - F_\nu^{-1})^-(v)\,dv$ too. We then deduce from (4.2.25) that $\chi_+(u) - \chi_-(u) = \Psi_+(u) - \Psi_-(u)$. By (2.3.8), $\Psi_+(u) > \Psi_-(u)$ for du-almost every $u \in \mathcal{U}^+$ and therefore $u \in \widetilde{\mathcal{U}}^+$. We deduce that

$$\chi_+(u) \ge \chi_-(u), \quad \text{for all } u \in [0,1] \backslash B_\mu, \text{ and the inequality is strict for } du\text{-almost every } u \in \widetilde{\mathcal{U}}^+ \backslash B_\mu. \tag{4.2.29}$$

Suppose now that $u \in \widetilde{\mathcal{U}}^+ \cap B_\mu$. Then there exists $x \in A_\mu$ such that $u \in (F_\mu(x-), F_\mu(x))$, or equivalently $u \in (a_x, F_\mu(x)) \cap \mathcal{U}^+$. Since F_μ^{-1} and F_ν^{-1} are left continuous, we have for $\varepsilon > 0$ small enough $[u-\varepsilon, u] \subset \mathcal{U}^+$ and of course $[u-\varepsilon, u] \subset (a_x, F_\mu(x))$, hence $[u-\varepsilon, u] \subset \widetilde{\mathcal{U}}^+$ and

$$\chi_+(u) > \chi_+(u-\varepsilon) \ge \chi_+(F_\mu(x-)).$$

Using (4.2.29) with $F_\mu(x-) \in [0,1] \backslash B_\mu$, we get $\chi_+(F_\mu(x-)) \ge \chi_-(F_\mu(x-))$. Moreover, the existence of u implies that $\mathcal{V}_x^+ \neq \mathcal{U}_x^+$ and therefore $\mathcal{V}_x^- = \mathcal{U}_x^-$ by (4.2.22), hence $\chi_-(u) = \chi_-(F_\mu(x-))$. We deduce that

$$\chi_+(u) > \chi_-(u), \quad \text{for all } u \in \widetilde{\mathcal{U}}^+ \cap B_\mu. \tag{4.2.30}$$

Since for all $u \in (0,1)$, $\chi_+(u) > \chi_-(u) \iff \Gamma(u) > u$, (4.2.28) follows directly from (4.2.29) and (4.2.30), which proves (4.2.27).

We now show that Q has the right marginals. On the one hand, $\mathcal{U}^+ = \widetilde{\mathcal{U}}^+ \cup \left(\bigcup_{x\in A_\mu} \mathcal{V}_x^+\right)$ where the unions are disjoint, hence its first marginal is

$$\begin{aligned}
&\frac{1}{\Psi_+(1)} d\chi_+(u) + \sum_{x\in A_\mu} \frac{1}{\Psi_+(1)} \mathbb{1}_{\mathcal{V}_x^+}(u)(x - F_\nu^{-1}(u))^+ \, du \\
&= \frac{1}{\Psi_+(1)} \left((F_\mu^{-1} - F_\nu^{-1})^+(u) \mathbb{1}_{\widetilde{\mathcal{U}}^+}(u) \, du + \sum_{x\in A_\mu} (F_\mu^{-1} - F_\nu^{-1})^+(u) \mathbb{1}_{\mathcal{V}_x^+}(u) \, du \right) \\
&= \frac{1}{\Psi_+(1)} (F_\mu^{-1} - F_\nu^{-1})^+(u) \mathbb{1}_{\mathcal{U}^+}(u) \, du = \frac{1}{\Psi_+(1)} d\Psi_+(u).
\end{aligned} \tag{4.2.31}$$

On the other hand, by Lemma 2.6.1 applied with $f_1 = (F_\mu^{-1} - F_\nu^{-1})^+ \mathbb{1}_{\widetilde{\mathcal{U}}^+}$, $f_2 = (F_\mu^{-1} - F_\nu^{-1})^- \mathbb{1}_{\widetilde{\mathcal{U}}^-}$, $u_0 = 1$ and $h : u \mapsto \mathbb{1}_{\{u \notin \widetilde{\mathcal{U}}^-\}}$, we get

$$\int_0^1 \mathbb{1}_{\{\Gamma(u)\notin\widetilde{\mathcal{U}}^-\}} \, d\chi_+(u) = \int_0^1 \mathbb{1}_{\{v\notin\widetilde{\mathcal{U}}^-\}} \, d\chi_-(v) = 0,$$

hence $\Gamma(u) \in \widetilde{\mathcal{U}}^-$ for du-almost all $u \in \widetilde{\mathcal{U}}^+$. By continuity of χ_- we have $\chi_-(\chi_-^{-1}(u)) = u$ for all $u \in [0, \chi_-(1)]$. Moreover using (4.2.13) after an appropriate normalisation, we get $\chi_+^{-1}(\chi_+(u)) = u$ for du-almost $u \in \widetilde{\mathcal{U}}^+$. We deduce that

$$u = \Gamma(\Gamma(u)), \quad du\text{-almost everywhere on } \widetilde{\mathcal{U}}^+. \tag{4.2.32}$$

Let $H : (0,1)^2 \to \mathbb{R}$ be a measurable and bounded map. Applying Lemma 2.6.1 with $f_1 = (F_\mu^{-1} - F_\nu^{-1})^+ \mathbb{1}_{\widetilde{\mathcal{U}}^+}$, $f_2 = (F_\mu^{-1} - F_\nu^{-1})^- \mathbb{1}_{\widetilde{\mathcal{U}}^-}$, $u_0 = 1$ and $h : u \mapsto H(\Gamma(u), u)$ then yields

$$\begin{aligned}
\int_{(0,1)^2} H(u,v) \, \widetilde{Q}(du, dv) &= \frac{1}{\Psi_+(1)} \int_0^1 H(u, \Gamma(u)) \, d\chi_+(u) \\
&= \frac{1}{\Psi_+(1)} \int_0^1 H(\Gamma(\Gamma(u)), \Gamma(u))(F_\mu^{-1} - F_\nu^{-1})^+(u) \mathbb{1}_{\widetilde{\mathcal{U}}^+}(u) \, du \\
&= \frac{1}{\Psi_+(1)} \int_0^1 H(\Gamma(v), v)(F_\mu^{-1} - F_\nu^{-1})^-(v) \mathbb{1}_{\widetilde{\mathcal{U}}^-}(v) \, dv \\
&= \frac{1}{\Psi_-(1)} \int_0^1 H(\Gamma(v), v) \, d\chi_-(v).
\end{aligned}$$

We deduce that

$$\widetilde{Q}(du, dv) = \frac{1}{\Psi_+(1)} d\chi_-(v) \, \delta_{\Gamma(v)}(du), \tag{4.2.33}$$

and we show with a calculation similar to (4.2.31) that its second marginal is $\frac{1}{\Psi_+(1)} d\Psi_-$. We conclude that $Q \in \mathcal{Q}$.

Step 3. For all $x \in \mathbb{R}$, let $(\eta_x(dy, dy'))_{x\in\mathbb{R}}$ be the probability kernel defined by

$$\begin{cases} \delta_x(dy)\,\delta_x(dy') \\ \qquad \text{if } F_\mu(x) = 0 \text{ or } F_\mu(x-) = 1; \\ \dfrac{1}{\mu(\{x\})}\left(\displaystyle\int_{u\in\mathcal{V}_x^+\cup\mathcal{V}_x^-} \widetilde{m}_u^Q(dy')\,\delta_{y'}(dy)\,du + \int_{u\in(a_x,b_x)} \widetilde{m}_u^Q(dy')\,\delta_{F_\nu^{-1}(u)}(dy)\,du\right) \\ \qquad \text{if } \mu(\{x\}) > 0; \\ \delta_{F_\nu^{-1}(F_\mu(x))}(dy)\,\widetilde{m}_{F_\mu(x)}^Q(dy') \\ \qquad \text{otherwise,} \end{cases}$$

where the probability kernel $(\widetilde{m}_u^Q)_{u\in(0,1)}$ is given by (4.2.3). Notice that in view of the definition (4.2.4) of M^Q, one can check that for $\mu(dx)$-almost all $x \in \mathbb{R}$,

$$M_x^Q(dy) = \begin{cases} \delta_x(dy) & \text{if } F_\mu(x) = 0 \text{ or } F_\mu(x_-) = 1; \\ \dfrac{1}{\mu(\{x\})}\displaystyle\int_{u=F_\mu(x-)}^{F_\mu(x)} \widetilde{m}_u^Q(dy)\,du & \text{if } \mu(\{x\}) > 0; \\ \widetilde{m}_{F_\mu(x)}^Q(dy) & \text{otherwise.} \end{cases} \tag{4.2.34}$$

Let us show that for $\mu(dx)$-almost all $x \in \mathbb{R}$, $\eta_x(dy, dy')$ is a coupling between P_x^{HF} and M_x^Q. Let $x \in \mathbb{R}$. By (4.2.13) we may suppose without loss of generality that $F_\mu(x) > 0$ and $F_\mu(x-) < 1$. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable and bounded map. Suppose first that $\mu(\{x\}) = 0$. Then by (4.2.15) for the second equality,

$$\begin{aligned} &\int_{\mathbb{R}\times\mathbb{R}} h(y)\,\eta_x(dy, dy') = h(F_\nu^{-1}(F_\mu(x))) = \int_{\mathbb{R}} h(y)\,P_x^{HF}(dy), \\ \text{and}\quad &\int_{\mathbb{R}\times\mathbb{R}} h(y')\,\eta_x(dy, dy') = \int_{\mathbb{R}} h(y')\,\widetilde{m}_{F_\mu(x)}^Q(dy') = \int_{\mathbb{R}} h(y')\,M_x^Q(dy'). \end{aligned}$$

Suppose now that that $\mu(\{x\}) > 0$. On the one hand, using the fact that $\mathcal{V}_x^+ \cup \mathcal{V}_x^- \cup (a_x, b_x] = (F_\mu(x-), F_\mu(x)]$ for the second equality, we have

$$\begin{aligned} \int_{\mathbb{R}\times\mathbb{R}} h(y')\,\eta_x(dy, dy') &= \frac{1}{\mu(\{x\})}\left(\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\int_{\mathbb{R}} h(y')\,\widetilde{m}_u^Q(dy')\,du + \int_{(a_x,b_x)}\int_{\mathbb{R}} h(y')\,\widetilde{m}_u^Q(dy')\,du\right) \\ &= \frac{1}{\mu(\{x\})}\int_{(F_\mu(x-),F_\mu(x)]}\int_{\mathbb{R}} h(y')\,\widetilde{m}_u^Q(dy')\,du = \int_{\mathbb{R}} h(y')\,M_x^Q(dy'). \end{aligned}$$

On the other hand,

$$\int_{\mathbb{R}\times\mathbb{R}} h(y)\,\eta_x(dy, dy') = \frac{1}{\mu(\{x\})}\left(\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\int_{\mathbb{R}} h(y')\,\widetilde{m}_u^Q(dy')\,du + \int_{(a_x,b_x)} h(F_\nu^{-1}(u))\,du\right). \tag{4.2.35}$$

Assume for a moment that

$$\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\int_{\mathbb{R}} h(y')\,\widetilde{m}_u^Q(dy')\,du = \int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-} h(F_\nu^{-1}(u))\,du. \tag{4.2.36}$$

We then deduce with (4.2.35) and (4.2.15) that

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} h(y)\,\eta_x(dy,dy') &= \frac{1}{\mu(\{x\})}\left(\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-} h(F_\nu^{-1}(u))\,du + \int_{(a_x,b_x)} h(F_\nu^{-1}(u))\,du\right)\\ &= \frac{1}{\mu(\{x\})}\int_{(F_\mu(x-),F_\mu(x)]} h(F_\nu^{-1}(u))\,du = \int_{\mathbb{R}} h(y)\,P_x^{HF}(dy).\end{aligned}$$

This proves that we indeed have $\eta_x(dy,dy') \in \Pi(P_x^{HF}, M_x^Q)$ for $\mu(dx)$-almost all $x\in\mathbb{R}$, hence (4.2.19) holds.

Let us then prove (4.2.36). By (4.2.2) there exists a probability kernel $(\pi_u^Q)_{u\in(0,1)}$ such that

$$Q(du,dv) = \frac{1}{\Psi_+(1)}\,d\Psi_+(u)\,\pi_u^Q(dv) = \frac{1}{\Psi_+(1)}\,d\Psi_-(v)\,\pi_v^Q(du).$$

Since the first, resp. second marginals of $Q-Q_x$ and Q_x are singular, i.e. supported by disjoint measurable subsets of $(0,1)$, we have

$$Q_x(du,dv) = \frac{1}{\Psi_+(1)}\mathbb{1}_{\mathcal{V}_x^+}(u)\,d\Psi_+(u)\,\pi_u^Q(dv) = \frac{1}{\Psi_+(1)}\mathbb{1}_{\mathcal{V}_x^-}(v)\,d\Psi_-(v)\,\pi_v^Q(du). \tag{4.2.37}$$

For $u,v\in(0,1)$ such that $F_\nu^{-1}(u)\neq F_\nu^{-1}(v)$, let $\Delta(u,v) = \frac{h(F_\nu^{-1}(v))-h(F_\nu^{-1}(u))}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}$. By Lemma 2.2.5, for du-almost all $u\in(0,1)$ we have

$$\begin{aligned}\int_{\mathbb{R}} h(y)\,\widetilde{m}_u^Q(dy) = h(F_\nu^{-1}(u)) &+ \int_{(0,1)} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^+(u)\,\pi_u^Q(dv)\\ &- \int_{(0,1)} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^-(u)\,\pi_u^Q(dv),\end{aligned} \tag{4.2.38}$$

where the integrals are well defined. Using the facts that $\mathcal{V}_x^+\subset\mathcal{U}^+$, $\mathcal{V}_x^-\subset\mathcal{U}^-$ and the symmetry of the function Δ, we get

$$\begin{aligned}&\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\int_{(0,1)} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^+(u)\,\pi_u^Q(dv)\,du\\ &= \int_{(0,1)^2} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^+(u)\mathbb{1}_{\mathcal{V}_x^+}(u)\,\pi_u^Q(dv)\,du\\ &= \Psi_+(1)\int_{(0,1)^2}\Delta(u,v)\,Q_x(du,dv)\\ &= \int_{(0,1)^2} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^-(v)\mathbb{1}_{\mathcal{V}_x^-}(v)\,\pi_v^Q(du)\,dv\\ &= \int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\int_{(0,1)} \Delta(u,v)(F_\mu^{-1}-F_\nu^{-1})^-(u)\,\pi_u^Q(dv)\,du.\end{aligned}$$

Then (4.2.36) is a direct consequence of (4.2.38) integrated on $\mathcal{V}_x^+ \cup \mathcal{V}_x^-$ with respect to the Lebesgue measure.

Step 4. As mentioned at the beginning of the proof, it remains only to show that for $\mu(dx)$-almost all $x \in \mathbb{R}$, (4.2.20) is satisfied. By (4.2.5) and (4.2.14) we have for $\mu(dx)$-almost all $x \in \mathbb{R}$ and dv-almost $v \in (0,1)$

$$\int_{\mathbb{R}} |y' - F_\nu^{-1}(F_\mu(x-) + v\mu(\{x\}))| \, \widetilde{m}^Q_{F_\mu(x-)+v\mu(\{x\})}(dy') = |x - F_\nu^{-1}(F_\mu(x-) + v\mu(\{x\}))|.$$

The latter equality implies that for $\mu(dx)$-almost all $x \in \mathbb{R}$ such that $\mu(\{x\}) = 0$,

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} |y - y'|\eta_x(dy, dy') &= \int_{\mathbb{R}} |y' - F_\nu^{-1}(F_\mu(x))| \, \widetilde{m}^Q_{F_\mu(x)}(dy') = |x - F_\nu^{-1}(F_\mu(x))| \\ &= \left| x - \int_{\mathbb{R}} y \, P_x^{HF}(dy) \right|.\end{aligned}$$

It remains to show (4.2.20) for $x \in \mathbb{R}$ such that $\mu(\{x\}) > 0$. For such an element x, we have by (4.2.22) either $\mathcal{V}_x^+ = \mathcal{U}_x^+$, which implies $(a_x, b_x) \cap \mathcal{U}_x^+ = \emptyset$, or $\mathcal{V}_x^- = \mathcal{U}_x^-$, which implies $(a_x, b_x) \cap \mathcal{U}_x^- = \emptyset$. In both cases we have that $u \mapsto F_\mu^{-1}(u) - F_\nu^{-1}(u)$ does not change sign on (a_x, b_x). Added to (4.2.5) and (4.2.21), we deduce that

$$\begin{aligned}&\int_{\mathbb{R}\times\mathbb{R}} |y - y'| \, \eta_x(dy, dy') \\ &= \frac{1}{\mu(\{x\})} \int_{(a_x,b_x)} \left(\int_{\mathbb{R}} |y' - F_\nu^{-1}(u)| \, \widetilde{m}^Q_u(dy') \right) du \\ &= \frac{1}{\mu(\{x\})} \int_{(a_x,b_x)} \left| F_\mu^{-1}(u) - F_\nu^{-1}(u) \right| du \\ &= \frac{1}{\mu(\{x\})} \left| \int_{(a_x,b_x)} (F_\mu^{-1}(u) - F_\nu^{-1}(u)) \, du \right| \\ &= \frac{1}{\mu(\{x\})} \left| \int_{\mathcal{V}_x^+} (F_\mu^{-1}(u) - F_\nu^{-1}(u)) \, du + \int_{(a_x,b_x)} (F_\mu^{-1}(u) - F_\nu^{-1}(u)) \, du + \int_{\mathcal{V}_x^-} (F_\mu^{-1}(u) - F_\nu^{-1}(u)) \, du \right| \\ &= \frac{1}{\mu(\{x\})} \left| \int_{(F_\mu(x-), F_\mu(x)]} (x - F_\nu^{-1}(u)) \, du \right| \\ &= \left| x - \frac{1}{\mu(\{x\})} \int_{(F_\mu(x-), F_\mu(x)]} F_\nu^{-1}(u) \, du \right| \\ &= \left| x - \int_{\mathbb{R}} y \, P_x^{HF}(dy) \right|,\end{aligned}$$

which shows (4.2.20) and completes the proof. □

Remark 4.2.3. Despite appearances, the martingale coupling M^Q constructed in the proof of Proposition 4.2.1 does not depend on the choice of the measures Q_x, $x \in A_\mu$, whose marginals are given by (4.2.23). Informally, we see that $(Q_x)_{x\in A_\mu}$ does not affect $(M_x^Q)_{x\in\mathbb{R}}$ outside the jumps of F_μ. Moreover, for all $x \in A_\mu$, Q_x describes the way the elements of $\mathcal{V}_x^+$

are matched with the elements of $\mathcal{V}_x^-$, but this level of detail is not seen by M_x^Q, which only retains the contribution $\frac{1}{\mu(\{x\})}\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\delta_{F_\nu^{-1}(u)}(dy)\,du$, as shown below.

Formally, since for all $x\in A_\mu$, the first, resp. second marginals of $\widetilde{Q}$ and Q_x are singular, i.e. supported by disjoint measurable subsets of $(0,1)$, we have $\pi_u^Q=\delta_{\Gamma(u)}$ for du-almost all $u\in\widetilde{\mathcal{U}}^+\cup\widetilde{\mathcal{U}}^-$. In view of the definition (4.2.3), we deduce that for all du-almost all $u\in\widetilde{\mathcal{U}}^+\cup\widetilde{\mathcal{U}}^-$, $\widetilde{m}_u^Q$ does not depend on $(Q_x)_{x\in A_\mu}$, neither does it for all $u\in(0,1)$ such that $F_\mu^{-1}(u)=F_\nu^{-1}(u)$. Moreover, the image of the continuous part $\mu-\sum_{x\in A_\mu}\mu(\{x\})\delta_x$ of μ by F_μ^{-1} is absolutely continuous with respect to the Lebesgue measure on $(0,1)$. This implies that for $\mu(dx)$-almost all $x\in\mathbb{R}$ such that $\mu(\{x\})=0$, $M_x^Q=\widetilde{m}^Q_{F_\mu(x)}$ does not depend on $(Q_{x'})_{x'\in A_\mu}$.

Let now $x\in A_\mu$. Since $(F_\mu(x-),F_\mu(x)]=\mathcal{V}_x^+\cup\mathcal{V}_x^-\cup(a_x,b_x]$, we have

$$\begin{aligned}M_x^Q(dy)&=\frac{1}{\mu(\{x\})}\int_{F_\mu(x-)}^{F_\mu(x)}\widetilde{m}_u^Q(dy)\,du\\&=\frac{1}{\mu(\{x\})}\left(\int_{\mathcal{V}_x^+}\widetilde{m}_u^Q(dy)\,du+\int_{\mathcal{V}_x^-}\widetilde{m}_u^Q(dy)\,du+\int_{(a_x,b_x]}\widetilde{m}_u^Q(dy)\,du\right).\end{aligned}$$

Using (4.2.3) for the first equality, the fact that $\mathbb{1}_{\mathcal{V}_x^+}(u)(F_\mu^{-1}(u)-F_\nu^{-1}(u))\,du=\mathbb{1}_{\mathcal{V}_x^+}(u)\,d\Psi_+(u)$ for the second equality, (4.2.37) for the next two equalities, the fact that $\mathbb{1}_{\mathcal{V}_x^-}(u)\,d\Psi_-(u)=-\mathbb{1}_{\mathcal{V}_x^-}(u)(F_\mu^{-1}(u)-F_\nu^{-1}(u))\,du$ for the last but one equality and (4.2.3) again for the last one, we have

$$\begin{aligned}&\int_{\mathcal{V}_x^+}\widetilde{m}_u^Q(dy)\,du\\&=\int_{\mathcal{V}_x^+\times(0,1)}\left(\frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\delta_{F_\nu^{-1}(v)}(dy)+\left(1-\frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\right)\delta_{F_\nu^{-1}(u)}(dy)\right)\,du\,\pi_u^Q(dv)\\&=\int_{\mathcal{V}_x^+}\delta_{F_\nu^{-1}(u)}(dy)\,du+\int_{(0,1)^2}\frac{\delta_{F_\nu^{-1}(v)}-\delta_{F_\nu^{-1}(u)}}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}(dy)\mathbb{1}_{\mathcal{V}_x^+}(u)\,d\Psi_+(u)\,\pi_u^Q(dv)\\&=\int_{\mathcal{V}_x^+}\delta_{F_\nu^{-1}(u)}(dy)\,du+\Psi_+(1)\int_{(0,1)^2}\frac{\delta_{F_\nu^{-1}(v)}-\delta_{F_\nu^{-1}(u)}}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}(dy)\,Q_x(du,dv)\\&=\int_{\mathcal{V}_x^+}\delta_{F_\nu^{-1}(u)}(dy)\,du+\int_{(0,1)^2}\frac{\delta_{F_\nu^{-1}(v)}-\delta_{F_\nu^{-1}(u)}}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}(dy)\mathbb{1}_{\mathcal{V}_x^-}(v)\,d\Psi_-(v)\,\pi_v^Q(du)\\&=\int_{\mathcal{V}_x^+}\delta_{F_\nu^{-1}(u)}(dy)\,du+\int_{(0,1)^2}\frac{\delta_{F_\nu^{-1}(v)}-\delta_{F_\nu^{-1}(u)}}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}(dy)\mathbb{1}_{\mathcal{V}_x^-}(u)\,d\Psi_-(u)\,\pi_u^Q(dv)\\&=\int_{\mathcal{V}_x^+}\delta_{F_\nu^{-1}(u)}(dy)\,du+\int_{\mathcal{V}_x^-}\delta_{F_\nu^{-1}(u)}(dy)\,du\\&\quad-\int_{\mathcal{V}_x^+\times(0,1)}\left(\frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\delta_{F_\nu^{-1}(v)}(dy)+\left(1-\frac{F_\mu^{-1}(u)-F_\nu^{-1}(u)}{F_\nu^{-1}(v)-F_\nu^{-1}(u)}\right)\delta_{F_\nu^{-1}(u)}(dy)\right)\,du\,\pi_u^Q(dv)\\&=\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\delta_{F_\nu^{-1}(u)}(dy)\,du-\int_{\mathcal{V}_x^-}\widetilde{m}_u^Q(dy)\,du.\end{aligned}$$

We deduce that $M_x^Q(dy)=\frac{1}{\mu(\{x\})}\left(\int_{\mathcal{V}_x^+\cup\mathcal{V}_x^-}\delta_{F_\nu^{-1}(u)}(dy)\,du+\int_{(a_x,b_x]}\widetilde{m}_u^Q(dy)\,du\right)$. Since $(a_x,b_x]\subset$

$\widetilde{\mathcal{U}}^+ \cup \widetilde{\mathcal{U}}^-$, we have from the foregoing that π_u^Q and therefore $\widetilde{m}_u^Q$ does not depend on $(Q_x)_{x\in A_\mu}$ for du-almost all $u \in (a_x, b_x]$, hence M_x^Q is independent of $(Q_{x'})_{x'\in A_\mu}$.

4.2.4 Stability of the inverse transform martingale coupling

Let $\widetilde{\Pi} = \{(\mu,\nu) \in \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R}) \mid \mu \leq_{cx} \nu\}$. For all $(\mu,\nu) \in \widetilde{\Pi}$, Proposition 4.2.1 and its proof provide a martingale rearrangement coupling $M_{\mu,\nu}$ of the Hoeffding-Fréchet coupling between μ and ν. According to Remark 4.2.2, the martingale coupling $M_{\mu,\nu}$ is the inverse transform martingale coupling between μ and ν as soon as the sign of $F_\mu^{-1} - F_\nu^{-1}$ is constant on each jump of F_μ, which is of course satisfied if F_μ is continuous, or equivalently μ is non-atomic. In the next proposition we prove that under a certain assumption, the inverse transform martingale coupling is stable in $\mathcal{AW}_1$. This directly implies that the map $(\mu,\nu) \mapsto M_{\mu,\nu}$ is continuous on the set $\widehat{\Pi}$ of all pairs $(\mu,\nu) \in \widetilde{\Pi}$ such that μ is non-atomic endowed with the product of the $\mathcal{W}_1$-distance topologies, where the codomain is endowed with the $\mathcal{AW}_1$-distance topology.

Proposition 4.2.4. *Let $(\mu,\nu) \in \widetilde{\Pi}$ and $((\mu_n,\nu_n))_{n\in\mathbb{N}} \in \widetilde{\Pi}^{\mathbb{N}}$ be such that μ_n and ν_n respectively converge to μ and ν in $\mathcal{W}_1$ as $n \to +\infty$. Suppose that asymptotically, any jump of F_μ is included in a jump of F_{μ_n}, that is*

$$\forall x \in \mathbb{R}, \quad \mu(\{x\}) > 0 \implies \exists (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}, \quad F_{\mu_n}(x_n) \wedge F_\mu(x) - F_{\mu_n}(x_n-) \vee F_\mu(x-) \underset{n\to+\infty}{\longrightarrow} \mu(\{x\}), \tag{4.2.39}$$

which is for instance satisfied if μ is non-atomic. Then

$$\mathcal{AW}_1(M_n^{IT}, M^{IT}) \underset{n\to+\infty}{\longrightarrow} 0, \tag{4.2.40}$$

where M_n^{IT}, resp. M^{IT}, denotes the inverse transform martingale coupling between μ_n and ν_n, resp. μ and ν.

Remark 4.2.5. If (4.2.39) is not satisfied, then (4.2.40) may not hold. Indeed, for $n \in \mathbb{N}^*$, let $\mu_n = \mathcal{U}((-1/n, 1/n))$, $\mu = \delta_0$ and $\nu_n = \nu = \mathcal{U}((-1,1))$. We trivially have $M^{IT}(dx,dy) = \mu(dx)\,\nu(dy)$, so $\mathcal{AW}_1(M_n^{IT}, M^{IT}) \geq \int_{x\in\mathbb{R}} \mathcal{W}_1((M_n^{IT})_x, \nu)\,\mu_n(dx)$. However, for $n \in \mathbb{N}^*$, since F_{μ_n} is continuous, we have that for all $x \in \mathbb{R}$, $(M_n^{IT})_x = (\widetilde{m}_n^{IT})_{F_\mu(x)}$, where according to (4.2.9), $((\widetilde{m}_n^{IT})_u(dy))_{u\in(0,1)}$ is a probability kernel such that for all $u \in (0,1)$, there exist $a, b \in [-1,1]$ and $p \in [0,1]$ which satisfy $\widetilde{m}_n^{IT}(u,dy) = p\delta_a + (1-p)\delta_b$. Using the fact that the comonotonic coupling is optimal for the $\mathcal{W}_1$-distance, we get

$$\mathcal{W}_1(p\delta_a + (1-p)\delta_b, \nu) = \int_0^p |a+1-2u|\,du + \int_p^1 |b+1-2u|\,du.$$

It is easy to show that $\int_0^p |a+1-2u|\,du$ is equal to $p(a+1-p) \geq p^2$ if $(a+1)/2 > p$, and equal to $(a+1)^2/2 - p(a+1) + p^2 \leq p^2$ if $(a+1)/2 \leq p$. Therefore, one can readily show that $\int_0^p |a+1-2u|\,du \geq p^2/2$, attained for $a = p-1$. Similarly, we have $\int_p^1 |b+1-2u|\,du \geq (1-p)^2/2$, attained for $b = p$. We deduce that for all $(a,b,p) \in \mathbb{R}^2 \times [0,1]$, $\mathcal{W}_1(p\delta_a + (1-p)\delta_b, \nu) \geq (p^2 + (1-p)^2)/2 \geq 1/4$, attained for $p = 1/2$, hence $\int_{x\in\mathbb{R}} \mathcal{W}_1((M_n^{IT})_x, \nu)\,\mu_n(dx) \geq 1/4$, which proves that (4.2.40) is not satisfied.

Corollary 4.2.6. *There exists a continuous map $\widehat{\Pi} \ni (\mu,\nu) \mapsto M_{\mu,\nu}$ such that for all $(\mu,\nu) \in \widehat{\Pi}$, $M_{\mu,\nu}$ is a martingale rearrangement coupling of the Hoeffding-Fréchet coupling between μ and ν, where $\widehat{\Pi}$ is endowed with the product of the $\mathcal{W}_1$-distance topologies and the codomain with the $\mathcal{AW}_1$-distance topology.*

Proof of Corollary 4.2.6. According to Remark 4.2.2 and Proposition 4.2.4, setting $M_{\mu,\nu}$ as the inverse transform sampling between μ and ν gives a solution. □

Proof of Proposition 4.2.4. Let $\lambda_{(0,1)}$ denote the Lebesgue measure on $(0,1)$. Then

$$\begin{aligned}\mathcal{AW}_1(M_n^{IT}, M^{IT}) &\le \int_0^1 \left(|F_{\mu_n}^{-1}(u) - F_\mu^{-1}(u)| + \mathcal{W}_1\left((M_n^{IT})_{F_{\mu_n}^{-1}(u)}, M^{IT}_{F_\mu^{-1}(u)}\right)\right) du \\ &= \mathcal{W}_1(\mu_n,\mu) + \int_0^1 \mathcal{W}_1\left((M_n^{IT})_{F_{\mu_n}^{-1}(u)}, M^{IT}_{F_\mu^{-1}(u)}\right) du.\end{aligned}$$

Let $(\widetilde{m}_u^{IT})_{u\in(0,1)}$ be the probability kernel defined by (4.2.9), and for all $n \in \mathbb{N}$, let $((\widetilde{m}_n^{IT})_u)_{u\in(0,1)}$ be its equivalent with μ and ν respectively replaced by μ_n and ν_n, so that (4.2.4) is satisfied with $(M^Q, \widetilde{m}^Q)$ replaced by $(M^{IT}, \widetilde{m}^{IT})$ and $(M_n^{IT}, \widetilde{m}_n^{IT})$. For $(x,v) \in \mathbb{R} \times [0,1]$ and $n \in \mathbb{N}$, let $\theta(x,v) = F_\mu(x-) + v\mu(\{x\})$, $\theta_n(x,v) = F_{\mu_n}(x-) + v\mu_n(\{x\})$ and

$$(M_n)_x(dy) = \int_{v=0}^1 \widetilde{m}^{IT}_{\theta_n(x,v)}(dy)\, dv.$$

Then (4.2.14) and the triangle inequality yield

$$\begin{aligned}&\int_{(0,1)} \mathcal{W}_1\left((M_n^{IT})_{F_{\mu_n}^{-1}(u)}, M^{IT}_{F_\mu^{-1}(u)}\right) du \\ &\le \int_{(0,1)} \left(\mathcal{W}_1\left((M_n^{IT})_{F_{\mu_n}^{-1}(u)}, (M_n)_{F_{\mu_n}^{-1}(u)}\right) + \mathcal{W}_1\left((M_n)_{F_{\mu_n}^{-1}(u)}, M^{IT}_{F_\mu^{-1}(u)}\right)\right) du \\ &\le \int_{(0,1)^2} \left(\mathcal{W}_1\left((\widetilde{m}_n^{IT})_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}\right) + \mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right)\right) du\, dv \\ &= \int_{(0,1)} \mathcal{W}_1\left((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}\right) du + \int_{(0,1)^2} \mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right) du\, dv.\end{aligned}$$

In order to show (4.2.40), it is therefore sufficient to prove that the right-hand side vanishes when n goes to $+\infty$. This is achieved in two steps. First, we prove that, on the probability space $(0,1)^2$ endowed with the Lebesgue measure, the random variables

$$\left(\mathcal{W}_1\left((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}\right)\right)_{n\in\mathbb{N}} \quad \text{and} \quad \left(\mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right)\right)_{n\in\mathbb{N}}$$

are uniformly integrable. Second, we show for $du\,dv$-almost every $(u,v) \in (0,1)^2$ that

$$\mathcal{W}_1\left((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}\right) \underset{n\to+\infty}{\longrightarrow} 0 \tag{4.2.41}$$

$$\text{and} \quad \mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right) \underset{n\to+\infty}{\longrightarrow} 0. \tag{4.2.42}$$

Let us begin with the uniform integrability. For $(u,v) \in (0,1)^2$, we can estimate

$$\mathcal{W}_1\left((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}\right) \leq \int_{\mathbb{R}} |y| \left((\widetilde{m}_n^{IT})_u(dy) + \widetilde{m}_u^{IT}(dy)\right),$$
$$\mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right) \leq \int_{\mathbb{R}} |y| \left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}(dy) + \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}(dy)\right).$$

For each nonnegative measurable function $f : \mathbb{R} \to \mathbb{R}$, we have by (4.2.14)

$$\begin{aligned}\int_{(0,1)^2} f\left(\int_{\mathbb{R}} |y| \, \widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}(dy)\right) du\, dv &= \int_{(0,1)^2} f\left(\int_{\mathbb{R}} |y| \, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}(dy)\right) du\, dv \\ &= \int_{(0,1)} f\left(\int_{\mathbb{R}} |y| \, \widetilde{m}_u^{IT}(dy)\right) du.\end{aligned}$$

According to Lemma 2.2.6, M^{IT} is the image of $\mathbb{1}_{(0,1)}(u)\, du\, \widetilde{m}_u^{IT}(dy)$ by $(u,y) \mapsto (F_\mu^{-1}(u), y)$ so that the second marginal of this measure is $\nu(dy)$. Therefore

$$\int_{(0,1)} \int_{\mathbb{R}} |y| \, \widetilde{m}_u^{IT}(dy)\, du = \int_{\mathbb{R}} |y| \, \nu(dy) < +\infty.$$

Hence the random variables $\left(\mathcal{W}_1\left(\widetilde{m}^{IT}_{\theta_n(F_{\mu_n}^{-1}(u),v)}, \widetilde{m}^{IT}_{\theta(F_\mu^{-1}(u),v)}\right)\right)_{n\in\mathbb{N}}$ are uniformly integrable and it is enough to check the uniform integrability of $\left(\int_{\mathbb{R}} |y| \, (\widetilde{m}_n^{IT})_u(dy)\right)_{n\in\mathbb{N}}$ to ensure that of $\left(\mathcal{W}_1\left((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}\right)\right)_{n\in\mathbb{N}}$. Since the second marginal of the measure $\mathbb{1}_{(0,1)}(u)\, du\, (\widetilde{m}_n^{IT})_u(dy)$ is $\nu_n(dy)$, this measure also writes $\nu_n(dy)k_y^n(du)$ for some probability kernel k^n on $\mathbb{R} \times (0,1)$. Let $\varepsilon > 0$ and A be a measurable subset of $(0,1)$ such that $\lambda(A) < \varepsilon$. For all $n \in \mathbb{N}$, we have

$$J_n(A) := \int_A \int_{\mathbb{R}} |y| \, (\widetilde{m}_n^{IT})_u(dy)\, du = \int_{\mathbb{R}} |y| \, \tau_n(dy),$$

where $\tau_n(dy) = \int_{u=0}^1 \mathbb{1}_A(u) k_y^n(du) \nu_n(dy)$ is such that $\tau_n \leq \nu_n$ and $\tau_n(\mathbb{R}) = \lambda(A)$. Hence

$$\sup_{A \in \mathcal{B}((0,1)),\ \lambda(A) \leq \varepsilon} J_n(A) \leq I_\varepsilon^1(\nu_n),$$

where $I_\varepsilon^1(\zeta)$ is defined for all $\zeta \in \mathcal{P}_1(\mathbb{R})$ as the supremum of $\int_{\mathbb{R}} |y| \, \tau(dy)$ over all finite measures τ on $\mathbb{R}$ such that $\tau \leq \zeta$ and $\tau(\mathbb{R}) \leq \varepsilon$. Let $\eta > 0$. By Lemma 5.3.1 (b) below, since $\nu \in \mathcal{P}_1(\mathbb{R})$, there exists $\varepsilon' > 0$ such that $I_{\varepsilon'}^1(\nu) < \eta$. Let then $N \in \mathbb{N}$ be such that for all $n > N$, $\mathcal{W}_1(\nu_n, \nu) < \eta$, so that by Lemma 5.3.1 (c) below, $I_{\varepsilon'}^1(\nu_n) \leq \mathcal{W}_1(\nu_n, \nu) + I_{\varepsilon'}^1(\nu) < 2\eta$. By Lemma 5.3.1 (b) again there exists $\varepsilon'' > 0$ such that for all $n \leq N$, $I_{\varepsilon''}^1(\nu_n) < 2\eta$. We deduce that for all $\varepsilon \in (0, \varepsilon' \wedge \varepsilon'')$,

$$\sup_{n\in\mathbb{N}} \sup_{A \in \mathcal{B}((0,1)),\ \lambda(A) \leq \varepsilon} J_n(A) \leq 2\eta,$$

which yields uniform integrability of $\left(\int_{\mathbb{R}} |y| \, (\widetilde{m}_n^{IT})_u(dy)\right)_{n\in\mathbb{N}}$.

Next, we show the $du\,dv$-almost everywhere pointwise convergence of (4.2.41). Since, by monotonicity, $u \mapsto (F_\mu^{-1}(u), F_\nu^{-1}(u))$ is continuous du-almost everywhere on $(0,1)$ and, then, the weak convergence implies that

$$(F_{\mu_n}^{-1}(u), F_{\nu_n}^{-1}(u)) \underset{n\to+\infty}{\longrightarrow} (F_\mu^{-1}(u), F_\nu^{-1}(u)), \tag{4.2.43}$$

we suppose without loss of generality that this convergence holds. Let $n \in \mathbb{N}$. We have

$$(\widetilde{m}_n^{IT})_u = p_n(u)\delta_{\alpha_n(u)} + (1-p_n(u))\delta_{F_{\nu_n}^{-1}(u)} \quad \text{with} \quad p_n(u) = \frac{F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)}{\alpha_n(u) - F_{\nu_n}^{-1}(u)} \in [0,1]$$

and, by convention, $\alpha_n(u) = F_{\nu_n}^{-1}(u) + 1$ if $F_{\mu_n}^{-1}(u) = F_{\nu_n}^{-1}(u)$.

Suppose first that $u \in \mathcal{U}_0$ i.e. $F_\mu^{-1}(u) = F_\nu^{-1}(u)$, so that $\widetilde{m}_u^{IT} = \delta_{F_\nu^{-1}(u)}$. We have

$$\begin{aligned}
\mathcal{W}_1((\widetilde{m}_n^{IT})_u, \widetilde{m}_u^{IT}) &= \frac{F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)}{\alpha_n(u) - F_{\nu_n}^{-1}(u)}|\alpha_n(u) - F_\nu^{-1}(u)| + \frac{\alpha_n(u) - F_{\mu_n}^{-1}(u)}{\alpha_n(u) - F_{\nu_n}^{-1}(u)}|F_{\nu_n}^{-1}(u) - F_\nu^{-1}(u)| \\
&\le \frac{F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)}{\alpha_n(u) - F_{\nu_n}^{-1}(u)}|\alpha_n(u) - F_{\nu_n}^{-1}(u)| + |F_{\nu_n}^{-1}(u) - F_\nu^{-1}(u)| \\
&\le |F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)| + |F_{\nu_n}^{-1}(u) - F_\nu^{-1}(u)| \\
&\le |F_{\mu_n}^{-1}(u) - F_\mu^{-1}(u)| + 2|F_{\nu_n}^{-1}(u) - F_\nu^{-1}(u)|,
\end{aligned} \tag{4.2.44}$$

where the right-hand side goes to 0 as $n \to \infty$ by (4.2.43).

Suppose next that $u \in \mathcal{U}_+$ i.e. $F_\mu^{-1}(u) > F_\nu^{-1}(u)$, the case $u \in \mathcal{U}_-$ being treated in a similar way. Then without loss of generality

$$\widetilde{m}_u^{IT} = p(u)\delta_{F_\nu^{-1}(\varphi(u))} + (1-p(u))\delta_{F_\nu^{-1}(u)} \text{ with } p(u) = \frac{F_\mu^{-1}(u) - F_\nu^{-1}(u)}{F_\nu^{-1}(\varphi(u)) - F_\nu^{-1}(u)}$$

and $\varphi(u) = \Psi_-^{-1}(\Psi_+(u))$. By (4.2.43), for n large enough, $u \in \mathcal{U}_{n+}$ so that without loss of generality, $\alpha_n(u) = F_{\nu_n}^{-1}(\varphi_n(u))$ with $\varphi_n(u) = \Psi_{n-}^{-1}(\Psi_{n+}(u))$ and checking (4.2.41) amounts to show that

$$F_{\nu_n}^{-1}(\varphi_n(u)) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}(\varphi(u)). \tag{4.2.45}$$

It was shown in the proof of Proposition 2.5.10 that Ψ_{n+} converges uniformly to Ψ_+ on $[0,1]$ and for dv-almost every $v \in (0,1)$,

$$F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n+}(1)v)) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}(\Psi_-^{-1}(\Psi_+(1)v)). \tag{4.2.46}$$

Let $\mathcal{D}$ be the set of discontinuities of $F_\nu^{-1} \circ \Psi_-^{-1}$, which is at most countable by monotonicity. Then [164, Proposition 4.10, Chapter 0] yields

$$0 = \int_{\Psi_+(0)}^{\Psi_+(1)} \mathbb{1}_{\mathcal{D}}(v)\,dv = \int_0^1 \mathbb{1}_{\{\Psi_+(u)\in\mathcal{D}\}}\,d\Psi_+(u).$$

We deduce that for du-almost all $u \in \mathcal{U}_+$, $F_\nu^{-1} \circ \Psi_-^{-1}$ is continuous at $\Psi_+(u)$, which we suppose from now. According to (4.2.46), there exists $\varepsilon > 0$ arbitrarily small such that

$$\begin{aligned}
&F_{\nu_n}^{-1}\left(\Psi_{n-}^{-1}\left(\Psi_{n+}(1)\frac{\Psi_+(u)-\varepsilon}{\Psi_+(1)}\right)\right) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}\left(\Psi_-^{-1}\left(\Psi_+(u)-\varepsilon\right)\right)\\
\text{and}\quad &F_{\nu_n}^{-1}\left(\Psi_{n-}^{-1}\left(\Psi_{n+}(1)\frac{\Psi_+(u)+\varepsilon}{\Psi_+(1)}\right)\right) \underset{n\to+\infty}{\longrightarrow} F_\nu^{-1}\left(\Psi_-^{-1}\left(\Psi_+(u)+\varepsilon\right)\right).
\end{aligned}$$

For n large enough, we have $\Psi_+(u) \in \left[\Psi_{n+}(1)\frac{\Psi_+(u)-\varepsilon}{\Psi_+(1)}, \Psi_{n+}(1)\frac{\Psi_+(u)+\varepsilon}{\Psi_+(1)}\right]$. Therefore, by monotonicity, we have

$$\begin{aligned}
F_\nu^{-1}\left(\Psi_-^{-1}\left(\Psi_+(u)-\varepsilon\right)\right) &= \liminf_{n\to+\infty} F_{\nu_n}^{-1}\left(\Psi_{n-}^{-1}\left(\Psi_{n+}(1)\frac{\Psi_+(u)-\varepsilon}{\Psi_+(1)}\right)\right)\\
&\le \liminf_{n\to+\infty} F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n+}(u)))\\
&\le \limsup_{n\to+\infty} F_{\nu_n}^{-1}(\Psi_{n-}^{-1}(\Psi_{n+}(u)))\\
&\le \limsup_{n\to+\infty} F_{\nu_n}^{-1}\left(\Psi_{n-}^{-1}\left(\Psi_{n+}(1)\frac{\Psi_+(u)+\varepsilon}{\Psi_+(1)}\right)\right)\\
&= F_\nu^{-1}\left(\Psi_-^{-1}\left(\Psi_+(u)+\varepsilon\right)\right).
\end{aligned}$$

Since $F_\nu^{-1} \circ \Psi_-^{-1}$ is continuous at $\Psi_+(u)$, we get when ε vanishes the convergence (4.2.45), which concludes the proof of (4.2.41).

Finally, let us show (4.2.42). Let $w \in (0,1)$ be in the set of continuity points of F_μ^{-1}, F_ν^{-1}, $F_\nu^{-1} \circ \varphi$ and $F_\nu^{-1} \circ \widetilde{\varphi}$. Recall that we have

$$\widetilde{m}_w^{IT} = p(w)\delta_{\alpha(w)} + (1-p(w))\delta_{F_\nu^{-1}(w)} \quad \text{with} \quad p(w) = \frac{F_\mu^{-1}(w) - F_\nu^{-1}(w)}{\alpha(w) - F_\nu^{-1}(w)} \in [0,1]$$

and, by convention, $\alpha(w) = F_\nu^{-1}(w) + 1$ if $F_\mu^{-1}(w) = F_\nu^{-1}(w)$.

Let $(w_n)_{n\in\mathbb{N}}$ be a sequence with values in $(0,1)$ converging to w and let us show that

$$\mathcal{W}_1(\widetilde{m}_{w_n}^{IT}, \widetilde{m}_w^{IT}) \underset{n\to+\infty}{\longrightarrow} 0. \tag{4.2.47}$$

Suppose first that $w \in \mathcal{U}_0$ i.e. $F_\mu^{-1}(w) = F_\nu^{-1}(w)$. Then a computation similar to (4.2.44) yields

$$\mathcal{W}_1(\widetilde{m}_{w_n}^{IT}, \widetilde{m}_w^{IT}) \le |F_\mu^{-1}(w_n) - F_\mu^{-1}(w)| + 2|F_\nu^{-1}(w_n) - F_\nu^{-1}(w)|,$$

where the right-hand side goes to 0 as $n \to \infty$ by continuity of F_μ^{-1} and F_ν^{-1} at w.

Suppose next that $w \in \mathcal{U}_+$ i.e. $F_\mu^{-1}(w) > F_\nu^{-1}(w)$, the case $w \in \mathcal{U}_-$ being treated in a similar way. Then by continuity of F_μ^{-1} and F_ν^{-1} at w, $w_n \in \mathcal{U}_+$ for n large enough so that without loss of generality

$$p(w) = \frac{F_\mu^{-1}(w) - F_\nu^{-1}(w)}{\alpha(w) - F_\nu^{-1}(w)}, \quad p(w_n) = \frac{F_\mu^{-1}(w_n) - F_\nu^{-1}(w_n)}{\alpha(w_n) - F_\nu^{-1}(w_n)},$$

$\alpha(w) = F_\nu^{-1}(\varphi(w))$ and $\alpha(w_n) = F_\nu^{-1}(\varphi(w_n))$, hence (4.2.47) follows from the continuity at w of F_μ^{-1}, F_ν^{-1} and $F_\nu^{-1} \circ \varphi$. Since the set of discontinuity points of the non-decreasing functions F_μ^{-1}, F_ν^{-1}, $F_\nu^{-1} \circ \varphi$ and $F_\nu^{-1} \circ \tilde{\varphi}$ are at most countable, we deduce by (4.2.14) and (4.2.47) that it is sufficient to show for $du\,dv$-almost every $(u, v) \in (0, 1)^2$

$$\theta_n(F_{\mu_n}^{-1}(u), v) \underset{n\to+\infty}{\longrightarrow} \theta(F_\mu^{-1}(u), v),$$

or equivalently

$$(F_{\mu_n}(x_u^n-), F_{\mu_n}(x_u^n)) \underset{n\to+\infty}{\longrightarrow} (F_\mu(x_u-), F_\mu(x_u)) \tag{4.2.48}$$

for du-almost every $u \in (0, 1)$, where $x_u := F_\mu^{-1}(u)$ and $x_u^n := F_{\mu_n}^{-1}(u)$.

Let then $u \in (0, 1)$. We suppose without loss of generality that (4.2.43) still holds. For $n \in \mathbb{N}$, define $l_n = \inf_{k\geq n} x_u^k$ and $r_n = \sup_{k\geq n} x_u^k$. Since (4.2.43) holds we find that $(l_n)_{n\in\mathbb{N}}$, resp. $(r_n)_{n\in\mathbb{N}}$, is a nondecreasing, resp. nonincreasing, sequence converging to x_u. Due to right continuity of F_μ and left continuity of $x \mapsto F_\mu(x-)$ we have

$$F_\mu(x_u-) = \lim_{p\to+\infty} F_\mu(l_p-) \quad \text{and} \quad \lim_{p\to+\infty} F_\mu(r_p) = F_\mu(x_u).$$

By Portmanteau's theorem and monotonicity of cumulative distribution functions we have

$$F_\mu(l_p-) \leq \liminf_{n\to+\infty} F_{\mu_n}(l_p-) \leq \liminf_{n\to+\infty} F_{\mu_n}(x_u^n-) \leq \limsup_{n\to+\infty} F_{\mu_n}(x_u^n) \leq \limsup_{n\to+\infty} F_{\mu_n}(r_p) \leq F_\mu(r_p).$$

By taking the limit $p \to +\infty$, we find

$$F_\mu(x_u-) \leq \liminf_{n\to+\infty} F_{\mu_n}(x_u^n-) \leq \limsup_{n\to+\infty} F_{\mu_n}(x_u^n) \leq F_\mu(x_u).$$

This implies (4.2.48) as soon as F_μ is continuous at x_u. Suppose now that F_μ is discontinuous at x_u. Since μ has countably many atoms, we may suppose without loss of generality that $u \in (F_\mu(x_u-), F_\mu(x_u))$. Let $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^\mathbb{N}$ be the sequence associated to $x = x_u$ by (4.2.39). For n large enough, we have $u \in (F_{\mu_n}(x_n-), F_{\mu_n}(x_n))$, hence $x_n = x_u^n$. Using the assumption made in (4.2.39), we get

$$\begin{aligned}
\liminf_{n\to+\infty} F_{\mu_n}(x_u^n) &= \liminf_{n\to+\infty} (F_{\mu_n}(x_u^n) \wedge F_\mu(x_u)) \\
&= \liminf_{n\to+\infty} (F_{\mu_n}(x_u^n) \wedge F_\mu(x_u) - F_{\mu_n}(x_u^n-) \vee F_\mu(x_u-) + F_{\mu_n}(x_u^n-) \vee F_\mu(x_u-)) \\
&= \mu(\{x_u\}) + \liminf_{n\to+\infty} (F_{\mu_n}(x_u^n-) \vee F_\mu(x_u-)) \geq F_\mu(x_u),
\end{aligned}$$

hence $F_{\mu_n}(x_u^n) \underset{n\to+\infty}{\longrightarrow} F_\mu(x_u)$. Similarly, $F_{\mu_n}(x_u^n-) \underset{n\to+\infty}{\longrightarrow} F_\mu(x_u-)$, which shows (4.2.48) and concludes the proof. $\square$

4.2.5 Non-optimality of the inverse transform martingale coupling

Suppose that F_ν^{-1} is constant on the jumps of F_μ, so that the Hoeffding-Fréchet coupling between μ and ν is concentrated on the graph of the Monge transport map $T = F_\nu^{-1} \circ F_\mu$. According to Proposition 2.2.11, the inverse transform martingale coupling M^{IT} between μ and ν minimises $\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M(dx, dy)$ among all martingale couplings M between μ and ν and satisfies $\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|\, M^{IT}(dx, dy) = \mathcal{W}_1(\mu, \nu)$. The latter fact was used in (4.2.16) to see that the inverse transform sampling constitutes a martingale rearrangement coupling of the Hoeffding-Fréchet coupling. In the attempt to generalise Proposition 4.2.1 to the $\mathcal{AW}_\rho$-distance for $\rho \geq 1$, we naturally raised the question of the optimality of $\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|^\rho\, M(dx, dy)$ among all martingale couplings M between μ and ν. According to Proposition 2.3.5 the inverse transform martingale coupling M^{IT} minimises (resp. maximises) $\int_{\mathbb{R}\times\mathbb{R}} |y - T(x)|^\rho\, M^Q(dx, dy)$ for $\rho > 2$ (resp. $1 < \rho < 2$) among all martingale couplings M^Q parametrised by $Q \in \mathcal{Q}$. However the following example shows that this optimality property holds within the family $(M^Q)_{Q\in\mathcal{Q}}$ but not necessarily when we consider the optimisation over the whole set of martingale couplings between μ and ν.

Example 4.2.7. Let $f : \mathbb{R} \to \mathbb{R}$, $q : \mathbb{R} \to [0, 1]$ and $G : \mathbb{R} \to \mathbb{R}$ be defined for all $y \in \mathbb{R}$ by

$$\begin{aligned}
f(y) &= \frac{1+\mathrm{e}}{6}\left(\mathrm{e}^{-|y|}\mathbb{1}_{\{|y|\geq 1\}} + \frac{\mathrm{e}^{-|y|}+1}{1+\mathrm{e}}\mathbb{1}_{\{|y|<1\}}\right);\\
q(y) &= \frac{\mathrm{e}}{1+\mathrm{e}}\mathbb{1}_{\{y\leq -1\}} + \frac{1}{1+\mathrm{e}^y}\mathbb{1}_{\{-1<y<1\}} + \frac{1}{1+\mathrm{e}}\mathbb{1}_{\{y\geq 1\}};\\
G(y) &= y + 2q(y) - 1.
\end{aligned}$$

The function G is increasing on $\mathbb{R}$. Let then T denote its inverse. Let $\nu(dy) = f(y)\, dy$ and $\mu = G_\sharp\nu$, so that $T_\sharp\mu = \nu$. For all $x \in \mathbb{R}$, let $m_x^0(dy) = q(T(x))\, \delta_{T(x)+1}(dy) + (1 - q(T(x)))\, \delta_{T(x)-1}(dy)$. Let us show that $M^0(dx, dy) = \mu(dx)\, m_x^0(dy)$ is a martingale coupling between μ and ν. Let $h : \mathbb{R} \to \mathbb{R}$ be a measurable and bounded map. Then

$$\begin{aligned}
\int_{\mathbb{R}\times\mathbb{R}} h(y)\, \mu(dx)\, m_x^0(dy) &= \int_{\mathbb{R}\times\mathbb{R}} \left(q(T(x))h(1+T(x)) + (1-q(T(x)))h(T(x)-1)\right)\, \mu(dx)\\
&= \int_{\mathbb{R}} \left(q(y)h(1+y) + (1-q(y))h(y-1)\right)\, \nu(dy)\\
&= \int_{\mathbb{R}} q(y)h(1+y)f(y)\, dy + \int_{\mathbb{R}} (1-q(y))h(y-1)f(y)\, dy\\
&= \int_{\mathbb{R}} \left(q(y-1)f(y-1) + (1-q(y+1))f(y+1)\right) h(y)\, dy.
\end{aligned}$$

By considering the cases $y \leq -2$, $-2 < y \leq -1$, $-1 < y \leq 0$, $0 < y \leq 1$, $1 < y \leq 2$ and $2 \leq y$, it is easy to check that for all $y \in \mathbb{R}$, $q(y-1)f(y-1) + (1-q(y+1))f(y+1) = f(y)$. Therefore,

$$\int_{\mathbb{R}\times\mathbb{R}} h(y)\, M^0(dx, dy) = \int_{\mathbb{R}} h(y)f(y)\, dy = \int_{\mathbb{R}} h(y)\, \nu(dy),$$

and M^0 is a coupling between μ and ν. Moreover, for all $x \in \mathbb{R}$ we have

$$x = G(T(x)) = T(x) + 2q(T(x)) - 1 = q(T(x))(1+T(x)) + (1-q(T(x)))(T(x)-1) = \int_{\mathbb{R}} y\, m_x^0(dy),$$

so M^0 is a martingale coupling between μ and ν.

In order to conclude that

$$\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^0(dx,dy) \; < \; (\text{resp. } >) \; \int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^{IT}(dx,dy) \quad \text{if } \rho > 2 \text{ (resp. } 1 < \rho < 2),$$

we can follow the reasoning made in [3, Section 6.1] which determines the martingale optimal transport between the two uniform distributions $\mathcal{U}([-1,1])$ and $\mathcal{U}([-2,2])$. For $\rho = 2$, $\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^2 \, M(dx,dy)$ does not depend on the choice of the martingale coupling M. Let M be a martingale coupling between μ and ν. Then Jensen's inequality gives

$$\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M(dx,dy) \geq \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^2 \, M(dx,dy) \right)^{\rho/2}$$

for $\rho > 2$ and

$$\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M(dx,dy) \leq \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^2 \, M(dx,dy) \right)^{\rho/2}$$

for $1 < \rho < 2$, with equality iff $|y-T(x)|$ is constant $M(dx,dy)$-almost everywhere. On the one hand, $|y-T(x)| = 1$, $M^0(dx,dy)$-almost everywhere. On the other hand, for all $x \in \mathbb{R}$ we have

$$F_\nu(T(x)) = \nu((-\infty, T(x))) = T_\sharp\mu((-\infty, T(x))) = F_\mu(x).$$

Since ν is clearly non-atomic, F_ν is invertible and the latter equality implies $T = F_\nu^{-1} \circ F_\mu$, hence T is the Monge transport map between μ and ν. By construction, M^{IT} is such that $M^{IT}_x(\{T(x)\}) > 0$, $\mu(dx)$-almost everywhere. Therefore, for $\rho > 2$,

$$\begin{aligned} \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^{IT}(dx,dy) \right)^{1/\rho} &> \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^2 \, M^{IT}(dx,dy) \right)^{1/2} \\ &= \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^2 \, M^0(dx,dy) \right)^{1/2} \\ &= \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^0(dx,dy) \right)^{1/\rho}. \end{aligned}$$

Similarly, for $1 < \rho < 2$,

$$\left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^{IT}(dx,dy) \right)^{1/\rho} < \left(\int_{\mathbb{R}\times\mathbb{R}} |y-T(x)|^\rho \, M^0(dx,dy) \right)^{1/\rho}.$$

4.3 Martingale rearrangement couplings of couplings which satisfy the barycentre dispersion assumption

We recall that a coupling $\pi \in \Pi(\mu,\nu)$ between two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ in the convex order satisfies the barycentre dispersion assumption formulated by Wiesel [194] iff

$$\forall a \in \mathbb{R}, \quad \int_{\mathbb{R}} \mathbb{1}_{[a,+\infty)}(x) \left(x - \int_{\mathbb{R}} y \, \pi_x(dy) \right) \mu(dx) \leq 0. \tag{4.3.1}$$

First we briefly recall Wiesel's construction [194] of a martingale rearrangement coupling of a coupling π which satisfies (4.3.1), which is well perceivable as soon as π has finite support but becomes rather implicit in the general case. Then we design our own construction of such a martingale rearrangement coupling, whose intelligibility does not depend on the finiteness of the support of π. Since the Hoeffding-Fréchet satisfies (4.3.1) [194, Lemma 2.3], this construction extends the study made in Section 4.2.

Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$ and $\mu \neq \nu$ and $\pi \in \Pi(\mu, \nu)$ be a coupling between μ and ν which satisfies the barycentre assumption (4.3.1). Suppose first that π has finite support and is not a martingale coupling between μ and ν. Denoting S the finite support of μ and for all $x \in S$, S_x the finite support of π_x, the latter is equivalent to say that there exists $x \in S$ such that $\int_{\mathbb{R}} y\, \pi_x(dy) \neq x$. As Wiesel [194] points out, the barycentre dispersion assumption (4.3.1) and the convex order between μ and ν imply the existence of $x^-, x^+ \in S$, $y^- \in S_{x^-}$ and $y^+ \in S_{x^+}$ such that

$$\int_{\mathbb{R}} y\, \pi_{x^-}(dy) < x^-, \quad x^+ < \int_{\mathbb{R}} y\, \pi_{x^+}(dy) \quad \text{and} \quad y^- < y^+.$$

He then switches as much as possible the mass at y^- and y^+ of π_{x^-} and π_{x^+} in order to rectify the barycentres. More precisely, he defines for all $x \in S$

$$\pi^{(1)}_x = \mathbb{1}_{\{x \notin \{x^-, x^+\}\}}\, \pi_x + \mathbb{1}_{\{x = x^-\}} \left(\pi_{x^-} + \frac{\lambda}{\mu(\{x^-\})}(\delta_{y^+} - \delta_{y^-}) \right) + \mathbb{1}_{\{x = x^+\}} \left(\pi_{x^+} + \frac{\lambda}{\mu(\{x^+\})}(\delta_{y^-} - \delta_{y^+}) \right),$$

where $\lambda \geq 0$ is taken as large as possible, so that

$$\text{either} \quad \pi^{(1)}_{x^-}(\{y^-\}) = 0, \quad \pi^{(1)}_{x^+}(\{y^+\}) = 0, \quad \int_{\mathbb{R}} y\, \pi^{(1)}_{x^+}(dy) = x^+ \quad \text{or} \quad \int_{\mathbb{R}} y\, \pi_{x^-}(dy) = x^-.$$

Then the measure $\pi^{(1)}(dx, dy) = \mu(dx)\, \pi^{(1)}_x(dy)$ is a coupling between μ and ν which satisfies the barycentre dispersion assumption (4.3.1). Repeating inductively this process yields a sequence $(\pi^{(n)}_x)_{x \in \mathbb{R}}$ of couplings between μ and ν which satisfy (4.3.1). By finiteness of S, the latter sequence is constant for n large enough and the limit is precisely a martingale rearrangement coupling of π.

In the general case, there exists by [194, Lemma 4.1] a sequence $(\pi^n)_{n \in \mathbb{N}^*}$ of finitely supported measures such that $\mathcal{W}^{nd}_1(\pi^n, \pi) \leq 1/n$ for all $n \in \mathbb{N}^*$. The marginals μ_n and ν_n of π^n are not in the convex order, but a mere adaptation of the previous reasoning yields the existence of a coupling π^n_{mr} between μ_n and ν_n which is almost a martingale rearrangement coupling of π^n, in the sense that

$$\int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, (\pi^n_{mr})_x(dy) \right| \mu^n(dx) \leq \frac{1}{n} \quad \text{and} \quad \mathcal{W}^{nd}_1(\pi^n_{mr}, \pi^n) \leq \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, \pi^n_x(dy) \right| \mu^n(dx). \tag{4.3.2}$$

Then Wiesel shows the existence of a coupling π_{mr} between μ and ν such that $\mathcal{W}^{nd}_1\left(\frac{1}{n}\sum_{k=1}^n \pi^k_{mr}, \pi_{mr}\right)$ vanishes as n goes to $+\infty$. By (4.3.2) taken to the limit $n \to +\infty$ and [194, Lemma 2.1] he deduces that π_{mr} is a martingale rearrangement coupling of π.

We propose an alternate construction of a martingale rearrangement coupling of π, regardless of the finiteness of its support. For all $u \in [0,1]$, let

$$G(u) = \int_{\mathbb{R}} y\, \pi_{F_\mu^{-1}(u)}(dy), \quad \Psi_+(u) = \int_0^u (F_\mu^{-1} - G)^+(v)\, dv \quad \text{and} \quad \Psi_-(u) = \int_0^u (F_\mu^{-1} - G)^-(v)\, dv.$$

Let us show that (4.3.1) is equivalent to

$$\forall u \in [0,1], \quad \Psi_+(u) \geq \Psi_-(u). \tag{4.3.3}$$

Using (4.2.10) we see that for all $u \in (0,1)$ and $a \in \mathbb{R}$, $u > F_\mu(a-) \implies F_\mu^{-1}(u) \geq a \implies u \geq F_\mu(a-)$. By the latter implications and the inverse transform sampling we deduce that (4.3.1) is equivalent to

$$\forall a \in \mathbb{R}, \quad \int_{F_\mu(a-)}^1 (F_\mu^{-1}(u) - G(u))\, du \leq 0.$$

Since $\Psi_+(1) = \Psi_-(1)$, consequence of the equality of the respective means of μ and ν, we deduce that it is equivalent to

$$\forall a \in \mathbb{R}, \quad \Psi_+(F_\mu(a-)) \geq \Psi_-(F_\mu(a-)).$$

By right continuity of F_μ, for all $a \in \mathbb{R}$ we have $F_\mu(a) = \lim_{h\to 0, h>0} F_\mu((a+h)-)$, so by continuity of Ψ_+ and Ψ_- we also have $\Psi_+(F_\mu(a)) \geq \Psi_-(F_\mu(a))$ for all $a \in \mathbb{R}$. Moreover, for all $a \in \mathbb{R}$ such that $\mu(\{a\}) > 0$ and $u \in (F_\mu(a-), F_\mu(a)]$, we have by (4.2.11) that $F_\mu^{-1}(u) = a$, so Ψ_+ and Ψ_- are affine on $(F_\mu(a-), F_\mu(a)]$. We deduce that we also have $\Psi_+ \geq \Psi_-$ on $(F_\mu(a-), F_\mu(a)]$, hence the equivalence with (4.3.3).

We define

$$\begin{aligned} &\mathcal{U}^+ = \{u \in (0,1) \mid F_\mu^{-1}(u) > G(u)\}, \quad \mathcal{U}^- = \{u \in (0,1) \mid F_\mu^{-1}(u) < G(u)\}, \\ &\text{and} \quad \mathcal{U}^0 = \{u \in (0,1) \mid F_\mu^{-1}(u) = G(u)\}, \end{aligned}$$

and thanks to the equality $\Psi_+(1) = \Psi_-(1)$ we can set for all $u \in [0,1]$

$$\varphi(u) = \begin{cases} \Psi_-^{-1}(\Psi_+(u)) & \text{if} \quad u \in \mathcal{U}^+; \\ \Psi_+^{-1}(\Psi_-(u)) & \text{if} \quad u \in \mathcal{U}^-; \\ u & \text{if} \quad u \in \mathcal{U}^0. \end{cases}$$

Applying Lemma 2.6.1 again with $f_1 = (F_\mu^{-1} - G)^+$, $f_2 = (F_\mu^{-1} - G)^-$, $u_0 = 1$ and $h : u \mapsto \mathbb{1}_{\{G(\varphi(u)) \leq F_\mu^{-1}(\varphi(u))\}}$ yields

$$\int_0^1 \mathbb{1}_{\{G(\varphi(u)) \leq F_\mu^{-1}(\varphi(u))\}}\, d\Psi_+(u) = \int_0^1 \mathbb{1}_{\{G(v) \leq F_\mu^{-1}(v)\}}\, d\Psi_-(u) = 0.$$

Similarly, we get $\int_0^1 \mathbb{1}_{\{G(\varphi(u)) \geq F_\mu^{-1}(\varphi(u))\}}\, d\Psi_-(u) = 0$. We deduce that

$$\varphi(u) \in \mathcal{U}^-, \quad \text{resp. } \varphi(u) \in \mathcal{U}^+, \quad \text{for } du\text{-almost all } u \in \mathcal{U}^+, \quad \text{resp. } \mathcal{U}^-. \tag{4.3.4}$$

This allows us to define for du-almost all $u \in \mathcal{U}^+ \cup \mathcal{U}^-$

$$p(u) = \frac{G(\varphi(u)) - F_\mu^{-1}(\varphi(u))}{F_\mu^{-1}(u) - G(u) + G(\varphi(u)) - F_\mu^{-1}(\varphi(u))}. \tag{4.3.5}$$

Reasoning like in the derivation of (4.2.32) with $(\varphi, \Psi_+, \Psi_-, \mathcal{U}^+, \mathcal{U}^-)$ replacing $(\Gamma, \chi_+, \chi_-, \widetilde{\mathcal{U}}^+, \widetilde{\mathcal{U}}^-)$, we get that $\varphi(\varphi(u)) = u$ for du-almost all $u \in \mathcal{U}^+$. Similarly, $\varphi(\varphi(u)) = u$ for all du-almost all $u \in \mathcal{U}^-$. We deduce that

$$\text{for } du\text{-almost all } u \in \mathcal{U}^+ \cup \mathcal{U}^-, \quad \varphi(\varphi(u)) = u, \tag{4.3.6}$$

and

$$\text{for } du\text{-almost all } u \in \mathcal{U}^+ \cup \mathcal{U}^-, \quad p(\varphi(u)) = \frac{G(u) - F_\mu^{-1}(u)}{F_\mu^{-1}(\varphi(u)) - G(\varphi(u)) + G(u) - F_\mu^{-1}(u)} = 1 - p(u). \tag{4.3.7}$$

In order to define the appropriate martingale kernel, we rely on the following Lemma which allows us to inject some stochastic order in the construction, a convenient tool for the computation of Wasserstein distances. Its proof is moved to the end of the present section. We recall that two probability measures μ and ν on the real line are said to be in the stochastic order, denoted $\mu \leq_{st} \nu$, iff $F_\mu^{-1}(u) \leq F_\nu^{-1}(u)$ for all $u \in [0,1]$. Since the Hoeffding-Fréchet coupling between μ and ν is optimal for $\mathcal{W}_1(\mu, \nu)$, this implies by the inverse transform sampling that $\mathcal{W}_1(\mu, \nu) = \int_\mathbb{R} y\, \nu(dy) - \int_\mathbb{R} x\, \mu(dx)$.

Lemma 4.3.1. *Let $\mathfrak{B}$ be the set of all quadruples $(y, \widetilde{y}, \mu, \widetilde{\mu}) \in \mathbb{R} \times \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R})$ such that μ and $\widetilde{\mu}$ have respective means x and $\widetilde{x}$ and $x < y \leq \widetilde{y} < \widetilde{x}$. Endow $\mathcal{P}_1(\mathbb{R})$ with the Borel σ-algebra of the weak convergence topology and $\mathfrak{B}$ with the trace of the product σ-algebra on $\mathbb{R} \times \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R})$.*

Then there exist two measurable maps $\beta, \widetilde{\beta} : \mathfrak{B} \to \mathcal{P}_1(\mathbb{R})$ such that for all $(y, \widetilde{y}, \mu, \widetilde{\mu})$, denoting $\nu = \beta(y, \widetilde{y}, \mu, \widetilde{\mu})$, $\widetilde{\nu} = \widetilde{\beta}(y, \widetilde{y}, \mu, \widetilde{\mu})$ and $p = \frac{\widetilde{x} - \widetilde{y}}{y - x + \widetilde{x} - \widetilde{y}}$ where x and $\widetilde{x}$ are the respective means of μ and $\widetilde{\mu}$, we have

$$\int_\mathbb{R} w\, \nu(dw) = y, \quad \int_\mathbb{R} w\, \widetilde{\nu}(dw) = \widetilde{y}, \quad \mu \leq_{st} \nu, \quad \widetilde{\nu} \leq_{st} \widetilde{\mu} \quad \textit{and} \quad p\nu + (1-p)\widetilde{\nu} = p\mu + (1-p)\widetilde{\mu}. \tag{4.3.8}$$

In particular, $p\, \delta_y(dz)\, \nu(dw) + (1-p)\, \delta_{\widetilde{y}}(dz)\, \widetilde{\nu}(dw)$ is a martingale coupling between $p\delta_y(dz) + (1-p)\delta_{\widetilde{y}}(dz)$ and $p\mu(dw) + (1-p)\widetilde{\mu}(dw)$ and $\mathcal{W}_1(\mu, \nu) = y - x$, $\mathcal{W}_1(\widetilde{\mu}, \widetilde{\nu}) = \widetilde{x} - \widetilde{y}$.

In order to use Lemma 4.3.1 we need to compare φ to the identity function. The inequality (4.3.3) is equivalent by appropriate normalisation of (4.2.10) to $u \geq \Psi_+^{-1}(\Psi_-(u))$ for all $u \in [0,1]$, hence

$$\forall u \in \mathcal{U}^-, \quad \varphi(u) \leq u. \tag{4.3.9}$$

Moreover, by (4.3.6), Lemma 2.6.1 applied with $f_1 = (F_\mu^{-1} - G)^+$, $f_2 = (F_\mu^{-1} - G)^-$, $u_0 = 1$ and $h : u \mapsto \mathbb{1}_{\{u < \varphi(u)\}}$ we have

$$\int_0^1 \mathbb{1}_{\{\varphi(u) < u\}}\, d\Psi_+(u) = \int_0^1 \mathbb{1}_{\{\varphi(u) < \varphi(\varphi(u))\}}\, d\Psi_+(u) = \int_0^1 \mathbb{1}_{\{u < \varphi(u)\}}\, d\Psi_-(u).$$

By (4.3.9) the right-hand side is 0, hence

$$\text{for } du\text{-almost all } u \in \mathcal{U}^+, \quad \varphi(u) \geq u. \tag{4.3.10}$$

Let

$$A^+ = \{u \in \mathcal{U}^+ \mid F_\mu^{-1}(\varphi(u)) < G(\varphi(u)),\ \varphi(\varphi(u)) = u \text{ and } p(\varphi(u)) = 1 - p(u)\}$$
$$\text{and} \quad A^- = \{u \in \mathcal{U}^- \mid F_\mu^{-1}(\varphi(u)) > G(\varphi(u)),\ \varphi(\varphi(u)) = u \text{ and } p(\varphi(u)) = 1 - p(u)\}.$$

For all $u \in A^+$, we have by definition

$$\varphi(u) \in \mathcal{U}^-, \quad F_\mu^{-1}(\varphi(\varphi(u))) = F_\mu^{-1}(u) > G(u) = G(\varphi(\varphi(u))),$$
$$\varphi(\varphi(\varphi(u))) = \varphi(u) \quad \text{and} \quad p(\varphi(\varphi(u))) = p(u) = 1 - p(\varphi(u)),$$

hence $\varphi(u) \in A^-$. Similarly, for all $u \in A^-$, $\varphi(u) \in A^+$. By (4.3.4), (4.3.6), (4.3.7), (4.3.10) and the monotonicity of F_μ^{-1}, we deduce that A^+ and A^- are two disjoint Borel sets such that the Lebesgue measure of $(\mathcal{U}^+ \backslash A^+) \cup (\mathcal{U}^- \backslash A^-)$ is 0 and

$$\forall u \in A^+, \quad G(u) < F_\mu^{-1}(u) \leq F_\mu^{-1}(\varphi(u)) < G(\varphi(u)). \tag{4.3.11}$$

For all $u \in A^+$, $\pi_{F_\mu^{-1}(u)}$ and $\pi_{F_\mu^{-1}(\varphi(u))}$ have by definition respective means $G(u)$ and $G(\varphi(u))$, so by (4.3.11) we can apply Lemma 4.3.1 with

$$(y, \widetilde{y}, \mu, \widetilde{\mu}) = (F_\mu^{-1}(u), F_\mu^{-1}(\varphi(u)), \pi_{F_\mu^{-1}(u)}, \pi_{F_\mu^{-1}(\varphi(u))}).$$

Hence there exist two probability measures $m_u, \widetilde{m}_u \in \mathcal{P}_1(\mathbb{R})$ with respective means $F_\mu^{-1}(u)$, $F_\mu^{-1}(\varphi(u))$ and such that

$$\begin{aligned} &\pi_{F_\mu^{-1}(u)} \leq_{st} m_u, \quad \widetilde{m}_u \leq_{st} \pi_{F_\mu^{-1}(\varphi(u))}, \\ \text{and} \quad &p(u)m_u + (1-p(u))\widetilde{m}_u = p(u)\pi_{F_\mu^{-1}(u)} + (1-p(u))\pi_{F_\mu^{-1}(\varphi(u))}. \end{aligned} \tag{4.3.12}$$

Since $A^+ = \varphi(A^-)$ and $A^- = \varphi(A^+)$, for all $u \in A^-$ we can set $m_u = \widetilde{m}_{\varphi(u)}$, so that

$$\forall u \in A^+, \quad \pi_{F_\mu^{-1}(u)} \leq_{st} m_u \quad \text{and} \quad \forall u \in A^-, \quad m_u \leq_{st} \pi_{F_\mu^{-1}(u)}, \tag{4.3.13}$$

and for all $u \in A^+ \cup A^-$,

$$p(u)m_u + p(\varphi(u))m_{\varphi(u)} = p(u)\pi_{F_\mu^{-1}(u)} + p(\varphi(u))\pi_{F_\mu^{-1}(\varphi(u))}. \tag{4.3.14}$$

Finally, for all $u \in \mathcal{U}^0 \cup (\mathcal{U}^+ \backslash A^+) \cup (\mathcal{U}^- \backslash A^-)$ set $m_u = \pi_{F_\mu^{-1}(u)}$. By composition of the measurable map $u \mapsto (F_\mu^{-1}(u), F_\mu^{-1}(\varphi(u)), \pi_{F_\mu^{-1}(u)}, \pi_{F_\mu^{-1}(\varphi(u))})$ and the measurable map β defined in Lemma 4.3.1, the map $u \mapsto m_u$ is measurable. By [6, Theorem 19.12] it is equivalent to say that $(m_u)_{u \in (0,1)}$ is a probability kernel, hence we can define

$$M(dx, dy) = \int_0^1 \delta_{F_\mu^{-1}(u)}(dx)\, m_u(dy)\, du. \tag{4.3.15}$$

We now prove that M is indeed a martingale coupling between μ and ν, then that it is a martingale rearrangement coupling of π.

Proposition 4.3.2. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$ and $\mu \neq \nu$ and $\pi \in \Pi(\mu, \nu)$ be a coupling between μ and ν which satisfies the barycentre assumption* (4.3.1)*. Then the measure M defined by* (4.3.15) *is a martingale coupling between μ and ν. Moreover, M is a martingale rearrangement coupling of π:*

$$\mathcal{AW}_1(\pi, M) = \inf_{M' \in \Pi^{\mathrm{M}}(\mu,\nu)} \mathcal{AW}_1(\pi, M').$$

Proof. First we show that M is a martingale coupling between μ and ν. Let $f : \mathbb{R} \to \mathbb{R}$ be measurable and bounded. By the inverse transform sampling for the second equality, we have

$$\int_{\mathbb{R}\times\mathbb{R}} f(x)\, M(dx, dy) = \int_0^1 f(F_\mu^{-1}(u))\, du = \int_{\mathbb{R}} f(x)\, \mu(dx),$$

hence the first marginal of M is μ. On the other hand, let $H : [0, 1] \to \mathbb{R}$ be a measurable and bounded map. Using (4.3.5), (4.3.6) and Lemma 2.6.1 applied with $f_1 = (F_\mu^{-1} - G)^+$, $f_2 = (F_\mu^{-1} - G)^-$, $u_0 = 1$ and $h : u \mapsto \frac{H(\varphi(u))}{F_\mu^{-1}(\varphi(u)) - G(\varphi(u)) + G(u) - F_\mu^{-1}(u)}$ for the third equality, we get

$$\begin{aligned}
\int_{\mathcal{U}^+} (1 - p(u))H(u)\, du &= \int_0^1 \frac{(F_\mu^{-1} - G)^+(u)}{F_\mu^{-1}(u) - G(u) + G(\varphi(u)) - F_\mu^{-1}(\varphi(u))} H(u)\, du \\
&= \int_0^1 h(\varphi(u))\, d\Psi_+(u) \\
&= \int_0^1 h(v)\, d\Psi_-(v) \\
&= \int_0^1 \frac{(F_\mu^{-1} - G)^-(v)}{F_\mu^{-1}(\varphi(v)) - G(\varphi(v)) + G(v) - F_\mu^{-1}(v)} H(\varphi(v))\, dv \\
&= \int_{\mathcal{U}^-} p(\varphi(v))H(\varphi(v))\, dv.
\end{aligned}$$

Similarly, we have $\int_{\mathcal{U}^-} (1 - p(u))H(u)\, du = \int_{\mathcal{U}^+} p(\varphi(u))H(\varphi(u))\, du$. We deduce that

$$\begin{aligned}
\int_0^1 H(u)\, du &= \int_{\mathcal{U}^0} H(u)\, du + \int_{\mathcal{U}^+} p(u)H(u)\, du + \int_{\mathcal{U}^+} (1 - p(u))H(u)\, du \\
&\quad + \int_{\mathcal{U}^-} p(u)H(u)\, du + \int_{\mathcal{U}^-} (1 - p(u))H(u)\, du \\
&= \int_{\mathcal{U}^0} H(u)\, du + \int_{\mathcal{U}^+\cup\mathcal{U}^-} (p(u)H(u) + p(\varphi(u))H(\varphi(u)))\, du.
\end{aligned} \tag{4.3.16}$$

Using (4.3.16) applied with $H : u \mapsto \int_{\mathbb{R}} f(y)\, m_u(dy)$ for the second equality, the fact that $m_u = \pi_{F_\mu^{-1}(u)}$ for all $u \in \mathcal{U}^0$ and (4.3.14) for the third equality, (4.3.16) again applied with $H : u \mapsto \int_{\mathbb{R}} f(y)\, \pi_{F_\mu^{-1}(u)}(dy)$ for the fourth equality and the inverse transform sampling for the last equality, we get

$$\int_{\mathbb{R}\times\mathbb{R}} f(y)\, M(dx, dy)$$

$$
\begin{aligned}
&= \int_0^1 \int_{\mathbb{R}} f(y)\, m_u(dy)\, du \\
&= \int_{\mathcal{U}^0} \int_{\mathbb{R}} f(y)\, m_u(dy)\, du + \int_{\mathcal{U}^+\cup\mathcal{U}^-} \int_{\mathbb{R}} f(y)\, (p(u)\, m_u(dy) + p(\varphi(u))\, m_{\varphi(u)}(dy))\, du \\
&= \int_{\mathcal{U}^0} \int_{\mathbb{R}} f(y)\, \pi_{F_\mu^{-1}(u)}(dy)\, du \\
&\quad + \int_{\mathcal{U}^+\cup\mathcal{U}-} \int_{\mathbb{R}} f(y)\, (p(u)\, \pi_{F_\mu^{-1}(u)}(dy) + p(\varphi(u))\, \pi_{F_\mu^{-1}(\varphi(u))}(dy))\, du \\
&= \int_0^1 \int_{\mathbb{R}} f(y)\, \pi_{F_\mu^{-1}(u)}(dy)\, du \\
&= \int_{\mathbb{R}} f(y)\, \nu(dy).
\end{aligned}
$$

We deduce that ν is the second marginal of M, hence $M \in \Pi(\mu, \nu)$. Let us now show that M is a martingale coupling. By construction, for du-almost all $u \in (0, 1)$,

$$\int_{\mathbb{R}} y\, m_u(dy) = F_\mu^{-1}(u).$$

Therefore, for any measurable and bounded map $f : \mathbb{R} \to \mathbb{R}$,

$$\int_{\mathbb{R}\times\mathbb{R}} f(x)(y - x)\, M(dx, dy) = \int_0^1 f(F_\mu^{-1}(u)) \int_{\mathbb{R}} (y - F_\mu^{-1}(u))\, m_u(dy)\, du = 0,$$

hence $M \in \Pi^{\mathrm{M}}(\mu, \nu)$.

Let us now show that M is a martingale rearrangement coupling of π. By (4.1.6) and [194, Lemma 2.1] we have

$$\inf_{M' \in \Pi^{\mathrm{M}}(\mu,\nu)} \mathcal{AW}_1(\pi, M') \geq \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, \pi_x(dy) \right| \mu(dx),$$

hence it is sufficient to show that

$$\mathcal{AW}_1(\pi, M) \leq \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, \pi_x(dy) \right| \mu(dx). \tag{4.3.17}$$

For $(x, v) \in \mathbb{R} \times (0, 1)$, let $\theta(x, v) = F_\mu(x-) + v\mu(\{x\})$, so that for $\mu(dx)$-almost all $x \in \mathbb{R}$ we have

$$M_x(dy) = \int_0^1 m_{\theta(x,v)}(dy)\, dv.$$

Then

$$\mathcal{AW}_1(\pi, M) \leq \int_{\mathbb{R}} \mathcal{W}_1(\pi_x, M_x)\, \mu(dx) = \int_{\mathbb{R}} \mathcal{W}_1 \left(\pi_x, \int_0^1 m_{\theta(x,v)}\, dv \right) \mu(dx).$$

By (4.2.11) we have $x = F_\mu^{-1}(\theta(x, v))$ for all $(x, v) \in \mathbb{R} \times (0, 1)$, hence

$$\pi_x(dy) = \int_0^1 \pi_{F_\mu^{-1}(\theta(x,v))}(dy)\, dv.$$

Using (4.2.14) for the second equality, we get

$$\begin{aligned}\int_{\mathbb{R}} \mathcal{W}_1\left(\pi_x, \int_0^1 m_{\theta(x,v)}\,dv\right)\mu(dx) &= \int_{\mathbb{R}} \mathcal{W}_1\left(\int_0^1 \pi_{F_\mu^{-1}(\theta(x,v))}\,dv, \int_0^1 m_{\theta(x,v)}\,dv\right)\mu(dx)\\ &\le \int_{(0,1)\times\mathbb{R}} \mathcal{W}_1(\pi_{F_\mu^{-1}(\theta(x,v))}, m_{\theta(x,v)})\,dv\,\mu(dx)\\ &= \int_0^1 \mathcal{W}_1(\pi_{F_\mu^{-1}(u)}, m_u)\,du.\end{aligned}$$

According to (4.3.13), for du-almost all $u \in (0,1)$, either $\pi_{F_\mu^{-1}(u)} \le_{st} m_u$ or $m_u \le_{st} \pi_{F_\mu^{-1}(u)}$, hence

$$\mathcal{W}_1(\pi_{F_\mu^{-1}(u)}, m_u) = \left|\int_{\mathbb{R}} y\,m_u(dy) - \int_{\mathbb{R}} y\,\pi_{F_\mu^{-1}(u)}(dy)\right| = \left|F_\mu^{-1}(u) - \int_{\mathbb{R}} y\,\pi_{F_\mu^{-1}(u)}(dy)\right|.$$

By the inverse transform sampling, we deduce that

$$\mathcal{AW}_1(\pi, M) \le \int_0^1 \left|F_\mu^{-1}(u) - \int_{\mathbb{R}} y\,\pi_{F_\mu^{-1}(u)}(dy)\right|du = \int_{\mathbb{R}} \left|x - \int_{\mathbb{R}} y\,\pi_x(dy)\right|\mu(dx),$$

which implies (4.3.17) and ends the proof. □

Proof of Lemma 4.3.1. Let $(y, \widetilde{y}, \mu, \widetilde{\mu}) \in \mathfrak{B}$, x and $\widetilde{x}$ be the respective means of μ and $\widetilde{\mu}$ and $p = \frac{\widetilde{x}-\widetilde{y}}{y-x+\widetilde{x}-\widetilde{y}}$. First we construct two measures $\nu, \widetilde{\nu} \in \mathcal{P}_1(\mathbb{R})$ which satisfy (4.3.8). Then we show that ν and $\widetilde{\nu}$ are measurable in $(y, \widetilde{y}, \mu, \widetilde{\mu})$.

We set for $q \in [0, p \wedge (1-p)]$

$$\begin{aligned} J(q) &:= \int_0^{\frac{q}{p}} F_{\widetilde{\mu}}^{-1}\left(\frac{1-pu-p}{1-p}\right)du + \int_{\frac{q}{p}}^1 F_\mu^{-1}(u)\,du,\\ \text{and}\quad \widetilde{J}(q) &:= \int_0^{\frac{q}{1-p}} F_\mu^{-1}\left(\frac{(1-p)u}{p}\right)du + \int_0^{\frac{1-q-p}{1-p}} F_{\widetilde{\mu}}^{-1}(u)\,du.\end{aligned} \tag{4.3.18}$$

Since the quantile functions are non-decreasing, the function J is continuous and concave as the sum of two concave functions and the function $\widetilde{J}$ is continuous and convex as the sum of two convex functions. We have $J(0) = \int_0^1 F_\mu^{-1}(u)\,du = x$ and $\widetilde{J}(0) = \int_0^1 F_{\widetilde{\mu}}^{-1}(u)\,du = \widetilde{x}$ by the inverse transform sampling. If $p \le \frac{1}{2}$, then by the change of variables $v = \frac{1-pu-p}{1-p}$ and since $F_{\widetilde{\mu}}^{-1}$ is non-decreasing, we have

$$\begin{aligned} J(p) &= \frac{1-p}{p}\int_{\frac{1-2p}{1-p}}^1 F_{\widetilde{\mu}}^{-1}(v)\,dv\\ &= \int_0^1 F_{\widetilde{\mu}}^{-1}(v)\,dv + \frac{1-2p}{1-p}\left(\frac{1}{1-\frac{1-2p}{1-p}}\int_{\frac{1-2p}{1-p}}^1 F_{\widetilde{\mu}}^{-1}(v)\,dv - \frac{1-p}{1-2p}\int_0^{\frac{1-2p}{1-p}} F_{\widetilde{\mu}}^{-1}(v)\,dv\right)\\ &\ge \int_0^1 F_{\widetilde{\mu}}^{-1}(v)\,dv = \widetilde{x}.\end{aligned}$$

Since $x < y < \tilde{x}$ and the function J is concave, there is a unique $q_\star \in (0, p)$ such that $J(q_\star) = y$. Moreover, the left-hand derivative of J at $q_\star$ is positive which writes

$$F_{\tilde{\mu}}^{-1}\left(\frac{1 - q_\star - p}{1 - p}+\right) > F_\mu^{-1}\left(\frac{q_\star}{p}\right). \tag{4.3.19}$$

If $p > \frac{1}{2}$, we have $\tilde{J}(1 - p) = \int_0^1 F_\mu^{-1}\left(\frac{(1-p)u}{p}\right) du \leq \int_0^1 F_\mu^{-1}(v)\, dv = x$ and by convexity of $\tilde{J}$ there is a unique $q_\star \in (0, 1 - p)$ such that $\tilde{J}(q_\star) = \tilde{y}$. Moreover, the left-hand derivative of $\tilde{J}$ at $q_\star$ is negative so that (4.3.19) still holds.

Let ν and $\tilde{\nu}$ be the respective images of the Lebesgue measure on $[0, 1]$ by

$$\begin{aligned} &u \mapsto \mathbb{1}_{\{u < \frac{q_\star}{p}\}} F_{\tilde{\mu}}^{-1}\left(\frac{1 - pu - p}{1 - p}\right) + \mathbb{1}_{\{u \geq \frac{q_\star}{p}\}} F_\mu^{-1}(u) \\ \text{and} \quad &u \mapsto \mathbb{1}_{\{u < \frac{q_\star}{1-p}\}} F_\mu^{-1}\left(\frac{(1 - p)u}{p}\right) + \mathbb{1}_{\{u \geq \frac{q_\star}{1-p}\}} F_{\tilde{\mu}}^{-1}\left(u - \frac{q_\star}{1 - p}\right). \end{aligned} \tag{4.3.20}$$

With the definition of J and $\tilde{J}$, we easily check that $\int_{\mathbb{R}} z\, \nu(dz) = J(q_\star)$ and $\int_{\mathbb{R}} z\, \tilde{\nu}(dz) = \tilde{J}(q_\star)$. Moreover, the inequality (4.3.19) implies that for each $u \in (0, \frac{q_\star}{p})$,

$$F_{\tilde{\mu}}^{-1}\left(\frac{1 - pu - p}{1 - p}\right) \geq F_{\tilde{\mu}}^{-1}\left(\frac{1 - q_\star - p}{1 - p}+\right) > F_\mu^{-1}\left(\frac{q_\star}{p}\right) \geq F_\mu^{-1}(u),$$

so that ν dominates for the stochastic order the image μ of the Lebesgue measure on $(0, 1)$ by F_μ^{-1}. In a symmetric way, $\tilde{\nu} \leq_{st} \tilde{\mu}$.

For $h : \mathbb{R} \to \mathbb{R}$ measurable and bounded, we have using the changes of variables $v = \frac{1-pu-p}{1-p}$, $v = \frac{(1-p)u}{p}$ and $v = u - \frac{q_\star}{1-p}$ then the inverse transform sampling for the last equality

$$\begin{aligned} &p \int_{\mathbb{R}} h(z)\, \nu(dz) + (1 - p) \int_{\mathbb{R}} h(z)\, \tilde{\nu}(dz) \\ &= p\left(\int_0^{\frac{q_\star}{p}} h\left(F_{\tilde{\mu}}^{-1}\left(\frac{1 - pu - p}{1 - p}\right)\right) du + \int_{\frac{q_\star}{p}}^1 h(F_\mu^{-1}(u))\, du\right) \\ &\quad + (1 - p)\left(\int_0^{\frac{q_\star}{1-p}} h\left(F_\mu^{-1}\left(\frac{(1 - p)u}{p}\right)\right) du + \int_{\frac{q_\star}{1-p}}^1 h\left(F_{\tilde{\mu}}^{-1}\left(u - \frac{q_\star}{1 - p}\right)\right) du\right) \\ &= (1 - p) \int_{\frac{1-q_\star-p}{1-p}}^1 h\left(F_{\tilde{\mu}}^{-1}(v)\right) dv + p \int_{\frac{q_\star}{p}}^1 h(F_\mu^{-1}(u))\, du + p \int_0^{\frac{q_\star}{p}} h\left(F_\mu^{-1}(v)\right) dv \\ &\quad + (1 - p) \int_0^{\frac{1-q_\star-p}{1-p}} h\left(F_{\tilde{\mu}}^{-1}(v)\right) dv \\ &= p \int_0^1 h(F_\mu^{-1}(u))\, du + (1 - p) \int_0^1 h\left(F_{\tilde{\mu}}^{-1}(v)\right) dv \\ &= p \int_{\mathbb{R}} h(z)\, \mu(dz) + (1 - p) \int_{\mathbb{R}} h(z)\, \tilde{\mu}(dz). \end{aligned}$$

Hence $p\nu + (1-p)\widetilde{\nu} = p\mu + (1-p)\widetilde{\mu}$. Taking expectations then using the definition of p, we get

$$pJ(q_\star) + (1-p)\widetilde{J}(q_\star) = px + (1-p)\widetilde{x} = py + (1-p)\widetilde{y}.$$

When $p \le \frac{1}{2}$ (resp. $p > \frac{1}{2}$), $J(q_\star) = y$ (resp. $\widetilde{J}(q_\star) = \widetilde{y}$) and we deduce that $\widetilde{J}(q_\star) = \widetilde{y}$ (resp. $J(q_\star) = y$).

It remains to show that ν and $\widetilde{\nu}$ are measurable in $(y, \widetilde{y}, \mu, \widetilde{\mu})$. It is clear that p is a measurable function of $(y, \widetilde{y}, \mu, \widetilde{\mu})$. Moreover we always have by definition $p \in (0,1)$, so the relation $p\nu + (1-p)\widetilde{\nu} = p\mu + (1-p)\widetilde{\mu}$ implies that it suffices to show that ν is measurable in $(y, \widetilde{y}, \mu, \widetilde{\mu})$. Any quantile function is an element of the set $\mathcal{D}$ of the real-valued càglàd functions on $(0,1)$. Analogously to the Skorokhod space of the real-valued càdlàg functions on $(0,1)$, we endow $\mathcal{D}$ with the σ-field generated by the projection maps $\alpha_u : \mathcal{D} \ni f \mapsto f(u)$, $u \in (0,1)$, which coincides with the σ-field $\sigma(\alpha_u, u \in T)$ for any dense subset $T \subset (0,1)$. Let $(\mu_n)_{n\in\mathbb{N}} \in \mathcal{P}_1(\mathbb{R})^{\mathbb{N}}$ converge to μ for the weak convergence topology and T be the complement of the at most countable set of discontinuities of F_μ^{-1}. Then for all $u \in T$, $F_{\mu_n}^{-1}(u) = \alpha_u(F_{\mu_n}^{-1})$ converges to $F_\mu^{-1}(u) = \alpha_u(F_\mu^{-1})$. We deduce that F_μ^{-1} and $F_{\widetilde{\mu}}^{-1}$ are respectively measurable in μ and $\widetilde{\mu}$. By (4.3.20), ν is the image of the Lebesgue measure on $[0,1]$ by a measurable function of $p, F_\mu^{-1}, F_{\widetilde{\mu}}^{-1}$ and $q_\star$, hence it remains to prove the measurability of $q_\star$ in $(y, \widetilde{y}, \mu, \widetilde{\mu})$.

If $p \le \frac{1}{2}$ we saw that $J(0) = x < y < \widetilde{x} \le J(p)$ and $q_\star$ is the only real number in $[0,p]$ such that $J(q_\star) = y$. By concavity of J we necessarily have for all $q \in [0,p]$ that $q_\star \le q$ iff $J(q) \ge y$. Similarly, if $p > \frac{1}{2}$ then for all $q \in [0, 1-p]$, $q_\star \le q$ iff $\widetilde{J}(q) \le \widetilde{y}$. Fix then $q \in [0, p \wedge (1-p)]$. To conclude, it suffices to show that $J(q)$ and $\widetilde{J}(q)$ are measurable in $(y, \widetilde{y}, \mu, \widetilde{\mu})$. By Lemma 3.4.5, the Borel σ-algebras of the weak convergence topology and the $\mathcal{W}_1$-distance topology coincide on $\mathcal{P}_1(\mathbb{R})$. Moreover, since $(\mathcal{P}_1(\mathbb{R}), \mathcal{W}_1)$ is separable [191, Theorem 6.18], we deduce from [6, Theorem 4.44] that the σ-field on $\mathfrak{B}$ coincides with the trace of the Borel σ-algebra of $\mathbb{R} \times \mathbb{R} \times \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R})$ endowed with the metric $((y, \widetilde{y}, \mu, \widetilde{\mu}), (y', \widetilde{y}', \mu', \widetilde{\mu}')) \mapsto |y-y'| + |\widetilde{y} - \widetilde{y}'| + \mathcal{W}_1(\mu, \mu') + \mathcal{W}_1(\widetilde{\mu}, \widetilde{\mu}')$. Then the measurability of $J(q)$ and $\widetilde{J}(q)$ follows from their continuity in $(y, \widetilde{y}, \mu, \widetilde{\mu})$ with respect to the latter metric, which is clear in view of their definition (4.3.18) which implies by the changes of variables $v = \frac{1-pu-p}{1-p}$ and $v = \frac{(1-p)u}{p}$ that

$$J(q) = \frac{1-p}{p} \int_{\frac{1-q-p}{1-p}}^1 F_{\widetilde{\mu}}^{-1}(v)\, dv + \int_{\frac{q}{p}}^1 F_\mu^{-1}(u)\, du, \quad \widetilde{J}(q) = \frac{p}{1-p} \int_0^{\frac{q}{p}} F_\mu^{-1}(v)\, dv + \int_0^{\frac{1-q-p}{1-p}} F_{\widetilde{\mu}}^{-1}(u)\, du,$$

and the easy fact that if $(a_n)_{n\in\mathbb{N}} \in [0,1]^{\mathbb{N}}$ converges to $a \in [0,1]$ and $(f_n)_{n\in\mathbb{N}} \in L^1([0,1])^{\mathbb{N}}$ converges to $f \in L^1([0,1])$ in L^1, then $\int_0^{a_n} f_n(u)\, du$ converges to $\int_0^a f(u)\, du$ as $n \to +\infty$. $\square$

4.4 Lemma

Lemma 4.4.1. *Let π, π' be two probability measures on $\mathbb{R} \times \mathbb{R}$ with respective first marginals μ and μ'. Let $\eta \in \Pi(\pi, \pi')$. Then $\eta \in \Pi_{bc}(\pi, \pi')$ iff there exist $\chi \in \Pi(\mu, \mu')$ and $(\gamma_{(x,x')}(dy, dy'))_{(x,x')\in\mathbb{R}\times\mathbb{R}}$ such that* (4.1.5) *is satisfied.*

Proof. Let

$$\chi(dx, dx') = \int_{(y,y')\in\mathbb{R}\times\mathbb{R}} \eta(dx, dy, dx', dy'),$$

which is obviously a coupling between μ and μ', and $(\gamma_{(x,x')}(dy, dy'))_{(x,x')\in\mathbb{R}\times\mathbb{R}}$ be a probability kernel such that

$$\eta(dx, dy, dx', dy') = \chi(dx, dx')\,\gamma_{(x,x')}(dy, dy').$$

Let $f : \mathbb{R}^3 \to \mathbb{R}$ be a measurable and bounded function. Let $(\pi_x(dy))_{x\in\mathbb{R}}$ and $(\pi'_{x'}(dy'))_{x'\in\mathbb{R}}$ be probability kernels such that $\pi(dx, dy) = \mu(dx)\,\pi_x(dy)$ and $\pi'(dx', dy') = \mu'(dx')\,\pi'_{x'}(dy')$, and $(\eta_{(x,y)}(dx', dy'))_{(x,y)\in\mathbb{R}\times\mathbb{R}}$, $(\eta_{(x',y')}(dx, dy))_{(x',y')\in\mathbb{R}\times\mathbb{R}}$, $(\eta_x(dx', dy'))_{x\in\mathbb{R}}$ and $(\eta_{x'}(dx, dy))_{x'\in\mathbb{R}}$ be defined by (4.1.3). Then

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\eta_{(x,y)}(dx', dy')\right) \pi(dx, dy) &= \int_{\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\eta(dx, dx', dy, dy')\\ &= \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\gamma_{(x,x')}(dy, dy')\right) \chi(dx, dx'),\end{aligned}$$

and

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\eta_x(dx', dy')\right) \pi(dx, dy) &= \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}} f(x, x', y)\,\pi_x(dy)\right) \mu(dx)\,\eta_x(dx', dy')\\ &= \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}} f(x, x', y)\,\pi_x(dy)\right) \chi(dx, dx').\end{aligned}$$

Therefore we have the following equivalence

$$\begin{aligned}&\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\eta_{(x,y)}(dx', dy')\right) \pi(dx, dy) = \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\eta_x(dx', dy')\right) \pi(dx, dy)\\ \iff &\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y)\,\gamma_{(x,x')}(dy, dy')\right) \chi(dx, dx') = \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}} f(x, x', y)\,\pi_x(dy)\right) \chi(dx, dx').\end{aligned} \tag{4.4.1}$$

Similarly we have

$$\begin{aligned}&\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y')\,\eta_{(x',y')}(dx, dy)\right) \pi'(dx', dy') = \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y')\,\eta_{x'}(dx, dy)\right) \pi'(dx', dy')\\ \iff &\int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}\times\mathbb{R}} f(x, x', y')\,\gamma_{(x,x')}(dy, dy')\right) \chi(dx, dx') = \int_{\mathbb{R}\times\mathbb{R}} \left(\int_{\mathbb{R}} f(x, x', y')\,\pi'_{x'}(dy')\right) \chi(dx, dx').\end{aligned} \tag{4.4.2}$$

Then $\eta \in \Pi_{bc}(\pi, \pi')$ iff the left-hand sides of (4.4.1) and (4.4.2) hold, iff the right-hand sides of (4.4.1) and (4.4.2) hold, i.e. iff $\chi(dx, dx')$-almost everywhere, $\gamma_{x,x'}(dy, dy') \in \Pi(\pi_x, \pi'_{x'})$. $\square$

Part II
Approximation of martingale couplings in the weak adapted topology and applications

This part is the result of a collaboration with Mathias Beiglböck, Benjamin Jourdain and Gudmund Pammer.

Chapter 5

Approximation of martingale couplings on the line in the weak adapted topology

Abstract

The Martingale Optimal Transport (MOT) problem is a variant of the classical Optimal Transport problem. Its additional martingale constraint reflects the condition for a financial market to be arbitrage free. Therefore, it is suited to get model-independent bounds of option prices, as illustrated by Beiglböck, Henry-Labordère and Penkner [23] in a discrete time setting and Galichon, Henry-Labordère and Touzi [84] in a continuous time setting.

With its numerical resolution in mind, Alfonsi, Corbetta and Jourdain [4] asked whether the MOT problem is stable, i.e. if one discretises the problem by approximating the marginals with finitely supported probability measures, does one have convergence of the discrete optimal cost towards the continuous one? Backhoff-Veraguas and Pammer [20] and Wiesel [194] gave a positive answer under regularity assumption on the cost function.

Their stability result is formulated in terms of Wasserstein convergence. However, the topology it induces is not always well suited, especially since it does not reflect the temporal structure of some stochastic processes. Numerous and independent works were done in order to strengthen the usual weak topology and make it reflect the information of a stochastic process. Strikingly, all those topologies are equal, at least in a discrete time setting [15], hence we may focus on the so called adapted Wasserstein distance.

The objective of the present paper is to pave the way to the adaptation of some known results which involve Wasserstein convergence and make sure that they still hold true for the adapted Wasserstein convergence. Our main theorem states that on the real line, any martingale coupling whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance. Consequences of this theorem are discussed in a companion paper.

Keywords: Martingale Optimal Transport, Adapted Wasserstein distance, Robust finance, Convex order, Martingale couplings.

5.1 Introduction and main result

5.1.1 The Martingale Optimal Transport problem

Let (X, d_X), (Y, d_Y) be Polish spaces and $C : X \times Y \to \mathbb{R}_+$ be a nonnegative measurable function. Denote by $\mathcal{P}(X)$ the set of probability measures on X. For $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the classical Optimal Transport problem consists in minimising

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} C(x, y)\, \pi(dx, dy), \tag{OT}$$

where $\Pi(\mu, \nu)$ denotes the set of probability measures in $\mathcal{P}(X \times Y)$ with first marginal μ and second marginal ν. When $X = Y$ and $C = d_X^r$ for some $r \geq 1$, (OT) corresponds to the well-known Wasserstein distance with index r to the power r, denoted $\mathcal{W}_r^r(\mu, \nu)$, see [8, 168, 190, 191] for a study in depth.

The OT theory is a long story: formulated by Gaspard Monge [140] in 1781 and modernised by Kantorovich [114] in 1942, it was rediscovered many times under various forms and has an impressive scope of applications until recently where it became an unmissable tool of data sciences. However, this theory under its classical form is not sufficient to solve some major problems raised by the field of mathematical finance, such as robust model-independent pricing. Indeed, Beiglböck, Henry-Labordère and Penkner [23] showed in a discrete time setting and Galichon, Henry- Labordère and Touzi [84] in a continuous time setting that one would need an additional martingale constraint to (OT) in order to get model-free bounds of an option price. This martingale constraint reflects the condition for a financial market to be arbitrage free.

This leads to the formulation of the Martingale Optimal Transport (MOT) problem: for any $\pi \in \mathcal{P}(X \times Y)$, we denote by $(\pi_x)_{x \in X}$ its disintegration with respect to its first marginal μ. We then write $\pi(dx, dy) = \mu(dx)\, \pi_x(dy)$, or with a slight abuse of notation, $\pi = \mu \times \pi_x$ if the context is not ambiguous. Let $C : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ be a nonnegative measurable function and μ, ν be two probability distributions on the real line with finite first moment. Then the MOT problem consists in minimising

$$\inf_{\pi \in \Pi_M(\mu,\nu)} \int_{\mathbb{R} \times \mathbb{R}} C(x, y)\, \pi(dx, dy), \tag{MOT}$$

where $\Pi_M(\mu, \nu)$ denotes the set of martingale couplings between μ and ν, that is

$$\Pi_M(\mu, \nu) = \left\{ \pi = \mu \times \pi_x \in \Pi(\mu, \nu) \mid \mu(dx)\text{-almost everywhere, } \int_{\mathbb{R}} y\, \pi_x(dy) = x \right\}.$$

According to Strassen's theorem [183], the existence of a martingale coupling between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ with finite first moment is equivalent to $\mu \leq_{cx} \nu$, where $\leq_{cx}$ denotes the convex order. We recall that two finite positive measures μ, ν on $\mathbb{R}^d$ with finite first moment are said to be in the convex order iff we have

$$\int_{\mathbb{R}^d} f(x)\, \mu(dx) \leq \int_{\mathbb{R}^d} f(y)\, \nu(dy),$$

for every convex function $f : \mathbb{R}^d \to \mathbb{R}$. Note that there holds equality for all linear functions, from which we deduce that μ and ν have equal mass and satisfy $\int_{\mathbb{R}^d} x\,\mu(dx) = \int_{\mathbb{R}^d} y\,\nu(dy)$.

For adaptations of celebrated results on classical optimal transport theory to the MOT problem, we refer to Henry-Labordère, Tan and Touzi [105] and Henry-Labordère and Touzi [106]. On duality, we refer to Beiglböck, Nutz and Touzi [27], Beiglböck, Lim and Obłój [26] and De March [68]. We also refer to De March [67] and De March and Touzi [69] for the multi-dimensional case.

About the numerical resolution of the MOT problem, one can look at Alfonsi, Corbetta and Jourdain [3, 4], De March [66], Guo and Obłój [96] and Henry-Labordère [104]. When μ and ν are finitely supported, then the MOT problem amounts to linear programming. In the general case, once the MOT problem is discretised by approximating μ and ν by probability measures with finite support and in the convex order, Alfonsi, Corbetta and Jourdain raised the question of the convergence of the discrete optimal cost towards the continuous one. Partial results were first brought by Guo and Obłój [96] and the stability of left-curtain couplings obtained by Juillet [112]. More recently, Backhoff-Veraguas and Pammer [20] and Wiesel [194] independently gave a positive answer under mild regularity assumption.

5.1.2 The Adapted Wasserstein distance

The stability result shown by Backhoff-Veraguas and Pammer involves Wasserstein convergence. More precisely, let $\mu^k, \nu^k \in \mathcal{P}(\mathbb{R})$, $k \in \mathbb{N}$ be in the convex order and respectively converge to μ and ν in $\mathcal{W}_r$. Under mild assumption, for all $k \in \mathbb{N}$ there exists $\pi^k \in \Pi_M(\mu^k, \nu^k)$, optimal for (MOT), and any accumulation point of $(\pi^k)_{k\in\mathbb{N}}$ with respect to the $\mathcal{W}_r$-convergence is a martingale coupling between μ and ν optimal for (MOT).

However, it turns out that the topology induced by the Wasserstein distance is not always well suited for any setting, especially in mathematical finance. Indeed, the symmetry of this distance does not take into account the temporal structure of martingales. One can easily get convinced that two stochastic processes very close in Wasserstein distance can yield radically unalike information, as [14, Figure 1] illustrates very well. Therefore, one needs to strengthen, or adapt this usual topology. This can be done in many different ways, such as the adapted weak topology (see below), Hellwig's information topology [101], Aldous's extended weak topology [2] or the optimal stopping topology [15]. Strikingly, all those apparently independent topologies are actually equal, at least in discrete time [15, Theorem 1.1].

Hence we may focus on the so called adapted Wasserstein distance. For an extensive background, we refer to [149, 150, 151, 152, 129, 39]. Fix $x_0 \in X$ and $r \geq 1$. We denote the set of all probability measures on X with finite r-th moment by $\mathcal{P}_r(X)$, i.e.

$$\mathcal{P}_r(X) = \left\{ p \in \mathcal{P}(X) \mid \int_X d_X^r(x, x_0)\, p(dx) < +\infty \right\}.$$

Similarly, we denote by $\mathcal{M}(X)$ (resp. $\mathcal{M}_r(X)$) the set of all finite positive measures (resp. with finite r-th moment). The set $\mathcal{M}(X)$, resp. $\mathcal{M}_r(X)$, is equipped with the weak topology induced by the set $C_b(X)$ of all real-valued absolutely bounded continuous functions on X, resp. the set $\Phi_r(X)$ of all real-valued continuous functions on X, $C(X)$, which satisfy the growth constraint

$$\Phi_r(X) = \{f \in C(X) \mid \exists \alpha > 0,\ \forall x \in X,\ |f(x)| \leq \alpha\,(1 + d_X^r(x, x_0))\}\,.$$

A sequence $(\mu^k)_{k\in\mathbb{N}}$ converges in $\mathcal{M}_r(X)$ to μ iff

$$\forall f \in \Phi_r(X), \quad \mu^k(f) \underset{k\to+\infty}{\longrightarrow} \mu(f). \tag{5.1.1}$$

If moreover μ and μ^k, $k \in \mathbb{N}$, all have equal mass, then the convergence (5.1.1) can be equivalently formulated in terms of the Wasserstein distance with index r:

$$\mathcal{W}_r(\mu^k, \mu) := \inf_{\pi\in\Pi(\mu^k,\mu)} \left(\int_{X\times X} d_X^r(x, y)\,\pi(dx, dy)\right)^{\frac{1}{r}} \underset{k\to+\infty}{\longrightarrow} 0.$$

For any $m_0 > 0$, we can then equip the set of finite positive measures in $\mathcal{M}_r(X \times Y)$ with mass m_0 with the Wasserstein topology. However, we can also equip it with a stronger topology, namely the adapted Wasserstein topology. It is induced by the metric $\mathcal{AW}_r$ defined for all $\pi, \pi' \in \mathcal{M}_r(X \times Y)$ such that $\pi(X \times Y) = \pi'(X \times Y) = m_0$ by

$$\mathcal{AW}_r(\pi, \pi') = \inf_{\chi\in\Pi(\mu,\mu')} \left(\int_{X\times X} \left(d_X^r(x, x') + \mathcal{W}_r^r(\pi_x, \pi'_{x'})\right)\,\chi(dx, dx')\right)^{\frac{1}{r}}, \tag{5.1.2}$$

where μ, resp μ' is the first marginal of π, resp. π'. It is easy to check that $\mathcal{W}_r \leq \mathcal{AW}_r$, and therefore $\mathcal{AW}_r$ indeed induces a stronger topology than $\mathcal{W}_r$. Another useful point of view is the following: let $J : \mathcal{M}(X \times Y) \to \mathcal{M}(X \times \mathcal{P}(Y))$ be the inclusion map defined for all $\pi = \mu \times \pi_x \in \mathcal{M}(X \times Y)$ by

$$J(\pi)(dx, dp) = \mu(dx)\,\delta_{\pi_x}(dp).$$

For all $\pi, \pi' \in \mathcal{M}_r(X\times Y)$ with equal mass, their adapted Wasserstein distance coincides with

$$\mathcal{AW}_r(\pi, \pi') = \mathcal{W}_r(J(\pi), J(\pi')). \tag{5.1.3}$$

Therefore, the topology induced by $\mathcal{AW}_r$ coincides with the initial topology w.r.t. J.

Finally, let us mention the interpretation of adapted Wasserstein distance in terms of bicausal couplings as done in [16]. Let $\pi, \pi' \in \mathcal{P}_r(X \times Y)$. Let Z_1, Z_2, Z_1', Z_2' be random variables such that the distribution of (Z_1, Z_2, Z_1', Z_2') is a $\mathcal{W}_r$-optimal coupling between π and π'. In many cases, there exists a Monge transport map $T : X \times Y \to X \times Y$ such that $(Z_1', Z_2') = T(Z_1, Z_2)$. As mentioned in [14], the temporal structure of stochastic processes is then not taken into account since the present value Z_1' is determined from the future value Z_2. Therefore, it is more suitable to restrict to couplings (Z_1, Z_2, Z_1', Z_2') between π and π' such that the conditional distribution of (Z_1', Z_2') (resp. (Z_1, Z_2)) given (Z_1, Z_2) (resp. (Z_1', Z_2')) is equal to the conditional distribution of (Z_1', Z_2') (resp. (Z_1, Z_2)) given Z_1 (resp. Z_1').

Let μ and μ' denote the respective first marginal distributions of π and π' and $\eta \in \Pi(\pi, \pi')$ be a coupling between π and π'. Denote by $(\eta_x(dx', dy'))_{x \in X}$ and $(\eta_{x'}(dx, dy))_{x' \in X}$ the probability kernels such that

$$\int_{y \in Y} \eta(dx, dy, dx', dy') = \mu(dx)\, \eta_x(dx', dy') \text{ and } \int_{y' \in Y} \eta(dx, dy, dx', dy') = \mu'(dx')\, \eta_{x'}(dx, dy).$$

Then η is called bicausal iff for $\pi(dx, dy)$-almost every $(x, y) \in X \times Y$, resp. $\pi(dx', dy')$-almost every $(x', y') \in X \times Y$,

$$\eta_{(x,y)}(dx', dy') = \eta_x(dx', dy'), \quad \text{resp.} \quad \eta_{(x',y')}(dx, dy) = \eta_{x'}(dx, dy).$$

We denote by $\Pi_{bc}(\pi, \pi')$ the set of bicausal couplings between π and π'. Another useful characterisation is η is bicausal iff there exist $\chi \in \Pi(\mu, \mu')$ and $(\gamma_{(x,x')}(dy, dy'))_{(x,x') \in X \times X}$ such that

$$\begin{aligned} &\chi(dx, dx')\text{-almost everywhere}, \quad \gamma_{(x,x')}(dy, dy') \in \Pi(\pi_x, \pi_{x'}) \\ \text{and} \quad &\eta(dx, dy, dx', dy') = \chi(dx, dx')\, \gamma_{(x,x')}(dy, dy'). \end{aligned}$$

Then the adapted Wasserstein distance coincides with

$$\mathcal{AW}_r(\pi, \pi') = \inf_{\eta \in \Pi_{bc}(\pi, \pi')} \left(\int_{X \times Y} \left(d_X^r(x, x') + d_Y^r(y, y') \right) \eta(dx, dy, dx', dy') \right)^{\frac{1}{r}}.$$

One of the objectives of the present paper is to prove that some well-known stability results for the $\mathcal{W}_r$-convergence also hold for the $\mathcal{AW}_r$-convergence. More details are given in Section 5.2.

5.1.3 Outline

Section 5.2 presents the main result of this article, namely Theorem 5.2.5 below. We give an explanation of how the conclusion of this theorem was expectable. Moreover, we give a sketch of its proof in order to help seeing through its technical nature.

Section 5.3 deals with technical lemmas which allow us to deal with difficulties specific to the adapted Wasserstein distance with more ease. They mainly explore properties about approximations and when the addition is continuous.

Section 5.4 focuses on the convex order. It deals with potential functions which are a convenient tool to address the convex order in dimension one. Moreover, it contains the proof of a key result which allows us to see that the main theorem needs only to be proved for irreducible pairs of marginals.

Section 5.5 is as the name suggests devoted to the proof of the main theorem. Before entering its technical proof, we prove that is is enough to prove $\mathcal{AW}_1$-convergence for irreducible pairs of marginals.

5.2 Main result

Our main result is Theorem 5.2.5 below. Before, we state a proposition which enlightens us about how the conclusion of the theorem was actually foreseeable. We also state the generalisation of this proposition to Polish spaces. Then, we state a proposition which is a key result to argue that the theorem needs only to be proved when the limit pair is irreducible. Next, we state the theorem with a sketch of its proof. It is understood that (X, d_X) and (Y, d_Y) denote any Polish spaces, and (x_0, y_0) is a fixed element of $X \times Y$.

It is well-known that when one considers convergent sequences of marginals $(\mu^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$ all with equal mass respectively to $\mu, \nu \in \mathcal{M}_r(X)$, then, informally speaking, there holds

$$\Pi(\mu^k, \nu^k) \underset{k\to+\infty}{\longrightarrow} \Pi(\mu, \nu) \quad \text{in } \mathcal{W}_r, \tag{5.2.1}$$

i.e., any sequence with convergent marginals has accumulation points in $\Pi(\mu, \nu)$, and for any $\pi \in \Pi(\mu, \nu)$ there holds

$$\inf_{\pi^k \in \Pi(\mu^k, \nu^k)} \mathcal{W}_r^r(\pi, \pi^k) \leq \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k) \underset{k\to+\infty}{\longrightarrow} 0. \tag{5.2.2}$$

Indeed, let $\eta^k \in \Pi(\mu^k, \mu)$, resp. $\tau^k \in \Pi(\nu, \nu^k)$ be optimal for $\mathcal{W}_r(\mu^k, \mu)$, resp. $\mathcal{W}_r(\nu, \nu^k)$ and set

$$\pi^k(dx^k, dy^k) = \int_{(x,y)\in X\times Y} \eta^k(dx^k, dx)\, \pi_x(dy)\, \tau_y^k(dy^k).$$

The next two propositions establish (5.2.1) with respect to $\mathcal{AW}_r$ for finite positive measures with common mass. The first one is formulated for $X = Y = \mathbb{R}$ and provides under mild assumptions an estimate of $\inf_{\pi^k \in \Pi(\mu^k, \nu^k)} \mathcal{AW}_r^r(\pi, \pi^k)$ with respect to the marginals as in (5.2.2). Its proof relies on unidimensional tools, which we recall here. For η a probability distribution on $\mathbb{R}$, we denote by $F_\eta : x \mapsto \eta((-\infty, x])$ its cumulative distribution function, and by $F_\eta^{-1} : (0,1) \to \mathbb{R}$ its quantile function defined for all $u \in (0,1)$ by

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

The following properties are standard results (see for instance Section 2.6 for proofs):

(a) F_η is càdlàg, F_η^{-1} is càglàd ;

(b) For all $(x, u) \in \mathbb{R} \times (0,1)$,

$$F_\eta^{-1}(u) \leq x \iff u \leq F_\eta(x), \tag{5.2.3}$$

which implies

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{5.2.4}$$

$$\text{and} \quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{5.2.5}$$

(c) For $\mu(dx)$-almost every $x \in \mathbb{R}$,

$$0 < F_\eta(x), \quad F_\eta(x-) < 1 \quad \text{and} \quad F_\eta^{-1}(F_\eta(x)) = x;$$

(d) The image of the Lebesgue measure on $(0,1)$ by F_η^{-1} is η.

The property (d) is referred to as inverse transform sampling.

Proposition 5.2.1. *Let $\mu, \mu^k, \nu, \nu^k \in \mathcal{M}_r(\mathbb{R})$, $k \in \mathbb{N}$, be with equal mass such that μ^k (resp. ν^k) converges to μ (resp. ν) in $\mathcal{W}_r$. Let $\pi \in \Pi(\mu, \nu)$. Then:*

(a) There exists a sequence $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to π in $\mathcal{AW}_r$;

(b) If for all $x \in \mathbb{R}$ and $k \in \mathbb{N}$ with $\mu^k(\{x\}) > 0$, there exists $x' \in \mathbb{R}$ such that

$$\mu((-\infty, x')) \leq \mu^k((-\infty, x)) < \mu^k((-\infty, x]) \leq \mu((-\infty, x']),$$

which is for instance always satisfied when μ^k is non-atomic, then

$$\mathcal{AW}_r^r(\pi, \pi^k) \leq \mathcal{W}_r^r(\mu, \mu^k) + \mathcal{W}_r^r(\nu, \nu^k). \tag{5.2.6}$$

Remark 5.2.2. If π is a martingale coupling, i.e. $\int_\mathbb{R} y' \, \pi_{x'}(dy') = x'$, $\mu(dx')$-almost everywhere, then for $\chi^k \in \Pi(\mu^k, \mu)$ an optimal coupling for $\mathcal{AW}_r(\pi^k, \pi)$, we have

$$\begin{aligned}
\int_\mathbb{R} \left| x - \int_\mathbb{R} y \, \pi_x^k(dy) \right|^r \mu^k(dx) &= \int_{\mathbb{R}\times\mathbb{R}} \left| x - \int_\mathbb{R} y \, \pi_x^k(dy) \right|^r \chi^k(dx, dx') \\
&\leq 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \left| x' - \int_\mathbb{R} y \, \pi_x^k(dy) \right|^r \right) \chi^k(dx, dx') \\
&= 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \left| \int_\mathbb{R} y' \, \pi_{x'}(dy') - \int_\mathbb{R} y \, \pi_x^k(dy) \right|^r \right) \chi^k(dx, dx') \\
&\leq 2^{r-1} \int_{\mathbb{R}\times\mathbb{R}} \left(|x - x'|^r + \mathcal{W}_1^r(\pi_x^k, \pi_{x'}) \right) \chi^k(dx, dx') \\
&\leq 2^{r-1} \mathcal{AW}_r^r(\pi, \pi^k) \underset{k\to+\infty}{\longrightarrow} 0.
\end{aligned}$$

In that sense, $\pi^k, k \in \mathbb{N}$ is almost a sequence of martingale couplings.

In the setting of Proposition 5.2.1, if μ^k and ν^k are also in the convex order and π is a martingale coupling, then in view of Remark 5.2.2 one would naturally expect that π^k can be slightly modified into a martingale coupling and still converge to π in $\mathcal{AW}_r$. This actually requires a lot of work and is the whole purpose of Theorem 5.2.5 below. We mention that the previous proposition generalises to any Polish spaces X and Y, as the next proposition states, but unfortunately without providing an estimate.

Proposition 5.2.3. *Let $\mu, \mu^k \in \mathcal{M}_r(X), \nu, \nu^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$, all with equal mass and such that μ^k (resp. ν^k) converges to μ (resp. ν) in $\mathcal{W}_r$. Let $\pi \in \Pi(\mu, \nu)$. Then there exists a sequence $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to π in $\mathcal{AW}_r$.*

The next proposition is a key ingredient which allows us to reduce the proof of Theorem 5.2.5 below to the case of irreducible pairs of marginals. For $\mu \in \mathcal{M}_1(\mathbb{R})$, we denote by u_μ its potential function, that is the map defined for all $y \in \mathbb{R}$ by $u_\mu(y) = \int_{\mathbb{R}} |y-x|\,\mu(dx)$ (see Section 5.4 for more details). We recall that a pair (μ,ν) of finite positive measures in convex order is called irreducible if $I = \{u_\mu < u_\nu\}$ is an interval and, $\mu(I)$ and $\nu(\overline{I})$ have the total mass. If $a \in \mathbb{R}$ is such that $\nu([a,+\infty)) = 0$, then the convex order implies $\mu([a,+\infty)) = 0$, hence

$$u_\mu(a) = a - \int_{\mathbb{R}} x\,\mu(dx) = a - \int_{\mathbb{R}} y\,\nu(dy) = u_\nu(a),$$

so $a \notin I$. Similarly, $\nu((-\infty,a]) = 0 \implies a \notin I$. We deduce that ν must assign positive mass to any neighbourhood of each of the boundaries of I.

According to [25, Theorem A.4], for any pair (μ,ν) of probability measures in convex order, there exist $N \subset \mathbb{N}$ and a sequence $(\mu_n,\nu_n)_{n\in N}$ of irreducible pairs of sub-probability measures in convex order such that

$$\mu = \eta + \sum_{n\in N} \mu_n, \quad \nu = \eta + \sum_{n\in\mathbb{N}} \nu_n \quad \text{and} \quad \{u_\mu < u_\nu\} = \bigcup_{n\in N} \{u_{\mu_n} < u_{\nu_n}\},$$

where the union is disjoint and $\eta = \mu|_{\{u_\mu = u_\nu\}}$. The sequence $(\mu_n,\nu_n)_{n\in N}$ is unique up to rearrangement of the pairs and is called the decomposition of (μ,ν) into irreducible components. Moreover, for any martingale coupling $\pi \in \Pi_M(\mu,\nu)$, there exists a unique sequence of martingale couplings $\pi_n \in \Pi_M(\mu_n,\nu_n)$, $n \in N$ such that

$$\pi = \chi + \sum_{n\in N} \pi_n,$$

where $\chi = (\mathrm{id},\mathrm{id})_*\eta$ and $*$ denotes the pushforward operation. This sequence satisfies

$$\forall n \in N, \quad \pi_n(dx,dy) = \mu_n(dx)\,\pi_x(dy). \tag{5.2.7}$$

Proposition 5.2.4. *Let $(\mu^k,\nu^k)_{k\in\mathbb{N}}$ be a sequence of pairs of probability measures on the real line in convex order which converge to (μ,ν) in $\mathcal{W}_1$. Let $(\mu_n,\nu_n)_{n\in N}$ be the decomposition of (μ,ν) into irreducible components and $\eta = \mu|_{\{u_\mu=u_\nu\}}$. Then there exists for any $k \in \mathbb{N}$ a decomposition of (μ^k,ν^k) into pairs of sub-probability measures $(\mu^k_n,\nu^k_n)_{n\in N}$, (η^k,υ^k) which are in convex order such that*

$$\eta^k + \sum_{n\in N} \mu^k_n = \mu^k, \quad \upsilon^k + \sum_{n\in N} \nu^k_n = \nu^k, \quad k \in \mathbb{N}, \tag{5.2.8}$$

$$\lim_{k\to+\infty} \eta^k = \eta, \quad \lim_{k\to+\infty} \mu^k_n = \mu_n, \quad \lim_{k\to+\infty} \nu^k_n = \nu_n, \quad \lim_{k\to+\infty} \upsilon^k = \eta \quad \text{in } \mathcal{W}_1. \tag{5.2.9}$$

We can now state our main result, namely Theorem 5.2.5 below. Any martingale coupling whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapated Wasserstein distance.

Theorem 5.2.5. *Let $\mu^k,\nu^k \in \mathcal{P}_r(\mathbb{R})$, $k \in \mathbb{N}$, be in convex order and respectively converge to μ and ν in $\mathcal{W}_r$. Let $\pi \in \Pi_M(\mu,\nu)$. Then there exists a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k,\nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_r$.*

Sketch of the proof. We will first argue that it is enough to consider the case $r = 1$. Thanks to Proposition 5.2.4, we can also reduce the proof to the case of irreducible pairs of marginals (μ, ν), whose single irreducible component is denoted $(l, r) = I$.

Step 1. Fix any martingale coupling $\pi \in \Pi_M(\mu, \nu)$. By directly approximating π we would face technical obstacles. First, for K a compact subset of I, $\mu|_K \times \pi_x$ is not necessarily compactly supported. Moreover, ν may put mass on the boundary of I. To overcome this, the kernel π_x is first compactified to a compact set $[-R, R]$, where $R > 0$, and then pushed forward by the map $y \mapsto \alpha(y-x)+x$, where $\alpha \in (0, 1)$. This yields a martingale coupling $\pi^{R,\alpha}$ close to π and easier to approximate, between μ and a probability measure $\nu^{R,\alpha}$ dominated by ν in the convex order. We find compact sets $K, L \subset I$ such that the restriction $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is compactly supported on $K \times L$ and concentrated on $K \times \mathring{L}$, where $\mathring{L}$ denotes the interior of L. Since by irreducibility ν puts mass to any neighbourhood of the boundary of I, $\nu^{R,\alpha}$ assigns positive mass to two open sets L_-, L_+ on both sides of K with positive distance to K. This is summarised in Figure 5.1, where J denotes a compact subset of I large enough.

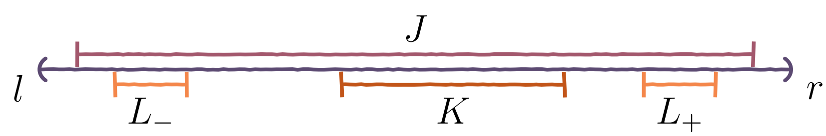

Figure 5.1: Intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parenthesis.

Step 2. It is possible to find an approximative sequence $(\hat{\pi}^k = \hat{\mu}^k \times \hat{\pi}^k_x)_{k\in\mathbb{N}}$ of the sub-probability martingale coupling $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ from step 1. Unfortunately $\hat{\pi}^k$ is not necessarily a martingale coupling. Therefore, we free up some mass, and use the one available on the left and right of K in L_- and L_+ to adjust the barycentres of the kernels π^k_x. Hence we find a sequence $(\tilde{\pi}^k = \hat{\mu}^k \times \tilde{\pi}^k_x)_{k\in\mathbb{N}}$ of sub-probability martingale couplings approximating $\pi^{R,\alpha}|_{K\times\mathbb{R}}$.

Step 3. By construction, up to multiplication by a factor smaller than and close to 1, the first marginal of $\tilde{\pi}^k$ satisfies $\hat{\mu}^k \leq \mu^k$. Moreover, its second marginal denoted $\tilde{\nu}^k$ is such that there exists a probability measure $\nu^{R,\alpha,k}$ which satisfies $\tilde{\nu}^k \leq \nu^{R,\alpha,k} \leq_{cx} \nu^k$. Then by using the uniform convergence of potential functions, we show that for k sufficiently large there exist sub-probability martingale couplings $\eta^k \in \Pi_M(\mu^k - \hat{\mu}^k, \nu^{R,\alpha,k} - \tilde{\nu}^k)$ so that the sum $\eta^k + \tilde{\pi}^k$ is a martingale coupling in $\Pi_M(\mu^k, \nu^{R,\alpha,k})$, where the second marginal is dominated by ν^k in the convex order.

Step 4. In the last step, we use the inverse-transform martingale coupling between $\nu^{R,\alpha,k}$ and ν^k, see Chapter 2, to change $\eta^k + \tilde{\pi}^k$ to a martingale coupling $\pi^k \in \Pi(\mu^k, \nu^k)$. Finally, we estimate the $\mathcal{AW}_1$-distance of π to π^k. $\square$

5.3 On the weak adapted topology

We begin this section with a Lemma on uniform integrability which will prove very handy throughout the paper. We formulate it for finite positive measures on X, but it is understood that (X, x_0) is replaced with (Y, y_0) for measures on Y.

Lemma 5.3.1. *Let $r \geq 1$ and $\mu \in \mathcal{M}_r(X)$. For $\varepsilon > 0$, let*

$$I_\varepsilon^r(\mu) := \sup_{\substack{\tau \in \mathcal{M}(X) \\ \tau \leq \mu,\ \tau(X) \leq \varepsilon}} \int_X d_X^r(x, x_0)\, \tau(dx). \tag{5.3.1}$$

(a) I_ε^r is monotone in μ, i.e., $\mu \leq \mu' \in \mathcal{M}_r(X)$ implies that $I_\varepsilon^r(\mu) \leq I_\varepsilon^r(\mu')$.

(b) The value of $I_\varepsilon^r(\mu)$ vanishes as $\varepsilon \to 0$.

(c) For any $\mu' \in \mathcal{M}_r(X)$ such that $\mu(X) = \mu'(X)$ we have

$$I_\varepsilon^r(\mu) \leq 2^{r-1} \left(I_\varepsilon^r(\mu') + \mathcal{W}_r^r(\mu, \mu') \right). \tag{5.3.2}$$

(d) Let $\mu, \mu^k \in \mathcal{M}_r(X)$, $k \in \mathbb{N}$ be with equal mass such that μ^k converges weakly to μ. Then

$$\mathcal{W}_r(\mu^k, \mu) \underset{k \to +\infty}{\longrightarrow} 0 \iff \sup_{k \in \mathbb{N}} I_\varepsilon^r(\mu^k) \underset{\varepsilon \to 0}{\longrightarrow} 0 \quad \textit{and} \quad \sup_{k \in \mathbb{N}} \int_X d_X^r(x, x_0)\, \mu^k(dx) < +\infty.$$

(e) Finally, if $X = \mathbb{R}^d$ and $\mu \leq_{cx} \nu$ with $\nu \in \mathcal{M}_1(\mathbb{R}^d)$, then $I_\varepsilon^1(\mu) \leq I_\varepsilon^1(\nu)$.

Remark 5.3.2. If $\mu(X) \leq \varepsilon$, then $I_\varepsilon^r(\mu)$ is simply the r-th moment of μ.

Proof. The first point (a) is an easy consequence of the definition of I_ε^r.

Next we check (b). Let $\mu \in \mathcal{M}_r(X)$ be such that $\mu(X) > 0$. Since

$$I_\varepsilon^r(\mu) = \mu(X) I^r_{\frac{\varepsilon}{\mu(X)}} \left(\frac{\mu}{\mu(X)} \right), \tag{5.3.3}$$

to check convergence of $I_\varepsilon^r(\mu)$ to 0 with ε, we may suppose that $\mu \in \mathcal{P}_r(X)$. Let $\varepsilon \in (0, 1)$. For $\eta \in \mathcal{M}_r(X)$, we denote by $\overline{\eta}$ the image of η by the map $x \mapsto d_X^r(x, x_0)$. It is then enough to show that

$$I_\varepsilon^r(\mu) = \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\, du. \tag{5.3.4}$$

Indeed, since $\mu \in \mathcal{P}_r(X)$ we have $\int_X d_X(x, x_0)^r\, \mu(dx) = \int_0^1 F_{\overline{\mu}}^{-1}(u)\, du < +\infty$, so we conclude by dominated convergence that $I_\varepsilon^r(\mu)$ vanishes with ε. Of course an appropriate upper bound would be sufficient, but (5.3.4) will come in handy for the proof of the last part. Let $\tau^* \in \mathcal{M}(X)$ be such that $\overline{\tau^*}$ is the right-most measure dominated by $\overline{\mu}$ with mass equal to ε, namely

$$\tau^*(dx) = \left(\mathbb{1}_{A_\varepsilon}(x) + \frac{F_{\overline{\mu}}(y_\varepsilon) - (1 - \varepsilon)}{\mu(B_\varepsilon)} \mathbb{1}_{B_\varepsilon}(x) \right) \mu(dx),$$

where

$$y_\varepsilon = F_{\overline{\mu}}^{-1}(1 - \varepsilon), \quad A_\varepsilon = \{x \in \mathbb{R} \mid d_X^r(x, x_0) > y_\varepsilon\} \quad \text{and} \quad B_\varepsilon = \{x \in \mathbb{R} \mid d_X^r(x, x_0) = y_\varepsilon\},$$

and the second summand of the right-hand side is zero if $\mu(B_\varepsilon) = 0$. By (5.2.5) and definition of y_ε, we have

$$F_{\overline{\mu}}(y_\varepsilon-) \leq 1-\varepsilon, \tag{5.3.5}$$

hence $\mu(B_\varepsilon) = \overline{\mu}(\{y_\varepsilon\}) \geq F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon)$ and therefore $\tau^* \leq \mu$. Moreover,

$$\tau^*(\mathbb{R}) = \mu(A_\varepsilon) + F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon) = \overline{\mu}((y_\varepsilon, +\infty)) - (1 - F_{\overline{\mu}}(y_\varepsilon)) + \varepsilon = \varepsilon.$$

Let us show that

$$\int_X d_X^r(x, x_0)\, \tau^*(dx) = \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\, du. \tag{5.3.6}$$

According to (5.3.5), $(1-\varepsilon, F_{\overline{\mu}}(y_\varepsilon)] \subset (F_{\overline{\mu}}(y_\varepsilon-), F_{\overline{\mu}}(y_\varepsilon)]$, so by (5.2.4), for all $u \in (1-\varepsilon, F_{\overline{\mu}}(y_\varepsilon)]$, $F_{\overline{\mu}}^{-1}(u) = y_\varepsilon$. Then, using the inverse transform sampling and (5.2.3) for the second equality, we get

$$\begin{aligned}
\int_X d_X^r(x, x_0)\, \tau^*(dx) &= \int_{\mathbb{R}} y \mathbb{1}_{(y_\varepsilon, +\infty)}(y)\, \overline{\mu}(dy) + (F_{\overline{\mu}}(y_\varepsilon) - (1-\varepsilon)) y_\varepsilon \\
&= \int_{F_{\overline{\mu}}(y_\varepsilon)}^1 F_{\overline{\mu}}^{-1}(u)\, du + \int_{1-\varepsilon}^{F_{\overline{\mu}}(y_\varepsilon)} F_{\overline{\mu}}^{-1}(u)\, du \\
&= \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\, du,
\end{aligned}$$

hence (5.3.6) holds.

Let $\tau \in \mathcal{M}(X)$ be such that $\tau \leq \mu$ and $0 < \tau(X) \leq \varepsilon$. In view of (5.3.6), it is enough in order to prove (5.3.4) to show that

$$\int_X d_X^r(x, x_0)\, \tau(dx) \leq \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\, du. \tag{5.3.7}$$

Since $\tau \leq \mu$, we have $\overline{\tau} \leq \overline{\mu}$. Using (5.2.5) for the last inequality, we get for all $u \in (0, 1)$

$$1 - F_{\overline{\tau}/\tau(X)}\left(F_{\overline{\mu}}^{-1}(1 - \tau(X)u)\right) = \frac{\overline{\tau}((F_{\overline{\mu}}^{-1}(1-\tau(X)u), +\infty))}{\tau(X)} \leq \frac{\overline{\mu}((F_{\overline{\mu}}^{-1}(1-\tau(X)u), +\infty))}{\tau(X)} \leq u,$$

hence $F_{\overline{\tau}/\tau(X)}(F_{\overline{\mu}}^{-1}(1-\tau(X)u)) \geq 1-u$ and by (5.2.3), $F_{\overline{\tau}/\tau(X)}^{-1}(1-u) \leq F_{\overline{\mu}}^{-1}(1-\tau(X)u)$. Using the inverse transform sampling, we deduce

$$\begin{aligned}
\int_X d_X^r(x, x_0)\, \tau(dx) &= \tau(X) \int_0^1 F_{\overline{\tau}/\tau(X)}^{-1}(u)\, du = \tau(X) \int_0^1 F_{\overline{\tau}/\tau(X)}^{-1}(1-u)\, du \\
&\leq \tau(X) \int_0^1 F_{\overline{\mu}}^{-1}(1-\tau(X)u)\, du = \int_{1-\tau(X)}^1 F_{\overline{\mu}}^{-1}(u)\, du \leq \int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(u)\, du,
\end{aligned}$$

hence (5.3.7) holds.

To see (c), fix $\mu' \in \mathcal{M}(X)$ with $\mu(X) = \mu'(X)$. We denote by $\pi(dx, dx') = \mu(dx)\, \pi_x(dx') \in \Pi(\mu, \mu')$ a $\mathcal{W}_r$-optimal coupling. Let $\tau \in \mathcal{M}(X)$ be such that $\tau \leq \mu$ and $\tau(X) \leq \varepsilon$. Let $\tau' \in \mathcal{M}(X)$ be defined by

$$\tau'(dx') = \int_{x \in X} \pi_x(dx')\, \tau(dx).$$

Since π is element of $\Pi(\mu, \mu')$, we find $\tau' \le \mu'$ and $\tau(X) = \tau'(X)$. Then

$$\int_X d_X^r(x, x_0)\, \tau(dx) \le 2^{r-1} \int_{X\times X} \left(d_X^r(x', x_0) + d_X^r(x, x')\right)\, \pi_x(dx')\, \tau(dx)$$
$$\le 2^{r-1} \left(I_\varepsilon^r(\mu') + \int_{X\times X} d_X^r(x, x')\, \pi(dx, dx') \right),$$

which shows by optimality of π the assertion.

We now show (d). Let $\mu, \mu^k \in \mathcal{M}_r(X)$ be with equal mass such that μ^k converges weakly to μ. According to (5.3.3), we may suppose that $\mu, \mu^k \in \mathcal{P}_r(X)$.

Suppose that $\mathcal{W}_r(\mu^k, \mu)$ vanishes as k goes to $+\infty$. Then the sequence of the r-th moments of μ^k, $k \in \mathbb{N}$ is bounded since it converges to the r-th moment of μ. Let $\eta > 0$. Let $k_0 \in \mathbb{N}$ be such that for all $k > k_0$, $\mathcal{W}_r^r(\mu^k, \mu) < \eta$. Then (c) yields for $\varepsilon > 0$

$$\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) \le \sum_{k\le k_0} I_\varepsilon^r(\mu^k) + \sup_{k>k_0} I_\varepsilon^r(\mu^k) \le \sum_{k\le k_0} I_\varepsilon^r(\mu^k) + 2^{r-1}(I_\varepsilon^r(\mu) + \eta).$$

According to (b) we then get

$$\limsup_{\varepsilon\to 0} \sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) \le 2^{r-1}\eta.$$

Since $\eta > 0$ is arbitrary, we deduce that $\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k)$ vanishes with ε.

Conversely, suppose that $\sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k)$ vanishes with ε and the sequence of the r-th moments of μ^k, $k \in \mathbb{N}$ is bounded. By Skorokhod's representation theorem, there exist random variables X and X^k, $k \in \mathbb{N}$, defined on a common probability space such that X, resp. X^k is distributed according to μ, resp. μ^k and X^k converges almost surely to X. Then for all $M > 0$,

$$\mathcal{W}_r^r(\mu^k, \mu) \le \mathbb{E}[d_X^r(X^k, X)] = \mathbb{E}[d_X^r(X^k, X)\mathbb{1}_{\{d_X^r(X^k,X)<M\}}] + \mathbb{E}[d_X^r(X^k, X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}].$$

By the dominated convergence theorem, we deduce

$$\limsup_{k\to+\infty} \mathcal{W}_r^r(\mu^k, \mu) \le \limsup_{k\to+\infty} \mathbb{E}[d_X^r(X^k, X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}].$$

Let us then prove that the right-hand side vanished as M goes to $+\infty$. Let $\eta > 0$. Let $\varepsilon > 0$ be such that $I_\varepsilon^r(\mu) + \sup_{k\in\mathbb{N}} I_\varepsilon^r(\mu^k) < \eta$. By Markov's inequality, we have

$$\sup_{k\in\mathbb{N}} \mathbb{E}[\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] \le \sup_{k\in\mathbb{N}} \frac{\mathbb{E}[d_X^r(X^k, X)]}{M} \le \frac{2^{r-1}}{M} \sup_{k\in\mathbb{N}} \int_X d_X^r(x, x_0)\, (\mu^k + \mu)(dx),$$

where the right-hand side vanishes as M goes to $+\infty$. Therefore, there exists $M_0 > 0$ such that for all $k \in \mathbb{N}$ and $M > M_0$,

$$\mathbb{E}[d_X^r(X^k, X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] \le 2^{r-1} \left(\mathbb{E}[d_X^r(X^k, x_0)\, \mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] + \mathbb{E}[d_X^r(x_0, X)\mathbb{1}_{\{d_X^r(X^k,X)\ge M\}}] \right)$$
$$\le 2^{r-1} \left(I_\varepsilon^r(\mu^k) + I_\varepsilon^r(\mu) \right) < 2^{r-1}\eta.$$

Therefore, for all $M > M_0$,

$$\limsup_{k\to+\infty} \mathbb{E}[d_X^r(X^k, X)\mathbb{1}_{\{d_X^r(X^k,X)\geq M\}}] \leq 2^{r-1}\eta.$$

Since η is arbitrary, this proves the assertion.

Finally, we want to show (e). Let $X = \mathbb{R}^d$ and $\mu \leq_{cx} \nu$ with $\nu \in \mathcal{M}_1(\mathbb{R}^d)$. According to (5.3.3), we may suppose that $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$. Again, we write $\overline{\mu}$ and $\overline{\nu}$ for the pushforward measures of μ and ν under the map $(x \mapsto |x - x_0|^r)$. First, we note that $\overline{\mu}$ is dominated by $\overline{\nu}$ in the increasing convex order. Indeed, let $f \in C(X)$ be convex and nondecreasing, then $x \mapsto f(|x - x_0|^r)$ constitutes a convex, continuous function. Thus,

$$\int_{\mathbb{R}} f(y)\,\overline{\mu}(dy) = \int_{\mathbb{R}^d} f(|x-x_0|^r)\,\mu(dx) \leq \int_{\mathbb{R}^d} f(|x-x_0|^r)\,\nu(dx) = \int_{\mathbb{R}} f(y)\,\overline{\nu}(dy).$$

The convex increasing order is characterised by the following family of inequalities (see for instance [[3], Theorem 2.4]): for all $0 \leq \varepsilon \leq 1$,

$$\int_{1-\varepsilon}^1 F_{\overline{\mu}}^{-1}(y)\,dy \leq \int_{1-\varepsilon}^1 F_{\overline{\nu}}^{-1}(y)\,dy.$$

The identity (5.3.4) concludes the proof. □

We now prove Proposition 5.2.1. A handy tool in the construction of the approximative sequence $(\pi^k)_{k\in\mathbb{N}}$ are copulas. Recall that a two-dimensional copula is an element C of $\Pi(\lambda, \lambda)$ where λ is the uniform distribution on $(0, 1)$. A coupling π is element of $\Pi(\mu, \nu)$ if and only if it can be written as the push-forward of a copula C under the quantile map $(F_\mu^{-1}, F_\nu^{-1})\colon (0,1) \times (0,1) \to \mathbb{R} \times \mathbb{R}$. Clearly, if C is a copula then $\pi = (F_\mu^{-1}, F_\nu^{-1})_*C$ is contained in $\Pi(\mu, \nu)$. Contrary, when $\pi \in \Pi(\mu, \nu)$ is given, we can construct a copula C by

$$C(du, dv) = \mathbb{1}_{(0,1)}(u)\,du\,C_u(dv),$$

where C_u is given by

$$C_u = ((y, w) \mapsto F_\nu(y-) + w\nu(\{y\}))_*(\pi_{F_\mu^{-1}(u)} \times \lambda). \tag{5.3.8}$$

In particular, we have that $u \mapsto C_u$ is constant on the jumps on F_μ. By the inverse transform sampling, showing that $\pi = (F_\mu^{-1}, F_\nu^{-1})_*C$ amounts to showing that the second marginal distribution of C is indeed uniformly distributed on $(0, 1)$, which is a direct consequence of the inverse transform sampling and the well-known result (see for instance Lemma 2.6.6 for a proof) that for any $\eta \in \mathcal{P}(\mathbb{R})$,

$$((z, w) \mapsto F_\eta(z-) + w\eta(\{z\}))_*(\eta \times \lambda) = \lambda. \tag{5.3.9}$$

Moreover, we have by (5.2.4) that $F_\nu^{-1}(F_\nu(y-) + w\nu(\{y\})) = y$ for all $(y, w) \in \mathbb{R} \times (0, 1]$, hence for all $u \in (0, 1)$,

$$\pi_{F_\mu^{-1}(u)} = (F_\nu^{-1})_*C_u. \tag{5.3.10}$$

Proof of Proposition 5.2.1. Because of homogeneity of the $\mathcal{AW}_r$- and $\mathcal{W}_r$-distances, we can suppose w.l.o.g. that μ, μ^k, ν, ν^k and π are probability measures. Let C be the copula defined by $C(du, dv) = \mathbb{1}_{(0,1)}(u)\, du\, C_u(dv)$, where C_u is given by (5.3.8).

In order to define π^k, we construct associated copulas C^k where $u \mapsto C_u^k$ is constant on the jumps of F_{μ^k}. Let

$$\theta_k\colon \mathbb{R} \times (0,1) \to (0,1),\ (x, w) \mapsto F_{\mu^k}(x-) + w\mu^k(\{x\}),$$
$$C_u^k(dv) = \int_{w=0}^{1} C_{\theta_k\left(F_{\mu^k}^{-1}(u),w\right)}(dv)\, dw,$$
$$\pi^k = (F_{\mu^k}^{-1}, F_{\nu^k}^{-1})_* C^k = (F_{\mu^k}^{-1}, F_{\nu^k}^{-1})_*(\mathbb{1}_{(0,1)}(u)\, du\, C_u^k(dv)).$$

The fact that C^k is a copula, and therefore $\pi^k \in \Pi(\mu^k, \nu^k)$, is a direct consequence of (5.3.9) and the inverse transform sampling. Since $u \mapsto C_u$ and $u \mapsto C_u^k$ are constant on the jumps of F_μ and F_{μ^k} respectively, reasoning like in the derivation of (5.3.10), we have for du-almost every u in $(0,1)$

$$\pi_{F_\mu^{-1}(u)} = (F_\nu^{-1})_* C_u, \quad \pi^k_{F_{\mu^k}^{-1}(u)} = (F_{\nu^k}^{-1})_* C_u^k.$$

Moreover, since $(u \mapsto (F_\mu^{-1}(u), F_{\mu^k}^{-1}(u)))_*\lambda$ is a coupling between μ and μ^k, namely the comonotonous coupling, we have

$$\begin{aligned}\mathcal{AW}_r^r(\pi, \pi^k) &\le \int_0^1 \left(|F_\mu^{-1}(u) - F_{\mu^k}^{-1}(u)|^r + \mathcal{W}_r^r(\pi_{F_\mu^{-1}(u)}, \pi^k_{F_{\mu^k}^{-1}(u)})\right) du \\ &= \mathcal{W}_r^r(\mu, \mu^k) + \int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u, (F_{\nu^k}^{-1})_* C_u^k\right) du.\end{aligned} \tag{5.3.11}$$

By Minkowski's inequality we have

$$\begin{aligned}\left(\int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u, (F_{\nu^k}^{-1})_* C_u^k\right) du\right)^{\frac{1}{r}} &\le \left(\int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k\right) du\right)^{\frac{1}{r}} \\ &\quad + \left(\int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u^k, (F_{\nu^k}^{-1})_* C_u^k\right) du\right)^{\frac{1}{r}}.\end{aligned} \tag{5.3.12}$$

Since for any $\eta \in \mathcal{P}(\mathbb{R})$ the map $F_\eta^{-1} \circ F_{C_u^k}^{-1}$ is non-decreasing, we have (see for instance [4, Lemma A.3]) that for dw-almost every $w \in (0,1)$,

$$F_\eta^{-1}(F_{C_u^k}^{-1}(w)) = F^{-1}_{(F_\eta^{-1})_* C_u^k}(w).$$

Hence, we deduce

$$\begin{aligned}\int_{(0,1)} \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u^k, (F_{\nu^k}^{-1})_* C_u^k\right) du &= \int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(F_{C_u^k}^{-1}(w)) - F_{\nu^k}^{-1}(F_{C_u^k}^{-1}(w))|^r\, dw\, du \\ &= \int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(v) - F_{\nu^k}^{-1}(v)|^r\, C_u^k(dv)\, du \\ &= \int_{(0,1)} |F_\nu^{-1}(v) - F_{\nu^k}^{-1}(v)|^r\, dv = \mathcal{W}_r^r(\nu, \nu^k) \to 0,\end{aligned} \tag{5.3.13}$$

where we used inverse transform sampling in the second equality. At this stage, we can already show (b). Indeed, the assumption made in (b) ensures that any jump of F_{μ_k} is included in a jump of F_μ. We already noted that $u \mapsto C_u$ is constant on the jumps of F_μ and therefore also constant on the jumps of F_{μ^k}. This yields for all $u, w \in (0,1)$ that $C_{\theta_k(F_{\mu^k}^{-1}(u),w)} = C_u$ and particularly $C_u^k = C_u$, which lets vanish the first term in the right-hand side of (5.3.12). Then the estimate (5.2.6) follows immediately from (5.3.11), (5.3.12) and (5.3.13).

To obtain (a) and in view of (5.3.11), (5.3.12) and (5.3.13), it is sufficient to show

$$\int_0^1 \mathcal{W}_r^r\left((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k\right) du \to 0.$$

This is achieved in two steps: First, we show for du-almost every $u \in (0,1)$ that

$$\mathcal{W}_r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \to 0. \tag{5.3.14}$$

Second, we prove that

$$u \mapsto \mathcal{W}_r^r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \quad k \in \mathbb{N}, \tag{5.3.15}$$

is uniformly integrable on $(0,1)$ with respect to λ.

To show (5.3.14), note that $\mathcal{W}_r$-convergence is already determined by a countable family $\mathcal{C} \subset \Phi_r(\mathbb{R})$ (see [78, Theorem 4.5.(b)]). For this reason, it is sufficient to show that for all $f \in \mathcal{C}$, for du-almost every $u \in (0,1)$,

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) \to g(u) := \int_{(0,1)} f(F_\nu^{-1}(v))\, C_u(dv), \quad k \to +\infty, \tag{5.3.16}$$

where the integrals are du-almost everywhere well defined because of the inverse transform sampling, the fact that $f \in \Phi_r(\mathbb{R})$ and $\nu \in \mathcal{P}_r(\mathbb{R})$. For $u \in (0,1)$, let $x_u = F_\mu^{-1}(u)$ and $x_u^k = F_{\mu^k}^{-1}(u)$. Let $\mathcal{U} \subset (0,1)$ be the set of continuity points of F_μ^{-1} and define

$$\mathcal{U}_c = \{u \in \mathcal{U} \mid F_\mu \text{ is continuous at } x_u\} \quad \text{and} \quad \mathcal{U}_d = \{u \in \mathcal{U} \backslash \mathcal{U}_c \mid u \in (F_\mu(x_u-), F_\mu(x_u))\}.$$

By monotonicity of F_μ^{-1}, the complement of $\mathcal{U}$ in $(0,1)$ is at most countable, and since μ has countably many atoms, the complement of $\mathcal{U}_d$ in $\mathcal{U}\backslash\mathcal{U}_c$ is also at most countable. We deduce that it is sufficient to show (5.3.16) for du-almost all $u \in \mathcal{U}_c \cup \mathcal{U}_d$.

Let then $u \in \mathcal{U}$. If $\mu^k(\{x_u^k\}) = 0$, then $C_u^k = C_u$ and

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) = g(u).$$

From now on and until (5.3.16) is proved, we suppose w.l.o.g. that $\mu^k(\{x_u^k\}) > 0$ for all $k \in \mathbb{N}$. Then

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) = \frac{1}{\mu^k(\{x_u^k\})} \int_{F_{\mu^k}(x_u^k-)}^{F_{\mu^k}(x_u^k)} g(w)\, dw. \tag{5.3.17}$$

Define $l_k = \inf_{n\geq k} x_u^n$ and $r_k = \sup_{n\geq k} x_u^n$. Since $u \in \mathcal{U}$ we find $l_k \nearrow x_u$ and $r_k \searrow x_u$ when k goes to $+\infty$. Due to right continuity of F_μ and left continuity of $x \mapsto F_\mu(x-)$ we have

$$F_\mu(x_u-) = \lim_p F_\mu(l_p-) \quad \text{and} \quad \lim_p F_\mu(r_p) = F_\mu(x_u).$$

By Portmanteau's theorem and monotonicity of cumulative distribution functions we have

$$F_\mu(l_p-) \leq \liminf_k F_{\mu^k}(l_p-) \leq \liminf_k F_{\mu^k}(x_u^k-) \leq \limsup_k F_{\mu^k}(x_u^k) \leq \limsup_k F_{\mu^k}(r_p) \leq F_\mu(r_p).$$

By taking the limit $p \to +\infty$, we find

$$F_\mu(x_u-) \leq \liminf_k F_{\mu^k}(x_u^k-) \leq \limsup_k F_{\mu^k}(x_u^k) \leq F_\mu(x_u). \tag{5.3.18}$$

By (5.2.5), the interval $[F_{\mu^k}(x_u^k-), F_{\mu^k}(x_u^k)]$ contains u, and if $u \in \mathcal{U}_c$, then (5.3.18) implies that its length $\mu^k(\{x_u^k\})$ vanishes when k goes to $+\infty$. Consequently, (5.3.17) and the Lebesgue differentiation theorem yield that for du-almost every $u \in \mathcal{U}_c$,

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) \to g(u).$$

Suppose now $u \in \mathcal{U}_d$ and define

$$a_k = F_{\mu^k}(x_u^k-) \vee F_\mu(x_u-), \quad b_k = F_{\mu^k}(x_u^k) \wedge F_\mu(x_u).$$

Note that on the interval (a_k, b_k) the function g is constant equal to $g(u)$, so (5.3.17) writes

$$\int_{(0,1)} f(F_\nu^{-1}(v))\, C_u^k(dv) = \frac{1}{\mu^k(\{x_u^k\})}\left(\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\,dw + \int_{a_k}^{b_k} g(u)\,dw + \int_{F_{\mu^k}(x_u^k-)}^{a_k} g(w)\,dw\right).$$

According to (5.3.18),

$$a_k - F_{\mu^k}(x_u^k-) \to 0 \quad \text{and} \quad F_{\mu^k}(x_u^k) - b_k \to 0, \quad k \to +\infty. \tag{5.3.19}$$

Moreover, it is clear with (5.2.5) in mind that

$$\begin{aligned} F_{\mu^k}(x_u^k-) < a_k &\implies \mu^k(\{x_u^k\}) \geq u - F_\mu(x_u-), \\ \text{and} \quad b_k < F_{\mu^k}(x_u^k) &\implies \mu^k(\{x_u^k\}) \geq F_\mu(x_u) - u. \end{aligned} \tag{5.3.20}$$

Using the latter fact and the equality

$$b_k - a_k = \mu_k(\{x_u^k\}) - (F_{\mu_k}(x_u^k) - b_k) - (a_k - F_{\mu_k}(x_u^k-)),$$

we get

$$1 - \frac{F_{\mu_k}(x_u^k) - b_k}{F_\mu(x_u) - u} - \frac{a_k - F_{\mu_k}(x_u^k-)}{u - F_\mu(x_u-)} \leq \frac{b_k - a_k}{\mu_k(\{x_u^k\})} \leq 1.$$

Hence by (5.3.19) we have $\frac{b_k - a_k}{\mu_k(\{x_u^k\})} \to 1$ as k goes to $+\infty$, which implies that $\frac{1}{\mu^k(\{x_u^k\})} \int_{a_k}^{b_k} g(u)\, dw \to g(u)$ as $k \to +\infty$. Therefore, we just have to show that

$$\frac{1}{\mu^k(\{x_u^k\})} \left(\int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\, dw + \int_{F_{\mu^k}(x_u^k -)}^{a_k} g(w)\, dw \right) \to 0, \quad k \to +\infty. \tag{5.3.21}$$

Note that we can assume w.l.o.g. that for all $k \in \mathbb{N}$ either $F_{\mu^k}(x_u^k -) < a_k$ or $b_k < F_{\mu^k}(x_u^k)$. Let $d = (u - F_\mu(x_u -)) \wedge (F_\mu(x_u) - u)$, which is positive since $u \in \mathcal{U}_d$. Then we have by (5.3.20)

$$\begin{aligned} \frac{1}{\mu^k(\{x_u^k\})} &\left| \int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\, dw + \int_{F_{\mu^k}(x_u^k -)}^{a_k} g(w)\, dw \right| \\ &\leq \frac{1}{d} \left| \int_{b_k}^{F_{\mu^k}(x_u^k)} g(w)\, dw + \int_{F_{\mu^k}(x_u^k -)}^{a_k} g(w)\, dw \right|. \end{aligned} \tag{5.3.22}$$

By the inverse transform sampling and the facts that $f \in \Phi_r(\mathbb{R})$ and $\nu \in \mathcal{P}_r(\mathbb{R})$, we have $\int_0^1 |g(w)|\, dw = \int_{\mathbb{R}} |f(y)|\, \nu(dy) < +\infty$. Then (5.3.21) is a direct consequence of (5.3.22), (5.3.19) and the dominated convergence theorem. Hence (5.3.14) is proved for du-almost every $u \in (0,1)$.

Next, we show uniform integrability of (5.3.15). We can estimate

$$\mathcal{W}_r^r((F_\nu^{-1})_* C_u, (F_\nu^{-1})_* C_u^k) \leq 2^{r-1} \left(\int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u(dv) + \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv) \right).$$

Since by inverse transform sampling we have

$$\int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u(dv)\, du = \int_{\mathbb{R}} |y|^r\, \nu(dy) < \infty,$$

it is enough to show uniform integrability of $u \mapsto \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)$, $k \in \mathbb{N}$.

On the one hand, using the inverse transform sampling and $\nu \in \mathcal{P}_r(\mathbb{R})$, we have

$$\sup_k \int_{(0,1)} \int_{(0,1)} |F_\nu^{-1}(v)|^r C_u^k(dv)\, du = \int_{\mathbb{R}} |y|^r\, \nu(dy) < +\infty.$$

On the other hand, let $\varepsilon > 0$ and A be a measurable subset of $(0,1)$ such that $\lambda(A) < \varepsilon$. We have

$$\int_A \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)\, du = \int_{\mathbb{R}} |y|^r\, (F_\nu^{-1})_* \tau^k(dy),$$

where $\tau^k(dv) = \int_{u=0}^1 \mathbb{1}_A(du)\, C_u^k(dv)\, du$. Note that $\tau^k \leq \lambda$, $(F_\nu^{-1})_* \tau^k \leq \nu$ and $(F_\nu^{-1})_* \tau^k(\mathbb{R}) = \tau^k((0,1)) = \lambda(A)$. Therefore,

$$\sup_{\substack{A \in \mathcal{B}((0,1)), \\ \lambda(A) \leq \varepsilon}} \sup_k \int_A \int_{(0,1)} |F_\nu^{-1}(v)|^r\, C_u^k(dv)\, du \leq I_\varepsilon^r(\nu),$$

where $I_\varepsilon^r(\nu)$ is defined by (5.3.1). By Lemma 5.3.1, the right-hand side converges to 0 with $\varepsilon \to 0$. This yields uniform integrability of (5.3.15), which completes the proof. □

As mentioned in Section 5.2, Proposition 5.2.1 generalises to Polish spaces. Unsurprisingly, the proof of Proposition 5.2.3 requires radically different tools from its unidimensional equivalent. In particular, we need to recall the so called Weak Optimal Transport (WOT) problem introduced by Gozlan, Roberto, Samson and Tetali [90] and studied in [89], formulated according to our needs. Let $C : X \times \mathcal{P}_r(Y) \to \mathbb{R}_+$ be nonnegative, continuous, strictly convex in the second argument and such that there exists a constant $K > 0$ which satisfies

$$\forall (x,p) \in X \times \mathcal{P}_r(Y), \quad C(x,p) \leq K \left(1 + d_X^r(x,x_0) + \int_Y d_Y^r(y,y_0)\, p(dy)\right). \tag{5.3.23}$$

Then the WOT problem consists in minimising

$$V_C(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x,\pi_x)\, \mu(dx). \tag{WOT}$$

In view of the definition (5.1.3) of the adapted Wasserstein distance which involves measures on the extended space $X \times \mathcal{P}(Y)$, it is natural to consider an extension of (WOT) which also involves this space. Hence we also consider the extended problem

$$V'_C(\mu,\nu) := \inf_{P \in \Lambda(\mu,\nu)} \int_{X \times \mathcal{P}(Y)} C(x,p)\, P(dx,dp), \tag{WOT'}$$

where $\Lambda(\mu,\nu)$ is the set of couplings between μ and any measure on $\mathcal{P}(Y)$ with mean ν, that is

$$\Lambda(\mu,\nu) = \left\{ P \in \mathcal{P}(X \times \mathcal{P}(Y)) \mid \int_{(x',p) \in X \times \mathcal{P}(Y)} \delta_{x'}(dx)\, p(dy)\, P(dx',dp) \in \Pi(\mu,\nu) \right\}. \tag{5.3.24}$$

Remark 5.3.3. We gather here useful results on weak transport problems which hold under the assumption we made on C:

(a) (WOT) admits a unique minimiser π^* [18, Theorem 1.2];

(b) As a consequence of the necessary optimality criterion [20, Theorem 2.2] and [20, Remark 2.3.(a)], $J(\pi^*)$ is the only minimiser of (WOT')

(c) $V(\mu,\nu) = V'(\mu,\nu)$ [18, Lemma 2.1];

(d) Stability of (WOT) and (WOT'): Let $\mu^k \in \mathcal{P}_r(X), \nu^k \in \mathcal{P}_r(Y)$, $k \in \mathbb{N}$ converge respectively to $\mu \in \mathcal{P}_r(X)$ and $\nu \in \mathcal{P}_r(Y)$ in $\mathcal{W}_r$. For $k \in \mathbb{N}$, let $\pi^k \in \Pi(\mu^k,\nu^k)$ be optimal for $V(\mu^k,\nu^k)$. Then π^k, resp. $J(\pi^k)$, converges to the unique minimiser π^*, resp. $J(\pi^*)$, in $\mathcal{W}_r$ [20, Theorem 1.3 and Corollary 2.8]. In particular, this shows that π^k converges to π^* even in $\mathcal{AW}_r$.

Proof of Proposition 5.2.3. Let $\varepsilon > 0$ and $y_0 \in Y$. Define for $R > 0$ the $\mathcal{W}_r$-open ball B_R of radius $R^{1/r}$ and centre δ_{y_0} and the set

$$A_R = \{x \in X \mid \pi_x \in B_R\} = \left\{ x \in X \;\middle|\; \int_Y d_Y^r(y,y_0)\, \pi_x(dy) < R \right\}.$$

Since $\nu \in \mathcal{P}_r(\mathbb{R})$, $\int_X \int_Y d_Y^r(y, y_0)\, \pi_x(dy)\, \mu(dx) = \int_Y d_Y^r(y, y_0)\, \nu(dy) < +\infty$, hence μ is concentrated on $\bigcup_{R>0} A_R$ and we can choose R large enough such that

$$\mu(X \backslash A_R) < \varepsilon.$$

By Lusin's theorem, there exists a closed set $F \subset A_R$ such that

$$\mu(X \backslash F) < \varepsilon \quad \text{and} \quad x \mapsto \pi_x \text{ restricted to } F \text{ is continuous.}$$

Let $\widetilde{\mathcal{M}}_r(Y)$ be the linear space of all finite signed measures on Y, the positive and negative parts of which are contained in $\mathcal{M}_r(Y)$, equipped with the weak topology induced by $\Phi_r(Y)$. Since weak topologies are locally convex, an extension of Tietze's theorem [73, Theorem 4.1] yields the existence of a continuous map $x \mapsto \overline{\pi}_x$ defined on X with values in $\widetilde{\mathcal{M}}_r(Y)$ such that $\overline{\pi}_x = \pi_x$ for all $x \in F$ and

$$\{\overline{\pi}_x \mid x \in X\} \subset \mathrm{co}\{\pi_x \mid x \in F\} \subset B_R,$$

where co denotes the convex hull.

Next, we define a nonnegative, continuous, strictly convex in the second argument function which satisfies a condition of the form (5.3.23) in order to use the results on weak transport problems detailed in Remark 5.3.3. Let $\{g_k \mid k \in \mathbb{N}\} \subset \Phi_1(Y)$ be a family of 1-Lipschitz continuous functions and absolutely bounded by 1, which separates $\mathcal{P}(Y)$ (see [78, Theorem 4.5.(a)]). We have for any pair $p, p' \in \mathcal{P}(Y)$, $p \neq p'$ that there is $l \in \mathbb{N}$ such that

$$\int_Y g_l(y)\, p(dy) \neq \int_Y g_l(y)\, p'(dy). \tag{5.3.25}$$

Define $C : X \times \mathcal{P}_r(Y) \to \mathbb{R}_+$ for all $(x, p) \in X \times \mathcal{P}_r(Y)$ by

$$C(x, p) := \rho(\overline{\pi}_x, p) + \sum_{k \in \mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) \right|^2,$$

where $\rho : \mathcal{P}(Y) \times \mathcal{P}(Y) \to [0, 1]$ is defined for all $p, p' \in \mathcal{P}(Y)$ by

$$\rho(p, p') = \inf_{\chi \in \Pi(p, p')} \int_{Y \times Y} (d_Y(y, y') \wedge 1)\, \chi(dy, dy').$$

Since ρ can be interpreted as a Wasserstein distance with respect to a bounded distance, it is immediate that it is a metric on $\mathcal{P}(Y)$ which induces the weak convergence topology. On the one hand, the map $(x, p) \mapsto \rho(\overline{\pi}_x, p)$ is continuous by continuity of $x \mapsto \overline{\pi}_x$. On the other hand, by Kantorovich and Rubinstein's duality theorem and Jensen's inequality, we have for all $(x, p), (x', p') \in X \times \mathcal{P}_r(Y)$

$$\sum_{k \in \mathbb{N}} \frac{1}{2^k} \left| \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) \right|^2 - \left| \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) - \int_Y g_k(y)\, p'(dy) \right|^2 \right|$$

$$
\begin{aligned}
&= \sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) + \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) - \int_Y g_k(y)\, p'(dy) \right| \\
&\qquad \times \left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, p(dy) - \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) + \int_Y g_k(y)\, p'(dy) \right| \\
&\le \sum_{k\in\mathbb{N}} \frac{4}{2^k} \left(\left| \int_Y g_k(y)\, \overline{\pi}_x(dy) - \int_Y g_k(y)\, \overline{\pi}_{x'}(dy) \right| + \left| \int_Y g_k(y)\, p(dy) - \int_Y g_k(y)\, p'(dy) \right| \right) \\
&\le 8 \left(\mathcal{W}_1(\overline{\pi}_x, \overline{\pi}_{x'}) + \mathcal{W}_1(p, p') \right) \le 8 \left(\mathcal{W}_r(\overline{\pi}_x, \overline{\pi}_{x'}) + \mathcal{W}_r(p, p') \right),
\end{aligned}
$$

where the right-hand side vanishes when (x', p') converges to (x, p) by continuity of $x \mapsto \overline{\pi}_x$. We deduce that C is continuous.

Note that ρ is convex in the second argument. Therefore, to obtain strict convexity of $C(x, \cdot)$ in the second argument, it is sufficient to verify that

$$
F(p) = \sum_{k\in\mathbb{N}} \frac{1}{2^k} \left| \int_Y g_k(y)\, p(dy) \right|^2
$$

is strictly convex. Let $p, p' \in \mathcal{P}(Y)$, $p \neq p'$ and $l \in \mathbb{N}$ such that (5.3.25) holds. Hence, strict convexity of the square proves

$$
\left| \alpha \int_Y g_l(y)\, p(dy) + (1-\alpha) \int_Y g_l(y)\, p'(dy) \right|^2 < \alpha \left| \int_Y g_l(y)\, p(dy) \right|^2 + (1-\alpha) \left| \int_Y g_l(y)\, p'(dy) \right|^2,
$$

which yields strict convexity of F on $\mathcal{P}(Y)$.

Moreover, we have for all $(x, p) \in X \times \mathcal{P}_r(Y)$, $C(x, p) \le 1 + 8 = 9$, hence C satisfies (5.3.23). Remember the definitions of V and V' given in (WOT) and (WOT'). Since for all $x \in F$, $C(x, \pi_x) = C(x, \overline{\pi}_x) = 0$, we have

$$
V(\mu, \nu) \le \int_{X \setminus F} C(x, \pi_x)\, \mu(dx) < 9\varepsilon.
$$

Let $\pi^{*,\varepsilon} \in \Pi(\mu, \nu)$ be optimal for $V(\mu, \nu)$. For $P, P' \in \mathcal{P}(X \times \mathcal{P}(Y))$, let

$$
\tilde{\rho}(P, P') = \inf_{\chi \in \Pi(P, P')} \int_{X \times \mathcal{P}(Y) \times X \times \mathcal{P}(Y)} \left((d_X(x, x') + \rho(p, p')) \wedge 1 \right) \, \chi(dx, dp, dx', dp').
$$

Since $\mu(dx)\, \delta_{\pi_x}(dp)\, \delta_x(dx')\, \delta_{\pi^{*,\varepsilon}_{x'}}(dp')$ is a coupling between $J(\pi)$ and $J(\pi^{*,\varepsilon})$, we can estimate

$$
\begin{aligned}
\tilde{\rho}(J(\pi), J(\pi^{*,\varepsilon})) &\le \int_X \rho(\pi_x, \pi^{*,\varepsilon}_x)\, \mu(dx) \\
&\le \int_F \rho(\pi_x, \pi^{*,\varepsilon}_x)\, \mu(dx) + \int_{X \times F} \int_Y (d_Y(y, y_0) \wedge 1)\, (\pi_x + \pi^{*,\varepsilon}_x)(dy)\, \mu(dx) \\
&\le V(\mu, \nu) + 2\varepsilon < 11\varepsilon.
\end{aligned}
$$

For $k \in \mathbb{N}$, let $\pi^{k,\varepsilon} \in \Pi(\mu^k, \nu^k)$ be optimal for $V(\mu^k, \nu^k)$. Then $J(\pi^{k,\varepsilon})$ is optimal for $V'(\mu^k, \nu^k)$ by Remark 5.3.3 (b), and converges to $J(\pi^{*,\varepsilon})$ in $\mathcal{W}_r$ and therefore weakly by Remark 5.3.3 (d). Then we get

$$
\limsup_{k\to+\infty} \tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi)) \le \limsup_{k\to+\infty} \left(\tilde{\rho}(J(\pi^{k,\varepsilon}), J(\pi^{*,\varepsilon})) + \tilde{\rho}(J(\pi^{*,\varepsilon}), J(\pi)) \right) \le 11\varepsilon. \tag{5.3.26}
$$

So far $\varepsilon > 0$ was arbitrary. Therefore, there exists a strictly increasing sequence $(k_N)_{N\in\mathbb{N}^*}$ of positive integers such that

$$\forall N \in \mathbb{N}^*, \quad \forall k \geq k_N, \quad \tilde{\rho}(J(\pi^{k,2^{-rN}}), J(\pi)) \leq 12 \cdot 2^{-rN}.$$

For $k \in \mathbb{N}$, let $N_k = \max\{N \in \mathbb{N} \mid k \geq k_N\}$, where the maximum of the empty set is defined as 0. Since k_N is strictly increasing, we find that $N_k \to +\infty$ as $k \to +\infty$. Then the sequence of couplings

$$\pi^k = \pi^{k,2^{-rN_k}} \in \Pi(\mu^k, \nu^k), \ k \in \mathbb{N}$$

is such that $\tilde{\rho}(J(\pi^k), J(\pi))$ vanishes as k goes to $+\infty$, and therefore $J(\pi^k)$ converges weakly to $J(\pi)$. Moreover, since $\mathcal{W}_r$-convergence is equivalent to weak convergence coupled with convergence of the r-moments, we have that the r-moments of μ^k and ν^k respectively converge to the r-moments of μ and ν, which implies

$$\begin{aligned}\int_{X\times\mathcal{P}(Y)} \mathcal{W}_r^r(p, \delta_{y_0})\, J(\pi^k)(dx, dp) = \int_X \mathcal{W}_r^r(\pi_x^k, \delta_{y_0})\, \mu^k(dx) = \int_Y d_Y^r(y, y_0)\, \nu^k(dy) \\ \underset{k\to+\infty}{\longrightarrow} \int_Y d_Y^r(y, y_0)\, \nu(dy) = \int_{X\times\mathcal{P}(Y)} \mathcal{W}_r^r(p, \delta_{y_0})\, J(\pi)(dx, dp).\end{aligned}$$

We deduce that $J(\pi^k)$ converges to $J(\pi)$ in $\mathcal{W}_r$ as $k \to +\infty$. According to (5.1.3), $\pi^{k,\varepsilon}$ converges to $\pi^{*,\varepsilon}$ in $\mathcal{AW}_r$, which concludes the proof. □

In the proof of Theorem 5.2.5 we need to be able to confine approximative sequences of couplings to certain sets. The next result provides all necessary tools for this.

Lemma 5.3.4. *Let $\mu, \mu^k \in \mathcal{M}_r(X)$, $\nu, \nu^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ all with equal mass and $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converge to $\pi \in \Pi(\mu, \nu)$ in $\mathcal{AW}_r$.*

(i) Let $A \subset X$ be measurable and $B \supset A$ be open. There are $\tilde{\mu}^k \leq \mu^k|_B$ and $\varepsilon_k \geq 0$, $k \in \mathbb{N}$ such that $\tilde{\pi}^k := \tilde{\mu}^k \times \pi_x^k$ satisfies

$$\mathcal{AW}_r(\tilde{\pi}^k, (1-\varepsilon_k)\pi|_{A\times Y}) + \varepsilon_k \underset{k\to+\infty}{\longrightarrow} 0.$$

(ii) Let $C \subset X$ be an open set on which ν is concentrated. There are $\hat{\mu}^k \leq \tilde{\mu}^k$, $\hat{\nu}^k \leq \nu^k$, $\hat{\pi}^k = \hat{\mu}^k \times \hat{\pi}_x^k \in \Pi(\hat{\mu}^k, \hat{\nu}^k)$ concentrated on $B \times C$ and $\varepsilon'_k \geq 0$, $k \in \mathbb{N}$ such that

$$\mathcal{AW}_r^r(\hat{\pi}^k, (1-\varepsilon'_k)\pi|_{A\times Y}) + \int_X \mathcal{W}_r^r(\hat{\pi}_x^k, \pi_x^k)\, \hat{\mu}^k(dx) + \varepsilon'_k \underset{k\to+\infty}{\longrightarrow} 0.$$

Proof. Both assertions are trivial if $\mu(A) = 0$. So assume that $\mu(A) > 0$.

(i) Let $\chi^k \in \Pi(\mu^k, \mu)$ be optimal for $\mathcal{AW}_r(\pi^k, \pi)$ and $\chi = (\mathrm{id}, \mathrm{id})_*\mu$.
Since $\chi^k(dx_1, dx_2)\, \delta_{(x_2,x_2)}(dx_3, dx_4)$ defines a coupling in $\Pi(\chi^k, \chi)$, we find

$$\mathcal{W}_r^r(\chi^k, \chi) \leq \int_{X^4} (d_X(x_1, x_3)^r + d_X(x_2, x_4)^r)\, \chi^k(dx_1, dx_2)\, \delta_{(x_2,x_2)}(dx_3, dx_4)$$

$$= \int_{X\times X} d_X(x_1,x_2)^r\,\chi^k(dx_1,dx_2) \le \mathcal{AW}_r^r(\pi^k,\pi) \to 0, \quad k\to+\infty.$$

Further, let $P : \mathcal{P}_r(X\times X) \to \mathcal{P}(X\times X)$ be the homeomorphism given by

$$P(\eta)(dx_1,dx_2) = \frac{(1+d_X(x_1,x_0)^r + d_X(x_2,x_0)^r)\,\eta(dx_1,dx_2)}{\int_{X\times X}(1+d_X(x_1',x_0)^r + d_X(x_2',x_0)^r)\,\eta(dx_1',dx_2')},$$

for $\eta \in \mathcal{P}_r(X\times X)$. Recall (5.1.1), then it is easily deductible that $P(\eta') \to P(\eta)$ weakly if and only if $\eta' \to \eta$ in $\mathcal{W}_r$. In particular, we find $P(\chi^k) \to P(\chi)$ weakly as k goes to $+\infty$. Let $f \in \Phi_r(X\times X)$ and

$$\varphi\colon X\times X\colon (x_1,x_2) \mapsto \frac{\mathbb{1}_{X\times A}(x_1,x_2) f(x_1,x_2)}{1+d_X(x_1,x_0)^r + d_X(x_2,x_0)^r}.$$

Then φ is a bounded measurable map which is continuous w.r.t. the first component. As a consequence of [127, Lemma 2.1], we find

$$\int_{X\times X}\varphi(x_1,x_2)\,P(\chi^k)(dx_1,dx_2) \to \int_{X\times X}\varphi(x_1,x_2)\,P(\chi)(dx_1,dx_2), \quad k\to+\infty,$$

which amounts to

$$\int_{X\times X} f(x_1,x_2)\,\chi^k|_{X\times A}(dx_1,dx_2) \to \int_{X\times X} f(x_1,x_2)\,\chi|_{X\times A}(dx_1,dx_2), \quad k\to+\infty.$$

Therefore (5.1.1) yields $\mathcal{W}_r$-convergence of $\chi^k|_{X\times A}$ to $\chi|_{X\times A}$. By Portmanteau's theorem, we have

$$\chi^k(B\times A) \le \mu(A) = \chi|_{X\times A}(B\times B) \le \liminf_{k\to+\infty}\chi^k|_{X\times A}(B\times B) = \liminf_{k\to+\infty}\chi^k(B\times A),$$

hence $\varepsilon_k := 1 - \frac{\chi^k(B\times A)}{\mu(A)}$, $k\in\mathbb{N}$ is a null sequence of nonnegative real numbers. Denote by $\tilde\mu^k$, resp. $\bar\mu^k$ the first resp. second, marginal of $\chi^k|_{B\times A}$, $k\in\mathbb{N}$. Note that

$$1-\varepsilon_k = \frac{\tilde\mu^k(X)}{\mu(A)}. \tag{5.3.27}$$

We want to show that

$$\mathcal{AW}_r(\tilde\mu^k\times\pi_x^k, (1-\varepsilon_k)\mu|_A\times\pi_x) \to 0. \tag{5.3.28}$$

On the one hand, note that

$$\begin{aligned}\mathcal{AW}_r^r(\tilde\mu^k\times\pi_x^k, \bar\mu^k\times\pi_x) &\le \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\chi^k|_{B\times A}(dx,dx') \\ &\le \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\chi^k(dx,dx') \\ &= \mathcal{AW}_r^r(\pi^k,\pi) \to 0, \quad k\to+\infty.\end{aligned} \tag{5.3.29}$$

On the other hand, let

$$\check{\mu}^k = (1-\varepsilon_k)\mu|_A, \quad \zeta^k = \check{\mu}^k \wedge \bar{\mu}^k \quad \text{and} \quad \alpha_k = \bar{\mu}^k(X) - \zeta^k(X) = \check{\mu}^k(X) - \zeta^k(X).$$

Let $\overline{\chi}^k \in \Pi(\bar{\mu}^k - \zeta^k, \check{\mu}^k - \zeta^k)$ be optimal for $\mathcal{AW}_r^r((\bar{\mu}^k - \zeta^k)\times\pi_x, (\check{\mu}^k - \zeta^k)\times\pi_x)$. Since $((\mathrm{id},\mathrm{id})_*\zeta^k + \overline{\chi}^k)$ is a coupling between $\bar{\mu}^k$ and $\check{\mu}^k$, we find

$$\begin{aligned}
\mathcal{AW}_r(\bar{\mu}^k\times\pi_x, \check{\mu}^k\times\pi_x) &\le \int_X (d_X^r(x,x') + \mathcal{W}_r^r(\pi_x,\pi_{x'}))\,\overline{\chi}^k(dx,dx') \\
&= \mathcal{AW}_r^r((\bar{\mu}^k - \zeta^k)\times\pi_x, (\check{\mu}^k - \zeta^k)\times\pi_x) \\
&\le \mathcal{AW}_r((\bar{\mu}^k - \zeta^k)\times\pi_x, \alpha_k\delta_{(x_0,y_0)}) + \mathcal{AW}_r((\check{\mu}^k - \zeta^k)\times\pi_x, \alpha_k\delta_{(x_0,y_0)}).
\end{aligned}$$

In the next estimates we use (5.3.1). Note that the first marginal of $(\bar{\mu}^k - \zeta^k)\times\pi_x$ is dominated by μ whereas its second marginal is dominated by ν. Thus, denoting $\tau^k(dy) = \int_X \pi_x(dy)\,(\bar{\mu}^k - \zeta^k)(dx)$, we find

$$\begin{aligned}
\mathcal{AW}_r^r((\bar{\mu}^k - \zeta^k)\times\pi_x, \alpha_k\delta_{(x_0,y_0)}) &= \int_X (d_X^r(x,x_0) + \mathcal{W}_r^r(\pi_x,\delta_{y_0}))\,(\bar{\mu}^k - \zeta^k)(dx) \\
&= \int_X d_X^r(x,x_0)\,(\bar{\mu}^k - \zeta^k)(dx) + \int_Y d_Y^r(y,y_0)\,\tau^k(dy) \\
&\le I_{\alpha_k}^r(\mu) + I_{\alpha_k}^r(\nu).
\end{aligned}$$

Similarly, we find

$$\mathcal{AW}_r^r((\check{\mu}^k - \zeta^k)\times\pi_x, \alpha_k\delta_{(x_0,y_0)}) \le I_{\alpha_k}^r(\mu) + I_{\alpha_k}^r(\nu).$$

If we can show that α_k vanishes for $k\to+\infty$, then we find by Lemma 5.3.1 (b) that

$$\mathcal{AW}_r(\bar{\mu}^k\times\pi_x, \check{\mu}^k\times\pi_x) \underset{k\to+\infty}{\longrightarrow} 0, \tag{5.3.30}$$

and the triangle inequality with (5.3.29) and (5.3.30) yield the assertion, (5.3.28).

Since $\check{\mu}^k, \bar{\mu}^k \le \mu|_A$, the respective densities of $\check{\mu}^k$ and $\bar{\mu}^k$ with respect to $\mu|_A$ satisfy $\frac{d\check{\mu}^k}{d\mu|_A}, \frac{d\bar{\mu}^k}{d\mu|_A} \le 1$. Then

$$\begin{aligned}
\alpha_k = \bar{\mu}^k(X) - \zeta^k(X) &= \int_A \left(\frac{d\bar{\mu}^k}{d\mu|_A}(x) - \frac{d\check{\mu}^k}{d\mu|_A}(x)\right)^+ \mu(dx) \le \int_A \left(1 - \frac{d\check{\mu}^k}{d\mu|_A}(x)\right)\mu(dx) \\
&= \mu(A) - \check{\mu}^k(A) = \varepsilon_k\mu(A) \underset{k\to+\infty}{\longrightarrow} 0,
\end{aligned}$$

which is conclusive.

(ii) To show the last assertion, we denote by $\tilde{\nu}^k$ and $\tilde{\nu}$ the second marginal of $\tilde{\mu}^k\times\pi_x^k$ and $\mu|_A\times\pi_x$ respectively. Since we have to remove mass outside of C from $\tilde{\nu}^k$ and the corresponding kernels, we consider the $\mathcal{W}_r$-optimal couplings

$$\mathring{\chi}^k \in \Pi(\tilde{\nu}^k, (1-\varepsilon_k)\tilde{\nu}) \quad \text{and} \quad \hat{\chi}^k \in \Pi(\tilde{\nu}^k(C)\tilde{\nu}, \tilde{\nu}(C)\tilde{\nu}^k|_C).$$

From now on, we denote by $\hat{\mu}^k = \tilde{\mu}^k \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)} \le \tilde{\mu}^k$. Using the couplings $\mathring{\chi}^k$ and $\hat{\chi}^k$ we define the kernels

$$\mathring{\pi}^k_x(dz) = \int_Y \mathring{\chi}^k_y(dz)\, \pi^k_x(dy) \quad \text{and} \quad \hat{\pi}^k_x(dt) = \int_Y \hat{\chi}^k_z(dt)\, \mathring{\pi}^k_x(dz).$$

Let then $\hat{\pi}^k := \hat{\mu}^k \times \hat{\pi}^k_x$, whose second marginal denoted $\hat{\nu}^k$ satisfies

$$\begin{aligned}\hat{\nu}^k(dt) &= \int_{x\in X}\int_{y\in Y}\int_{z\in Z} \hat{\mu}^k(dx)\, \pi^k_x(dy)\, \mathring{\chi}^k_y(dz)\, \hat{\chi}^k_z(dt)\\ &= \frac{\tilde{\nu}(C)(1-\varepsilon_k)}{\tilde{\mu}^k(X)} \tilde{\nu}^k|_C(dt) = \frac{\tilde{\nu}(C)}{\mu(A)} \tilde{\nu}^k|_C(dt).\end{aligned}$$

So $\hat{\nu}^k$ is concentrated on C. Moreover, we have $\tilde{\nu}^k|_C \le \nu^k$, and since ν is concentrated on C, $\tilde{\nu}(C) = \tilde{\nu}(Y) = \mu|_A(X) = \mu(A)$, hence $\hat{\nu}^k = \tilde{\nu}^k|_C \le \nu^k$. For $k \in \mathbb{N}$, let $\varepsilon'_k \in \mathbb{R}$ be such that

$$1 - \varepsilon'_k = \frac{\tilde{\nu}^k(C)}{\mu(A)}\cdot \tag{5.3.31}$$

For $k \in \mathbb{N}$, using (5.3.27), we have $\tilde{\nu}^k(C) \le \tilde{\mu}^k(X) = (1-\varepsilon_k)\mu(A) \le \mu(A)$, hence ε'_k is nonnegative. Then it remains to show that

$$\mathcal{AW}_r(\hat{\pi}^k, (1-\varepsilon'_k)\pi|_{A\times Y}) + \int_X \mathcal{W}^r_r(\hat{\pi}^k_x, \pi^k_x)\, \hat{\mu}^k(dx) + \varepsilon'_k \underset{k\to+\infty}{\longrightarrow} 0. \tag{5.3.32}$$

Since we have

$$\begin{gathered}\pi^k_x(dy)\, \mathring{\chi}^k_y(dz) \in \Pi(\pi^k_x, \mathring{\pi}^k_x), \quad \mathring{\pi}^k_x(dz)\, \hat{\chi}^k_z(dt) \in \Pi(\mathring{\pi}^k_x, \hat{\pi}^k_x),\\ \int_{x\in X} \hat{\mu}^k(dx)\, \pi^k_x(dy)\, \mathring{\chi}^k_y(dz) = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)} \mathring{\chi}^k(dy,dz),\\ \int_{x\in X} \hat{\mu}^k(dx)\, \mathring{\pi}^k_x(dz)\, \hat{\chi}^k_z(dt) = \frac{(1-\varepsilon_k)}{\tilde{\mu}^k(X)} \hat{\chi}^k(dz,dt),\end{gathered}$$

we find plugging the expressions (5.3.27) and (5.3.31) that

$$\begin{aligned}&\mathcal{AW}^r_r(\hat{\mu}^k \times \pi^k_x, \hat{\mu}^k \times \hat{\pi}^k_x)\\ &\le \int_X \mathcal{W}^r_r(\pi^k_x, \hat{\pi}^k_x)\, \hat{\mu}^k(dx)\\ &\le 2^{r-1} \int_X \left(\mathcal{W}^r_r(\pi^k_x, \mathring{\pi}^k_x) + \mathcal{W}^r_r(\mathring{\pi}^k_x, \hat{\pi}^k_x)\right) \hat{\mu}^k(dx)\\ &\le 2^{r-1} \int_X \left(\int_{Y\times Y} d^r_Y(y,z)\, \pi^k_x(dy)\, \mathring{\chi}^k_y(dz) + \int_{Y\times Y} d^r_Y(z,t)\, \mathring{\pi}^k_x(dz)\, \hat{\chi}^k_z(dt)\right) \hat{\mu}^k(dx)\\ &= 2^{r-1} \left(\frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)} \int_{Y\times Y} d^r_Y(y,z)\, \mathring{\chi}^k(dy,dz) + \frac{(1-\varepsilon_k)}{\tilde{\mu}^k(X)} \int_{Y\times Y} d^r_Y(z,t)\, \chi^k(dz,dt)\right)\end{aligned}$$

$$= 2^{r-1}\left(\frac{1-\varepsilon'_k}{1-\varepsilon_k}\mathcal{W}_r^r(\tilde{\nu}^k,(1-\varepsilon_k)\tilde{\nu}) + \frac{1}{\mu(A)}\mathcal{W}_r^r(\tilde{\nu}^k(C)\tilde{\nu},\tilde{\nu}(C)\tilde{\nu}^k|_C)\right).$$

To see convergence to 0, note that since $\mathcal{AW}_r$ dominates $\mathcal{W}_r$, we find by continuity of the projection on the second marginal that (5.3.28) implies

$$\mathcal{W}_r(\tilde{\nu}^k,(1-\varepsilon_k)\tilde{\nu}) \to 0, \quad k \to +\infty.$$

Using Portmanteau's theorem and the fact that $(1-\varepsilon_k) \to 1$ as k goes to $+\infty$, we have for all nonnegative function $f \in \Phi_r(Y)$

$$\limsup_{k\to+\infty} \tilde{\nu}^k(\mathbb{1}_C f) \le \limsup_{k\to+\infty} \tilde{\nu}^k(f) = \tilde{\nu}(f) = \tilde{\nu}(\mathbb{1}_C f) \le \liminf_{k\to+\infty} \tilde{\nu}^k(\mathbb{1}_C f),$$

hence

$$\tilde{\nu}^k|_C(f) \to \tilde{\nu}(f), \quad k \to +\infty. \tag{5.3.33}$$

Moreover, (5.3.33) applied with $f = 1$ yields $\tilde{\nu}^k(C) \to \tilde{\nu}(C) = \mu(A)$ as k goes to $+\infty$, hence ε'_k vanishes as k goes to $+\infty$ and

$$\mathcal{W}_r(\tilde{\nu}^k(C)\tilde{\nu},\tilde{\nu}(C)\tilde{\nu}^k|_C) \to 0, \quad k \to +\infty.$$

We deduce that

$$\mathcal{AW}_r^r(\hat{\mu}^k \times \pi_x^k, \hat{\mu}^k \times \hat{\pi}_x^k) \le \int_X \mathcal{W}_r^r(\pi_x^k,\hat{\pi}_x^k)\,\hat{\mu}^k(dx) \to 0, \quad k \to +\infty.$$

On the other hand, by (5.3.27) and (5.3.31) we have $\hat{\mu}^k = \frac{\tilde{\nu}^k(C)}{\tilde{\mu}^k(X)}\tilde{\mu}^k = \frac{1-\varepsilon'_k}{1-\varepsilon_k}\tilde{\mu}^k$, hence

$$\mathcal{AW}_r(\hat{\mu}^k \times \pi_x^k, (1-\varepsilon'_k)\mu|_A \times \pi_x) = \frac{1-\varepsilon'_k}{1-\varepsilon_k}\mathcal{AW}_r(\tilde{\mu}^k \times \pi_x^k, (1-\varepsilon_k)\mu|_A \times \pi_x),$$

where the right-hand side vanishes as k goes to $+\infty$ by the first part. Then (5.3.32) follows by triangle inequality and the latter convergences, which completes the proof. □

The addition of measures is continuous with respect to the weak and Wasserstein topology. More precisely, we have the estimate

$$\mathcal{W}_r^r(\mu+\mu',\nu+\nu') \le \mathcal{W}_r^r(\mu,\nu) + \mathcal{W}_r^r(\mu',\nu')$$

for all measures $\mu,\mu',\nu,\nu' \in \mathcal{P}_r(X)$ such that μ and ν, resp. μ' and ν' have equal mass.

When considering the adapted weak topology, the next example disproves a comparable statement.

Example 5.3.5. Let $X = Y = \mathbb{R}$, and $\pi^k = \delta_{\left(\frac{1}{k},1\right)}$, $\chi^k = \delta_{\left(-\frac{1}{k},-1\right)}$, $k \in \mathbb{N}$. Then both sequences are convergent in $\mathcal{AW}_1$, but

$$\mathcal{AW}_1(\pi^k+\chi^k, \delta_{(0,1)}+\delta_{(0,-1)}) = \frac{2}{k}+2$$

does not vanish.

However, we show in the next Lemma that the addition of measures with respect to the weak adapted topology could still be somehow considered continuous if one of the limits has mass significantly smaller than the other.

Lemma 5.3.6. *Let $\hat{\mu}, \hat{\mu}^k, \hat{\nu}, \hat{\nu}^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ be with equal mass and $\tilde{\mu}, \tilde{\mu}^k, \tilde{\nu}, \tilde{\nu}^k \in \mathcal{M}_r(Y)$, $k \in \mathbb{N}$ be with equal mass smaller than ε. Let $\hat{\pi}^k \in \Pi(\hat{\mu}^k, \hat{\nu}^k), \tilde{\pi}^k \in \Pi(\tilde{\mu}^k, \tilde{\nu}^k)$, $k \in \mathbb{N}$, $\hat{\pi} \in \Pi(\hat{\mu}, \hat{\nu})$ and $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$. Let $\mu = \hat{\mu} + \tilde{\mu}$ and $\nu = \hat{\nu} + \tilde{\nu}$. Then*

(a) We have for all $k \in \mathbb{N}$

$$\begin{aligned}
&\mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\
&\quad \leq \mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) + 2^{r-1}\left(I_\varepsilon^r(\tilde{\mu}) + I_\varepsilon^r(\tilde{\mu}^k) + I_\varepsilon^r(\tilde{\nu}) + I_\varepsilon^r(\tilde{\nu}^k) + 2I_\varepsilon^r(\hat{\nu}) + 2I_\varepsilon^r(\hat{\nu}^k)\right) \\
&\quad \leq \mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) + (2^{r-1})^2\left(\mathcal{W}_r^r(\tilde{\mu}^k, \tilde{\mu}) + \mathcal{W}_r^r(\tilde{\nu}^k, \tilde{\nu}) + 2\mathcal{W}_r^r(\hat{\nu}^k, \hat{\nu})\right) \\
&\qquad + 2^{r-1}(1 + 2^{r-1})I_\varepsilon^r(\mu) + 3 \cdot 2^{r-1}(1 + 2^{r-1})I_\varepsilon^r(\nu),
\end{aligned} \tag{5.3.34}$$

where $I_\varepsilon^r(\cdot)$ is defined by (5.3.1).

(b) If $(\hat{\pi}^k)_{k\in\mathbb{N}}$ converges to $\hat{\pi}$ in $\mathcal{AW}_r$ and $(\mu^k = \hat{\mu}^k + \tilde{\mu}^k)_{k\in\mathbb{N}}$, resp. $(\nu^k = \hat{\nu}^k + \tilde{\nu}^k)_{k\in\mathbb{N}}$, converges to μ, resp. ν, in $\mathcal{W}_r$, then

$$\limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \leq C(I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)), \tag{5.3.35}$$

where $C > 0$ depends only on r.

Proof. The second inequality of (5.3.34) is easily deduced from the first one, (5.3.2) and the fact that $I_\varepsilon^r(\tilde{\mu}) \leq I_\varepsilon^r(\mu)$, $I_\varepsilon^r(\tilde{\nu}) \leq I_\varepsilon^r(\nu)$ and $I_\varepsilon^r(\hat{\nu}) \leq I_\varepsilon^r(\nu)$.

To see (b), assume for a moment that the first inequality of (5.3.34) holds true and suppose

$$\hat{\pi}^k \to \hat{\pi} \text{ in } \mathcal{AW}_r, \quad \mu^k = \hat{\mu}^k + \tilde{\mu}^k \to \mu \quad \text{and} \quad \nu^k = \hat{\nu}^k + \tilde{\nu}^k \to \nu \text{ in } \mathcal{W}_r$$

as $k \to +\infty$. Using Lemma 5.3.1 (a) and then (5.3.2), we get

$$\begin{aligned}
\limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) &\leq C' \limsup_{k\to+\infty}\left(I_\varepsilon^r(\mu^k) + I_\varepsilon^r(\nu^k) + I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)\right) \\
&\leq C \limsup_{k\to+\infty}\left(\mathcal{W}_r^r(\mu^k, \mu) + \mathcal{W}_r^r(\nu^k, \nu) + I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)\right) \\
&= C(I_\varepsilon^r(\mu) + I_\varepsilon^r(\nu)),
\end{aligned}$$

where $C, C' > 0$ depend only on r. Hence (b) is proved.

To conclude the proof, it remains to show first inequality of (5.3.34). Let $\hat{\rho}^k \in \Pi(\hat{\mu}^k, \hat{\mu})$ be optimal for $\mathcal{AW}_r(\hat{\pi}^k, \hat{\pi})$ and $\tilde{\rho}^k \in \Pi(\tilde{\mu}^k, \tilde{\mu})$ be arbitrary. We write $\rho^k = \hat{\rho}^k + \tilde{\rho}^k$. Then

$$\mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \leq \int_{X\times X}\left(d_X^r(x, x') + \mathcal{W}_r^r((\hat{\pi}^k + \tilde{\pi}^k)_x, (\hat{\pi} + \tilde{\pi})_{x'})\right)\rho^k(dx, dx'). \tag{5.3.36}$$

Let $\hat p = \frac{d\hat\mu}{d\mu}$ and $\hat p^k = \frac{d\hat\mu^k}{d\mu^k}$. Notice that $\hat p$ and $\hat p^k$ take values in $[0,1]$. The identities

$$(\hat\pi + \tilde\pi)(dx, dx') = \mu(dx)\left(\hat p(x)\,\hat\pi_x(dx') + (1-\hat p(x))\,\tilde\pi_x(dx')\right),$$
$$(\hat\pi^k + \tilde\pi^k)(dx, dx') = \mu^k(dx)\left(\hat p^k(x)\,\hat\pi^k_x(dx') + (1-\hat p^k(x))\,\tilde\pi^k_x(dx')\right),$$

provide representations for the disintegrations of $(\hat\pi+\tilde\pi)$ and $(\hat\pi^k+\tilde\pi^k)$ respectively for $\mu(dx)$- and $\mu^k(dx)$-almost every x:

$$(\hat\pi+\tilde\pi)_x = \hat p(x)\,\hat\pi_x + (1-\hat p(x))\tilde\pi_x, \quad (\hat\pi^k+\tilde\pi^k)_x = \hat p^k(x)\,\hat\pi^k_x + (1-\hat p^k(x))\,\tilde\pi^k_x.$$

Thus, we have when letting $\alpha^k_+(x,x') = (\hat p^k(x) - \hat p(x'))^+$, $\alpha^k_-(x,x') = (\hat p^k(x) - \hat p(x'))^-$ and $\beta^k(x,x') = \hat p^k(x) \wedge \hat p(x')$ that

$$\begin{aligned}
&\mathcal{W}_r^r((\hat\pi^k+\tilde\pi^k)_x, (\hat\pi+\tilde\pi)_{x'}) \\
&\quad \le \mathcal{W}_r^r(\beta^k(x,x')\,\hat\pi^k_x, \beta^k(x,x')\,\hat\pi_{x'}) \\
&\qquad + \mathcal{W}_r^r\Big(\alpha^k_+(x,x')\,\hat\pi^k_x + (1-\hat p^k(x))\,\tilde\pi^k_x, \alpha^k_-(x,x')\,\hat\pi_{x'} + (1-\hat p(x'))\,\tilde\pi_{x'}\Big) \\
&\quad \le \beta^k(x,x')\mathcal{W}_r^r(\hat\pi^k_x, \hat\pi_{x'}) + 2^{r-1}\Big(\alpha^k_+(x,x')\mathcal{W}_r^r(\hat\pi^k_x, \delta_{y_0}) + (1-\hat p^k(x))\mathcal{W}_r^r(\tilde\pi^k_x, \delta_{y_0}) \\
&\qquad + \alpha^k_-(x,x')\mathcal{W}_r^r(\hat\pi_{x'}, \delta_{y_0}) + (1-\hat p(x'))\mathcal{W}_r^r(\tilde\pi_{x'}, \delta_{y_0})\Big).
\end{aligned} \tag{5.3.37}$$

Since $\beta^k(x,x') = \hat p^k(x) \wedge \hat p(x') \le 1$, we deduce from (5.3.36), (5.3.37) and $\mathcal{AW}_r$-optimality of $\hat\rho^k$

$$\begin{aligned}
\mathcal{AW}_r^r(\hat\pi^k+\tilde\pi^k, \hat\pi+\tilde\pi) \le \mathcal{AW}_r^r(\hat\pi^k, \hat\pi) &+ \int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p^k(x)\mathcal{W}_r^r(\hat\pi^k_x, \delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \hat p(x')\mathcal{W}_r^r(\hat\pi_{x'}, \delta_{y_0})\,\tilde\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha^k_+(x,x')\mathcal{W}_r^r(\hat\pi^k_x, \delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p^k(x))\mathcal{W}_r^r(\tilde\pi^k_x, \delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} \alpha^k_-(x,x')\mathcal{W}_r^r(\hat\pi_{x'}, \delta_{y_0})\,\rho^k(dx,dx') \\
&+ 2^{r-1}\int_{X\times X} (1-\hat p(x'))\mathcal{W}_r^r(\tilde\pi_{x'}, \delta_{y_0})\,\rho^k(dx,dx').
\end{aligned} \tag{5.3.38}$$

Recall that $\tilde\rho^k$ has marginals $\tilde\mu^k$ and $\tilde\mu$ with total mass smaller than ε. By (5.3.1) we find

$$\int_{X\times X} d_X(x,x')^r\,\tilde\rho^k(dx,dx') \le 2^{r-1}\left(I^r_\varepsilon(\tilde\mu^k) + I^r_\varepsilon(\tilde\mu)\right). \tag{5.3.39}$$

Concerning the marginals of $\hat{p}^k(x)\,\rho(dx,dx')$ and $\hat{p}(x')\,\rho(dx,dx')$, we find the relation

$$\hat{p}^k(x)\,\tilde{\mu}^k(dx) = (1-\hat{p}^k(x))\,\hat{\mu}^k(dx), \quad \hat{p}(x')\,\tilde{\mu}(dx') = (1-\hat{p}(x'))\,\hat{\mu}(dx').$$

Again by (5.3.1), we find since $\tilde{\rho}^k \in \Pi(\tilde{\mu}^k,\tilde{\mu})$, $\hat{\pi}^k \in \Pi(\hat{\mu}^k,\hat{\nu}^k)$ and $\hat{\pi} \in \Pi(\hat{\mu},\hat{\nu})$ that

$$\int_{X\times X} \hat{p}^k(x)\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\tilde{\rho}^k(dx,dx') = \int_{X\times X}(1-\hat{p}^k(x))\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\hat{\mu}^k(dx) \le I_\varepsilon^r(\hat{\nu}^k), \quad (5.3.40)$$

$$\int_{X\times X} \hat{p}(x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\tilde{\rho}^k(dx,dx') = \int_{X\times X}(1-\hat{p}(x'))\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\hat{\mu}(dx') \le I_\varepsilon^r(\hat{\nu}). \quad (5.3.41)$$

We deduce from (5.3.38) and (5.3.39)-(5.3.41) that it is sufficient to show

$$\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\hat{\nu}^k), \quad (5.3.42)$$

$$\int_{X\times X}(1-\hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}^k), \quad (5.3.43)$$

$$\int_{X\times X} \alpha_-^k(x,x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\hat{\nu}), \quad (5.3.44)$$

$$\int_{X\times X}(1-\hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') \le I_\varepsilon^r(\tilde{\nu}). \quad (5.3.45)$$

To see (5.3.43) and (5.3.45), note that

$$(1-\hat{p}^k(x))\,\mu^k(dx) = \tilde{\mu}^k(dx) \quad \text{and} \quad (1-\hat{p}(x'))\,\mu(dx') = \tilde{\mu}(dx'). \quad (5.3.46)$$

As a consequence, the first marginal of $(1-\hat{p}^k(x))\,\rho^k(dx,dx')$ is $\tilde{\mu}^k$, whereas the second marginal of $(1-\hat{p}(x'))\,\rho^k(dx,dx')$ coincides with $\tilde{\mu}$. Hence, as the mass of $\tilde{\mu}^k$ and $\tilde{\mu}$ does not exceed ε, we have

$$\int_{X\times X}(1-\hat{p}^k(x))\mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\tilde{\pi}_x^k,\delta_{y_0})\,\tilde{\mu}^k(dx) = \mathcal{W}_r^r(\tilde{\nu}^k,\delta_{y_0}) = I_\varepsilon^r(\tilde{\nu}^k),$$

$$\int_{X\times X}(1-\hat{p}(x'))\mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\tilde{\pi}_{x'},\delta_{y_0})\,\tilde{\mu}(dx') = \mathcal{W}_r^r(\tilde{\nu},\delta_{y_0}) = I_\varepsilon^r(\tilde{\nu}).$$

Next, we show (5.3.42) and (5.3.44). To this end, denoting with a slight abuse of notation $\rho^k(dx,dx') = \mu^k(dx)\,\rho_x^k(dx') = \mu(dx')\,\rho_{x'}^k(dx)$, we have

$$\alpha_+^k(x,x')\,\rho^k(dx,dx') \le \hat{p}^k(x)\,\rho^k(dx,dx') = \frac{d\hat{\mu}^k}{d\mu^k}(x)\,\mu^k(dx)\,\rho_x^k(dx') = \hat{\mu}^k(dx)\,\rho_x^k(dx'),$$

$$\alpha_-^k(x,x')\,\rho^k(dx,dx') \le \hat{p}(x')\,\rho^k(dx,dx') = \frac{d\hat{\mu}}{d\mu}(x')\,\mu(dx')\,\rho_{x'}^k(dx) = \hat{\mu}(dx')\,\rho_{x'}^k(dx).$$

In particular, the first marginal of $\alpha_+^k(x,x')\,\rho^k(dx,dx')$, denoted here by τ, is dominated by $\hat{\mu}^k$, whereas the second marginal of $\alpha_-^k(x,x')\,\rho^k(dx,dx')$, denoted here by τ', is dominated by $\hat{\mu}$. Concerning the mass of τ and τ', remember (5.3.46), $\alpha_+^k(x,x') \le 1-\hat{p}(x')$ and $\alpha_-^k(x,x') \le 1-\hat{p}^k(x)$, thus,

$$\tau(X) = \int_{X\times X} \alpha_+^k(x,x')\,\rho^k(dx,dx') \le \int_X (1-\hat{p}(x'))\,\mu(dx') = \tilde{\mu}(X) \le \varepsilon,$$

$$\tau'(X) = \int_{X\times X} \alpha_-^k(x,x')\,\rho^k(dx,dx') \le \int_X (1-\hat{p}^k(x))\,\mu^k(dx) = \tilde{\mu}^k(X) \le \varepsilon.$$

Using (5.3.1), we conclude with

$$\int_{X\times X} \alpha_+^k(x,x')\mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\hat{\pi}_x^k,\delta_{y_0})\,\tau(dx) \le I_\varepsilon^r(\hat{\nu}^k),$$
$$\int_{X\times X} \alpha_-^k(x,x')\mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\rho^k(dx,dx') = \int_X \mathcal{W}_r^r(\hat{\pi}_{x'},\delta_{y_0})\,\tau'(dx') \le I_\varepsilon^r(\hat{\nu}).$$

□

The addition on $\mathcal{M}_r(X\times Y)$ is continuous with respect to the adapted weak topology as soon as the limits have singular first marginal distributions. We recall that two positive measures μ,ν are called singular iff there exists a measurable set $A\subset X$ such that $\mu(A^\complement) = 0 = \nu(A)$.

Lemma 5.3.7. *Let $\pi,\chi \in \mathcal{M}_r(X\times Y)$ be such that their respective first marginals are singular. Let $\pi^k,\chi^k \in \mathcal{M}_r(X\times Y)$, $k\in\mathbb{N}$ converge to π and χ respectively in $\mathcal{AW}_r$. Then*

$$\pi^k+\chi^k \underset{k\to+\infty}{\longrightarrow} \pi+\chi \quad \textit{in } \mathcal{AW}_r.$$

Proof. Let μ_1, μ_2, μ_1^k and μ_2^k denote the respective first marginals of π, χ, π^k and χ^k. Due to singularity, there is a measurable set $A\subset X$ such that $\mu_1(A^\complement) = 0 = \mu_2(A)$.

Suppose first that for all $k\in\mathbb{N}$, $\mu_1^k(A^\complement) = 0 = \mu_2^k(A)$. Let $\rho_1^k \in \Pi(\mu_1^k,\mu_1)$, resp. $\rho_2^k\in\Pi(\mu_2^k,\mu_2)$, be an optimal coupling for $\mathcal{AW}_r(\pi^k,\pi)$, resp. $\mathcal{AW}_r(\chi^k,\chi)$. Since almost surely

$$(\pi^k+\chi^k)_x = \mathbb{1}_A(x)\,\pi_x^k + \mathbb{1}_{A^\complement}(x)\chi_x^k \quad\text{and}\quad (\pi+\chi)_x = \mathbb{1}_A(x)\,\pi_x + \mathbb{1}_{A^\complement}(x)\chi_x,$$

we have

$$\begin{aligned}
\mathcal{AW}_r^r(\pi^k+\chi^k,\pi+\chi) &\le \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r((\pi^k+\chi^k)_x,(\pi+\chi)_{x'})\right)(\rho_1^k+\rho_2^k)(dx,dx')\\
&= \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\pi_x^k,\pi_{x'})\right)\rho_1^k(dx,dx')\\
&\quad + \int_{X\times X}\left(d_X^r(x,x') + \mathcal{W}_r^r(\chi_x^k,\chi_{x'})\right)\rho_2^k(dx,dx')\\
&= \mathcal{AW}_r^r(\pi^k,\pi) + \mathcal{AW}_r^r(\chi^k,\chi) \to 0,\quad k\to+\infty.
\end{aligned}$$

Let us now go back to the general case. Since X is a Polish space, μ_1 and μ_2 are inner regular, so there exist two compact sets $K_1\subset A$ and $K_2\subset A^\complement$ such that

$$\mu_1(K_1^\complement) < \varepsilon \quad\text{and}\quad \mu_2(K_2^\complement) < \varepsilon.$$

Since X is Polish, it is normal, hence we can seperate the closed, disjoint sets K_1 and K_2 by open, disjoint sets $\tilde{K}_1$ and $\tilde{K}_2$ where $K_1\subset\tilde{K}_1$ and $K_2\subset\tilde{K}_2$. Then Lemma 5.3.4

(i) provides sequences $(\tilde{\mu}_1^k \times \pi_x^k)_{k\in\mathbb{N}}$ and $(\tilde{\mu}_2^k \times \chi_x^k)_{k\in\mathbb{N}}$ with values in $\mathcal{M}(X \times Y)$ and null sequences $(\varepsilon_k)_{k\in\mathbb{N}}$ and $(\eta_k)_{k\in\mathbb{N}}$ with values in $[0,1]$, such that $\tilde{\mu}_1^k \le \mu_1^k|_{\tilde{K}_1}$, $\tilde{\mu}_2^k \le \mu_2^k|_{\tilde{K}_2}$ and

$$\mathcal{AW}_r^r(\tilde{\mu}_1^k \times \pi_x^k, (1-\varepsilon_k)\pi|_{K_1\times Y}) + \mathcal{AW}_r^r(\tilde{\mu}_2^k \times \chi_x^k, (1-\eta_k)\chi|_{K_2\times Y}) \to 0, \quad k \to +\infty.$$

To apply Lemma 5.3.6 (b), let $0 < \varepsilon' \le \varepsilon$ be such that $\varepsilon'(\mu_1(K_1) + \mu_2(K_2)) < \varepsilon$. Let k be sufficiently large such that $\varepsilon^k \wedge \eta^k < \varepsilon'$. We consider the sequences

$$\hat{\pi}^k = \frac{1-\varepsilon'}{1-\varepsilon^k}\tilde{\mu}_1^k \times \pi_x^k + \frac{1-\varepsilon'}{1-\eta^k}\tilde{\mu}_2^k \times \chi_x^k, \quad \hat{\pi} = (1-\varepsilon')\left(\pi|_{K_1\times Y} + \chi|_{K_2\times Y}\right),$$
$$\tilde{\pi}^k = \pi^k + \chi^k - \hat{\pi}^k, \quad \tilde{\pi} = \pi + \chi - \hat{\pi},$$

where $\tilde{\pi}^k$ is well-defined in $\mathcal{M}_r(X \times Y)$ since $\varepsilon^k < \varepsilon'$ and $\eta^k < \varepsilon'$. Note that as $k \to +\infty$,

$$\mathcal{AW}_r^r\left(\frac{1-\varepsilon'}{1-\varepsilon^k}\tilde{\mu}_1^k \times \pi_x^k, (1-\varepsilon')\pi|_{K_1\times Y}\right) = \frac{1-\varepsilon'}{1-\varepsilon^k}\mathcal{AW}_r^r\left(\tilde{\mu}_1^k \times \pi_x^k, (1-\varepsilon^k)\pi|_{K_1\times Y}\right) \to 0,$$
$$\mathcal{AW}_r^r\left(\frac{1-\varepsilon'}{1-\eta^k}\tilde{\mu}_2^k \times \chi_x^k, (1-\varepsilon')\chi|_{K_2\times Y}\right) = \frac{1-\varepsilon'}{1-\eta^k}\mathcal{AW}_r^r\left(\tilde{\mu}_2^k \times \chi_x^k, (1-\eta^k)\chi|_{K_2\times Y}\right) \to 0.$$

Since the first marginal distributions of $\tilde{\mu}_1^k \times \pi_x^k$ and $(1-\varepsilon_k)\pi|_{K_1\times Y}$, resp. $\tilde{\mu}_2^k \times \chi_x^k$ and $(1-\eta_k)\chi|_{K_2\times Y}$, are concentrated on $\tilde{K}_1$, resp. $\tilde{K}_2$, and since $\tilde{K}_1$ and $\tilde{K}_2$ are disjoint, we have according to the preceding part that

$$\mathcal{AW}_r^r(\hat{\pi}^k, \hat{\pi}) \to 0, \quad k \to +\infty.$$

Due to $\mathcal{AW}_r$-convergence of $(\pi^k)_{k\in\mathbb{N}}$ and $(\chi^k)_{k\in\mathbb{N}}$, we obtain $\mathcal{W}_r$-convergence of the marginals of $\pi^k + \chi^k$ to the marginals of $\pi + \chi$. Furthermore, we have

$$\tilde{\pi}^k(X \times Y) = \tilde{\pi}(X \times Y) \le \mu_1(K_1^c) + \mu_2(K_2^c) + \varepsilon'(\mu_1(K_1) + \mu_2(K_2)) < 3\varepsilon.$$

Then (5.3.35) yields

$$\begin{aligned}\limsup_{k\to+\infty} \mathcal{AW}_r^r(\pi^k + \chi^k, \pi + \chi) &= \limsup_{k\to+\infty} \mathcal{AW}_r^r(\hat{\pi}^k + \tilde{\pi}^k, \hat{\pi} + \tilde{\pi}) \\ &\le C\left(I_{3\varepsilon}^r(\mu_1 + \mu_2) + I_{3\varepsilon}^r(\nu_1 + \nu_2)\right),\end{aligned}$$

where ν_1 and ν_2 denote the respective second marginals of π and χ, and the constant C only depends on r. Therefore, the right-hand side vanishes with ε according to Lemma 5.3.1 (b), which concludes the proof. □

5.4 Auxiliary results on the convex order in dimension one

We recall that the convex order on $\mathcal{M}_1(\mathbb{R})$ is defined by

$$\mu \le_{cx} \nu \iff \quad \forall f\colon \mathbb{R} \to \mathbb{R} \text{ convex}, \quad \mu(f) \le \nu(f).$$

The following assertions can be found in [25, Section 4]: for all $(m_0, m_1) \in \mathbb{R}_+^* \times \mathbb{R}$, there is a one-to-one correspondence between finite positive measures $\mu \in \mathcal{M}_1(\mathbb{R})$ with mass m_0 such that $\int_{\mathbb{R}} y\, \mu(dy) = m_1$ and the set of functions $u\colon \mathbb{R} \to \mathbb{R}^+$ which satisfy

(i) u is convex;

(ii) $u(y) - m_0|y - m_1|$ goes to 0 as $|y|$ tends to $+\infty$.

Any function which suffices (i) and (ii) is then called a potential function. The potential function of μ is denoted by

$$u_\mu(y) = \int_{\mathbb{R}} |y - x|\, \mu(dx).$$

A sequence $(\mu^k)_{k\in\mathbb{N}}$ of finite positive measures with equal mass on the line converges in $\mathcal{W}_1$ to μ iff the sequence of potential functions $(u_{\mu^k})_{k\in\mathbb{N}}$ converges pointwise to u_μ. In that case, since for all $y \in \mathbb{R}$ the map $x \mapsto |y - x|$ is Lipschitz continuous with constant 1, we have by Kantorovich and Rubinstein's duality theorem that

$$\sup_{y\in\mathbb{R}} |u_{\mu^k}(y) - u_\mu(y)| \leq \mathcal{W}_1(\mu^k, \mu) \to 0, \quad k \to +\infty,$$

hence we even have uniform convergence on $\mathbb{R}$ of potential functions.

In dimension one, for all $m_1 \in \mathbb{R}$, the set of all finite positive measures with mean m_1 is a lattice [120, Proposition 1.6], and even a complete lattice [121]. Then all $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ with mean m_1 have a supremum, denoted $\mu \vee_{cx} \nu$, and an infimum, denoted $\mu \wedge_{cx} \nu$, with respect to the convex order. In that context it is convenient to work with potential functions since they provide simple characterisations of those bounds:

$$\mu \vee_c \nu \text{ is defined as the measure with potential function } u_\mu \vee u_\nu,$$
$$\mu \wedge_c \nu \text{ is defined as the measure with potential function } \operatorname{co}(u_\mu \wedge u_\nu),$$

where co is the convex hull.

Lemma 5.4.1. *Let $(\mu^k)_{k\in\mathbb{N}}$, $(\nu^k)_{k\in\mathbb{N}}$ be two sequences of $\mathcal{M}_1(\mathbb{R})$ converging respectively to μ and ν in $\mathcal{W}_1$. Suppose that there exists $(m_0, m_1) \in \mathbb{R}_+^* \times \mathbb{R}$ such that $\mu^k(\mathbb{R}) = \nu^k(\mathbb{R}) = m_0$ and $\int_{\mathbb{R}} x\, \mu^k(dx) = \int_{\mathbb{R}} y\, \nu^k(dy) = m_1$ for all $k \in \mathbb{N}$. Then*

$$\lim_{k\to+\infty} \mathcal{W}_1(\mu^k \vee_{cx} \nu^k, \mu \vee_{cx} \nu) = 0 \quad \textit{and} \quad \lim_{k\to+\infty} \mathcal{W}_1(\mu^k \wedge_{cx} \nu^k, \mu \wedge_{cx} \nu) = 0.$$

Proof. Convergence in $\mathcal{W}_1$ is equivalent to pointwise convergence of the potential functions. Thus, the convergence of $\mu^k \vee_{cx} \nu^k$ to $\mu \vee_{cx} \nu$ in $\mathcal{W}_1$ is a consequence of the pointwise convergence of $u_{\mu^k \vee_{cx} \nu^k} = u_{\mu^k} \vee u_{\nu^k}$ to $u_\mu \vee u_\nu = u_{\mu\vee_{cx}\nu}$.

To show convergence of $\mu^k \wedge_{cx} \nu^k$ to $\mu \wedge_{cx} \nu$ in $\mathcal{W}_1$, it is sufficient to show for all $x \in \mathbb{R}$

$$u_{\mu^k\wedge_{cx}\nu^k}(x) = \operatorname{co}(u_{\mu^k} \wedge u_{\nu^k})(x) \to \operatorname{co}(u_\mu \wedge u_\nu)(x) = u_{\mu\wedge_{cx}\nu}(x), \quad k \to +\infty. \tag{5.4.1}$$

Since u_{μ^k} and u_{ν^k} converge uniformly on $\mathbb{R}$ to u_μ and u_ν respectively, we have uniform convergence of $u_{\mu^k} \wedge u_{\nu^k}$ to $u_\mu \wedge u_\nu$. Let $\varepsilon > 0$ and $k_0 \in \mathbb{N}$ be such that for all $k \geq k_0$,

$$\sup_{x\in\mathbb{R}} |(u_{\mu^k} \wedge u_{\nu^k})(x) - (u_\mu \wedge u_\nu)(x)| \leq \varepsilon.$$

For all $k \geq k_0$, we find

$$\begin{aligned} \operatorname{co}(u_\mu \wedge u_\nu) - \varepsilon \leq (u_\mu \wedge u_\nu) - \varepsilon \leq u_{\mu^k} \wedge u_{\nu^k}, \\ \operatorname{co}(u_{\mu^k} \wedge u_{\nu^k}) - \varepsilon \leq (u_{\mu^k} \wedge u_{\nu^k}) - \varepsilon \leq u_\mu \wedge u_\nu. \end{aligned}$$

Thus, as the convex hull is the supremum over all dominated, convex functions, this yields

$$\operatorname{co}(u_\mu \wedge u_\nu) - \varepsilon \leq \operatorname{co}(u_{\mu^k} \wedge u_{\nu^k}) \leq \operatorname{co}(u_\mu \wedge u_\nu) + \varepsilon,$$

which proves (5.4.1) and completes the proof. □

We now prove the key argument to see that it is enough to prove Theorem 5.2.5 for irreducible pairs of marginals.

Proof of Proposition 5.2.4. To construct the desired decomposition, pick for all $k \in \mathbb{N}$ a coupling $\pi^k \in \Pi_M(\mu^k, \nu^k)$. Let l_n and r_n denote the left and right boundary of the open interval $\{u_{\mu_n} < u_{\nu_n}\}$ on which μ_n is concentrated, and set

$$\mu_n^k(dx) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F^{-1}_{\mu^k}(u)}(dx)\, du, \quad \nu_n^k(dy) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \pi^k_{F^{-1}_{\mu^k}(u)}(dy)\, du.$$

These are the respective marginals of π_n^k on $\mathbb{R}^2$ given by

$$\pi_n^k(dx, dy) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F^{-1}_{\mu^k}(u)}(dx)\, \pi^k_{F^{-1}_{\mu^k}(u)}(dy)\, du. \tag{5.4.2}$$

Since π^k is a martingale coupling, we have $\mu_n^k \leq_{cx} \nu_n^k$. Finally define

$$J = [0,1] \setminus \bigcup_{n\in N} (F_\mu(l_n), F_\mu(r_n-)),$$

and set

$$\eta^k(dx) = \int_{u\in J} \delta_{F^{-1}_{\mu^k}(u)}(dx)\, du, \quad \upsilon^k(dy) = \int_{u\in J} \pi^k_{F^{-1}_{\mu^k}(u)}(dy)\, du.$$

These are the respective marginals of $\tilde{\pi}^k$ defined by

$$\tilde{\pi}^k(dx, dy) = \int_{u\in J} \delta_{F^{-1}_{\mu^k}(u)}(dx)\, \pi^k_{F^{-1}_{\mu^k}(u)}(dy)\, du,$$

which is again a martingale coupling with marginals (η^k, υ^k), thus, $\eta^k \leq_{cx} \upsilon^k$.

Using inverse transform sampling for the second equality, we find

$$\left(\tilde{\pi}^k + \sum_{n\in N} \pi_n^k\right)(dx, dy) = \int_{u=0}^{1} \delta_{F^{-1}_{\mu^k}(u)}(dx)\, \pi^k_{F^{-1}_{\mu^k}(u)}(dy)\, du = \int_{x^k\in\mathbb{R}} \delta_{x^k}(dx)\, \pi^k_{x^k}(dy)\, \mu^k(dx^k)$$

$$= \mu^k(dx)\, \pi^k_x(dy) = \pi^k(dx, dy).$$

Concerning the marginals, we deduce

$$\eta^k + \sum_{n\in N} \mu^k_n = \mu^k \quad \text{and} \quad \upsilon^k + \sum_{n\in N} \nu^k_n = \nu^k.$$

For all $(\tau, u, l, r) \in \mathcal{P}_1(\mathbb{R}) \times (0,1) \times \mathbb{R} \times \mathbb{R}$, we have by (5.2.3):

$$F_\tau(l) < u < F_\tau(r-) \implies l < F^{-1}_\tau(u) < r \implies F_\tau(l) < u \le F_\tau(r-). \tag{5.4.3}$$

Since $\mu_n(dx) = \mathbb{1}_{(l_n,r_n)}(x)\, \mu(dx)$, using (5.4.3) for the second equality we find

$$\mu_n(dx) = \int_{x'\in(l_n,r_n)} \delta_{x'}(dx)\, \mu(dx) = \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F^{-1}_\mu(u)}(dx)\, du.$$

We deduce that

$$\begin{aligned}\eta(dx) = \left(\mu - \sum_{n\in N} \mu_n\right)(dx) &= \int_{u=0}^{1} \delta_{F^{-1}_\mu(u)}(dx)\, du - \sum_{n\in N} \int_{u=F_\mu(l_n)}^{F_\mu(r_n-)} \delta_{F^{-1}_\mu(u)}(dx)\, du \\ &= \int_{u\in J} \delta_{F^{-1}_\mu(u)}(dx)\, du.\end{aligned}$$

Since the monotone rearrangement gives an optimal coupling, we have

$$\mathcal{W}_1(\eta^k, \eta) + \sum_{n\in N} \mathcal{W}_1(\mu^k_n, \mu_n) = \int_0^1 |F^{-1}_{\mu^k}(u) - F^{-1}_\mu(u)|\, du = \mathcal{W}_1(\mu^k, \mu),$$

hence

$$\lim_{k\to+\infty} \mathcal{W}_1(\eta^k, \eta) = 0 = \lim_{k\to+\infty} \mathcal{W}_1(\mu^k_n, \mu_n), \quad \forall n \in N.$$

Since the marginals of π^k converge weakly, the sequences $(\mu^k)_{k\in\mathbb{N}}$ and $(\nu^k)_{k\in\mathbb{N}}$ are tight, and so is $(\pi^k)_{k\in\mathbb{N}}$. For any $n \in N$, π^k_n is dominated by π^k, hence $(\pi^k_n)_{k\in\mathbb{N}}$ is tight and therefore relatively compact. Moreover, by $\mathcal{W}_1$-convergence of $(\mu^k)_{k\in\mathbb{N}}$ and $(\nu^k)_{k\in\mathbb{N}}$, the sequences $\left(\int_\mathbb{R} |x|\, \mu^k(dx)\right)_{k\in\mathbb{N}}$ and $\left(\int_\mathbb{R} |y|\, \nu^k(dy)\right)_{k\in\mathbb{N}}$ are convergent and therefore bounded. Hence the sequences $\left(\int_\mathbb{R} |x|\, \mu^k_n(dx)\right)_{k\in\mathbb{N}}$ and $\left(\int_\mathbb{R} |y|\, \nu^k_n(dy)\right)_{k\in\mathbb{N}}$ are bounded as well and admit convergent subsequences. Since the $\mathcal{W}_1$-convergence is equivalent to the weak convergence and the convergence of the first moments, we deduce that the sequence $(\pi^k_n)_{k\in\mathbb{N}}$ is relatively compact in $\mathcal{W}_1$. There are subsequences $(\pi^{k_j}_n)_{j\in\mathbb{N}}$ converging in $\mathcal{W}_1$ to a measure $\tilde{\pi}_n$, where $\tilde{\pi}_n \le \pi$. The first marginal of $\tilde{\pi}_n$ coincides with μ_n due to continuity of the projection, thus,

$$\tilde{\pi}_n \le \pi|_{I_n\times\mathbb{R}} = \pi_n.$$

As $\tilde{\pi}_n(\mathbb{R} \times \mathbb{R}) = \mu_n(I_n) = \pi_n(\mathbb{R} \times \mathbb{R})$, there must hold equality, i.e., $\tilde{\pi}_n = \pi_n$. In particular, $(\pi^k_n)_{k\in\mathbb{N}}$ has only a single accumulation point and is therefore $\mathcal{W}_1$-convergent. Due to continuity of the projection, $(\nu^k_n)_{k\in\mathbb{N}}$ converges in $\mathcal{W}_1$ to ν_n for each $n \in N$. Analogously, we find that $(\upsilon^k)_{k\in\mathbb{N}}$ converges to η. $\square$

The next two lemmas explore certain scaling and restrictions of measure on condition that the transformed measures are in convex order.

Lemma 5.4.2. *Let $r \geq 1$ and $\mu \in \mathcal{M}_r(\mathbb{R}^d)$ be a finite positive measure. Let $m_1 = \int_{\mathbb{R}} x\,\mu(dx)$ and μ^α, $\alpha \in \mathbb{R}_+$ be the image of μ by $y \mapsto \alpha(y - m_1) + m_1$. Then for all $\alpha, \beta \in \mathbb{R}_+$,*

$$\mathcal{W}_r(\mu^\alpha, \mu^\beta) = |\beta - \alpha| \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu(dx) \right)^{\frac{1}{r}}. \tag{5.4.4}$$

Moreover, $(\mu^\alpha)_{\alpha \in \mathbb{R}^+}$ constitutes a peacock, i.e., $\alpha \leq \beta \in \mathbb{R}_+$ implies $\mu^\alpha \leq_{cx} \mu^\beta$.

Remark 5.4.3. Let $\eta_0, \eta_1 \in \mathcal{M}_r(\mathbb{R}^d)$ be with equal mass and γ be an optimal coupling for $\mathcal{W}_r(\eta_0, \eta_1)$. For all $t \in [0,1]$, let η_t be the image of γ by $(y, z) \mapsto (1-t)y + tz$. It is well known that the curve $[0,1] \ni t \mapsto \eta_t$ is a constant speed geodesic for W_r connecting η_0 to η_1. Moreover, for all $0 \leq s \leq t \leq 1$, the image of γ by $((1-s)y + sz, (1-t)y + tz)$ is an optimal coupling for $\mathcal{W}_r(\eta_s, \eta_t)$.

In particular for $\alpha < \beta \in \mathbb{R}_+$, $\eta_0 = \delta_{m_1}$ and $\eta_1 = \mu^\beta$, we have $\gamma(dy, dz) = \delta_x(dy)\,\mu^\beta(dz)$ and for $t = \frac{\alpha}{\beta}$, $\eta_t = \mu^\alpha$. Therefore, the image of γ by $(y, z) \mapsto ((1-t)y + tz, z)$, that is the image of μ by $y \mapsto (\alpha(y - x) + x, \beta(y - x) + x)$, is an optimal coupling for $\mathcal{W}_r(\mu^\alpha, \mu^\beta)$, hence (5.4.4) is proved.

Since the proof is brief, we add it here for the sake of completeness.

Proof of Lemma 5.4.2. Let $\alpha \leq \beta \in \mathbb{R}_+$. We have by the triangle inequality for $\mathcal{W}_r$ that

$$\begin{aligned} \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu^\beta(dx) \right)^{\frac{1}{r}} &= \mathcal{W}_r(\delta_{m_1}, \mu^\beta) \leq \mathcal{W}_r(\delta_{m_1}, \mu^\alpha) + \mathcal{W}_r(\mu^\alpha, \mu^\beta) \\ &= \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu^\alpha(dx) \right)^{\frac{1}{r}} + \mathcal{W}_r(\mu^\alpha, \mu^\beta). \end{aligned}$$

Thus,

$$\begin{aligned} \mathcal{W}_r(\mu^\alpha, \mu^\beta) &\geq \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu^\beta(dx) \right)^{\frac{1}{r}} - \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu^\alpha(dx) \right)^{\frac{1}{r}} \\ &= (\beta - \alpha) \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu(dx) \right)^{\frac{1}{r}}. \end{aligned}$$

Since the image of μ by $x \mapsto (\alpha(x - m_1) + m_1, \beta(y - m_1) + m_1)$ is a coupling between μ^α and μ^β, we also have the reverse inequality

$$\mathcal{W}_r(\mu^\alpha, \mu^\beta) \leq (\beta - \alpha) \left(\int_{\mathbb{R}^d} |x - m_1|^r \,\mu(dx) \right)^{\frac{1}{r}},$$

which proves (5.4.4).

To see that $(\mu^\alpha)_{\alpha \in \mathbb{R}^+}$ is a peacock, we fix again $\alpha \leq \beta \in \mathbb{R}_+$ and any convex function f on $\mathbb{R}^d$. Due to convexity, we have

$$\mu^\alpha(f) = \int_{\mathbb{R}^d} f(\alpha(x - m_1) + m_1)\,\mu(dx)$$

$$\leq \int_{\mathbb{R}^d} \left(\frac{\alpha}{\beta} f(\beta(x-m_1)+m_1) + \left(1-\frac{\alpha}{\beta}\right) f(m_1) \right) \mu(dx) \leq \mu^\beta(f).$$

$\square$

Lemma 5.4.4. *For all $p \in \mathcal{P}_1(\mathbb{R})$ with barycentre $m_1 \in \mathbb{R}$ and $R \geq 0$, let p^R be defined by*

$$p^R = p \wedge_{cx} \left(\frac{R-m_1}{2R}\, \delta_{-R} + \frac{R+m_1}{2R}\, \delta_R \right) \quad \text{if} \quad R \geq |m_1|,$$

and $p^R = \delta_{m_1}$ otherwise. Then

(a) For all $R > 0$, $p_R \leq_{cx} p$, and if $R \geq |m_1|$, then p^R is concentrated on $[-R, R]$.

(b) We have

$$\mathcal{W}_1(p^R, p) \underset{R\to+\infty}{\longrightarrow} 0.$$

Proof. Let $p \in \mathcal{P}_1(\mathbb{R})$ be with barycentre $m_1 \in \mathbb{R}$. For all $R \geq |m_1|$, let $\eta^R = \frac{R-m_1}{2R}\, \delta_{-R} + \frac{R+m_1}{2R}\, \delta_R$, so that $p^R = p \wedge_{cx} \eta^R$. If $R < |m_1|$ then $p^R = \delta_{m_1}$ so we clearly have $p^R \leq_{cx} p$. Else, $p^R \leq_{cx} p$ still holds by definition of the convex infimum. Moreover, since η^R is concentrated on $[-R, R]$, so is p^R by domination in the convex order, hence (a) is proved.

To show (b), we show that for all $y \in \mathbb{R}$,

$$u_{p\wedge\eta^R}(y) = \mathrm{co}(u_p \wedge u_{\eta^R})(y) \to u_p(y), \quad R \to +\infty. \tag{5.4.5}$$

Let $\varepsilon > 0$. Since $u_p(y) - |y - m_1|$ vanishes as $|y| \to +\infty$, there exists $M > 0$ such that

$$\forall y \in \mathbb{R}, \quad |y| > M \implies u_p(y) \leq |y - m_1| + \varepsilon.$$

Let $R_0 = |m_1| + \sup_{y\in[-M,M]} u_p(y)$ and $R \geq R_0$. The map u_{η^R} is a piecewise affine function which changes slope at $-R$ and R and such that $u_{\eta^R}(y) \to +\infty$ as $|y| \to +\infty$. It therefore attains its minimum either at $-R$ where it is equal to $R + m_1$ or at R where it is equal to $R - m_1$, and this minimum is equal to $R - |m_1|$. We deduce that for all $y \in \mathbb{R}$, $u_{\eta^R}(y) \geq R - |m_1|$. Moreover, $\delta_{m_1} \leq_{cx} u_{\eta^R}$, hence we also have $u_p(y) \geq |y - m_1|$ for all $y \in \mathbb{R}$. Let $y \in \mathbb{R}$. If $|y| \leq M$, then

$$u_p(y) \leq \sup_{[-M,M]} u_p = R_0 - |m_1| \leq R - m_1 \leq u_{\eta^R}(y).$$

Else if $|y| > M$ and $u_{\eta^R}(y) < u_p(y)$, then

$$u_p(y) \leq |y - m_1| + \varepsilon \leq u_{\eta^R}(y) + \varepsilon.$$

We deduce that for all $y \in \mathbb{R}$ and $R \geq R_0$, $u_p(y) - \varepsilon \leq (u_p \wedge u_{\eta^R})(y)$. Thus, as the convex hull is the supremum over all dominated, convex functions, this yields

$$u_p - \varepsilon \leq \mathrm{co}(u_p \wedge u_{\eta^R}) \leq u_p,$$

which proves (5.4.5) and completes the proof. $\square$

5.5 Proof of the main theorem

We consider the setting of Theorem 5.2.5. Before entering its technical proof, we argue that it is sufficient to consider the case $r = 1$ and that we can consider without loss of generality that (μ, ν) is irreducible.

When considering a weakly convergent sequence of couplings $(\pi^k)_{k\in\mathbb{N}}$, whose sequence of marginal distributions $(\mu^k, \nu^k)_{k\in\mathbb{N}}$ is converging in $\mathcal{W}_r$, one can deduce $\mathcal{W}_r$-convergence for the sequence of couplings. This is due to $\mathcal{W}_r$-convergence being equivalent to weak convergence coupled with convergence of the r-moments. To see the latter, we find (when equipping $X \times Y$ with the product metric $(d_X^r(x, x') + d_Y^r(y, y'))^{1/r}$)

$$\begin{aligned} &\int_{X\times Y} (d_X^r(x, x_0) + d_Y^r(y, y_0))\ \pi^k(dx, dy) = \mathcal{W}_r^r(\mu^k, \delta_{x_0}) + \mathcal{W}_r^r(\nu^k, \delta_{y_0}) \\ \underset{k\to+\infty}{\longrightarrow}\ &\mathcal{W}_r^r(\mu, \delta_{x_0}) + \mathcal{W}_r^r(\nu, \delta_{y_0}) = \int_{X\times Y} (d_X^r(x, x_0) + d_Y^r(y, y_0))\ \pi(dx, dy). \end{aligned} \tag{5.5.1}$$

By a similar reasoning, we show that convergence in $\mathcal{AW}_r$ holds as soon as $\mathcal{AW}_1$ holds and we have $\mathcal{W}_r$-convergence of the marginals. This is accomplished in the next Lemma.

Lemma 5.5.1. *In the setting of Theorem 5.2.5, assume that there exists a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_1$. Then this sequence also converges to π in $\mathcal{AW}_r$.*

Proof. Since its marginals are convergent, we have relative compactness of $(\pi^k)_{k\in\mathbb{N}} \subset \mathcal{P}_r(\mathbb{R}\times\mathbb{R})$. By [18, Lemma 2.6] this is equivalent to relative compactness of $(J(\pi^k))_{k\in\mathbb{N}}$ in $\mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$, where J is the inclusion map defined above (5.1.3). Let $P \in \mathcal{P}_r(\mathbb{R} \times \mathcal{P}_r(\mathbb{R}))$ be a $\mathcal{W}_r$-accumulation point of the sequence. Again by (5.1.3), we have that $J(\pi^k) \to J(\pi)$ in $\mathcal{W}_1$, and therefore also weakly in $\mathcal{P}(\mathbb{R} \times \mathcal{P}_1(\mathbb{R}))$. Note that $f \in C_b(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$ implies $f \in C_b(\mathbb{R}\times\mathcal{P}_1(\mathbb{R}))$, and therefore, $J(\pi^k)$ converges weakly to $J(\pi)$ in $\mathcal{P}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$. Similarly, we find that P is a weak accumulation point in $\mathcal{P}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$. But the weak topology is Hausdorff, and therefore P has to coincide with $J(\pi)$. As a consequence we have $\mathcal{W}_r$-convergence of $J(\pi^k)$ to $J(\pi)$, and due to (5.1.3) the assertion. □

Next, Proposition 5.2.4 is the key ingredient to show that it is enough to prove Theorem 5.2.5 when (μ, ν) is irreducible.

Lemma 5.5.2. *If the conclusion of Theorem 5.2.5 holds for $r = 1$ and for any irreducible pair of marginals (μ, ν), then it holds for $r = 1$ and for any pair (μ, ν) in the convex order.*

Proof. In the setting of Theorem 5.2.5, fix $\pi \in \Pi_M(\mu, \nu)$. Denote by $(\mu_n, \nu_n)_{n\in N}$ the decomposition of (μ, ν) into irreducible components with

$$\mu = \eta + \sum_{n\in N} \mu_n, \quad \nu = \eta + \sum_{n\in N} \nu_n.$$

By Proposition 5.2.4, we can find sub-probability measures $(\eta^k, \upsilon^k)_{k\in\mathbb{N}}$, $(\mu_n^k)_{(k,n)\in\mathbb{N}\times N}$, $(\nu_n^k)_{(k,n)\in\mathbb{N}\times N}$ such that

$$\eta^k \leq_{cx} \upsilon^k, \quad \mu_n^k \leq_{cx} \nu_n^k \quad \forall (k,n) \in \mathbb{N}\times N,$$
$$\eta^k \to \eta, \quad \upsilon^k \to \eta, \quad \mu_n^k \to \mu_n, \quad \nu_n^k \to \nu_n \quad \text{in } \mathcal{W}_1, \quad k \to +\infty.$$

For $k \in \mathbb{N}$, let $\chi^k \in \Pi_M(\eta^k, \upsilon^k)$ be any martingale coupling between η^k and υ^k. Since the marginals both converge to η in $\mathcal{W}_1$, $(\chi^k)_{k\in\mathbb{N}}$ is tight and any accumulation point with respect to the weak topology belongs to $\Pi_M(\eta,\eta)$. Since $\chi := (\mathrm{id},\mathrm{id})_*\eta$ is the only martingale coupling between η and itself, $(\chi^k)_{k\in\mathbb{N}}$ converges weakly to χ as k goes to $+\infty$ and even in $\mathcal{W}_1$ according to (5.5.1). We can show that this convergence also holds in $\mathcal{AW}_1$. Indeed, according to Proposition 5.2.1, there exists a sequence $\tilde\chi^k \in \Pi(\eta^k, \upsilon^k)$, $k \in \mathbb{N}$, converging to χ in $\mathcal{AW}_1$. Then

$$\begin{aligned}\mathcal{AW}_1(\chi^k, \tilde\chi^k) &\leq \int_{\mathbb{R}} \mathcal{W}_1(\chi_x^k, \tilde\chi_x^k)\, \eta^k(dx) \leq \int_{\mathbb{R}} \left(\mathcal{W}_1(\chi_x^k, \delta_x) + \mathcal{W}_1(\delta_x, \tilde\chi_x^k)\right)\, \eta^k(dx)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} |x'-x|\, (\chi_x^k + \tilde\chi_x^k)(dx')\, \eta^k(dx) = \int_{\mathbb{R}} |x'-x|\, (\chi^k + \tilde\chi^k)(dx,dx').\end{aligned}$$

Since $(x,x') \mapsto |x'-x| \in \Phi_1(\mathbb{R}^2)$ and χ^k and $\tilde\chi^k$ converge to χ in $\mathcal{W}_1$, we deduce using (5.1.1) that

$$\int_{\mathbb{R}} |x'-x|\, (\chi^k + \tilde\chi^k)(dx,dx') \to 2\int_{\mathbb{R}} |x'-x|\, \chi(dx,dx') = 0, \quad k \to +\infty,$$

hence,

$$\mathcal{AW}_1(\chi^k, \chi) \leq \mathcal{AW}_1(\chi^k, \tilde\chi^k) + \mathcal{AW}_1(\tilde\chi^k, \chi) \to 0, \quad k \to +\infty.$$

By assumption, we can find for any $n \in N$ a sequence $(\pi_n^k)_{k\in\mathbb{N}}$ of martingale couplings between μ_n^k and ν_n^k, $k \in \mathbb{N}$, which converges in $\mathcal{AW}_1$ to π_n as k goes to $+\infty$, where π_n denotes π restricted to the n-th irreducible component given by (5.2.7). By Lemma 5.3.7, we have for all $p \in N$ that

$$\chi^k + \sum_{n\in N,\ n\leq p} \pi_n^k \to \chi + \sum_{n\in N,\ n\leq p} \pi_n \quad \text{in } \mathcal{AW}_1, \quad k \to +\infty.$$

Moreover, the respective marginals of $\chi^k + \sum_{n\in N} \pi_n^k$, namely μ^k and ν^k, converge in $\mathcal{W}_1$ to the respective marginals of $\chi + \sum_{n\in N} \pi_n$, namely μ and ν. Therefore, according to Lemma 5.3.6 (b), there exists a constant $C > 0$ such that

$$\limsup_k \mathcal{AW}_1\left(\chi^k + \sum_{n\in N} \pi_n^k, \chi + \sum_{n\in N} \pi_n\right) \leq C\left(I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu)\right),$$

where $\varepsilon_p = \sum_{n\in N, n>p} \mu_n(\mathbb{R})$ where by convention the sum over an empty set is 0. Clearly, $(\varepsilon_p)_{p\in N}$ is a null sequence, thus Lemma 5.3.1 (b) reveals that the right-hand side vanishes as p goes to $\sup N$. This proves that $\pi^k = \chi^k + \sum_{n\in N} \pi_n^k \in \Pi_M(\mu^k, \nu^k)$ converges in $\mathcal{AW}_1$ to $\pi = \chi + \sum_{n\in N} \pi^k \in \Pi_M(\mu,\nu)$. $\square$

Proof of Theorem 5.2.5. Step 1. Due to Lemma 5.5.1 and Lemma 5.5.2, we may suppose without loss of generality that $r = 1$ and (μ, ν) is irreducible with component $I = (l, r)$, $l \in \mathbb{R} \cup \{-\infty\}$, $r \in \mathbb{R} \cup \{+\infty\}$. Next, we define auxiliary martingale couplings close to π which will be easier to approximate in the limit. We define them with the same first marginal distribution whereby the second marginal distribution is smaller with respect to the convex order. These auxiliary couplings will satisfy two key properties: first, their second marginal distribution must be concentrated on a compact subset of I when the first marginal distribution is itself concentrated on a certain compact subset K of I. Second, it is essential that their second marginal distribution has positive mass on two compact subsets of I on both sides of K.

Fix $\varepsilon \in (0, \frac{1}{2})$. Choose a compact subset $K = [a, b]$ of I with

$$\mu(K^{\complement}) < \varepsilon. \tag{5.5.2}$$

Instead of directly approximating π, we initially consider the martingale coupling $\pi^{R,\alpha}$ whose definition is given below. For any $R > 0$, let $(\pi_x^R)_{x\in\mathbb{R}}$ be the probability kernel obtained by virtue of Lemma 5.4.4. By Lemma 5.4.4 (a) we have for all $x \in \mathbb{R}$ that $\pi_x^R \leq_{cx} \pi_x$. Therefore,

$$\mathcal{W}_1(\pi_x^R, \pi_x) \leq 2 \int_{\mathbb{R}} |y| \, \pi_x(dy),$$

where the right-hand side is a μ-integrable function of x. By Lemma 5.4.4 (b) we find $\pi_x^R \to \pi_x$ in $\mathcal{W}_1$ as $R \to +\infty$. Let $\pi^R := \mu \times \pi_x^R$, then dominated convergence yields

$$\mathcal{AW}_1(\pi^R, \pi) \leq \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x) \, \mu(dx) \to 0, \quad R \to +\infty.$$

Denote by ν^R the second marginal of π^R. Consequently, ν^R converges to ν for the $\mathcal{W}_1$-distance and $\nu^R \leq_{cx} \nu$ for all $R > 0$. Let $\tilde{a}$ and $\tilde{b}$ be real numbers such that $\tilde{a} \in (l, a)$ and $\tilde{b} \in (b, r)$, for instance

$$\tilde{a} = \frac{l + a}{2} \vee (a - 1) \text{ and } \tilde{b} = (b + 1) \wedge \frac{b + r}{2}.$$

Since (μ, ν) is irreducible on I, ν assigns positive mass to any neighbourhood of the endpoints l and r of I. From now on, we use the notational convention that for all $c \in \mathbb{R} \cup \{\pm\infty\}$,

$$[-\infty, c) = \{x \in \mathbb{R} \mid x < c\}, \quad (c, +\infty] = \{x \in \mathbb{R} \mid c < x\} \quad \text{and} \quad [-\infty, +\infty] = \mathbb{R}.$$

In particular, $\overline{I} = [l, r] \subset \mathbb{R}$. Then $[l, \tilde{a})$ and $(\tilde{b}, r]$ are relatively open on the common support $\overline{I}$ of the probability measures ν^R and ν, so Portmanteau's theorem yields

$$\liminf_{R\to+\infty} \nu^R([l, \tilde{a})) \geq \nu([l, \tilde{a})) > 0, \quad \liminf_{R\to+\infty} \nu^R((\tilde{b}, r]) \geq \nu((\tilde{b}, r]) > 0.$$

Thus, we deduce that we can choose $R > 0$ such that

$$R \geq (-a) \vee b, \quad \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x) \, \mu(dx) < \varepsilon, \quad \nu^R([l, \tilde{a})) > 0, \quad \text{and} \quad \nu^R((\tilde{b}, r]) > 0. \tag{5.5.3}$$

Let $\pi_x^{R,\alpha}$ be the image of π_x^R by $y \mapsto \alpha(y-x)+x$ when $\alpha \in (0,1)$. Then $\pi^{R,\alpha} := \mu \times \pi_x^{R,\alpha}$ satisfies by Lemma 5.4.2

$$\begin{aligned}
\mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi) &\le \int_{\mathbb{R}} \mathcal{W}_1(\varepsilon\pi_x + (1-\varepsilon)\pi_x^{R,\alpha}, \pi_x)\,\mu(dx) \\
&\le (1-\varepsilon)\int_{\mathbb{R}} \mathcal{W}_1(\pi_x^{R,\alpha}, \pi_x)\,\mu(dx) \\
&\le \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^{R,\alpha}, \pi_x^R)\,\mu(dx) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\,\mu(dx) \\
&= (1-\alpha)\int_{\mathbb{R}}\int_{\mathbb{R}} |x-y|\,\pi_x^R(dy)\,\mu(dx) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\,\mu(dx) \\
&\le (1-\alpha)\left(\int_{\mathbb{R}} |x|\,\mu(dx) + \int_{\mathbb{R}} |y|\,\nu^R(dy)\right) + \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\,\mu(dx),
\end{aligned}$$

where the right-hand side converges to $\int_{\mathbb{R}} \mathcal{W}_1(\pi_x^R, \pi_x)\,\mu(dx) < \varepsilon$ for $\alpha \to 1$. Note that $\frac{2R-a-\tilde{a}}{2R-2\tilde{a}}, \frac{b+\tilde{b}+2R}{\tilde{2}b+2R} \in (0,1)$, so we can choose $\alpha \in (0,1)$ such that

$$\mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi) < \varepsilon \quad \text{and} \quad \alpha \ge \frac{2R-a-\tilde{a}}{2R-2\tilde{a}} \vee \frac{b+\tilde{b}+2R}{\tilde{2}b+2R}. \tag{5.5.4}$$

Let L be a compact subset of I such that the interior $\mathring{L}$ of L satisfies

$$[(-R) \vee (\alpha l + (1-\alpha)a), R \wedge (\alpha r + (1-\alpha)b)] \subset \mathring{L}.$$

Because $R \ge (-a) \vee b$ and thereby $[a,b] = K \subset [-R,R]$, we have that $\mu|_K \times \pi_x^R$ is concentrated on $K \times ([-R,R] \cap \overline{I})$. Furthermore, for any $(x,y) \in K \times ([-R,R] \cap \overline{I})$, we find $\alpha y + (1-\alpha)x \in \mathring{L}$. Hence, $\mu|_K \times \pi_x^{R,\alpha}$ is concentrated on $K \times \mathring{L}$.

Denote the second marginal of $\pi^{R,\alpha}$ by $\nu^{R,\alpha}$. Since

$$(x,y) \in (l,R) \times [l,\tilde{a}) \implies l < (1-\alpha)x + \alpha y < R - \alpha(R-\tilde{a}) \le \frac{a+\tilde{a}}{2},$$

we have that

$$\begin{aligned}
\nu^{R,\alpha}\left(\left(l, \frac{a+\tilde{a}}{2}\right)\right) &= \int_{\mathbb{R}^2} \mathbb{1}_{\left(l,\frac{a+\tilde{a}}{2}\right)}(y)\,\pi^{R,\alpha}(dx,dy) = \int_{\mathbb{R}^2} \mathbb{1}_{\left(l,\frac{a+\tilde{a}}{2}\right)}(\alpha y + (1-\alpha)x)\,\pi^R(dx,dy) \\
&\ge \int_{\mathbb{R}^2} \mathbb{1}_{(l,R)\times[l,\tilde{a})}(y)\,\pi^R(dx,dy) = \int_{(l,R)} \pi_x^R((-\infty,\tilde{a}))\,\mu(dx).
\end{aligned}$$

If $x \in [R,+\infty)$, then $\pi_x^R = \delta_x$ and since $R \ge \tilde{a}$, $\pi_x^R((-\infty,\tilde{a})) = 0$. Added to the fact that μ is concentrated on I, we obtain

$$\int_{(l,R)} \pi_x^R((-\infty,\tilde{a}))\,\mu(dx) = \int_{\mathbb{R}} \pi_x^R((-\infty,\tilde{a}))\,\mu(dx) = \nu^R((-\infty,\tilde{a})) = \nu^R([l,\tilde{a})) > 0.$$

We deduce that

$$\nu^{R,\alpha}\left(\left(l, \frac{a+\tilde{a}}{2}\right)\right) > 0, \text{ and similarly, } \nu^{R,\alpha}\left(\left(\frac{b+\tilde{b}}{2}, r\right)\right) > 0. \tag{5.5.5}$$

To summarise, we have constructed a martingale coupling $\pi^{R,\alpha} \in \Pi_M(\mu, \nu^{R,\alpha})$ close to π with respect to the $\mathcal{AW}_1$-distance, whose restriction $\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is compactly supported on $K \times L$ and concentrated on $K \times \mathring{L}$. Moreover, the second marginal distribution $\nu^{R,\alpha}$ is dominated by ν in the convex order and assigns positive mass on both sides of K.

Step 2. In the next step we construct a sequence of sub-probability martingale couplings supported on a compact cube $J \times J$ ($K \subset J \subset I$) converging to $\pi^{R,\alpha}|_{K\times\mathbb{R}}$.

Our first goal is to find a sequence $\nu^{R,\alpha,k}$, $k \in \mathbb{N}$, such that $\mu^k \leq_{cx} \nu^{R,\alpha,k} \leq_{cx} \nu^k$ and

$$\mathcal{W}_1(\nu^{R,\alpha,k}, \nu^{R,\alpha}) \to 0, \quad k \to \infty. \tag{5.5.6}$$

Defining $\nu^{R,\alpha,k}$ by

$$\nu^{R,\alpha,k} = \nu^k \wedge_c (\mu^k \vee_c T_k(\nu^{R,\alpha})),$$

where T_k denotes the translation by the difference between the common barycentre of μ^k and ν^k and the common barycentre of ν and $\nu^{R,\alpha}$, i.e., $\int_{\mathbb{R}} y\, \nu^k(dy) - \int_{\mathbb{R}} y\, \nu^{R,\alpha}(dy)$, fulfils these requirements. Indeed

$$\mathcal{W}_1(T_k(\nu^{R,\alpha}), \nu^{R,\alpha}) = \left| \int_{\mathbb{R}} y\, \nu^k(dy) - \int_{\mathbb{R}} y\, \nu(dy) \right| \leq \mathcal{W}_1(\nu^k, \nu) \to 0,$$

as k goes to $+\infty$. Then Lemma 5.4.1 provides $\nu^{R,\alpha,k} \to \nu \wedge_{cx} (\mu \vee_c \nu^{R,\alpha}) = \nu^{R,\alpha}$ in $\mathcal{W}_1$ as k goes to $+\infty$. By Proposition 5.2.1 we can approximate $\pi^{R,\alpha}$ with couplings $\pi^{R,\alpha,k} \in \Pi(\mu^k, \nu^{R,\alpha,k})$ in $\mathcal{AW}_1$. Unfortunately the sequence $\pi^{R,\alpha,k}$, $k \in \mathbb{N}$ does not have to consist of solely martingale couplings. Therefore, we have to adjust the barycentres of its disintegrations, $(\pi_x^{R,\alpha,k})$ to obtain martingale kernels and thereby martingale couplings. Due to (5.5.5) and inner regularity of $\nu^{R,\alpha}$, we find compact sets

$$L_- \subset \left(l, \frac{a+\tilde{a}}{2}\right), \quad L_+ \subset \left(\frac{b+\tilde{b}}{2}, r\right)$$

with $\nu^{R,\alpha}$-positive measure. Let $\tilde{l}, \tilde{r} \in I$, be such that $\tilde{l} < \inf(L \cup L-)$ and $\sup(L \cup L_+) < \tilde{r}$. Then define

$$\tilde{L}_- = \left(\tilde{l}, \frac{a+\tilde{a}}{2}\right), \quad \tilde{L}_+ = \left(\frac{b+\tilde{b}}{2}, \tilde{r}\right) \quad \text{and} \quad \tilde{K} = \left(\frac{3a+\tilde{a}}{4}, \frac{3b+\tilde{b}}{4}\right), \tag{5.5.7}$$

so that $\tilde{L}_-$, $\tilde{L}_+$ and $\tilde{K}$ are bounded and open intervals covering respectively L_-, L_+ and K and such that the distance e between $\tilde{L}_- \cup \tilde{L}_+$ and $\tilde{K}$ is positive:

$$e = \inf\left\{|x-y| \mid (x,y) \in (\tilde{L}_- \cup \tilde{L}_+) \times \tilde{K}\right\} \geq \frac{a-\tilde{a}}{4} \wedge \frac{\tilde{b}-b}{4} > 0.$$

Denoting $J = [\tilde{l}, \tilde{r}]$, Figure 5.2 summarises the construction.

The respective restrictions of $\nu^{R,\alpha,k}$ to $\tilde{L}_-$ and $\tilde{L}_+$ are denoted by ν_-^k and ν_+^k, respectively. Since $\tilde{L}_-$ and $\tilde{L}_+$ are open, Portmanteau's theorem ensures that eventually (for k sufficiently large) ν_-^k and ν_+^k each have more total mass than some constant $\delta > 0$.

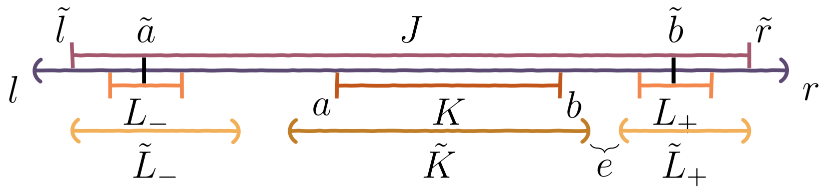

Figure 5.2: Points and intervals involved in the proof. The boundaries of the closed intervals are vertical bars and those of the open intervals are parenthesis.

By Lemma 5.3.4 (ii) there are $\hat\mu^k \le \mu^k$, $\hat\nu^k \le \nu^{R,\alpha,k}$, $\hat\pi^k = \hat\mu^k \times \hat\pi^k_x \in \Pi(\hat\mu^k, \hat\nu^k)$ concentrated on $\tilde K \times \mathring{L}$, and $\varepsilon_k \ge 0$ such that

$$\mathcal{AW}_1(\hat\pi^k, (1-\varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) + \varepsilon_k \to 0, \quad k \to +\infty. \tag{5.5.8}$$

The following procedure shows that there are for $\hat\mu^k(dx)$-almost every x unique constants $c^k_-(x), c^k_+(x) \in [0,+\infty)$ and $d^k(x) \in [1,+\infty)$ such that

$$\tilde\pi^k_x := \frac{\hat\pi^k_x + c^k_+(x)\nu^k_+ + c^k_-(x)\nu^k_-}{d^k(x)} \in \mathcal{P}(\mathbb{R}), \quad \int_{\mathbb{R}} y\,\tilde\pi^k_x(dy) = x, \quad c^k_-(x) \wedge c^k_+(x) = 0.$$

Note that the constraint $c^k_-(x) \wedge c^k_+(x) = 0$ provides

$$\int_{\mathbb{R}} y\,\hat\pi^k_x(dy) \le x \implies c^k_-(x) = 0, \quad \int_{\mathbb{R}} y\,\hat\pi^k_x(dy) \ge x \implies c^k_+(x) = 0. \tag{5.5.9}$$

We require $\tilde\pi^k_x$ to be a probability measure with mean x, thus,

$$1 + c^k_+(x)\nu^k_+(\mathbb{R}) + c^k_-(x)\nu^k_-(\mathbb{R}) = d^k(x), \tag{5.5.10}$$

$$\int_{\mathbb{R}} y\,\hat\pi^k_x(dy) + c^k_+(x)\int_{\mathbb{R}} y\,\nu^k_+(dy) + c^k_-(x)\int_{\mathbb{R}} y\,\nu^k_-(dy) = xd^k(x). \tag{5.5.11}$$

Combining (5.5.9) with (5.5.10) and (5.5.11) yields

$$\begin{aligned}
c^k_-(x) &= \frac{\int_{\mathbb{R}} y\,\hat\pi^k_x(dy) - x}{\int_{\mathbb{R}}(x-y)\,\nu^k_-(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\,\hat\pi^k_x(dy)\right|}{e\nu^k_-(\mathbb{R})}\right], \\
c^k_+(x) &= \frac{x - \int_{\mathbb{R}} y\,\hat\pi^k_x(dy)}{\int_{\mathbb{R}}(y-x)\,\nu^k_+(dy)} \vee 0 \in \left[0, \frac{\left|x - \int_{\mathbb{R}} y\,\hat\pi^k_x(dy)\right|}{e\nu^k_+(\mathbb{R})}\right], \\
d^k(x) &= 1 + c^k_-(x)\nu^k_-(\mathbb{R}) + c^k_+(x)\nu^k_+(\mathbb{R}) \in \left[1, 1 + \frac{\left|x - \int_{\mathbb{R}} y\,\hat\pi^k_x(dy)\right|}{e}\right].
\end{aligned}$$

Remember from (5.5.7) that $L \cup \tilde L_- \cup \tilde L_+ \subset [\tilde l, \tilde r] \subset I$. Then we obtain for $\hat\mu^k(dx)$-almost every x the estimate

$$\mathcal{W}_1(\tilde\pi^k_x, \hat\pi^k_x) \le \mathcal{W}_1\left(\frac{c^k_+(x)\nu^k_+ + c^k_-(x)\nu^k_-}{d^k(x)}, \frac{d^k(x)-1}{d^k(x)}\hat\pi^k_x\right) \le \frac{d^k(x)-1}{d^k(x)}|\tilde r - \tilde l|$$

$$\leq \frac{\left|x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy)\right|}{e} |\tilde{r} - \tilde{l}|.$$

Hence, the adapted Wasserstein distance between $\hat{\pi}^k$ and $\tilde{\pi}^k = \hat{\mu}^k \times \tilde{\pi}^k_x$ satisfies

$$\begin{aligned}\mathcal{AW}_1(\tilde{\pi}^k, \hat{\pi}^k) &\leq \int_{\mathbb{R}} \mathcal{W}_1(\tilde{\pi}^k_x, \hat{\pi}^k_x)\, \hat{\mu}^k(dx) \leq \frac{|\tilde{r} - \tilde{l}|}{e} \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy) \right| \hat{\mu}^k(dx) \\ &\leq \frac{|\tilde{r} - \tilde{l}|}{e} \mathcal{AW}_1(\hat{\pi}^k, (1 - \varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}),\end{aligned}$$

where we used Remark 5.2.2 in the last inequality. The triangle inequality and (5.5.8) then yield

$$\lim_k \mathcal{AW}_1(\tilde{\pi}^k, (1 - \varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) \to 0, \quad k \to \infty. \tag{5.5.12}$$

Next we bound the total mass which we require to fix the barycentres. We find that

$$\begin{aligned}\int_{\mathbb{R}} \frac{c^k_-(x) + c^k_+(x)}{d^k(x)}\, \hat{\mu}^k(dx) &\leq \frac{1}{e(\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}))} \int_{\mathbb{R}} \left| x - \int_{\mathbb{R}} y\, \hat{\pi}^k_x(dy) \right| \hat{\mu}^k(dx) \\ &\leq \frac{\mathcal{AW}_1(\hat{\pi}^k, (1 - \varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}})}{e(\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}))} \to 0, \quad k \to +\infty,\end{aligned}$$

where we used Remark 5.2.2 again for the last inequality and the fact that $\nu^k_-(\mathbb{R}) \wedge \nu^k_+(\mathbb{R}) \geq \delta$ for k large enough for the limit. Consequently, when $\tilde{\nu}^k$ denotes the second marginal of $\tilde{\pi}^k$, we have for k sufficiently large that

$$(1 - 2\varepsilon)\tilde{\nu}^k \leq (1 - 2\varepsilon)\hat{\nu}^k + (1 - 2\varepsilon)(\nu^k_- + \nu^k_+) \int_{\mathbb{R}} \frac{c^k_-(x) + c^k_+(x)}{d^k(x)}\, \hat{\mu}^k(dx) \leq (1 - \varepsilon)\nu^{R,\alpha,k}.$$

Step 3. In this step, we complement the martingale coupling $(1 - 2\varepsilon)\tilde{\pi}^k$ to a martingale coupling with marginals μ^k and $\varepsilon\nu^k + (1 - \varepsilon)\nu^{R,\alpha,k}$ for k sufficiently large. Recall that $\tilde{\pi}^k \in \Pi_M(\hat{\mu}^k, \tilde{\nu}^k)$ and $\pi^{R,\alpha}|_{K\times\mathbb{R}} \in \Pi_M(\mu|_K, \check{\nu}^{R,\alpha})$, where $\check{\nu}^{R,\alpha}$ is the second marginal distribution of $\pi^{R,\alpha}|_{K\times\mathbb{R}}$, are concentrated on the compact cube $J \times J$ and

$$\mathcal{AW}_1(\tilde{\pi}^k, (1 - \varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}) \to 0, \quad k \to +\infty.$$

Furthermore, since $(1-\varepsilon)\pi^{R,\alpha} - (1-2\varepsilon)\pi^{R,\alpha}|_{K\times\mathbb{R}}$ is a martingale coupling with marginals

$$(1 - \varepsilon)\mu - (1 - 2\varepsilon)\mu|_K \quad \text{and} \quad (1 - \varepsilon)\nu^{R,\alpha} - (1 - 2\varepsilon)\check{\nu}^{R,\alpha},$$

we deduce by irreducibility of the pair (μ, ν) on I irreducibility of the pair of sub-probability measures

$$\varepsilon\mu + (1 - \varepsilon)\mu - (1 - 2\varepsilon)\mu|_K \quad \text{and} \quad \varepsilon\nu + (1 - \varepsilon)\nu^{R,\alpha} - (1 - 2\varepsilon)\check{\nu}^{R,\alpha},$$

whose potential functions satisfy

$$0 \leq u_\mu - u_{(1-2\varepsilon)\mu|_K} < u_{\varepsilon\nu+(1-\varepsilon)\nu^{R,\alpha}} - u_{(1-2\varepsilon)\check{\nu}^{R,\alpha}} \quad \text{on } I.$$

Since those potential functions are continuous, there exists $\tau > 0$ such that they have distance greater τ on J. By uniform convergence of potential functions, for $k \in \mathbb{N}$ sufficiently large we have

$$0 \leq u_{\mu^k} - u_{(1-2\varepsilon)\hat{\mu}^k} + \frac{\tau}{2} \leq u_{\varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k}} - u_{(1-2\varepsilon)\tilde{\nu}^k} \quad \text{on } J.$$

On the complement of J we have $u_{(1-2\varepsilon)\hat{\mu}^k} = u_{(1-2\varepsilon)\tilde{\nu}^k}$ since both measures are concentrated on J and satisfy $(1-2\varepsilon)\int_{\mathbb{R}} x\, \hat{\mu}^k(dx) = (1-2\varepsilon)\int_{\mathbb{R}} y\, \tilde{\nu}^k(dy)$. Therefore,

$$0 \leq u_{\mu^k} - u_{(1-2\varepsilon)\hat{\mu}^k} \leq u_{\varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k}} - u_{(1-2\varepsilon)\tilde{\nu}^k} \quad \text{on } J^c.$$

By Strassen's theorem [183], there exists $\eta^k \in \Pi_M(\mu^k - (1-2\varepsilon)\hat{\mu}^k, \varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k} - (1-2\varepsilon)\tilde{\nu}^k)$. Finally, we write

$$\overline{\pi}^k = \eta^k + (1-2\varepsilon)\tilde{\pi}^k \in \Pi_M(\mu^k, \varepsilon\nu^k + (1-\varepsilon)\nu^{R,\alpha,k}).$$

Step 4. In the last step, we show that the sequence constructed in this way is eventually close to the original martingale coupling π in adapted Wasserstein distance.

The marginals of $\overline{\pi}^k$ are converging in $\mathcal{W}_1$ to $(\mu, \varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha})$ as k goes to $+\infty$. We have according to (5.5.12) that

$$(1-2\varepsilon)\mathcal{AW}_1(\tilde{\pi}^k, (1-\varepsilon_k)\pi^{R,\alpha}) \to 0, \quad k \to \infty.$$

Since $\mu(K) \geq 1-\varepsilon$, $(1-2\varepsilon)(1-\varepsilon) > 1-3\varepsilon$, and ε_k vanishes as k goes to $+\infty$, we have for $k \in \mathbb{N}$ large enough

$$\begin{aligned}
\eta^k(\mathbb{R}^2) &= \left(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha} - (1-2\varepsilon)(1-\varepsilon_k)\pi^{R,\alpha}|_{K\times\mathbb{R}}\right)(\mathbb{R}^2) \\
&= 1 - (1-2\varepsilon)(1-\varepsilon_k)\mu(K) \\
&\leq 3\varepsilon,
\end{aligned}$$

hence (5.3.35) writes

$$\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) \leq C(I_{3\varepsilon}(\mu) + I_{3\varepsilon}(\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha})),$$

where C depends only on the exponent $r = 1$. Since $\nu^{R,\alpha} \leq_{cx} \nu$, then $\varepsilon\nu + (1-\varepsilon)\nu^{R,\alpha} \leq_{cx} \nu$, so using Lemma 5.3.1 (e), the triangle inequality and (5.5.4), we get

$$\begin{aligned}
\limsup_k \mathcal{AW}_1(\overline{\pi}^k, \pi) &\leq \limsup_k \left(\mathcal{AW}_1(\overline{\pi}^k, \varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}) + \mathcal{AW}_1(\varepsilon\pi + (1-\varepsilon)\pi^{R,\alpha}, \pi)\right) \\
&\leq C(I_{3\varepsilon}(\mu) + I_{3\varepsilon}(\nu)) + \varepsilon.
\end{aligned}$$

Since the right-hand side only depends on ε and vanishes as ε goes to 0, we can reason like in the proof of Proposition 5.2.3 (from (5.3.26)) to find a null sequence $(\tilde{\varepsilon}_k)_{k\in\mathbb{N}}$, two sequences $(R_k)_{k\in\mathbb{N}}$, $(\alpha_k)_{k\in\mathbb{N}}$ with values respectively in $\mathbb{R}_+^*$ and $(0,1)$, and martingale couplings

$$\mathring{\pi}^k \in \Pi_M(\mu^k, \tilde{\varepsilon}_k\nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}), \quad k \in \mathbb{N}$$

such that

$$\mathcal{AW}_1(\mathring{\pi}^k, \pi) \to 0, \quad k \to +\infty. \tag{5.5.13}$$

In particular, the $\mathcal{W}_1$-distance of their second marginal distributions vanishes as k goes to $+\infty$, hence the triangle inequality yields

$$\mathcal{W}_1(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k) \leq \mathcal{W}_1(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu) + \mathcal{W}_1(\nu, \nu^k) \to 0, \quad k \to +\infty.$$

Remember that $\nu^{R_k,\alpha_k,k} \leq_{cx} \nu^k$, hence $\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k} \leq_{cx} \nu^k$. Then by Theorem 2.2.12, there exist martingale couplings $M^k \in \Pi_M(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k)$, $k \in \mathbb{N}$ such that

$$\int_{\mathbb{R}\times\mathbb{R}} |x-y| \, M^k(dx, dy) \leq 2\mathcal{W}_1(\tilde{\varepsilon}_k \nu^k + (1-\tilde{\varepsilon}_k)\nu^{R_k,\alpha_k,k}, \nu^k) \to 0, \quad k \to +\infty. \tag{5.5.14}$$

Let then π^k be the joint distribution of the first and third marginals of $\mathring{\pi}^k \dot{\otimes} M^k$, that is

$$\pi^k(dx, dy) = \mu^k(dx) \int_{z\in\mathbb{R}} M^k_z(dy) \, \mathring{\pi}^k_x(dz) \in \Pi_M(\mu^k, \nu^k).$$

Using the fact that for $\mu^k(dx)$-almost every x, $\mathring{\pi}^k_x(dz) \, M^k_z(dy) \in \Pi(\mathring{\pi}^k_x, \pi^k_x)$, we get

$$\begin{aligned}\mathcal{AW}_1(\pi^k, \mathring{\pi}^k) &\leq \int_{\mathbb{R}} \mathcal{W}_1(\pi^k_x, \mathring{\pi}^k_x) \, \mu^k(dx) \leq \int_{\mathbb{R}\times\mathbb{R}\times\mathbb{R}} |z-y| \, \mu^k(dx) \, M^k_z(dy) \, \mathring{\pi}^k_x(dz) \\ &= \int_{\mathbb{R}\times\mathbb{R}} |z-y| \, M^k(dy, dz),\end{aligned}$$

where the right-hand side vanishes by (5.5.14) as k goes to $+\infty$. Then (5.5.13) and the triangle inequality yield

$$\mathcal{AW}_1(\pi^k, \pi) \leq \mathcal{AW}_1(\pi^k, \mathring{\pi}^k) + \mathcal{AW}_1(\mathring{\pi}^k, \pi) \to 0, \quad k \to +\infty,$$

which concludes the proof. $\square$

Chapter 6

Stability of weak martingale optimal transport and weak optimal transport

Abstract

Motivated by the study of the concentration of measure phenomenon and their connection with transport-entropy inequalities, Gozlan, Roberto, Samson and Tetali [90] introduced a general notion of transport cost problem called the Weak Optimal Transport (WOT) problem. Backhoff-Veraguas, Beiglböck and Pammer [18] and Backhoff-Veraguas and Pammer [20] proved that under some regularity assumption on the cost function, existence, uniqueness and stability hold for the WOT problem. Thanks to a result proved in the companion paper (Chapter 5), we recover those result in a different way.

Because the martingale constraint reflects the condition for a financial market to be arbitrage free, it is natural in the context of mathematical finance to consider the martingale counterpart of the WOT problem, namely the Weak Martingale Optimal Transport (WMOT) problem. Thanks to the main theorem of the companion paper, we prove the existence, the uniqueness and most importantly the stability of the WMOT problem under mild regularity assumption of the cost function.

We also prove that martingale C-monotonicity is sufficient for optimality of the WMOT problem, that the so called Wasserstein projections are Lipschitz continuous in dimension 1 and finally we establish the convergence in a space of extended martingale couplings. We discuss a consequence on the surreplication bound for VIX futures.

Keywords: Martingale Optimal Transport, Adapted Wasserstein distance, Robust finance, Weak transport, Stability, Convex order, Martingale couplings.

6.1 Introduction and motivations

6.1.1 The Weak Optimal Transport problem

Motivated by the study of the concentration of measure phenomenon and their connection with transport-entropy inequalities, whose extensive study can be found in [130, 88, 44], Gozlan, Roberto, Samson and Tetali [90] introduced a general notion of transport cost problem called the Weak Optimal Transport (WOT) problem, which they studied with Shu in [89]. In order to define it we introduce some notation. Let X and Y be two Polish spaces respectively endowed with the compatible and complete metrics d_X and d_Y. Let μ be in the set $\mathcal{P}(X)$ of probability measures on X, $\nu \in \mathcal{P}(Y)$ and $C : X \times \mathcal{P}(Y) \to \mathbb{R}_+$ be a nonnegative measurable function. We denote by $\Pi(\mu, \nu)$ the set of couplings between μ and ν, that is $\pi \in \Pi(\mu, \nu)$ iff $\pi \in \mathcal{P}(X \times Y)$ is such that for any measurable subsets $A \subset X$ and $B \subset Y$, $\pi(A \times Y) = \mu(A)$ and $\pi(X \times B) = \nu(B)$. Then the WOT problem consists in the minimisation

$$V_C(\mu, \nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int_X C(x, \pi_x)\, \mu(dx), \tag{WOT}$$

where for all $\pi \in \Pi(\mu, \nu)$, $(\pi_x)_{x \in \mathbb{R}}$ denotes a disintegration of π with respect to its first marginal, which we write $\pi(dx, dy) = \mu(dx)\, \pi_x(dy)$, or with a slight abuse of notation, $\pi = \mu \times \pi_x$ if the context is not ambiguous. Note that for a measurable map $c : X \times Y \to \mathbb{R}_+$, the WOT problem with the cost function $C : (x, p) \mapsto \int_Y c(x, y)\, p(dy)$ amounts to the classical Optimal Transport (OT) problem already discussed in the companion paper (Chapter 5). In particular it was mentioned that the OT theory covers an impressive range of applications. This particularity seems to be shared with the WOT problem, which benefits of high flexibility. One could for instance see the recent work of Backhoff-Veraguas and Pammer [21] and the references inside for an investigation of a connection of the WOT problem with the Schrödinger problem, the Brenier-Strassen Theorem, optimal mechanism design, linear transfers and semimartingale transport.

In order to gain some insight on the WOT problem, we recall some results of paramount importance, namely existence, uniqueness and stability. To formulate those results we need to introduce a more specific setting. From now on, we fix $r \geq 1$ and x_0, y_0 two arbitrary elements of X and Y respectively, their specific value having no impact on our study. Let $\mathcal{P}^r(X)$ denote the set of all probability measures on X with finite r-th moment, i.e.

$$\mathcal{P}^r(X) = \left\{ p \in \mathcal{P}(X) \mid \int_X d_X^r(x, x_0)\, p(dx) < +\infty \right\}. \tag{6.1.1}$$

Let $\mathcal{C}(X)$ denote the set of all real-valued continuous functions on X. The set $\mathcal{P}^r(X)$ is equipped with the weak topology induced by

$$\Phi^r(X) = \{ f \in C(X) \mid \exists \alpha > 0,\ \forall x \in X,\ |f(x)| \leq \alpha(1 + d_X^r(x, x_0)) \}.$$

Then a sequence $(p^k)_{k \in \mathbb{N}}$ converges in $\mathcal{P}^r(X)$ to p iff

$$\forall g \in \Phi^r(X), \quad p^k(g) := \int_X g(x)\, p^k(dx) \underset{k \to +\infty}{\longrightarrow} p(g) := \int_X g(x)\, p(dx).$$

It is well known that the latter is equivalent to

$$\mathcal{W}_r(p^k, p) := \inf_{\pi \in \Pi(p^k, p)} \left(\int_{X \times X} d_X^r(x, y)\, \pi(dx, dy) \right)^{\frac{1}{r}} \underset{k \to +\infty}{\longrightarrow} 0,$$

where $\mathcal{W}_r$ is the Wasserstein distance with index r, which is a metric on $\mathcal{P}^r(X)$ compatible with its topology, turning $\mathcal{P}^r(X)$ into a complete separable metric space, see [8, 168, 190, 191] for much more details.

Back to the WOT problem, consider a cost function $C : X \times \mathcal{P}^r(Y) \to \mathbb{R}_+$ being lower semicontinuous and convex in its second argument. Backhoff-Veraguas, Beiglböck and Pammer [18] prove that for all $(\mu, \nu) \in \mathcal{P}(X) \times \mathcal{P}^r(Y)$, the infimum $V_C(\mu, \nu)$ is attained. Moreover, the map $(\mu, \nu) \mapsto V_C(\mu, \nu)$ is lower semicontinuous. Backhoff-Veraguas and Pammer [20] also prove the stability of the WOT problem: suppose that $C \in \Phi^r(X \times \mathcal{P}^r(Y))$ and let $\mu^k \in \mathcal{P}(X), \nu^k \in \mathcal{P}^r(Y)$, $k \in \mathbb{N}$ converge respectively weakly to $\mu \in \mathcal{P}(X)$ and in $\mathcal{W}_r$ to $\nu \in \mathcal{P}^r(Y)$ as k goes to $+\infty$. For $k \in \mathbb{N}$, let $\pi^k \in \Pi(\mu^k, \nu^k)$ be optimal for $V(\mu^k, \nu^k)$. Then any accumulation point of $(\pi^k)_{k \in \mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C(\mu, \nu)$. If the latter has a unique minimiser π^*, which happens for instance if C is strictly convex in its second argument, then π^k converges weakly to π^* as k goes to $+\infty$. In the latter case, when μ^k, μ belong to $\mathcal{P}^r(X)$ and the convergence of μ^k to μ holds in $\mathcal{W}_r$, then one can easily show that

$$\pi^k \underset{k \to +\infty}{\longrightarrow} \pi^* \text{ in } \mathcal{W}_r. \tag{6.1.2}$$

However, the topology induced by the Wasserstein distance in not always well suited for any setting, especially in mathematical finance, since its symmetry does not take into account the temporal structure of martingales. As explained in the companion paper (Chapter 5), it is sometimes necessary to strengthen the usual topology and therefore consider the adapted Wasserstein distance $\mathcal{AW}_r$ of index r defined for all couplings $\pi = \mu \times \pi_x, \pi' = \mu' \times \pi'_x \in \mathcal{P}(X \times Y)$ by

$$\mathcal{AW}_r(\pi, \pi') = \inf_{\chi \in \Pi(\mu, \mu')} \left(\int_{X \times X} \left(d_X^r(x, x') + \mathcal{W}_r^r(\pi_x, \pi'_{x'}) \right) \chi(dx, dx') \right)^{\frac{1}{r}}.$$

It will prove important to observe that

$$\mathcal{AW}_r(\pi, \pi') = \mathcal{W}_r(J(\pi), J(\pi')), \tag{6.1.3}$$

where J is the trivial embedding map from $\mathcal{P}(X \times Y)$ to $\mathcal{P}(X \times \mathcal{P}(Y))$, namely

$$J : \mathcal{P}(X \times Y) \ni \pi = \mu \times \pi_x \mapsto \mu(dx)\, \delta_{\pi_x}(dp) \in \mathcal{P}(X \times \mathcal{P}(Y)). \tag{6.1.4}$$

There exist other ways to adapt the usual weak topology: Hellwig's information topology [101], Aldous's extended weak topology [2] or the optimal stopping topology [15]. But we do not lose generality by using the topology induced by the adapted Wasserstein distance since strikingly, all those apparently independent topologies are actually equal, at least in discrete time [15, Theorem 1.1]. By the connection Backhoff-Veraguas and Pammer establish between the WOT problem and an extended version of it, which minimises a cost function over a set $\Lambda(\mu,\nu)$ of extended couplings (see Section 6.2.6 below for a proper definition), in the setting of (6.1.2) when C is strictly convex in its second argument, we can derive the convergence of $J(\pi^k)$ to $J(\pi^*)$ in $\mathcal{W}_r$ as k goes to $+\infty$, which by (6.1.3) is equivalent to the convergence of the minimiser π^k of $V_C(\mu^k,\nu^k)$ to the only minimiser π^* of $V_C(\mu,\nu)$ in $\mathcal{AW}_r$.

In the companion paper we prove that any coupling whose marginals are approximated by probability measures can be approximated by couplings with respect to the adapted Wasserstein distance (see Proposition 6.2.3 below). We show in the present paper that this result allows us to recover the existence, uniqueness and most importantly the stability of the WOT problem under less regularity assumption on the cost function.

6.1.2 The Martingale Weak Optimal Transport problem

The classical OT theory is not sufficient to solve some major problems raised by the field of mathematical finance, such as robust model-independent pricing. Indeed, Beiglböck, Henry-Labordère and Penkner [23] showed in a discrete time setting and Galichon, Henry- Labordère and Touzi [84] in a continuous time setting that one would need an additional martingale constraint to the OT problem in order to get model-free bounds of an option price. This martingale constraint reflects the condition for a financial market to be arbitrage free. This leads to the formulation of the Martingale Optimal Transport (MOT) problem recalled in the companion paper (Chapter 5). With regard to this, it is natural to study a martingale counterpart of the WOT problem.

Let $d\in\mathbb{N}^*$, $C:\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}_+$ and $\mu,\nu\in\mathcal{P}^1(\mathbb{R}^d)$. Then the Weak Martingale Optimal Transport (WMOT) problem consists in the minimisation

$$V_C^M(\mu,\nu):=\inf_{\pi\in\Pi_M(\mu,\nu)}\int_{\mathbb{R}^d}C(x,\pi_x)\,\mu(dx), \tag{WMOT}$$

where $\Pi_M(\mu,\nu)$ denotes the set of martingale couplings between μ and ν, that is

$$\Pi_M(\mu,\nu)=\left\{\pi=\mu\times\pi_x\in\Pi(\mu,\nu)\mid\mu(dx)\text{-almost everywhere, }\int_{\mathbb{R}}y\,\pi_x(dy)=x\right\}.$$

According to Strassen's theorem [183], the existence of a martingale coupling between two probability measures $\mu,\nu\in\mathcal{P}^1(\mathbb{R}^d)$ is equivalent to $\mu\le_{cx}\nu$, where $\le_{cx}$ denotes the convex order. We recall that two finite positive measures μ,ν on $\mathbb{R}^d$ with finite first moment are said to be in the convex order iff we have

$$\int_{\mathbb{R}^d}f(x)\,\mu(dx)\le\int_{\mathbb{R}^d}f(y)\,\nu(dy),$$

for every convex function $f:\mathbb{R}^d\to\mathbb{R}$. Note that by evaluating this inequality for the constant function equal to 1, the identity function and their opposites, we have that μ and ν have equal mass and satisfy $\int_{\mathbb{R}^d}x\,\mu(dx)=\int_{\mathbb{R}^d}y\,\nu(dy)$. For a measurable map $c:X\times Y\to\mathbb{R}_+$, the WMOT problem with the cost function $C:(x,p)\mapsto\int_Y c(x,y)\,p(dy)$ amounts to the MOT problem already discussed in the companion paper (Chapter 5).

The main result of the companion paper is that any martingale coupling whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance, see Theorem 6.2.4 below. Similarly as for the WOT problem, we prove in the present paper thanks to the latter theorem the existence, the uniqueness and most importantly the stability of the WMOT problem under reasonable regularity assumptions on the cost function.

In particular we recover the stability of the MOT problem proved by Backhoff-Veraguas and Pammer [20]. To do so, they used the tool of martingale C-monotonicity by proving that it was a stable necessary optimality criterion. However the question remained open whether any martingale coupling satisfying this condition is optimal. We show here that it is indeed the case under mild regularity assumptions on the cost function.

6.1.3 Outline

We state in Section 6.2 the main results of the present paper, namely the stability of the WOT and the WMOT problems, the sufficient optimality criterion of martingale C-monotonicity for the WMOT problem, the Lipschitz continuity of the so called Wasserstein projections and the convergence in an extended space of martingale couplings. We also connect this work with an application on the surreplication bound for VIX futures. Those results have their own devoted section. Section 6.3 consists of the unified proof of the stability of the WOT and the WMOT problems. Section 6.4 consists in showing that martingale C-monotonicity is sufficient for optimality for the WMOT problem. In Section 6.5 we establish auxiliary results which help us see that the Wasserstein projections are Lipschitz continuous in dimension 1. In Section 6.6 we prove a result similar to the main theorem of the companion paper, namely the convergence in an extended space of martingale couplings. Finally Section 6.7 is an appendix which gathers the proofs of useful lemmas.

6.2 Main results

6.2.1 An extension of the weak and adapted topologies

For $r \geq 1$, the Wasserstein distance $\mathcal{W}_r$ is widely used to measure the distance between two probability measures with finite r-th moment. In order to measure the distance between two couplings, one could also use the stronger adapted Wasserstein distance for reasons discussed above. Despite being very handy, those distances sometimes lack topological convenience. For example, the $\mathcal{W}_r$-balls $\{p \in \mathcal{P}^r(X) \mid \mathcal{W}_r(p, \delta_{x_0}) \leq R\}$, $R > 0$, are not compact for the $\mathcal{W}_r$-distance topology. This observation is not without consequences since it stood in the way of our proof that martingale C-monotonicity is a sufficient optimality criterion for the WMOT problem (see Section 6.2.4 below).

In order to overcome that hurdle, we choose in the present paper to work in a finer topology which benefits of more convenient and flexible properties. We give the definition here as well as some insight to understand its basic properties. All proofs and technical details are deferred to Section 6.7.2 below.

Definition 6.2.1. Let $f\colon X \to [1, +\infty)$ be continuous. We consider the space

$$\mathcal{P}_f(X) = \{p \in \mathcal{P}(X) \mid p(f) < +\infty\}.$$

We equip $\mathcal{P}_f(X)$ with the topology induced by the following convergence: a sequence $(p_k)_{k\in\mathbb{N}} \in \mathcal{P}_f(X)^{\mathbb{N}}$ converges in $\mathcal{P}_f(X)$ to p iff one of the two following assertions is satisfied:

(i) $p_k \underset{k\to+\infty}{\longrightarrow} p$ in $\mathcal{P}(X)$ and $p_k(f) \underset{k\to+\infty}{\to} p(f)$.

(ii) $p_k(h) \underset{k\to+\infty}{\longrightarrow} p(h)$ for all $h \in \Phi_f(X) := \{h \in \mathcal{C}(X) \mid \exists \alpha > 0,\ \forall x \in X,\ |h(x)| \le \alpha f(x)\}$.

Unless explicitly stated otherwise, $\mathcal{P}(X)$ is endowed with the weak convergence topology; for $r \ge 1$, $\mathcal{P}^r(X)$ is endowed with the $\mathcal{W}_r$-distance topology; for $f : X \to [1, +\infty)$ continuous, $\mathcal{P}_f(X)$ is endowed with the topology induced by the convergence described in Definition 6.2.1. When f is the map $x \mapsto 1 + d_X^r(x, x_0)$, then $\mathcal{P}_f(X) = \mathcal{P}^r(X)$ and the two topologies match. Hence the reader who is not willing to consider this extension may completely disregard it and consistently view $\mathcal{P}_f(X)$ as the usual Wasserstein space $\mathcal{P}^r(X)$.

We will mainly address convergences of probability measures in terms of topology. However it will sometimes prove useful to consider the metric $\overline{\mathcal{W}}_f$ defined on $\mathcal{P}_f(X)$ by

$$\forall p, q \in \mathcal{P}_f(X), \quad \overline{\mathcal{W}}_f(p, q) := \sup_{\substack{h\colon X\to[-1,1],\\ h \text{ is 1-Lipschitz}}} (p(fh) - q(fh)), \tag{6.2.1}$$

which is a complete metric compatible with the topology on $\mathcal{P}_f(X)$.

A continuous function $g : Y \to [1, +\infty)$ can naturally be lifted to a continuous function $\hat{g} : \mathcal{P}_g(Y) \to [1, +\infty)$ by setting

$$\forall p \in \mathcal{P}_f(Y), \quad \hat{g}(p) = p(g). \tag{6.2.2}$$

We will often consider probability measures on $\mathcal{P}(Y)$. A very convenient aspect of the extended topology is that the spaces $\mathcal{P}_{\hat{g}}(\mathcal{P}(Y))$ and $\mathcal{P}_{\hat{g}}(\mathcal{P}_g(Y))$ and their respective topologies, a priori different, are actually equal. If moreover $\mathcal{P}_g(Y)$ is endowed with the metric $\overline{\mathcal{W}}_g$, then those topological spaces are also equal to $\mathcal{P}^1(\mathcal{P}_g(Y))$, whose definition is given by (6.1.1) with $(1, \mathcal{P}_g(Y), \overline{\mathcal{W}}_g)$ replacing (r, X, d_X). Therefore one can freely switch between the topological spaces $\mathcal{P}_{\hat{g}}(\mathcal{P}(Y))$, $\mathcal{P}_{\hat{g}}(\mathcal{P}_g(Y))$ and $\mathcal{P}^1(\mathcal{P}_g(Y))$.

It is also possible to extend the adapted weak topology, in the spirit of (6.1.3). Recall the map J defined by (6.1.4) which embeds $\mathcal{P}(X \times Y)$ into $\mathcal{P}(X \times \mathcal{P}(Y))$. For two real-valued functions f and g respectively defined on X and Y, we denote by $f \oplus g$ the map $X \times Y \ni (x, y) \mapsto f(x) + g(y)$.

Definition 6.2.2. Let $f : X \to [1, +\infty)$ and $g : Y \to [1, +\infty)$ be continuous. For $k \in \mathbb{N}$, let $\mu^k, \mu \in \mathcal{P}_f(X)$, $\nu^k, \nu \in \mathcal{P}_g(Y)$, $\pi^k \in \Pi(\mu^k, \nu^k)$ and $\pi \in \Pi(\mu, \nu)$. We say that $(\pi^k)_{k\in\mathbb{N}}$ converges in $\mathcal{AW}_{f\oplus\hat{g}}$ to π if one of the two following equivalent assertions is satisfied:

(i) $J(\pi^k) \underset{k\to+\infty}{\longrightarrow} J(\pi)$ in $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y))$.

(ii) $J(\pi^k) \underset{k\to+\infty}{\longrightarrow} J(\pi)$ in $\mathcal{P}(X \times \mathcal{P}(Y))$, $\mu^k(f) \underset{k\to+\infty}{\longrightarrow} \mu(f)$ and $\nu^k(g) \underset{k\to+\infty}{\longrightarrow} \nu(g)$.

There also holds the convenient fact that $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y))$ and $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$ and their respective topologies are equal, hence we can rephrase (i) as $J(\pi^k) \underset{k\to+\infty}{\longrightarrow} J(\pi)$ in $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$. When f and g are respectively the maps $x \mapsto 1 + d_X^r(x, x_0)$ and $y \mapsto 1 + d_Y^r(y, y_0)$, then $(\pi^k)_{k\in\mathbb{N}}$ converges in $\mathcal{AW}_{f\oplus\hat{g}}$ to π iff it converges in $\mathcal{AW}_r$. Once again, the reader may skip this extension and consider as he wishes that convergences in $\mathcal{AW}_{f\oplus\hat{g}}$ mean convergences in $\mathcal{AW}_r$.

6.2.2 Stability

The proof of the stability of the WOT problem relies on the following extension to the finer topology of the approximation of couplings on the line in the weak adapted topology proved in the companion paper (Proposition 5.2.3). This extension is an easy consequence of the equivalence stated in Definition 6.2.2.

Proposition 6.2.3. *Let $f\colon X \to [1, +\infty)$ and $g\colon Y \to [1, +\infty)$ be continuous. Let $\mu^k \in \mathcal{P}_f(\mathbb{R})$, $\nu^k \in \mathcal{P}_g(\mathbb{R})$, $k \in \mathbb{N}$, respectively converge to μ and ν in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$ respectively. Then there is for any $\pi \in \Pi(\mu, \nu)$ a sequence of couplings $\pi^k \in \Pi(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_{f\oplus\hat{g}}$.*

In the martingale setting, we recall the main theorem of the companion paper, namely that any martingale couplings whose marginals are approximated by probability measures in the convex order can be approximated by martingale couplings with respect to the adapted Wasserstein distance. We state it in the setting of our extended topology, which is also a direct consequence of the equivalence stated in Definition 6.2.2. For $r \geq 1$, we denote by $\mathcal{F}^r(X)$ the set of continuous functions $f : X \to [1, +\infty)$ which dominate $x \mapsto 1 + d_X^r(x, x_0)$, that is

$$\forall x \in X, \quad f(x) \geq 1 + d_X^r(x, x_0). \tag{6.2.3}$$

Theorem 6.2.4. *Let $f \in \mathcal{F}^1(\mathbb{R})$ and $g \in \mathcal{F}^1(\mathbb{R})$. Let $\mu^k \in \mathcal{P}_f(\mathbb{R})$ and $\nu^k \in \mathcal{P}_g(\mathbb{R})$, $k \in \mathbb{N}$, be in convex order and respectively converge to μ and ν in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$. Let $\pi \in \Pi_M(\mu, \nu)$. Then there exists a sequence of martingale couplings $\pi^k \in \Pi_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$ converging to π in $\mathcal{AW}_{f\oplus\hat{g}}$.*

We recall that a sequence $(\mu^k)_{k\in\mathbb{N}}$ of probability measures on X is said to converge strongly to some $\mu \in \mathcal{P}(X)$ iff for any measurable subset $A \subset X$, $\mu^k(A)$ converges to $\mu(A)$ as k goes to $+\infty$.

Theorem 6.2.5. *Let $f : X \to [1, +\infty)$ and $g : Y \to [1, +\infty)$ be continuous. Let X and Y be Polish spaces, $C : X \times \mathcal{P}_g(Y) \to \mathbb{R}$ be convex in the second argument, lower semicontinuous and such that there exists a constant $K > 0$ which satisfies for all $(x, p) \in X \times \mathcal{P}_g(Y)$*

$$|C(x, p)| \leq K \left(f(x) + \int_Y g(y)\, p(dy) \right). \tag{6.2.4}$$

For $k \in \mathbb{N}$, let $\mu^k \in \mathcal{P}_f(X)$ and $\nu^k \in \mathcal{P}_g(Y)$ converge in $\mathcal{P}_f(X)$ and $\mathcal{P}_g(Y)$ as $k \to +\infty$ to μ and ν respectively. Suppose moreover that one of the following holds true:

(A) C is continuous.

(B) C is continuous in the second argument and μ^k converges strongly to μ as $k \to +\infty$.

Then

(a) Existence: there exists $\pi^* \in \Pi(\mu, \nu)$ which minimises $V_C(\mu, \nu)$.

(b) Stability: we have

$$V_C(\mu^k, \nu^k) \underset{k\to+\infty}{\longrightarrow} V_C(\mu, \nu).$$

Moreover, for $k \in \mathbb{N}$ let $\pi^{k,*} \in \Pi(\mu^k, \nu^k)$ be a minimiser of $V_C(\mu^k, \nu^k)$. Then any accumulation point of $(\pi^{k,*})_{k\in\mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C(\mu, \nu)$. If the latter has a unique minimiser π^*, then

$$\pi^{k,*} \underset{k\to+\infty}{\longrightarrow} \pi^* \quad \text{in } \mathcal{P}_{f\oplus g}(X \times Y). \tag{6.2.5}$$

(c) Uniqueness: if C is strictly convex in the second argument, then the minimisers are unique and the convergence (6.2.5) *holds in $\mathcal{AW}_{f\oplus\hat{g}}$.*

Note that the WMOT problem is in fact a particular case of the WOT problem. Indeed, the resolution of the WMOT problem between $\mu, \nu \in \mathcal{P}^1(\mathbb{R}^d)$ for a cost function $C : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ amounts to the resolution of the WOT problem between the same marginals and the cost function

$$\tilde{C} : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \ni (x, p) \mapsto \begin{cases} C(x,p) & \text{if } \int_Y y\, p(dy) = x \\ +\infty & \text{else} \end{cases} \in \mathbb{R}_+ \cup \{+\infty\}.$$

Since infinite-valued cost functions are not admissible in the setting of Theorem 6.2.5, the case of the stability of the WMOT problem requires its own statement.

Theorem 6.2.6. *Let $f \in \mathcal{F}^1(\mathbb{R}^d)$ and $g \in \mathcal{F}^1(\mathbb{R}^d)$. Let $C : \mathbb{R}^d \times \mathcal{P}_g(\mathbb{R}^d) \to \mathbb{R}$ be convex in the second argument, lower semicontinuous and such that there exists a constant $K > 0$ which satisfies for all $(x, p) \in \mathbb{R}^d \times \mathcal{P}_g(\mathbb{R}^d)$*

$$|C(x, p)| \leq K \left(f(x) + \int_Y g(y)\, p(dy) \right). \tag{6.2.6}$$

For $k \in \mathbb{N}$, let $\mu^k \in \mathcal{P}_f(\mathbb{R}^d)$ and $\nu^k \in \mathcal{P}_g(\mathbb{R}^d)$ be such that $\mu^k \leq_{cx} \nu^k$ and μ^k, resp. ν^k, converges to μ in $\mathcal{P}_f(\mathbb{R}^d)$, resp. ν in $\mathcal{P}_g(\mathbb{R}^d)$ as $k \to +\infty$. Suppose moreover that one of the following holds true:

(A) C is continuous.

(B) C is continuous in the second argument and μ^k converges strongly to μ as $k \to +\infty$.

Then

(a') Existence: there exists $\pi^* \in \Pi_M(\mu, \nu)$ which minimises $V_C^M(\mu, \nu)$.

(b') Stability in dimension 1: when $d = 1$ there holds

$$V_C^M(\mu^k, \nu^k) \underset{k\to+\infty}{\longrightarrow} V_C^M(\mu, \nu). \tag{6.2.7}$$

(c') Stability: suppose that (6.2.7) holds. For $k \in \mathbb{N}$ let $\pi^{k,*} \in \Pi_M(\mu^k, \nu^k)$ be a minimiser of $V_C^M(\mu^k, \nu^k)$. Then any accumulation point of $(\pi^{k,*})_{k\in\mathbb{N}}$ for the weak convergence topology is a minimiser of $V_C^M(\mu, \nu)$. If the latter has a unique minimiser π^*, then

$$\pi^{k,*} \underset{k\to+\infty}{\longrightarrow} \pi^* \quad \text{in } \mathcal{P}_{f\oplus g}(\mathbb{R}^d \times \mathbb{R}^d). \tag{6.2.8}$$

(d') Uniqueness: if C is strictly convex in the second argument, then the minimisers are unique and the convergence (6.2.8) holds in $\mathcal{AW}_{f\oplus\hat{g}}$.

Remark 6.2.7. We actually show that (a'), (c') and (d') are still valid when there exist two Polish subspaces X and Y of $\mathbb{R}^d$ such that f is defined on X, g on Y, C on $X \times \mathcal{P}_g(Y)$, (6.2.6) holds for all $(x, p) \in X \times \mathcal{P}_g(Y)$, $\mu^k \in \mathcal{P}_f(X)$ converges in $\mathcal{P}_f(X)$ to $\mu \in \mathcal{P}_f(X)$ and $\nu^k \in \mathcal{P}_f(Y)$ converges in $\mathcal{P}_f(Y)$ to $\nu \in \mathcal{P}_f(Y)$ as $k \to +\infty$.

We can exhibit a strong connection betwen the (WMOT) problem and a Benamou-Brenier type formulation of the MOT problem suggested by [17]. In dimension one, this problem consists for two probability measures μ, ν in the convex order in maximising

$$MT(\mu, \nu) := \sup \mathbb{E}\left[\int_0^1 \sigma_t \, dt\right] \tag{MBB}$$

over all filtered probability spaces $(\Omega, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P})$, real-valued $(\mathcal{F}_t)_{t\in[0,1]}$-progressive process $(\sigma_t)_{t\in[0,1]}$ and real-valued $(\mathcal{F}_t)_{t\in[0,1]}$-Brownian motions $(B_t)_{t\in[0,1]}$ such that the process

$$(M_t)_{t\in[0,1]} = \left(M_0 + \int_0^t \sigma_s \, dB_s\right)_{t\in[0,1]}$$

is a continuous martingale which satisfies $M_0 \sim \mu$ and $M_1 \sim \nu$. When the second moment of ν is finite, then (MBB) has a unique maximiser $(M_t^*)_{t\in[0,1]}$ [17, Theorem 1.5] called the stretched Brownian motion from μ to ν, since it is the martingale subject to the constraints $M_0^* \sim \mu$ and $M_1^* \sim \nu$ which correlates the most with the Brownian motion.

Let $C_2 : \mathbb{R}\times\mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ be defined for all $(x, p) \in \mathbb{R}\times\mathcal{P}_2(\mathbb{R})$ by $C_2(x, p) = \mathcal{W}_2^2(p, \mathcal{N}(0, 1))$, where $\mathcal{N}(0, 1)$ denotes the unidimensional standard normal distribution. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ be in the convex order and $V_{C_2}^M(\mu, \nu)$ be the value function given by (WMOT) for the cost function C_2. Let $\pi^* \in \Pi_M(\mu, \nu)$ be optimal for $V_{C_2}^M(\mu, \nu)$ and M^* be the stretched Brownian motion from μ to ν. Then Remark 2.1, Theorem 2.2 and Remark 2.3 from [17] imply that

(a) $MT(\mu,\nu) = \frac{1}{2}\left(1 + \int_{\mathbb{R}} |y|^2\,\nu(dy) - V_{C_2}^M(\mu,\nu)\right)$;

(b) π^* is the joint probability distribution of (M_0^*, M_1^*), and conversely

$$\forall t \in [0,1], \quad M_t^* = \mathbb{E}\left[F^{-1}_{\pi^*_X}(F_{\mathcal{N}(0,1)}(B_1))|X, (B_s)_{0\le s\le t}\right], \tag{6.2.9}$$

where $X \sim \mu$ is a random variable independent of the Brownian motion $(B_t)_{t\in[0,1]}$, and F_η, resp. F_η^{-1} denotes the cumulative distribution function, resp. the quantile function of a probability distribution $\eta \in \mathcal{P}(\mathbb{R})$ (see Section 6.2.5 below for more details).

As a consequence of Theorem 6.2.6, for $r \ge 2$, the stretched Brownian motion between converging marginals in $\mathcal{W}_r$ converges in $\mathcal{AW}_r$ to the stretched Brownian motion between the limits. Because of the martingale structure, we even have an estimate of the $\mathcal{AW}_r$-distance between the joint probability distributions of the initial position and the whole trajectory.

Corollary 6.2.8 (Stability of the unidimensional stretched Brownian motion)**.** *Let $r \ge 2$ and $\mu^k, \nu^k, \mu, \nu \in \mathcal{P}^r(\mathbb{R})$, $k \in \mathbb{N}$ be such that for all $k \in \mathbb{N}$, $\mu^k \le_{cx} \nu^k$ and μ^k, resp. ν^k, converges to μ, resp. ν, in $\mathcal{W}_r$.*

For $k \in \mathbb{N}$, let M^k be the stretched Brownian motion from μ^k to ν^k and M^* be the stretched Brownian motion from μ to ν. Equipping $C([0,1])$ with the supremum distance and denoting by $\mathcal{L}(Z)$ the law of any random variable Z, we have the estimate

$$\mathcal{AW}_r^r\left(\mathcal{L}(M_0^k, (M_t^k)_{t\in[0,1]}), \mathcal{L}(M_0^*, (M_t^*)_{t\in[0,1]})\right) \le \left(\frac{r}{r-1}\right)^r \mathcal{AW}_r^r\left(\mathcal{L}(M_0^k, M_1^k), \mathcal{L}(M_0^*, M_1^*)\right),$$

and the right-hand side vanishes as k goes to $+\infty$.

6.2.3 Stability of the superreplication bound for VIX futures

The Volatility Index (VIX), often referred to as the Fear Index, is a popular measure to determine market sentiment. When investors expect the market to move vigorously, they typically tend to purchase more options, which has an impact on implied volatility levels. The VIX is by definition the implied volatility calculated on a 30 days horizon on the S&P 500. The more the VIX increases, the more demand is expressed for options, which become more expensive. In that case the market is described as volatile. Conversely, a decreasing VIX often means less demand and therefore decreasing option prices, hence the market is perceived as calm.

We consider a financial market composed of two financial assets: the risk-free asset and the S&P 500 $(S_t)_{t\in\{T_1,T_2\}}$, tradable at dates T_1 and $T_2 = T_1 + 30$ days. We suppose known the market price of call options for any strike $K \ge 0$, so that by the Breeden-Litzenberger formula [46] we get the respective probability distributions μ and ν of S_{T_1} and S_{T_2}. We allow trading at time 0 in vanilla options with maturities T_1 and T_2, and trading at time T_1 in the S&P 500 and the forward-starting log-contract, that is the option with payoff $\frac{-2}{T_2-T_1}\ln\frac{S_{T_2}}{S_{T_1}}$ at

T_2. In this setting, Guyon, Menegaux and Nutz [99] derive the model-independent arbitrage-free upper bound for the VIX future expiring at T_1, given by the smallest superreplication price at time 0

$$P_{\text{super}}(\mu,\nu) = \inf\left(\int_{(0,+\infty)} u_1(x)\,\mu(dx) + \int_{(0,+\infty)} u_2(y)\,\nu(dy)\right), \tag{6.2.10}$$

where the infimum is taken over all $(u_1,u_2) \in L^1(\mu)\times L^1(\nu)$ and measurable maps Δ^S, Δ^L such that for all $(x,y,v) \in (0,+\infty)^2 \times [0,+\infty)$,

$$u_1(x) + u_2(y) + \Delta^S(x,v)(y-x) + \Delta^L(x,v)\left(-\frac{2}{T_2-T_1}\ln\frac{y}{x} - v^2\right) - v \tag{6.2.11}$$

is nonnegative. Similarly, the model-independent arbitrage-free lower bound for the VIX future expiring at T_1 is given by the largest subreplication price at time 0

$$P_{\text{sub}}(\mu,\nu) = \sup\left(\int_{(0,+\infty)} u_1(x)\,\mu(dx) + \int_{(0,+\infty)} u_2(y)\,\nu(dy)\right),$$

where the supremum is taken over all $(u_1,u_2) \in L^1(\mu)\times L^1(\nu)$ and measurable maps Δ^S, Δ^L such that for all $(x,y,v) \in (0,+\infty)^2 \times [0,+\infty)$, (6.2.11) is nonpositive.

Note that the primal problem $P_{\text{super}}(\mu,\nu)$ involves in (6.2.11) three variables x, y, s, which stand respectively for the S&P 500 at time T_1, the S&P 500 at time T_2, and the VIX at time T_1. We would then naturally expect the dual formulation to involve three marginals. Strikingly, the dual side of the superreplication of the VIX takes the form of a WMOT problem with 2 marginals only thanks to concavity of the square root, see [99, Proposition 4.10].

Proposition 6.2.9 (Guyon, Menegaux, Nutz, 2017)**.** *Let $0 < T_1 < T_2$ and $f\colon [1,+\infty) \to \mathbb{R}_+$ be given by $f(x) = |\ln(x)| + |x|$. Let $\mu,\nu \in \mathcal{P}_f((0,+\infty))$ be in the convex order, then the dual problem D_{super} consists of*

$$D_{super}(\mu,\nu) = \sup_{\pi\in\Pi_M(\mu,\nu)} \int_{(0,+\infty)} C_{VIX}(x,\pi_x)\,\mu(dx), \tag{6.2.12}$$

when $C_{VIX}\colon (0,+\infty)\times\mathcal{P}_f((0,+\infty)), (x,p) \mapsto \sqrt{-\frac{2}{T_2-T_1}\int_{(0,+\infty)} \ln\left(\frac{y}{x}\right)\,p(dy)}$. The values of $P_{super}(\mu,\nu)$ and $D_{super}(\mu,\nu)$ coincide.

The fact that μ and ν are defined on $(0,+\infty)$ motivated Remark 6.2.7, that is the consideration of the stability of the WMOT problem in the setting of Polish subspaces of $\mathbb{R}^d$. Note that C_{VIX} is indeed an admissible weak transport cost: on $\{(x,p) \in (0,+\infty)\times \mathcal{P}_f((0,+\infty)) \mid \int_{(0,+\infty)} y\,p(dy) = x\}$ it is well-defined and continuous, and the map $p \mapsto C_{VIX}(x,p)$ is concave on $\{p \in \mathcal{P}_f((0,+\infty)) \mid \int_{\mathbb{R}} y\,p(dy) = x\}$ for fixed $x \in (0,+\infty)$. Hence, we can apply Theorem 6.2.6 and find that the robust superreplication bound for VIX futures depends continuously on the marginals:

Corollary 6.2.10. *In the setting of Proposition 6.2.9, let the pairs $\mu^k, \nu^k \in \mathcal{P}_f((0,+\infty))$, $k \in \mathbb{N}$, be in convex order and converge in $\mathcal{P}_f(\mathbb{R})$ to μ and ν respectively. Then there exist maximisers $\pi^{k,*} \in \Pi_M(\mu^k, \nu^k)$,*

$$\lim_{k\to+\infty} D_{super}(\mu^k, \nu^k) = D_{super}(\mu, \nu),$$

and any weak accumulation point of $(\pi^{k,*})_{k\in\mathbb{N}}$ maximises $D_{super}(\mu, \nu)$.

6.2.4 Martingale monotonicity

A remarkable tool in the theory of optimal transport is cyclical monotonicity. It allows to determine optimality of a coupling only by knowing its support. In its spirit the notion of finite optimality was developed in context of martingale optimal transport in [25] and [93].

Definition 6.2.11 (Competitor)**.** Let $\alpha = \mu \times \alpha_x \in \mathcal{P}^1(\mathbb{R} \times \mathbb{R})$. We call $\alpha' = \mu' \times \alpha'_x \in \mathcal{P}^1(\mathbb{R} \times \mathbb{R})$ a competitor of α, if

$$\mu = \mu' \quad \text{and} \quad \int_{\mathbb{R}} y\, \alpha'_x(dy) = \int_{\mathbb{R}} y\, \alpha_x(dy), \quad \mu(dx)\text{-almost everywhere.}$$

Definition 6.2.12 (Finite optimality)**.** Let $c\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a cost function. We say that a Borel set $\Gamma \subset \mathbb{R} \times \mathbb{R}$ is *finitely optimal* for c if for every probability measure $\alpha \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ finitely supported on Γ, we have

$$\int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, \alpha(dx,dy) \le \int_{\mathbb{R}\times\mathbb{R}} c(x,y)\, \alpha'(dx,dy),$$

for every competitor α' of α.

Under the assumption that c is continuous and sufficiently integrable, there was shown in [25, Lemma A.2] and [93, Theorem 1.3] that a martingale coupling is optimal if it is concentrated on a finitely optimal set.

Recently the notion of martingale C-monotonicity, c.f. [20], was introduced for Weak Martingale Optimal Transport (WMOT), which was therein used to show stability of the Martingale Optimal Transport problem.

Definition 6.2.13 (Martingale C-monotonicity)**.** We say that a Borel set $\Gamma \subset \mathbb{R} \times \mathcal{P}^1(\mathbb{R})$ is *martingale C-monotone* iff for any $N \in \mathbb{N}$, any collection $(x_1, p_1), \ldots, (x_N, p_N) \in \Gamma$ and $q_1, \ldots, q_N \in \mathcal{P}^1(\mathbb{R})$ such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ and $\int_{\mathbb{R}} y\, p_i(dy) = \int_{\mathbb{R}} y\, q_i(dy)$, we have

$$\sum_{i=1}^N C(x_i, p_i) \le \sum_{i=1}^N C(x_i, q_i).$$

So far, it was known that martingale C-monotonicity is a necessary optimality criterion in the following sense, c.f. [20, Theorem 3.4]: let $\pi^* \in \Pi_M(\mu, \nu)$ be a martingale coupling which minimises (WMOT), then $J(\pi^*)$ is concentrated on a martingale C-monotone set. This means explicitly that there is a martingale C-monotone set Γ with

$$(x, \pi_x) \in \Gamma \quad \text{for } \mu(dx)\text{-almost every } x. \tag{6.2.13}$$

Remark 6.2.14. Conversely, if $\pi \in \Pi_M(\mu,\nu)$ is a finitely supported coupling of the form $\frac{1}{N}\sum_{i=1}^N \delta_{x_i}(dx)\,p_i(dy)$ for $x_1 < \cdots < x_n \in \mathbb{R}$ and $p_1, \cdots, p_N \in \mathcal{P}^1(\mathbb{R})$ and satisfies (6.2.13), then it is optimal. Indeed, in that case $(x_1,p_1), \cdots, (x_N,p_N) \in \Gamma$ and any martingale coupling $\pi' \in \Pi_M(\mu,\nu)$ is of the form $\pi'(dx,dy) = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}(dx)\,q_i(dy)$, where $q_1, \cdots, q_N \in \mathcal{P}^1(\mathbb{R})$ are such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ and for all $i \in \{1, \cdots, N\}$, $\int_{\mathbb{R}} y\,p_i(dy) = x_i = \int_{\mathbb{R}} y\,q_i(dy)$. By definition of martingale C-monotonicity, we get

$$\int_{\mathbb{R}\times\mathbb{R}} C(x,\pi_x)\,\mu(dx) = \frac{1}{N}\sum_{i=1}^N C(x_i,p_i) \le \frac{1}{N}\sum_{i=1}^N C(x_i,q_i) = \int_{\mathbb{R}^2} C(x,\pi'_x)\,\mu(dx),$$

hence π is optimal.

However, the question remained open if any martingale coupling satisfying (6.2.13) is optimal. The stability result, Theorem 6.2.6, allows us to confirm that this is indeed the case.

Theorem 6.2.15 (Sufficiency)**.** *Let $f\colon \mathbb{R} \to [1,+\infty)$ and $g\colon \mathbb{R} \to [1,+\infty)$ be continuous. Let $\mu \in \mathcal{P}_f(\mathbb{R})$ and $\nu \in \mathcal{P}_g(\mathbb{R})$ be in convex order, and $C\colon \mathbb{R}\times\mathcal{P}_g(\mathbb{R}) \to \mathbb{R}$ be a measurable cost function, continuous in the second argument and such that there exists a constant $K > 0$ which satisfies*

$$\forall (x,p) \in \mathbb{R}\times\mathcal{P}_g(\mathbb{R}), \quad C(x,p) \le K\left(f(x) + \int_{\mathbb{R}} g(y)\,p(dy)\right),$$

Let Γ be martingale C-monotone and $\pi \in \Pi_M(\mu,\nu)$ be such that we have (6.2.13)*. Then π is optimal for* (WMOT)*.*

In turn Theorem 6.2.15 allows us to strengthen [25, Lemma A.2] and [93, Theorem 1.3] by assuming less continuity of the cost function.

Corollary 6.2.16. *Let $f\colon \mathbb{R} \to [1,+\infty)$ and $g\colon \mathbb{R} \to [1,+\infty)$ be continuous. Let $\mu \in \mathcal{P}_f(\mathbb{R})$ and $\nu \in \mathcal{P}_g(\mathbb{R})$ be in convex order, $c\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be measurable and $y \mapsto c(x,y)$ be continuous for all $x \in \mathbb{R}$. Furthermore, let $K > 0$ be a constant such that*

$$c(x,y) \le K(f(x) + g(y)), \quad \forall (x,y) \in \mathbb{R}\times\mathbb{R}.$$

Then $\pi \in \Pi_M(\mu,\nu)$ is finitely optimal if and only if π is optimal for the MOT problem.

6.2.5 On the Wasserstein projection

The continuity of the so called Wasserstein projection is essential in our proof of a desired convergence in a space of extended marginale couplings, see Section 6.2.6 below. It will also prove useful to show that the stretched Brownian motion provides a convenient tool to approximate two probability measures in the convex order with atomless ones still in the convex order, see Section 6.7.4 below. In this context we will work in dimension 1 and rely on the very convenient tools of cumulative distribution functions and quantile functions.

For η a probability distribution on $\mathbb{R}$, we denote by $F_\eta : x \mapsto \eta((-\infty, x])$ its cumulative distribution function, and by $F_\eta^{-1} : (0,1) \to \mathbb{R}$ its quantile function defined for all $u \in (0,1)$ by

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

The following properties are standard results (see for instance Section 2.6 for proofs):

(a) F_η is càdlàg, F_η^{-1} is càglàd ;

(b) For all $(x,u) \in \mathbb{R} \times (0,1)$,

$$F_\eta^{-1}(u) \leq x \iff u \leq F_\eta(x), \tag{6.2.14}$$

which implies

$$F_\eta(x-) < u \leq F_\eta(x) \implies x = F_\eta^{-1}(u), \tag{6.2.15}$$
$$\text{and} \quad F_\eta(F_\eta^{-1}(u)-) \leq u \leq F_\eta(F_\eta^{-1}(u)); \tag{6.2.16}$$

(c) For $\eta(dx)$-almost every $x \in \mathbb{R}$,

$$0 < F_\eta(x), \quad F_\eta(x-) < 1 \quad \text{and} \quad F_\eta^{-1}(F_\eta(x)) = x;$$

(d) The image of the Lebesgue measure on $(0,1)$ by F_η^{-1} is η.

The property (d) is referred to as inverse transform sampling.

For $\mu, \nu \in \mathcal{P}^r(\mathbb{R})$, let $\mathcal{I}(\mu,\nu), \mathcal{J}(\mu,\nu) \in \mathcal{P}^r(\mathbb{R})$ be the probability measures whose respective quantile functions are defined for all $u \in (0,1)$ by

$$F_{\mathcal{I}(\mu,\nu)}^{-1}(u) = F_\mu^{-1}(u) - \partial_- \operatorname{co}(G)(u) \quad \text{and} \quad F_{\mathcal{J}(\mu,\nu)}^{-1}(u) = F_\nu^{-1}(u) + \partial_- \operatorname{co}(G)(u), \tag{6.2.17}$$

where $G : v \mapsto \int_0^v (F_\mu^{-1} - F_\nu^{-1})(u)\,du$, co denotes the convex hull and ∂_- the left-hand derivative. By [4, Theorem 2.6 and Proposition 4.2], $\mathcal{I}(\mu,\nu)$, resp. $\mathcal{J}(\mu,\nu)$, is a projection of μ, resp. ν, on the set of probability measures dominated by ν, resp. dominating μ, in the convex order:

$$\mathcal{W}_r(\mu, \mathcal{I}(\mu,\nu)) = \inf\{\mathcal{W}_r(\mu,\eta) \mid \eta \in \mathcal{P}^r(\mathbb{R}),\ \eta \leq_{cx} \nu\}$$
$$\text{and} \quad \mathcal{W}_r(\mathcal{J}(\mu,\nu), \nu) = \inf\{\mathcal{W}_r(\eta,\nu) \mid \eta \in \mathcal{P}^r(\mathbb{R}),\ \mu \leq_{cx} \eta\}.$$

Notice that for $r > 1$, the projections $\mathcal{I}(\mu,\nu)$ and $\mathcal{J}(\mu,\nu)$ are unique and do not depend on r. For $r = 1$, those projections are not unique and the maps $\mathcal{I}$ and $\mathcal{J}$ provide a measurable way of choosing those projections.

Remark 6.2.17. By (6.2.17), the map $u \mapsto F_{\mathcal{I}(\mu,\nu)}^{-1}(u) - F_\mu^{-1}(u)$ is the opposite of the left-hand derivative of a convex map, hence it is nonincreasing. Therefore, for all $u \in (0,1)$, its mean on $(0,u)$ is bigger than or equal to its mean on $(0,1)$, that is

$$\frac{1}{u}\int_0^u (F_{\mathcal{I}(\mu,\nu)}^{-1}(v) - F_\mu^{-1}(v))\,dv \geq \int_0^1 (F_{\mathcal{I}(\mu,\nu)}^{-1}(v) - F_\mu^{-1}(v))\,dv = \int_{\mathbb{R}} y\,\nu(dy) - \int_{\mathbb{R}} x\,\mu(dx),$$

where the last equality comes from the inverse transform sampling and the fact that $\mathcal{I}(\mu,\nu)$ and ν have the same barycentre, consequence of $\mathcal{I}(\mu,\nu) \leq_{cx} \nu$. If $\int_{\mathbb{R}} y\,\nu(dy) = \int_{\mathbb{R}} x\,\mu(dx)$, then $\int_0^u F^{-1}_{\mathcal{I}(\mu,\nu)}(v)\,dv \geq \int_0^u F^{-1}_\mu(v)\,dv$ for all $u \in (0,1)$, and [169, Theorem 3.A.5] implies that $\mathcal{I}(\mu,\nu) \leq_{cx} \mu$, and therefore $\mathcal{I}(\mu,\nu) \leq_{cx} \mu \wedge_{cx} \nu$. In general, there is no equality.

Similarly, we have that $\mu \leq_{cx} \mathcal{J}(\mu,\nu)$, and if $\int_{\mathbb{R}} y\,\nu(dy) = \int_{\mathbb{R}} x\,\mu(dx)$ then $\nu \leq_{cx} \mathcal{J}(\mu,\nu)$, hence $\mu \vee_{cx} \nu \leq \mathcal{J}(\mu,\nu)$.

Theorem 6.2.18. *The maps $\mathcal{I}$ and $\mathcal{J}$ are Lipschitz continuous with*

$$\mathcal{W}_r(\mathcal{I}(\mu,\nu),\mathcal{I}(\mu',\nu')) \leq 2\mathcal{W}_r(\mu,\mu') + \mathcal{W}_r(\nu,\nu') \tag{6.2.18}$$

$$\textit{and} \quad \mathcal{W}_r(\mathcal{J}(\mu,\nu),\mathcal{J}(\mu',\nu')) \leq \mathcal{W}_r(\mu,\mu') + 2\mathcal{W}_r(\nu,\nu') \tag{6.2.19}$$

for all $\mu,\nu,\mu',\nu' \in \mathcal{P}^r(\mathbb{R})$.

Remark 6.2.19. The inequalities (6.2.18) and (6.2.19) respectively generalise [4, Proposition 3.1 and Proposition 4.3] which state that when $\mu \leq_{cx} \nu$ so that $\mathcal{I}(\mu,\nu) = \mu$ and $\mathcal{J}(\mu,\nu) = \nu$,

$$\begin{aligned} &\mathcal{W}_r(\mu,\mathcal{I}(\mu',\nu')) \leq 2\mathcal{W}_r(\mu,\mu') + \mathcal{W}_r(\nu,\nu') \\ \text{and} \quad &\mathcal{W}_r(\nu,\mathcal{J}(\mu',\nu')) \leq \mathcal{W}_r(\mu,\mu') + 2\mathcal{W}_r(\nu,\nu'). \end{aligned}$$

6.2.6 Convergence in an extended space of martingale couplings

The adapted weak topology is defined as the initial topology under the map J, which is given by (6.1.4) and embeds $\mathcal{P}(X \times Y)$ into $\mathcal{P}(X \times \mathcal{P}(Y))$. Conversely, it is widely known that we can associate to a probability measure $P \in \mathcal{P}(\mathcal{P}(Y))$ its intensity $I(P)(dy) = \int_{p\in\mathcal{P}(Y)} p(dy)\,P(dp) \in \mathcal{P}(Y)$. For the extended space $\mathcal{P}(X \times \mathcal{P}(Y))$ we naturally define the extended intensity $\hat{I}$ by

$$\hat{I} : \mathcal{P}(X \times \mathcal{P}(Y)) \ni P \mapsto \int_{p\in\mathcal{P}(Y)} p(dy)\,P(dx,dp) \in \mathcal{P}(X \times Y), \tag{6.2.20}$$

which associates to each $P \in \mathcal{P}(X \times \mathcal{P}(Y))$ a coupling $\hat{I}(P) \in \mathcal{P}(X \times Y)$. We note that $\hat{I}$ is the left-inverse of J.

For $(\mu,\nu) \in \mathcal{P}(X)\times\mathcal{P}(Y)$, we define the set of extended couplings $\Lambda(\mu,\nu)$ between μ and ν as the set of probability measures on $\mathcal{P}(X \times \mathcal{P}(Y))$ whose extended intensity is a coupling between μ and ν, that is

$$\Lambda(\mu,\nu) = \left\{ P = \mu \times P_x \in \mathcal{P}(X \times \mathcal{P}(Y)) \mid \int_{(x,p)\in X\times\mathcal{P}(Y)} p(dy)\,P(dx,dp) = \nu(dy) \right\}.$$

If $f\colon X \to \mathbb{R}^+$ and $g\colon Y \to \mathbb{R}^+$ are measurable functions, then any $P \in \Lambda(\mu,\nu)$ satisfies

$$\begin{aligned} &\int_{X\times\mathcal{P}(Y)} f(x)\,P(dx,dp) = \int_X f(x)\,\mu(dx), \\ \text{and} \quad &\int_{X\times\mathcal{P}(Y)} \int_Y g(y)\,p(dy)\,P(dx,dp) = \int_Y g(y)\,\nu(dy). \end{aligned} \tag{6.2.21}$$

For $\mu, \nu \in \mathcal{P}^1(\mathbb{R}^d)$, we define the martingale counterpart $\Lambda_M(\mu, \nu)$ of $\Lambda(\mu, \nu)$ as the set of probability measures on $\mathcal{P}^1(\mathbb{R}^d \times \mathcal{P}^1(\mathbb{R}^d))$, whose extended intensity is a martingale coupling between μ and ν, that is

$$\Lambda_M(\mu, \nu) = \left\{ P \in \Lambda(\mu, \nu) \mid \int_{\mathbb{R}^d} y\, p(dy) = x,\ \ P(dx, dp)\text{-almost everywhere} \right\}. \tag{6.2.22}$$

The following theorem may be seen as a complement of Theorem 6.2.4.

Theorem 6.2.20. *Let $f, g \in \mathcal{F}^1(\mathbb{R})$, and for $k \in \mathbb{N}$, let $\mu^k \in \mathcal{P}_f(\mathbb{R})$ and $\nu^k \in \mathcal{P}_g(\mathbb{R})$ be in convex order and respectively converge to μ and ν in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$. Let $P \in \Lambda_M(\mu, \nu)$. Then there exists a sequence $P^k \in \Lambda_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to P in $\mathcal{P}_{f \oplus \hat{g}}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$.*

Remark 6.2.21. In the light of Theorem 6.2.4, we can add in the statement of this theorem that when $P \in J(\Lambda_M(\mu, \nu))$, then P^k can be chosen in $J(\Lambda_M(\mu^k, \nu^k))$.

6.3 Stability

This section is devoted to the proof of Theorem 6.2.5 and Theorem 6.2.6 about the stability of (WOT) and (WMOT), and the corollary on the stability of the stretched Brownian motion in dimension one.

Proof of Theorem 6.2.5 and Theorem 6.2.6. First, we prove (a) and (a'). Let $\hat{\Pi}(\mu, \nu) = \Pi(\mu, \nu)$ and $\hat{V}_C(\mu, \nu) = V_C(\mu, \nu)$ in the setting of Theorem 6.2.5, and $\hat{\Pi}(\mu, \nu) = \Pi_M(\mu, \nu)$ and $\hat{V}_C(\mu, \nu) = V_C^M(\mu, \nu)$ in the setting of Theorem 6.2.6.

Let $(\pi^n)_{n \in \mathbb{N}} \in \hat{\Pi}(\mu, \nu)^{\mathbb{N}}$ be such that $\int_X C(x, \pi^n_x)\, \mu(dx)$ converges to $\hat{V}_C(\mu, \nu)$ as $n \to +\infty$. By tightness of μ and ν we deduce the existence of a subsequence $(\pi^{n_l})_{l \in \mathbb{N}}$ of $(\pi^n)_{n \in \mathbb{N}}$ which converges to some $\pi^* \in \hat{\Pi}(\mu, \nu)$ with respect to the weak convergence topology and therefore the topology of $\mathcal{P}_{f \oplus g}(X \times Y)$ since $\pi^{n_l}(f \oplus g) = \mu(f) + \nu(f) = \pi^*(f \oplus g)$ for all $l \in \mathbb{N}$. By Proposition 6.7.9 (b) below we then have

$$\hat{V}_C(\mu, \nu) \leq \int_X C(x, \pi^*_x)\, \mu(dx) \leq \liminf_{l \to +\infty} \int_X C(x, \pi^{n_l}_x)\, \mu(dx) = \hat{V}_C(\mu, \nu),$$

which shows that π^* is a minimiser for $\hat{V}_C(\mu, \nu)$ and proves (a) and (a').

We now show that the convergence

$$\hat{V}_C(\mu^k, \nu^k) \underset{k \to +\infty}{\longrightarrow} \hat{V}_C(\mu, \nu) \tag{6.3.1}$$

holds in the setting of Theorem 6.2.5, and in the setting of Theorem 6.2.6 as soon as $d = 1$, which will prove (b') and the first part of (b). Let π^* be a minimiser of $\hat{V}_C(\mu, \nu)$. By Proposition 6.2.3 in the setting of Theorem 6.2.5 and Theorem 6.2.4 in the setting of Theorem 6.2.6 if $d = 1$, there exists a sequence $\pi^k \in \hat{\Pi}(\mu^k, \nu^k)$, $k \in \mathbb{N}$, which converges to π^* in $\mathcal{AW}_{f \oplus \hat{g}}$, which is equivalent to $J(\pi^k)$ converging to $J(\pi^*)$ in $\mathcal{P}_{f \oplus \hat{g}}(X \times \mathcal{P}(Y))$.

Under Assumption (A), we then have by Lemma 6.7.13 (b) that

$$\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^k)(dx,dp) \underset{k\to+\infty}{\longrightarrow} \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^*)(dx,dp). \tag{6.3.2}$$

Under Assumption (B), the strong convergence of $(\mu^k)_{k\in\mathbb{N}}$ to μ and the weak convergence of $(J(\pi^k))_{k\in\mathbb{N}}$ to $J(\pi^*)$ imply by Lemma 6.7.12 (b) that $(J(\pi^k))_{k\in\mathbb{N}}$ converges stably to $J(\pi^*)$, hence (6.3.2) still holds by Lemma 6.7.13 (d).

Using (6.3.2) for the second equality, we then have

$$\begin{aligned}\limsup_{k\to+\infty} \hat{V}_C(\mu^k,\nu^k) &\le \limsup_{k\to+\infty}\int_X C(x,\pi^k_x)\,\mu^k(dx)\\ &= \limsup_{k\to+\infty}\int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^k)(dx,dp)\\ &= \int_{X\times\mathcal{P}_g(Y)} C(x,p)\,J(\pi^*)(dx,dp)\\ &= \hat{V}_C(\mu,\nu).\end{aligned} \tag{6.3.3}$$

Let $(\hat{V}_C(\mu^{k_l},\nu^{k_l}))_{l\in\mathbb{N}}$ be a subsequence of $(\hat{V}_C(\mu^k,\nu^k))_{k\in\mathbb{N}}$ converging to $\liminf_{k\to+\infty}\hat{V}_C(\mu^k,\nu^k)$. Let $\tilde{\pi}\in\hat{\Pi}(\mu,\nu)$ be an accumulation point of $(\pi^{k_l,*})_{l\in\mathbb{N}}$ with respect to the weak convergence topology, which exists by tightness of the marginals. Note that in the martingale case, the fact that $\tilde{\pi}$ is a martingale coupling is guaranteed by the $\mathcal{W}_1$-convergence of the marginals. Then by Proposition 6.7.9 (b) below, we find that

$$\liminf_{k\to+\infty}\hat{V}_C(\mu^k,\nu^k) = \lim_{l\to+\infty}\int_X C(x,\pi^{k_l,*}_x)\,\mu^{k_l}(dx) \ge \int_X C(x,\tilde{\pi}_x)\,\mu(dx) \ge \hat{V}_C(\mu,\nu).$$

With (6.3.3), we conclude that $\lim_{k\to+\infty}\hat{V}_C(\mu^k,\nu^k)=\hat{V}_C(\mu,\nu)$, which proves (6.3.1).

Let us now prove (c') and complete the proof of (b), assuming if we are in the setting of Theorem 6.2.6 with $d>1$ that (6.3.1) holds. For $k\in\mathbb{N}$, let $\pi^{k,*}\in\hat{\Pi}(\mu^k,\nu^k)$ be a minimiser of $\hat{V}_C(\mu^k,\nu^k)$. For any subsequence $(\pi^{k_j,*})_{j\in\mathbb{N}}$ of $(\pi^{k,*})_{k\in\mathbb{N}}$ converging weakly to some $\tilde{\pi}$, Proposition 6.7.9 (b) below ensures that

$$\hat{V}_C(\mu,\nu) = \lim_{j\to+\infty}\hat{V}_C(\mu^{k_j},\nu^{k_j}) = \lim_{j\to+\infty}\int_X C(x,\pi^{k_j,*}_x)\,\mu(dx) \ge \int_X C(x,\tilde{\pi}_x)\,\mu(dx) \ge \hat{V}_C(\mu,\nu),$$

so $\tilde{\pi}$ is a minimiser of $\hat{V}_C(\mu,\nu)$. In particular if $\hat{V}_C(\mu,\nu)$ has a unique minimiser π^*, it is the unique accumulation point with respect to the weak convergence topology of the tight sequence $(\pi^{k,*})_{k\in\mathbb{N}}$, which therefore converges to π^* weakly and even in $\mathcal{P}_{f\oplus g}(X\times Y)$ since its marginals converge in $\mathcal{P}_f(X)$ and $\mathcal{P}_g(Y)$ respectively. Hence (b) and (c') are proved.

Finally, let us show (c) and (d'). The strict convexity of $C(x,\cdot)$ for all $x\in X$ yields uniqueness of the minimisers. Indeed when $\pi,\tilde{\pi}\in\hat{\Pi}(\mu,\nu)$ then $\frac{1}{2}(\pi+\tilde{\pi})\in\hat{\Pi}(\mu,\nu)$. When, moreover, $\pi\neq\tilde{\pi}$, then $\mu(\{x\in X\mid \pi_x\neq\tilde{\pi}_x\})>0$ and since $C(x,\frac{1}{2}(\pi_x+\tilde{\pi}_x))\le\frac{1}{2}(C(x,\pi_x)+C(x,\tilde{\pi}_x))$ with strict inequality when $\pi_x\neq\tilde{\pi}_x$,

$$\int_X C\left(x,\frac{\pi_x+\tilde{\pi}_x}{2}\right)\mu(dx) < \frac{1}{2}\left(\int_X C(x,\pi_x)\,\mu(dx)+\int_X C(x,\tilde{\pi}_x)\,\mu(dx)\right). \tag{6.3.4}$$

Let then π^* be the only minimiser of $\hat{V}_C(\mu,\nu)$. To conclude the proof, it is enough to show that $J(\pi^{k,*})$ converges to $J(\pi^*)$ in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ as k goes to $+\infty$. Let $P^*\in\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ be an accumulation point of $(J(\pi^{k,*}))_{k\in\mathbb{N}}$, which exists by Lemma 6.7.8.

To conclude, it suffices to show that $P^*=J(\pi^*)$, which is achieved in three steps. Let $\hat{\Lambda}(\mu,\nu)=\Lambda(\mu,\nu)$ in the setting of Theorem 6.2.5 and $\hat{\Lambda}(\mu,\nu)=\Lambda_M(\mu,\nu)$ in the setting of Theorem 6.2.6. First we show that

$$P^*\in\hat{\Lambda}(\mu,\nu). \tag{6.3.5}$$

Next, we show that $J(\pi^*)$ and P^* both minimise

$$\tilde{V}_C(\mu,\nu):=\inf_{P\in\hat{\Lambda}(\mu,\nu)}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P(dx,dp).$$

Finally, we show the uniqueness of minimisers of $\tilde{V}_C(\mu,\nu)$.

Let $(J(\pi^{k_l,*}))_{l\in\mathbb{N}}$ be a subsequence converging to P^* in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$. By Lemma 6.7.10 below we have

$$\int_{(x,p)\in X\times\mathcal{P}_g(Y)}p(dy)\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{(x,p)\in X\times\mathcal{P}_g(Y)}p(dy)\,P^*(dx,dp),$$

where the convergence holds in $\mathcal{P}_{f\oplus g}(X\times Y)$ as l goes to $+\infty$. Since the left-hand side is ν^{k_l}, which converges to ν in $\mathcal{W}_g$ and therefore in the weak topology, we deduce by uniqueness of the limit that the right-hand side is ν, hence $P^*\in\Lambda(\mu,\nu)$. In the setting of Theorem 6.2.6, since, as $f,g\in\mathcal{F}^1(\mathbb{R}^d)$, $X\times\mathcal{P}_g(Y)\ni(x,p)\mapsto|x-\int_Y y\,p(dy)|\in\Phi_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$, we have that

$$0=\int_{X\times\mathcal{P}_g(Y)}\left|x-\int_Y y\,p(dy)\right|\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{X\times\mathcal{P}_g(Y)}\left|x-\int_Y y\,p(dy)\right|\,P^*(dx,dp),$$

hence $P^*\in\Lambda_M(\mu,\nu)$.

Let us show that $J(\pi^*)$ and P^* both minimise $\tilde{V}_C(\mu,\nu)$. Note that since $P^*\in\hat{\Lambda}(\mu,\nu)$, we have $P^*(X\times\mathcal{P}_g(Y))=1$. Since $(J(\pi^{k_l,*}))_{l\in\mathbb{N}}$ converges to P^* in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$, we find with Lemma 6.7.13 like in the derivation of (6.3.2) that

$$\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,J(\pi^{k_l,*})(dx,dp)\underset{l\to+\infty}{\longrightarrow}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P^*(dx,dp). \tag{6.3.6}$$

Then (6.3.6), the definition of $\pi^{k_l,*}$ and last (a), resp. (a'), yield

$$\begin{aligned}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,P^*(dx,dp)&=\lim_{l\to+\infty}\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,J(\pi^{k_l,*})(dx,dp)\\&=\lim_{l\to+\infty}\hat{V}_C(\mu^{k_l},\nu^{k_l})\\&=\hat{V}_C(\mu,\nu)\\&=\int_{X\times\mathcal{P}_g(Y)}C(x,p)\,J(\pi^*)(dx,dp).\end{aligned}$$

Let $P(dx, dp) = \mu(dx)\, P_x(dp) \in \hat{\Lambda}(\mu, \nu)$. Then $\mu(dx) \int_{p \in \mathcal{P}_f(Y)} p(dy)\, P_x(dp) \in \hat{\Pi}(\mu, \nu)$, so by Proposition 6.7.11 below for the last inequality,

$$
\begin{aligned}
\int_{X \times \mathcal{P}_f(Y)} C(x, p)\, J(\pi^*)(dx, dp) &= \int_X C(x, \pi^*_x)\, \mu(dx) \\
&= \hat{V}_C(\mu, \nu) \\
&\leq \int_X C\left(x, \int_{p \in \mathcal{P}_f(Y)} p(dy)\, P_x(dp)\right) \mu(dx) \\
&\leq \int_X \int_{\mathcal{P}_f(Y)} C(x, p)\, P_x(dp)\, \mu(dx),
\end{aligned}
\tag{6.3.7}
$$

which proves that $J(\pi^*)$ minimises $\tilde{V}_C(\mu, \nu)$, and so does P^*.

We now prove that $J(\pi^*)$ is the only minimiser of $\tilde{V}_C(\mu, \nu)$. To do so, we first prove that any minimiser of $\tilde{V}_C(\mu, \nu)$ is in the image of J. Let then $\tilde{P}$ be such a minimiser. For $x \in X$, let $\tilde{\pi}_x(dy) = \int_{p \in \mathcal{P}_f(Y)} p(dy)\, \tilde{P}_x(dp)$ and $\tilde{\pi}(dx, dy) = \mu(dx)\, \tilde{\pi}_x(dy)$. Then $J(\tilde{\pi}) \in \hat{\Lambda}(\mu, \nu)$ and Proposition 6.7.11 below yields

$$
\int_{X \times \mathcal{P}_f(Y)} C(x, p)\, J(\tilde{\pi})(dx, dp) = \int_X C(x, \tilde{\pi}_x)\, \mu(dx) \leq \int_X \int_{\mathcal{P}_f(Y)} C(x, p)\, \tilde{P}_x(dp)\, \mu(dx).
$$

By optimality of $\tilde{P}$, this inequality is an equality, hence for $\mu(dx)$-almost every $x \in X$ we have

$$
C(x, \tilde{\pi}_x) = \int_{\mathcal{P}_f(Y)} C(x, p)\, \tilde{P}_x(dp),
$$

and therefore $\tilde{P}_x = \delta_{\tilde{\pi}_x}$ by the equality case of Proposition 6.7.11 below, or equivalently $\tilde{P} = J(\tilde{\pi})$. Therefore any minimiser of $\tilde{V}_C(\mu, \nu)$ is contained in $J(\hat{\Pi}(\mu, \nu))$.

Recall that for all $\pi \in \hat{\Pi}(\mu, \nu)$ we have

$$
\int_X C(x, \pi_x)\, \mu(dx) = \int_{X \times \mathcal{P}_g(Y)} C(x, p)\, J(\pi)(dx, dp).
$$

With (6.3.7), we deduce that $P \in \hat{\Lambda}(\mu, \nu)$ is a minimiser of $\tilde{V}_C(\mu, \nu)$ iff P is the image of a minimiser of $\hat{V}_C(\mu, \nu)$ by J. By (6.3.4) the minimiser of $\hat{V}_C(\mu, \nu)$ is unique. This shows the uniqueness of minimisers of $\tilde{V}_C(\mu, \nu)$, and therefore the uniqueness of accumulation points of $(J(\pi^{k,*}))_{k \in \mathbb{N}}$, which is conclusive. □

The proof of Corollary 6.2.8 relies on the following Lemma.

Lemma 6.3.1. *Let $\rho > 1$, and $C_\rho : \mathbb{R} \times \mathcal{P}^\rho(\mathbb{R}) \to \mathbb{R}$ be defined for all $(x, p) \in \mathbb{R} \times \mathcal{P}^\rho(\mathbb{R})$ by $C_\rho(x, p) = \mathcal{W}_\rho^\rho(p, \gamma)$, where $\gamma \in \mathcal{P}^\rho(\mathbb{R})$ does not weight points. Let $V^M_{C_\rho}$ be the value function given by* (WMOT) *for the cost function C_ρ.*

Let $r \geq \rho$ and $\mu^k, \nu^k \in \mathcal{P}^r(\mathbb{R})$, $k \in \mathbb{N}$ be in convex order and converge respectively to μ and ν in $\mathcal{W}_r$. Then $\lim_{k \to +\infty} V^M_{C^\rho}(\mu^k, \nu^k) = V^M_{C^\rho}(\mu, \nu)$ and the optimisers are converging in $\mathcal{AW}_r$.

Proof. By Theorem 6.2.6 it is sufficient to show that $p \mapsto \mathcal{W}_\rho^\rho(\gamma, p)$ is strictly convex. Since γ does not weight points, the unique $\mathcal{W}_\rho$-optimal coupling between γ and $p \in \mathcal{P}_\rho(\mathbb{R})$ is the comonotonous coupling χ^p given by the map $x \mapsto F_p^{-1}(F_\gamma(x))$ i.e. the image of γ by $x \mapsto (x, F_p^{-1}(F_\gamma(x)))$. For $q \in \mathcal{P}_\rho(\mathbb{R})$ and $\lambda \in (0,1)$ the coupling $\chi = (1-\lambda)\chi^p + \lambda\chi^q$ between γ and $(1-\lambda)p + \lambda q$ is not given by a map unless $F_q^{-1}(u) = F_p^{-1}(u)$ for all $u \in (0,1)$ i.e. $p = q$. Therefore, when $p \neq q$,

$$(1-\lambda)\mathcal{W}_\rho^\rho(\gamma, p) + \lambda\mathcal{W}_\rho^\rho(\gamma, q) = \int |x-y|^\rho \, \chi(dx, dy) > \mathcal{W}_\rho^\rho(\gamma, (1-\lambda)p + \lambda q).$$

□

We can now prove the stability of the unidimensional stretched Brownian motion.

Proof of Corollary 6.2.8. Let $\gamma = \mathcal{N}(0,1)$ be the unidimensional standard normal distribution and $C_2 : \mathbb{R} \times \mathcal{P}^2(\mathbb{R}) \to \mathbb{R}$ be defined for all $(x,p) \in \mathbb{R} \times \mathcal{P}^2(\mathbb{R})$ by $C_2(x,p) = \mathcal{W}_2^2(p, \gamma)$. Let $V_{C_2}^M$ be the value function given by (WMOT) for the cost function C_2.

In the setting of Corollary 6.2.8, let $\pi^* \in \Pi_M(\mu, \nu)$, resp. $\pi^k \in \Pi_M(\mu^k, \nu^k)$ be optimal for $V_{C_2}^M(\mu, \nu)$, resp. $V_{C_2}^M(\mu^k, \nu^k)$. For $(x, b) \in \mathbb{R} \times \mathbb{R}^{[0,1]}$, let $B = (B_t)_{t\in[0,1]}$ be a Brownian motion and

$$G^k(x,b) = \left(\mathbb{E}\left[F_{\pi_x^k}^{-1}(F_\gamma(B_1 - B_t + b_t))\right]\right)_{t\in[0,1]} \text{ and } G^*(x,b) = \left(\mathbb{E}\left[F_{\pi_x^*}^{-1}(F_\gamma(B_1 - B_t + b_t))\right]\right)_{t\in[0,1]}.$$

According to (6.2.9), $(M_0^k, (M_t^k)_{t\in[0,1]})$ and $(M_0^*, (M_t^*)_{t\in[0,1]})$ are respectively distributed according to

$$\eta^k(dx, df) := \mu^k(dx)\,(G^k(x,\cdot)_* W)(df) \quad \text{and} \quad \eta^*(dx, df) := \mu(dx)\,(G^*(x,\cdot)_* W)(df),$$

where W denotes the Wiener measure on $C([0,1])$. Let $\chi^k \in \Pi(\mu^k, \mu)$ be optimal for $\mathcal{AW}_r(\pi^k, \pi)$. Then

$$\mathcal{AW}_r^r(\eta^k, \eta^*) \leq \int_{\mathbb{R}\times\mathbb{R}} \Big(|x-x'|^r + \mathcal{W}_r^r(G^k(x,\cdot)_* W, G(x',\cdot)_* W)\Big)\, \chi^k(dx, dx').$$

According to (6.2.9), for $\mu^k(dx)$-almost every $x \in \mathbb{R}$, $G^k(x, B)$ is the stretched Brownian motion from δ_x to π_x^k, hence it is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale, where $(\mathcal{F}_t)_{t\in[0,1]}$ is the natural filtration associated to B. Similarly, for $\mu(dx')$-almost every $x \in \mathbb{R}$, $G^*(x', B)$ is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. Therefore, for $\chi^k(dx, dx')$-almost every $(x, x') \in \mathbb{R} \times \mathbb{R}$, $G^k(x, B) - G^*(x', B)$ is a continuous $(\mathcal{F}_t)_{t\in[0,1]}$-martingale. Using Doob's martingale inequality for the second inequality, the fact that $F_\gamma(B_1)$ is uniformly distributed on $(0,1)$ for the first equality and the fact that the comonotonous coupling between π_x^k and $\pi_{x'}^*$ is optimal for $\mathcal{W}_r(\pi_x^k, \pi_{x'}^*)$ for the second equality, we get for $\chi^k(dx, dx')$-almost every $(x, x') \in \mathbb{R}\times\mathbb{R}$

$$\begin{aligned}
\mathcal{W}_r^r(G^k(x,\cdot)_* W, G(x',\cdot)_* W) &\leq \mathbb{E}\left[\sup_{t\in[0,1]} \Big|G^k(x,B)_t - G^*(x',B)_t\Big|^r\right] \\
&\leq \left(\frac{r}{r-1}\right)^r \mathbb{E}[|G^k(x,B)_1 - G^*(x',B)_1|^r]
\end{aligned}$$

$$= \left(\frac{r}{r-1}\right)^r \mathbb{E}[|F^{-1}_{\pi^k_x}(F_\gamma(B_1)) - F^{-1}_{\pi^*_{x'}}(F_\gamma(B_1)|^r]$$
$$= \left(\frac{r}{r-1}\right)^r \mathcal{W}_r^r(\pi^k_x, \pi^*_{x'}).$$

We deduce that

$$\mathcal{AW}_r^r(\eta^k, \eta^*) \leq \left(\frac{r}{r-1}\right)^r \int_{\mathbb{R}\times\mathbb{R}} \left(|x-x'|^r + \mathcal{W}_r^r(\pi^k_x, \pi^*_{x'})\right) \chi^k(dx, dx') = \left(\frac{r}{r-1}\right)^r \mathcal{AW}_r^r(\pi^k, \pi^*),$$

where the right-hand side vanishes as k goes to $+\infty$ in virtue of Lemma 6.3.1. □

6.4 Martingale monotonicity

In this section we prove the claim that martingale C-monotonicity is sufficient for optimality for (WMOT).

For $g : Y \to [1, +\infty)$ continuous, we denote

$$\mathcal{F}_g(Y) := \{f : Y \to [1, +\infty) \text{ continuous } \mid \forall y \in Y,\ f(y) \geq g(y)\}. \tag{6.4.1}$$

When $Y = \mathbb{R}^d$ for some $d \in \mathbb{N}^*$, we denote

$$\mathcal{F}_g^+(\mathbb{R}^d) := \left\{ f \in \mathcal{F}_g(\mathbb{R}^d) \mid \exists h : \mathbb{R}_+ \to [1, +\infty),\ \frac{h(t)}{t} \underset{t\to+\infty}{\longrightarrow} +\infty \text{ and } f = h \circ g \right\}. \tag{6.4.2}$$

Proof of Theorem 6.2.15. Let $h \in \mathcal{F}_g(\mathbb{R})$ be such that $\nu(h) < +\infty$, whose purpose will be revealed later in the proof. To demonstrate the main idea without further technical details, we assume for now that μ is concentrated on a Polish subset $\tilde{K} \subset \mathbb{R}$ and the restriction $C|_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})}$ is continuous. Let $X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$ be independent random variables identically distributed according to μ and $\mathcal{G} \subset \Phi_{f\oplus\hat{g}}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$ be a countable family which determines the convergence in $\mathcal{P}_{f\oplus\hat{g}}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$ (see [78, Theorem 4.5.(b)]). By the law of large numbers, almost surely, for all $\psi \in \mathcal{G}$,

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^n \int_{\mathbb{R}\times\mathcal{P}(\mathbb{R})} \psi(x,p)\, \delta_{(X_k, \pi_{X_k})}(dx, dp) &= \frac{1}{n}\sum_{k=1}^n \psi(X_k, \pi_{X_k}) \\ &\underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\psi(X_1, \pi_{X_1})] \\ &= \int_{\mathbb{R}} \psi(x, \pi_x)\, \mu(dx) \\ &= \int_{\mathbb{R}\times\mathcal{P}(\mathbb{R})} \psi(x,p)\, J(\pi)(dx, dp),\end{aligned} \tag{6.4.3}$$

Moreover, almost surely, for all $n \in \mathbb{N}$,

$$(X_n, \pi_{X_n}) \in \Gamma \cap (\tilde{K} \times \mathcal{P}_g(\mathbb{R})), \tag{6.4.4}$$

and by the law of large numbers again, we have almost surely

$$\frac{1}{n}\sum_{k=1}^{n}\pi_{X_k}(h) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\pi_{X_1}(h)] = \int_{\mathbb{R}}\pi_x(h)\,\mu(dx) = \nu(h). \tag{6.4.5}$$

Let then $\omega\in\Omega$ be such that (6.4.3), (6.4.4) and (6.4.5) hold when evaluated at ω and set $x_n = X_n(\omega)$ and $\pi^n(dx,dy) = \frac{1}{n}\sum_{k=1}^n \delta_{x_k}(dx)\,\pi_{x_k}(dy)$ for $n\in\mathbb{N}$. Then π^n has first marginal $\mu^n = \frac{1}{n}\sum_{k=1}^n\delta_{x_k}$ and second marginal $\nu^n = \int_{x\in\mathbb{R}}\pi_x(dy)\,\mu^n(dx)$. We deduce that π^n is a martingale C-monotone coupling between μ^n and ν^n which satisfies

$$J(\pi^n) = \frac{1}{n}\sum_{k=1}^{n}\delta_{(x_k,\pi_{x_k})} \underset{n\to+\infty}{\longrightarrow} J(\pi) \quad \text{in } \mathcal{P}_{f\oplus\hat{g}}(\tilde{K}\times\mathcal{P}(\mathbb{R})),$$

Thus π^n converges to π in $\mathcal{AW}_{f\oplus\hat{g}}$ as n goes to $+\infty$. In particular we have convergence of the marginals in $\mathcal{P}_f(\mathbb{R})$ and $\mathcal{P}_g(\mathbb{R})$ respectively, and even convergence in $\mathcal{P}_h(\mathbb{R})$ of the second marginals since $\nu^n(h)$ converges to $\nu(h)$ as $n\to+\infty$, consequence of (6.4.5) evaluated at ω. Note that due to martingale C-monotonicity of π^n, we have according to Remark 6.2.14 that

$$V_C^M(\mu^n,\nu^n) = \int_{\mathbb{R}} C(x,\pi^n_x)\,\mu^n(dx),$$

where we recall that the value function V_C^M is defined in (WMOT). Since μ^n, resp. ν^n converges to μ, resp. ν in $\mathcal{P}_f(\mathbb{R})$, resp. $\mathcal{P}_h(\mathbb{R})$, by Theorem 6.2.6 we have convergence of the optimal values $V_C^M(\mu^n,\nu^n)\to V_C^M(\mu,\nu)$ as n goes to $+\infty$. By convergence of $(\nu^n(h))_{n\in\mathbb{N}}$ to $\nu(h)$, the convergence $J(\pi^n)\to J(\pi)$ is not only in $\mathcal{P}_{f\oplus\hat{g}}(\tilde{K}\times\mathcal{P}(\mathbb{R}))$, but even in $\mathcal{P}_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}(\mathbb{R}))$ (see Definition 6.2.2) and therefore $\mathcal{P}_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}_h(\mathbb{R}))$ by Lemma 6.7.3 (b) below. In that context, $C|_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})}\in\Phi_{f\oplus\hat{h}}(\tilde{K}\times\mathcal{P}_h(\mathbb{R}))$, so

$$\begin{aligned}
\int_{\mathbb{R}} C(x,\pi_x)\,\mu(dx) &= \int_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\,J(\pi)(dx,dp)\\
&= \lim_{n\to+\infty}\int_{\tilde{K}\times\mathcal{P}_h(\mathbb{R})} C(x,p)\,J(\pi^n)(dx,dp)\\
&= \lim_{n\to+\infty}\int_{\mathbb{R}} C(x,\pi^n_x)\,\mu^n(dx)\\
&= \lim_{n\to+\infty} V_C^M(\mu^n,\nu^n)\\
&= V_C^M(\mu,\nu),
\end{aligned}$$

hence π is optimal for $V_C^M(\mu,\nu)$.

Next, we drop the additional joint-continuity assumption on C. Since $\nu(g)<+\infty$, there exists by the de La Vallée Poussin theorem $h\in\mathcal{F}_g^+(\mathbb{R})$ such that $\nu(h)<+\infty$. For $N\in\mathbb{N}^*$, let $B_N = \{p\in\mathcal{P}_g(\mathbb{R}) \mid p(h)\le N\}$, which is a compact subset of $\mathcal{P}_g(\mathbb{R})$ by Lemma 6.7.7 below, and $\mathcal{C}(B_N)$ be the set of continuous functions from B_N to $\mathbb{R}$, endowed with the topology of uniform convergence. The map $\phi^N\colon\mathbb{R}\to\mathcal{C}(B_N)$ given by $\phi^N(x) = C(x,\cdot)|_{B_N}$ is Borel measurable due to [6, Theorem 4.55]. Let $\varepsilon\in(0,1)$. By Lusin's theorem there is for every $N \in \mathbb{N}^*$ a compact set $K^N \subset \mathbb{R}$ such that the restriction $\phi^N|_{K^N}$ is continuous and $\mu(K^N) \geq 1 - \frac{\varepsilon}{2^N}$. We have

$$\mu\left(\bigcap_{N\in\mathbb{N}^*} K^N\right) \geq 1 - \sum_{N\in\mathbb{N}^*} \mu\left((K^N)^c\right) \geq 1 - \sum_{N\in\mathbb{N}^*} \frac{\varepsilon}{2^N} = 1 - \varepsilon.$$

Let $K^\varepsilon = \bigcap_{N\in\mathbb{N}^*} K^N$, then for all $N \in \mathbb{N}^*$ the restriction $\phi^N|_{K^\varepsilon}$ is continuous. We claim that $C|_{K^\varepsilon\times\mathcal{P}_h(\mathbb{R})}$ is continuous w.r.t. the product topology of $\mathbb{R}\times\mathcal{P}_h(\mathbb{R})$. To this end, take any sequence $(x_k, p_k)_{k\in\mathbb{N}} \in (K^\varepsilon \times \mathcal{P}_h(\mathbb{R}))^{\mathbb{N}}$ with limit point $(x, p) \in K^\varepsilon \times \mathcal{P}_h(\mathbb{R})$. Since $p_k \to p$ in $\mathcal{P}_h(\mathbb{R})$ as k goes to $+\infty$, the sequence $(p_k(h))_{k\in\mathbb{N}}$ is convergent and therefore bounded so there exists $N \in \mathbb{N}$ such that $p, p_k \in B_N$ for all $k \in \mathbb{N}$. As $\phi^N(x_k)$ converges uniformly to $\phi^N(x)$, we have

$$C(x_k, p_k) = \phi^N(x_k)(p_k) \underset{k\to+\infty}{\longrightarrow} \phi^N(x, p) = C(x, p).$$

Therefore, $C|_{K^\varepsilon\times\mathcal{P}_h(\mathbb{R})}$ is continuous.

Let $\mu^\varepsilon = \frac{1}{\mu(K^\varepsilon)}\mu|_{K^\varepsilon}$, $\pi^\varepsilon = \mu^\varepsilon \times \pi_x = \frac{1}{\mu(K^\varepsilon)}\pi|_{K^\varepsilon\times\mathbb{R}}$ and ν^ε be the second marginal of π^ε. Obviously μ^ε is concentrated on K^ε. Since $\mu(K^\varepsilon)\mu^\varepsilon \leq \mu$ and $\pi^\varepsilon_x = \pi_x$, π^ε is martingale C-monotone and satisfies $(x, \pi^\varepsilon_x) \in \Gamma$ for $\mu^\varepsilon(dx)$-almost every x. Finally, $\mu(K^\varepsilon)\nu^\varepsilon(h) = \int_{K^\varepsilon} \pi_x(h)\,\mu(dx) \leq \nu(h) < +\infty$, hence $\nu^\varepsilon \in \mathcal{P}_h(\mathbb{R})$. Therefore the reasoning of the first part applied with $(K^\varepsilon, \mu^\varepsilon, \nu^\varepsilon, \pi^\varepsilon)$ replacing $(\tilde{K}, \mu, \nu, \pi)$ proves that π^ε is optimal for $V^M_C(\mu^\varepsilon, \nu^\varepsilon)$.

Next, we convince ourselves that $J(\pi^\varepsilon)$ converges to $J(\pi)$ stably in $\mathcal{P}_{f\oplus\hat{h}}(\mathbb{R}\times\mathcal{P}(\mathbb{R}))$: let $\psi\colon \mathbb{R}\times\mathcal{P}(\mathbb{R})$ be measurable and absolutely dominated by a positive multiple of $f \oplus \hat{h}$, then

$$J(\pi^\varepsilon)(\psi) = \int_{\mathbb{R}} \psi(x, \pi_x)\,\mu^\varepsilon(dx) = \frac{1}{\mu(K^\varepsilon)}\int_{K^\varepsilon} \psi(x, \pi_x)\,\mu(dx) \underset{\varepsilon\to 0}{\longrightarrow} \int_{\mathbb{R}} \psi(x, \pi_x)\,\mu(dx) = J(\pi)(\psi),$$

where we employed dominated convergence and that $1 - \varepsilon \leq \mu(K^\varepsilon) \leq 1$. In particular, the marginals $(\mu^\varepsilon)_{\varepsilon>0}$ converge to μ in $\mathcal{P}_f(\mathbb{R})$ and strongly, whereas the marginals $(\nu^\varepsilon)_{\varepsilon>0}$ converge to ν in $\mathcal{P}_h(\mathbb{R})$ for $\varepsilon \searrow 0$. Using item (d) of Lemma 6.7.13 yields

$$\begin{aligned}\lim_{n\to+\infty} V^M_C(\mu^{1/n}, \nu^{1/n}) &= \lim_{n\to+\infty} \int_{\mathbb{R}\times\mathcal{P}_h(\mathbb{R})} C(x, p)\,J(\pi^{1/n})(dx, dp) \\ &= \int_{\mathbb{R}\times\mathcal{P}_h(\mathbb{R})} C(x, p)\,J(\pi)(dx, dp) = \int_{\mathbb{R}} C(x, \pi_x)\,\mu(dx).\end{aligned}$$

The marginal sequences $(\mu^{1/n})_{n\in\mathbb{N}^*}$ and $(\nu^{1/n})_{n\in\mathbb{N}^*}$ satisfy the assumptions of Theorem 6.2.6 with (B). Hence, by item (b') of the very same theorem we have that

$$V^M_C(\mu, \nu) = \lim_{n\to+\infty} V^M_C(\mu^{1/n}, \nu^{1/n}) = \int_{\mathbb{R}} C(x, \pi_x)\,\mu(dx),$$

proving optimality of π. □

6.5 On the Wasserstein projection

In order to prove Theorem 6.2.18, we need first to establish some auxiliary results, especially Proposition 6.5.4 which shows that the convex hull is a contraction.

Lemma 6.5.1. *Let $0 \leq a < b$ and $F, G : [0, b] \to \mathbb{R}$ be continuous on $[0, b]$, convex on $[0, a)$ and affine on $[a, b]$. Then for any increasing convex function $\theta \colon \mathbb{R} \to \mathbb{R}$ we have*

$$\int_0^b \theta\left(|\partial_+(\mathrm{co}(F) - \mathrm{co}(G))|\right)(u)\, du \leq \int_0^b \theta\left(|\partial_+(F - G)|\right)(u)\, du \tag{6.5.1}$$

where co *denotes the convex hull and ∂_+ the right-hand derivative.*

Proof of Lemma 6.5.1. Let f and g be the respective right-hand derivatives of F and G, so that f and g are right-continuous functions, nondecreasing on $[0, a)$, which satisfy $F(x) = F(0) + \int_0^x f(u)\, du$ and $G(x) = G(0) + \int_0^x g(u)\, du$ for all $x \in [0, b]$.

First, we make the convex hull of F explicit. Let

$$c = \inf\left\{x \in [0, a] \mid f(x) \geq \frac{F(b) - F(x)}{b - x}\right\}.$$

Since $f(a) = \frac{F(b)-F(a)}{b-a}$, c is the infimum of a nonempty set. By continuity of F and right-continuity of f, we have $f(c) \geq \frac{F(b)-F(c)}{b-c}$. Let $\Delta : [0, b] \to \mathbb{R}$ be the affine function which coincides with F at c and b. Let us show that $\Delta \leq F$. We have for all $x \in [c, a]$

$$F(x) = F(c) + \int_c^x f(u)\, du \geq F(c) + (x - c)f(c) \geq F(c) + (x - c)\frac{F(b) - F(c)}{b - c} = \Delta(x).$$

If $c > 0$, then by definition of the infimum and left-continuity of $x \mapsto f(x-)$ on $(0, a]$ we have $f(c-) \leq \frac{F(b)-F(c)}{b-c}$. Similarly to the previous calculation, we have for all $x \in [0, c]$

$$\begin{aligned} F(x) &= F(c) + \int_c^x f(u-)\, du = F(c) - \int_x^c f(u-)\, du \geq F(c) - (c - x)f(c-) \\ &\geq F(c) - (c - x)\frac{F(b) - F(c)}{b - c} \\ &= \Delta(x), \end{aligned}$$

hence $\Delta \leq F$ on $[0, a]$. Moreover, F is affine on $[a, b]$ and is equal to $F(a) \geq \Delta(a)$ at a and $\Delta(b)$ at b. We deduce that $\Delta \leq F$ on $[a, b]$ and therefore on $[0, b]$. This implies that for all $x \in [0, b]$,

$$\frac{F(b) - F(x)}{b - x} \leq \frac{F(b) - \Delta(x)}{b - x} = \frac{\Delta(b) - \Delta(x)}{b - x} = \frac{\Delta(b) - \Delta(c)}{b - c} = \frac{F(b) - F(c)}{b - c}. \tag{6.5.2}$$

Let $x \in [0, a)$. If $x \geq c$ then $f(x) \geq f(c) \geq \frac{F(b)-F(c)}{b-c}$. If $x < c$ then by definition of c and (6.5.2), $f(x) < \frac{F(b)-F(x)}{b-x} \leq \frac{F(b)-F(c)}{b-c}$. We deduce that

$$\forall x \in [0, a), \quad f(x) \geq \frac{F(b) - F(c)}{b - c} \iff x \geq c. \tag{6.5.3}$$

Let $\tilde{F} : [0, b] \to \mathbb{R}$ be the function which coincides with F on $[0, c]$ and with Δ on $[c, b]$. Let us show that $\tilde{F} = \mathrm{co}(F)$. First, we deduce from (6.5.3) that

$$\forall x \in [0, b], \quad \tilde{F}(x) = F(0) + \int_0^x \left(f(u \wedge c) \wedge \left(\frac{F(b) - F(c)}{b - c} \right) \right) du, \tag{6.5.4}$$

hence $\tilde{F}$ is the sum of a constant and an integrated nondecreasing function and is therefore convex. Moreover, since $\Delta \leq F$ we have $\tilde{F} \leq F$, so by definition of the convex hull, $\tilde{F} \leq \mathrm{co}(F)$. On the other hand, let $x \in [0, b]$. If $x \in [0, c] \cup \{b\}$, then $\tilde{F}(x) = F(x) \geq \mathrm{co}(F)(x)$. If $x \in (c, b)$, then by convexity of $\mathrm{co}(F)$ for the second inequality and the fact that $\mathrm{co}(F)(x) \geq \tilde{F}(x) = \Delta(x)$ and $\mathrm{co}(F)(c) \leq F(c) = \Delta(c)$ for the third inequality, we have

$$\begin{aligned}
F(b) \geq \mathrm{co}(F)(b) &= \mathrm{co}(F)(x) + (b - x)\frac{\mathrm{co}(F)(b) - \mathrm{co}(F)(x)}{b - x} \\
&\geq \mathrm{co}(F)(x) + (b - x)\frac{\mathrm{co}(F)(x) - \mathrm{co}(F)(c)}{x - c} \\
&\geq \mathrm{co}(F)(x) + (b - x)\frac{\Delta(x) - \Delta(c)}{x - c} \\
&= \mathrm{co}(F)(x) + (b - x)\frac{\Delta(b) - \Delta(x)}{b - x} \\
&\geq \Delta(x) + (b - x)\frac{\Delta(b) - \Delta(x)}{b - x} \\
&= \Delta(b) = F(b),
\end{aligned}$$

so those inequalities are equalities and $\tilde{F}(x) = \Delta(x) = \mathrm{co}(F)(x)$. We deduce that $\tilde{F} = \mathrm{co}(F)$, or equivalently

$$co(F)|_{[0,c]} = F|_{[0,c]}, \quad co(F)(x) = \frac{b - x}{b - c} F(c) + \frac{x - c}{b - c} F(b) \quad \text{for } x \in [c, b].$$

Reasoning in the same way for G, we deduce the existence of $d \in [0, a]$ such that for all $x \in [0, a)$, $g(x) \geq \frac{G(b) - G(d)}{b - d} \iff x \geq d$ and $\mathrm{co}(G)(x) = G(0) + \int_0^x \left(g(u \wedge d) \wedge \left(\frac{G(b) - G(d)}{b - d} \right) \right) du$, or equivalently

$$co(G)|_{[0,d]} = G|_{[0,d]}, \quad co(G)(x) = \frac{b - x}{b - d} G(d) + \frac{x - d}{b - d} G(b) \quad \text{for } x \in [d, b].$$

Without loss of generality we assume that $c \leq d$. Let $\varphi = \frac{F(b) - F(c)}{b - c}$ and $\gamma = \frac{G(b) - G(d)}{b - d}$. Then the left-hand side of (6.5.1) is given by

$$\begin{aligned}
&\int_0^b \theta\left(|(f(u \wedge c) \wedge \varphi|) - \theta(|g(u \wedge d) \wedge \gamma|)\right) du \\
&\quad = \int_0^c \theta\left(|f(u) - g(u)|\right) du + \int_c^d \theta\left(|\varphi - g(u)|\right) du + \int_d^b \theta\left(|\varphi - \gamma|\right) du.
\end{aligned} \tag{6.5.5}$$

It remains to show that

$$\int_c^d \theta\left(|\varphi - g(u)|\right)\, du + \int_d^b \theta\left(|\varphi - \gamma|\right)\, du \leq \int_c^b \theta\left(|f(u) - g(u)|\right)\, du. \tag{6.5.6}$$

Suppose first that $\varphi \geq \gamma$. It was shown that for all $x \in [0, a)$, $g(x) \geq \gamma$ iff $x \geq d$. Therefore,

$$g(x) < \gamma \leq \varphi \leq f(x) \quad \forall x \in [c, d). \tag{6.5.7}$$

Hence, we find by increasing convexity of θ and convexity of the modulus that

$$\begin{aligned}
\int_c^b \theta\left(|f(u) - g(u)|\right)\, du &\geq \int_c^d \theta\left(f(u) - g(u)\right)\, du + \int_d^b \theta\left(\frac{1}{b-d}\int_d^b |f(v) - g(v)|\, dv\right)\, du \\
&\geq \int_c^d \theta\left(f(u) - g(u)\right)\, du + \int_d^b \theta\left(\left|\int_d^b \frac{f(v) - g(v)}{b-d}\, dv\right|\right)\, du \\
&= \int_c^d \theta\left(f(u) - g(u)\right)\, du + \int_d^b \theta\left(|\tilde{\phi} - \gamma|\right),
\end{aligned}$$

where $\tilde{\phi} = \frac{F(b)-F(d)}{b-d}$. To obtain (6.5.6) it is sufficient to show the following inequality

$$\int_c^d \left(\theta(f(u) - g(u)) - \theta(\phi - g(u))\right)\, du + \int_d^b \left(\theta(|\tilde{\phi} - \gamma|) - \theta(\phi - \gamma)\right)\, du \geq 0. \tag{6.5.8}$$

By convexity of θ we have $\theta(y) - \theta(x) \leq (y-x)\frac{\theta(z)-\theta(x)}{z-x}$, resp. $(z-y)\frac{\theta(z)-\theta(x)}{z-x} \leq \theta(z) - \theta(y)$ for all $(x, y, z) \in \mathbb{R}^3$ such that $x \leq y < z$, resp. $x < y \leq z$. By (6.5.7) we can apply the former inequality with $(x, y, z) = (\varphi - \gamma, f(u) - \gamma, f(u) - g(u))$ and the latter inequality with $(x, y, z) = (\varphi - \gamma, \varphi - g(u), f(u) - g(u))$ for all $u \in [c, d]$, which yields $\theta(f(u) - \gamma) - \theta(\varphi - \gamma) \leq \theta(f(u) - g(u)) - \theta(\varphi - g(u))$. By integration on $[c, d]$ we get

$$\int_c^d \left(\theta(f(u) - g(u)) - \theta(\phi - g(u))\right)\, du \geq \int_c^d \left(\theta(f(u) - \gamma) - \theta(\phi - \gamma)\right)\, du. \tag{6.5.9}$$

Applying (6.5.9) yields the first inequality, Jensen's inequality the second and Jensen's inequality coupled with θ being increasing the final one

$$\begin{aligned}
&\int_c^d \theta(f(u) - g(u)) - \theta(\phi - g(u))\, du + \int_d^b \theta(|\tilde{\phi} - \gamma|) - \theta(\phi - \gamma)\, du \\
&\quad \geq \int_c^d \theta(f(u) - \gamma) - \theta(\phi - \gamma)\, du + \int_d^b \theta(|\tilde{\phi} - \gamma|) - \theta(\phi - \gamma)\, du \\
&\quad \geq \int_c^b \theta\left(\frac{1}{b-c}\left(\int_c^d f(v) - \gamma\, dv + \int_d^b |\tilde{\phi} - \gamma|\, dv\right)\right) - \theta(\phi - \gamma)\, du \\
&\quad \geq \int_c^b \theta\left(\left|\frac{1}{b-c}\left(\int_c^d f(v) - \gamma\, dv + \int_d^b f(v) - g(v)\, dv\right)\right|\right) - \theta(\phi - \gamma)\, du \\
&\quad = \int_c^b \theta\left(\phi - \gamma\right) - \theta(\phi - \gamma)\, du = 0.
\end{aligned}$$

Suppose now $\varphi < \gamma$. Let $e = \inf\{u \in [c,d] \mid g(u) \geq \varphi\}$, where $\inf \emptyset = d$ by convention, so that

$$\int_c^d \theta\left(|\varphi - g(u)|\right)\, du + \int_d^b \theta\left(|\varphi - \gamma|\right)\, du$$
$$= \int_c^e \theta\left(\varphi - g(u)\right)\, du + \int_e^d \theta\left(g(u) - \varphi\right)\, du + \int_d^b \theta\left(\gamma - \varphi\right)\, du.$$

On the one hand, (6.5.3) implies that $\int_c^e \theta(\varphi - g(u))\, du \leq \int_c^e \theta(f(u) - g(u))\, du = \int_c^e \theta(|f(u) - g(u)|)\, du$. On the other hand, (6.5.2) implies that $\tilde{\phi} = \frac{F(b)-F(d)}{b-d} \leq \varphi$. Again due to convexity and monotonicity of θ, we get

$$\int_d^b \theta(|f(u) - g(u)|)\, du \geq \int_d^b \theta(|\tilde{\phi} - \gamma|)\, du = \int_d^b \theta(\gamma - \tilde{\phi})\, du.$$

Hence, it is sufficient to show

$$\frac{1}{b-e}\left(\int_e^d \delta_{g(u)-\phi}\, du + (b-d)\, \delta_{\gamma-\phi}\right) \leq_{icx} \frac{1}{b-e}\left(\int_e^d \delta_{|g(u)-f(u)|}\, du + (b-d)\, \delta_{\gamma-\tilde{\phi}}\right), \quad (6.5.10)$$

where $\leq_{icx}$ denotes the increasing convex order. Let η_1, resp. η_2 denote the left-hand, resp. right-hand side of (6.5.10). By [169, Theorem 4.A.3] (6.5.10) is then equivalent to

$$\forall p \in [0,1], \quad \int_p^1 F_{\eta_1}^{-1}(u)\, du \leq \int_p^1 F_{\eta_2}^{-1}(u)\, du. \quad (6.5.11)$$

On $[0,d)$, the map g is nondecreasing and dominated by γ, so the map

$$u \mapsto (g((e + (b-e)u)-) - \varphi)\mathbb{1}_{\left\{u \leq \frac{d-e}{b-e}\right\}} + (\gamma - \varphi)\mathbb{1}_{\left\{\frac{d-e}{b-e} < u\right\}} \quad (6.5.12)$$

is nondecreasing on $(0,1)$. Since the image of the Lebesgue measure on $(0,1)$ by the latter map is η_1, by [4, Lemma A.3] the expression of $F_{\eta_1}^{-1}$ is given by (6.5.12). Moreover the image of the Lebesgue measure on $(0,1)$ by the map

$$u \mapsto |g((e + (b-e)u)-) - f((e + (b-e)u)-)|\mathbb{1}_{\left\{u \leq \frac{d-e}{b-e}\right\}} + (\gamma - \tilde{\varphi})\mathbb{1}_{\left\{\frac{d-e}{b-e} < u\right\}}$$

is η_2. By Lemma 6.7.1 below we deduce that

$$J(p) := \int_{(e+(b-e)p)\wedge d}^d |g(u) - f(u)|\, du + (b - ((e + (b-e)p) \vee d))(\gamma - \tilde{\varphi}) \leq \int_p^1 F_{\eta_2}^{-1}(u)\, du,$$

and (6.5.11) is satisfied if for all $p \in [0,1]$,

$$\int_p^1 F_{\eta_1}^{-1}(u)\, du = \int_{(e+(b-e)p)\wedge d}^d (g(u) - \phi)\, du + (b - ((e + (b-e)p) \vee d))(\gamma - \varphi) \leq J(p).$$

First let $p \in [0,1]$ be such that $p > \frac{d-e}{b-e}$. Since $\tilde{\varphi} \leq \varphi$, we have

$$\int_p^1 F_{\eta_1}^{-1}(u)\, du = (b-e)(1-p)(\gamma - \varphi) \leq (b-e)(1-p)(\gamma - \tilde{\varphi}) = J(p).$$

Next let $p \in [0,1]$ be such that $p \leq \frac{d-e}{b-e}$. We have

$$\int_{e+(b-e)p}^{d} g(u)\,du \leq \int_{e+(b-e)p}^{d} \left(f(u) + |f(u)-g(u)|\right)\,du,$$

hence

$$\int_p^1 F_{\eta_1}^{-1}(u)\,du \leq \int_{e+(b-e)p}^{d} |f(u)-g(u)|\,du + \int_{e+(b-e)p}^{d} (f(u)-\varphi)\,du + (b-d)(\gamma-\phi).$$

Recall that $(b-d)\tilde{\phi} = F(b) - F(d)$ and $\tilde{\phi} \leq \phi$. Therefore, the inequality

$$\int_{e+(b-e)p}^{d} \left(f(u)-\phi\right)\,du + (b-d)(\gamma-\phi) \leq (b-d)(\gamma-\tilde{\phi})$$

is equivalent to

$$F(d) - F(e+(b-e)p) + (b-d)\tilde{\phi} = F(b) - F(e+(b-e)p) \leq (b-(e+(b-e)p))\phi,$$

which is always satisfied by (6.5.2). This yields

$$\int_p^1 F_{\eta_1}^{-1}(u)\,du \leq \int_{e+(b-e)p}^{d} |g(u)-f(u)|\,du + (b-d)(\gamma-\tilde{\varphi}) = J(p),$$

which concludes the proof. $\square$

Lemma 6.5.2. *Let f be real-valued and càdlàg on $[0,1]$ and F be an antiderivative of f. Let I be an interval of $[0,1]$ and $(c,d) = (\inf I, \sup I)$. Then*

(a) For all $x \in (c,d)$, $|\partial_+ \operatorname{co}(F|_I)(x)| \leq \|f\|_\infty < +\infty$.

(b) $\operatorname{co}(F|_I)(c+)$ and $\operatorname{co}(F|_I)(d-)$ exist and are respectively equal to $F(c)$ and $F(d)$.

Proof. Let $K = \|f\|_\infty$, which is finite since f is càdlàg on a compact interval. Let I be an interval of $[0,1]$ and $(c,d) = (\inf I, \sup I)$. Then the maps $x \mapsto F(x) - K(x-c)$ and $x \mapsto F(d) - K(d-x)$ are convex and dominated by F on I. By definition of the convex hull we find for all $x \in I$

$$F(c) - K(x-c) \leq \operatorname{co}(F|_I)(x) \leq F(x) \leq F(c) + K(x-c),$$

$$F(d) - K(d-x) \leq \operatorname{co}(F|_I)(x) \leq F(x) \leq F(c) + K(d-x),$$

which clearly implies (b). By convexity of $\operatorname{co}(F|_I)$ we then have for all $x \in (c,d)$

$$\begin{aligned} -K \leq \frac{\operatorname{co}(F|_I)(x) - F(c)}{x-c} &\leq \frac{\operatorname{co}(F|_I)(x) - \operatorname{co}(F|_I)(c+)}{x-c} \\ &\leq \partial_+ \operatorname{co}(F|_I)(x) \leq \frac{\operatorname{co}(F|_I)(d-) - \operatorname{co}(F|_I)(x)}{d-x} = \frac{F(d) - \operatorname{co}(F|_I)(x)}{d-x} \leq K, \end{aligned}$$

which proves ((a)). $\square$

Lemma 6.5.3. *Let $0 \le a < b$, $F : [0,a) \to \mathbb{R}$ and $G : [a,b) \to \mathbb{R}$. Then*

$$\mathrm{co}(\mathrm{co}(F)\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)}) = \mathrm{co}(F\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)}).$$

Proof of Lemma 6.5.3. Let $u = \mathrm{co}(\mathrm{co}(F)\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)})$ and $v = \mathrm{co}(F\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)})$. The function u is convex and satisfies $u \le \mathrm{co}(F)\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)} \le F\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)}$ so by definition of the convex hull, $u \le v$.

Conversely, $v|_{[a,b)} \le G$ and $v|_{[0,a)} \le F$. Since $v|_{[0,a)}$ is convex, the latter inequality implies $v|_{[0,a)} \le \mathrm{co}(F)$, hence $v \le \mathrm{co}(F)\mathbb{1}_{[0,a)} + G\mathbb{1}_{[a,b)}$. Since v is convex, we get by definition of the convex hull that $v \le u$, which proves the equality. $\square$

Proposition 6.5.4. *Let f and g be real-valued càdlàg functions on $[0,1]$ with respective antiderivatives F and G. We have*

$$\left\|\partial_+ \left(\mathrm{co}(F) - \mathrm{co}(G)\right)\right\|_r \le \left\|f - g\right\|_r. \tag{6.5.13}$$

Proof of Proposition 6.5.4. Assume for a moment that (6.5.13) holds true for antiderivatives of piecewise constant functions. Let $n \in \mathbb{N}^*$. Since f and g are càdlàg, there exist $p \in \mathbb{N}^*$, $0 = a_0 < a_1 < \cdots < a_p = 1$ and two piecewise constant functions $f_n : [0,1] \to \mathbb{R}$ and $g_n : [0,1] \to \mathbb{R}$ which coincide respectively with f and g on $\{a_0, \cdots, a_p\}$, are constant on $[a_k, a_{k+1})$ for $k \in \{0, \cdots, p-1\}$ and satisfy

$$\sup_{x\in\mathbb{R}} \left(|f(x) - f_n(x)| + |g(x) - g_n(x)|\right) \le \frac{1}{n}\cdot$$

In particular, we have that $\|f - f_n\|_r \to 0$ and $\|g - g_n\|_r \to 0$. Let $F_n(v) = F(0) + \int_0^v f_n(u)\,du$ and $G_n(v) = G(0) + \int_0^v g_n(u)\,du$ where $v \in [0,1]$ and $n \in \mathbb{N}$. Since $F \le F_n + \frac{1}{n}$, we have $\mathrm{co}(F) \le \mathrm{co}(F_n) + \frac{1}{n}$, and similarly $\mathrm{co}(F_n) \le \mathrm{co}(F) + \frac{1}{n}$. By the same reasoning for G and G_n, we deduce that

$$\sup_{x\in\mathbb{R}} \left(|\mathrm{co}(F)(x) - \mathrm{co}(F_n)(x)| + |\mathrm{co}(G)(x) - \mathrm{co}(G_n)(x)|\right) \le \frac{2}{n}\cdot$$

Thus, $\mathrm{co}(F_n)$ and $\mathrm{co}(G_n)$ converge to $\mathrm{co}(F)$ and $\mathrm{co}(G)$ respectively. Moreover, since f, g are càdlàg on a compact interval they are uniformly bounded. Thus, there is a constant such that $\|g_n\|_\infty \vee \|f_n\|_\infty \le K$ for all $n \in \mathbb{N}$. By Lemma 6.5.2 (a) we find that for all $n \in \mathbb{N}$, $\partial_+ \mathrm{co}(F), \partial_+ \mathrm{co}(G), \partial_+ \mathrm{co}(F_n), \partial_+ \mathrm{co}(G_n) \in [-K, K]$. By monotonicity and therefore continuity almost everywhere of their right-hand derivatives, $\mathrm{co}(F)$ and $\mathrm{co}(G)$ are differentiable almost everywhere. Because $\mathrm{co}(F_n) \to \mathrm{co}(F)$ and $\mathrm{co}(G_n) \to \mathrm{co}(G)$ pointwise, we have by [108, Theorem 6.2.7] that $\partial_+ \mathrm{co}(F_n) \to \partial_+ \mathrm{co}(F)$ and $\partial_+ \mathrm{co}(G_n) \to \partial_+ \mathrm{co}(G)$ almost everywhere on $[0,1]$. By dominated convergence we get

$$\lim_{n\to+\infty} \left\|\partial_+(\mathrm{co}(F) - \mathrm{co}(F_n))\right\|_r = 0 = \lim_{n\to+\infty} \left\|\partial_+(\mathrm{co}(G) - \mathrm{co}(G_n))\right\|_r.$$

Applying then the triangle inequality reveals

$$\left\|\partial_+(\mathrm{co}(F) - \mathrm{co}(G))\right\|_r \le \lim_{n\to+\infty} \left(\left\|\partial_+(\mathrm{co}(F) - \mathrm{co}(F_n))\right\|_r + \left\|\partial_+(\mathrm{co}(F_n) - \mathrm{co}(G_n))\right\|_r\right.$$

$$
\begin{aligned}
&\quad + \|\partial_+(\mathrm{co}(G_n) - \mathrm{co}(G))\|_r) \\
&\leq \lim_{n\to+\infty} \|f_n - g_n\|_r \leq \lim_{n\to+\infty} (\|f_n - f\|_r + \|f-g\|_r + \|g-g_n\|_r) \\
&= \|f-g\|_r \,,
\end{aligned}
$$

which shows the statement.

We now prove the assertion assuming that f and g are piecewise constant. Let then $(a_k)_{0\leq k\leq n}$ be a partition of $[0,1]$ adapted to f and g, i.e. $0 = a_0 < \ldots < a_n = 1$ and for all $k \in \{0, \cdots, n-1\}$, $f|_{(a_k,a_{k+1})}$ and $g|_{(a_k,a_{k+1})}$ are constant. For $k \in \{0,\cdots,n\}$, we consider the functions

$$
F^k : x \mapsto \begin{cases} \mathrm{co}(F|_{[0,a_k)})(x) & \text{if } x \in [0,a_k), \\ F(x) & \text{else;} \end{cases} \qquad G^k : x \mapsto \begin{cases} \mathrm{co}(G|_{[0,a_k)})(x) & \text{if } x \in [0,a_k), \\ G(x) & \text{else,} \end{cases}
$$

and we denote $f^k = \partial_+ F^k$ and $g^k = \partial_+ G^k$.

Note that $F^0 = F^1 = F$, $G^0 = G^1 = G$ and $F^n = \mathrm{co}(F)$, $G^n = \mathrm{co}(G)$. We will show by induction that $\|f^{k+1} - g^{k+1}\|_r \leq \|f^k - g^k\|_r$ for $k \in \{0,\cdots,n-1\}$. As the initial case is trivial, we assume that the assumption holds for $0 \leq k \leq n-2$. We have

$$
\begin{aligned}
\left\|f^{k+1} - g^{k+1}\right\|_r^r &= \left\|(f^{k+1} - g^{k+1})|_{[0,a_{k+1})}\right\|_r^r + \left\|(f^{k+1} - g^{k+1})|_{[a_{k+1},1]}\right\|_r^r \\
&= \left\|\partial_+(\mathrm{co}(F|_{[0,a_{k+1})}) - \mathrm{co}(G|_{[0,a_{k+1})})\right\|_r^r + \left\|(f^k - g^k)|_{[a_{k+1},1]}\right\|_r^r .
\end{aligned} \tag{6.5.14}
$$

Applying Lemma 6.5.3 with $a = a_k$, $b = a_{k+1}$ and the maps $F|_{[0,a_{k+1})}$ and $G|_{[a_k,a_{k+1})}$ yields

$$
\mathrm{co}(F|_{[0,a_{k+1})}) = \mathrm{co}(\mathrm{co}(F|_{[0,a_k)})\mathbb{1}_{[0,a_k)} + F\mathbb{1}_{[a_k,a_{k+1})}) = \mathrm{co}(F^k|_{[0,a_{k+1})}).
$$

Similarly, we have $\mathrm{co}(G|_{[0,a_{k+1})}) = \mathrm{co}(G^k|_{[0,a_{k+1})})$. By Lemma 6.5.2 (b), F^k and G^k are continuous at a_k and therefore on $[0, a_{k+1})$. Moreover they are convex on $[0,a_k)$ and affine on $[a_k, a_{k+1})$, so we can apply Lemma 6.5.1 with $a = a_k$, $b = a_{k+1}$ and the maps F^k and G^k to get

$$
\begin{aligned}
\left\|\partial_+(\mathrm{co}(F|_{[0,a_{k+1})}) - \mathrm{co}(G|_{[0,a_{k+1})}))\right\|_r &= \left\|\partial_+(\mathrm{co}(F^k|_{[0,a_{k+1})}) - \mathrm{co}(G^k|_{[0,a_{k+1})}))\right\|_r \\
&\leq \left\|(f^k - g^k)|_{[0,a_{k+1})}\right\|_r .
\end{aligned} \tag{6.5.15}
$$

We deduce from (6.5.14) and (6.5.15) that

$$
\left\|f^{k+1} - g^{k+1}\right\|_r \leq \left\|f^k - g^k\right\|_r .
$$

Hence, we have shown that

$$
\|\partial_+(\mathrm{co}(F) - \mathrm{co}(G))\|_r = \|f^n - g^n\|_r \leq \left\|f^0 - g^0\right\|_r = \|f-g\|_r \,,
$$

which shows (6.5.13) for antiderivatives of piecewise constant functions and completes the proof. □

Proof of Theorem 6.2.18. Fix $\mu, \mu', \nu, \nu' \in \mathcal{P}^r(\mathbb{R})$ and assume for a moment that the statement holds for compactly supported measures $\mu_n, \mu'_n, \nu_n, \nu'_n \in \mathcal{P}^r(\mathbb{R})$, $n \in \mathbb{N}$ where

$$\lim_{n\to+\infty} \left(\mathcal{W}_r(\mu,\mu_n) + \mathcal{W}_r(\mu',\mu'_n) + \mathcal{W}_r(\nu,\nu_n) + \mathcal{W}_r(\nu',\nu'_n)\right) = 0. \tag{6.5.16}$$

Let Θ be the set of all even and strictly convex functions $\theta : \mathbb{R} \to [1,+\infty)$ such that $\mu(\theta) + \nu(\theta) + \sup_{n\in\mathbb{N}}(\mu_n(\theta) + \nu_n(\theta)) < +\infty$. For all $\theta \in \Theta$, let $\mathcal{W}_\theta : \mathcal{P}_\theta(\mathbb{R}) \times \mathcal{P}_\theta(\mathbb{R}) \to \mathbb{R}$ be defined for all $(\eta,\tau) \in \mathcal{P}_\theta(\mathbb{R}) \times \mathcal{P}_\theta(\mathbb{R})$ by

$$\mathcal{W}_\theta(\eta,\tau) = \inf_{\pi\in\Pi(\eta,\tau)} \int_{\mathbb{R}\times\mathbb{R}} \theta\left(\frac{x-y}{2}\right) \pi(dx,dy). \tag{6.5.17}$$

Note that for all $(\eta,\tau) \in \mathcal{P}_\theta(\mathbb{R}) \times \mathcal{P}_\theta(\mathbb{R})$ and $\pi \in \Pi(\eta,\tau)$,

$$\int_{\mathbb{R}\times\mathbb{R}} \theta\left(\frac{x-y}{2}\right) \pi(dx,dy) \le \int_{\mathbb{R}\times\mathbb{R}} \frac{1}{2}(\theta(x)+\theta(-y))\,\pi(dx,dy) = \frac{1}{2}\eta(\theta) + \frac{1}{2}\tau(\theta) < +\infty, \tag{6.5.18}$$

hence $\mathcal{W}_\theta(\eta,\tau) < +\infty$. We deduce by [19, Theorem 1.2], resp [19, Theorem 1.4] that for all $\theta \in \Theta$, $\mathcal{I}(\mu,\nu)$ and $\mathcal{J}(\mu,\nu)$ coincide respectively with the unique minimiser of

$$\inf_{\eta\in\mathcal{P}_\theta(\mathbb{R}),\ \eta\le_{cx}\nu} \mathcal{W}_\theta(\mu,\eta), \quad \text{and} \quad \inf_{\eta\in\mathcal{P}_\theta(\mathbb{R}),\ \mu\le_{cx}\eta} \mathcal{W}_\theta(\eta,\nu), \tag{6.5.19}$$

which by [19, (1.2) and (1.7)] have a common value $V_\theta(\mu,\nu)$. Let us show that $(\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ and $(\mathcal{J}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ respectively converge to $\mathcal{I}(\mu,\nu)$ and $\mathcal{J}(\mu,\nu)$ in $\mathcal{W}_r$ as $n \to +\infty$. On the probability space $(0,1)$ endowed with the Lebesgue measure, by (6.5.16) the random variables $F_{\mu_n}^{-1}$ and $F_{\nu_n}^{-1}$ converge in L^r to F_μ^{-1} and F_ν^{-1} respectively, hence $(|F_{\mu_n}^{-1}|^r)_{n\in\mathbb{N}}$, $|F_\mu^{-1}|^r$, $(|F_{\nu_n}^{-1}|^r)_{n\in\mathbb{N}}$ and $|F_\nu^{-1}|^r$ are uniformly integrable. By the de La Vallée Poussin theorem, there exists an increasing and strictly convex map $h : \mathbb{R}_+ \to [1,+\infty)$ such that for all $t \in \mathbb{R}_+$, $h(t) \ge t$, $\frac{h(t)}{t}$ goes to $+\infty$ with t and $\sup_{n\in\mathbb{N}} \int_\mathbb{R} h(1+|x|^r)\,(\mu_n+\nu_n+\mu+\nu)(dx) < +\infty$. Denoting by $(\mathcal{F}^r)^+(\mathbb{R})$ the set $\mathcal{F}_g^+(\mathbb{R})$ defined by (6.4.2) with $x \mapsto 1+|x|^r$ replacing g, we deduce that the map $f : x \mapsto h(1+|x|^r)$ belongs to $\Theta \cap (\mathcal{F}^r)^+(\mathbb{R})$.

For all $\lambda > 0$ and $\eta \in \mathcal{P}(\mathbb{R})$, we denote by $(\frac{\cdot}{\lambda})_*\eta$ the image of η by $x \mapsto \frac{x}{\lambda}$. Using the inequality $f\left(\frac{x-z}{4}\right) \le \frac{1}{2}f\left(\frac{x-y}{2}\right) + \frac{1}{2}f\left(\frac{y-z}{2}\right)$ valid for all $x,y,z \in \mathbb{R}$ for the first inequality, we get

$$\begin{aligned}
\left((\frac{\cdot}{4})_*\mathcal{I}(\mu_n,\nu_n)\right)(f) &= \mathcal{W}_f\left(\delta_0, (\frac{\cdot}{2})_*\mathcal{I}(\mu_n,\nu_n)\right) \\
&\le \frac{1}{2}\mathcal{W}_f(\delta_0,\mu_n) + \frac{1}{2}\mathcal{W}_f(\mu_n,\mathcal{I}(\mu_n,\nu_n)) \\
&\le \frac{1}{2}\mathcal{W}_f(\delta_0,\mu_n) + \frac{1}{2}\mathcal{W}_f(\mu_n,\nu_n) \\
&\le \frac{1}{4}\theta(0) + \frac{1}{2}\mu_n(f) + \frac{1}{4}\nu_n(f).
\end{aligned}$$

We deduce the existence of $R > 0$ such that for all $n \in \mathbb{N}$, $(\frac{\cdot}{4})_*\mathcal{I}(\mu_n,\nu_n)$ belongs to $B_R := \{p \in \mathcal{P}(\mathbb{R}) \mid p(f) < +\infty\}$. Since $f \in (\mathcal{F}^r)^+(\mathbb{R})$, we have by Lemma 6.7.7 below that B_R

is compact for the $\mathcal{W}_r$-distance topology. Therefore the sequence $((\frac{\cdot}{4})_*\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ admits a $\mathcal{W}_r$-accumulation point. This represents no challenge to show that the image by $x\mapsto 4x$ of this accumulation point is a $\mathcal{W}_r$-accumulation point of the sequence $(\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$. With a similar reasoning we show that the sequence $(\mathcal{J}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ admits a $\mathcal{W}_r$-accumulation point as well.

We now choose θ even, strictly convex and dominated by a multiple of $x\mapsto(1+|x|^r)$. Let $(\zeta_{n'})_{n\in\mathbb{N}}$ and $(\xi_{n'})_{n\in\mathbb{N}}$ be subsequences of $(\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ and $(\mathcal{J}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ converging to ζ and ξ in $\mathcal{W}_r$, respectively. Then by [19, Theorem 1.5],

$$\begin{aligned}\mathcal{W}_\theta(\mu,\zeta)&=\lim_{n\to+\infty}\mathcal{W}_\theta(\mu_{n'},\zeta_{n'})=\lim_{n\to+\infty}V_\theta(\mu_{n'},\nu_{n'})=V_\theta(\mu,\nu),\\ \mathcal{W}_\theta(\xi,\nu)&=\lim_{n\to+\infty}\mathcal{W}_\theta(\xi_{n'},\nu_{n'})=\lim_{n\to+\infty}V_\theta(\mu_{n'},\nu_{n'})=V_\theta(\mu,\nu).\end{aligned}$$

Moreover, for all $n\in\mathbb{N}$, $\mathcal{I}(\mu_{n'},\nu_{n'})\le_{cx}\nu_{n'}$ and $\mu_{n'}\le_{cx}\mathcal{J}(\mu_{n'},\nu_{n'})$, so by $\mathcal{W}_r$-convergence we have $\zeta\le_{cx}\nu$ and $\mu\le_{cx}\xi$, and both measures are therefore admissible for (6.5.19). But, for strictly convex θ the optimisers of (6.5.19) are unique. Thus, $(\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ is not only $\mathcal{W}_r$-relatively compact, but also has only a single accumulation point, $\mathcal{I}(\mu,\nu)$. This means in other words that $(\mathcal{I}(\mu_n,\nu_n))_{n\in\mathbb{N}}$ converges in $\mathcal{W}_r$ to $\mathcal{I}(\mu,\nu)$. We also deduce that $\mathcal{J}(\mu^n,\nu^n)\to\mathcal{J}(\mu,\nu)$ in $\mathcal{W}_r$ as $n\to+\infty$. Analogously we obtain that $\mathcal{I}(\mu_n',\nu_n')\to\mathcal{I}(\mu',\nu')$ and $\mathcal{J}(\mu_n',\nu_n')\to\mathcal{J}(\mu',\nu')$ in $\mathcal{W}_r$ as $n\to+\infty$. Therefore,

$$\begin{aligned}\mathcal{W}_r\left(\mathcal{I}(\mu,\nu),\mathcal{I}(\mu',\nu')\right)&\le\lim_{n\to+\infty}\left(\mathcal{W}_r\left(\mathcal{I}(\mu,\nu),\mathcal{I}(\mu_n,\nu_n)\right)+\mathcal{W}_r\left(\mathcal{I}(\mu_n,\nu_n),\mathcal{I}(\mu_n',\nu_n')\right)\right.\\ &\qquad\left.+\mathcal{W}_r\left(\mathcal{I}(\mu_n',\nu_n'),\mathcal{I}(\mu',\nu')\right)\right)\\ &=\lim_{n\to+\infty}\mathcal{W}_r\left(\mathcal{I}(\mu_n,\nu_n),\mathcal{I}(\mu_n',\nu_n')\right)\\ &\le\lim_{n\to+\infty}\left(2\mathcal{W}_r(\mu_n,\mu_n')+\mathcal{W}_r(\nu_n,\nu_n')\right)\\ &=2\mathcal{W}_r(\mu,\mu')+\mathcal{W}_r(\nu,\nu'),\end{aligned}$$

and similarly $\mathcal{W}_r\left(\mathcal{J}(\mu,\nu),\mathcal{J}(\mu',\nu')\right)\le\mathcal{W}_r(\mu,\mu')+2\mathcal{W}_r(\nu,\nu')$.

Since compactly supported measures are dense in $\mathcal{P}^r(\mathbb{R})$ (to see it, consider for $\eta\in\mathcal{P}^r(\mathbb{R})$ the image of η by $x\mapsto(-R)\vee(x\wedge R)$, where $R>0$ is arbitrarily large), the preceeding allows us to only consider compactly supported $\mu,\mu',\nu,\nu'\in\mathcal{P}^r(\mathbb{R})$. We define

$$G:v\mapsto\int_0^v(F_\mu^{-1}-F_\nu^{-1})(u)\,du\quad\text{and}\quad G':v\mapsto\int_0^v(F_{\mu'}^{-1}-F_{\nu'}^{-1})(u)\,du.$$

By convexity of $\operatorname{co}(G)$ and $\operatorname{co}(G')$, we have that $\partial_-\operatorname{co}(G)(u)$ and $\partial_-\operatorname{co}(G')(u)$ respectively coincide du-almost everywhere on $(0,1)$ with $\partial_+\operatorname{co}(G)(u)$ and $\partial_+\operatorname{co}(G')(u)$, so by (6.2.17) we have for all du-almost all $u\in(0,1)$

$$\begin{aligned}F^{-1}_{\mathcal{I}(\mu,\nu)}(u)=F_\mu^{-1}(u)-\partial_+\operatorname{co}(G)(u),&\quad F^{-1}_{\mathcal{J}(\mu,\nu)}(u)=F_\nu^{-1}(u)+\partial_+\operatorname{co}(G)(u),\\ F^{-1}_{\mathcal{I}(\mu',\nu')}(u)=F_{\mu'}^{-1}(u)-\partial_+\operatorname{co}(G')(u),&\quad\text{and}\quad F^{-1}_{\mathcal{J}(\mu',\nu')}(u)=F_{\nu'}^{-1}(u)+\partial_+\operatorname{co}(G')(u).\end{aligned}$$

Since F_μ^{-1} is nondecreasing, it has at most countably many discontinuities, and since μ is compactly supported, $F_\mu^{-1}(0+)$ and $F_\mu^{-1}(1-)$ are finite. Denoting $F_\mu^{-1}(1+)=F_\mu^{-1}(1-)$,

F_μ^{-1} coincides therefore du-almost everywhere with the real-valued càdlàg map $[0,1] \ni u \mapsto F_\mu^{-1}(u+)$. Reasoning the same for μ', ν and ν', we deduce that G and G' can be interpreted as antiderivatives of real-valued càdlàg functions on $[0,1]$. Therefore by Proposition 6.5.4 and we get

$$\begin{aligned}\|\partial_+ (\mathrm{co}(G) - \mathrm{co}(G'))\|_r \leq \|\partial_+(G - G')\|_r &\leq \left\|F_\mu^{-1} - F_\nu^{-1} - F_{\mu'}^{-1} + F_{\nu'}^{-1}\right\|_r \\ &\leq \mathcal{W}_r(\mu,\mu') + \mathcal{W}_r(\nu,\nu').\end{aligned}$$

By the latter inequality and Minkowski's inequality we then have

$$\begin{aligned}\mathcal{W}_r(\mathcal{I}(\mu,\nu),\mathcal{I}(\mu',\nu')) &= \left(\int_0^1 |F_{\mathcal{I}(\mu,\nu)}^{-1}(u) - F_{\mathcal{I}(\mu',\nu')}^{-1}(u)|^r \, du\right)^{\frac{1}{r}} \\ &= \left\|F_\mu^{-1} - \partial_+ \mathrm{co}(G) - F_{\mu'}^{-1} + \partial_+ \mathrm{co}(G')\right\|_r \\ &\leq \left\|F_\mu^{-1} - F_{\mu'}^{-1}\right\|_r + \|\partial_+ (\mathrm{co}(G) - \mathrm{co}(G'))\|_r \\ &\leq 2\mathcal{W}_r(\mu,\mu') + \mathcal{W}_r(\nu,\nu'),\end{aligned}$$

and

$$\begin{aligned}\mathcal{W}_r(\mathcal{J}(\mu,\nu),\mathcal{J}(\mu',\nu')) &= \left(\int_0^1 |F_{\mathcal{J}(\mu,\nu)}^{-1}(u) - F_{\mathcal{J}(\mu',\nu')}^{-1}(u)|^r \, du\right)^{\frac{1}{r}} \\ &= \left\|F_\nu^{-1} + \partial_+ \mathrm{co}(G) - F_{\nu'}^{-1} - \partial_+ \mathrm{co}(G')\right\|_r \\ &\leq \left\|F_\nu^{-1} - F_{\nu'}^{-1}\right\|_r + \|\partial_+ (\mathrm{co}(G) - \mathrm{co}(G'))\|_r \\ &\leq \mathcal{W}_r(\mu,\mu') + 2\mathcal{W}_r(\nu,\nu').\end{aligned}$$

□

6.6 Convergence in an extended space of martingale couplings

We recall that a pair (μ,ν) of finite positive measures in convex order is called irreducible if $I = \{u_\mu < u_\nu\}$ is an interval and, $\mu(I)$ and $\nu(\overline{I})$ have the total mass. If $a \in \mathbb{R}$ is such that $\nu([a,+\infty)) = 0$, then the convex order implies $\mu([a,+\infty)) = 0$, hence

$$u_\mu(a) = a - \int_{\mathbb{R}} x\,\mu(dx) = a - \int_{\mathbb{R}} y\,\nu(dy) = u_\nu(a),$$

so $a \notin I$. Similarly, $\nu((-\infty,a]) = 0 \implies a \notin I$. We deduce that ν must assign positive mass to any neighbourhood of each of the boundaries of I.

According to [25, Theorem A.4], for any pair (μ,ν) of probability measures in convex order, there exist $N \subset \mathbb{N}$ and a sequence $(\mu_n,\nu_n)_{n\in N}$ of irreducible pairs of sub-probability measures in convex order such that

$$\mu = \eta + \sum_{n\in N} \mu_n, \quad \nu = \eta + \sum_{n\in N} \nu_n \quad \text{and} \quad \{u_\mu < u_\nu\} = \bigcup_{n\in N} \{u_{\mu_n} < u_{\nu_n}\},$$

where the union is disjoint and $\eta = \mu|_{\{u_\mu = u_\nu\}}$. The sequence $(\mu_n, \nu_n)_{n\in N}$ is unique up to rearrangement of the pairs and is called the decomposition of (μ, ν) into irreducible components. Moreover, for any martingale coupling $\pi \in \Pi_M(\mu, \nu)$, there exists a unique sequence of martingale couplings $\pi_n \in \Pi_M(\mu_n, \nu_n)$, $n \in N$ such that

$$\pi = \chi + \sum_{n\in N} \pi_n,$$

where $\chi = (\mathrm{id}, \mathrm{id})_* \eta$ and $*$ denotes the pushforward operation. This sequence satisfies

$$\forall n \in N, \quad \pi_n(dx, dy) = \mu_n(dx)\, \pi_x(dy). \tag{6.6.1}$$

The following assertions can be found in [25, Section 4]: for all $(m_0, m_1) \in \mathbb{R}_+^* \times \mathbb{R}$, there is a one-to-one correspondence between finite positive measures μ with mass m_0 such that $\int_{\mathbb{R}} y\, \mu(dy) = m_1$ and the set of functions $u\colon \mathbb{R} \to \mathbb{R}^+$ which satisfy

(i) u is convex;

(ii) $u(y) - m_0|y - m_1|$ goes to 0 as $|y|$ tends to $+\infty$.

Any function which suffices (i) and (ii) is then called a potential function. The potential function of μ is denoted by

$$u_\mu(y) = \int_{\mathbb{R}} |y - x|\, \mu(dx).$$

A sequence $(\mu^k)_{k\in\mathbb{N}}$ of finite positive measures with equal mass on the line converges in $\mathcal{W}_1$ to μ iff the sequence of potential functions $(u_{\mu^k})_{k\in\mathbb{N}}$ converges pointwise to u_μ. In that case, since for all $y \in \mathbb{R}$ the map $x \mapsto |y - x|$ is Lipschitz continuous with constant 1, we have by Kantorovich and Rubinstein's duality theorem that

$$\sup_{y\in\mathbb{R}} |u_{\mu^k}(y) - u_\mu(y)| \le \mathcal{W}_1(\mu^k, \mu) \to 0, \quad k \to +\infty,$$

hence we even have uniform convergence on $\mathbb{R}$ of potential functions.

In dimension one, for all $m_1 \in \mathbb{R}$, the set of all finite positive measures with mean m_1 is a lattice [120, Proposition 1.6], and even a complete lattice [121]. Then all $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ with mean m_1 have a supremum, denoted $\mu \vee_{cx} \nu$, and an infimum, denoted $\mu \wedge_{cx} \nu$, with respect to the convex order. In that context it is convenient to work with potential functions since they provide simple characterisations of those bounds:

$$\mu \vee_c \nu \text{ is defined as the measure with potential function } u_\mu \vee u_\nu,$$
$$\mu \wedge_c \nu \text{ is defined as the measure with potential function } \operatorname{co}(u_\mu \wedge u_\nu),$$

where co is the convex hull.

Before proving Theorem 6.2.20, consider $f, g \in \mathcal{F}^1(\mathbb{R})$. Note that convergence in $\mathcal{P}_{f\oplus\hat{g}}(\mathbb{R}\times \mathcal{P}(\mathbb{R}))$ is equivalent to weak convergence coupled with the convergence of the $f \oplus \hat{g}$-integrals, whereby we can suppose without loss of generality that $f(x) = g(x) = 1 + |x|$. Indeed, if there exists $P^k \in \Lambda_M(\mu^k, \nu^k)$, $k \in \mathbb{N}$, converging to P in $\mathcal{W}_1$, then it converges weakly to P, and

$$P^k(f \oplus \hat{g}) = \hat{I}(P^k)(f \oplus g) = \mu^k(f) + \nu^k(g) \to \mu(f) + \nu(g) = \hat{I}(P)(f \oplus g) = P(f \oplus \hat{g}),$$

as k tends to $+\infty$.

We can also suppose without loss of generality that P is a finite sum of elements of $J(\mathcal{P}(\mathbb{R} \times \mathbb{R}))$, and that (μ, ν) is irreducible, as shown in the next two lemmas.

Lemma 6.6.1. *Let $f, g : x \mapsto 1 + |x|$. Assume that the conclusion of Theorem 6.2.20 holds for any $P \in \Lambda_M(\mu, \nu)$ of the form*

$$P = \frac{1}{n} \sum_{j=1}^{n} J(\pi^j), \tag{6.6.2}$$

where $n \in \mathbb{N}^*$ and $\pi^1, \cdots, \pi^n \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ are all martingale couplings with first marginal μ. Then it holds for any $P \in \Lambda_M(\mu, \nu)$.

Proof of Lemma 6.6.1. Fix $P = \mu \times P_x \in \Lambda_M(\mu, \nu)$. First we show that P can be splitted into a family $(P^t)_{t \in (0,1)} \in \mathcal{P}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))^{(0,1)}$ such that $t \mapsto P^t$ is measurable, for each $t \in (0, 1)$, P^t is concentrated on the graph of a function, i.e. there exists $\pi^t \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ with first marginal μ which satisfies $J(\pi^t) = P^t$, and

$$P(dx, dp) = \int_{t=0}^{1} P^t(dx, dp)\, dt = \int_{t=0}^{1} J(\pi^t)(dx, dp)\, dt. \tag{6.6.3}$$

Let λ denote the Lebesgue measure on $(0, 1)$. Since $\mathcal{P}^1(\mathbb{R})$ is Polish, there exists by [113, Lemma 3.22] a measurable map $h : \mathbb{R} \times (0, 1) \to \mathcal{P}^1(\mathbb{R})$ such that

$$P = ((x, t) \mapsto (x, h(x, t)))_*(\mu \otimes \lambda). \tag{6.6.4}$$

For all $t \in (0, 1)$, let then

$$P^t(dx, dp) = \mu(dx)\, \delta_{h(x,t)}(dp).$$

The fact that h is measurable ensures that P^t is well defined and $t \mapsto P^t$ is measurable, and by setting $\pi^t(dx, dy) = \mu(dx)\, h(x, t)(dy)$ we have $P^t = J(\pi^t)$. Moreover, for all nonnegative measurable map $f : \mathbb{R} \times \mathcal{P}(\mathbb{R}) \to \mathbb{R}$ we have using (6.6.4) for the last equality that

$$\begin{aligned}
\int_{t=0}^{1} \int_{\mathbb{R} \times \mathcal{P}(\mathbb{R})} f(x, p)\, P^t(dx, dp)\, dt &= \int_{\mathbb{R} \times \mathcal{P}(\mathbb{R}) \times (0,1)} f(x, p)\, \mu(dx)\, \delta_{h(x,t)}(dp)\, dt \\
&= \int_{\mathbb{R} \times (0,1)} f(x, h(x, t))\, \mu(dx)\, dt \\
&= \int_{\mathbb{R} \times \mathcal{P}(\mathbb{R})} f(x, p)\, P(dx, dp),
\end{aligned}$$

which proves (6.6.3). For $n \in \mathbb{N}^*$, let

$$P_n = \frac{1}{n}\sum_{j=1}^n J\left(n\int_{\frac{j-1}{n}}^{\frac{j}{n}} \pi^t\,dt\right) = \frac{1}{n}\sum_{j=1}^n \mu \times \delta_{n\int_{\frac{j-1}{n}}^{\frac{j}{n}} \pi^t_x\,dt}.$$

By (6.6.3) and definition of P_n, we have for all measurable nonnegative or with at most linear growth function $f : \mathbb{R}\times\mathcal{P}^1(\mathbb{R}) \to \mathbb{R}$ that

$$\begin{aligned} &\int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})} f(x,p)\,P(dx,dp) = \int_{\mathbb{R}\times(0,1)} f(x,\pi^t_x)\,\mu(dx)\,dt \\ \text{and}\quad &\int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})} f(x,p)\,P_n(dx,dp) = \frac{1}{n}\sum_{j=1}^n\int_{\mathbb{R}} f\left(x, n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi^t_x\,dt\right)\mu(dx) \end{aligned} \tag{6.6.5}$$

Applying (6.6.5) with $f : (x,p) \mapsto p$ and $f : (x,p) \mapsto |x - \int_{\mathbb{R}} y\,p(dy)|$ yields

$$\begin{aligned} \int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})} p\,P_n(dx,dp) &= \frac{1}{n}\sum_{j=1}^n\int_{\mathbb{R}} n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi^t_x\,dt\,\mu(dx) = \int_{\mathbb{R}\times(0,1)}\pi^t_x\,\mu(dx)\,dt \\ &= \int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})} p\,P(dx,dp) = \nu, \end{aligned}$$

and

$$\begin{aligned} \int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})}\left|x - \int_{\mathbb{R}} y\,p(dy)\right| P_n(dx,dp) &= \frac{1}{n}\sum_{j=1}^n\int_{\mathbb{R}}\left|x - n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\int_{\mathbb{R}} y\,\pi^t_x(dy)\,dt\right|\mu(dx) \\ &= \sum_{j=1}^n\int_{\mathbb{R}}\left|\int_{\frac{j-1}{n}}^{\frac{j}{n}}\left(x - \int_{\mathbb{R}} y\,\pi^t_x(dy)\right)dt\right|\mu(dx) \\ &\le \sum_{j=1}^n\int_{\mathbb{R}}\int_{\frac{j-1}{n}}^{\frac{j}{n}}\left|x - \int_{\mathbb{R}} y\,\pi^t_x(dy)\right| dt\,\mu(dx) \\ &= \int_{\mathbb{R}\times(0,1)}\left|x - \int_{\mathbb{R}} y\,\pi^t_x(dy)\right|\mu(dx)\,dt = 0 \\ &= \int_{\mathbb{R}\times\mathcal{P}^1(\mathbb{R})}\left|x - \int_{\mathbb{R}} y\,p(dy)\right| P(dx,dp) = 0, \end{aligned}$$

hence for all $j \in \{1,\cdots,n\}$, $n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi^t\,dt$ is a martingale coupling and $P_n \in \Lambda_M(\mu,\nu)$. Moreover we show that P_n converges to P in $\mathcal{W}_1$ for $n \to \infty$. This can be achieved by means of the modulus of continuity [77]:

$$\begin{aligned} \mathcal{W}_1(P,P_n) &= \mathcal{W}_1\left(\frac{1}{n}n\sum_{j=1}^n\int_{\frac{j-1}{n}}^{\frac{j}{n}} J(\pi^t)\,dt, \frac{1}{n}\sum_{j=1}^n J\left(n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi^t\,dt\right)\right) \\ &\le \frac{1}{n}\sum_{j=1}^n\mathcal{W}_1\left(n\int_{\frac{j-1}{n}}^{\frac{j}{n}} J(\pi^t)\,dt, J\left(n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi^t\,dt\right)\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n}\mathcal{W}_1\left(\mu\times n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\delta_{\pi_x^t}\,dt, \mu\times\delta_{n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi_x^t\,dt}\right)\\
&\le \frac{1}{n}\sum_{j=1}^{n}\int_{x\in\mathbb{R}}\mathcal{W}_1\left(n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\delta_{\pi_x^t}\,dt, \delta_{n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi_x^t\,dt}\right)\mu(dx)\\
&= \frac{1}{n}\int_{x\in\mathbb{R}}\sum_{j=1}^{n}n\int_{t=\frac{j-1}{n}}^{\frac{j}{n}}\mathcal{W}_1\left(\pi_x^t, n\int_{\frac{j-1}{n}}^{\frac{j}{n}}\pi_x^s\,ds\right)dt\,\mu(dx)\\
&\le \int_{x\in\mathbb{R}}\sum_{j=1}^{n}\int_{t=\frac{j-1}{n}}^{\frac{j}{n}}\int_{s=\frac{j-1}{n}}^{\frac{j}{n}}n\mathcal{W}_1\left(\pi_x^t,\pi_x^s\right)ds\,dt\,\mu(dx).
\end{aligned}
$$

Denote by λ the Lebesgue measure on $(0,1)$ and by ω_x^1 the 1-modulus of continuity of $\mathbb{1}_{(0,1)}(t)\,dt\,\delta_{\pi_x^t}(dp)\in\mathcal{P}^1((0,1)\times\mathcal{P}^1(\mathbb{R}))$, that is

$$
\omega_x^1(\delta)=\sup\left\{\int_{(0,1)^2}\mathcal{W}_1(\pi_x^t,\pi_x^s)\,\chi(dt,ds)\mid\chi\in\Pi(\lambda,\lambda)\text{ and }\int_{(0,1)^2}|t-s|\,\chi(dt,ds)\le\delta\right\}.
$$

Let

$$
\chi_n(dt,ds)=dt\,n\mathbb{1}_{\left[\frac{\lfloor nt\rfloor}{n},\frac{\lceil nt\rceil}{n}\right]}(s)\,ds=n\sum_{j=1}^{n}\mathbb{1}_{\left[\frac{j-1}{n},\frac{j}{n}\right]^2}(t,s)\,dt\,ds.
$$

We compute $n^2\int_{\left(\frac{j-1}{n},\frac{j}{n}\right)^2}|t-s|\,dt\,ds=\frac{1}{3n}$, so that $\int_{(0,1)^2}|t-s|\,\chi_n(dt,ds)=\frac{1}{3n}$. Then by definition of ω_x^1 we have

$$
\sum_{j=1}^{n}\int_{t=\frac{j-1}{n}}^{\frac{j}{n}}\int_{s=\frac{j-1}{n}}^{\frac{j}{n}}n\mathcal{W}_1\left(\pi_x^t,\pi_x^s\right)ds\,dt\le\omega_x^1\left(\frac{1}{3n}\right),
$$

hence $\mathcal{W}_1(P,P_n)\le\int_{\mathbb{R}}\omega_x^1\left(\frac{1}{3n}\right)\mu(dx)$. By [77, Lemma 2.7] we have $\omega_x^1\left(\frac{1}{3n}\right)\searrow 0$ as $n\to\infty$, and note that by the triangle inequality we have

$$
\omega_x^1\left(\frac{1}{3n}\right)\le\sup_{\chi\in\Pi(\lambda,\lambda)}\int_{(0,1)^2}\mathcal{W}_1(\pi_x^t,\pi_x^s)\,\chi(dt,ds)\le 2\int_{\mathbb{R}\times(0,1)}|y|\,\pi_x^t(dy)\,dt. \tag{6.6.6}
$$

By (6.6.5) applied with $f:(x,p)\mapsto\int_{\mathbb{R}}|y|\,p(dy)$, the definition of $\Lambda_M(\mu,\nu)$ and the fact that $\nu\in\mathcal{P}^1(\mathbb{R})$, we get that the right-hand side of (6.6.6) is μ-integrable. Hence, by dominated convergence we have

$$
\mathcal{W}_1(P,P_n)\le\int_{\mathbb{R}}\omega_x^1\left(\frac{\delta}{n}\right)\mu(dx)\searrow 0\quad\text{as }n\to\infty.
$$

By assumption, for all $n\in\mathbb{N}$ there exists a sequence $P_n^k\in\Lambda_M(\mu^k,\nu^k)$, $k\in\mathbb{N}$, such that $P_n^k\to P_n$ in $\mathcal{W}_1$ as $k\to+\infty$. Then there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ of positive integers such that

$$
\forall n\in\mathbb{N},\quad\forall k\ge k_n,\quad\mathcal{W}_1(P_n,P_n^{k_n})\le 2^{-n}.
$$

For $k \in \mathbb{N}$, let $n_k = \max\{n \in \mathbb{N} \mid k \geq k_n\}$, where we set by convention $\max \emptyset = 0$. Since $(k_n)_{n\in\mathbb{N}}$ is strictly increasing, we find that $n_k \to +\infty$ as $k \to +\infty$. Then the sequence

$$P^k = P^k_{n_k} \in \Lambda_M(\mu^k, \nu^k), \quad k \in \mathbb{N}$$

satisfies

$$\mathcal{W}_1(P, P^k) \leq \mathcal{W}_1(P, P_{n_k}) + \mathcal{W}_1(P_{n_k}, P^k_{n_k}) \underset{k\to+\infty}{\longrightarrow} 0.$$

□

Lemma 6.6.2. *Let $f, g : x \mapsto 1 + |x|$. If the conclusion of Theorem 6.2.20 holds for any irreducible pair (μ, ν), then it holds for any pair (μ, ν) in the convex order.*

Proof of Lemma 6.6.2. In the setting of Theorem 6.2.20, fix $P \in \Lambda_M(\mu, \nu)$. Denote by $(\mu_n, \nu_n)_{n\in N}$ the decomposition of (μ, ν) into irreducible components with

$$\mu = \eta + \sum_{n\in N} \mu_n, \quad \nu = \eta + \sum_{n\in N} \nu_n.$$

By Proposition 5.2.4, we can find sub-probability measures $(\eta^k, \upsilon^k)_{k\in\mathbb{N}}$, $(\mu^k_n)_{(k,n)\in\mathbb{N}\times N}$, $(\nu^k_n)_{(k,n)\in\mathbb{N}\times N}$ such that

$$\eta^k \leq_{cx} \upsilon^k, \quad \mu^k_n \leq_{cx} \nu^k_n \quad \forall (k, n) \in \mathbb{N} \times N,$$
$$\eta^k \to \eta, \quad \upsilon^k \to \eta, \quad \mu^k_n \to \mu_n, \quad \nu^k_n \to \nu_n \quad \text{in } \mathcal{W}_1, \quad k \to +\infty.$$

Since η^k and υ^k are in convex order and both converge to η in $\mathcal{W}_1$, there exists by Theorem 6.2.4 a sequence of martingale couplings $\chi^k \in \Pi_M(\eta^k, \upsilon^k)$, $k \in \mathbb{N}$ converging in $\mathcal{AW}_1$ to the only martingale coupling between η and itself, namely $\chi = (\mathrm{id}, \mathrm{id})_*\eta$. Equivalently, $J(\chi^k)$ converges to $J(\chi)$ in $\mathcal{W}_1$ as $k \to +\infty$.

Let $\pi = \mu \times \pi_x$, where $\pi_x = \int_{\mathcal{P}(\mathbb{R})} p\, P_x(dp)$. Then π is a martingale coupling between μ and ν, hence for all $n \in N$, the measure $\pi_n = \mu_n \times \pi_x$ is a martingale coupling between μ_n and ν_n. We deduce that for all $n \in N$, $P_n = \mu_n \times P_x \in \Lambda_M(\mu_n, \nu_n)$. By assumption, we can find for any $n \in N$ a sequence $P^k_n \in \Lambda_M(\mu^k_n, \nu^k_n)$, $k \in \mathbb{N}$, which converges in $\mathcal{W}_1$ to P_n as k goes to $+\infty$. For $k \in \mathbb{N}$, let $P^k = J(\chi^k) + \sum_{n\in N} P^k_n$. Then for all $p \in N$ we have

$$\begin{aligned}\mathcal{W}_1(P^k, P) &= \mathcal{W}_1\left(J(\chi^k) + \sum_{n\in N} P^k_n, J(\chi) + \sum_{n\in N} P_n\right)\\ &\leq \mathcal{W}_1(J(\chi^k), J(\chi)) + \sum_{n\in N,\ n\leq p} \mathcal{W}_1(P^k_n, P_n) + \mathcal{W}_1\left(\sum_{n\in N,\ n>p} P^k_n, \sum_{n\in N,\ n>p} P_n\right).\end{aligned} \tag{6.6.7}$$

By the triangle inequality we have

$$\mathcal{W}_1\left(\sum_{n\in N,\ n>p} P^k_n, \sum_{n\in N,\ n>p} P_n\right) \leq \mathcal{W}_1\left(\sum_{n\in N,\ n>p} P^k_n, \delta_{(0,\delta_0)}\right) + \mathcal{W}_1\left(\delta_{(0,\delta_0)}, \sum_{n\in N,\ n>p} P_n\right). \tag{6.6.8}$$

Let $\varepsilon_p = \sum_{n\in N, n>p} \mu_n(\mathbb{R}) = \sum_{n\in N, n>p} \mu_n^k(\mathbb{R})$ where by convention the sum over an empty set is 0, and for $\zeta \in \mathcal{P}(\mathbb{R})$, let

$$I^1_{\varepsilon_p}(\zeta) := \sup \int_{\mathbb{R}} |x|\, \tau(dx),$$

where the supremum is taken over all positive measures τ on $\mathbb{R}$ such that $\tau \leq \zeta$ and $\tau(\mathbb{R}) \leq \varepsilon_p$. Then by (6.2.21) we have

$$\begin{aligned}\mathcal{W}_1\left(\sum_{n\in N,\ n>p} P_n^k, \delta_{(0,\delta_0)}\right) &= \sum_{n\in N,\ n>p} \int_{\mathbb{R}\times\mathcal{P}^r(\mathbb{R})} \left(|x| + \int_{\mathbb{R}} |y|\, p(dy)\right) P_n^k(dx, dp) \\ &= \sum_{n\in N,\ n>p} \left(\int_{\mathbb{R}} |x|\, \mu_n^k(dx) + \int_{\mathbb{R}} |y|\, \nu_n^k(dy)\right) \\ &\leq I^1_{\varepsilon_p}(\mu^k) + I^1_{\varepsilon_p}(\nu^k).\end{aligned}$$

The same reasoning yields $\mathcal{W}_1\left(\delta_{(0,\delta_0)}, \sum_{n\in N,\ n>p} P_n\right) \leq I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu)$. So by (6.6.7) and (6.6.8) we have

$$\mathcal{W}_1(P^k, P) \leq \mathcal{W}_1(J(\chi^k), J(\chi)) + \sum_{n\in N,\ n\leq p} \mathcal{W}_1(P_n^k, P_n) + I^1_{\varepsilon_p}(\mu^k) + I^1_{\varepsilon_p}(\nu^k) + I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu).$$

By Lemma 5.3.1 (c) we have $I^1_{\varepsilon_p}(\mu^k) + I^1_{\varepsilon_p}(\nu^k) \leq I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu) + \mathcal{W}_1(\mu, \mu^k) + \mathcal{W}_1(\nu, \nu^k)$, from which we deduce

$$\limsup_{k\to+\infty} \mathcal{W}_1(P^k, P) \leq 2(I^1_{\varepsilon_p}(\mu) + I^1_{\varepsilon_p}(\nu)).$$

Clearly, $(\varepsilon_p)_{p\in N}$ is a null sequence, thus Lemma 5.3.1 (b) reveals that the right-hand side vanishes as p goes to $\sup N$. This proves that $P^k \in \Lambda_M(\mu^k, \nu^k)$ converges in $\mathcal{W}_1$ to $P \in \Lambda_M(\mu, \nu)$. □

The proof of Theorem 6.2.20 also relies on the following approximation lemma.

Lemma 6.6.3. *Let $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$ be such that $a < b$ and $p \in \mathcal{P}^1(\mathbb{R})$ be supported on $[a, b]$ with mean $m \in \mathbb{R}$. Then for any sequences $(a^k)_{k\in\mathbb{N}}, (b^k)_{k\in\mathbb{N}}$ of real numbers such that*

$$a < a^k < m < b^k < b, \quad a^k \underset{k\to+\infty}{\searrow} a \quad \text{and} \quad b^k \underset{k\to+\infty}{\nearrow} b,$$

there are measures $p^k \leq_{cx} p$ supported on $[a^k, b^k]$ such that $\mathcal{W}_1(p^k, p) \underset{k\to+\infty}{\longrightarrow} 0$.

Proof of Lemma 6.6.3. First, we let $a, b \in \mathbb{R}$ and consider for each $k \in \mathbb{N}$ the measure

$$p^k := \left(\frac{b^k - m}{b^k - a^k}\delta_{a^k} + \frac{m - a^k}{b^k - a^k}\delta_{b^k}\right) \wedge_c p.$$

Clearly, p^k is supported on $[a^k, b^k]$ and is dominated by p in the convex order. Moreover, due to $\mathcal{W}_1$-continuity of $\wedge_c$ (Lemma 5.4.1), we have that

$$\lim_{k\to+\infty} p^k = \left(\frac{b-m}{b-a}\delta_a + \frac{m-a}{b-a}\delta_b\right) \wedge_c p = p \quad \text{in } \mathcal{W}_1.$$

Suppose now $a, b \in \mathbb{R} \cup \{-\infty, +\infty\}$. Let $\varepsilon > 0$. By Lemma 5.4.4, there exists a compactly supported probability measure $p^\varepsilon \leq_{cx} p$ such that $\mathcal{W}_1(p, p^\varepsilon) \leq \varepsilon$. Since p^ε is compactly supported, we can find like in the previous case measures $p^{\varepsilon,k} \leq_{cx} p^\varepsilon$ supported on $[a^k, b^k]$ such that $\mathcal{W}_1(p^{\varepsilon,k}, p^\varepsilon)$ vanishes as k goes to $+\infty$, which implies

$$p^{\varepsilon,k} \leq_{cx} p \quad \text{and} \quad \limsup_{k\to+\infty} \mathcal{W}_1(p^{\varepsilon,k}, p) \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary we deduce the existence of a strictly increasing sequence $(k_N)_{N\in\mathbb{N}^*}$ of positive integers such that

$$\forall N \in \mathbb{N}^*, \quad \forall k \geq k_N, \quad \mathcal{W}_1(p^{2^{-N},k}, p) \leq 3 \cdot 2^{-N}.$$

For $k \in \mathbb{N}$, let $N_k = \max\{N \in \mathbb{N} \mid k \geq k_N\}$, where the maximum of the empty set is defined as 0. Since k_N is strictly increasing, we find that $N_k \to +\infty$ as $k \to +\infty$. Then the sequence of probability measures

$$p^k := p^{2^{-N_k},k}, \quad k \in \mathbb{N}$$

gives a solution. □

Proof of Theorem 6.2.20. According to the discussion on f and g above, Lemma 6.6.1 and Lemma 6.6.2, we can assume from now on that $f(x) = g(x) = 1 + |x|$, P is given in the form (6.6.2), that is

$$P = \frac{1}{n}\sum_{j=1}^{n} J(\pi^j),$$

where $n \in \mathbb{N}^*$ and $\pi^1, \cdots, \pi^n \in \mathcal{P}(\mathbb{R}\times\mathbb{R})$ are all martingale couplings with first marginal μ, and (μ, ν) consists of a single irreducible component $I = (a, b)$. We call the second martingal of π^j by ν_j.

Assume for a moment that we can find for all $\varepsilon \in (0, 1)$ and $j \in \{1, \cdots, n\}$ a sequence $(\nu_j^{k,\varepsilon})_{k\in\mathbb{N}}$ of probability measures which satisfies

$$\limsup_{k\to+\infty} \mathcal{W}_1(\nu_j^{k,\varepsilon}, \nu_j) \leq M\varepsilon, \quad \mu^k \leq_{cx} \nu_j^{k,\varepsilon} \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^{n} \nu_j^{k,\varepsilon} = \nu^k, \tag{6.6.9}$$

where $M \in \mathbb{R}$ is independent of ε.

As $\varepsilon > 0$ is arbitrary we deduce the existence of a strictly increasing sequence $(k_N)_{N\in\mathbb{N}^*}$ of positive integers such that

$$\forall N \in \mathbb{N}^*, \quad \forall k \geq k_N, \quad \forall j \in \{1, \cdots, n\}, \quad \mathcal{W}_1(\nu_j^{k,2^{-N}}, \nu_j) \leq (M+1) \cdot 2^{-N}.$$

For $k \in \mathbb{N}$, let $N_k = \max\{N \in \mathbb{N} \mid k \geq k_N\}$, where the maximum of the empty set is defined as 0. Since k_N is strictly increasing, we find that $N_k \to +\infty$ as $k \to +\infty$. Then for all $j \in \{1, \cdots, n\}$ the sequence of probability measures

$$\nu_j^k := \nu^{k,2^{-N_k}}, \quad k \in \mathbb{N}$$

is such that

$$\nu_j^k \underset{k\to+\infty}{\longrightarrow} \nu_j \text{ in } \mathcal{W}_1, \quad \mu^k \leq_{cx} \nu_j^k \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^n \nu_j^k = \nu^k.$$

By Theorem 5.2.5, there is for each $j \in \{1, \cdots, n\}$ a sequence $\pi^{k,j} \in \Pi_M(\mu^k, \nu_j^k)$, $k \in \mathbb{N}$ which converges in $\mathcal{AW}_1$ to π^j. The sequence of extended martingale couplings is given by

$$P^k := \frac{1}{n}\sum_{j=1}^n J(\pi^{k,j}) \in \Lambda(\mu^k, \nu^k),$$

and conclude with

$$\mathcal{W}_1(P^k, P) \leq \frac{1}{n}\sum_{j=1}^n \mathcal{W}_1(J(\pi^{k,j}), J(\pi^j)) = \frac{1}{n}\sum_{j=1}^n \mathcal{AW}_1(\pi^{k,j}, \pi^j) \underset{k\to+\infty}{\longrightarrow} 0.$$

Fix $\varepsilon \in (0,1)$. In order to conclude, it remains to show the existence for all $j \in \{1, \cdots, n\}$ of a sequence $(\nu_j^{k,\varepsilon})_{k\in\mathbb{N}}$ which satisfies (6.6.9).

Let $a^m, b^m \in I$, $m \in \mathbb{N}$, be such that $a^m < b^m$, $\mu([a^m, b^m]) > 0$, $\mu(\{a^m, b^m\}) = 0$, $a^m \searrow a$ and $b^m \nearrow b$ as $m \to +\infty$, and define p_x^m when it exists as the only probability measure supported on $\{a^m, b^m\}$ with barycentre x, δ_x else:

$$p_x^m := \begin{cases} \frac{b^m - x}{b^m - a^m}\delta_{a^m} + \frac{x - a^m}{b^m - a^m}\delta_{b^m} & \text{if } x \in (a^m, b^m), \\ \delta_x & \text{else.} \end{cases}$$

Consider for each $j \in \{1, \cdots, n\}$ and $x \in \mathbb{R}$ the probability measure

$$\pi_x^{j,m} := \mathcal{I}(\pi_x^j, p_x^m),$$

and set $\pi^{j,m}(dx, dy) := \mu(dx)\,\pi_x^{j,m}(dy)$. Since $\pi_x^{j,m}$ is dominated in the convex order by p_x^m, it has barycentre x and $\pi^{j,m}$ is a martingale coupling between μ and its second marginal $\nu_{j,m}$. Moreover, as π_x^j and p_x^m have the same barycentre, we find according to Remark 6.2.17 that $\pi_x^{j,m} \leq_{cx} \pi_x^j$, thus $\nu_{j,m} \leq_{cx} \nu_j$. Moreover,

$$\mathcal{AW}_1(\pi^j, \pi^{j,m}) \leq \int_{\mathbb{R}} \mathcal{W}_1(\pi_x^j, \pi_x^{j,m})\,\mu(dx). \tag{6.6.10}$$

Up to modifying $x \mapsto \pi_x^j$ on a μ-null set, we may suppose w.l.o.g. that for all $x \in \mathbb{R}$, π_x^j is concentrated on the closure of I. By Lemma 6.6.3, π_x^j can be approximated by measures $q_x^{j,m} \leq_{cx} \pi_x^j$, $m \in \mathbb{N}$, concentrated on $[a^m, b^m]$ such that $\mathcal{W}_1(q^{j,m}, \pi_x^j)$ vanishes as m goes to $+\infty$. By definition of the Wasserstein projection, there holds

$$\mathcal{W}_1(\pi_x^j, \pi_x^{j,m}) \leq \mathcal{W}_1(\pi_x^j, q_x^{j,m}) \underset{m\to+\infty}{\longrightarrow} 0.$$

By the triangle inequality and convexity of the absolute value,

$$\mathcal{W}_1(\pi_x^j, q_x^{j,m}) \leq \mathcal{W}_1(\pi_x^j, \delta_0) + \mathcal{W}_1(\delta_0, q_x^{j,m}) \leq 2\mathcal{W}_1(\pi_x^j, \delta_0),$$

whereby we can apply dominated convergence in (6.6.10) to find that $\pi^{j,m}$ converges to π^j in $\mathcal{AW}_1$ as $m \to +\infty$. In particular, we have $\mathcal{W}_1$-convergence of $\nu_{j,m}$ to ν_j as $m \to +\infty$.

Choose $m \in \mathbb{N}$ such that $\tilde{\nu}_j := \nu_{j,m}$ satisfies

$$\mathcal{W}_1(\tilde{\nu}_j, \nu_j) < \varepsilon, \quad \forall j \in \{1, \cdots, n\}. \tag{6.6.11}$$

Since $\mu(\{a^m, b^m\}) = 0$, we have that for any continuous map $h : \mathbb{R} \to \mathbb{R}$ with at most linear growth, the set of discontinuities of $h\mathbb{1}_{[a^m,b^m]}$ is a μ-null set, so Portmanteau's theorem implies that $\mu^k|_{[a^m,b^m]}(h)$ converges to $\mu|_{[a^m,b^m]}(h)$ as $k \to +\infty$, hence

$$\frac{\mu([a^m, b^m])}{\mu^k([a^m, b^m])}\mu^k|_{[a^m,b^m]} \underset{k\to+\infty}{\longrightarrow} \mu|_{[a^m,b^m]} \text{ in } \mathcal{W}_1 \quad \text{and} \quad \mu^k([a^m, b^m]) \underset{k\to+\infty}{\longrightarrow} \mu([a^m, b^m]). \tag{6.6.12}$$

Since $\pi_x^{j,m} \leq_{cx} p_x^m$, $\pi_x^{j,m}([a^m, b^m]) = 1$ for $x \in [a^m, b^m]$ and $\pi_x^{j,m} = \delta_x$ for $x \in \mathbb{R}\backslash[a^m, b^m]$. So for any measurable and bounded map $h : \mathbb{R} \to \mathbb{R}$, we have

$$\tilde{\nu}_j|_{\mathbb{R}\backslash[a^m,b^m]}(h) = \int_{\mathbb{R}} \pi_x^{j,m}(h\mathbb{1}_{\mathbb{R}\backslash[a^m,b^m]})\, \mu(dx) = \int_{\mathbb{R}} h(x)\mathbb{1}_{\mathbb{R}\backslash[a^m,b^m]}(x)\, \mu(dx) = \mu|_{\mathbb{R}\backslash[a^m,b^m]}(h),$$

$$\text{and} \quad \tilde{\nu}_j|_{[a^m,b^m]}(h) = \int_{\mathbb{R}} \pi_x^{j,m}(h\mathbb{1}_{[a^m,b^m]})\, \mu(dx) = \int_{\mathbb{R}} \pi_x^{j,m}(h)\, \mu|_{[a^m,b^m]}(dx),$$

hence $\tilde{\nu}_j|_{\mathbb{R}\backslash[a^m,b^m]} = \mu|_{\mathbb{R}\backslash[a^m,b^m]}$ and $\tilde{\nu}_j|_{[a^m,b^m]}$ is the second marginal of $\mu|_{[a^m,b^m]} \times \pi_x^{j,m}$, which yields $\mu|_{[a^m,b^m]} \leq_{cx} \tilde{\nu}_j|_{[a^m,b^m]}$. In particular this implies $\tilde{\nu}_j([a^m, b^m]) = \mu([a^m, b^m])$, and the barycentres $\frac{1}{\mu([a^m,b^m])} \int_{\mathbb{R}} x\, \tilde{\nu}^j|_{([a^m,b^m])}(x)$ and $\frac{1}{\mu[a^m,b^m]} \int_{\mathbb{R}} x\, \mu|_{[a^m,b^m]}(dx)$ coincide, we denote their common value by $\tilde{x}$. By (6.6.12), for k large enough $\mu^k([a^m, b^m]) > 0$ so we can define

$$\hat{\nu}_j^k := \mathcal{J}\left(\frac{\mu^k|_{[a^m,b^m]}}{\mu^k([a^m, b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu([a^m, b^m])}\right) \wedge_c p_{x^k}^m,$$

where $x^k := \frac{1}{\mu^k([a^m,b^m])} \int_{\mathbb{R}} x\, \mu^k|_{[a^m,b^m]}(dx)$. As $\mathcal{J}$ is $\mathcal{W}_1$-continuous (Theorem 6.2.18), by (6.6.12) then the fact that $\mu|_{[a^m,b^m]} \leq_{cx} \tilde{\nu}_j|_{[a^m,b^m]}$, we have that

$$\mathcal{J}\left(\frac{\mu^k|_{[a^m,b^m]}}{\mu^k([a^m, b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu([a^m, b^m])}\right) \underset{k\to+\infty}{\longrightarrow} \mathcal{J}\left(\frac{\mu|_{[a^m,b^m]}}{\mu([a^m, b^m])}, \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu([a^m, b^m])}\right) = \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu([a^m, b^m])} \quad \text{in } \mathcal{W}_1.$$

Since $\tilde{\nu}_j|_{[a^m,b^m]}$ is concentrated on $[a^m, b^m]$ with mass $\mu([a^m, b^m])$ and barycentre $\tilde{x}$, we get that $\tilde{\nu}_j|_{[a^m,b^m]} \leq_{cx} \mu([a^m, b^m])p_{\tilde{x}}^m$. Due to (6.6.12), we have that x^k converges to $\tilde{x}$, thus, $p_{x^k}^m \underset{k\to+\infty}{\longrightarrow} p_{\tilde{x}}^m$ in $\mathcal{W}_1$. Therefore, $\mathcal{W}_1$-continuity of $\wedge_c$ (Lemma 5.4.1) yields that

$$\lim_{k\to+\infty} \hat{\nu}_j^k = \frac{\tilde{\nu}_j|_{[a^m,b^m]}}{\mu([a^m, b^m])} \quad \text{in } \mathcal{W}_1. \tag{6.6.13}$$

We set

$$\tilde{\nu}_j^k := (1-\varepsilon)\left(\mu^k|_{\mathbb{R}\setminus[a^m,b^m]} + \mu^k([a^m,b^m])\hat{\nu}_j^k\right) + \varepsilon\mu^k. \tag{6.6.14}$$

Note that $\mu^k([a^m,b^m])\hat{\nu}_j^k$ is the convex order infimum of two measures which dominate $\mu^k|_{[a^m,b^m]}$ in the convex order, hence $\mu^k|_{[a^m,b^m]} \leq_{cx} \mu^k([a^m,b^m])\hat{\nu}_j^k$, which yields

$$\mu^k = (1-\varepsilon)\left(\mu^k|_{\mathbb{R}\setminus[a^m,b^m]} + \mu^k|_{[a^m,b^m]}\right) + \varepsilon\mu^k \leq_{cx} \tilde{\nu}_j^k. \tag{6.6.15}$$

Furthermore, there holds by (6.6.12) and (6.6.13) that

$$\tilde{\nu}_j^k \underset{k\to+\infty}{\longrightarrow} (1-\varepsilon)\left(\mu|_{\mathbb{R}\setminus[a^m,b^m]} + \tilde{\nu}_j|_{[a^m,b^m]}\right) + \varepsilon\mu = (1-\varepsilon)\tilde{\nu}_j + \varepsilon\mu \quad \text{in } \mathcal{W}_1. \tag{6.6.16}$$

Due to uniform convergence of potential functions, there is for any $\delta > 0$ an index $k(\delta) \in \mathbb{N}$ such that for all $k \geq k(\delta)$ and $j \in \{1,\cdots,n\}$ we have

$$u_{\tilde{\nu}_j^k} \leq (1-\varepsilon)u_{\tilde{\nu}_j} + \varepsilon u_\mu + \delta \quad \text{and} \quad u_\nu \leq u_{\nu^k} + \delta. \tag{6.6.17}$$

Recall that (μ,ν) is irreducible, therefore we can fix a $\delta > 0$ such that

$$\varepsilon u_\mu|_{[a^m,b^m]} \leq \varepsilon u_\nu|_{[a^m,b^m]} - 2\delta. \tag{6.6.18}$$

Recall moreover that $\tilde{\nu}_j \leq_{cx} \nu_j$ for all $j \in \{1,\cdots,n\}$, hence

$$\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j \leq_{cx} \frac{1}{n}\sum_{j=1}^n \nu_j = \nu. \tag{6.6.19}$$

So, let $k \geq k(\delta)$, then for $y \in [a^m,b^m]$ we obtain by (6.6.17) for the first and last inequalities and by (6.6.18) and (6.6.19) for the second inequality that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n u_{\tilde{\nu}_j^k}(y) &\leq (1-\varepsilon)\frac{1}{n}\sum_{j=1}^n u_{\tilde{\nu}_j}(y) + \varepsilon u_\mu(y) + \delta \leq (1-\varepsilon)u_\nu(y) + \varepsilon u_\nu(y) - \delta \\ &= u_\nu(y) - \delta \leq u_{\nu^k}(y).\end{aligned}$$

On the other hand, recall that $\hat{\nu}_j^k \leq_{cx} p_{x^k}^m$. Moreover, $p_{x^k}^m$ and $\frac{1}{\mu^k([a^m,b^m])}\mu^k|_{[a^m,b^m]}$ are two probability measures concentrated on $[a^m,b^m]$ with the same barycentre, which yields $\mu^k([a^m,b^m])u_{p_{x^k}^m} = u_{\mu^k|_{[a^m,b^m]}}$ on $\mathbb{R}\setminus[a^m,b^m]$. Using the definition (6.6.14) for the first equality, we then find for $y \notin [a^m,b^m]$

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n u_{\tilde{\nu}_j^k}(y) &= (1-\varepsilon)\left(u_{\mu^k|_{\mathbb{R}\setminus[a^m,b^m]}}(y) + \mu^k([a^m,b^m])\frac{1}{n}\sum_{j=1}^n u_{\hat{\nu}_j^k}(y)\right) + \varepsilon u_{\mu^k}(y) \\ &\leq (1-\varepsilon)\left(u_{\mu^k|_{\mathbb{R}\setminus[a^m,b^m]}}(y) + \mu^k([a^m,b^m])u_{p_{x^k}^m}(y)\right) + \varepsilon u_{\mu^k}(y) = u_{\mu^k}(y).\end{aligned}$$

This shows that $\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j^k \leq_{cx} \nu^k$ when $k \geq k(\delta)$. Let then $\chi^k \in \Pi_M(\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j^k, \nu^k)$ be the inverse transform martingale coupling (see Section 2.3.1), which by Theorem 2.2.12 is such that

$$\int_{\mathbb{R}\times\mathbb{R}} |y-x|\,\chi^k(dx,dy) \leq 2\mathcal{W}_1\left(\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j^k, \nu^k\right). \tag{6.6.20}$$

We can define $\nu_j^{k,\varepsilon}$ as the second marginal of $\tilde{\nu}_j^k \times \chi_x^k$, that is

$$\nu_j^{k,\varepsilon}(dy) := \int_{\mathbb{R}} \chi_x^k(dy)\,\tilde{\nu}_j^k(dx).$$

In particular we have

$$\tilde{\nu}_j^k \leq_{cx} \nu_j^{k,\varepsilon} \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^n \nu_j^{k,\varepsilon} = \nu^k. \tag{6.6.21}$$

We can estimate by the triangle inequality

$$\begin{aligned}\mathcal{W}_1(\nu_j^{k,\varepsilon}, \nu_j) &\leq \mathcal{W}_1(\nu_j^{k,\varepsilon}, \tilde{\nu}_j^k) + \mathcal{W}_1(\tilde{\nu}_j^k, \tilde{\nu}_j) + \mathcal{W}_1(\tilde{\nu}_j, \nu_j)\\ &\leq \int_{\mathbb{R}}\int_{\mathbb{R}} |y-x|\,\chi_x^k(dy)\,\tilde{\nu}_j^k(dx) + \mathcal{W}_1(\tilde{\nu}_j^k, \tilde{\nu}_j) + \mathcal{W}_1(\tilde{\nu}_j, \nu_j).\end{aligned}$$

Summing over j, drawing the limit $k \to +\infty$ and using (6.6.20), (6.6.16) and (6.6.11) leads to

$$\begin{aligned}\limsup_{k\to+\infty} \frac{1}{n}\sum_{j=1}^n \mathcal{W}_1(\nu_j^{k,\varepsilon}, \nu_j) &\leq \limsup_{k\to+\infty}\left(\int_{\mathbb{R}\times\mathbb{R}} |y-x|\,\chi^k(dx,dy) + \frac{1}{n}\sum_{j=1}^n \mathcal{W}_1(\tilde{\nu}_j^k, \tilde{\nu}_j)\right) + \varepsilon\\ &\leq 2\limsup_{k\to+\infty} \mathcal{W}_1\left(\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j^k, \nu^k\right) + \frac{\varepsilon}{n}\sum_{j=1}^n \mathcal{W}_1(\mu, \tilde{\nu}_j) + \varepsilon\\ &\leq 2\mathcal{W}_1\left((1-\varepsilon)\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j + \varepsilon\mu, \nu\right) + \frac{\varepsilon}{n}\sum_{j=1}^n (\mathcal{W}_1(\mu, \nu_j) + \varepsilon) + \varepsilon\\ &\leq 2(1-\varepsilon)\mathcal{W}_1\left(\frac{1}{n}\sum_{j=1}^n \tilde{\nu}_j, \nu\right) + 2\varepsilon\mathcal{W}_1(\mu,\nu) + \frac{\varepsilon}{n}\sum_{j=1}^n \mathcal{W}_1(\mu,\nu_j) + \varepsilon^2 + \varepsilon\\ &\leq \frac{2(1-\varepsilon)}{n}\sum_{j=1}^n \mathcal{W}_1(\tilde{\nu}_j, \nu_j) + 2\varepsilon\mathcal{W}_1(\mu,\nu) + \frac{\varepsilon}{n}\sum_{j=1}^n \mathcal{W}_1(\mu,\nu_j) + \varepsilon^2 + \varepsilon\\ &\leq 2\varepsilon\mathcal{W}_1(\mu,\nu) + \frac{\varepsilon}{n}\sum_{j=1}^n \mathcal{W}_1(\mu,\nu_j) + 3\varepsilon.\end{aligned} \tag{6.6.22}$$

Then by (6.6.22), (6.6.15) and (6.6.21), for all $j \in \{1, \cdots, n\}$ the sequence $(\nu_j^{k,\varepsilon})_{k\in\mathbb{N}}$ satisfies (6.6.9) with $M = n(2\mathcal{W}_1(\mu,\nu) + \frac{1}{n}\sum_{j=1}^n \mathcal{W}_1(\mu,\nu_j) + 4)$, which is conclusive. □

6.7 Appendix

6.7.1 Lemma

Lemma 6.7.1. *Let $\mu \in \mathcal{P}^1(\mathbb{R})$ and $f : \mathbb{R} \to \mathbb{R}$ be a measurable map such that μ is the image of Lebesgue measure on $(0, 1)$ by f. Then*

$$\forall p \in [0,1], \quad \int_p^1 f(u)\,du \le \int_p^1 F_\mu^{-1}(u)\,du.$$

Proof. Let $p \in [0, 1]$, $c : (x, y) \mapsto \mathbb{1}_{\{x<1-p\}}y$ and U be a uniform random variable on $(0, 1)$. For all $x, x', y, y' \in \mathbb{R}$ such that $x \le x'$ and $y \le y'$ we have

$$c(x', y') - c(x, y') - c(x', y) + c(x, y) \le 0.$$

Then by [161, Theorem 3.1.2] we have

$$\int_p^1 f(u)\,du = \mathbb{E}[c(U, f(1 - U))] \le \mathbb{E}[c(U, F_\mu^{-1}(1 - U))] = \int_p^1 F_\mu^{-1}(u)\,du.$$

□

6.7.2 Extension from $\mathcal{P}^r$ to $\mathcal{P}_f$.

We recall that unless explicitly stated otherwise, $\mathcal{P}(Y)$ is endowed with the weak convergence topology, and for any continuous map $f : Y \to [1, +\infty)$ we endow the space $\mathcal{P}_f(Y) = \{p \in \mathcal{P}(Y) \mid p(f) < +\infty\}$ with the topology induced by the following convergence: a sequence $(p_k)_{k\in\mathbb{N}} \in \mathcal{P}_f(Y)^{\mathbb{N}}$ converges in $\mathcal{P}_f(Y)$ to p iff p_k converges weakly to p and $p_k(f)$ converges to $p(f)$ as $k \to +\infty$.

As mentioned in Section 6.2, this extension emerged from the need to overcome the inconvenience of the non-compacity of the $\mathcal{W}_r$-balls $\{p \in \mathcal{P}^r(Y) \mid \mathcal{W}_r(p, \delta_{y_0}) \le R\}$, $R > 0$ for the $\mathcal{W}_r$-distance topology. All the following lemmas together show that this extension enjoys nearly the same flexibility as the usual Wasserstein distance topology and most importantly benefits of a helpful compacity result, see Lemma 6.7.7 below.

Remark 6.7.2. We continue with some remarks on the structure of $\mathcal{P}_f(Y)$:

(1) Convergence in $\mathcal{P}_f(Y)$ can be described differently: let $(p_k)_{k\in\mathbb{N}}$ converge to p in $\mathcal{P}_f(Y)$, and let $g \in \mathcal{C}(Y)$ be such that $0 \le g \le f$. By Portmanteau's theorem we have $p(g) \le \liminf_{k\to+\infty} p_k(g)$ and $p(f) - p(g) = p(f - g) \le \liminf_{k\to+\infty} p_k(f - g) = p(f) - \limsup_{k\to+\infty} p_k(g)$, hence $\limsup_{k\to+\infty} p_k(g) \le p(g)$. We deduce that

$$p_k \underset{k\to+\infty}{\longrightarrow} p \text{ in } \mathcal{P}_f(Y) \iff p_k(g) \underset{k\to+\infty}{\longrightarrow} p(g), \quad \forall g \in \Phi_f(Y), \tag{6.7.1}$$

when $\Phi_f(Y) := \{g \in \mathcal{C}(Y) \mid g \text{ is absolutely dominated by a positive multiple of } f\}$.

It is immediate that for $r \ge 1$, this topology is finer than the one induced by $\mathcal{W}_r$ on $\mathcal{P}_f(Y)$ if f belongs to the set $\mathcal{F}^r(Y)$ of real-valued continuous functions defined on Y and bounded from below by $y \mapsto 1 + d_Y^r(y, y_0)$.

(2) The set $\mathcal{P}_f(Y)$ is naturally embedded into the set $\mathcal{M}_+(Y)$ of all bounded positive Borel measures on Y, endowed with the weak topology, via the following continuous injection

$$\iota\colon \mathcal{P}_f(Y) \to \mathcal{M}_+(Y), \quad \iota(p)(dy) = f(y)\, p(dy).$$

Clearly, the topology on $\mathcal{P}_f(Y)$ coincides with the initial topology under ι. Even more, the set $\iota(\mathcal{P}_f(Y)) = \{m \in \mathcal{M}_+(Y)\colon m(\frac{1}{f}) = 1\}$ is a closed subset of $\mathcal{M}_+(Y)$ since $\frac{1}{f}$ is continuous and bounded. As such, we deduce that $\mathcal{P}_f(Y)$ is a Polish space.

(3) By [41, Theorem 8.3.2 and the preceding discussion], we have that the weak topology on $\mathcal{M}_+(Y)$ is induced by the norm

$$\|m_1 - m_2\|_0 := \sup_{\substack{g\colon Y\to[-1,1] \\ g \text{ is 1-Lipschitz}}} (m_1(g) - m_2(g)).$$

This permits us to define a metric on $\mathcal{P}_f(Y)$ via

$$\overline{\mathcal{W}}_f(p,q) := \sup_{\substack{g\colon Y\to[-1,1], \\ g \text{ is 1-Lipschitz}}} (p(fg) - q(fg)) = \|\iota(p) - \iota(q)\|_0. \tag{6.7.2}$$

Thus, $\overline{\mathcal{W}}_f$ is a complete metric compatible with the topology on $\mathcal{P}_f(Y)$.

From now on, we equip $\mathcal{P}_f(Y)$ with $\overline{\mathcal{W}}_f$. A continuous function $f\colon Y \to [1,+\infty)$ can naturally be lifted to a continuous function $\hat{f}\colon \mathcal{P}_f(Y) \to [1,+\infty)$ by setting

$$\hat{f}(p) := p(f). \tag{6.7.3}$$

Let us recall some notation. For any probability $P \in \mathcal{P}(\mathcal{P}(Y))$ we denote its intensity $I(P) \in \mathcal{P}(Y)$, defined by $I(P)(dy) = \int_{\mathcal{P}(Y)} p(dy)\, P(dp)$. Then we have $P(\hat{f}) = I(P)(f)$. For two maps $f : X \to \mathbb{R}$ and $g : Y \to \mathbb{R}$, we denote $f \oplus g : X \times Y \ni (x,y) \mapsto f(x) + g(y)$.

As we are solely interested in topological properties, the next lemma shows that we can freely switch between the spaces $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$, and $\mathcal{P}^1(\mathcal{P}_f(Y))$, the latter's definition being given by (6.1.1) with $(1, \mathcal{P}_f(Y), \overline{\mathcal{W}}_f)$ replacing (r, X, d_X).

Lemma 6.7.3. *(a) Let $f : Y \to [1,+\infty)$ be continuous. Then*

$$\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) = \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)), \tag{6.7.4}$$

and their topologies are equal. If moreover one endows $\mathcal{P}_f(Y)$ with the metric $\overline{\mathcal{W}}_f$ defined by (6.2.1)*, then*

$$\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) = \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)) = \mathcal{P}^1(\mathcal{P}_f(Y)), \tag{6.7.5}$$

and their topologies are equal.

(b) Let $f : X \to [1, +\infty)$ and $g : Y \to [1, +\infty)$ be continuous. Then

$$\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y)) = \mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y)), \tag{6.7.6}$$

and their topologies are equal.

Remark 6.7.4. The equalities (6.7.4), (6.7.5) and (6.7.6) are to be understood up to an identification, namely we consider that for two measurable sets $Z' \subset Z$, a probability measure $p \in \mathcal{P}(Z)$ belongs to $\mathcal{P}(Z')$ if $p(Z') = 1$, the underlying identification being of course between $p \in \mathcal{P}(Z)$ and the probability measure $p' \in \mathcal{P}(Z')$ defined for any measurable subset $A \subset Z'$ by $p'(A) = p(A \cap Z')$.

Proof. Let us prove (a). The inclusion $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) \supset \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ is straightforward. Conversely, let $P \in \mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$. Then by definition,

$$P(\hat{f}) = \int_{\mathcal{P}(Y)} p(f)\, P(dp) < +\infty,$$

which can only hold if $p(f)$ is $P(dp)$-almost everywhere finite, or equivalently $P(\mathcal{P}_f(Y)) = 1$, hence $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y)) \subset \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ and therefore we have equality. To see that the two topologies match, let us show that

$$P^k \underset{k\to+\infty}{\longrightarrow} P \text{ in } \mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y)) \iff P^k \underset{k\to+\infty}{\longrightarrow} P \text{ in } \mathcal{P}_{\hat{f}}(\mathcal{P}(Y)).$$

Since the topology on $\mathcal{P}_f(Y)$ is finer than the weak topology on $\mathcal{P}(Y)$, we have $\mathcal{C}(\mathcal{P}(Y)) \subset \mathcal{C}(\mathcal{P}_f(Y))$, so the direct implication is trivial. Conversely, suppose that P^k converges in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ to P as k goes to $+\infty$. Let $h \in \mathcal{C}(Y)$ be bounded. Then $\hat{h} \in \mathcal{C}(\mathcal{P}(Y))$ is bounded, and $I(P^k)(h) = P^k(\hat{h})$ converges to $P(\hat{h}) = I(P)(h)$ as k goes to $+\infty$. Moreover $I(P^k)(f) = P^k(\hat{f})$ converges to $P(\hat{f}) = I(P)(f)$. This shows that $(I(P^k))_{k\in\mathbb{N}}$ converges in $\mathcal{P}_f(Y)$ to $I(P)$. Therefore $\{I(P^k) \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_f(Y)$. We deduce by Lemma 6.7.5 below that $\{P^k \mid k \in \mathbb{N}\}$ is relatively compact in $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$. Let Q be an accumulation point of $(P^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$. In particular Q is by the direct implication shown above an accumulation point of $(P^k)_{k\in\mathbb{N}}$ in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, hence $Q = P$ by uniqueness of the limit since the topology is metrisable and therefore Hausdorff.

Let us now prove the second part of (a). We endow $\mathcal{P}_f(Y)$ with the metric $\overline{\mathcal{W}}_f$. To see that the sets $\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ and $\mathcal{P}^1(\mathcal{P}_f(Y))$ are the same, we find

$$P(\hat{f}) < +\infty \iff \int_{\mathcal{P}(Y)} p(f)\, P(dp) < +\infty \iff \int_{\mathcal{P}(Y)} \overline{\mathcal{W}}_f(p, \delta_{y_0})\, P(dp) < +\infty,$$

which is an easy consequence, as well as the equality of the topologies, of

$$\forall p \in \mathcal{P}_f(Y), \quad p(f) - f(y_0) \le \overline{\mathcal{W}}_f(p, \delta_{y_0}) \le p(f) + f(y_0).$$

Let us now prove (b). We derive the equality $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}(Y)) = \mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$ as in (a) since

$$P(f \oplus \hat{g}) = \int_{X\times\mathcal{P}(Y)} (f(x) + p(g))\, P(dx, dp) < +\infty,$$

which can only hold if the second marginal of P is concentrated on $\mathcal{P}_g(Y)$. To see that the topologies are equal, the only nontrivial part is, as in (a), to show that if $(P^k)_{k\in\mathbb{N}}$ converges in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$, then $\{P^k \mid k\in\mathbb{N}\}$ is relatively compact in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$. Let then $(P^k)_{k\in\mathbb{N}}$ converge in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ to some P. Recall moreover the definition of the extended intensity $\hat{I}$ given by (6.2.20). Let $h_1: X\to\mathbb{R}$ and $h_2: Y\to\mathbb{R}$ be two continuous and bounded maps. Then the map $H: X\times\mathcal{P}(Y)\ni(x,p)\mapsto\int_Y h_1(x)h_2(y)\,p(dy)$ is continuous and bounded. Denoting $h:(x,y)\mapsto h_1(x)h_2(y)$, we deduce that $\hat{I}(P^k)(h) = P^k(H)$ converges to $P(H)=\hat{I}(P)(h)$ as k goes to $+\infty$. Hence $(\hat{I}(P^k))_{k\in\mathbb{N}}$ converges weakly to $\hat{I}(P)$. Then by continuity of the projections the first marginal μ^k, resp. the second marginal ν^k of $\hat{I}(P^k)$ converges weakly to the first marginal μ, resp. the second marginal ν of $\hat{I}(P)$. Since the maps $f\oplus\hat{0}:(x,p)\mapsto f(x)$ and $0\oplus\hat{g}:(x,p)\mapsto\hat{g}(p)$ belong to $\mathcal{C}(X\times\mathcal{P}(Y))$ and are dominated by $f\oplus\hat{g}$, we also have that

$$\mu^k(f) = P^k(f\oplus\hat{0}) \underset{k\to+\infty}{\longrightarrow} P(f\oplus\hat{0}) = \mu(f) \quad\text{and}\quad \nu^k(g) = P^k(0\oplus g) \underset{k\to+\infty}{\longrightarrow} P(0\oplus g) = \nu(g),$$

which shows that $(\mu^k,\nu^k)_{k\in\mathbb{N}}$ converges in $\mathcal{P}_f(X)\times\mathcal{P}_g(Y)$ to (μ,ν). Therefore $(\hat{I}(P^k))_{k\in\mathbb{N}}$ is tight in $\mathcal{P}(X\times Y)$ and both projections $\{\mu^k \mid k\in\mathbb{N}\}$ and $\{\nu^k \mid k\in\mathbb{N}\}$ are relatively compact respectively in $\mathcal{P}_f(X)$ and $\mathcal{P}_f(Y)$, so by Lemma 6.7.8 below $\{P^k \mid k\in\mathbb{N}\}$ is relatively compact in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}_g(Y))$, which proves the claim. □

Lemma 6.7.5. *A set $\mathcal{A}\subset\mathcal{P}_{\hat{f}}(\mathcal{P}_f(Y))$ is relatively compact if and only if the set of its intensities $I(\mathcal{A})\subset\mathcal{P}_f(Y)$ is relatively compact.*

Proof. The first implication follows as in [18, Lemma 2.4] by continuity of I, c.f. Lemma 6.7.10 below. The reverse implication can be shown by pursuing the same idea as in [18, Lemma 2.4] with slight modification: instead of considering the map $y\mapsto d_Y(y,y')^t$ we use $y\mapsto f(y)$. □

Lemma 6.7.6. *A set $\mathcal{A}\subset\mathcal{P}_f(Y)$ is relatively compact if and only if it is tight and*

$$\forall\varepsilon>0,\quad \exists R>0,\quad \sup_{\mu\in\mathcal{A}}\int_{\{y\in Y\mid f(y)\geq R\}} f(y)\,\mu(dy)<\varepsilon.$$

Proof. The proof of this lemma runs along the lines of [18, Lemma 2.5] when replacing $y\mapsto d_Y(y,y')^t$ by $y\mapsto f(y)$. □

For $g:\mathbb{R}^d\to[1,+\infty)$, recall the definition (6.4.2) of the set $\mathcal{F}_g^+(\mathbb{R}^d)$.

Lemma 6.7.7. *Let $d\in\mathbb{N}^*$ and $\mathbb{R}^d$ be endowed with a norm $|\cdot|$, and let $g:\mathbb{R}^d\to[1,+\infty)$ be continuous. Then for all $f\in\mathcal{F}_g^+(\mathbb{R}^d)$, the set $B_R:=\{p\in\mathcal{P}(\mathbb{R})\mid p(f)\leq R\}$ is a compact subset of $\mathcal{P}_g(\mathbb{R}^d)$.*

Proof. Let $R\geq 0$, $(p_n)_{n\in\mathbb{N}}$ be a sequence in $B_R^{\mathbb{N}}$ and $\varepsilon>0$. There exists $r>0$ such that for all $x\in\mathbb{R}^d$, $|x|\geq r$ implies $f(x)\geq\frac{R}{\varepsilon}$. Let $K=\{x\in\mathbb{R}^d\mid |x|\leq r\}$. For all $n\in\mathbb{N}$, we have $R\geq p_n(f)\geq p_n(\mathbb{R}^d\backslash K)\frac{R}{\varepsilon}$, hence $p_n(\mathbb{R}^d\backslash K)\leq\varepsilon$. So $(p_n)_{n\in\mathbb{N}}$ is tight, and by Prokhorov's theorem there exists a subsequence, still denoted $(p_n)_{n\in\mathbb{N}}$ for notational simplicity, which converges weakly to $p \in \mathcal{P}(\mathbb{R})$. Since f is continuous and nonnegative, we have by Portmanteau's theorem

$$p(f) \le \liminf_{n\to+\infty} p_n(f) \le R,$$

so $p(f) \in B_R$. It remains to show that this convergence also holds in $\mathcal{W}_g$. By Skorokhod's representation theorem, there exists for all $n \in \mathbb{N}$ a random variable $Z_n \sim p_n$, such that $(Z_n)_{n\in\mathbb{N}}$ converges almost surely to a random variable $Z \sim p$. For all $n \in \mathbb{N}$ we have

$$p_n(g) = \mathbb{E}[g(Z_n)] \le \mathbb{E}[f(Z_n)] = p_n(f) \le R,$$

so by the de La Vallée Poussin theorem, $(g(Z_n))_{n\in\mathbb{N}}$ is uniformly integrable. We deduce by

$$\lim_{n\to+\infty} p_n(g) = p(g)$$

and (6.7.1) that $(p_n)_{n\in\mathbb{N}}$ converges in $\mathcal{P}_g(\mathbb{R}^d)$ to p, so B_R is compact. $\square$

For a probability measure $\pi \in \mathcal{P}(X \times Y)$, we denote by $\mathrm{proj}_X(\pi)$ and $\mathrm{proj}_Y(\pi)$ its X-marginal and Y-marginal, respectively. Recall moreover the definition of the extended intensity $\hat{I}$ given by (6.2.20).

Lemma 6.7.8. *Let $f\colon X \to [1,+\infty)$ and $g\colon Y \to [1,+\infty)$ be continuous. The following are equivalent:*

(a) A set $\Pi \subset \mathcal{P}(X\times Y)$ is tight and both projections, $\mathrm{proj}_X(\Pi) \subset \mathcal{P}_f(X)$ *and* $\mathrm{proj}_Y(\Pi) \subset \mathcal{P}_g(Y)$*, are relatively compact.*

(b) $J(\Pi)$ as a subset of $\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$ is relatively compact.

Conversely, the following are equivalent:

(a') $\Lambda \subset \mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y))$ is relatively compact.

(b') $\hat{I}(\Lambda) \subset \mathcal{P}(X\times Y)$ is tight, and both projections, $\mathrm{proj}_X(\hat{I}(\Lambda)) \subset \mathcal{P}_f(X)$ *and* $\mathrm{proj}_Y(\hat{I}(\Lambda)) \subset \mathcal{P}_g(Y)$*, are relatively compact.*

Proof. For this lemma works the same proof as in [18, Lemma 2.6] when using Lemma 6.7.5, the characterisation of relative compactness given in Lemma 6.7.6 and continuity of $\hat{I}$, see Lemma 6.7.10. $\square$

Proposition 6.7.9. *Let $f\colon X \to [1,+\infty)$ and $g\colon Y \to [1,+\infty)$ be continuous functions, and $C\colon X \times \mathcal{P}_g(Y) \to \mathbb{R}\cup\{+\infty\}$ be lower semicontinuous and bounded from below by a negative multiple of $f \oplus \hat{g}$. Then*

(a) The map

$$\mathcal{P}_{f\oplus\hat{g}}(X \times \mathcal{P}_g(Y)) \ni P \mapsto \int_{X\times\mathcal{P}_g(Y)} C(x,p)\, P(dx,dp) \tag{6.7.7}$$

is lower semicontinuous.

(b) *Suppose in addition that for all $x \in X$, the map $p \mapsto C(x,p)$ is convex. Then*

$$\mathcal{P}_{f\oplus g}(X \times Y) \ni \pi \mapsto \int_X C(x, \pi_x)\, \mu(dx), \tag{6.7.8}$$

where μ denotes the X-marginal of π, is lower semicontinuous.

Proof. Lower semicontinuity of (6.7.7) is obtained by standard arguments. To see (6.7.8), let $(\pi_k)_{k\in\mathbb{N}} \in \mathcal{P}_{f\oplus g}(X \times Y)^{\mathbb{N}}$ converge in $\mathcal{P}_{f\oplus g}(X\times Y)$ to some π. We find by the first part of Lemma 6.7.8 an accumulation point $P \in \mathcal{P}_{f\oplus \hat{g}}(X \times \mathcal{P}(Y))$ of $(J(\pi^k))_{k\in\mathbb{N}}$. By possibly passing to a subsequence we can assume that $P^k := J(\pi^k)$ converges to P in $\mathcal{P}_{f\oplus\hat{g}}(X \times P(Y))$ as k goes to $+\infty$. Write μ^k, $k \in \mathbb{N}$ and μ for the X-marginal of π^k and π, respectively. Due to (6.7.7), we obtain

$$\begin{aligned}
\liminf_{k\to+\infty} \int_X C(x, \pi^k_x)\, \mu^k(dx) &= \liminf_{k\to+\infty} \int_{X\times\mathcal{P}_f(Y)} C(x,p)\, P^k(dx,dp) \\
&\geq \int_{X\times\mathcal{P}_f(Y)} C(x,p)\, P(dx,dp) \\
&\geq \int_X C\left(x, I(P_x)\right)\, \mu(dx) \\
&= \int_X C\left(x, \hat{I}(P)_x\right)\, \mu(dx),
\end{aligned}$$

where we used Proposition 6.7.11 below for the last inequality. Since $\hat{I}$ is continuous by Lemma 6.7.10 below, we find that $\pi^k = \hat{I}(P^k) \to \hat{I}(P)$ and $\hat{I}(P^k) = \pi^k \to \pi$ as $k \to +\infty$. But the weak topology is Hausdorff and therefore $\pi = \hat{I}(P)$ yielding

$$\liminf_{k\to+\infty} \int_X C(x, \pi^k_x)\, \mu^k(dx) \geq \int_X C(x, \pi_x)\, \mu(dx),$$

and thus (6.7.8). $\square$

Lemma 6.7.10. *Let $f\colon X \to [1,+\infty)$ and $g\colon Y \to [1,+\infty)$ be continuous. The maps*

$$I\colon \mathcal{P}_{\hat{g}}(\mathcal{P}(Y)) \to \mathcal{P}_g(Y), \quad I(P)(dy) := \int_{\mathcal{P}(Y)} p(dy)\, P(dp), \tag{6.7.9}$$

$$\hat{I}\colon \mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y)) \to P_{f\oplus g}(X\times Y), \quad \hat{I}(P)(dx,dy) := \int_{X\times\mathcal{P}(Y)} p(dy)\, P(dx,dp), \tag{6.7.10}$$

are continuous.

Proof. Let $(P^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_{\hat{g}}(\mathcal{P}(Y))$ with limit point P. Let $h \in \mathcal{C}_b(Y)$, then $\hat{h} \in \mathcal{C}_b(P(Y))$. Thus,

$$\begin{aligned}
\lim_{k\to+\infty} I(P^k)(h) &= \lim_{k\to+\infty} P^k(\hat{h}) = P(\hat{h}) = I(P)(h), \\
\lim_{k\to+\infty} I(P^k)(g) &= \lim_{k\to+\infty} P^k(\hat{g}) = P(\hat{g}) = I(P)(g),
\end{aligned}$$

which shows by (6.7.1) continuity of I.

Next, let $(P^k)_{k\in\mathbb{N}}$ be a sequence in $\mathcal{P}_{f\oplus\hat{g}}(X\times\mathcal{P}(Y))$ converging to P. Let $h \in \mathcal{C}_b(X\times Y)$, then $\check{h}(x,p) := \int_Y h(x,y)\,p(dy)$ is contained in $\mathcal{C}_b(X\times\mathcal{P}(Y))$. Again, we find

$$\lim_{k\to+\infty} \hat{I}(P^k)(h) = \lim_{k\to+\infty} P^k(\check{h}) = P(\check{h}) = \hat{I}(P)(h),$$
$$\lim_{k\to+\infty} \hat{I}(P^k)(f\oplus g) = \lim_{k\to+\infty} P^k(f\oplus\hat{g}) = P(f\oplus\hat{g}) = \hat{I}(P)(f\oplus g),$$

whereby we derive continuity of $\hat{I}$ by virtue of (6.7.1). □

Proposition 6.7.11. *Let $f\colon X\to[1,+\infty)$ be continuous, $C:\mathcal{P}_f(Y)\to\mathbb{R}$ be convex, lower semicontinuous and lower bounded by a negative multiple of $\hat{f}$. Then for all $Q\in\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ holds*

$$C\left(I(Q)\right) \le \int_{\mathcal{P}_f(Y)} C(p)\,Q(dp). \tag{6.7.11}$$

If moreover C is strictly convex, then (6.7.11) *is an equality iff $Q=\delta_{I(Q)}$.*

Proof. Let $Q\in\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$, $P_n:\Omega\to\mathcal{P}(Y)$, $n\in\mathbb{N}^*$ be independent random variables identically distributed according to Q and $\mathcal{G}\subset\Phi_{\hat{f}}(\mathcal{P}(Y))$ be a countable family which determines the convergence in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ (see [78, Theorem 4.5.(b)]). By the law of large numbers, almost surely, for all $\psi\in\mathcal{G}$,

$$\frac{1}{n}\sum_{k=1}^n \psi(P_k) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[\psi(P_1)] = Q(\psi) \quad\text{and}\quad \frac{1}{n}\sum_{k=1}^n C(P_k) \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[C(P_1)] = Q(C). \tag{6.7.12}$$

Let $\omega\in\Omega$ be such that (6.7.12) holds when evaluated at ω and set $p_n = P_n(\omega)$ for $n\in\mathbb{N}^*$. Then $\left(\frac{1}{n}\sum_{k=1}^n \delta_{p_k}\right)_{n\in\mathbb{N}}$ converges in $\mathcal{P}_{\hat{f}}(\mathcal{P}(Y))$ to Q. By Lemma 6.7.10, $\frac{1}{n}\sum_{k=1}^n p_k$ converges to $I(Q)$ as $n\to+\infty$. By lower semicontinuity of C for the first inequality, convexity of C for the second one and (6.7.12) evaluated at ω for the equality, we get

$$C(I(Q)) \le \liminf_{n\to+\infty} C\left(\frac{1}{n}\sum_{k=1}^n p_k\right) \le \liminf_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n C(p_k) = Q(C). \tag{6.7.13}$$

If $Q=\delta_{I(Q)}$ we have trivially equality in (6.7.11). So, assume that Q is not concentrated on a single point, and that C is strictly convex. There are $h\in\Phi_f(Y)$ and $b\in\mathbb{R}$ such that $A=\{p\in\mathcal{P}_f(Y)\mid p(h)\le b\}$ satisfies

$$Q(A)>0 \text{ and } Q(A^c)>0. \tag{6.7.14}$$

Indeed, pick any points $p_1,p_2\in\mathcal{P}_f(Y)$, $p_1\ne p_2$ in the support of Q, then the Hahn-Banach separation theorem provides $h\in\Phi_f(Y)$ and $b\in\mathbb{R}$ such that $p_1(h)<b<p_2(h)$. As both points lie in the support of Q, and $\{p\in\mathcal{P}_f(Y)\mid p(f)<b\}$ and $\{p\in\mathcal{P}_f(Y)\mid p(f)>b\}$ are open subsets containing p_1 and p_2, respectively, we obtain (6.7.14). Write

$Q_1(dp) := \mathbb{1}_A \frac{Q(dp)}{Q(A)}$ and $Q_2(dp) := \mathbb{1}_{A^c} \frac{Q(dp)}{Q(A^c)}$. By the definition of A, we have that $I(Q_1)(h) < b < I(Q_2)(h)$ and especially $I(Q_1) \neq I(Q_2)$. By (6.7.11) we find

$$\int_{\mathcal{P}_f(y)} C(p)\, Q_1(dp) \geq C(I(Q_1)) \text{ and } \int_{\mathcal{P}_f(Y)} C(p)\, Q_2(dp) \geq C(I(Q_2)).$$

Hence, as $Q = Q(A)Q_1 + (1 - Q(A))Q_2$ we get

$$\begin{aligned}\int_{\mathcal{P}_f(Y)} C(p)\, Q(dp) &= \int_{\mathcal{P}_f(Y)} C(p)Q(A)\, Q_1(dp) + \int_{\mathcal{P}_f(Y)} C(p)Q(A^c)\, Q_2(dp)\\ &\geq Q(A)C(I(Q_1)) + (1 - Q(A))C(I(Q_2))\\ &> C(Q(A)I(Q_1) + (1 - Q(A))I(Q_2)) = C(I(Q)),\end{aligned}$$

where we used $I(Q_1) \neq I(Q_2)$ and strict convexity for the last inequality. □

6.7.3 A Portmanteau-like theorem for Carathéodory maps

Let $(\pi^k)_{k\in\mathbb{N}}$ be a sequence of probability measures defined on $X \times Y$ converging in $\mathcal{P}_{f\oplus g}(X \times Y)$ to π, and $c : X \times Y \to \mathbb{R}$ be a (lower) Carathéodory map, that is a measurable function which is (lower semi-)continuous in its second argument. The goal of the present section is to determine in which situation we can connect the asymptotic behaviour of $\int_{X\times Y} c(x,y)\, \pi^k(dx,dy)$ and $\int_{X\times Y} c(x,y)\, \pi(dx,dy)$. We recall that $(\pi^k)_{k\in\mathbb{N}}$ is said to converge stably to π iff for every bounded measurable map $g : X \to \mathbb{R}$ and bounded continuous map $h : Y \to \mathbb{R}$

$$\int_{X\times Y} g(x)h(y)\, \pi^k(dx,dy) \underset{k\to+\infty}{\longrightarrow} \int_{X\times Y} g(x)h(y)\, \pi(dx,dy). \tag{6.7.15}$$

We say that a sequence $(\mu^k)_{k\in\mathbb{N}}$ of probability measures on $\mathcal{P}(X)$ K-converges in total variation to μ iff for every subsequence $(\mu^{k_i})_{i\in\mathbb{N}}$ we have

$$\frac{1}{n}\sum_{i=1}^{n} \mu^{k_i} \underset{n\to+\infty}{\longrightarrow} \mu \quad \text{in total variation.}$$

Lemma 6.7.12. *Let $\pi, \pi^k \in \mathcal{P}(X \times Y)$, $k \in \mathbb{N}$ be with respective first marginal μ, μ^k. All of the following statements are equivalent:*

(a) $(\pi^k)_{k\in\mathbb{N}}$ converges to π stably.

(b) $(\pi^k)_{k\in\mathbb{N}}$ converges to π weakly and $(\mu^k)_{k\in\mathbb{N}}$ converges strongly to μ.

(c) $(\pi^k)_{k\in\mathbb{N}}$ converges to π weakly and every subsequence of $(\mu^k)_{k\in\mathbb{N}}$ has an in total variation K-convergent sub-subsequence with limit μ.

Proof. We prove "(a) $\Longrightarrow$ (b)". The definition of stable convergence given by (6.7.15) is in the Polish set-up by [42, Theorem 8.10.65 (ii)] equivalent to

$$\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) \underset{k\to+\infty}{\longrightarrow} \int_{X\times Y} c(x,y)\,\pi(dx,dy)$$

for all $c\colon X\times Y\to\mathbb{R}$ which are bounded and Carathéodory. Thus, stable convergence is stronger than weak convergence. For all measurable subsets $A\subset X$, we find by setting $g=\mathbb{1}_A$ and $h=1$ in (6.7.15) that

$$\mu^k(A) \underset{k\to+\infty}{\longrightarrow} \mu(A).$$

Next we show "(b) $\Longrightarrow$ (c)".

Let $\mu^k(dx)=\rho^k(x)\,\mu(dx)+\eta^k(dx)$ be the Lebesgue decomposition of μ^k w.r.t. μ. Since η^k is singular to μ there is $N^k\in\mathcal{B}(X)$ such that $\eta^k(N^k)=\eta^k(X)$ and $\mu(N^k)=0$. Define $N=\bigcup_{k\in\mathbb{N}}N^k\in\mathcal{B}(X)$, then $\eta^k(N)=\eta^k(X)$ for all $k\in\mathbb{N}$ and $\mu(N)$ vanishes as a countable union of null sets. Thus, $\eta^k(X)=\mu^k(N)\to\mu(N)=0$ as $k\to+\infty$. Since $(\rho^k)_{k\in\mathbb{N}}$ is bounded in $L^1(\mu)$ there is by Komlós theorem a K-convergent subsequence to some limiting function $\rho\in L^1(\mu)$. We have

$$\frac{1}{n}\sum_{l=1}^{n}\rho^{k_l} \underset{n\to+\infty}{\longrightarrow} \rho, \quad \mu\text{-a.s.}$$

By [41, Corollary 4.5.7] the above convergence even holds in $L^1(\mu)$. We find for any measurable subset $A\subset X$

$$\int_A \frac{1}{n}\sum_{l=1}^{n}\rho^{k_l}(x)\,\mu(dx) \underset{n\to+\infty}{\longrightarrow} \int_A \rho(x)\,\mu(dx)=\mu(A).$$

Hence, $\rho(x)=1$, $\mu(dx)$-almost surely and

$$\mathrm{TV}\left(\frac{1}{n}\sum_{l=1}^{n}\mu^{k_l},\mu\right)=\eta^k(X)+\int_X\left|\frac{1}{n}\sum_{l=1}^{n}\rho^{k_l}(x)-1\right|\mu(dx) \underset{n\to+\infty}{\longrightarrow} 0.$$

Finally we show "(c) $\Longrightarrow$ (a)". If $(\pi^k)_{k\in\mathbb{N}}$ does not converge stably to π, then there is a bounded Carathéodory function $c\colon X\times Y\to\mathbb{R}$, such that

$$\limsup_{k\to+\infty}\left|\int_{X\times Y} c(x,y)\,\pi^k(dx,dy)-\int_{X\times Y} c(x,y)\,\pi(dx,dy)\right|>0.$$

Hence, w.l.o.g. there is a subsequence $(\pi^{k_j})_{j\in\mathbb{N}}$ such that $\pi^{k_j}(c)\ge\pi(c)+\delta$ for some $\delta>0$. Especially, we have for any sub-subsequence $(\pi^{k_{j_i}})_{i\in\mathbb{N}}$ of $(\pi^{k_j})_{j\in\mathbb{N}}$ that

$$\frac{1}{n}\sum_{i=1}^{n}\pi^{k_{j_i}}(c)\ge\pi(c)+\delta, \tag{6.7.16}$$

whereby the Cesàro-means of the sub-subsequence are not stably convergent. By assumption there exists a subsequence $(\mu^{k_{j_i}})_{i\in\mathbb{N}}$ of $(\mu^{k_j})_{j\in\mathbb{N}}$ which K-converges in total variation to μ. For $n \in \mathbb{N}^*$ define

$$\hat{\mu}^n = \frac{1}{n}\sum_{i=1}^n \mu^{k_{j_i}} \quad \text{and} \quad \hat{\pi}^n = \frac{1}{n}\sum_{i=1}^n \pi^{k_{j_i}}.$$

We will show that $(\hat{\pi}^n)_{n\in\mathbb{N}^*}$ converges stably to π, which will contradict (6.7.16) and end the proof. Let $\hat{\mu}^n(dx) = \hat{\pi}^n(x)\,\mu(dx) + \hat{\eta}^n(dx)$ be the Lebesgue decomposition of $\hat{\mu}^n$ w.r.t. μ. Define the auxiliary sequence

$$\tilde{\pi}^n(dx,dy) = \Big((1\wedge\hat{\rho}^n(x))\,\hat{\pi}^n_x(dy) + (1-\hat{\rho}^n(x))^+\,\pi_x(dy)\Big)\,\mu(dx).$$

Let $c\colon X\times Y\to\mathbb{R}$ be Carathéodory and absolutely bounded by K, then

$$\begin{aligned}
&\left|\int_{X\times Y} c(x,y)\;\tilde{\pi}^n(dx,dy) - \int_{X\times Y} c(x,y)\,\hat{\pi}^n(dx,dy)\right| \\
&\quad\le K\left(\int_X |\hat{\rho}^n(x) - 1\wedge\hat{\rho}^n(x)|\;\mu(dx) + \int_X (1-\hat{\rho}^n(x))^+\,\mu(dx) + \hat{\eta}^n(X)\right) \\
&\quad\le K\left(\int_X |\hat{\rho}^n(x)-1|\;\mu(dx) + 2\hat{\eta}^n(X)\right) \\
&\quad\le 2K\,\mathrm{TV}(\hat{\mu}^n,\mu) \underset{n\to+\infty}{\longrightarrow} 0.
\end{aligned} \tag{6.7.17}$$

In particular, we have found that $(\tilde{\pi}^n)_{n\in\mathbb{N}^*}$ converges to π weakly. Note that the first marginal $\tilde{\pi}^n$ is μ, and therefore [127, Lemma 2.1] yields stable convergence of $\tilde{\pi}^n$ to π as $n\to+\infty$. By (6.7.17), we find that $(\hat{\pi}^n)_{n\in\mathbb{N}^*}$ also stably converges to π. □

Lemma 6.7.13. *Let $f\colon X\to[1,+\infty)$ and $g\colon Y\to[1,+\infty)$ be continuous, and let $(\pi^k)_{k\in\mathbb{N}}$ converge to π in $\mathcal{P}_{f\oplus g}(X\times Y)$.*

(a) If $c\colon X\times Y\to\mathbb{R}\cup\{+\infty\}$ is lower semicontinuous and bounded from below by a negative multiple of $g\oplus h$, then

$$\liminf_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) \ge \int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(b) If $c\colon X\times Y\to\mathbb{R}$ is continuous and absolutely bounded by positive multiple of $g\oplus h$, then

$$\lim_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) = \int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(c) If $c\colon X\times Y\to\mathbb{R}\cup\{+\infty\}$ is lower Carathéodory and bounded from below by a negative multiple of $g\oplus h$, and π^k converges to π stably, then

$$\liminf_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy) \ge \int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

(d) If $c\colon X\times Y\to\mathbb{R}$ is Carathéodory and absolutely bounded by a positive multiple of $g\oplus h$, and π^k converges to π stably, then

$$\lim_{k\to+\infty}\int_{X\times Y} c(x,y)\,\pi^k(dx,dy)=\int_{X\times Y} c(x,y)\,\pi(dx,dy).$$

Proof. These results are well-known. Note that by [41, Theorem 8.10.65] we have for every bounded lower Carathéodory map c that $\pi\mapsto\pi(c)$ is lower semicontinuous w.r.t. the topology of stable convergence. $\square$

6.7.4 On the continuity of the marginal distributions of the stretched Brownian motion

The following Lemma shows that the stretched Brownian motion provides a convenient tool to approximate two probability measures in the convex order with atomless ones still in the convex order.

Lemma 6.7.14. *Let $\mu,\nu\in\mathcal{P}^2(\mathbb{R})$ be such that $\mu\le_{cx}\nu$ and (μ,ν) consists of a single irreducible component $I=(l,r)$. Let $(M^*_t)_{t\in[0,1]}$ be the unidimensional stretched Brownian motion from μ to ν. Then*

(a) For each $t\in(0,1)$ the distribution ν_t of M^*_t is atomless.

(b) For all $s,t\in[0,1]$ such that $s<t$, (ν_s,ν_t) consists of the single irreducible component I.

Proof. Let us first prove (a). Let $t\in(0,1)$ and $y\in\mathbb{R}$. Let $\gamma=\mathcal{N}(0,1)$ be the unidimensional standard normal distribution and $C_2:\mathbb{R}\times\mathcal{P}^2(\mathbb{R})\to\mathbb{R}$ be defined for all $(x,p)\in\mathbb{R}\times\mathcal{P}^2(\mathbb{R})$ by $C_2(x,p)=\mathcal{W}_2^2(p,\gamma)$. Let $V^M_{C_2}$ be the value function given by (WMOT) for the cost function C_2 and $\pi^*\in\Pi_M(\mu,\nu)$ be optimal for $V^M_{C_2}(\mu,\nu)$. According to (6.2.9),

$$M^*_t=\varphi_t(X,B_t),$$

where $X\sim\mu$ is a random variable independent of the Brownian motion $(B_s)_{s\in[0,1]}$ and $\varphi_t:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is defined for all $(x,b)\in\mathbb{R}^2$ by

$$\varphi_t(x,b)=\int_{\mathbb{R}} F^{-1}_{\pi^*_x}\left(F_\gamma\left(\sqrt{1-t}y+b\right)\right)\,\gamma(dy). \tag{6.7.18}$$

In order to prove that $\mathbb{P}(\{M^*_t=y\})=0$, it clearly suffices to show that for $\mu(dx)$-almost every $x\in\mathbb{R}$,

$$\mathbb{P}(\{\varphi_t(x,B_t)=y\})=0. \tag{6.7.19}$$

The map $T=F^{-1}_{\pi^*_x}\circ F_\gamma$ is nondecreasing. Let $b>b'$ and assume that $T(\sqrt{1-t}y+b)=T(\sqrt{1-t}y+b')$ for all $y\in\mathbb{R}$. By monotonicity of T we deduce that T has to be constant on all intervals of $\mathbb{R}$ of length $b-b'$ and therefore on $\mathbb{R}$. So assume that T is not constant.

Then there exists $y \in \mathbb{R}$ such that $T(\sqrt{1-t}y + b) > T(\sqrt{1-t}y + b')$. As F_γ is continuous and $F_{\pi_x^*}^{-1}$ is left-continuous, we find an $\varepsilon > 0$ such that

$$\forall y' \in (y - \varepsilon, y], \quad T(\sqrt{1-t}y' + b) > T(\sqrt{1-t}y' + b'),$$

hence $\varphi_t(x, b) > \varphi_t(x, b')$. We deduce that $b \mapsto \varphi_t(x, b)$ is increasing and therefore one-to-one, hence the equation $\varphi_t(x, b) = y$ has at most one solution b^*. Denoting by b^* any real number if the latter equation has no solution, we then have

$$\mathbb{P}(\{\varphi_t(x, B_t) = y\}) \leq \mathbb{P}(\{B_t = b^*\}) = 0.$$

In order to prove (6.7.19) and conclude the proof, it remains to show that for $\mu(dx)$-almost every $x \in \mathbb{R}$, the map $F_{\pi_x^*}^{-1} \circ F_\gamma$ is not constant. Since γ is the unidimensional standard normal distribution and π_x^* is a martingale kernel, it is equivalent to show that for $\mu(dx)$-almost every $x \in \mathbb{R}$, $\pi_x^* \neq \delta_x$. This is done using the WMOT monotonicity principle. By (6.2.13) there exists a martingale C_2-monotone set $\Gamma \subset \mathbb{R} \times \mathcal{P}^1(\mathbb{R})$ such that $(x, \pi_x^*) \in \Gamma$ for all x in a μ-full set $A \subset \mathbb{R}$. This implies that for all $x, x' \in A$ and $p, p' \in \mathcal{P}^1(\mathbb{R})$ such that $\pi_x^* + \pi_{x'}^* = p + p'$, $\int_\mathbb{R} y\, p(dy) = x$ and $\int_\mathbb{R} y\, p'(dy) = x'$, we have

$$\mathcal{W}_2^2(\pi_x^*, \gamma) + \mathcal{W}_2^2(\pi_{x'}^*, \gamma) \leq \mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma). \tag{6.7.20}$$

Let $x \in A$. To conclude, it suffices to show that $\pi_x^* \neq \delta_x$. Note that if (p, p') is admissible for (6.7.20), so is $(\frac{1}{2}(\pi_x^* + p), \frac{1}{2}(\pi_{x'}^* + p'))$. In the proof of Lemma 6.3.1 we show that $q \mapsto \mathcal{W}_2^2(q, \gamma)$ is strictly convex. Therefore, if $p \neq \pi_x^*$ or $p' \neq \pi_{x'}^*$, then

$$\begin{aligned} \mathcal{W}_2^2(\pi_x^*, \gamma) + \mathcal{W}_2^2(\pi_{x'}^*, \gamma) &\leq \mathcal{W}_2^2\left(\frac{1}{2}(\pi_x^* + p), \gamma\right) + \mathcal{W}_2^2\left(\frac{1}{2}(\pi_{x'}^* + p'), \gamma\right) \\ &< \frac{1}{2}\mathcal{W}_2^2(\pi_x^*, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p, \gamma) + \frac{1}{2}\mathcal{W}_2^2(\pi_{x'}^*, \gamma) + \frac{1}{2}\mathcal{W}_2^2(p', \gamma), \end{aligned}$$

and the inequality (6.7.20) is strict. To show that $\pi_x^* \neq \delta_x$ and thereby end the proof, we deduce that it suffices to find $x' \in A$ and two measures $p, p' \in \mathcal{P}^1(\mathbb{R})$ such that

$$\delta_x + \pi_{x'}^* = p + p', \quad \int_\mathbb{R} y\, p(dy) = x, \quad \int_\mathbb{R} y\, p'(dy) = x', \quad p \neq \delta_x, \tag{6.7.21}$$

$$\text{and} \quad \mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma) \leq \mathcal{W}_2^2(\pi_{x'}^*, \gamma) + \mathcal{W}_2^2(\delta_x, \gamma). \tag{6.7.22}$$

Suppose that

$$\mu(\{x' \in (l, x] \mid \pi_{x'}^*((x, r)) > 0\}) + \mu(\{x' \in (x, r) \mid \pi_{x'}^*((l, x)) > 0\}) = 0. \tag{6.7.23}$$

Then for $\mu(dx')$-almost every $x' \in (l, r)$, the sign of $y - x$ is $\pi_{x'}^*(dy)$-almost everywhere constant equal to the sign of $x' - x$, so using the martingale property of $\pi_{x'}^*$ in the third equality, we get that

$$u_\nu(x) = \int_\mathbb{R} |y - x|\, \nu(dy) = \int_\mathbb{R} \left(\int_\mathbb{R} |y - x|\, \pi_{x'}^*(dy)\right) \mu(dx')$$

$$= \int_{(l,x]} \left| \int_{\mathbb{R}} y \, \pi^*_{x'}(dy) - x \right| \mu(dx') + \int_{(x,r)} \left| \int_{\mathbb{R}} y \, \pi^*_{x'}(dy) - x \right| \mu(dx')$$
$$= \int_{\mathbb{R}} |x' - x| \, \mu(dx') = u_\mu(x),$$

which contradicts the irreducibility of (μ, ν). We deduce that (6.7.23) does not hold, hence there exists $x' \in A$ such that $x' \leq x$ and $\pi^*_{x'}((x, r)) > 0$, or $x' > x$ and $\pi^*_{x'}((l, x)) > 0$. Since $\pi^*_{x'}$ has mean x', we can find in both cases $\tilde{x} < x < \tilde{y}$ such that $\pi^*_{x'}((\tilde{x}, x)) > 0$ and $\pi^*_{x'}((x, \tilde{y})) > 0$, which implies

$$0 < F_{\pi^*_{x'}}(x-) \leq F_{\pi^*_{x'}}(x) < 1. \tag{6.7.24}$$

Define for $\alpha, \beta \in [0, 1]$

$$p_\alpha = \int_0^\alpha \delta_{F^{-1}_{\pi^*_{x'}}(u)} \, du, \quad q_\beta = \int_{1-\beta}^1 \delta_{F^{-1}_{\pi^*_{x'}}(u)} \, du.$$

Let $c = \left(\int_0^{F_{\pi^*_{x'}}(x-)} (x - F^{-1}_{\pi^*_{x'}}(u)) \, du \right) \wedge \left(\int_{F_{\pi^*_{x'}}(x)}^1 (F^{-1}_{\pi^*_{x'}(x)}(u) - x) \, du \right)$. Since for all $u \in (0, 1)$, $u > F_{\pi^*_{x'}}(x) \iff F^{-1}_{\pi^*_{x'}}(u) > x$ and $u < F_{\pi^*_{x'}}(x-) \implies F^{-1}_{\pi^*_{x'}}(u) < x \implies u \leq F_{\pi^*_{x'}}(x-)$, the maps $\alpha \mapsto \int_0^\alpha (x - F^{-1}_{\pi^*_{x'}}(u)) \, du$ and $\beta \mapsto \int_{1-\beta}^1 (F^{-1}_{\pi^*_{x'}}(u) - x) \, du$ are nondecreasing respectively on $[0, F_{\pi^*_{x'}}(x-)]$ and $[0, 1 - F_{\pi^*_{x'}}(x)]$. Moreover, those two maps are continuous, so we deduce the existence of $\alpha' \in (0, F_{\pi^*_{x'}}(x-)]$ and $\beta' \in (0, 1 - F_{\pi^*_{x'}}(x)]$ such that they both equal c respectively at $\alpha = \alpha'$ and $\beta = \beta'$, hence

$$\int_0^{\alpha'} (F^{-1}_{\pi^*_{x'}}(u) - x) \, du + \int_{1-\beta'}^1 (F^{-1}_{\pi^*_{x'}}(u) - x) \, du = \int_{\mathbb{R}} y \, p_{\alpha'}(dy) + \int_{\mathbb{R}} y \, q_{\beta'}(dy) - (\alpha' + \beta')x = 0.$$

Note that (6.7.24) implies that $\alpha' + \beta' \in (0, 1]$. Then the measures $p = (1 - \alpha' - \beta')\delta_x + p_{\alpha'} + q_{\beta'}$ and $p' = (\alpha' + \beta')\delta_x + \pi^*_{x'} - p_{\alpha'} - q_{\beta'}$ satisfy (6.7.21). Let $\chi \in \Pi(\pi^*_{x'}, \gamma)$ be the $\mathcal{W}_2$-optimal coupling and denote by $\hat{p} = \pi^*_{x'} - p_{\alpha'} - q_{\beta'}$ and $\tilde{p} = p_{\alpha'} + q_{\beta'}$. Then

$$\left(\tilde{p}(dy) \, \chi_y(dz) + \delta_x(dy) \int_{t \in \mathbb{R}} \chi_t(dz) \, \hat{p}(dt) \right) \in \Pi(p, \gamma),$$
$$\left(\hat{p}(dy) \, \chi_y(dz) + \delta_x(dy) \int_{t \in \mathbb{R}} \chi_t(dz) \, \tilde{p}(dt) \right) \in \Pi(p', \gamma),$$

hence

$$\mathcal{W}_2^2(p, \gamma) \leq \int_{\mathbb{R} \times \mathbb{R}} |y - z|^2 \, \tilde{p}(dy) \, \chi_y(dz) + \int_{\mathbb{R} \times \mathbb{R}} |x - z|^2 \, \hat{p}(dt) \, \chi_t(dz),$$
$$\mathcal{W}_2^2(p', \gamma) \leq \int_{\mathbb{R} \times \mathbb{R}} |y - z|^2 \, \hat{p}(dy) \, \chi_y(dz) + \int_{\mathbb{R} \times \mathbb{R}} |x - z|^2 \, \tilde{p}(dt) \, \chi_t(dz).$$

Combining these inequalities yields

$$\mathcal{W}_2^2(p, \gamma) + \mathcal{W}_2^2(p', \gamma) \leq \int_{\mathbb{R} \times \mathbb{R}} |y - z|^2 \, (\tilde{p} + \hat{p})(dy) \, \chi_y(dz) + \int_{\mathbb{R} \times \mathbb{R}} |x - z|^2 \, (\hat{p} + \tilde{p})(dt) \, \chi_t(dz)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}\times\mathbb{R}} |y-z|^2 \, \chi(dy,dz) + \int_{\mathbb{R}} |x-z|^2 \, \gamma(dz) \\
&= \mathcal{W}_2^2(\pi^*_{x'}, \gamma) + \mathcal{W}_2^2(\delta_x, \gamma),
\end{aligned}
$$

which proves (6.7.22) and completes the proof.

Let us now prove (b). Let $s, t \in [0,1]$ be such that $s < t$. Since $\mu \leq_{cx} \nu_s \leq_{cx} \nu_t \leq_{cx} \nu$, we have $u_\mu \leq u_{\nu_s} \leq u_{\nu_t} \leq u_\nu$, hence $u_{\nu_s} = u_{\nu_t}$ on $I^{\complement}$. Let $z \in I$. Then

$$
\begin{aligned}
u_{\nu_s}(z) &= \mathbb{E}[|M^*_s - z|] = \mathbb{E}[|\mathbb{E}[M^*_t - z | X, (B_u)_{u\in[0,s]}]|] \\
&\leq \mathbb{E}[\mathbb{E}[|M^*_t - z| | X, (B_u)_{u\in[0,s]}] = \mathbb{E}[|M^*_t - z|] = u_{\nu_t}(z).
\end{aligned}
\tag{6.7.25}
$$

Let us show that the inequality above is strict. This is equivalent to show that given X and $(B_u)_{u\in[0,s]}$, the sign of $M^*_t - z$ is not almost surely constant. Suppose that $\mathbb{P}(M^*_s \leq z) > 0$, the case $\mathbb{P}(M^*_s \geq z) > 0$ being treated symmetrically. Then it suffices to find a Borel subset $A \subset \mathbb{R}$ such that

$$
\mathbb{P}(M^*_t > z, X \in A, M^*_s \leq z) > 0, \tag{6.7.26}
$$

since the martingale property would then imply that $\mathbb{P}(M^*_t < z, X \in A, M^*_s \leq z)$ is positive as well. The pair (μ, ν) being irreducible, we have that

$$
\mu(A := \{x \in (-\infty, z] \mid \pi^*_x((z, +\infty)) > 0\}) > 0.
$$

For fixed $x, y \in \mathbb{R}$, the map $b \mapsto T^t_{x,y}(b) = F^{-1}_{\pi^*_x}(F_\gamma(\sqrt{1-t}y + b))$ is non-decreasing where $\lim_{b\to+\infty} T^t_{x,y}(b) = \lim_{u\nearrow 1} F^{-1}_{\pi^*_x}(u)$. Recall that $y \mapsto T^t_{x,y}(b)$ and $y \mapsto T^s_{x,y}$ are γ-integrable, therefore we have due to monotone convergence

$$
\begin{aligned}
\lim_{b\to+\infty} \varphi_t(x,b) &= \lim_{b\to+\infty} \int_{\mathbb{R}} T^t_{x,y}(b) \, \gamma(dy) = \lim_{u\nearrow 1} F^{-1}_{\pi^*_x}(u), \\
\lim_{b\to-\infty} \varphi_s(x,b) &= \lim_{b\to-\infty} \int_{\mathbb{R}} T^s_{x,y}(b) \, \gamma(dy) = \lim_{u\searrow 0} F^{-1}_{\pi^*_x}(u),
\end{aligned}
$$

and, in particular,

$$
\forall x \in A, \quad \lim_{b\to+\infty} \varphi_t(x,b) > z \quad \text{and} \quad \lim_{b\to-\infty} \varphi_s(x,b) < z.
$$

Again, recall that $x \mapsto \varphi_t(x,b)$ and $x \mapsto \varphi_s(x,b)$ are μ-integrable, therefore we find due to monotone convergence

$$
\begin{aligned}
\lim_{b\to+\infty} \int_A \varphi_t(x,b) \, \mu(dx) &= \int_A \lim_{b\to+\infty} \varphi_t(x,b) \, \mu(dx) > z\mu(A), \\
\lim_{b\to-\infty} \int_A \varphi_s(x,b) \, \mu(dx) &= \int_A \lim_{b\to-\infty} \varphi_s(x,b) \, \mu(dx) < z\mu(A).
\end{aligned}
$$

Hence, there are $b_0, b_1 \in \mathbb{R}$ and $A' \subset A$, $\mu(A') > 0$, such that

$$
\varphi_t(x,b) > z \text{ and } \varphi_s(x,b') < z,
$$

for every $x \in A'$, $b \geq b_0$ and $b' \leq b_1$. Then

$$\begin{aligned}\mathbb{P}(M_t^* > z, X \in A,\ M_s^* \leq z) &\geq \mathbb{P}(B_t \geq b_0, X \in A', M_s^* \leq z)\\ &= \mathbb{P}(B_t \geq b_0, X \in A', B_s \leq b_1) > 0,\end{aligned}$$

which proves (6.7.26). Hence the inequality in (6.7.25) is strict and $u_{\nu_s} < u_{\nu_t}$ on I. □

Corollary 6.7.15. *Let $(\mu^k, \nu^k)_{k\in\mathbb{N}}$ be any sequence in $\mathcal{P}^1(\mathbb{R}) \times \mathcal{P}^1(\mathbb{R})$ converging to an irreducible pair $(\mu, \nu) \in \mathcal{P}^1(\mathbb{R}) \times \mathcal{P}^1(\mathbb{R})$ with component I. Let $\varepsilon > 0$, then there are $(\tilde{\nu}^k)_{k\in\mathbb{N}}$ and an atomless measure $\tilde{\nu}$ in $\mathcal{P}^1(\mathbb{R})$ with $\lim_{k\to+\infty} \tilde{\nu}^k = \tilde{\nu}$ in $\mathcal{W}_1$, such that*

$$\mathcal{W}_1(\tilde{\nu}, \nu) < \varepsilon, \quad \mu^k \leq_{cx} \tilde{\nu}^k \leq_{cx} \nu^k \text{ for all } k \in \mathbb{N} \quad \text{and} \quad (\mu, \tilde{\nu}) \text{ is irreducible with component } I.$$

Proof. It is clear that whenever two measures (μ, ν) have finite second moment, the stretched Brownian motion provides by Corollary 6.2.8 and Lemma 6.7.14 (a) a continuous interpolation $(\mu_t)_{t\in[0,1]}$, where $\mu_0 = \mu$ and $\mu_1 = \nu$, such that μ_t is atomless for $t \in (0, 1)$. We are going to extend such an interpolation to a case where only first moments are finite. To work around this issue, assume for a moment that we can introduce an intermediary measure $\bar{\nu}$ with $\mu \leq_{cx} \bar{\nu} \leq_{cx} \nu$, where the decomposition into irreducible components $(I_n)_{n\in N}$ of $(\bar{\nu}, \nu)$ consists only of bounded intervals, and $\bar{\nu}(J) = 0$ for $J = \mathbb{R} \setminus \bigcup_{n\in N} I_n$. For all $n \in N$, let $(\bar{\nu}|_{I_n}, \nu_n)$ be the irreducible pair associated with I_n in the decomposition of $(\bar{\nu}, \nu)$. Since I_n is bounded, $\nu_n \in \mathcal{P}^2(\mathbb{R})$ so we can consider the stretched Brownian motion $(M_t^n)_{t\in[0,1]}$ from $\frac{1}{\bar{\nu}(I_n)}\bar{\nu}|_{I_n}$ to $\frac{1}{\bar{\nu}(I_n)}\nu_n$. Since $t \mapsto M_t^n$ is almost surely continuous on $[0, 1]$ and I_n is bounded, we find by dominated convergence that the law of M_t^n converges in $\mathcal{W}_1$ to $\frac{1}{\bar{\nu}(I_n)}\nu_n$ as t tends to 1. Therefore we find for each I_n a time $t_n \in (0, 1)$ such that for each $n \in N$ the distribution $\frac{1}{\bar{\nu}(I_n)}\bar{\nu}_n$ of $M_{t_n}^n$ satisfies

$$\mathcal{W}_1(\bar{\nu}_n, \nu_n) < \frac{\varepsilon}{2^{n+1}}.$$

In particular, $\bar{\nu}_n$ is atomless by Lemma 6.7.14 (a). Recall the definition of the Wasserstein projection $\mathcal{I}(\cdot, \cdot)$ given in Section 6.2.5. We set

$$\tilde{\nu} := \sum_{n\in N} \bar{\nu}_n \quad \text{and} \quad \tilde{\nu}^k := \mathcal{I}(\tilde{\nu}, \nu^k) \vee_{cx} \mu^k.$$

Thus, $\mu^k \leq_{cx} \tilde{\nu}^k \leq_{cx} \nu^k$ and

$$\mathcal{W}_1(\tilde{\nu}, \nu) < \sum_{n\in N} \frac{\varepsilon}{2^{n+1}} \leq \varepsilon.$$

Moreover there holds

$$u_\mu \leq u_{\bar{\nu}} = \sum_{n\in N} u_{\bar{\nu}|_{I_n}} \leq \sum_{n\in N} u_{\bar{\nu}_n} = u_{\tilde{\nu}} \leq \sum_{n\in N} u_{\nu_n} = u_\nu, \tag{6.7.27}$$

which implies $\mu \leq_{cx} \tilde{\nu} \leq_{cx} \nu$. For all $n \in N$, $(\bar{\nu}|_{I_n}, \bar{\nu}_n)$ is irreducible by Lemma 6.7.14 (b), so the second inequality in (6.7.27) is strict on $\bigcup_{n\in N} I_n$. Since $u_\mu < u_\nu = u_{\bar{\nu}}$ on $I \setminus \bigcup_{n\in N} I_n$, we deduce that $u_\mu < u_{\tilde\nu}$ on I. On $I^\complement$, we have $u_\mu = u_\nu$, which implies $u_\mu = u_{\tilde\nu}$. Therefore $\{u_\mu < u_{\tilde\nu}\} = I$ and $(\mu, \tilde\nu)$ is irreducible with component I. Finally, since $\mu \leq_{cx} \tilde\nu$, $\tilde\nu \vee_{cx} \mu = \tilde\nu$, and since $\tilde\nu \leq_{cx} \nu$, $\mathcal{I}(\tilde\nu, \nu) = \tilde\nu$. Due to $\mathcal{W}_1$-continuity of $\mathcal{I}$ (Theorem 6.2.18) and $\vee_{cx}$ (Lemma 5.4.1), we derive that $\lim_{k\to+\infty} \tilde\nu^k = \tilde\nu$ in $\mathcal{W}_1$. It remains to show that there is a measure $\hat\nu$ with the above mentioned properties.

If $\nu \notin \mathcal{P}^2(\mathbb{R})$, then I has to be unbounded. For simplicity, we assume that $I = \mathbb{R}$, since if $I = (-\infty, b)$ or $I = (a, +\infty)$ with $a, b \in \mathbb{R}$ the construction below also works in these cases with the obvious modifications. To this end, we define iteratively $u_1^1 = \frac{1}{2} = u_2^1$, and for $n \in \mathbb{N}^*$, we choose $u_1^{n+1} \in (0, \frac{1}{2^{n+1}} \wedge u_1^n)$ and $u_2^{n+1} \in ((1 - \frac{1}{2^{n+1}}) \vee u_2^n, 1)$ such that

$$|\{x \in \mathbb{R} \mid F_\nu(x) \in ((u_1^{n+1}, u_1^n))\}| > 1 \quad \text{and} \quad |\{x \in \mathbb{R} \mid F_\nu(x) \in ((u_2^n, u_2^{n+1}))\}| > 1, \tag{6.7.28}$$

which is possible as $\nu((-\infty, R)) \wedge \nu((R, +\infty)) > 0$ for all $R \in \mathbb{R}$. We have that $\lim_{n\to+\infty} u_1^n = 0$ and $\lim_{n\to+\infty} u_2^n = 1$, and set

$$\hat\nu := \sum_{n\in\mathbb{N}^*} \left((u_1^n - u_1^{n+1}) \delta_{\int_{u_1^{n+1}}^{u_1^n} F_\nu^{-1}(u) \frac{du}{u_1^n - u_1^{n+1}}} + (u_2^{n+1} - u_2^n) \delta_{\int_{u_2^n}^{u_2^{n+1}} F_\nu^{-1}(u) \frac{du}{u_2^{n+1} - u_2^n}} \right),$$

which entails us to define $\bar\nu := \hat\nu \vee_c \mu$. For all $x \in \mathbb{R}$ we have by inverse transform sampling for the last equality

$$\begin{aligned} u_{\hat\nu}(x) &= \sum_{n\in\mathbb{N}^*} \left(\left| \int_{u_1^{n+1}}^{u_1^n} (F_\nu^{-1}(u) - x)\, du \right| + \left| \int_{u_2^n}^{u_2^{n+1}} (F_\nu^{-1}(u) - x)\, du \right| \right) \\ &\leq \sum_{n\in\mathbb{N}^*} \left(\int_{u_1^{n+1}}^{u_1^n} |F_\nu^{-1}(u) - x|\, du + \int_{u_2^n}^{u_2^{n+1}} |F_\nu^{-1}(u) - x|\, du \right) \\ &= \int_0^1 |F_\nu^{-1}(u) - x|\, du = u_\nu(x), \end{aligned}$$

where the inequality is strict iff there exists $n \in \mathbb{N}^*$ such that $F_\nu^{-1} - x$ is not constant on (u_1^{n+1}, u_1^n) or (u_2^n, u_2^{n+1}). By monotonicity of F_ν^{-1} the strict inequality is equivalent to $x \in (F_\nu^{-1}(u_1^{n+1}), F_\nu^{-1}(u_1^n))$ or $(F_\nu^{-1}(u_2^n), F_\nu^{-1}(u_2^{n+1}))$ for some $n \in \mathbb{N}^*$. We deduce that $\hat\nu \leq_{cx} \nu$ and therefore $\mu \leq_{cx} \bar\nu \leq_{cx} \nu$, and the irreducible components of $(\hat\nu, \nu)$ are given by the intervals

$$I_n^1 = (F_\nu^{-1}(u_1^{n+1}), F_\nu^{-1}(u_1^n)) \quad \text{and} \quad I_n^2 = (F_\nu^{-1}(u_2^n), F_\nu^{-1}(u_2^{n+1})), \quad n \in \mathbb{N}^*, \tag{6.7.29}$$

which are indeed nonempty by (6.7.28) and bounded. In particular for all $n \in \mathbb{N}^*$ we have

$$u_{\hat\nu}(F_\nu^{-1}(u_1^n)) = u_\nu(F_\nu^{-1}(u_1^n)) \quad \text{and} \quad u_{\hat\nu}(F_\nu^{-1}(u_2^n)) = u_\nu(F_\nu^{-1}(u_2^n)). \tag{6.7.30}$$

Since $u_{\hat\nu} \leq u_\mu \vee u_{\hat\nu} = u_{\bar\nu} \leq u_\nu$ and (μ, ν) is irreducible, (6.7.29) is also the decomposition into irreducible components of $(\hat\nu, \nu)$, which consists solely of bounded intervals.

To conclude, it remains to show that

$$\bar\nu \left(\mathbb{R} \setminus \bigcup_{n\in\mathbb{N}^*} (I_n^1 \cup I_n^2) \right) = \bar\nu \left(\left\{ F_\nu^{-1}(u_1^n) \mid n \in \mathbb{N}^* \right\} \cup \left\{ F_\nu^{-1}(u_2^n) \mid n \in \mathbb{N}^* \right\} \right) = 0. \tag{6.7.31}$$

For each $n \in \mathbb{N}^*$ and bounded neighbourhood of $F_\nu^{-1}(u_1^n)$, there is by irreducibility of (μ, ν) and continuity of potential functions a $\delta > 0$ such that $u_\mu + \delta < u_\nu$ on this neighbourhood. Thus, for y close enough to $F_\nu^{-1}(u_1^n)$, we have $u_\mu(y) < u_{\hat{\nu}}(y)$ due to (6.7.30), hence

$$u_{\bar{\nu}}(y) = u_{\hat{\nu}}(y) \vee u_\mu(y) = u_{\hat{\nu}}(y), \quad \text{for } y \text{ close enough to } F_\nu^{-1}(u_1^n). \tag{6.7.32}$$

For each $n \in \mathbb{N}^*$, it is clear from the definition of $\hat{\nu}$ that its restriction to the closure of I_n^1 is concentrated on a single point in I_n^1 and therefore does not charges the boundaries of I_n^1. We recall the easy fact that the potential function of a probability measure is linear on an open intervall iff this measure does not charge this interval. We deduce, with use of (6.7.32), that we can find an open neighbourhood of $F_\nu^{-1}(u_n^1)$ such that $u_{\hat{\nu}}$ and therefore $u_{\bar{\nu}}$ is linear, which implies that $\bar{\nu}$ does not put mass on $\{F_\nu^{-1}(u_1^n) \mid n \in \mathbb{N}^*\}$. Analogously, we find that $\tilde{\nu}$ does not charge $\{F_\nu^{-1}(u_2^n) \mid n \in \mathbb{N}^*\}$, which proves (6.7.31). $\square$

Bibliography

[1] M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

[2] D. J. Aldous. Weak convergence and general theory of processes. Unpublished incomplete draft of monograph; Department of Statistics, University of California, Berkeley, CA 94720, July 1981.

[3] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds. *International Journal of Theoretical and Applied Finance*, 22(3), 2019.

[4] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order by Wasserstein projection. *Annales de l'Institut Henri Poincaré B, Probability and Statistics*, 56(3):1706–1729, 2020.

[5] A. Alfonsi and B. Jourdain. A remark on the optimal transport between two probability measures sharing the same copula. *Statistics and Probability Letters*, 84:131–134, 2014.

[6] C. D. Aliprantis and K. C. Border. *Infinite dimensional analysis: A hitchhiker's guide.* Springer, 3rd edition, 2006.

[7] L. Ambrosio. Lecture notes on optimal transport problems. In *Mathematical aspects of evolving interfaces*, pages 1–52. Springer, 2003.

[8] L. Ambrosio and N. Gigli. A user's guide to optimal transport. In *Modelling and optimisation of flows on networks*, volume 2062 of *Lecture Notes in Math.*, pages 1–155. Springer, Heidelberg, 2013.

[9] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows: in metric spaces and in the space of probability measures.* Springer Science & Business Media, 2008.

[10] P. Appell. Mémoire sur les déblais et les remblais des systemes continus ou discontinus. *Mémoires présentes par divers Savants à l'Académie des Sciences de l'Institut de France*, 29:1–208, 1887.

[11] P. Appell. *Le probleme geometrique des déblais et remblais.* Gauthier-Villars, 1928.

[12] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34 th International Conference on Machine Learning, Sydney, Australia*, 2017.

[13] L. Bachelier. Théorie de la spéculation. In *Annales scientifiques de l'École normale supérieure*, volume 17, pages 21–86, 1900.

[14] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. Adapted wasserstein distances and stability in mathematical finance. *Finance and Stochastics*, 24(3):601–632, 2020.

[15] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probability Theory and Related Fields*, pages 1–48, 2020.

[16] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and J. Wiesel. Estimating processes in adapted wasserstein distance. *arXiv e-prints:2002.07261*, February 2020.

[17] J. Backhoff-Veraguas, M. Beiglböck, M. Huesmann, and S. Källblad. Martingale Benamou–Brenier: a probabilistic perspective. *Annals of Probability*, 48(5):2258–2289, 2020.

[18] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Existence, duality, and cyclical monotonicity for weak transport costs. *Calculus of Variations and Partial Differential Equations*, 58(6):203, 2019.

[19] J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer. Weak monotone rearrangement on the line. *Electronic Communications in Probability*, 25(18):1–16, 2020.

[20] J. Backhoff-Veraguas and G. Pammer. Stability of martingale optimal transport and weak optimal transport. *arXiv e-prints:1904.04171*, April 2019.

[21] J. Backhoff-Veraguas and G. Pammer. Applications of weak transport theory. *arXiv preprint arXiv:2003.05338*, 2020.

[22] M. Beiglböck, A. M. Cox, and M. Huesmann. Optimal transport and skorokhod embedding. *Inventiones mathematicae*, 208(2):327–400, 2017.

[23] M. Beiglböck, P. Henry-Labordère, and F. Penkner. Model-independent bounds for option prices: A mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[24] M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and skorokhod embedding. *Stochastic Processes and their Applications*, 127(9):3005–3013, 2017.

[25] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Annals of Probability*, 44(1):42–106, 2016.

[26] M. Beiglböck, T. Lim, and J. Obłój. Dual attainment for the martingale transport problem. *Bernoulli*, 25(3):1640–1658, 2019.

[27] M. Beiglböck, M. Nutz, and N. Touzi. Complete Duality for Martingale Optimal Transport on the Line. *Annals of Probability*, 45(5):3038–3074, 2017.

[28] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[29] J.-D. Benamou and Y. Brenier. Mixed L2-wasserstein optimal mapping between prescribed density functions. *Journal of Optimization Theory and Applications*, 111(2):255–271, 2001.

[30] J.-D. Benamou, Y. Brenier, and K. Guittet. The Monge–Kantorovitch mass transfer and its computational fluid mechanics formulation. *International Journal for Numerical methods in fluids*, 40(1-2):21–30, 2002.

[31] J.-D. Benamou and G. Carlier. Augmented lagrangian methods for transport optimization, mean field games and degenerate elliptic equations. *Journal of Optimization Theory and Applications*, 167(1):1–26, 2015.

[32] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138, 2015.

[33] J.-D. Benamou, F. Collino, and J.-M. Mirebeau. Monotone and consistent discretization of the Monge-Ampere operator. *Mathematics of computation*, 85(302):2743–2775, 2016.

[34] J.-D. Benamou and V. Duval. Minimal convex extensions and finite difference discretisation of the quadratic Monge–Kantorovich problem. *European Journal of Applied Mathematics*, 30(6):1041–1078, 2019.

[35] J.-D. Benamou, B. D. Froese, and A. M. Oberman. Numerical solution of the optimal transportation problem using the Monge–Ampère equation. *Journal of Computational Physics*, 260:107–126, 2014.

[36] D. P. Bertsekas. A new algorithm for the assignment problem. *Mathematical Programming*, 21(1):152–171, 1981.

[37] D. P. Bertsekas. Auction algorithms for network flow problems: A tutorial introduction. *Computational optimization and applications*, 1(1):7–66, 1992.

[38] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[39] J. Bion-Nadal and D. Talay. On a wasserstein-type distance between solutions to stochastic differential equations. *Annals of Applied Probability*, 29(3):1609–1639, 2019.

[40] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[41] V. I. Bogachev. *Measure theory*, volume 1. Springer Science and Business Media, 2007.

[42] V. I. Bogachev. *Measure theory*, volume 2. Springer Science and Business Media, 2007.

[43] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. *Probability Theory and Related Fields*, 137(3-4):541–593, 2007.

[44] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[45] G. Bouchitté, C. Jimenez, and M. Rajesh. Asymptotique d'un problème de positionnement optimal. *Comptes Rendus Mathematique*, 335(10):853–858, 2002.

[46] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651, 1978.

[47] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *CR Acad. Sci. Paris Sér. I Math.*, 305:805–808, 1987.

[48] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[49] G. Buttazzo, C. Jimenez, and E. Oudet. An optimization problem for mass transportation with congested dynamics. *SIAM Journal on Control and Optimization*, 48(3):1961–1976, 2009.

[50] S. Cambanis, G. Simons, and W. Stout. Inequalities for ek (x, y) when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 36(4):285–294, 1976.

[51] G. Carlier, V. Duval, G. Peyré, and B. Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2):1385–1418, 2017.

[52] M. Cassandro, E. Olivieri, A. Pellegrinotti, and E. Presutti. Existence and uniqueness of dlr measures for unbounded spin systems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 41(4):313–334, 1978.

[53] P. Cattiaux, A. Guillin, and F. Malrieu. Probabilistic approach for granular media equations in the non-uniformly convex case. *Probability theory and related fields*, 140(1-2):19–40, 2008.

[54] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018.

[55] R. Cominetti and J. San Martín. Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1-3):169–187, 1994.

[56] N. Courty, R. Flamary, D. Tuia, and A. Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2016.

[57] M. J. P. Cullen. *A mathematical theory of large-scale atmosphere/ocean flow*. World Scientific, 2006.

[58] M. J. P. Cullen and R. Douglas. Applications of the Monge–Ampère equation and Monge transport problem to meteorology and oceanography. *Monge–Ampère Equation: Applications to Geometry and Optimization*, 226:33, 1999.

[59] M. J. P. Cullen and M. Feldman. Lagrangian solutions of semigeostrophic equations in physical space. *SIAM journal on mathematical analysis*, 37(5):1371–1395, 2006.

[60] M. J. P. Cullen and W. Gangbo. A variational approach for the 2-dimensional semigeostrophic shallow water equations. *Archive for rational mechanics and analysis*, 156(3):241–273, 2001.

[61] M. J. P. Cullen and R. J. Purser. Properties of the lagrangian semigeostrophic equations. *Journal of the atmospheric sciences*, 46(17):2684–2697, 1989.

[62] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013.

[63] G. B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. in ed. tjalling c. koopmans. activity analysis of production and allocation, cowles commission monograph no. 13. In *Proceedings of Linear Programming Conference*, 1949.

[64] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *The annals of mathematical statistics*, pages 1470–1480, 1972.

[65] F. De Goes, K. Breeden, V. Ostromoukhov, and M. Desbrun. Blue noise through optimal transport. *ACM Transactions on Graphics (TOG)*, 31(6):1–11, 2012.

[66] H. De March. Entropic approximation for multi-dimensional martingale optimal transport. *arXiv e-prints:1812.11104*, December 2018.

[67] H. De March. Local structure of multi-dimensional martingale optimal transport. *arXiv e-prints:1805.09469*, November 2018.

[68] H. De March. Quasi-sure duality for multi-dimensional martingale optimal transport. *arXiv e-prints:1805.01757*, May 2018.

[69] H. De March and N. Touzi. Irreducible convex paving for decomposition of multi-dimensional martingale transport plans. *Annals of Probability*, 47(3):1726–1774, 2019.

[70] L. Denis and C. Martini. A theoretical framework for the pricing of contingent claims in the presence of model uncertainty. *The Annals of Applied Probability*, 16(2):827–852, 2006.

[71] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability and Its Applications*, 15(3):458–486, 1970.

[72] R. M. Dudley. Probability and metrics. *Aarhus Lecture Notes, Aarhus University*, 1976.

[73] J. Dugundji. An extension of tietze's theorem. *Pacific Journal of Mathematics*, 1(3):353–367, 1951.

[74] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[75] B. Dupire. Pricing and hedging with smiles. *Mathematics of derivative securities*, 1(1):103–111, 1997.

[76] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Applied Mathematics and Optimization*, pages 1–29, 2019.

[77] M. Eder. Compactness in adapted weak topologies. *arXiv e-prints:1905.00856*, May 2019.

[78] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley-Blackwell, Hoboken, 2nd revised edition, 2005.

[79] H. Föllmer and A. Schied. *Stochastic finance: an introduction in discrete time.* Walter de Gruyter, 2011.

[80] R. Fortet and E. Mourier. Convergence de la répartition empirique vers la répartition théorique. In *Annales scientifiques de l'École Normale Supérieure*, volume 70, pages 267–285, 1953.

[81] M. Fréchet. Sur la distance de deux lois de probabilité. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 244(6):689–692, 1957.

[82] B. D. Froese. A numerical method for the elliptic Monge–Ampère equation with transport boundary conditions. *SIAM Journal on Scientific Computing*, 34(3):A1432–A1459, 2012.

[83] C. Frogner, C. Zhang, H. Mobahi, M. Araya-Polo, and T. Poggio. Learning with a wasserstein loss. In *Advances in neural information processing systems*, pages 2053–2061, 2015.

[84] A. Galichon, P. Henry-Labordère, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *Annals of Applied Probability*, 24(1):312–336, 2014.

[85] A. Galichon and B. Salanié. Matching with trade-offs: Revealed preferences over competing characteristics. 2010.

[86] W. Gangbo and R. J. McCann. Shape recognition via wasserstein distance. *Quarterly of Applied Mathematics*, pages 705–737, 2000.

[87] A. d. S. Germaine. Études sur le problème des déblais et remblais. *Extrait des Mémoires de l'Académie nationale de Caen, 1SS6*, 1886.

[88] N. Gozlan and C. Léonard. Transport inequalities. a survey. *Markov Processes And Related Fields*, 16(4):635–736, 2010.

[89] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 54(3):1667–1693, 2018.

[90] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327–3405, 2017.

[91] R. M. Gray, D. L. Neuhoff, and P. C. Shields. A generalization of ornstein's d distance with applications to information theory. *The Annals of Probability*, pages 315–328, 1975.

[92] R. M. Gray, D. S. Ornstein, and R. L. Dobrushin. Block synchronization, sliding-block coding, invulnerable sources and zero error codes for discrete noisy channels. *The Annals of Probability*, pages 639–674, 1980.

[93] C. Griessler. An extended footnote on finitely minimal martingale measures. *arXiv preprint arXiv:1606.03106*, 2016.

[94] L. A. Grzelak and C. W. Oosterlee. On the heston model with stochastic interest rates. *SIAM Journal on Financial Mathematics*, 2(1):255–286, 2011.

[95] X. Gu, F. Luo, J. Sun, and S.-T. Yau. Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge–Ampère equations. *Asian Journal of Mathematics*, 20(2):383–398, 2016.

[96] G. Guo and J. Obłój. Computational methods for martingale optimal transport problems. *Annals of Applied Probability*, 29(6):3311–3347, 2019.

[97] G. Guo, X. Tan, and N. Touzi. Tightness and duality of martingale transport on the skorokhod space. *Stochastic Processes and their Applications*, 127(3):927–956, 2017.

[98] C. E. Gutiérrez and H. Brezis. *The Monge–Ampère equation*, volume 44. Springer, 2001.

[99] J. Guyon, R. Menegaux, and M. Nutz. Bounds for vix futures given S&P 500 smiles. *Finance and Stochastics*, 21(3):593–630, 2017.

[100] G. H. Hardy, J. Littlewood, and G. Pólya. Inequalities. Cambridge Univ. Press., 1951.

[101] M. F. Hellwig. Sequential decisions under uncertainty and the maximum theorem. *Journal of Mathematical Economics*, 25(4):443–464, 1996.

[102] P. Henry-Labordère. Automated option pricing: Numerical methods. *International Journal of Theoretical and Applied Finance*, 16(08):1350042, 2013.

[103] P. Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint.* CRC Press, 2017.

[104] P. Henry-Labordère. (Martingale) optimal transport and anomaly detection with neural networks: A primal-dual algorithm. *arXiv preprint arXiv:1904.04546*, 2019.

[105] P. Henry-Labordère, X. Tan, and N. Touzi. An explicit martingale version of the one-dimensional Brenier's theorem with full marginals constraint. *Stochastic Processes and their Applications*, 126(9):2800–2834, 2016.

[106] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance and Stochastics*, 20(3):635–668, 2016.

[107] S. L. Heston. 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, 1993.

[108] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I: Fundamentals*, volume 305. Springer-Verlag Berlin Heidelberg, 1993.

[109] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance and Stochastics*, 19(1):189–214, 2015.

[110] D. Hobson and A. Neuberger. Robust Bounds for Forward Start Options. *Mathematical Finance*, 22(1):31–56, 2012.

[111] P. J. Huber. *Robust statistical procedures.* SIAM, 1996.

[112] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 52(4):1823–1843, 2016.

[113] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[114] L. V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk SSSR*, 37(7-8):199–201, 1942.

[115] L. V. Kantorovich. On a problem of Monge. *CR (Doklady) Acad. Sci. URSS (NS)*, 3:225–226, 1948.

[116] L. V. Kantorovich and G. P. Akilov. *Functional Analysis*. Moscow: Nauka, 1977.

[117] L. V. Kantorovich and M. L. Gavurin. Application of mathematical methods to problems of analysis of freight flows. *Problems of raising the efficiency of transport performance*, pages 110–138, 1949.

[118] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[119] J. Kemperman. On the role of duality in the theory of moments. In *Semi-infinite programming and applications*, pages 63–92. Springer, 1983.

[120] R. P. Kertz and U. Rösler. Stochastic and convex orders and lattices of probability measures, with a martingale interpretation. *Israel Journal of Mathematics*, 77(1-2):129–164, 1992.

[121] R. P. Kertz and U. Rösler. Complete lattices of probability measures with applications to martingale theory. *Lecture Notes-Monograph Series*, 35:153–177, 2000.

[122] L. G. Khachiyan. A polynomial algorithm in linear programming. In *Doklady Akademii Nauk*, volume 244, pages 1093–1096. Russian Academy of Sciences, 1979.

[123] J. Kitagawa. An iterative scheme for solving the optimal transportation problem. *Calculus of Variations and Partial Differential Equations*, 51(1-2):243–263, 2014.

[124] J. Kitagawa, Q. Mérigot, and B. Thibert. Convergence of a newton algorithm for semi-discrete optimal transport. *Journal of the European Mathematical Society*, 21(9):2603–2651, 2019.

[125] S. G. Kou. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101, 2002.

[126] H. W. Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

[127] D. Lacker. Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport. *arXiv e-prints:1805.03185*, May 2018.

[128] D. Lamberton and B. Lapeyre. *Introduction to stochastic calculus applied to finance.* CRC press, 2007.

[129] R. Lassalle. Causal transference plans and their Monge-Kantorovich problems. *Stochastic Analysis and Applications*, 36(3):452–484, 2018.

[130] M. Ledoux. *The concentration of measure phenomenon.* Number 89. American Mathematical Soc., 2001.

[131] C. Léonard. From the schrödinger problem to the Monge–Kantorovich problem. *Journal of Functional Analysis*, 262(4):1879–1920, 2012.

[132] V. L. Levin and A. A. Milyutin. The mass transfer problem with discontinuous cost function and a mass setting for the problem of duality of convex extremum problems. *Trans Russian Math. Surveys*, 34:1–78, 1979.

[133] B. Lévy. A numerical algorithm for l2 semi-discrete optimal transport in 3d. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1693–1715, 2015.

[134] D. Lombardi and E. Maitre. Eulerian models and algorithms for unbalanced optimal transport. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1717–1744, 2015.

[135] G. G. Lorentz. An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179, 1953.

[136] F. Malrieu. Logarithmic sobolev inequalities for some nonlinear PDE's. *Stochastic processes and their applications*, 95(1):109–132, 2001.

[137] Q. Mérigot. A multiscale approach to optimal transport. In *Computer Graphics Forum*, volume 30, pages 1583–1592. Wiley Online Library, 2011.

[138] R. C. Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[139] J.-M. Mirebeau. Discretization of the 3d Monge–Ampère operator, between wide stencils and power diagrams. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(5):1511–1523, 2015.

[140] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[141] T. F. Móri. Sharp inequalities between centered moments. *Journal of inequalities in pure and applied mathematics*, 10(4):Art. 99, 2009.

[142] M. Neilan, A. J. Salgado, and W. Zhang. *The Monge–Ampère equation*, volume 21 of *Handbook of Numerical Analysis*, pages 105–219. Elsevier B.V., 2020.

[143] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electronic Journal of Probability*, 18, 2013.

[144] J. Obłój. The skorokhod embedding problem and its offspring. *Probability Surveys*, 1:321–392, 2004.

[145] V. I. Oliker and L. D. Prussner. On the numerical solution of the equation and its discretizations, i. *Numerische Mathematik*, 54(3):271–293, 1989.

[146] I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra and its Applications*, 48:257–263, 1982.

[147] D. S. Ornstein. Ergodic theory, randomness and dynamical systems. *Yale Math. Monographs*, 5, 1974.

[148] N. Papadakis, G. Peyre, and E. Oudet. Optimal transport with proximal splitting. *SIAM Journal on Imaging Sciences*, 7(1):212, 2014.

[149] G. C. Pflug and A. Pichler. A distance for multistage stochastic optimization models. *SIAM Journal on Optimization*, 22(1):1–23, 2012.

[150] G. C. Pflug and A. Pichler. *Multistage stochastic optimization.* Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2014.

[151] G. C. Pflug and A. Pichler. Dynamic generation of scenario trees. *Computational Optimization and Applications*, 62(3):641–668, 2015.

[152] G. C. Pflug and A. Pichler. From empirical observations to tree models for stochastic optimization: convergence properties. *SIAM Journal on Optimization*, 26(3):1715–1740, 2016.

[153] A. Pilipenko. *An introduction to stochastic differential equations with reflection.* Lectures in pure and applied mathematics. Universitätsverlag Potsdam, Potsdam, 2014.

[154] D. Possamaï, G. Royer, and N. Touzi. On the robust superhedging of measurable claims. *Electronic Communications in Probability*, 18, 2013.

[155] N. Privault. *Stochastic finance: An introduction with market examples.* CRC Press, 2013.

[156] Y. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory of Probability and Its Applications*, 1(2):157–214, 1956.

[157] J. Rabin and N. Papadakis. Convex color image segmentation with optimal transport distances. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 256–269. Springer, 2015.

[158] J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

[159] S. T. Rachev. On minimal metrics in a space of real-valued random variables. In *Doklady Akademii Nauk*, volume 257, pages 1067–1070. Russian Academy of Sciences, 1981.

[160] S. T. Rachev. The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and Its Applications*, 29(4):647–676, 1985.

[161] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory.* Springer Science & Business Media, 1998.

[162] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume II: Applications.* Springer Science & Business Media, 1998.

[163] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001.

[164] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg, 3 edition, 1999.

[165] E. Rio. Upper bounds for minimal distances in the central limit theorem. In *Annales de l'IHP Probabilités et statistiques*, volume 45, pages 802–817, 2009.

[166] G. S. Rubinstein. Duality in mathematical programming and some questions of convex analysis. *Uspekhi mat. nauk*, 25(5):155, 1970.

[167] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99–121, 2000.

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

[169] M. Shaked and G. Shanthikumar. *Stochastic Orders.* Springer Series in Statistics. Springer-Verlag, New York, 2007.

[170] S. Shirdhonkar and D. W. Jacobs. Approximate earth mover's distance in linear time. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[171] S. Shreve. *Stochastic calculus for finance I: the binomial asset pricing model.* Springer Science & Business Media, 2005.

[172] S. E. Shreve. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer Science & Business Media, 2004.

[173] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.

[174] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.

[175] M. Sklar. Fonctions de repartition à n dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

[176] A. V. Skorokhod. *Issledovaniia po teorii sluchainykh protsessov.* Izd-vo Kievskogo universiteta, 1961.

[177] A. V. Skorokhod. *Studies in the theory of random processes*, volume 7021. Courier Dover Publications, 1982.

[178] S. Smale. On the average number of steps of the simplex method of linear programming. *Mathematical programming*, 27(3):241–262, 1983.

[179] J. Solomon, F. De Goes, G. Peyré, M. Cuturi, A. Butscher, A. Nguyen, T. Du, and L. Guibas. Convolutional wasserstein distances: Efficient optimal transportation on geometric domains. *ACM Transactions on Graphics (TOG)*, 34(4):1–11, 2015.

[180] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. Earth mover's distances on discrete surfaces. *ACM Transactions on Graphics (TOG)*, 33(4):1–12, 2014.

[181] H. M. Soner, N. Touzi, and J. Zhang. Dual formulation of second order target problems. *The Annals of Applied Probability*, 23(1):308–347, 2013.

[182] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[183] V. Strassen. The existence of probability measures with given marginals. *Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[184] V. N. Sudakov. *Geometric problems in the theory of infinite-dimensional probability distributions*, volume 141. American Mathematical Soc., 1979.

[185] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[186] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *The annals of probability*, 41(5):3201–3240, 2013.

[187] M. Thorpe, S. Park, S. Kolouri, G. K. Rohde, and D. Slepčev. A transportation Lp distance for signal analysis. *Journal of mathematical imaging and vision*, 59(2):187–210, 2017.

[188] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Physical review*, 36(5):823, 1930.

[189] A. M. Vershik. Some remarks on the infinite-dimensional linear programming problems. *Uspekhi Mat. Nauk*, 25:117–124, 1970.

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

[191] C. Villani. *Optimal Transport, Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2009.

[192] M. Y. Vygodskii. The origin of the differential geometry. *Monge G. Prilozhenie analiza k geometrii*, pages 7–70, 1936.

[193] L. N. Wasserstein. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72, 1969.

[194] J. Wiesel. Continuity of the martingale optimal transport problem on the real line. *arXiv e-prints:1905.04574*, January 2020.

[195] J. E. Yukich. Optimal matching and empirical measures. *Proceedings of the American Mathematical Society*, 107(4):1051–1059, 1989.

[196] V. M. Zolotarev. General problems of the stability of mathematical models. *Bull. Int. Stat. Inst*, 47(2):382–401, 1977.

[197] V. M. Zolotarev. Probability metrics. *Teoriya Veroyatnostei i ee Primeneniya*, 28(2):264–287, 1983.