Optimal Transport

Benjamin JOURDAIN

January 19, 2026

Contents

Acknowledgements: These notes are strongly inspired from [1], that I thank Gudmund Pammer for sharing.

Chapter 1

Optimal Transport

1.1 From Monge to Kantorovich

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces for the distances $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$. We denote by $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ the sets of probability measures on $\mathcal{X}$ and $\mathcal{Y}$ respectively.

Definition 1.1. *The image of the probability measure $\mu \in \mathcal{P}(\mathcal{X})$ by $T : \mathcal{X} \to \mathcal{Y}$ measurable is the probability measure $T\#\mu \in \mathcal{P}(\mathcal{Y})$ defined by*

$$\forall B \in \mathcal{B}(\mathcal{Y}),\ T\#\mu(B) = \mu(T^{-1}(B)),$$

where $T^{-1}(B)$ is the preimage of B by T.

The Monge formulation [13] of the optimal transport problem from $\mu \in \mathcal{P}(\mathcal{X})$ to $\nu \in \mathcal{P}(\mathcal{Y})$ with cost $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ measurable is

$$V_c^{\text{Monge}}(\mu, \nu) = \inf_{T:T\#\mu=\nu} \int_{\mathcal{X}} c(x, T(x))\mu(dx).$$

This formulation has several drawbacks. First when $\mathcal{X} = \mathcal{Y} = \mathbb{R}$, $\mu = \delta_0$ and $\nu = \frac{1}{2}(\delta_0 + \delta_1)$, there is no transport map T such that $T\#\mu = \nu$ since $T\#\mu = \delta_{T(0)}$. Moreover, the set of transport maps is neither convex nor sequentially compact. Kantorovich [12] introduced a relaxation of this problem by considering couplings instead of maps. Let

$$\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \int_{y\in\mathcal{Y}} \pi(dx, dy) = \mu(dx) \text{ and } \int_{x\in\mathcal{X}} \pi(dx, dy) = \nu(dy) \right\}.$$

denote the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ with first marginal equal to μ and second marginal equal to ν. The Kantorovich formulation is

$$V_c(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \pi(c).$$

Remark 1.2.

- *Let $i_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$ denote the identity function on $\mathcal{X}$ defined by $i_{\mathcal{X}}(x) = x$ for $x \in \mathcal{X}$. Since when $T\#\mu = \nu$, then $(i_{\mathcal{X}}, T)\#\mu \in \Pi(\mu, \nu)$ and $(i_{\mathcal{X}}, T)\#\mu(c) = \int_{\mathcal{X}} c(x, T(x))\mu(dx)$, we have*

$$V_c^{\text{Monge}}(\mu, \nu) \geq V_c(\mu, \nu).$$

- *The set* $\Pi(\mu,\nu)$ *is never empty since* $\mu\otimes\nu(dx,dy)=\mu(dx)\nu(dy)\in\Pi(\mu,\nu)$. *Moreover this set is convex and has nice compactness properties stated in Lemma 1.7 below.*

Proposition 1.3. *When* μ *is atomless (i.e.* $\mu(\{x\})=0$ *for all* $x\in\mathcal{X}$*), then the set of Monge couplings* $\{(i_\mathcal{X},T)\#\mu : T:\mathcal{X}\to\mathcal{Y}$ *measurable*$\}$ *is dense in* $\Pi(\mu,\nu)$ *and* $V_c^{\text{Monge}}(\mu,\nu)=V_c(\mu,\nu)$.

Proof: Let $\pi(dx,dy)=\mu(dx)\pi_x(dy)\in\Pi(\mu,\nu)$. By the fundamental theorem of simulation, there exists $S:\mathcal{X}\times[0,1]\to\mathcal{Y}$ measurable such that $S(x,\cdot)\#1_{[0,1]}(u)du=\pi_x$, $\mu(dx)$ a.e.. Let $p_\mathcal{X}:\mathcal{X}\times[0,1]\to\mathcal{X}$ be defined by $p_\mathcal{X}(x,u)=x$. Then $(p_\mathcal{X},S)\#(\mu\otimes 1_{[0,1]}(u)du)=\mu\otimes\pi_x=\pi$. Let $(x_m)_{m\in\mathbb{N}}$ be dense in $\mathcal{X}$ and for $n\in\mathbb{N}^*$, $B_0^n=\mathcal{B}(x_0,\frac{1}{n})$ and for $m\geq 1$, $B_m^n=\mathcal{B}(x_m,\frac{1}{n})\cap\left\{\bigcup_{k=0}^{m-1}\mathcal{B}(x_k,\frac{1}{n})\right\}^c$.

Since μ is atomless, we can find $R^{n,m}:B_m^n\to B_m^n\times[0,1]$ measurable such that $R^{n,m}\#\mu|_{B_m^n}=\mu|_{B_m^n}\otimes 1_{[0,1]}(u)du$. We set

$$R^n(x)=\sum_{m\in\mathbb{N}}1_{B_m^n}(x)R^{n,m}(x).$$

When $X\sim\mu$, $R^n(X)\sim\mu\otimes 1_{[0,1]}(u)du$ and $(p_\mathcal{X}(R^n(X)),S(R^n(X)))\sim\pi$.

When $x\in B_m^n$, $d_\mathcal{X}(x,p_\mathcal{X}(R^n(x)))\leq d_\mathcal{X}(x,x_m)+d_\mathcal{X}(x_m,p_\mathcal{X}(R_m^n(x)))\leq\frac{2}{n}$. Hence $\forall x\in\mathcal{X}$, $d_\mathcal{X}(x,p_\mathcal{X}(R^n(x)))\leq\frac{2}{n}$. Therefore for $\varphi:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ Lipschitz continuous and bounded

$$\begin{aligned}|\mathbb{E}[\varphi(X,S(R^n(X)))]-\pi(\varphi)|&=|\mathbb{E}[\varphi(X,S(R^n(X)))]-\mathbb{E}[\varphi(p_\mathcal{X}(R^n(X)),S(R^n(X)))]|\\&\leq\mathbb{E}\left[|\varphi(X,S(R^n(X)))-\varphi(p_\mathcal{X}(R^n(X)),S(R^n(X)))|\right]\leq\text{Lip}(\varphi)\times\frac{2}{n}\overset{n\to\infty}{\longrightarrow}0.\end{aligned}$$

By the Portmanteau theorem (see Theorem 5.2), we conclude that the law of $(X,S(R^n(X)))$ converges weakly to π as $n\to\infty$. ∎

The following uniqueness criterion for optimal couplings will be used in the proof of Brenier's theorem.

Proposition 1.4. *If any optimal coupling* $\pi_\star\in\Pi(\mu,\nu)$ *for* $V_c(\mu,\nu)$ *is a Monge coupling, then there is at most one optimal coupling.*

Proof: Let $\pi_\star,\tilde{\pi}_\star\in\Pi(\mu,\nu)$ be optimal. Then $\pi_\star(dx,dy)=\mu(dx)\delta_{T(x)}(dy)$ and $\tilde{\pi}_\star(dx,dy)=\mu(dx)\delta_{\tilde{T}(x)}(dy)$ for some measurable maps $T,\tilde{T}:\mathcal{X}\to\mathcal{Y}$. Since $V_c(\mu,\nu)=(\frac{1}{2}(\pi+\tilde{\pi}_\star))(c)$, $\frac{1}{2}(\pi+\tilde{\pi}_\star)(dx,dy)=\mu(dx)\frac{1}{2}\left(\delta_{T(x)}+\delta_{\tilde{T}(x)}\right)(dy)$ also is optimal and therefore a Monge coupling. As a consequence $\mu(dx)$ a.e., $\tilde{T}(x)=T(x)$ and $\tilde{\pi}_\star=\pi_\star$. ∎

1.2 The case of finitely supported probability measures

In this case, the Kantorovich formulation of the optimal transport problem is a linear programming problem with finitely many variables and can be solved efficiently by the simplex algorithm or the interior-point method.

Let $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_I \end{pmatrix} \in [0,1]^I$ and $\nu = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_J \end{pmatrix} \in [0,1]^J$ be such that $\sum_{i=1}^I \mu_i = 1 = \sum_{j=1}^J \nu_j$. For $c = (c_{ij})_{1\le i\le I, 1\le j\le J}$, we want to solve

$$\min \sum_{i=1}^I \sum_{j=1}^J c_{ij}\pi_{ij} \text{ under the constraints } \forall (i,j) \in \{1,\cdots,I\}\times\{1,\cdots,J\},\ \pi_{ij} \ge 0,$$

$$\forall i \in \{1,\cdots,I\},\ \sum_{j=1}^J \pi_{ij} = \mu_i \text{ and } \forall j \in \{1,\cdots,J\},\ \sum_{i=1}^I \pi_{ij} = \nu_i.$$

Since the under the constraints, $\pi = (\pi_{ij})_{1\le i\le I, 1\le j\le J}$ describes a closed and bounded and therefore compact subset of $\mathbb{R}^{I\times J}$ and $\mathbb{R}^{I\times J} \ni \pi \mapsto c.\pi = \sum_{i=1}^I \sum_{j=1}^J c_{ij}\pi_{ij}$ is continuous, there exists an optimizer $\pi^\star$.

Let us now characterize the set of optimizers. Up to replacing c_{ij} by $c_{ij} - \min_{1\le k\le I, 1\le \ell\le J} c_{k\ell}$, we may suppose that the entries of c are non-negative, which we write $c \ge 0$. Then the optimization problem is equivalent to

$$(P)\ \min c.\pi \text{ under the constraints } A\pi \ge \begin{pmatrix} \mu \\ \nu \\ O_{I\times J} \end{pmatrix},$$

with $A = \begin{pmatrix} B \\ I_{IJ} \end{pmatrix} \in \mathbb{R}^{(I+J+IJ)\times IJ}$ and $B \in \mathbb{R}^{(I+J)\times IJ}$ giving the marginal constraints that one should saturate since $c \ge 0$. Let us extract from A the rows such that the corresponding entries of $A\pi^\star$ and $\begin{pmatrix} \mu \\ \nu \\ O_{I\times J} \end{pmatrix}$ are equal and in particular the $I+J$ first rows. Let $\tilde{A} \in \mathbb{R}^{d\times IJ}$ with $d \ge I+J$ denote the matrix with these rows. The next lemma (see for instance Theorem 22.3 [15]) is a consequence of the separation of convexes and therefore of the Hahn-Banach theorem.

Lemma 1.5 (Farkas). *Either $\exists \eta \in \mathbb{R}^{IJ}$ such that $\tilde{A}\eta \ge 0$ and $c.\eta < 0$ or $\exists h \in \mathbb{R}^d$ such that $\tilde{A}^T h = c$ and $h \ge 0$.*

Indeed, $\{\tilde{A}^T h : h \ge 0\}$ is the closed cone spanned by the columns of $\tilde{A}^T$ and when c is not in this cone, we can separate $\{c\}$ and the cone by the hyperplane directed by η. Because of the optimality of $\pi^\star$,

$$\forall \eta \in \mathbb{R}^{IJ} \text{ such that } \tilde{A}\eta \ge 0 \text{ coordinate-wise},\ c.\eta \ge 0.$$

Otherwise, we could get a smaller value and still respect the constraints by adding $\varepsilon\eta$ to $\pi^\star$ with $\varepsilon > 0$ small enough. Therefore, by Farkas lemma, there exists $h = \begin{pmatrix} \phi^\star \\ \psi^\star \\ \tilde{h} \end{pmatrix}$ with $\phi^\star \in \mathbb{R}^I$, $\psi^\star \in \mathbb{R}^J$ and $\tilde{h} \in \mathbb{R}^{d-(I+J)}$ such that $\tilde{A}^T h = c$ and $h \ge 0$. In the row of $\tilde{A}^T$ corresponding to π_{ij}, we have coefficient 1 on the i-th column (constraint $\sum_{\ell=1}^J \pi_{i\ell} = \mu_i$), on the $I+j$-th column (constraint $\sum_{k=1}^{I} \pi_{kj} = \nu_j$) and possibly on one of the $d-(I+J)$ last columns when $\pi^\star_{ij} = 0$ while all the other entries are equal to 0. Therefore

$$c_{ij} = \begin{cases} \phi^\star_i + \psi^\star_j \text{ when } \pi^\star_{ij} > 0 \\ \phi^\star_i + \psi^\star_j + \tilde{h}_{ij} \text{ when } \pi^\star_{ij} = 0 \end{cases}.$$

Let us consider the dual problem

$$(D)\ \max \left\{ \sum_{i=1}^{I} \phi_i \mu_i + \sum_{j=1}^{J} \psi_j \nu_j \right\} \text{ under the constraints}$$
$$\forall (i,j) \in \{1,\cdots,I\} \times \{1,\cdots,J\},\ \phi_i + \psi_j \leq c_{ij}.$$

For π satisfying the constraints in the primal problem and ϕ, ψ satisfying those in the dual problem,

$$\sum_{i=1}^{I} \phi_i \mu_i + \sum_{j=1}^{J} \psi_j \nu_j \leq \sum_{i=1}^{I} \sum_{j=1}^{J} (\phi_i + \psi_j) \pi_{ij} \leq c.\pi.$$

We deduce that $(D) \leq (P)$. For π, ϕ, ψ replaced by $(\pi^\star, \phi^\star, \psi^\star)$ both inequalities are equalities so that $(D) = (P)$.

Moreover, the complementary slackness property (propriété des écarts complémentaires en français) holds : if $\pi^\star$ is optimal for (P) and $(\phi^\star, \psi^\star)$ for (D), then $\phi^\star_i + \psi^\star_j = c_{ij}$ for each (i,j) such that $\pi^\star_{ij} > 0$. Last, the support $\{(i,j) : \pi^\star_{ij} > 0\}$ of $\pi^\star$ is c-cyclically monotone in the sense that if $(i_1, j_1), \cdots, (i_k, j_k)$ are in the support, then

$$\sum_{\ell=1}^{k} c_{i_\ell j_\ell} = \sum_{\ell=1}^{k} (\phi^\star_{i_\ell} + \psi^\star_{j_\ell}) = \sum_{\ell=1}^{k} (\phi^\star_{i_\ell} + \psi^\star_{j_{\ell+1}}) \leq \sum_{\ell=1}^{k} c_{i_\ell j_{\ell+1}},$$

under the convention $j_{k+1} = j_1$. We will next investigate the dual formulation, the complementary slackness condition and the c-cyclical monotonicity for general probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$.

1.3 Study of the primal Kantorovich formulation

Theorem 1.6. *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous bounded from below. Then there exists $\pi_\star \in \Pi(\mu,\nu)$ such that $\pi_\star(c) = V_c(\mu,\nu)$. Moreover, $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \in (\mu,\nu) \mapsto V_c(\mu,\nu)$ is lower semi-continuous and convex.*

The proof relies on the next lemma.

Lemma 1.7. *The set $\Pi(\mu,\nu)$ is compact for the weak convergence topology. Moreover when $\mu_n \to \mu$ and $\nu_n \to \nu$ weakly as $n \to \infty$, then from any sequence $\pi_n \in \Pi(\mu_n,\nu_n)$ we may extract a subsequence converging weakly to $\pi_\infty \in \Pi(\mu,\nu)$.*

Proof of Theorem 1.6: Let $(\pi_n)_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$ be a minimizing sequence for $V_c(\mu,\nu)$. By Lemma 1.7, we may extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ converging weakly to $\pi_\star \in \Pi(\mu,\nu)$. By the Portmanteau theorem (see Theorem 5.2 7)),

$$V_c(\mu,\nu) = \liminf_{k\to\infty} \pi_{n_k}(c) \geq \pi_\star(c) \geq V_c(\mu,\nu).$$

Let $\tilde{\mu} \in \mathcal{P}(\mathcal{X})$, $\tilde{\nu} \in \mathcal{P}(\mathcal{Y})$, $\tilde{\pi}_\star \in \Pi(\tilde{\mu}, \tilde{\nu})$ be optimal for $V_c(\tilde{\mu}, \tilde{\nu})$ and $\alpha \in [0,1]$. Then $\alpha\pi_\star + (1-\alpha)\tilde{\pi}_\star \in \Pi(\alpha\mu + (1-\alpha)\tilde{\mu}, \alpha\nu + (1-\alpha)\tilde{\nu})$ so that

$$V_c(\alpha\mu + (1-\alpha)\tilde{\mu}, \alpha\nu + (1-\alpha)\tilde{\nu}) \leq (\alpha\pi_\star + (1-\alpha)\tilde{\pi}_\star)(c) = \alpha V_c(\mu,\nu) + (1-\alpha)V_c(\tilde{\mu}, \tilde{\nu}).$$

Therefore V_c is convex. Let us now suppose that $\mu_n \to \mu$ and $\nu_n \to \nu$ weakly as $n \to \infty$. Let $\pi_n \in \Pi(\mu_n, \nu_n)$ be optimal for $V_c(\mu_n, \nu_n)$. By Lemma 1.7, from any subsequence attaining $\liminf_{n\to\infty} V_c(\mu_n, \nu_n)$, we may extract a further subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ converging weakly to $\pi_\infty \in \Pi(\mu, \nu)$. Using the Portmanteau theorem (see Theorem 5.2 7)) for the second inequality, we have

$$V_c(\mu,\nu) \leq \pi_\infty(c) \leq \lim_{k\to\infty} \pi_{n_k}(c) = \liminf_{n\to\infty} V_c(\mu_n, \nu_n)$$

and V_c is lower semi-continuous. ∎

Proof of Lemma 1.7: The first assertion is a consequence of the second for the choice $(\mu_n, \nu_n) = (\mu, \nu)$ for each n. To prove the second assertion, we set $\varepsilon \in (0,1)$. Since $\mathcal{X}$ (resp. $\mathcal{Y}$) is Polish, by the Prokhorov theorem (see Theorem 5.3), there exists a compact subset $K_\mathcal{X} \in \mathcal{X}$ (resp. $K_\mathcal{Y} \in \mathcal{Y}$) such that $\sup_n \mu_n(K_\mathcal{X}^c) \leq \frac{\varepsilon}{2}$ (resp. $\sup_n \nu_n(K_\mathcal{Y}^c) \leq \frac{\varepsilon}{2}$). Since $\pi_n \in \Pi(\mu_n, \nu_n)$,

$$\begin{aligned}\pi_n(\{K_\mathcal{X} \times K_\mathcal{Y}\}^c) &= \pi_n(\{K_\mathcal{X}^c \times \mathcal{Y}\} \cup \{\mathcal{X} \times K_\mathcal{Y}^c\}) \\ &\leq \pi_n(K_\mathcal{X}^c \times \mathcal{Y}) + \pi_n(\mathcal{X} \times K_\mathcal{Y}^c) = \mu_n(K_\mathcal{X}^c) + \nu_n(K_\mathcal{Y}^c) \leq \varepsilon.\end{aligned}$$

By Tykhonov's theorem, $K_\mathcal{X} \times K_\mathcal{Y}$ is a compact subset of $\mathcal{X} \times \mathcal{Y}$. Using the other direction in the Prokhorov theorem (see Theorem 5.3), we deduce that we can extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ converging weakly to π_∞. Let μ_∞ and ν_∞ denote the marginals of π_∞. By continuity of the projections $\mathcal{X}\times\mathcal{Y} \ni (x,y) \mapsto x \in \mathcal{X}$ and $\mathcal{X}\times\mathcal{Y} \ni (x,y) \mapsto y \in \mathcal{Y}$, $\mu_\infty = \lim_{k\to\infty} \mu_{n_k} = \mu$ and $\nu_\infty = \lim_{k\to\infty} \nu_{n_k} = \nu$ so that $\pi_\infty \in \Pi(\mu, \nu)$. ∎

1.4 Dual formulation

Theorem 1.8. *Let $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}\cup\{+\infty\}$ be lower semi-continuous bounded from below. Then*

$$V_c(\mu,\nu) = \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y}) \\ \phi\oplus\psi\leq c}} \{\mu(\phi) + \nu(\psi)\}, \text{ where } \phi\oplus\psi(x,y) = \phi(x) + \psi(y).$$

Remark 1.9. *For $\pi \in \Pi(\mu,\nu)$ and $(\phi,\psi) \in C_b(\mathcal{X}) \times C_b(\mathcal{Y})$ such that $\phi\oplus\psi \leq c$, we have*

$$\pi(c) \geq \pi(\phi\oplus\psi) = \mu(\phi) + \nu(\psi).$$

Therefore the weak duality inequality $V_c(\mu,\nu) \geq \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y}) \\ \phi\oplus\psi\leq c}} \{\mu(\phi) + \nu(\psi)\}$ holds. The converse inequality is the main statement in the theorem.

The proof relies on the Fenchel-Moreau theorem (see Theorem 2.3.3 [16]).

Theorem 1.10 (Fenchel-Moreau). *Let V be a Hausdorff topological vector space (2 distinct points in V have distinct neighbourhoods) with topological dual V' (space of continuous linear forms on V), $f : V \to \mathbb{R} \cup \{+\infty\}$ and $f^\star(v') = \sup_{v\in V}\{v'(v) - f(v)\}$, $v' \in V'$ denote its convex conjugate. Then f is lower semi-continuous and convex iff*

$$f(v) = \sup_{v'\in V'}\{v'(v) - f^\star(v')\} = f^{\star\star}(v),\ v \in V.$$

Remark 1.11. *The convex conjugate $f^\star$ is convex lower semi-continuous on V' as the supremum of linear (hence convex) continuous (hence lower semi-continuous) functions.*

Exercise 1.12. *Prove that the supremum of convex (resp. lower semi-continuous) $\mathbb{R} \cup \{+\infty\}$-valued functions is convex (resp. lower semi-sontinuous). For the lower semi-continuity, you may first check that f is lower semi-continuous iff its level sets $\{f \leq \alpha\}$, $\alpha \in \mathbb{R}$ are closed.*

The next lemma ensures that a lower semi-continuous function bounded from below is the non-decreasing limit of Lipschitz functions obtained by inf-convolution.

Lemma 1.13. *A function $c : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ non constantly equal to $+\infty$, lower semi-continuous and bounded from below on a metric space $\mathcal{Z}$ with distance $d_\mathcal{Z}$ is the non-decreasing limit as $n \to \infty$ of the n-Lipschitz functions $c_n(z) = \inf_{\tilde z\in\mathcal{Z}}\{c(\tilde z) + nd_\mathcal{Z}(\tilde z, z)\}$ which are bounded from below by the same constant as c.*

Proof: Let $z_0 \in \mathcal{Z}$ satisfy $c(z_0) < +\infty$. Clearly, c_n is bounded from below by the same constant as c, non-decreasing with n and bounded from above by c (choice $\tilde z = z$) and by $c(z_0) + nd_\mathcal{Z}(\cdot, z_0) < +\infty$ (choice $\tilde z = z_0$). Let $z, \hat z \in \mathcal{Z}$ be such that $c_n(z) \leq c_n(\hat z)$ and for $k \in \mathbb{N}^*$, $\tilde z_{n,k}$ such that $c_n(z) \geq c(\tilde z_{n,k}) + nd_\mathcal{Z}(\tilde z_{n,k}, z) - \frac{1}{k}$. Then, by the definition of $c_n(\hat z)$ and the triangle inequality,

$$c_n(\hat z) - c_n(z) \leq c(\tilde z_{n,k}) + nd_\mathcal{Z}(\tilde z_{n,k}, \hat z) - \left(c(\tilde z_{n,k}) + nd_\mathcal{Z}(\tilde z_{n,k}, z) - \frac{1}{k}\right) \leq nd_\mathcal{Z}(\hat z, z) + \frac{1}{k}.$$

By letting $k \to \infty$, we deduce that c_n is n-Lipschitz. We have

$$c(z) \geq c_n(z) \geq c(\tilde z_{n,n}) + nd_\mathcal{Z}(\tilde z_{n,n}, z) - \frac{1}{n} \geq \inf c + nd_\mathcal{Z}(\tilde z_{n,n}, z) - \frac{1}{n}.$$

If $\tilde z_{n,n} \to z$ as $n \to \infty$, then, by the first two inequalities,

$$c(z) \geq \lim_{n\to\infty} c_n(z) \geq \liminf_{n\to\infty} c(\tilde z_{n,n}) \geq c(z)$$

where the last inequality follows from the lower semi-continuity of c. Otherwise,

$$c(z) \geq \limsup_{n\to\infty} c_n(z) \geq \limsup_{n\to\infty} (\inf c + nd_\mathcal{Z}(\tilde z_{n,n}, z)) = +\infty \geq c(z).$$

∎

Lemma 1.14. *Let $f \in \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be bounded from below and Lipschitz continuous in its first variable. For each $\varepsilon > 0$, there exists a measurable map $S_\varepsilon : \mathcal{X} \to \mathcal{Y}$ such that*

$$\forall x \in \mathcal{X},\ f(x, S_\varepsilon(x)) \leq \inf_{y\in\mathcal{Y}} f(x, y) + \varepsilon.$$

Proof: Let $(x_n)_{n\in\mathbb{N}}$ be dense in $\mathcal{X}$. For each $n \in \mathbb{N}$, there exists y_n such that $f(x_n, y_n) \leq \inf_{y\in\mathcal{Y}} f(x_n, y) + \frac{\varepsilon}{3}$. Let L denote the Lipschitz constant of f in its first variable. If $L = 0$, $S_\varepsilon(x) = x_0$ does the job. Let us suppose that $L > 0$. By the argument given in the proof of Lemma 1.13, we can check that the function $g(x) = \inf_{y\in\mathcal{Y}} f(x, y)$ is Lipschitz with constant L. For $n \in \mathbb{N}$, we set $A_n = B(x_n, \frac{\varepsilon}{3L}) \cap \left\{\bigcup_{k=0}^{n-1} B(x_k, \frac{\varepsilon}{3L})\right\}^c \in \mathcal{B}(\mathcal{X})$. By density of $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$, this set is the disjoint union of the $(A_n)_{n\in\mathbb{N}}$. Let us define $S_\varepsilon(x) = \sum_{n\in\mathbb{N}} 1_{A_n}(x) y_n$. For $n \in \mathbb{N}$ and $x \in A_n$, we have

$$\begin{aligned} f(x, S_\varepsilon(x)) - g(x) &= f(x, y_n) - f(x_n, y_n) + f(x_n, y_n) - g(x_n) + g(x_n) - g(x) \\ &\leq L \times \frac{\varepsilon}{3L} + \frac{\varepsilon}{3} + L \times \frac{\varepsilon}{3L} = \varepsilon. \end{aligned}$$

∎

Proof of Theorem 1.8: When c is constantly equal to $+\infty$, then the conclusion holds since $V_c(\mu, \nu) = +\infty$ and the supremum of $\{\mu(\phi) + \nu(\psi)\}$ over constant functions ϕ and ψ such that $\phi \otimes \psi \leq c$ also is $+\infty$. We now suppose that c is not constantly equal to $+\infty$. We apply the Fenchel-Moreau theorem to V equal to the space of bounded signed measures on $\mathcal{Y}$ endowed with the bounded Lipschitz norm :

$$\|q\|_{\mathrm{BL}} = \sup_{\substack{f:\mathcal{Y}\to\mathbb{R} \\ 1-\text{ Lipschitz bounded by } 1}} |q(f)|.$$

By Theorem 5.2 3), this norm metricizes the weak convergence topology on $\mathcal{P}(\mathcal{Y})$. We have $V' = C_b(\mathcal{Y})$ with $g(q) = q(g) = \int_\mathcal{Y} g(y)q(dy)$ for $(g, q) \in C_b(\mathcal{Y}) \times V$. We fix $\mu \in \mathcal{P}(\mathcal{X})$ and set

$$F(q) = \begin{cases} V_c(\mu, q) \text{ if } q \in \mathcal{P}(\mathcal{Y}) \\ +\infty \text{ if } q \in V \setminus \mathcal{P}(\mathcal{Y}) \end{cases}.$$

Since $\mathcal{P}(\mathcal{Y})$ is a closed convex subset of V and $\mathcal{P}(\mathcal{Y}) \ni \nu \mapsto V_c(\mu, \nu)$ is lower semi-continuous and convex by Theorem 1.6, F is lower semi-continuous and convex. Hence by the Fenchel-Moreau theorem,

$$\forall \nu \in \mathcal{P}(\mathcal{Y}),\ V_c(\mu, \nu) = F(\nu) = \sup_{\psi\in C_b(\mathcal{Y})} \{\nu(\psi) - F^\star(\psi)\}. \tag{1.1}$$

Case c Lipschitz. For $\psi \in C_b(\mathcal{Y})$, we define $\psi^{\bar{c}}(x) = \inf_{y\in\mathcal{Y}}\{c(x, y) - \psi(y)\},\ x \in \mathcal{X}$. The function $\psi^{\bar{c}}$ is bounded from below. By Lemma 1.14, there exists a measurable map $S_\varepsilon : \mathcal{X} \to \mathcal{Y}$ such that $\forall x \in \mathcal{X},\ c(x, S_\varepsilon(x)) - \psi(S_\varepsilon(x)) \leq \psi^{\bar{c}}(x) + \varepsilon$. Setting $\pi_\varepsilon(dx, dy) = \mu(dx)\delta_{S_\varepsilon(x)}(dy)$, we have

$$\begin{aligned} \mu(\psi^{\bar{c}}) = \inf_{\substack{\nu\in\mathcal{P}(\mathcal{Y}) \\ \pi\in\Pi(\mu,\nu)}} \pi(\psi^{\bar{c}} \oplus 0) &\leq \inf_{\substack{\nu\in\mathcal{P}(\mathcal{Y}) \\ \pi\in\Pi(\mu,\nu)}} \pi(c - 0 \oplus \psi) \\ &\leq \pi_\varepsilon(c - 0 \oplus \psi) \leq \pi_\varepsilon(\psi^{\bar{c}} \oplus 0 + \varepsilon) = \mu(\psi^{\bar{c}}) + \varepsilon. \end{aligned}$$

Letting $\varepsilon \to 0$, we deduce that $\inf_{\substack{\nu\in\mathcal{P}(\mathcal{Y}) \\ \pi\in\Pi(\mu,\nu)}} \pi(c - 0 \oplus \psi) = \mu(\psi^{\bar{c}})$. Using the definitions of F and V_c for the second equality, we deduce that

$$-F^\star(\psi) = -\sup_{q\in V}\{q(\psi) - F(q)\} = -\sup_{\nu\in\mathcal{P}(\mathcal{Y})}\left\{\nu(\psi) - \inf_{\pi\in\Pi(\mu,\nu)}\pi(c)\right\}$$
$$= -\sup_{\substack{\nu\in\mathcal{P}(\mathcal{Y})\\ \pi\in\Pi(\mu,\nu)}}\pi(0\oplus\psi - c) = \inf_{\substack{\nu\in\mathcal{P}(\mathcal{Y})\\ \pi\in\Pi(\mu,\nu)}}\pi(c - 0\oplus\psi) = \mu(\psi^{\bar{c}}).$$

Plugging this equality in (1.1) and using the weak duality result in Remark 1.9, we conclude that

$$\forall\nu\in\mathcal{P}(\mathcal{Y}),\ V_c(\mu,\nu) = \sup_{\psi\in C_b(\mathcal{Y})}\{\nu(\psi)+\mu(\psi^{\bar{c}})\} \le \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})\\ \phi\oplus\psi\le c}}\{\mu(\phi)+\nu(\psi)\} \le V_c(\mu,\nu).$$

General case. We use the sequence $(c_n)_{n\geq 1}$ of n-Lipschitz functions growing to c given by Lemma 1.13. Using the weak duality inequality in Remark 1.9 for the first inequality and the previous case for the equality, we get that

$$V_c(\mu,\nu) \ge \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})\\ \phi\oplus\psi\le c}}\{\mu(\phi)+\nu(\psi)\}$$
$$\ge \sup_n \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})\\ \phi\oplus\psi\le c_n}}\{\mu(\phi)+\nu(\psi)\} = \sup_n V_{c_n}(\mu,\nu). \tag{1.2}$$

Let for $n\geq 1$, $\pi_n\in\Pi(\mu,\nu)$ be optimal for $V_{c_n}(\mu,\nu)$. By Lemma 1.7, we may extract a sequence $(\pi_{n_k})_k$ which converges weakly to $\pi_\infty\in\Pi(\mu,\nu)$. We have

$$\sup_n V_{c_n}(\mu,\nu) = \lim_{n\to\infty}\nearrow\pi_n(c_n) = \lim_{k\to\infty}\pi_{n_k}(c_{n_k}).$$

For fixed $m\in\mathbb{N}^*$, when k is large enough, we have $n_k\geq m$ and $c_{n_k}\geq c_m$. By the Portmanteau theorem (see Theorem 5.2), $\liminf_{k\to\infty}\pi_{n_k}(c_m)\geq\pi_\infty(c_m)$. By the monotone convergence theorem, $\sup_{m\geq 1}\pi_\infty(c_m)=\pi_\infty(c)$. Therefore

$$\sup_n V_{c_n}(\mu,\nu) = \lim_{k\to\infty}\pi_{n_k}(c_{n_k}) \ge \sup_{m\ge 1}\liminf_{k\to\infty}\pi_{n_k}(c_m) \ge \sup_{m\ge1}\pi_\infty(c_m) = \pi_\infty(c)\ge V_c(\mu,\nu).$$

With (1.2), we conclude that $V_c(\mu,\nu) = \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})\\ \phi\oplus\psi\le c}}\{\mu(\phi)+\nu(\psi)\}$.

∎

1.5 c-cyclical monotonicity and the super-differential

For ϕ,ψ such that $\phi\oplus\psi\le c$, we can always increase the dual value $\{\mu(\phi)+\nu(\psi)\}$ by replacing ϕ (resp. ψ) by the $\bar{c}$-conjugate $\psi^{\bar{c}}$ of ψ (resp. the c-conjugate ϕ^c of ϕ) defined by

$$\psi^{\bar{c}}(x) = \inf_{y\in\mathcal{Y}}\{c(x,y)-\phi(y)\} \text{ and } \phi^c(y) = \inf_{x\in\mathcal{X}}\{c(x,y)-\phi(x)\}.$$

For these definitions to make sense, we suppose that c is real-valued (it cannot take the value $+\infty$ like in Theorems 1.6 and 1.8) and that $\psi:\mathcal{Y}\to\{-\infty\}\cup\mathbb{R}$ (resp. $\phi:\mathcal{X}\to$

$\{-\infty\} \cup \mathbb{R}$). When $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$ (resp. $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$), then $\psi^{\bar{c}}$ (resp. ϕ^c) is $\{-\infty\} \cup \mathbb{R}$-valued iff ψ (resp. ϕ) is not constantly equal to $-\infty$.
When $\mathcal{Y} = \mathcal{X}$ and the cost function c is symmetric, the distinction between these $\bar{c}$ and the c transforms is no longer needed.

Definition 1.15.
- *We say that $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$ is c-concave if $\psi = \phi^c$ for some $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$.*
- *We say that $\phi : \mathcal{X} \to \{-\infty\}\cup\mathbb{R}$ is $\bar{c}$-concave if $\phi = \psi^{\bar{c}}$ for some $\psi : \mathcal{Y} \to \{-\infty\}\cup\mathbb{R}$.*

Note that when $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$ is c-concave (resp. $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ is $\bar{c}$-concave), then any function $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ such that $\psi = \phi^c$ ($\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$ such that $\phi = \psi^{\bar{c}}$) is non constantly equal to $-\infty$. According to proof of the next lemma, it is not useful to iterate the transform

Lemma 1.16. *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. The function $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ (resp. $\psi : \mathcal{Y} \to \{-\infty\}\cup\mathbb{R}$) non constantly equal to $-\infty$ is $\bar{c}$-concave (resp. c-concave) iff $\phi = (\phi^c)^{\bar{c}}$ (resp. $\psi = (\psi^{\bar{c}})^c$).*

Proof: We only deal with the statement concerning ϕ since the one concerning ψ is proved in a symmetric way. Since ϕ is not constant equal to $-\infty$, ϕ^c is $\{-\infty\} \cup \mathbb{R}$-valued and $\phi = (\phi^c)^{\bar{c}}$ implies that ϕ is $\bar{c}$-concave. To prove the converse implication, it is enough to check that for $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$, $((\psi^{\bar{c}})^c)^{\bar{c}} = \psi^{\bar{c}}$. Indeed, when ϕ is $\bar{c}$-concave, then there exists $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$ such that $\phi = \psi^{\bar{c}} = ((\psi^{\bar{c}})^c)^{\bar{c}} = (\phi^c)^{\bar{c}}$. We have

$$\begin{aligned} ((\psi^{\bar{c}})^c)^{\bar{c}}(x) &= \inf_{y\in\mathcal{Y}} \left\{ c(x,y) - \inf_{\tilde{x}\in\mathcal{X}} \left\{ c(\tilde{x},y) - \inf_{\tilde{y}\in\mathcal{Y}} \{c(\tilde{x},\tilde{y}) - \psi(\tilde{y})\} \right\} \right\} \\ &= \inf_{y\in\mathcal{Y}} \sup_{\tilde{x}\in\mathcal{X}} \inf_{\tilde{y}\in\mathcal{Y}} \{c(x,y) - c(\tilde{x},y) + c(\tilde{x},\tilde{y}) - \psi(\tilde{y})\}. \end{aligned}$$

Restricting $\tilde{x}$ to be equal to x, we deduce that $((\psi^{\bar{c}})^c)^{\bar{c}}(x) \le \inf_{\tilde{y}\in\mathcal{Y}} \{c(x,\tilde{y}) - \psi(\tilde{y})\} = \psi^{\bar{c}}(x)$. Restricting y to be equal to $\tilde{y}$, we also deduce that $((\psi^{\bar{c}})^c)^{\bar{c}}(x) \ge \inf_{\tilde{y}\in\mathcal{Y}} \{c(x,\tilde{y}) - \psi(\tilde{y})\} = \psi^{\bar{c}}(x)$. Hence $((\psi^{\bar{c}})^c)^{\bar{c}} = \psi^{\bar{c}}$. ∎

Example 1.17. *Let $\mathcal{Y} = \mathcal{X}$ and $c(x,y) = d_{\mathcal{X}}(x,y)$. Let $\phi : \mathcal{X} \to \{-\infty\}\cup\mathbb{R}$ non constantly equal to $-\infty$. Let us check that ϕ is $d_{\mathcal{X}}$-concave iff it is $\mathbb{R}$-valued and 1-Lipschitz. If ϕ is $d_{\mathcal{X}}$-concave, then, by Lemma 1.16, $\phi(x) = \inf_{y\in\mathcal{Y}}\{d_{\mathcal{X}}(x,y) - \phi^{d_{\mathcal{X}}}(y)\}$. If $\phi(x) \le \phi(\hat{x})$ with $\phi(\hat{x}) > -\infty$, then choosing $(y_n)_{n\in\mathbb{N}}$ such that $\phi(x) = \lim_{n\to\infty}\{d_{\mathcal{X}}(x,y_n) - \phi^{d_{\mathcal{X}}}(y_n)\}$ and using that $\phi(\hat{x}) \le d_{\mathcal{X}}(\hat{x},y_n) - \phi^{d_{\mathcal{X}}}(y_n)$, we obtain that*

$$\phi(\hat{x}) - \phi(x) \le \liminf_{n\to\infty} \left\{ d_{\mathcal{X}}(\hat{x},y_n) - \phi^{d_{\mathcal{X}}}(y_n) - d_{\mathcal{X}}(x,y_n) + \phi^{d_{\mathcal{X}}}(y_n) \right\} \le d_{\mathcal{X}}(\hat{x},x),$$

so that ϕ is $\mathbb{R}$-valued and 1-Lipschitz. Conversely, $\phi^{d_{\mathcal{X}}}(x) = \inf_{y\in\mathcal{X}}\{d_{\mathcal{X}}(x,y) - \phi(y)\} \le d_{\mathcal{X}}(x,x) - \phi(x) = -\phi(x)$, and if $\phi : \mathcal{X} \to \mathbb{R}$ is 1-Lipschitz, then

$$\phi^{d_{\mathcal{X}}}(x) \ge \inf_{y\in\mathcal{X}}\{d_{\mathcal{X}}(x,y) - (\phi(x) + d_{\mathcal{X}}(x,y))\} = -\phi(x) \text{ so that } \phi = -\phi^{d_{\mathcal{X}}}.$$

Since $-\phi$ also is 1-Lipschitz, we deduce that $-\phi = -(-\phi)^{d_{\mathcal{X}}}$ so that $\phi = (-\phi)^{d_{\mathcal{X}}}$ and ϕ is $d_{\mathcal{X}}$-concave.

Example 1.18. *Let $\mathcal{Y} = \mathcal{X} = \mathbb{R}^d$ and $c(x, y) = -x.y$. Then for $\phi : \mathbb{R}^d \to \{-\infty\} \cup \mathbb{R}$,*

$$\phi^c(y) = \inf_{x\in\mathbb{R}^d}\{-x.y - \phi(x)\} = -\sup_{x\in\mathbb{R}^d}\{x.y - (-\phi(x))\} = -(-\phi)^\star(y).$$

Let us moreover suppose that ϕ is not constantly equal to $-\infty$. With Lemma 1.16, we deduce that

$$\phi \text{ c-concave} \Leftrightarrow \phi = (\phi^c)^c = (-(-\phi)^\star)^c = -(-\phi)^{\star\star}.$$

Since $-(-\phi)^{\star\star}$ is upper semi-continuous and concave, ϕ c-concave $\Rightarrow$ ϕ upper semi-continuous and concave. Conversely, if ϕ is upper semi-continuous and concave, then $-\phi$ is lower semi-continuous and convex and, by Fenchel-Moreau duality (see Theorem 1.10), $-\phi = (-\phi)^{\star\star}$ so that $\phi = (\phi^c)^c$ and ϕ is c-concave.

Example 1.19. *Let $\mathcal{Y} = \mathcal{X} = \mathbb{R}^d$ and $c(x, y) = |x - y|^2$. Let $\phi : \mathbb{R}^d \to \{-\infty\} \cup \mathbb{R}$. We set $\bar\phi(x) = \frac{|x|^2}{2} - \phi(x)$ with values in $\mathbb{R} \cup \{+\infty\}$. We have*

$$\begin{aligned}\frac{|y|^2}{2} - \phi^c(y) &= \frac{|y|^2}{2} - \inf_{x\in\mathbb{R}^d}\left\{\frac{1}{2}|x-y|^2 - \phi(x)\right\} = -\inf_{x\in\mathbb{R}^d}\left\{-x.y + \bar\phi(x)\right\}\\ &= \sup_{x\in\mathbb{R}^d}\{x.y - \bar\phi(x)\} = (\bar\phi)^\star(y).\end{aligned}$$

Therefore $\phi^c(y) = \frac{|y|^2}{2} - (\bar\phi)^\star(y)$ and

$$\begin{aligned}(\phi^c)^c(x) &= \inf_{y\in\mathbb{R}^d}\left\{\frac{1}{2}|x-y|^2 - \left(\frac{|y|^2}{2} - (\bar\phi)^\star(y)\right)\right\} = \frac{|x|^2}{2} - \sup_{y\in\mathbb{R}^d}\left\{x.y - (\bar\phi)^\star(y)\right\}\\ &= \frac{|x|^2}{2} - (\bar\phi)^{\star\star}(x).\end{aligned}$$

With Lemma 1.16 and Fenchel-Moreau duality (see Theorem 1.10), we conclude that when ϕ is not constantly equal to $-\infty$,

$$\phi \text{ c-concave} \Leftrightarrow \phi = (\phi^c)^c \Leftrightarrow \bar\phi = (\bar\phi)^{\star\star} \Leftrightarrow \bar\phi \text{ lower semi-continuous convex.}$$

Let us now introduce the superdifferentials and the cyclic monotonicity.

Definition 1.20. *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$.*

- *Let $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ (resp. $\psi : \mathcal{Y} \to \{-\infty\} \cup \mathbb{R}$) be $\bar c$-concave (resp. c-concave). Its $\bar c$-superdifferential (resp. c-superdifferential) is defined as*

$$\begin{aligned}\partial^{\bar c}\phi &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} :\ \phi(x) + \phi^c(y) = c(x, y)\}\\ (\textit{resp. } \partial^c\psi &= \{(x, y) \in \mathcal{X} \times \mathcal{Y} :\ \psi^{\bar c}(x) + \psi(y) = c(x, y)\}).\end{aligned}$$

  We also set $\partial^{\bar c}\phi(x) = \{y \in \mathcal{Y} :\ \phi(x) + \phi^c(y) = c(x, y)\}$ for $x \in \mathcal{X}$ and $\partial^c\psi(y) = \{x \in \mathcal{X} :\ \psi^{\bar c}(x) + \psi(y) = c(x, y)\}$ for $y \in \mathcal{Y}$.

- *A subset Γ of $\mathcal{X} \times \mathcal{Y}$ is called c-cyclically monotone if for all $N \in \mathbb{N}^*$, $(x_1, y_1), \cdots, (x_N, y_N) \in \Gamma$, we have $\sum_{i=1}^N c(x_i, y_i) \le \sum_{i=1}^N c(x_i, y_{i+1})$ with convention $y_{N+1} = y_1$.*

- *A coupling $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is called c-cyclically monotone if there exists some c-cyclically monotone set $\Gamma \in \mathcal{B}(\mathcal{X} \times \mathcal{Y})$ such that $\pi(\Gamma) = 1$.*

Example 1.21. *Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $c(x,y) = -x.y$ and $\phi : \mathbb{R}^d \to \{-\infty\} \cup \mathbb{R}$ non constantly equal to $-\infty$ be c-concave. By Example 1.18, ϕ is upper semi-continuous and concave. By definition of $\phi^c(y)$,*

$$y \in \partial^{\bar{c}}\phi(x) \Leftrightarrow \phi^c(y) = -x.y - \phi(x) \Leftrightarrow \inf_{\tilde{x}\in\mathbb{R}^d}\{-\tilde{x}.y - \phi(\tilde{x})\} = -x.y - \phi(x)$$
$$\Leftrightarrow \sup_{\tilde{x}\in\mathbb{R}^d}\{\phi(\tilde{x}) + y.\tilde{x}\} = \phi(x) + y.x \Leftrightarrow \forall \tilde{x} \in \mathbb{R}^d,\ \phi(\tilde{x}) \le \phi(x) - y.(\tilde{x} - x)$$

and $\partial^{\bar{c}}\phi(x)$ is minus the super-differential of the concave function ϕ at x.

Proposition 1.22. *Let $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. The set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is c-cyclically monotone iff $\Gamma \subset \partial^{\bar{c}}\phi$ for some $\bar{c}$-concave function $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$.*

Proof: To prove the sufficient condition, it is enough to check that if $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ is $\bar{c}$-concave, then $\partial^{\bar{c}}\phi$ is c-cyclically monotone. This follows from the fact that for $N \in \mathbb{N}^*$ and $(x_1, y_1), \cdots, (x_N, y_N) \in \partial^{\bar{c}}\phi$,

$$\sum_{i=1}^N c(x_i, y_i) = \sum_{i=1}^N \{\phi(x_i) + \phi^c(y_i)\} = \sum_{i=1}^N \{\phi(x_i) + \phi^c(y_{i+1})\} \le \sum_{i=1}^N c(x_i, y_{i+1}) \text{ as } \phi \oplus \phi^c \le c.$$

To show the necessary condition, we suppose that $\Gamma \neq \emptyset$ is c-cyclically monotone. We fix $(x_0, y_0) \in \Gamma$ and set

$$\phi(x) = \inf\left\{c(x, y_N) + \sum_{i=0}^{N-1} c(x_{i+1}, y_i) - \sum_{i=0}^N c(x_i, y_i) :\ N \in \mathbb{N},\ (x_1, y_1), \cdots, (x_N, y_N) \in \Gamma\right\}. \tag{1.3}$$

Clearly, ϕ is $\{-\infty\} \cup \mathbb{R}$-valued. Moreover, we have $x_{N+1} = x_0$ in the constraints of the minimization problem giving $\phi(x_0)$, so that, by c-cyclical monotony of Γ, the infimum is taken on non-negative $\sum_{i=0}^N c(x_{i+1}, y_i) - \sum_{i=0}^N c(x_i, y_i)$. Hence $\phi(x_0) \ge 0$. We also set

$$-\psi(y) = \inf\left\{\sum_{i=0}^{N-1} c(x_{i+1}, y_i) - \sum_{i=0}^N c(x_i, y_i) :\ N \in \mathbb{N},\ (x_1, y_1), \cdots, (x_N, y_N) \in \Gamma \text{ and } y_N = y\right\}$$

under the convention $\inf \emptyset = +\infty$ (the existence of $x_N \in \mathcal{X}$ such that $(x_N, y) \in \Gamma$ is not guaranteed). Using that $c(x, y) = c(x, y_N)$ under the constraint $y_N = y$ for the minimization problem for $-\psi(y)$, we get

$$\inf_{y\in\mathcal{Y}}\{c(x, y) - \psi(y)\} = \phi(x)$$

Hence $\phi = \psi^{\bar{c}}$ and, since $\phi(x_0) \ge 0$, ψ is $\{-\infty\} \cup \mathbb{R}$-valued so that ϕ is $\bar{c}$-concave.

To show that $\Gamma \subset \partial^{\bar{c}}\phi$, it is enough to check that $\phi(x) + \psi(y) \ge c(x, y)$ for $(x, y) \in \Gamma$ since $\phi^c = (\psi^{\bar{c}})^c \ge \psi$ and $\phi \oplus \phi^c \le c$. Let $(x, y) \in \Gamma$. There exists a sequence $(\tilde{y}_n)_{n\in\mathbb{N}} \subset \mathcal{Y}$ such that $\phi(x) = \lim_{n\to\infty}\{c(x, \tilde{y}_n) - \psi(\tilde{y}_n)\}$. Choosing $N \in \mathbb{N}^*$ and $(x_N, y_{N-1}) = (x, \tilde{y}_n)$ in the minimization problem giving $-\psi(y)$, we get

$$-\psi(y) \le \inf\Bigg\{c(x, \tilde{y}_n) - c(x, y) + \sum_{i=0}^{N-2} c(x_{i+1}, y_i) - \sum_{i=0}^{N-1} c(x_i, y_i) :$$
$$N \in \mathbb{N}^*,\ (x_1, y_1), \cdots, (x_{N-1}, y_{N-1}) \in \Gamma \text{ and } y_{N-1} = \tilde{y}_n\Bigg\}$$
$$= c(x, \tilde{y}_n) - c(x, y) - \psi(\tilde{y}_n) = -c(x, y) + \{c(x, \tilde{y}_n) - \psi(\tilde{y}_n)\}.$$

Taking the limit $n \to \infty$ in the right-hand side, we conclude that $-\psi(y) \leq -c(x,y)+\phi(x)$ i.e. $\phi(x)+\psi(y) \geq c(x,y)$. ∎

Remark 1.23.

- *In* (1.3) *with* $x = x_0$, *we get for the choice* $N = 1$ *and* $(x_1, y_1) = (x_0, y_0)$, *using* $x_2 = x_0$,

$$\begin{aligned}\phi(x_0) &\leq c(x_1,y_0)+c(x_2,y_1)-c(x_0,y_0)-c(x_1,y_1)\\ &= c(x_0,y_0)+c(x_0,y_0)-c(x_0,y_0)-c(x_0,y_0)=0.\end{aligned}$$

  Hence $\phi(x_0) = 0$.

- *The functions* ϕ *and* ϕ^c *constructed in the proof of Proposition 1.22 may not be measurable. But when* c *is measurable and* $\Gamma \in \mathcal{B}(\mathcal{X}\times\mathcal{Y})$, *then* ϕ *and (resp.* ϕ^c*) is universally measurable which means that for any* $\mu \in \mathcal{P}(\mathcal{X})$ *(resp.* $\nu \in \mathcal{P}(\mathcal{Y})$*), there exists a Borel function* $\tilde{\phi} : \mathcal{X} \to \{-\infty\}\cup\mathbb{R}$ *(resp.* $\tilde{\psi} : \mathcal{Y} \to \{-\infty\}\cup\mathbb{R}$*) such that* $\phi(x) = \tilde{\phi}(x)$, $\mu(dx)$ *a.e. (resp.* $\phi^c(y) = \psi(y)$, $\nu(dy)$ *a.e.).*

Definition 1.24. *The support* $\operatorname{supp}(\eta)$ *of a probability measure* η *on* $(\mathcal{Z},\mathcal{B}(\mathcal{Z}))$ *is the smallest closed subset* A *of* $\mathcal{Z}$ *such that* $\eta(A) = 1$.

Remark 1.25. *This definition makes sense since an arbitrary intersection of closed subsets remains closed. The complementary* $\operatorname{supp}(\eta)^c$ *of* $\operatorname{supp}(\eta)$ *is an open subset such that* $\eta(\operatorname{supp}(\eta)^c) = 0$. *Therefore* $\operatorname{supp}(\eta)^c \subset \{z \in \mathcal{Z} : \exists\, O \text{ open containing } z \text{ such that } \eta(O) = 0\}$. *The converse inclusion holds since when there is an open set* O *containing* z *with* $\eta(O) = 0$ *then* $\operatorname{supp}(\eta)$ *is included in the closed set* O^c *and therefore* $z \in O \subset \operatorname{supp}(\eta)^c$. *Hence*

$$\operatorname{supp}(\eta) = \{z \in \mathcal{Z} : \forall\, O \text{ open containing } z,\ \eta(O) > 0\}.$$

Theorem 1.26 (Fundamental Theorem of OT)**.** *Let* $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$, $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ *be continuous and bounded from below and such that* $c \leq a \oplus b$ *for some* $(a,b) \in L^1(\mu)\times L^1(\nu)$. *For* $\pi_\star \in \Pi(\mu,\nu)$ *the following assertions are equivalent :*

(i) $\pi_\star$ *is optimal for* $V_c(\mu,\nu)$,

(ii) *the support of* $\pi_\star$ *is c-cyclically monotone,*

(iii) *there exists a* $\bar{c}$*-concave function* $\phi : \mathcal{X} \to \{-\infty\}\cup\mathbb{R}$ *such that* $\phi^+ \in L^1(\mu)$ *and the support of* $\pi_\star$ *is included in* $\partial^{\bar{c}}\phi$.

Remark 1.27. *Under the assumptions, if* $\pi_\star$ *is optimal for* $V_c(\pi_\star(dx\times\mathcal{Y}), \pi_\star(\mathcal{X}\times dy))$ *and* $\pi \in \Pi(\mu,\nu)$ *such that* $\operatorname{supp}(\pi) \subset \operatorname{supp}(\pi_\star)$, *then* π *is optimal for* $V_c(\mu,\nu)$ *as soon as there exists* $(a,b) \in L^1(\mu)\times L^1(\nu)$ *such that* $c \leq a \oplus b$.

The proof relies on the next lemma.

Lemma 1.28. *Let* $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ *and* $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ *be continuous. If* $\pi_\star \in \Pi(\mu,\nu)$ *is such* $\pi_\star(c) = V_c(\mu,\nu) < +\infty$, *then the support of* $\pi_\star$ *is c-cyclically monotone.*

Proof of Theorem 1.26: By Lemma 1.28, we have $(i) \Rightarrow (ii)$.
For $(ii) \Rightarrow (iii)$, we remark that Proposition 1.22 ensures the existence of a $\bar{c}$-concave function ϕ such that $\text{supp}\,(\pi_\star) \subset \partial^{\bar{c}}\phi$. Moreover, the definition (1.3) of ϕ in its proof together with the inequality $c \leq a \oplus b$ ensure that

$$\forall x \in \mathcal{X},\ \phi(x) \leq c(x, y_0) - c(x_0, y_0) \leq a(x) + b(y_0) - c(x_0, y_0)$$

so that $\phi^+ \in L^1(\mu)$ since $a \in L^1(\mu)$.
To prove $(iii) \Rightarrow (i)$, we are going to check that, under (iii), $\pi^\star(c) \leq \pi(c)$ where π is an arbitrary coupling in $\Pi(\mu, \nu)$. Note that since c is bounded from below and such that $c \leq a \oplus b$ with $(a, b) \in L^1(\mu) \times L^1(\nu)$, $c \in L^1(\pi)$. By definition of ϕ^c, $\phi \oplus \phi^c \leq c$ so that $\phi \oplus \phi^c$ is semi-integrable with respect to π and $\pi(\phi \oplus \phi^c) \leq \pi(c)$. Since $\text{supp}(\pi^\star) \subset \partial^{\bar{c}}\phi$, we have $\pi^\star(c) = \pi^\star(\phi \oplus \phi^c)$. With the equality

$$\pi^\star(\phi \oplus \phi^c) = \mu(\phi) + \nu(\phi^c) = \pi(\phi \oplus \phi^c), \tag{1.4}$$

which seems formally obvious but that we are next going establish rigorously, we conclude that

$$\pi^\star(c) = \pi^\star(\phi \oplus \phi^c) = \mu(\phi) + \nu(\phi^c) = \pi(\phi \oplus \phi^c) \leq \pi(c).$$

Let $\Gamma = \text{supp}(\pi^\star)$ and for $x \in \mathcal{X}$, $\Gamma_x = \{y \in \mathcal{Y} : (x, y) \in \Gamma\}$. For $(x, y) \in \partial^{\bar{c}}\phi$, we have $\phi(x) + \phi^c(y) = c(x, y)$ and therefore $\phi(x) > -\infty$ and $\phi^c(y) > -\infty$. With $\Gamma \subset \partial^{\bar{c}}\phi$, we deduce that $\{x \in \mathcal{X} : \pi^\star_x(\Gamma_x) > 0\} \subset \{x \in \mathbb{R} : \phi(x) > -\infty\}$. Therefore

$$\mu\left(\{x \in \mathbb{R} : \phi(x) > -\infty\}\right) \geq \mu\left(\{x \in \mathcal{X} : \pi^\star_x(\Gamma_x) > 0\}\right) \geq \int_{\mathcal{X}} \pi^\star_x(\Gamma_x)\mu(dx) = \pi^\star(\Gamma) = 1.$$

Hence there exists $x_0 \in \mathcal{X}$ such that both $\phi(x_0)$ and $a(x_0)$ belong to $\mathbb{R}$. With the definition of ϕ^c and the inequality $c \leq a \oplus b$, we deduce that

$$\forall y \in \mathcal{Y},\ (\phi^c(y))^+ \leq (c(x_0, y) - \phi(x_0))^+ \leq (a(x_0) + b(y) - \phi(x_0))^+$$

Since $b \in L^1(\nu)$, we conclude that $(\phi^c)^+ \in L^1(\nu)$. In a symmetric way, we obtain that $\phi^+ \in L^1(\mu)$ and deduce that $\phi^+ \oplus (\phi^c)^+ \in L^1(\pi)$. For $k \in \mathbb{N}$, $(\phi \vee -k) \in L^1(\mu)$, $(\phi^c \vee -k) \in L^1(\nu)$ and $(\phi \vee -k) \oplus (\phi^c \vee -k) \in L^1(\pi)$ so that

$$\begin{aligned}
\pi\left(\phi^+ \oplus (\phi^c)^+ - (\phi \vee -k) \oplus (\phi^c \vee -k)\right) &= \pi\left(\phi^+ \oplus (\phi^c)^+\right) - \pi\left((\phi \vee -k) \oplus (\phi^c \vee -k)\right) \\
&= \mu(\phi^+) + \nu((\phi^c)^+) - \mu((\phi \vee -k)) - \nu((\phi^c \vee -k)) \\
&= \mu(\phi^+ - (\phi \vee -k)) + \nu((\phi^c)^+ - (\phi^c \vee -k)).
\end{aligned}$$

With the monotone convergence theorem, we conclude that

$$\begin{aligned}
\pi\left(\phi^+ \oplus (\phi^c)^+\right) - \pi\left(\phi \oplus \phi^c\right) &= \pi\left(\phi^+ \oplus (\phi^c)^+ - \phi \oplus \phi^c\right) \\
&= \lim_{k\to\infty} \pi\left(\phi^+ \oplus (\phi^c)^+ - (\phi \vee -k) \oplus (\phi^c \vee -k)\right) \\
&= \lim_{k\to\infty} \left\{\mu\left(\phi^+ - (\phi \vee -k)\right) + \nu\left((\phi^c)^+ - (\phi^c \vee -k)\right)\right\} \\
&= \mu(\phi^+ - \phi) + \nu((\phi^c)^+ - \phi^c) \\
&= \mu(\phi^+) - \mu(\phi) + \nu((\phi^c)^+) - \nu(\phi^c) \\
&= \pi\left(\phi^+ \oplus (\phi^c)^+\right) - \mu(\phi) - \nu(\phi^c).
\end{aligned}$$

Hence $\pi(\phi \oplus \phi^c) = \mu(\phi) + \nu(\phi^c)$ and, since $\pi_\star$ also belongs to $\Pi(\mu,\nu)$, $\pi_\star(\phi \oplus \phi^c) = \mu(\phi) + \nu(\phi^c)$ so that (1.4) holds.

∎

Proof of Lemma 1.28: Let us suppose that there exist $N \geq 2$ and $(x_1, y_1), \cdots, (x_N, y_N) \in \mathrm{supp}(\pi_\star)$ such that $\sum_{i=1}^N c(x_i, y_i) > \sum_{i=1}^N c(x_i, y_{i+1})$ with $y_{N+1} = y_1$. By continuity of c there are open neighbourhoods U_i of x_i and V_i of y_i such that

$$\inf_{((u_i,v_i))_{1\leq i\leq N}\in\prod_{i=1}^N U_i\times V_i} \sum_{i=1}^N c(u_i, v_i) > \sup_{((u_i,v_i))_{1\leq i\leq N}\in\prod_{i=1}^N U_i\times V_i} \sum_{i=1}^N c(u_i, v_{i+1}) \text{ with } v_{N+1} = v_1. \tag{1.5}$$

Since $(x_i, y_i) \in \mathrm{supp}(\pi_\star)$, by Remark 1.25, $m_i := \pi_\star(U_i \times V_i) > 0$. Let μ_i and ν_i denote the marginals of the probability measure $\frac{1}{m_i}\pi_\star|_{U_i\times V_i}$ and

$$\pi = \pi_\star + \frac{\min_{1\leq i\leq N} m_i}{N} \sum_{i=1}^N \left(\mu_i \otimes \nu_{i+1} - \frac{1}{m_i}\pi_\star|_{U_i\times V_i}\right) \text{ where } \nu_{N+1} = \nu_1.$$

Since for each $i \in \{1, \cdots, N\}$, $\pi_\star \geq \pi_\star|_{U_i\times V_i}$, we have

$$\pi_\star \geq \frac{1}{N}\sum_{i=1}^N \pi_\star|_{U_i\times V_i} \geq \frac{1}{N}\sum_{i=1}^N \frac{\min_{1\leq j\leq N} m_j}{m_i}\pi_\star|_{U_i\times V_i}$$

so that π is a non-negative measure. Since $\mu_i\otimes\nu_{i+1}, \frac{1}{m_i}\pi_\star|_{U_i\times V_i} \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. The marginals of π are

$$\mu + \frac{\min_{1\leq i\leq N} m_i}{N}\sum_{i=1}^N(\mu_i - \mu_i) = \mu \text{ and } \nu + \frac{\min_{1\leq i\leq N} m_i}{N}\sum_{i=1}^N(\nu_{i+1} - \nu_i) = \nu,$$

so that $\pi \in \Pi(\mu,\nu)$. Since

$$\begin{aligned}\pi(c) &\leq \pi_\star(x)\\ &+ \frac{\min_{1\leq i\leq N} m_i}{N}\left(\sup_{((u_i,v_i))_{1\leq i\leq N}\in\prod_{i=1}^N U_i\times V_i}\sum_{i=1}^N c(u_i, v_{i+1}) - \inf_{((u_i,v_i))_{1\leq i\leq N}\in\prod_{i=1}^N U_i\times V_i}\sum_{i=1}^N c(u_i, v_i)\right),\end{aligned}$$

the inequality (1.5) ensures that $\pi_\star$ is not optimal for $V_c(\mu,\nu)$. By contraposition, we conclude that $\mathrm{supp}(\pi_\star)$ is c-cyclically monotone. ∎

Theorem 1.29 (Stability of OT)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$, $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ be continuous and bounded from below with $c \leq a \oplus b$ for some $(a,b) \in L^1(\mu)\times L^1(\nu)$. If $(\pi^n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ is a sequence of optimal couplings for $V_C(\pi^n(dx\times\mathcal{Y}), \pi^n(\mathcal{X}\times dy))$ which converges weakly to $\pi \in \Pi(\mu,\nu)$, then $V_c(\mu,\nu) = \pi(c)$.*

Corollary 1.30. *If $c \in C_b(\mathcal{X}\times\mathcal{Y})$, then V_c is continuous on $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$.*

Proof of Corollary 1.30: By Theorem 1.6, V_c is lower semi-continuous. Let $(\mu,\nu) \in$

$\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ and $(\mu_n,\nu_n)_{n\in\mathbb{N}}\subset\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ such that μ_n (resp. ν_n) converges weakly to μ (resp. ν) as $n\to\infty$ and $\lim_{n\to\infty}V_c(\mu_n,\nu_n)=\limsup_{(\tilde\mu,\tilde\nu)\to(\mu,\nu)}V_c(\tilde\mu,\tilde\nu)$. By Theorem 1.6, for each $n\in\mathbb{N}$, there exists $\pi_n\in\Pi(\mu_n,\nu_n)$ optimal for $V_c(\mu_n,\nu_n)$ i.e. $V_c(\mu_n,\nu_n)=\pi_n(c)$. By Lemma 1.7, we may extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ converging weakly to $\pi\in\Pi(\mu,\nu)$ as $k\to\infty$. Using Theorem 1.29 for the first equality and the lower semi-continuity of V_c for the last inequality, we conclude that

$$\begin{aligned} V_c(\mu,\nu)=\pi(c)=\lim_{k\to\infty}\pi_{n_k}(c)=\lim_{k\to\infty}V_c(\mu_{n_k},\nu_{n_k})&=\limsup_{(\tilde\mu,\tilde\nu)\to(\mu,\nu)}V_c(\tilde\mu,\tilde\nu)\\ \geq\liminf_{(\tilde\mu,\tilde\nu)\to(\mu,\nu)}V_c(\tilde\mu,\tilde\nu)\geq V_c(\mu,\nu).& \end{aligned}$$

∎

The proof of Theorem 1.29 relies on the next lemma

Lemma 1.31. *For $N\in\mathbb{N}^*$, $\mathrm{supp}(\pi^{\otimes N})=\mathrm{supp}(\pi)^N$.*

Proof of Lemma 1.31: Let $(z_1,\cdots,z_N)\in\mathrm{supp}(\pi)^N$. Any open neighbourhood B of $(z_1,\cdots,z_N)$ contains a product $\prod_{i=1}^N B_i$ of open neighbourhoods B_i of the z_i which, by Remark 1.25, are such that $\pi(B_i)>0$ since $z_i\in\mathrm{supp}(\pi)$. Therefore

$$\pi^{\otimes N}(B)\geq\pi^{\otimes N}\left(\prod_{i=1}^N B_i\right)=\prod_{i=1}^N\pi(B_i)>0$$

and $\mathrm{supp}(\pi)^N\subset\mathrm{supp}(\pi^{\otimes N})$. On the other hand, $\mathrm{supp}(\pi)^N$ is closed and satisfies $\pi^{\otimes N}(\mathrm{supp}(\pi)^N)=(\pi(\mathrm{supp}(\pi))^N=1$. By definition of the support (see Definition 1.24), we conclude that $\mathrm{supp}(\pi)^N=\mathrm{supp}(\pi^{\otimes N})$. ∎

Proof of Theorem 1.29: For $N\in\mathbb{N}^*$, $(\mathcal{X}\times\mathcal{Y})^N\ni((x_1,y_1),\cdots,(x_N,y_N))\mapsto c_N((x_i,y_i)_{1\leq i\leq N})=\sum_{i=1}^N(c(x_i,y_{i+1})-c(x_i,y_i))$ (with the usual convention $y_{N+1}=y_1$) is continuous. The support Γ_n of π_n is c-cyclically monotone by Lemma 1.28. Hence $\Gamma_n^N\subset\{c_N\geq 0\}$ and

$$\pi_n^{\otimes N}(\{c_N\geq 0\})\geq\pi_n^{\otimes N}\left(\Gamma_n^N\right)=(\pi_n(\Gamma_n))^N=1.$$

Since $\{c_N\geq 0\}$ is closed by continuity of c_N, using the Portmanteau theorem (see Theorem 5.2 4)), we deduce that

$$\pi^{\otimes N}(\{c_N\geq 0\})\geq\limsup_{n\to\infty}\pi_n^{\otimes N}(\{c_N\geq 0\})=1.$$

The closed set $\{c_N\geq 0\}$ has full $\pi^{\otimes N}$ measure and therefore contains the support of $\pi^{\otimes N}$, which, by Lemma 1.31, is equal to the product Γ^N of the support Γ of π. Hence Γ is c-cyclical monotonic and, by Theorem 1.26, π is optimal for $V_c(\mu,\nu)$. ∎

1.6 Brenier's theorem

Theorem 1.32. *Let* $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $c(x,y) = \frac{1}{2}|x-y|^2$ *and* $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ *be such that* $\int_{\mathbb{R}^d} |x|^2 \mu(dx) + \int_{\mathbb{R}^d} |y|^2 \nu(dy) < \infty$ *and* $\mu << \lambda$ *where* λ *denotes the Lebesgue measure on* $\mathbb{R}^d$*. Then there is a unique optimal coupling* $\pi_\star$ *and it writes*

$$\pi_\star(dx,dy) = \mu(dx)\delta_{\nabla\varphi(x)}(dy)$$

for some convex function $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *such that* $\mu(\{\varphi < \infty\}) = \mu(\{\varphi \text{ differentiable}\}) = 1$.

Definition 1.33. *The domain of a convex function* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *is the convex set* $\mathrm{dom}(\varphi) = \{x \in \mathbb{R}^d : \varphi(x) < +\infty\}$.

The proof of Brenier's theorem relies on the following properties of convex functions from $\mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$.

Proposition 1.34. *Let* $\varphi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *be a convex function and*

$$\mathrm{diff}(\varphi) = \{x \in \mathrm{int}(\mathrm{dom}(\varphi)) : \varphi \text{ is differentiable at } x\}.$$

Then we have

- $\lambda(\mathrm{dom}(\varphi) \setminus \mathrm{diff}(\varphi)) = 0$,
- $\forall (x,y) \in \mathrm{diff}(\varphi) \times \mathbb{R}^d$, $\varphi(y) \geq \varphi(x) + \nabla\varphi(x).(y-x)$.

Remark 1.35. *When* $x \in \mathrm{diff}(\varphi)$, *then* $x \in \mathrm{int}(\mathrm{dom}(\varphi))$ *and for each* $y \in \mathbb{R}^d$, $(1-\varepsilon)x + \varepsilon y \in \mathrm{dom}(\varphi)$ *when* $\varepsilon \in [0,1)$ *is small enough. The convexity then ensures that*

$$0 \leq \frac{1}{\varepsilon}\left((1-\varepsilon)\varphi(x) + \varepsilon\varphi(y) - \varphi((1-\varepsilon)x + \varepsilon y)\right) = \frac{1}{\varepsilon}(\varphi(x) - \varphi((1-\varepsilon)x + \varepsilon y) - \varphi(x) + \varphi(y)$$

and by taking the limit $\varepsilon \to 0+$, *we conclude that* $0 \leq -\nabla\varphi(x).(y-x) - \varphi(x) + \varphi(y)$. *The fact that* $\lambda\left(\mathrm{int}(\mathrm{dom}(\varphi)) \setminus \mathrm{diff}(\varphi)\right) = 0$ *is a standard result of convex analysis (see for instance Theorem 25.5 [15]). To deduce that* $\lambda(\mathrm{dom}(\varphi) \setminus \mathrm{diff}(\varphi)) = 0$, *it is enough to check that the Lebesgue measure of the boundary of* $\mathrm{dom}(\varphi)$ *is* 0 *(see for instance Theorem 5.2 [11]).*

Proof of Theorem 1.32: Since $c(x,y) = \frac{1}{2}|x-y|^2$ is continuous and bounded from below, there exists an optimal coupling $\pi_\star$ by Theorem 1.6. Since $c(x,y) \leq |x|^2 + |y|^2$ and $\int_{\mathbb{R}^d} |x|^2\mu(dx) + \int_{\mathbb{R}^d} |y|^2\nu(dy) < \infty$, according to Theorem 1.26, there exists a c-concave function ϕ such that $\Gamma := \mathrm{supp}(\pi_\star) \subset \partial^c\phi$. Moreover, $\phi > -\infty$ on

$$\Gamma_{\mathcal{X}} = \{x \in \mathbb{R}^d : \exists y \in \mathbb{R}^d \text{ such that } (x,y) \in \Gamma\}.$$

Since $\Gamma \subset \Gamma_{\mathcal{X}} \times \mathbb{R}^d$, $1 = \pi_\star(\Gamma) \leq \pi_\star(\Gamma_{\mathcal{X}} \times \mathbb{R}^d) = \mu(\Gamma_{\mathcal{X}})$. By Example 1.19, the function $\varphi(x) = \frac{|x|^2}{2} - \phi(x)$ is lower semi-continuous and convex. We are going to check that $\pi_\star = \mu(dx)\delta_{\nabla\varphi(x)}(dy)$. Since this ensures that each optimal coupling is a Monge coupling, uniqueness follows from Proposition 1.4. Let $x \in \mathbb{R}^d$ be such that $\phi(x) > -\infty$ i.e. $x \in \text{dom}(\varphi)$. We have

$$y \in \partial^c\phi(x) \Leftrightarrow \phi(x) + \phi^c(y) = \frac{1}{2}|x-y|^2 \Leftrightarrow \inf_{z\in\mathbb{R}^d}\left\{\frac{1}{2}|z-y|^2 - \phi(z)\right\} = \frac{1}{2}|x-y|^2 - \phi(x)$$
$$\Leftrightarrow \forall z \in \mathbb{R}^d,\ \varphi(z) - y.z \geq \varphi(x) - y.x \Leftrightarrow \forall z \in \mathbb{R}^d,\ \varphi(z) \geq \varphi(x) + y.(z-x).$$

When $x \in \text{diff}(\varphi)$, then for $w \in \mathbb{R}^d$, $\lim_{h\to 0+}\frac{1}{h}(\varphi(x+hw) - \varphi(x)) = \nabla\varphi(x).w$ and we deduce that

$$y \in \partial^c\phi(x) \Rightarrow \forall w \in \mathbb{R}^d,\ \nabla\varphi(x).w \geq y.w \Rightarrow y = \nabla\varphi(x).$$

Hence for $x \in \text{diff}(\varphi)$, $\partial^c\phi(x) \subset \{\nabla\varphi(x)\}$. In view of the previous equivalence and the last assertion in Proposition 1.34, we even have

$$\forall x \in \text{diff}(\varphi),\ \partial^c\phi(x) = \{\nabla\varphi(x)\}. \tag{1.6}$$

Since $\text{diff}(\varphi) \subset \text{dom}(\varphi)$, we have

$$\mathbb{R}^d \setminus \text{diff}(\varphi) \subset \{\mathbb{R}^d \setminus \text{dom}(\varphi)\} \cup \{\text{dom}(\varphi) \setminus \text{diff}(\varphi)\}.$$

By Proposition 1.34, $\lambda(\{\text{dom}(\varphi) \setminus \text{diff}(\varphi)\}) = 0$ so that $\mu(\{\text{dom}(\varphi) \setminus \text{diff}(\varphi)\}) = 0$ since $\mu << \lambda$. Since $\{\mathbb{R}^d \setminus \text{dom}(\varphi)\} = \{\phi = -\infty\} \subset \mathbb{R}^d \setminus \Gamma_{\mathcal{X}}$, $\mu(\mathbb{R}^d \setminus \text{dom}(\varphi)) = 0$ and therefore $\mu(\mathbb{R}^d \setminus \text{diff}(\varphi)) = 0$. Hence $\pi_\star(dx,dy)$, $x \in \text{diff}(\varphi)$. Since $\text{supp}(\pi_\star) \subset \partial^c\phi$, $\pi_\star(dx,dy)$ a.e., $y \in \partial^c\phi(x)$. With (1.6), we conclude that $\pi_\star(dx,dy)$ a.e., $y = \nabla\varphi(x)$ i.e. $\pi_\star(dx,dy) = \mu(dx)\delta_{\nabla\varphi(x)}(dy)$.

∎

1.7 The Wasserstein distance

1.7.1 General case

Let $q \geq 1$,

$$\mathcal{P}_\rho(\mathcal{X}) = \left\{\eta \in \mathcal{P}(\mathcal{X}) :\ \exists x_0 \in \mathcal{X},\ \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\eta(dx) < \infty\right\},$$
$$\text{and for } \mu,\nu \in \mathcal{P}_\rho(\mathcal{X}),\ \mathcal{W}_\rho(\mu,\nu) = \left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{X}} d^\rho_{\mathcal{X}}(x,y)\pi(dx,dy)\right)^{1/\rho}.$$

Since, by the triangle and Jensen inequalities, $d^\rho_{\mathcal{X}}(x_1,x) \leq 2^{\rho-1}\left(d^\rho_{\mathcal{X}}(x_1,x_0) + d^\rho_{\mathcal{X}}(x_0,x)\right)$ for $x_0,x_1,x \in \mathcal{X}$, $\int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\eta(dx) < \infty \Leftrightarrow \forall x_1 \in \mathcal{X},\ \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_1,x)\eta(dx) < \infty$. Therefore $\mathcal{P}_\rho(\mathcal{X}) = \left\{\eta \in \mathcal{P}(\mathcal{X}) :\ \forall x_0 \in \mathcal{X},\ \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\eta(dx) < \infty\right\}$. Moreover, $d^\rho_{\mathcal{X}}(x,y) \leq 2^{\rho-1}\left(d^\rho_{\mathcal{X}}(x,x_0) + d^\rho_{\mathcal{X}}(x_0,y)\right)$ implies that $\mathcal{W}_\rho(\mu,\nu) < \infty$ for $\mu,\nu \in \mathcal{P}_\rho(\mathcal{X})$.
Moreover, since $\mathcal{X}\times\mathcal{X} \ni (x,y) \mapsto d_{\mathcal{X}}(x,y)$ is bounded from below by 0 and 1-Lipshitz and therefore lower semi-continuous, according to Theorem 1.6,

$$\forall \mu,\nu \in \mathcal{P}_\rho(\mathcal{X}),\ \exists \pi_\star \in \Pi(\mu,\nu),\ \mathcal{W}^\rho_\rho(\mu,\nu) = \int_{\mathcal{X}\times\mathcal{X}} d^\rho_{\mathcal{X}}(x,y)\pi_\star(dx,dy). \tag{1.7}$$

When $1 \leq \tilde{\rho} \leq \rho$, by Jensen's inequality, we deduce that for $\mu, \nu \in \mathcal{P}_\rho(\mathcal{X})$,

$$\mathcal{W}_\rho(\mu,\nu) = \left(\int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}^\rho(x,y)\pi_\star(dx,dy)\right)^{1/\rho} \geq \left(\int_{\mathcal{X}\times\mathcal{X}} d_{\mathcal{X}}^{\tilde{\rho}}(x,y)\pi_\star(dx,dy)\right)^{1/\tilde{\rho}} \geq \mathcal{W}_{\tilde{\rho}}(\mu,\nu). \tag{1.8}$$

Theorem 1.36. *For $\rho \geq 1$, $\mathcal{W}_\rho$ is a metric on $\mathcal{P}_\rho(\mathcal{X})$.*

Proof: Since the image of $\pi \in \Pi(\mu,\nu)$ by $\mathcal{X}\times\mathcal{X} \ni (x,y) \mapsto (y,x) \in \mathcal{X}\times\mathcal{X}$ belongs to $\Pi(\nu,\mu)$, $\mathcal{W}_\rho$ is symmetric : $\mathcal{W}_\rho(\mu,\nu) = \mathcal{W}_\rho(\nu,\mu)$.
Since for the identity function $i_{\mathcal{X}}$ introduced in Remark 1.2, $(i_{\mathcal{X}}, i_{\mathcal{X}})\#\mu \in \Pi(\mu,\mu)$, we have $\mathcal{W}_\rho(\mu,\mu) \leq \int_{\mathcal{X}} |x-x|^\rho \mu(dx) = 0$. If conversely $\mathcal{W}_\rho(\mu,\nu) = 0$, then the optimal coupling $\pi_\star \in \Pi(\mu,\nu)$ given by (1.7) is such that $d_{\mathcal{X}}(x,y) = 0$ and therefore $x = y$, $\pi_\star(dx,dy)$ a.e.. Therefore, the image ν of $\pi_\star$ by $\mathcal{X}\times\mathcal{X} \ni (x,y) \mapsto y \in \mathcal{X}$ is equal to its image μ by $\mathcal{X}\times\mathcal{X} \ni (x,y) \mapsto x \in \mathcal{X}$.
Let $\mu,\nu,\tilde{\mu} \in \mathcal{P}_\rho(\mathcal{X})$. By (1.7), there exist $\pi \in \Pi(\mu,\nu)$ optimal for $\mathcal{W}_\rho(\mu,\nu)$ and $\tilde{\pi}(dy,d\tilde{x}) = \nu(dy)\tilde{\pi}_y(d\tilde{x}) \in \Pi(\nu,\tilde{\mu})$ optimal for $\mathcal{W}_\rho(\mu,\nu)$. The coupling $\int_{y\in\mathcal{X}} \tilde{\pi}_y(d\tilde{x})\pi(dx,dy)$ belongs to $\Pi(\mu,\tilde{\mu})$. Therefore, using the Minkowski inequality for the second inequality, we have

$$\begin{aligned}
\mathcal{W}_\rho(\mu,\tilde{\mu}) &\leq \left(\int_{\mathcal{X}^3} (d_{\mathcal{X}}(x,y) + d_{\mathcal{X}}(y,\tilde{x}))^\rho\, \tilde{\pi}_y(d\tilde{x})\pi(dx,dy)\right)^{1/\rho} \\
&\leq \left(\int_{\mathcal{X}^3} d_{\mathcal{X}}^\rho(x,y)\tilde{\pi}_y(d\tilde{x})\pi(dx,dy)\right)^{1/\rho} + \left(\int_{\mathcal{X}^3} d_{\mathcal{X}}^\rho(y,\tilde{x})\tilde{\pi}_y(d\tilde{x})\pi(dx,dy)\right)^{1/\rho} \\
&= \left(\int_{\mathcal{X}^2} d_{\mathcal{X}}^\rho(x,y)\pi(dx,dy)\right)^{1/\rho} + \left(\int_{\mathcal{X}^2} d_{\mathcal{X}}^\rho(y,\tilde{x})\tilde{\pi}_y(d\tilde{x})\nu(dy)\right)^{1/\rho} \\
&= \mathcal{W}_\rho(\mu,\nu) + \mathcal{W}_\rho(\nu,\tilde{\mu}).
\end{aligned}$$

∎

Proposition 1.37.

$$\forall \mu,\nu \in \mathcal{P}_1(\mathcal{X}),\ \mathcal{W}_1(\mu,\nu) = \sup_{f:\mathcal{X}\to\mathbb{R}\ 1-\text{Lipschitz}} \{\mu(f) - \nu(f)\}.$$

Proof: Let $f : \mathcal{X} \to \mathbb{R}$ be 1-Lipschitz. For $x,y \in \mathcal{X}$, $f \ominus f(x,y) = f(x) - f(y) \leq d_{\mathcal{X}}(x,y)$. Therefore,

$$\forall \pi \in \Pi(\mu,\nu),\ \{\mu(f) - \nu(f)\} = \pi(f \ominus f) \leq \pi(c),$$

so that $\sup_{f:\mathcal{X}\to\mathbb{R}\ 1-\text{Lipschitz}}\{\mu(f) - \nu(f)\} \leq \mathcal{W}_1(\mu,\nu)$. By (1.7), there exist $\pi_\star \in \Pi(\mu,\nu)$ optimal for $\mathcal{W}_1(\mu,\nu)$. Since $d_{\mathcal{X}}(x,y) \leq d_{\mathcal{X}}(x,x_0) + d_{\mathcal{X}}(x_0,y)$ with $d_{\mathcal{X}}(x_0,\cdot) \in L^1(\mu) \cap L^1(\nu)$, by the fundamental theorem of Optimal Transport (see Theorem 1.26), there exists $\phi : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ $d_{\mathcal{X}}$-concave such that $\text{supp}(\pi_\star) \in \partial^{d_{\mathcal{X}}}\phi$ i.e. $\pi_\star(dx,dy)$ a.e., $d_{\mathcal{X}}(x,y) = \phi(x) + \phi^{d_{\mathcal{X}}}(y)$. In particular $\phi(x) > -\infty$ for some $x \in \mathcal{X}$ and, by Example 1.17, ϕ is $\mathbb{R}$-valued and 1-Lipschitz and $\phi^{d_{\mathcal{X}}} = -\phi$. Therefore,

$$\mathcal{W}_1(\mu,\nu) = \pi_\star(c) = \pi_\star(\phi \ominus \phi) = \mu(\phi) - \nu(\phi) \leq \sup_{f:\mathcal{X}\to\mathbb{R}\ 1-\text{Lipschitz}} \{\mu(f) - \nu(f)\}.$$

∎

Definition 1.38. *We say that a sequence* $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}_\rho(\mathcal{X})$ *converges to* $\mu \in \mathcal{P}_\rho(\mathcal{X})$ *in* $\mathcal{P}_\rho(\mathcal{X})$ *and denote* $\mu_n \overset{\mathcal{P}_\rho(\mathcal{X})}{\longrightarrow} \mu$ *when* μ_n *converges weakly to* μ *as* $n \to \infty$ *and*

$$\exists x_0 \in \mathcal{X},\ \lim_{n\to\infty} \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu_n(dx) = \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu(dx).$$

Example 1.39. *Let* $(X_i)_{i\geq 1}$ *be* $\mathbb{R}^d$*-valued random variables i.i.d. according to some element of* $\mathcal{P}_2(\mathbb{R}^d)$ *and* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ *for* $n \in \mathbb{N}^*$*. By the central limit theorem, the distribution* μ_n *of* $\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1])$ *converges weakly to* $G \sim \mathcal{N}_d(0, \mathrm{Cov}(X_1))$*, where* $\mathrm{Cov}(X_1) \in \mathbb{R}^{d\times d}$ *denotes the covariance matrix of* X_1*. Since*

$$\mathbb{E}[|\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1])|^2] = n\mathrm{tr}\left(\mathrm{Cov}(\bar{X}_n)\right) = \mathrm{tr}\left(\mathrm{Cov}(X_1)\right) = \mathbb{E}[|G|^2],$$

we have that $\mu_n \overset{\mathcal{P}_2(\mathbb{R}^d)}{\longrightarrow} \mathcal{N}_d(0, \mathrm{Cov}(X_1))$*. In view of Proposition 1.41 below, we deduce that*

$$\forall f : \mathbb{R}^d \to \mathbb{R} \text{ continuous s.t. } \sup_{x\in\mathbb{R}^d} \frac{|f(x)|}{1+|x|^2} < \infty,\ \lim_{n\to\infty} \mathbb{E}\left[f(\sqrt{n}(\bar{X}_n - \mathbb{E}[X_1]))\right] = \mathbb{E}[f(G)].$$

Theorem 1.40. *For* $\rho \geq 1$*, the metric* $\mathcal{W}_\rho$ *metricizes the convergence in* $\mathcal{P}_\rho(\mathcal{X})$*.*

Proof of Theorem 1.40: Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{X})$. By (1.8) and Proposition 1.37,

$$\mathcal{W}_\rho(\mu_n,\mu) \geq \mathcal{W}_1(\mu_n,\nu) \geq \sup_{f:\mathcal{X}\to\mathbb{R}\ \text{1-Lipschitz bounded by 1}} \{\mu_n(f) - \mu(f)\}.$$

With Theorem 5.2 3), we deduce that $\lim_{n\to\infty} \mathcal{W}_\rho(\mu_n,\mu) = 0$ implies that μ_n converges weakly to μ as $n \to \infty$. Moreover, since by the triangle inequality $|\mathcal{W}_\rho(\mu_n,\delta_{x_0}) - \mathcal{W}_\rho(\mu,\delta_{x_0})| \leq \mathcal{W}_\rho(\mu_n,\mu)$, it also implies that

$$\int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu_n(dx) = \mathcal{W}^\rho_\rho(\mu_n,\delta_{x_0}) \overset{n\to\infty}{\longrightarrow} \mathcal{W}^\rho_\rho(\mu,\delta_{x_0}) = \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu(dx).$$

Conversely, let us suppose that μ_n converges weakly to μ and $\lim_{n\to\infty}\int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu_n(dx) = \int_{x\in\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu(dx)$. By Proposition 5.4, there exist $X_n \sim \mu_n$ and $X \sim \mu$ such that $\lim_{n\to\infty} X_n = X$ a.s. and $\lim_{n\to\infty}\mathbb{E}[d^\rho_{\mathcal{X}}(x_0,X_n)] = \mathbb{E}[d^\rho_{\mathcal{X}}(x_0,X)]$. Since $0 \leq 2^{\rho-1}(d^\rho_{\mathcal{X}}(x_0,X) + d^\rho_{\mathcal{X}}(x_0,X_n)) - d^\rho_{\mathcal{X}}(X,X_n) \overset{n\to\infty}{\longrightarrow} 2^\rho d^\rho_{\mathcal{X}}(x_0,X)$ a.s., by Fatou Lemma, we get

$$\begin{aligned}2^\rho\mathbb{E}[d^\rho_{\mathcal{X}}(x_0,X)] &\leq \liminf_{n\to\infty} \mathbb{E}\left[2^{\rho-1}(d^\rho_{\mathcal{X}}(x_0,X) + d^\rho_{\mathcal{X}}(x_0,X_n)) - d^\rho_{\mathcal{X}}(X,X_n)\right] \\ &= 2^\rho\mathbb{E}[d^\rho_{\mathcal{X}}(x_0,X)] - \limsup_{n\to\infty}\mathbb{E}\left[d^\rho_{\mathcal{X}}(X,X_n)\right].\end{aligned}$$

Since $\mathcal{W}^\rho_\rho(\mu_n,\mu) \leq \mathbb{E}\left[d^\rho_{\mathcal{X}}(X,X_n)\right]$, we conclude that

$$\limsup_{n\to\infty} \mathcal{W}^\rho_\rho(\mu_n,\mu) \leq \limsup_{n\to\infty}\mathbb{E}\left[d^\rho_{\mathcal{X}}(X,X_n)\right] = 0.$$

■

Proposition 1.41. *The convergence* $\mu_n \overset{\mathcal{P}_\rho(\mathcal{X})}{\longrightarrow} \mu$ *is equivalent to*

$$\forall f \in C_\rho(\mathcal{X}),\ \lim_{n\to\infty} \mu_n(f) = \mu(f),$$

where $C_\rho(\mathcal{X})$ *denotes the set of functions* $f : \mathcal{X} \to \mathbb{R}$ *continuous such that* $\sup_{x\in\mathcal{X}} \frac{|f(x)|}{1+d^\rho_{\mathcal{X}}(x_0,x)} < \infty$.

Proof: The sufficient condition is clear since $\sup_{x\in\mathcal{X}} \frac{|f(x)|}{1+d^\rho_{\mathcal{X}}(x_0,x)} < \infty$ when f is bounded and when $f(x) = d^\rho_{\mathcal{X}}(x_0,x)$ which is a continuous function.

To prove the necessary condition, we set $\nu_n = d^\rho_{\mathcal{X}}(x_0,\cdot)\#\mu_n$ and $\nu = d^\rho_{\mathcal{X}}(x_0,\cdot)\#\mu$ which are probability measures on $\mathbb{R}_+$ such that ν_n converges weakly to ν and $\lim_{n\to\infty}\int_{\mathbb{R}} x\nu_n(dx) = \int_{\mathbb{R}} x\nu(dx)$.
By Lemmas 1.42 and 1.43 below, we have that for $U \sim \mathcal{U}[0,1]$, $\lim_{n\to\infty} F^{-1}_{\nu_n}(U) = F^{-1}_\nu(U)$ a.s. and $\lim_{n\to\infty}\mathbb{E}[F^{-1}_{\nu_n}(U)] = \mathbb{E}[F^{-1}_\nu(U)]$. Since $(F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U))^+ \le F^{-1}_\nu(U)$, by Lebesgue's theorem, $\lim_{n\to\infty}\mathbb{E}[(F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U))^+] = 0$. Since

$$|F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U)| = 2(F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U))^+ + F^{-1}_{\nu_n}(U) - F^{-1}_\nu(U),$$

we deduce that $\lim_{n\to\infty}\mathbb{E}\left[|F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U)|\right] = 0$. Since for $k\in\mathbb{N}$, $\mathbb{R}\ni y \mapsto (|y|-k)^+$ is 1-Lipschitz,

$$\begin{aligned}&\sup_{k\in\mathbb{N}}\left|\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] - \mathbb{E}[(|F^{-1}_\nu(U)|-k)^+]\right|\\ &\le \sup_{k\in\mathbb{N}}\mathbb{E}\left[|(|F^{-1}_{\nu_n}(U)|-k)^+ - (|F^{-1}_\nu(U)|-k)^+|\right] \le \mathbb{E}\left[|F^{-1}_\nu(U) - F^{-1}_{\nu_n}(U)|\right] \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

For $\varepsilon > 0$, we may thus choose $n_\varepsilon \in \mathbb{N}$ such that

$$\sup_{n\ge n_\varepsilon}\sup_{k\in\mathbb{N}}\left|\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] - \mathbb{E}[(|F^{-1}_\nu(U)|-k)^+]\right| \le \frac{\varepsilon}{2}.$$

Since by Lebesgue's theorem, $\lim_{k\to\infty}\mathbb{E}[(|F^{-1}_\nu(U)|-k)^+] = 0$, we may choose $k_{n_\varepsilon}\in\mathbb{N}$ such that $\sup_{k\ge k_{n_\varepsilon}}\mathbb{E}[(|F^{-1}_\nu(U)|-k)^+] \le \frac{\varepsilon}{2}$ and therefore

$$\sup_{k\ge k_{n_\varepsilon}}\sup_{n\ge n_\varepsilon}\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] \le \frac{\varepsilon}{2}+\frac{\varepsilon}{2} = \varepsilon.$$

On the other hand, for $n \in \{0,\cdots,n_\varepsilon - 1\}$, since by Lebesgue's theorem, $\lim_{k\to\infty}\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] = 0$, we may choose $k_n\in\mathbb{N}$ such that $\sup_{k\ge k_n}\mathbb{E}[(|F^{-1}_{\nu_n}(U)| - k)^+] \le \varepsilon$. We then have $\sup_{k\ge\max(k_0,\cdots,k_{n_\varepsilon})}\sup_{n\in\mathbb{N}}\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] \le \varepsilon$ and deduce that

$$\lim_{k\to\infty}\sup_{n\in\mathbb{N}}\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] = 0.$$

Let now $f\in C_\rho(\mathcal{X})$, $C = \sup_{x\in\mathcal{X}}\frac{|f(x)|}{1+d^\rho_{\mathcal{X}}(x_0,x)}$, $X_n\sim\mu_n$, $X\sim\mu$ and $f_k(x) = (-C(1+k))\vee f(x)\wedge(C(1+k))$, for $k\in\mathbb{N}$. We have $|f(X_n)| \le C(1+d^\rho_{\mathcal{X}}(x_0,X_n))$ and $(|f(X_n)| - C(1+k))^+ \le C(d^\rho_{\mathcal{X}}(x_0,X_n)-k)^+$ and therefore

$$\sup_{n\in\mathbb{N}}\mathbb{E}[(|f(X_n) - C(1+k))^+] \le C\sup_{n\in\mathbb{N}}\mathbb{E}[(|F^{-1}_{\nu_n}(U)|-k)^+] \overset{k\to\infty}{\longrightarrow} 0.$$

Since $|f(x) - f_k(x)| \leq (|f(x)| - C(1+k))^+$, we deduce that

$$\sup_{n\in\mathbb{N}} |\mathbb{E}[f(X_n)] - \mathbb{E}[f_k(X_n)]| \overset{k\to\infty}{\longrightarrow} 0.$$

Moreover, by Lebesgue's theorem, $\lim_{k\to\infty} \mathbb{E}[f_k(X)] = \mathbb{E}[f(X)]$. We deduce that for $\varepsilon > 0$, we may choose k_ε such that

$$\sup_{n\in\mathbb{N}} |\mathbb{E}[f(X_n)] - \mathbb{E}[f_{k_\varepsilon}(X_n)]| \leq \frac{\varepsilon}{3} \text{ and } |\mathbb{E}[f(X)] - \mathbb{E}[f_{k_\varepsilon}(X)]| \leq \frac{\varepsilon}{3}.$$

Since f_{k_ε} is continuous and bounded, the weak convergence of μ_n to μ ensures that $\lim_{n\to\infty} \mathbb{E}[f_{k_\varepsilon}(X_n)] = \mathbb{E}[f_{k_\varepsilon}(X)]$ so that $\exists n_\varepsilon \in \mathbb{N}$, $\sup_{n\geq n_\varepsilon} |\mathbb{E}[f_{k_\varepsilon}(X_n)] - \mathbb{E}[f_{k_\varepsilon}(X)]| \leq \frac{\varepsilon}{3}$. We conclude that

$$\begin{aligned}\forall n \geq n_\varepsilon, |\mathbb{E}[f(X_n)] - \mathbb{E}[f(X)]| &\leq |\mathbb{E}[f(X_n)] - \mathbb{E}[f_{k_\varepsilon}(X_n)]| + |\mathbb{E}[f_{k_\varepsilon}(X_n)] - \mathbb{E}[f_{k_\varepsilon}(X)]| \\ &\quad + |\mathbb{E}[f_{k_\varepsilon}(X)] - \mathbb{E}[f(X)]| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

■

1.7.2 The real line case

We denote by $F_\eta(x) = \eta((-\infty, x])$, $x \in \mathbb{R}$ and $F_\eta^{-1}(u) = \inf\{x \in \mathbb{R} : F_\eta(x) \geq u\}$, $u \in (0,1)$ the cumulative distribution function and the quantile function of a probability measure $\eta \in \mathcal{P}(\mathbb{R})$. Let us recall the inverse sampling transform.

Lemma 1.42.

$$\forall (u,x) \in (0,1) \times \mathbb{R},\ F_\eta^{-1}(u) \leq x \Leftrightarrow u \leq F_\eta(x).$$

Moreover, for $U \sim \mathcal{U}[0,1]$, $F_\eta^{-1}(U) \sim \eta$.

Proof: The cumulative distribution function F_η is right-continuous and non-decreasing and, by definition of $F_\eta^{-1}(u)$, we have $\forall x > F_\eta^{-1}(u)$, $F_\eta(x) \geq u$. Therefore $F_\eta(F_\eta^{-1}(u)) \geq u$. As a consequence, with the monotonicity of F_η, we obtain

$$F_\eta^{-1}(u) \leq x \Rightarrow F_\eta(F_\eta^{-1}(u)) \leq F_\eta(x) \Rightarrow u \leq F_\eta(x).$$

Conversely, $u \leq F_\eta(x) \Rightarrow F_\eta^{-1}(u) \leq x$, by definition of $F_\eta^{-1}(u)$.

The equivalence implies that for $U \sim \mathcal{U}[0,1]$, $\mathbb{P}(F_\eta^{-1}(U) \leq x) = \mathbb{P}(U \leq F_\eta(x)) = F_\eta(x)$ so that $F_\eta^{-1}(U)$ has the cumulative distribution function F_η and therefore is distributed according to η. ■

Lemma 1.43. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}(\mathbb{R})$ which converges weakly to μ as $n \to \infty$. Then, for each continuity point u of F_μ^{-1} in $(0,1)$, we have $\lim_{n\to+\infty} F_{\mu_n}^{-1}(u) = F_\mu^{-1}(u)$. In particular, for $U \sim \mathcal{U}[0,1]$, $F_{\mu_n}^{-1}(U)$ converges a.s. to $F_\mu^{-1}(U)$ as $n \to \infty$.*

Proof: Since the function F_μ^{-1} is non-decreasing on $(0,1)$, it has at most countably many discontinuity points in $(0,1)$. Therefore the second assertion is a consequence of the first.

By definition of F_μ^{-1} we have

$$\forall u \in (0,1),\ \forall x < F_\mu^{-1}(u),\ F_\mu(x) < u. \tag{1.9}$$

Moreover, $F_\mu(F_\mu^{-1}(u)) \geq u$ by definition of F_μ^{-1} and right-continuity of F_μ. If $F_\mu(F_\mu^{-1}(u)) > u$ then $\forall x \geq F_\mu^{-1}(u)$, $F_\mu(x) \geq F_\mu(F_\mu^{-1}(u)) > u$. If $F_\mu(F_\mu^{-1}(u)) = u$ and there exists some $x > F_\mu^{-1}(u)$ with $F_\mu(x) = u$, then for each $v \in (u,1)$, $F_\mu^{-1}(v) \geq x > F_\mu^{-1}(u)$ and F_μ is not right-continuous at u. We deduce that

$$\forall u \in (0,1),\ F_\mu^{-1} \text{ right-continuous at } u \Rightarrow \forall x > F_\mu^{-1}(u),\ F_\mu(x) > u. \tag{1.10}$$

Let $u \in (0,1)$ be a continuity point of F_μ^{-1}. Since the non-decreasing function F_μ has at most countably many discontinuity points on $\mathbb{R}$, for $k \in \mathbb{N}^*$, there exists $\varepsilon_k \in (0, \frac{1}{k}]$ such that F_μ is continuous both at $F_\mu^{-1}(u) - \varepsilon_k$ and $F_\mu^{-1}(u) + \varepsilon_k$ so that, by Theorem 5.2 8),

$$\lim_{n\to\infty} F_{\mu_n}(F_\mu^{-1}(u) - \varepsilon_k) = F_\mu(F_\mu^{-1}(u) - \varepsilon_k) \text{ and } \lim_{n\to\infty} F_{\mu_n}(F_\mu^{-1}(u) + \varepsilon_k) = F_\mu(F_\mu^{-1}(u) + \varepsilon_k).$$

Moreover, by (1.9), $F_\mu(F_\mu^{-1}(u) - \varepsilon_k) < u$ and by (1.10), $F_\mu(F_\mu^{-1}(u) + \varepsilon_k) > u$. We deduce the existence of $N_k < \infty$ such that for $n \geq N_k$, $F_{\mu_n}(F_\mu^{-1}(u) - \varepsilon_k) < u < F_{\mu_n}(F_\mu^{-1}(u) + \varepsilon_k)$ so that, by definition of $F_{\mu_n}^{-1}(u)$,

$$\forall n \geq N_k,\ F_\mu^{-1}(u) - \frac{1}{k} \leq F_\mu^{-1}(u) - \varepsilon_k < F_{\mu_n}^{-1}(u) \leq F_\mu^{-1}(u) + \varepsilon_k \leq F_\mu^{-1}(u) + \frac{1}{k}.$$

We conclude that $\lim_{n\to\infty} F_{\mu_n}^{-1}(u) = F_\mu^{-1}(u)$. ∎

Proposition 1.44 (Hoeffding-Fréchet bounds)**.** *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ and $\pi \in \Pi(\mu,\nu)$. Then*

$$\forall x, y \in \mathbb{R},\ (F_\mu(x) + F_\nu(y) - 1)^+ \leq \pi\left((-\infty, x] \times (-\infty, y]\right) \leq F_\mu(x) \wedge F_\nu(y),$$

with the upper bound attained for the comonotonous coupling $\pi = \mathcal{L}(F_\mu^{-1}(U), F_\nu^{-1}(U))$ where $U \sim \mathcal{U}[0,1]$ and the lower bound for the anti-comonotonous coupling $\pi = \mathcal{L}(F_\mu^{-1}(U), F_\nu^{-1}(1-U))$.

Proof: The function $F_\pi(x,y) = \pi\left((-\infty, x] \times (-\infty, y]\right)$ is non-decreasing in each of its variables. Therefore

$$F_\pi(x,y) \leq F_\pi(x, +\infty) \wedge F_\pi(+\infty, y) = F_\mu(x) \wedge F_\nu(y).$$

On the other hand,

$$\pi\left((x, +\infty) \times (y, +\infty)\right) = F_\pi(x,y) - F_\pi(x,+\infty) - F_\pi(+\infty, y) + F_\pi(+\infty,+\infty),$$

which implies $F_\pi(x,y) \geq F_\pi(x,+\infty) + F_\pi(+\infty,y) - F_\pi(+\infty,+\infty) = F_\mu(x) + F_\nu(y) - 1$. Therefore $F_\pi(x,y) \geq (F_\mu(x) + F_\nu(y) - 1)^+$. Finally, by Lemma 1.42, we have

$$\begin{aligned}
\mathbb{P}(F_\mu^{-1}(U) \leq x, F_\nu^{-1}(U) \leq y) &= \mathbb{P}(U \leq F_\mu(x), U \leq F_\nu(y)) \\
&= \mathbb{P}(U \leq F_\mu(x) \wedge F_\nu(y)) = F_\mu(x) \wedge F_\nu(y) \text{ and} \\
\mathbb{P}(F_\mu^{-1}(U) \leq x, F_\nu^{-1}(1-U) \leq y) &= \mathbb{P}(U \leq F_\mu(x), 1 - U \leq F_\nu(y)) \\
&= \mathbb{P}(U \leq F_\mu(x), U \geq 1 - F_\nu(y)) = (F_\mu(x) + F_\nu(y) - 1)^+.
\end{aligned}$$

∎

Proposition 1.45. *Let $\rho > 1$.*

$$\forall \mu, \nu \in \mathcal{P}_\rho(\mathbb{R}),\ \mathcal{W}_\rho(\mu,\nu) = \left(\int_{u=0}^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^\rho du\right)^{1/\rho}.$$

The proof relies on the next lemma.

Lemma 1.46. *Let $\rho > 1$, $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R})$ and $\pi \in \Pi(\mu,\nu)$. Then*

$$\int_{\mathbb{R}^2} |x-y|^\rho(\pi(dx,dy)-\mu(dx)\nu(dy)) = \rho(\rho-1)\int_{\mathbb{R}^2} |z-w|^{\rho-2}\left(F_\mu(z)F_\nu(w) - F_\pi(z,w)\right) dzdw.$$

Proof of Proposition 1.45: According to Lemma 1.46, to minimize (resp. maximize) $\int_{\mathbb{R}^2} |x-y|^\rho \pi(dx,dy)$ over $\Pi(\mu,\nu)$, it is enough to choose π which maximizes (resp. minimizes) F_π, i.e. π equal to the comonotonous coupling $\mathcal{L}(F_\mu^{-1}(U), F_\nu^{-1}(U))$ (resp. anti-comonotonous coupling $\mathcal{L}(F_\mu^{-1}(U), F_\nu^{-1}(1-U))$) by Proposition 1.44. For $\pi = \mathcal{L}(F_\mu^{-1}(U), F_\nu^{-1}(U))$, we have

$$\int_{\mathbb{R}^2} |x-y|^\rho \pi(dx,dy) = \mathbb{E}\left[|F_\mu^{-1}(U) - F_\nu^{-1}(U)|^\rho\right] = \int_{u=0}^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^\rho du.$$

■

Remark 1.47. *Since the two-dimensional cumulative distribution function F_π characterizes the coupling π, the comonotonous coupling is the unique optimal coupling.*

Proof of Lemma 1.46: The function $\mathbb{R}^2 \ni (z,w) \mapsto |z-w|^\rho$ is continuously differntiable with $\partial_z |z-w|^\rho = \rho(1_{\{z>w\}} - 1_{\{w>z\}})|z-w|^{\rho-1}$. Moreover, $\mathbb{R}^2 \ni (z,w) \mapsto \partial_z|z-w|^\rho$ is differentiable on $\mathbb{R}^2 \setminus \{(z,w) \in \mathbb{R}^d :\ z = w\}$ with $\partial^2_{zw}|z-w|^\rho = -\rho(\rho-1)|z-w|^{\rho-2}$. We deduce that

$$\forall w_1, w_2 \in \mathbb{R},\ \partial_z|z-w_2|^\rho - \partial_z|z-w_1|^\rho = -\rho(\rho-1)\int_{w=w_1}^{w_2} |z-w|^{\rho-2}dw.$$

As a consequence, for all $x, y, \tilde{x}, \tilde{y} \in \mathbb{R}$,

$$\begin{aligned}
&|\tilde{x}-\tilde{y}|^\rho + |x-y|^\rho - |\tilde{x}-y|^\rho - |x-\tilde{y}|^\rho = -\rho(\rho-1)\int_{z=x}^{\tilde{x}}\int_{w=y}^{\tilde{y}} |z-w|^{\rho-2}dwdz \\
&= -\rho(\rho-1)\int_{\mathbb{R}^2} (1_{\{x\le z\}} - 1_{\{\tilde{x}\le z\}})(1_{\{y\le w\}} - 1_{\{\tilde{y}\le w\}})|z-w|^{\rho-2}dwdz \\
&= \rho(\rho-1)\int_{\mathbb{R}^2} \left(1_{\{x\le z\}}1_{\{\tilde{y}\le w\}} + 1_{\{\tilde{x}\le z\}}1_{\{y\le w\}} - 1_{\{x\le z, y\le w\}} - 1_{\{\tilde{x}\le z,\tilde{y}\le w\}}\right)|z-w|^{\rho-2}dwdz.
\end{aligned} \tag{1.11}$$

The equality between the second and last expressions ensures that the function $\mathbb{R}^2 \ni (z,w) \mapsto \left(1_{\{x\le z\}}1_{\{\tilde{y}\le w\}} + 1_{\{\tilde{x}\le z\}}1_{\{y\le w\}} - 1_{\{x\le z, y\le w\}} - 1_{\{\tilde{x}\le z,\tilde{y}\le w\}}\right)$ has constant sign equal to that of $(x-\tilde{x})(\tilde{y}-y)$. Therefore

$$\begin{aligned}
\rho(\rho-1)&\int_{\mathbb{R}^2} \left|1_{\{x\le z\}}1_{\{\tilde{y}\le w\}} + 1_{\{\tilde{x}\le z\}}1_{\{y\le w\}} - 1_{\{x\le z, y\le w\}} - 1_{\{\tilde{x}\le z,\tilde{y}\le w\}}\right| |z-w|^{\rho-2}dwdz \\
&= \left||\tilde{x}-\tilde{y}|^\rho + |x-y|^\rho - |\tilde{x}-y|^\rho - |x-\tilde{y}|^\rho\right| \le 2^\rho\left(|x|^\rho + |y|^\rho + |\tilde{x}|^\rho + |\tilde{y}|^\rho\right),
\end{aligned}$$

where the right-hand side and thus each expression is integrable with respect to $\pi(dx,dy)\pi(d\tilde{x},d\tilde{y})$ for $\pi\in\Pi(\mu,\nu)$. This permits to use Fubini's theorem to check that the integral of the right-hand side of (1.11) is equal to

$$2\rho(\rho-1)\int_{\mathbb{R}^2}|z-w|^{\rho-2}(F_\mu(z)F_\nu(w)-F_\pi(z,w))dzdw.$$

Since the integral of the left-hand side of (1.11) is equal to $2\int_{\mathbb{R}^2}|x-y|^\rho(\pi(dx,dy)-\mu(dx)\nu(dy))$, we conclude that the statement holds. ∎

Corollary 1.48.

$$\forall\mu,\nu\in\mathcal{P}_1(\mathbb{R}),\ \mathcal{W}_1(\mu,\nu)=\int_{u=0}^1|F_\mu^{-1}(u)-F_\nu^{-1}(u)|du=\int_{x\in\mathbb{R}}|F_\mu(x)-F_\nu(x)|dx.$$

Remark 1.49. *The comonotonous coupling is optimal for $\mathcal{W}_1$ but there may also exist other optimal couplings. For instance, if there exists $a\in\mathbb{R}$ such that $F_\mu(a)=1$ and $F_\nu(a)=0$, then for each $\pi\in\Pi(\mu,\nu)$,*

$$\pi\left((a,+\infty)\times\mathbb{R}\right)=\mu\left((a,+\infty)\right)=1-F_\mu(a)=0=F_\nu(a)=\pi\left(\mathbb{R}\times(-\infty,a]\right).$$

As a consequence, $\pi\left((-\infty,a]\times[a,+\infty)\right)=1$ so that

$$\int_{\mathbb{R}^2}|x-y|\pi(dx,dy)=\int_{\mathbb{R}^2}(x-y)\pi(dx,dy)=\int_{\mathbb{R}}x\mu(dx)-\int_{\mathbb{R}}y\nu(dy)$$

does not depend on $\pi\in\Pi(\mu,\nu)$.

Proof: By Lemma 1.42, $F_\mu^{-1}(u)\le x<F_\nu^{-1}(u)\Leftrightarrow F_\nu(x)<u\le F_\mu(x)$ and $F_\nu^{-1}(u)\le x<F_\mu^{-1}(u)\Leftrightarrow F_\mu(x)<u\le F_\nu(x)$. With Fubini's theorem, we deduce that

$$\begin{aligned}\int_{u=0}^1|F_\mu^{-1}(u)-F_\nu^{-1}(u)|du&=\int_{u=0}^1\int_{x\in\mathbb{R}}1_{\{F_\mu^1(u)\le x<F_\nu^{-1}(u)\}}+1_{\{F_\mu^1(u)\le x<F_\mu^{-1}(u)\}}dxdu\\&=\int_{x\in\mathbb{R}}\int_{u=0}^1 1_{\{F_\nu(x)<u\le F_\mu(x)\}}+1_{\{F_\mu(x)<u\le F_\nu(x)\}}dudx\\&=\int_{x\in\mathbb{R}}|F_\mu(x)-F_\nu(x)|dx.\end{aligned}$$

For $n\in\mathbb{N}$, let μ_n and ν_n denote the respective images of μ and ν by $\mathbb{R}\ni x\mapsto(-n)\vee x\wedge n$. We have $F_{\mu_n}^{-1}(u)=(-n)\vee F_\mu^{-1}(u)\wedge n$ and $F_{\nu_n}^{-1}(u)=(-n)\vee F_\nu^{-1}(u)\wedge n$ so that, by monotone convergence, $\int_{u=0}^1|F_{\mu_n}^{-1}(u)-F_{\nu_n}^{-1}(u)|du$ converges to $\int_{u=0}^1|F_\mu^{-1}(u)-F_\nu^{-1}(u)|du$ as $n\to\infty$. We have $\mathcal{W}_1(\mu_n,\mu)\le\int_{u=0}^1|F_{\mu_n}^{-1}(u)-F_\mu^{-1}(u)|du=\int_{u=0}^1 1_{\{|F_\mu^{-1}(u)|>n\}}(|F_\mu^{-1}(u)|-n)du$, where the right-hand side goes to 0 when $n\to\infty$ by Lebesgue's theorem. In the same way $\lim_{n\to\infty}\mathcal{W}_1(\nu_n,\nu)=0$. By the triangle inequality, we deduce that

$$|\mathcal{W}_1(\mu,\nu)-\mathcal{W}_1(\mu_n,\nu_n)|\le\mathcal{W}_1(\mu_n,\mu)+\mathcal{W}_1(\nu_n,\nu)\overset{n\to\infty}{\longrightarrow}0.$$

Therefore, to prove that $\mathcal{W}_1(\mu,\nu)=\int_{u=0}^1|F_\mu^{-1}(u)-F_\nu^{-1}(u)|du$, it is enough to check that the same equality holds with (μ,ν) replaced by (μ_n,ν_n). Let $\pi_n\in\Pi(\mu_n,\nu_n)$ be optimal for $\mathcal{W}_1(\mu_n,\nu_n)$. By Proposition 1.45,

$$\forall\rho>1,\ \int_{\mathbb{R}^2}|x-y|^\rho\pi_n(dx,dy)\ge\int_{u=0}^1|F_{\mu_n}^{-1}(u)-F_{\nu_n}^{-1}(u)|^\rho du.$$

The fact that $|F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)| \leq 2n$ and $|x-y| \leq 2n$, $\pi_n(dx,dy)$ a.e., permits to apply Lebesgue's theorem to take the limit $\rho \to 1$ in the inequality to conclude that

$$\mathcal{W}_1(\mu_n,\nu_n) = \int_{\mathbb{R}^2} |x-y|\pi_n(dx,dy) \geq \int_{u=0}^1 |F_{\mu_n}^{-1}(u) - F_{\nu_n}^{-1}(u)|du \geq \mathcal{W}_1(\mu_n,\nu_n).$$

■

1.7.3 Quadratic Wasserstein distance between Gaussian distributions

Let $\mathcal{S}_+(d)$ denote the set of symmetric positive semi-definite $d \times d$ matrices i.e. the set of covariance matrices of d-dimensional square integrable random vectors. For $\Sigma \in \mathcal{S}_+(d)$ we denote by $\Sigma^{1/2}$ the only element of $\mathcal{S}_+(d)$ such that $\Sigma^{1/2}\Sigma^{1/2} = \Sigma$, i.e. the symmetric square root of Σ. When Σ is non singular, we also denote by $\Sigma^{-1/2}$ the inverse of $\Sigma^{1/2}$. For $m \in \mathbb{R}^d$, $\Sigma \in \mathcal{S}_+(d)$ let finally $\mathcal{N}_d(\mu,\Sigma)$ denote the d-dimensional Gaussian distribution with expectation m and covariance matrix Σ.

Proposition 1.50.

$$\mathcal{W}_2^2(\mathcal{N}_d(m_\mu,\Sigma_\mu),\mathcal{N}_d(m_\nu,\Sigma_\nu)) = |m_\mu - m_\nu|^2 + \mathrm{tr}\left(\Sigma_\mu + \Sigma_\nu - 2(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\right).$$

Remark 1.51. *Since Σ_μ and Σ_ν play symmetric roles, we have* $\mathrm{tr}\left((\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\right) = \mathrm{tr}\left((\Sigma_\nu^{1/2}\Sigma_\mu\Sigma_\nu^{1/2})^{1/2}\right)$.

Let us denote by $m_\eta \in \mathbb{R}^d$ and $\Sigma_\eta \in \mathcal{S}_+(d)$ the expectation and the covariance matrix of $\eta \in \mathcal{P}_2(\mathbb{R}^d)$. For $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\pi \in \Pi(\mu,\nu) \subset \mathcal{P}_2(\mathbb{R}^{2d})$, we have $m_\pi = \begin{pmatrix} m_\mu \\ m_\nu \end{pmatrix}$ and $\Sigma_\pi = \begin{pmatrix} \Sigma_\mu & \Theta_\pi \\ \Theta_\pi^T & \Sigma_\nu \end{pmatrix}$ for some $\Theta_\pi \in \mathbb{R}^{d\times d}$ such that $\Sigma_\pi \in \mathcal{S}_+(2d)$. Moreover, by bias variance decomposition,

$$\begin{aligned}\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2\pi(dx,dy) =& |m_\mu - m_\nu|^2 \\ &+ \int_{\mathbb{R}^d\times\mathbb{R}^d} \left(|x-m_\mu|^2 + |y-m_\nu|^2 - 2(x-m_\mu).(y-m_\nu)\right)\pi(dx,dy) \\ =& |m_\mu - m_\nu|^2 + \mathrm{tr}(\Sigma_\mu + \Sigma_\nu - 2\Theta_\pi).\end{aligned}$$

Therefore

$$\mathcal{W}_2^2(\mu,\nu) = |m_\mu - m_\nu|^2 + \mathrm{tr}(\Sigma_\mu) + \mathrm{tr}(\Sigma_\nu) - 2\sup_{\pi\in\Pi(\mu,\nu)} \mathrm{tr}(\Theta_\pi).$$

The specificity of the Gaussian case $\mu = \mathcal{N}_d(m_\mu,\Sigma_\mu)$ and $\nu = \mathcal{N}_d(m_\nu,\Sigma_\nu)$ is that any $\Theta \in \mathbb{R}^{d\times d}$ such that $\begin{pmatrix} \Sigma_\mu & \Theta \\ \Theta^T & \Sigma_\nu \end{pmatrix} \in \mathcal{S}_+(2d)$ is attainable by the Gaussian coupling $\mathcal{N}_{2d}\left(\begin{pmatrix} m_\mu \\ m_\nu \end{pmatrix}, \begin{pmatrix} \Sigma_\mu & \Theta \\ \Theta^T & \Sigma_\nu \end{pmatrix}\right)$. As a consequence, to explicit $\mathcal{W}_2^2(\mathcal{N}_d(m_\mu,\Sigma_\mu),\mathcal{N}_d(m_\nu,\Sigma_\nu))$, it is enough to maximize $\mathrm{tr}(\Theta)$ over such matrices Θ.

This was done in [14, 9, 10].

To prove Proposition 1.50, we are rather going to use the approach with more probabilistic insight about correlation matrices introduced in [3]. Let $\mathcal{C}(d) = \{C \in \mathcal{S}_+(d) : \forall i \in \{1, \cdots, d\}, C_{ii} = 1\}$ denote the set of correlation matrices of d-dimensional square integrable random vectors.
We say that the correlation matrix C is associated with the covariance matrix Σ if $\Sigma_{ij} = \sqrt{\Sigma_{ii}\Sigma_{jj}}C_{ij}$ for all $i, j \in \{1, \cdots d\}$ i.e.

$$\Sigma = \mathrm{dg}(\Sigma)^{1/2} C \mathrm{dg}(\Sigma)^{1/2}$$

where $\mathrm{dg}(\Sigma)$ denotes the diagonal matrix with diagonal entries equal to those of Σ.
Note that denoting by D the diagonal matrix with diagonal entries $D_{ii} = 1_{\{\Sigma_{ii}>0\}}\Sigma_{ii}^{-1/2}$ for $i \in \{1, \cdots, d\}$, the correlation matrix $D\Sigma D$ is associated with Σ and, when Σ is non singular so that $D = \mathrm{dg}(\Sigma)^{-1/2}$, this is the only correlation matrix associated with Σ. Let $\mathcal{O}(d)$ denote the set of $d \times d$ orthogonal matrices i.e. $\mathcal{O}(d) = \{U \in \mathbb{R}^{d\times d} : UU^T = I_d\}$. The proof of Proposition 1.50 relies on the next lemma.

Lemma 1.52. *For $\Sigma_\mu, \Sigma_\nu \in \mathcal{S}_+(d)$, there exists $U \in \mathcal{O}(d)$ such that $U\Sigma_\mu U^T$ and $U\Sigma_\nu U^T$ share the same correlation matrix C.*

Proof of Proposition 1.50: According to the above discussion, it is enough to check that $V(\Sigma_\mu, \Sigma_\nu)$ defined by

$$V(\Sigma_\mu, \Sigma_\nu) = \sup_{\Theta\in\mathbb{R}^{d\times d}: \begin{pmatrix} \Sigma_\mu & \Theta \\ \Theta^T & \Sigma_\nu \end{pmatrix} \in\mathcal{S}_+(2d)} \mathrm{tr}(\Theta)$$

is equal to $\mathrm{tr}\left((\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\right)$.
For $U \in \mathcal{O}(d)$, since $\begin{pmatrix} U\Sigma_\mu U^T & U\Theta U^T \\ U\Theta^T U^T & U\Sigma_\nu U^T \end{pmatrix} = \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix}\begin{pmatrix} \Sigma_\mu & \Theta \\ \Theta^T & \Sigma_\nu \end{pmatrix}\begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix}^T$,

$$\begin{pmatrix} \Sigma_\mu & \Theta \\ \Theta^T & \Sigma_\nu \end{pmatrix} \in \mathcal{S}_+(2d) \Leftrightarrow \begin{pmatrix} U\Sigma_\mu U^T & U\Theta U^T \\ U\Theta^T U^T & U\Sigma_\nu U^T \end{pmatrix} \in \mathcal{S}_+(2d),$$

and, by the cyclicity of the trace, $\mathrm{tr}(U\Theta U^T) = \mathrm{tr}(U^T U\Theta) = \mathrm{tr}(\Theta)$. Therefore

$$\forall U \in \mathcal{O}(d),\ V(\Sigma_\mu, \Sigma_\nu) = V(U\Sigma_\mu U^T, U\Sigma_\nu U^T). \tag{1.12}$$

We now use Lemma 1.52 to choose $U \in \mathcal{O}(d)$ such that $\widetilde{\Sigma}_\mu := U\Sigma_\mu U^T$ and $\widetilde{\Sigma}_\nu := U\Sigma_\nu U^T$ share the correlation matrix C i.e.

$$\widetilde{\Sigma}_\mu = \mathrm{dg}(\widetilde{\Sigma}_\mu)^{1/2} C \mathrm{dg}(\widetilde{\Sigma}_\mu)^{1/2} \text{ and } \widetilde{\Sigma}_\nu = \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} C \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2}. \tag{1.13}$$

For $\Theta \in \mathbb{R}^{d\times d}$ such that $\begin{pmatrix} \widetilde{\Sigma}_\mu & \Theta \\ \Theta^T & \widetilde{\Sigma}_\nu \end{pmatrix} \in \mathcal{S}_+(2d)$ and $i, j \in \{1, \cdots, d\}$, we have $\begin{pmatrix} (\widetilde{\Sigma}_\mu)_{ii} & \Theta_{ii} \\ \Theta_{ii} & (\widetilde{\Sigma}_\nu)_{ii} \end{pmatrix} \in \mathcal{S}_+(2)$ and therefore $\Theta_{ii}^2 \leq (\widetilde{\Sigma}_\mu)_{ii}(\widetilde{\Sigma}_\nu)_{ii}$. We deduce that

$$V\left(\widetilde{\Sigma}_\mu, \widetilde{\Sigma}_\nu\right) \leq \sum_{i=1}^d \sqrt{(\widetilde{\Sigma}_\mu)_{ii}(\widetilde{\Sigma}_\nu)_{ii}}.$$

This inequality turns out to be an equality since $\Theta = \mathrm{dg}(\widetilde{\Sigma}_\mu)^{1/2} C \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2}$ attains the upper-bound and is such that $\begin{pmatrix} \widetilde{\Sigma}_\mu & \Theta \\ \Theta^T & \widetilde{\Sigma}_\nu \end{pmatrix} = \left(\mathrm{dg}(\widetilde{\Sigma}_\mu)^{1/2}, \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2}\right) C \begin{pmatrix} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{1/2} \\ \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \end{pmatrix} \in \mathcal{S}_+(2d)$.

When Σ_μ is non singular, the matrix $\widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \widetilde{\Sigma}_\mu^{1/2}$ is equal to $\widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \widetilde{\Sigma}_\mu^{1/2}$ and therefore symmetric. Moreover, using (1.13) for the second and third equalities, we have

$$\begin{aligned}
&\widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \widetilde{\Sigma}_\mu^{1/2} \\
&= \widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \widetilde{\Sigma}_\mu \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \widetilde{\Sigma}_\mu^{1/2} = \widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} C \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \widetilde{\Sigma}_\mu^{1/2} \\
&= \widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2}.
\end{aligned}$$

Therefore, when Σ_μ is non singular, $(\widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2})^{1/2} = \widetilde{\Sigma}_\mu^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu^{-1/2}) \widetilde{\Sigma}_\mu^{1/2}$ so that, by cyclicity of the trace,

$$\mathrm{tr}\left((\widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2})^{1/2}\right) = \mathrm{tr}\left(\mathrm{dg}(\widetilde{\Sigma}_\nu)^{1/2} \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2} \widetilde{\Sigma}_\mu\right) = \sum_{i=1}^d \sqrt{(\widetilde{\Sigma}_\mu)_{ii} (\widetilde{\Sigma}_\nu)_{ii}}.$$

Replacing (Σ_μ, Σ_ν) by

$$\left((\mathrm{dg}(\Sigma_\mu)^{1/2} + \varepsilon I_d)((1-\varepsilon)C + \varepsilon I_d)(\mathrm{dg}(\Sigma_\mu)^{1/2} + \varepsilon I_d), \mathrm{dg}(\Sigma_\nu)^{1/2}((1-\varepsilon)C + \varepsilon I_d)\mathrm{dg}(\Sigma_\nu)^{1/2}\right)$$

with $\varepsilon > 0$ in this equality and using the continuity of the symmetric square root and of the trace to take the limit $\varepsilon \to 0$, we conclude with (1.12) that even when Σ_μ is singular,

$$\mathrm{tr}\left((\widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2})^{1/2}\right) = \sum_{i=1}^d \sqrt{(\widetilde{\Sigma}_\mu)_{ii} (\widetilde{\Sigma}_\nu)_{ii}} = V\left(\widetilde{\Sigma}_\mu, \widetilde{\Sigma}_\nu\right) = V\left(\Sigma_\mu, \Sigma_\nu\right).$$

For $\Sigma \in \mathcal{S}_+(d)$, $(U\Sigma^{1/2}U^T)(U\Sigma^{1/2}U^T) = U\Sigma U^T$ so that $(U\Sigma U^T)^{1/2} = U\Sigma^{1/2}U^T$. Therefore,

$$\left(\widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2}\right)^{1/2} = \left(U\Sigma_\mu^{1/2}U^T U\Sigma_\nu U^T U\Sigma_\mu^{1/2}U^T\right)^{1/2} = U\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2} U^T,$$

so that, by the cyclicity of the trace,

$$\mathrm{tr}\left((\widetilde{\Sigma}_\mu^{1/2} \widetilde{\Sigma}_\nu \widetilde{\Sigma}_\mu^{1/2})^{1/2}\right) = \mathrm{tr}\left(U(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}U^T\right) = \mathrm{tr}\left((\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\right).$$

∎

Proof of Lemma 1.52: Let us first suppose that Σ_μ is non singular. We obtain U by diagonalization of the matrix $\Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}$ which belongs to $\mathcal{S}_+(d)$, i.e. we choose $U \in \mathcal{O}(d)$ such that $U\Sigma_\mu^{-1/2}(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2})^{1/2}\Sigma_\mu^{-1/2}U^T = D$ for some diagonal matrix D with positive diagonal coefficients.
Note that for $\Sigma \in \mathcal{S}_+(d)$, $(U\Sigma^{1/2}U^T)(U\Sigma^{1/2}U^T) = U\Sigma U^T$ so that $(U\Sigma U^T)^{1/2} = U\Sigma^{1/2}U^T$

and, in the same way, $(U\Sigma U^T)^{-1/2} = U\Sigma^{-1/2}U^T$ when Σ is moreover invertible. Setting $\widetilde{\Sigma}_\mu = U\Sigma_\mu U^T$ and $\widetilde{\Sigma}_\nu = U\Sigma_\nu U^T$, we deduce that

$$\begin{aligned}
\left(\widetilde{\Sigma}_\mu^{1/2}\widetilde{\Sigma}_\nu\widetilde{\Sigma}_\mu^{1/2}\right)^{1/2} &= \left(U\Sigma_\mu^{1/2}U^TU\Sigma_\nu U^TU\Sigma_\mu^{1/2}U^T\right)^{1/2} = U\left(\Sigma_\mu^{1/2}\Sigma_\nu\Sigma_\mu^{1/2}\right)^{1/2}U^T \\
&= U\Sigma_\mu^{1/2}U^TDU\Sigma_\mu^{1/2}U^T = \widetilde{\Sigma}_\mu^{1/2}D\widetilde{\Sigma}_\mu^{1/2} \\
\widetilde{\Sigma}_\mu^{1/2}\widetilde{\Sigma}_\nu\widetilde{\Sigma}_\mu^{1/2} &= \left(\widetilde{\Sigma}_\mu^{1/2}\widetilde{\Sigma}_\nu\widetilde{\Sigma}_\mu^{1/2}\right)^{1/2}\left(\widetilde{\Sigma}_\mu^{1/2}\widetilde{\Sigma}_\nu\widetilde{\Sigma}_\mu^{1/2}\right)^{1/2} = \widetilde{\Sigma}_\mu^{1/2}D\widetilde{\Sigma}_\mu^{1/2}\widetilde{\Sigma}_\mu^{1/2}D\widetilde{\Sigma}_\mu^{1/2}.
\end{aligned}$$

By multiplying the last equality to the left and to the right by $\widetilde{\Sigma}_\mu^{-1/2}$, we conclude that $\widetilde{\Sigma}_\nu = D\widetilde{\Sigma}_\mu D$. Therefore $\widetilde{\Sigma}_\nu$ and $\widetilde{\Sigma}_\mu$ share the correlation matrix $C = \mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2}\widetilde{\Sigma}_\mu\mathrm{dg}(\widetilde{\Sigma}_\mu)^{-1/2}$ where $\mathrm{dg}(\widetilde{\Sigma}_\mu)$ denotes the diagonal matrix with diagonal entries equal to those of $\widetilde{\Sigma}_\mu$. When Σ_μ is singular, we choose a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ of positive numbers converging to 0 as $n \to \infty$. By the previous case, for each $n \in \mathbb{N}$, there exists $(U_n, C_n) \in \mathcal{O}(d) \times \mathcal{C}(d)$ such that

$$\begin{aligned}
U_n(\Sigma_\mu + \varepsilon_n I_d)U_n^T &= \mathrm{dg}\left(U_n(\Sigma_\mu + \varepsilon_n I_d)U_n^T\right)^{1/2}C_n\mathrm{dg}\left(U_n(\Sigma_\mu + \varepsilon_n I_d)U_n^T\right)^{1/2} \\
\text{and } U_n\Sigma_\nu U_n^T &= \mathrm{dg}\left(U_n\Sigma_\nu U_n^T\right)^{1/2}C_n\mathrm{dg}\left(U_n\Sigma_\nu U_n^T\right)^{1/2}.
\end{aligned}$$

By compactness of the sets $\mathcal{O}(d)$ and $\mathcal{C}(d)$, we can extract from $(U_n, C_n)_{n\in\mathbb{N}}$ a subsequence converging to $(U, C) \in \mathcal{O}(d) \times \mathcal{C}(d)$ which does the job by taking the limit in the two previous equalities. ∎

Chapter 2

Weak optimal transport

Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces with respective metrics $d_\mathcal{X}$ and $d_\mathcal{Y}$ and $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$. For $\pi(dx, dy) = \mu(dx)\pi_x(dy) \in \Pi(\mu, \nu)$ and $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ bounded from below, we have

$$\pi(c) = \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\pi(dx, dy) = \int_\mathcal{X} \left(\int_\mathcal{Y} c(x, y)\pi_x(dy) \right) \mu(dx) = \int_\mathcal{X} C(x, \pi_x)\mu(dx),$$

where $C(x, p) = \int_\mathcal{Y} c(x, y)p(dy)$ for $(x, p) \in \mathcal{X} \times \mathcal{P}(\mathcal{Y})$. Weak Optimal transport is a generalization of Optimal Transport concerned with cost functions $C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$:

$$V_C(\mu, \nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_\mathcal{X} C(x, \pi_x)\mu(dx).$$

The Martingale Optimal Transport problem considered in Chapter 3 and the Entropic Optimal Transport problem considered in Chapter 4 are particular cases of WOT. The MOT problem is restricted to the case $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and relies on the next definition.

Definition 2.1.
- *Let $\bar{p} = \int_{\mathbb{R}^d} xp(dx)$ denote the mean of a probability measure $p \in \mathcal{P}_1(\mathbb{R}^d)$.*
- *For $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$, a coupling $\pi \in \Pi(\mu, \nu)$ is called a martingale coupling if $\mu(dx)$ a.e., $\bar{\pi}_x = x$.*
- *The set of martingale couplings between μ and ν is denoted by $\Pi_M(\mu, \nu)$.*

Since $\bar{\nu} = \int_\mathcal{X} \bar{\pi}_x\mu(dx)$, a necessary condition for $\Pi_M(\mu, \nu) \neq \emptyset$ is $\bar{\nu} = \bar{\mu}$. Strassen's theorem (see Theorem 2.16 below) gives a necessary and sufficient condition for $\Pi_M(\mu, \nu) \neq \emptyset$. For $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, measurable the value function of the martingale optimal transport problem is

$$V_c^M(\mu, \nu) = \inf_{\pi\in\Pi_M(\mu,\nu)} \pi(c).$$

Let $C : \mathbb{R} \times \mathcal{P}_1(\mathbb{R}^d)$ be defined by

$$C(x, p) = \begin{cases} \int_{\mathbb{R}^d} c(x, y)p(dy) \text{ if } \bar{p} = x \\ +\infty \text{ otherwise} \end{cases}.$$

When c is bounded from below, so that the use of Fubini's theorem is justified, we have $V_c^M(\mu, \nu) = V_C(\mu, \nu)$, which shows that MOT is a particular case of WOT. Let c be lower semi-continuous and bounded from below. Then C is bounded from below by the same constant. Moreover, for $x \in \mathbb{R}^d$, $p, q \in \mathcal{P}_1(\mathbb{R}^d)$ and $\alpha \in (0,1)$, $C(x, \alpha p + (1-\alpha q)$ is equal to

- $+\infty$ if $\overline{\alpha p + (1-\alpha q)} \neq x$ and, then, $\bar{p} \neq x$ or $\bar{q} \neq x$, so that $\alpha C(x,p) + (1-\alpha)C(x,q) = +\infty$,
- $\int_{\mathbb{R}^d} c(x,y)(\alpha p + (1-\alpha q))(dy) \leq \alpha C(x,p) + (1-\alpha)C(x,q)$ otherwise.

We deduce that C is convex in its measure (second) argument. Since $\{(x,p) \in \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) : \bar{p} \neq x\}$ is open, it is enough to check that $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \ni (x,p) \mapsto \int_{\mathbb{R}^d} c(x,y)p(dy)$ is lower semi-continuous to conclude that so is C. The case when c is constant equal to $+\infty$ is clear. Otherwise, by Lemma 1.13, c is the non-decreasing limit of n-Lipschitz functions c_n bounded from below by the same constant as c. Using the inequality (2.3) below with $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and $f = c_n$, we obtain that $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \ni (x,p) \mapsto \int_{\mathbb{R}^d} c_n(x,y)p(dy)$ is continuous. By monotone convergence, $\mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \ni (x,p) \mapsto \int_{\mathbb{R}^d} c(x,y)p(dy)$ is the supremum of these continuous functions and it is thus lower semi-continuous.

The EOT problem writes

$$V_{c,\varepsilon}(\mu,\nu) = \inf_{\pi \in \Pi(\mu, nu)} \left(\pi(c) + \varepsilon H(\pi|\mu \otimes \nu) \right),$$

where $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a cost function, $\varepsilon > 0$, $\mu \otimes \nu(dx,dy) = \mu(dx)\nu(dy)$ and, for tow probability measure η, γ on the the same measurable set $(\mathcal{Z}, \mathcal{B}(\mathcal{Z}))$,

$$H(\eta|\gamma) = \begin{cases} \int_{\mathcal{Z}} \ln\left(\frac{d\eta}{d\gamma}(z)\right) \eta(dz) \text{ if } \eta << \gamma \\ +\infty \text{ otherwise} \end{cases}.$$

If $\pi \in \Pi(\mu,\nu)$ satisfies $\pi << \mu \otimes \nu$ then writing $\pi(dx,dy) = \mu(dx)\pi_x(dy)$, we have $\frac{d\pi}{d\mu \otimes \nu}(x,y) = \frac{d\pi_x}{d\nu}(y)$ so that

$$H(\pi|\mu \otimes \nu) = \int_{\mathcal{X} \times \mathcal{Y}} \ln\left(\frac{d\pi_x}{d\nu}(y)\right) \pi_x(dy)\mu(dx) = \int_{\mathcal{X}} H(\pi_x|\nu)\mu(dx).$$

Therefore $V_{c,\varepsilon}(\mu,\nu) = V_C(\mu,\nu)$ with

$$C(x,p) = \int_{\mathcal{Y}} c(x,y)p(dy) + \varepsilon H(p|\nu).$$

The function $\mathcal{P}(\mathcal{Y}) \times \mathcal{P}(\mathcal{Y}) \in (p,\nu) \mapsto H(p|\nu)$ is non-negative by (4.1) and jointly lower-semi continuous and convex according to Lemma 4.2.

2.1 Properties of Weak Optimal Transport

Let $\rho \geq 1$, $C : \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous bounded from below and convex in the measure (i.e. second) argument. We set

$$\forall (\mu,\nu) \in \mathcal{P}_\rho(\mathcal{X}) \times \mathcal{P}_\rho(\mathcal{Y}),\ V_C(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} C(x,\pi_x)\mu(dx).$$

The convexity in the measure argument is needed because of the lack of continuity of $\mathcal{P}(\mathcal{X} \times \mathcal{Y}) \ni \pi(dx,dy) = \mu(dx)\pi_x(dy) \mapsto (\pi_x)_{x \in \mathcal{X}}$. We want to prove the following main results.

Theorem 2.2. *Let $\rho \geq 1$, $C : \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous bounded from below and convex in the measure argument. Then, for all $(\mu,\nu) \in \mathcal{P}_\rho(\mathcal{X}) \times \mathcal{P}_\rho(\mathcal{Y})$, there exists $\pi^\star \in \Pi(\mu,\nu)$ such that $V_C(\mu,\nu) = \int_{\mathcal{X}} C(x,\pi^\star_x)\mu(dx)$. Moreover, the function V_C is lower semi-continuous on $\mathcal{P}_\rho(\mathcal{X}) \times \mathcal{P}_\rho(\mathcal{Y})$ and convex in its second argument.*

Definition 2.3.
- *Let $\mathcal{M}_\rho(\mathcal{Y})$ denote the set of functions $f : \mathcal{Y} \to \mathbb{R}$ measurable and such that $\sup_{y \in \mathcal{Y}} \frac{|f(y)|}{1+d^\rho_{\mathcal{Y}}(y_0,y)} < \infty$ where y_0 is any element of $\mathcal{Y}$. Let $C_\rho(\mathcal{Y})$ denote the subset of $\mathcal{M}_\rho(\mathcal{Y})$ which consists in continuous functions.*
- *For $C : \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ and $\psi \in \mathcal{M}_\rho(\mathcal{Y})$, we define $\psi^C : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ by $\psi^C(x) = \inf_{p \in \mathcal{P}_\rho(\mathcal{Y})}\{C(x,p) - p(\psi)\}$.*
- *Let $C^\wedge_\rho(\mathcal{Y}) = \{\psi \in C_\rho(\mathcal{Y}) : \psi$ is bounded from above$\}$.*

Note that when C is bounded from below and $\psi \in C^\wedge_\rho(\mathcal{Y})$, then ψ^C is bounded from below and therefore semi-integrable with respect to $\mu \in \mathcal{P}(\mathcal{X})$. The following dual formulation is the second main result.

Theorem 2.4. *Let $\rho \geq 1$, $C : \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous bounded from below and convex in the measure argument. Then for $(\mu,\nu) \in \mathcal{P}_\rho(\mathcal{X}) \times \mathcal{P}_\rho(\mathcal{Y})$,*

$$V_C(\mu,\nu) = \sup_{\psi \in C^\wedge_\rho(\mathcal{Y})} \left\{\mu(\psi^C) + \nu(\psi)\right\} = \sup_{\substack{(\phi,\psi) \in C_1(\mathcal{X}) \times C^\wedge_\rho(\mathcal{Y}) \\ \phi \oplus \psi(\cdot) \leq C}} \{\mu(\phi) + \nu(\psi)\},$$

where $\phi \oplus \psi(\cdot) \leq C$ stands for $\forall (x,p) \in \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})$, $\phi(x) + p(\psi) \leq C(x,p)$.

Remark 2.5. *In the proof of Theorem 2.4, we will check that when C is moreover Lipschitz continuous, then for $\psi \in C_\rho(\mathcal{Y})$, ψ^C is either constant and equal to $-\infty$ or $\mathbb{R}$-valued and Lipschitz continuous so that $\mu(\psi^C)$ makes and*

$$V_C(\mu,\nu) = \sup_{\psi \in C_\rho(\mathcal{Y})} \left\{\mu(\psi^C) + \nu(\psi)\right\}.$$

The proof of Theorem 2.2 relies on the next proposition.

Proposition 2.6. *Let $\rho \geq 1$, $C : \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous bounded from below and convex in the measure argument. Then $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y}) \ni \pi \mapsto \int_{\mathcal{X}} C(x,\pi_x)\pi(dx \times \mathcal{P}_\rho(\mathcal{Y}))$ is lower semi-continuous.*

To prove the proposition, let us introduce

- $J : \mathcal{P}(\mathcal{X} \times \mathcal{Y}) \to \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ such that for $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, $J(\pi)$ is the image of π by $\mathcal{X} \times \mathcal{Y} \ni (x,y) \mapsto (x,\pi_x)$. Note that $J(\pi)$ also is the image of the first marginal $\pi(dx \times \mathcal{Y})$ of π by $\mathcal{X} \ni x \mapsto (x,\pi_x)$.
- The intensity $I : \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ defined for $P(dx,dp) = P(dx \times \mathcal{P}(\mathcal{Y}))P_x(dp) \in \mathcal{P}(\mathcal{X} \times \mathcal{P}(\mathcal{Y}))$ by

$$\begin{aligned} \forall f \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y}),\ I(P)(f) &= \int_{\mathcal{X} \times \mathcal{P}(\mathcal{Y}) \times \mathcal{Y}} f(x,y)p(dy)P(dx,dp) \\ &= \int_{\mathcal{X} \times \mathcal{Y}} f(x,y)\left(\int_{\mathcal{P}(Y)} pP_x(dp)\right)(dy)P(dx \times \mathcal{P}(\mathcal{Y})). \end{aligned} \tag{2.1}$$

Since, for $f \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y})$,

$$I(J(\pi))(f) = \int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\pi_x(dy)\pi(dx \times \mathcal{P}(Y)) = \int_{\mathcal{X}\times\mathcal{Y}} f(x,y)\pi(dx,dy) = \pi(f),$$

we have

$$\forall \pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y}),\ I(J(\pi)) = \pi.$$

Lemma 2.7. *The intensity $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})) \ni P \mapsto I(P) \in \mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$ is continuous.*

Proof: Let $x_0 \in \mathcal{X}$, $y_0 \in \mathcal{Y}$. Since for $p \in \mathcal{P}_\rho(\mathcal{Y})$, $\mathcal{W}_\rho^\rho(\delta_{y_0}, p) = \int_\mathcal{Y} d_\mathcal{Y}^\rho(y_0,y)p(dy)$, we have

$$\begin{aligned}\forall P \in \mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})),\ &\int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} \left(d_\mathcal{X}^\rho(x_0,x) + \mathcal{W}_\rho^\rho(\delta_{y_0},p)\right) P(dx,dp) \\ &= \int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})\times\mathcal{Y}} \left(d_\mathcal{X}^\rho(x_0,x) + d_\mathcal{Y}^\rho(y_0,y)\right) p(dy)P(dx,dp) \\ &= \int_{\mathcal{X}\times\mathcal{Y}} \left(d_\mathcal{X}^\rho(x_0,x) + d_\mathcal{Y}^\rho(y_0,y)\right) I(P)(dx,dy). \qquad (2.2)\end{aligned}$$

Hence convergence of the ρ-th order moment of P_n to that of P_∞ is equivalent to convergence of the ρ-th order moment of $I(P_n)$ to that of $I(P_\infty)$. According to the Portmanteau theorem (see Theorem 5.2 2)), the weak convergence is characterized by the convergence of integrals for bounded and Lipschitz continuous test functions. Let $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ be bounded and Lipschitz continuous. For $x, \tilde{x} \in \mathcal{X}$ and $p, \tilde{p} \in \mathcal{P}_\rho(\mathcal{Y})$, we have, using Proposition 1.37 for the second inequality and (1.8) for the third

$$\begin{aligned}\left|\int_\mathcal{Y} f(x,y)p(dy) - \int_\mathcal{Y} f(\tilde{x},y)\tilde{p}(dy)\right| \le& \int_\mathcal{Y} |f(x,y) - f(\tilde{x},y)|p(dy) \\ &+ \left|\int_\mathcal{Y} f(\tilde{x},y)p(dy) - \int_\mathcal{Y} f(\tilde{x},y)\tilde{p}(dy)\right| \\ \le& \mathrm{Lip}(f)\left(|x-\tilde{x}| + \mathcal{W}_1(p,\tilde{p})\right) \qquad (2.3) \\ \le& \mathrm{Lip}(f)\left(|x-\tilde{x}| + \mathcal{W}_\rho(p,\tilde{p})\right).\end{aligned}$$

As a consequence, $\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y}) \ni (x,p) \mapsto \int_\mathcal{Y} f(x,y)p(dy)$ is bounded and Lipschitz continuous with constant Lip(f). Since $I(P)(f) = \int_{\mathcal{X}\times\mathcal{P}(\mathcal{Y})} \left(\int_\mathcal{Y} f(x,y)p(dy)\right) P(dx,dp)$, we conclude that the weak convergence of P_n to P_∞ implies that of $I(P_n)$ to $I(P_\infty)$. ∎

Remark 2.8. *Replacing $d_\mathcal{X}$ and $d_\mathcal{Y}$ by $d_\mathcal{X}\wedge 1$ and $d_\mathcal{Y}\wedge 1$ in this proof ensures that $I : \mathcal{P}(\mathcal{X}\times\mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is continuous (for the weak convergence topology).*

Lemma 2.9. *A subset $\mathcal{E}$ of $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y}))$ is relatively compact iff its image $I(\mathcal{E})$ by I is relatively compact in $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$.*

The proof relies on the next lemma.

Lemma 2.10. *Let $\mathcal{Y}$ be a Polish space $\hat{\mathcal{E}} \subset \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ and $\hat{I} : \mathcal{P}(\mathcal{P}(\mathcal{Y})) \to \mathcal{P}(\mathcal{Y})$ be defined by $\hat{I}(Q) = \int_{\mathcal{P}(\mathcal{Y})} pQ(dp)$. Then, for the weak convergence topology,*

$$\hat{\mathcal{E}} \text{ tight} \Leftrightarrow \hat{\mathcal{E}} \text{ relatively compact} \Leftrightarrow \hat{I}(\hat{\mathcal{E}}) \text{ relatively compact} \Leftrightarrow \hat{I}(\hat{\mathcal{E}}) \text{ tight}.$$

Remark 2.11. *We do not need $\mathcal{P}(\mathcal{Y})$ to be Polish, to ensure that the relative compactness of $\hat{\mathcal{E}}$ implies its tightness (this would be needed to apply Theorem 5.3).*

Proof of Lemma 2.10: The implications

$$\hat{\mathcal{E}} \text{ tight} \Rightarrow \hat{\mathcal{E}} \text{ relatively compact, and , } \hat{I}(\hat{\mathcal{E}}) \text{ relatively compact} \Rightarrow \hat{I}(\hat{\mathcal{E}}) \text{ tight}$$

follow from Theorem 5.3. Like in Remark 2.8, we can check that $\hat{I}$ is continuous. Since the image of a relatively compact set by a continuous function is relatively compact, this ensures that

$$\hat{\mathcal{E}} \text{ relatively compact} \Rightarrow \hat{I}(\hat{\mathcal{E}}) \text{ relatively compact.}$$

Let us finally check that the tightness of $\hat{\mathcal{E}} \subset \mathcal{P}(\mathcal{P}(\mathcal{Y}))$ follows from that of $\hat{I}(\hat{\mathcal{E}})$. Let $\varepsilon > 0$. By the Prokhorov theorem (see Theorem 5.3), for each $n \in \mathbb{N}$ there exists some compact subset K_n of $\mathcal{Y}$ such that $\sup_{\nu \in \hat{I}(\hat{\mathcal{E}})} \nu(K_n^c) \leq \frac{\varepsilon^2}{2^{2n}}$. Let

$$\mathbb{K} = \left\{ p \in \mathcal{P}(\mathcal{Y}) : \forall n \in \mathbb{N},\ p(K_n^c) \leq \frac{\varepsilon}{2^n} \right\}.$$

For the weak convergence topology, this set is closed by Theorem 5.2 6) and relatively compact by Theorem 5.3. Hence $\mathbb{K}$ is a compact subset of $\mathcal{P}(\mathcal{Y})$. For $Q \in \hat{\mathcal{E}}$, we have, using $\mathbb{K}^c = \bigcup_{n \in \mathbb{N}} \left\{ p \in \mathcal{P}(\mathcal{Y}) : p(K_n^c) > \frac{\varepsilon}{2^n} \right\}$ then the Markov inequality,

$$\begin{aligned} Q(\mathbb{K}^c) &\leq \sum_{n \in \mathbb{N}} Q\left(\left\{ p \in \mathcal{P}(\mathcal{Y}) : p(K_n^c) > \frac{\varepsilon}{2^n} \right\} \right) \leq \sum_{n \in \mathbb{N}} \frac{2^n}{\varepsilon} \int_{\mathcal{P}(\mathcal{Y})} p(K_n^c) Q(dp) \\ &= \sum_{n \in \mathbb{N}} \frac{2^n}{\varepsilon} \hat{I}(Q)(K_n^c) \leq \sum_{n \in \mathbb{N}} \frac{2^n}{\varepsilon} \sup_{\nu \in \hat{I}(\hat{\mathcal{E}})} \nu(K_n^c) \leq \sum_{n \in \mathbb{N}} \frac{2^n}{\varepsilon} \times \frac{\varepsilon^2}{2^{2n}} = \varepsilon. \end{aligned}$$

∎

Proof of Lemma 2.9: Since the image of a relatively compact set by a continuous function is relatively compact, the necessary condition is a consequence of Lemma 2.7.

Let us suppose that $I(\mathcal{E})$ is relatively compact in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$. Of course, this set is relatively compact for the weak convergence topology and, by continuity of the projections, so are the sets of the first and second marginals of the elements of $I(\mathcal{E})$. Since $\mathcal{X}$ and $\mathcal{Y}$ are Polish, by Prokhorov's theorem (see Theorem 5.3), these two sets of marginals are tight. As the first marginal of P and $I(P)$ coincide, the set of first marginals of the elements of $\mathcal{E}$ are tight. Since the second marginal of $I(P)$ is the image of the second marginal of P by $\hat{I}$ introduced in Lemma 2.9, this lemma ensures that the set of second marginals of the elements of $\mathcal{E}$ is tight. Using that a product of compact sets is compact by Tikhonov's theorem, we deduce that $\mathcal{E}$ is tight and, in view of Theorem 5.3, relatively compact for the weak convergence topology.

From any sequence $(P_n)_{n \in \mathbb{N}}$ of elements of $\mathcal{E}$ we may extract a subsequence $(P_{n_k})_{k \in \mathbb{N}}$ such that P_{n_k} converges weakly to P_∞ as $k \to \infty$ and $I(P_{n_k})$ converges in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$ and therefore weakly to some limit, which, by Remark 2.8, is $I(P_\infty)$. According to (2.2), the convergence of the ρ-th order moment of $I(P_{n_k})$ to that of $I(P_\infty)$ implies the convergence of the ρ-th order moment of P_{n_k} to that of P_∞. We conclude that P_{n_k} converges to P_∞ in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$ and therefore $\mathcal{E}$ is relatively compact.

∎

Lemma 2.12. *Let $C : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous, bounded from below and convex in the measure argument. Then*

$$\forall P \in \mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})),\ P(C) \geq \int_{\mathcal{X}} C(x, I(P)_x) P(dx \times \mathcal{P}_\rho(\mathcal{Y})).$$

Proof: Since, by (2.1), $I(p)_x = \int_{\mathcal{P}_\rho(Y)} p P_x(dp)$, $P(dx \times \mathcal{P}(\mathcal{Y}))$ a.e., we have, using Jensen's inequality,

$$\begin{aligned}\int_{\mathcal{X}} C(x, I(P)_x) P(dx \times \mathcal{P}(\mathcal{Y})) &= \int_{\mathcal{X}} C\left(x, \int_{\mathcal{P}_\rho(\mathcal{Y})} p P_x(dp)\right) P(dx \times \mathcal{P}(\mathcal{Y})) \\ &\leq \int_{\mathcal{X}} C(x,p) P_x(dp) P(dx \times \mathcal{P}(\mathcal{Y})) = P(C).\end{aligned}$$

To justify the use of Jensen's inequality, let us prove that for $Q \in \mathcal{P}_\rho(\mathcal{P}_\rho(\mathcal{Y}))$, $C\left(x, \int_{\mathcal{P}_\rho(\mathcal{Y})} pQ(dp)\right) \leq \int_{\mathcal{P}_\rho(\mathcal{Y})} C(x,p) Q(dp)$. We approximate Q by $\frac{1}{n}\sum_{k=1}^n \delta_{p_k}$ where the $(p_k)_{k\geq 1}$ are i.i.d. according to Q and choose some ω in the underlying probability space such that $\frac{1}{n}\sum_{k=1}^n \delta_{p_k(\omega)}$ converges to Q in $\mathcal{P}_\rho(\mathcal{P}_\rho(\mathcal{Y}))$ as $n \to \infty$, which is possible according to the strong law of large numbers. By the continuity of $\mathcal{P}_\rho(\mathcal{P}_\rho(\mathcal{Y})) \ni R \mapsto \int_{\mathcal{P}_\rho(\mathcal{Y})} pR(dp) \in \mathcal{P}_\rho(\mathcal{Y})$ which can be established like in the proof of Lemma 2.7, $\frac{1}{n}\sum_{k=1}^n p_k(\omega)$ converges to $\int_{\mathcal{P}_\rho(\mathcal{Y})} pQ(dp)$ in $\mathcal{P}_\rho(\mathcal{Y})$. Then, by lower semi-continuity then convexity of C in its second argument and finally Theorem 5.2 7), we have

$$\begin{aligned}C\left(x, \int_{\mathcal{P}_\rho(\mathcal{Y})} pQ(dp)\right) &\leq \liminf_{n\to\infty} C\left(x, \frac{1}{n}\sum_{k=1}^n p_k(\omega)\right) \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n C(x, p_k(\omega)) \\ &= \liminf_{n\to\infty} \int_{\mathcal{P}_\rho(\mathcal{Y})} C(x,p) \frac{1}{n}\sum_{k=1}^n \delta_{p_k(\omega)}(dp) \leq \int_{\mathcal{P}_\rho(\mathcal{Y})} C(x,p) Q(dp).\end{aligned}$$

∎

Proof of Proposition 2.6: Let $(\pi^n)_{n\in\mathbb{N}}$ converge to π in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$. We denote by μ_n (resp. μ) the first marginal and by ν_n (resp. ν) the second marginal of π^n (resp. π). We have

$$\int_{\mathcal{X}} C(x, \pi_x^n) \mu_n(dx) = \int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} C(x,p) J(\pi^n)(dx, dp).$$

Since $I(J(\pi^n)) = \pi^n$, the sequence $(I(J(\pi^n)))_{n\in\mathbb{N}}$ is relatively compact in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$. By Lemma 2.9, we deduce that $(J(\pi^n))_{n\in\mathbb{N}}$ is relatively compact in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$. From any subsequence along which $\liminf_{n\to\infty} \int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} C(x,p) J(\pi^n)(dx, dp)$ is attained, we can extract a further subsequence $(J(\pi^{n_k}))_{k\in\mathbb{N}}$ such that $J(\pi^{n_k})$ converges to P in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$ as $k \to \infty$. The continuity of the intensity I stated in Lemma 2.7 ensures that $I(J(\pi^{n_k})) = \pi^{n_k}$ converges to $I(P)$ in $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$. We deduce that $I(P) = \pi$ and $P(dx \times \mathcal{P}_\rho(\mathcal{Y})) = \mu(dx)$.

Using Theorem 5.2 7) for the first inequality and Lemma 2.12 for the second, we deduce that

$$\begin{aligned}\liminf_{n\to\infty}\int_{\mathcal{X}} C(x,\pi^n_x)\mu_n(dx) &= \liminf_{n\to\infty}\int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} C(x,p)J(\pi^n)(dx,dp)\\ &= \lim_{k\to\infty}\int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} C(x,p)J(\pi^{n_k})(dx,dp) \geq \int_{\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})} C(x,p)P(dx,dp)\\ &\geq \int_{\mathcal{X}} C(x,I(P)_x)\mu(dx) = \int_{\mathcal{X}} C(x,\pi_x)\mu(dx).\end{aligned}$$

■

Proof of Theorem 2.2: Let $(\pi^n)_{n\in\mathbb{N}} \subset \Pi(\mu,\nu)$ be a minimizing sequence for $V_C(\mu,\nu)$. Since

$$\int_{\mathcal{X}\times\mathcal{Y}} \left(d^\rho_{\mathcal{X}}(x_0,x) + d^\rho_{\mathcal{X}}(y_0,y)\right)\pi^n(dx,dy) = \int_{\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu(dx) + \int_{\mathcal{Y}} d^\rho_{\mathcal{X}}(y_0,y)\nu(dy)$$

does not depend on n, Lemma 1.7 and Definition 1.38 ensure that we can extract a subsequence $(\pi^{n_k})_{k\in\mathbb{N}}$ converging to $\pi^\star$ in $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$. By Proposition 2.6, $\int_{\mathcal{X}} C(x,\pi^\star_x)\mu(dx) \leq \liminf_{k\to\infty}\int_{\mathcal{X}} C(x,\pi^{n_k}_x)\mu(dx) = V_C(\mu,\nu)$ so that the coupling $\pi^\star$ is optimal.

Let now $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{R}}$ converge to μ and ν as $n\to\infty$ respectively in $\mathcal{P}_\rho(\mathcal{X})$ and $\mathcal{P}_\rho(\mathcal{Y})$ and $\pi^n \in \Pi(\mu^n,\nu^n)$ be optimal for $V_C(\mu_n,\nu_n)$. By Lemma 1.7, from any subsequence of $(\pi^n)_{n\in\mathbb{N}}$ along which $\liminf_{n\to\infty}\int_{\mathcal{X}} C(x,\pi^n_x)\mu_n(dx)$ is attained, we can extract a further subequence $(\pi^{n_k})_{k\in\mathbb{N}}$ converging weakly to $\pi\in\Pi(\mu,\nu)$. Since

$$\int_{\mathcal{X}\times\mathcal{Y}} \left(d^\rho_{\mathcal{X}}(x_0,x) + d^\rho_{\mathcal{X}}(y_0,y)\right)\pi^{n_k}(dx,dy) = \int_{\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu_{n_k}(dx) + \int_{\mathcal{Y}} d^\rho_{\mathcal{X}}(y_0,y)\nu_{n_k}(dy)$$

converges to

$$\int_{\mathcal{X}} d^\rho_{\mathcal{X}}(x_0,x)\mu(dx) + \int_{\mathcal{Y}} d^\rho_{\mathcal{X}}(y_0,y)\nu(dy) = \int_{\mathcal{X}\times\mathcal{Y}} \left(d^\rho_{\mathcal{X}}(x_0,x) + d^\rho_{\mathcal{X}}(y_0,y)\right)\pi(dx,dy)$$

as $k\to\infty$, the sequence $(\pi^{n_k})_{k\in\mathbb{N}}$ converges to π in $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$. Using Proposition 2.6 for the second inequality, we deduce that

$$V_C(\mu,\nu) \leq \int_{\mathcal{X}} C(x,\pi_x)\mu(dx) \leq \lim_{k\to\infty}\int_{\mathcal{X}} C(x,\pi^{n_k}_x)\mu_{n_k}(dx) = \liminf_{n\to\infty} V_C(\mu_n,\nu_n).$$

Hence V_C is lower semi-continuous on $\mathcal{P}_\rho(\mathcal{X})\times\mathcal{P}_\rho(\mathcal{Y})$.
Let finally $\mu\in\mathcal{P}_\rho(\mathcal{X})$, $\nu,\tilde{\nu}\in\mathcal{P}_\rho(\mathcal{Y})$, $\pi\in\Pi(\mu,\nu)$, $\tilde{\pi}\in\Pi(\mu,\tilde{\nu})$ be optimal for $V_C(\mu,\nu)$ and $V_C(\mu,\tilde{\nu})$ respectively, and $\alpha\in[0,1]$. Using the convexity of C in the measure argument for the first inequality and $\alpha\pi + (1-\alpha)\tilde{\pi}\in\Pi(\mu,\alpha\nu+(1-\alpha)\tilde{\nu})$ for the second, we obtain

$$\begin{aligned}\alpha V_C(\mu,\nu) + (1-\alpha)V_C(\mu,\tilde{\nu}) &= \int_{\mathcal{X}} \left(\alpha C(x,\pi_x) + (1-\alpha)C(x,\tilde{\pi}_x)\right)\mu(dx)\\ &\geq \int_{\mathcal{X}} C(x,\alpha\pi_x + (1-\alpha)\tilde{\pi}_x)\mu(dx) \geq V_C(\mu,\alpha\nu+(1-\alpha)\tilde{\nu}).\end{aligned}$$

■

Lemma 2.13. *Let $f \in \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ be Lipschitz continuous in its first variable with constant L and $g(x) = \inf_{z\in\mathcal{Z}} f(x,z)$ for $x \in \mathcal{X}$. Then*

- *either g is $\mathbb{R}$-valued, Lipschitz continuous with the same constant L and for each $\varepsilon > 0$, there exists a measurable map $S_\varepsilon : \mathcal{X} \to \mathcal{Z}$ such that*
$$\forall x \in \mathcal{X},\ f(x, S_\varepsilon(x)) \le g(x) + \varepsilon,$$
- *or g is constantly equal to $-\infty$ and for each $\varepsilon > 0$, there exists a measurable map $S_\varepsilon : \mathcal{X} \to \mathcal{Z}$ such that*
$$\forall x \in \mathcal{X},\ f(x, S_\varepsilon(x)) \le -\frac{1}{\varepsilon}.$$

Proof: When $g(x_0) > -\infty$ for some x_0, we can check that g is $\mathbb{R}$-valued and Lipschitz continuous with constant L as in the proof of Lemma 1.13. The existence of S_ε then follows from the argument in the proof of Lemma 1.14. We now suppose that g is constantly equal to $-\infty$ and prove the existence of S_ε for $\varepsilon > 0$ in this case. Let $(x_n)_{n\in\mathbb{N}}$ be dense in $\mathcal{X}$. For each $n \in \mathbb{N}$, there exists y_n such that $f(x_n, y_n) \le -\frac{2}{\varepsilon}$. Since $S_\varepsilon(x) = x_0$ does the job when $L = 0$, we now suppose that $L > 0$. For $n \in \mathbb{N}$, we set $A_n = B(x_n, \frac{1}{L\varepsilon}) \cap \left\{\bigcup_{k=0}^{n-1} B(x_k, \frac{1}{L\varepsilon})\right\}^c \in \mathcal{B}(\mathcal{X})$. By density of $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$, this set is the disjoint union of the $(A_n)_{n\in\mathbb{N}}$. Let us define $S_\varepsilon(x) = \sum_{n\in\mathbb{N}} 1_{A_n}(x) y_n$. For $n \in \mathbb{N}$ and $x \in A_n$, we have, using the Lipschitz continuity of f in its first variable

$$f(x, S_\varepsilon(x)) = f(x, y_n) - f(x_n, y_n) + f(x_n, y_n) \le L \times \frac{1}{L\varepsilon} - \frac{2}{\varepsilon} = -\frac{1}{\varepsilon}.$$

∎

Proof of Theorem 2.4: Let

$$\widetilde{V}_C(\mu,\nu) = \inf\{P(C) : P \in \mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})) \text{ such that } I(P) \in \Pi(\mu,\nu)\}.$$

By Lemma 2.12, for $P \in \mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$ such that $I(P) \in \Pi(\mu,\nu)$, we have

$$P(C) \ge \int_{\mathcal{X}} C(x, I(P)_x)\mu(dx) \ge \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathcal{X}} C(x, \pi_x)\mu(dx) = V_C(\mu,\nu).$$

As a consequence, $\widetilde{V}_C(\mu,\nu) \ge V_C(\mu,\nu)$. The converse inequality holds since for $\pi^\star$ optimal for $V_C(\mu,\nu)$ (which exists by Theorem 2.2), $J(\pi^\star)(C) = \int_{\mathcal{X}} C(x, \pi^\star_x)\mu(dx) = V_C(\mu,\nu)$ while $I(J(\pi^\star)) = \pi^\star \in \Pi(\mu,\nu)$. Moreover, $J(\pi^\star)$ is an optimizer of $\widetilde{V}_C(\mu,\nu)$ and, by Theorem 2.2, $\widetilde{V}_C$ is lower semi-continuous and convex in its measure argument.

Since the case when C is constantly equal to $+\infty$ is obvious, we suppose that C takes finite values. We apply the Fenchel-Moreau theorem with V equal to the space of bounded signed measures q on $\mathcal{Y}$ such that $\int_{\mathcal{Y}} d^\rho_{\mathcal{Y}}(y_0, y)|q|(dy) < \infty$ where $|q|$ is the total variation of q. We have $V' = C_\rho(\mathcal{Y})$ with $g(q) = q(g) = \int_{\mathcal{Y}} g(y)q(dy)$ for $(g,q) \in C_\rho(\mathcal{Y}) \times V$. We fix $\mu \in \mathcal{P}_\rho(\mathcal{X})$ and set

$$F(q) = \begin{cases} \widetilde{V}_C(\mu, q) \text{ if } q \in \mathcal{P}_\rho(\mathcal{Y}) \\ +\infty \text{ if } q \in V \setminus \mathcal{P}_\rho(\mathcal{Y}) \end{cases}.$$

Since $\mathcal{P}_\rho(\mathcal{Y})$ is a closed convex subset of V and $\mathcal{P}_\rho(\mathcal{Y}) \ni \nu \mapsto \widetilde{V}_c(\mu,\nu)$ is lower semi-continuous and convex, F is lower semi-continuous and convex. Hence by the Fenchel-Moreau theorem (see Theorem 1.10),

$$\forall \nu \in \mathcal{P}_\rho(\mathcal{Y}),\ \widetilde{V}_C(\mu,\nu) = F(\nu) = \sup_{\psi\in C_\rho(\mathcal{Y})} \{\nu(\psi) - F^\star(\psi)\}. \tag{2.4}$$

Case C Lipschitz. Let $\psi \in C_\rho(\mathcal{Y})$. By Lemma 2.13, either ψ^C is constantly equal to $-\infty$ or it is $\mathbb{R}$-valued and Lipschitz continuous with the same constant as C. Let us first deal with the second case. For $\varepsilon > 0$, still by Lemma 2.13, there exists a measurable map

$$S_\varepsilon : \mathcal{X} \to \mathcal{P}_\rho(\mathcal{Y}) \text{ such that } \forall x \in \mathcal{X},\ C(x, S_\varepsilon(x)) - \psi(S_\varepsilon(x)) \le \psi^C(x) + \varepsilon,$$

where, by a small abuse of notation, $\psi(p) = p(\psi)$ for $p \in \mathcal{P}_\rho(\mathcal{Y})$. Setting $P_\varepsilon(dx,dp) = \mu(dx)\delta_{S_\varepsilon(x)}(dp)$, we have, using the definition of ψ^C for the first inequality

$$\begin{aligned}\mu(\psi^C) &= \inf_{\substack{\nu\in\mathcal{P}_\rho(\mathcal{Y})\\ P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)}} P(\psi^C \oplus 0) \le \inf_{\substack{\nu\in\mathcal{P}_\rho(\mathcal{Y})\\ P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)}} P(C - 0 \oplus \psi(\cdot))\\ &\le P_\varepsilon(C - 0\oplus\psi(\cdot)) \le P_\varepsilon(\psi^C \oplus 0 + \varepsilon) = \mu(\psi^C) + \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, we deduce that

$$\inf_{\substack{\nu\in\mathcal{P}_\rho(\mathcal{Y})\\ P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)}} P(C - 0\oplus\psi(\cdot)) = \mu(\psi^C).$$

This equality is preserved when ψ^C is constantly equal to $-\infty$ by applying the same reasoning with the measurable map

$$S_\varepsilon : \mathcal{X} \to \mathcal{P}_\rho(\mathcal{Y}) \text{ such that } \forall x \in \mathcal{X},\ C(x, S_\varepsilon(x)) - \psi(S_\varepsilon(x)) \le -\frac{1}{\varepsilon},$$

given by Lemma 2.13. Using the definitions of F and $\widetilde{V}_C$ for the second equality then that $P(0\oplus\psi(\cdot)) = \nu(\psi)$ for $P \in \mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y}))$ such that $I(P) \in \Pi(\mu,\nu)$, we deduce that

$$\begin{aligned}-F^\star(\psi) &= -\sup_{q\in V}\{q(\psi) - F(q)\} = -\sup_{\nu\in\mathcal{P}_\rho(\mathcal{Y})}\left\{\nu(\psi) - \inf_{P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)} P(C)\right\}\\ &= -\sup_{\substack{\nu\in\mathcal{P}_\rho(\mathcal{Y})\\ P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)}} P(0\oplus\psi(\cdot) - C)\\ &= \inf_{\substack{\nu\in\mathcal{P}_\rho(\mathcal{Y})\\ P\in\mathcal{P}_\rho(\mathcal{X}\times\mathcal{P}_\rho(\mathcal{Y})):I(P)\in\Pi(\mu,\nu)}} P(C - 0\oplus\psi(\cdot)) = \mu(\psi^C).\end{aligned}$$

Plugging this equality in (2.4), we conclude that

$$\widetilde{V}_C(\mu,\nu) = \sup_{\psi\in C_\rho(\mathcal{Y})}\{\mu(\psi^C) + \nu(\psi)\} \ge \sup_{\psi\in C_\rho^\wedge(\mathcal{Y})}\{\mu(\psi^C) + \nu(\psi)\}. \tag{2.5}$$

Let $\psi \in C_\rho(\mathcal{Y})$. For $k\in\mathbb{N}$, $\psi\wedge k \in C_\rho^\wedge(\mathcal{Y})$ and $(\psi\wedge k)^C(x)$ decreases with k to

$$\begin{aligned}\inf_{k\in\mathbb{N}}\inf_{p\in\mathcal{P}_\rho(\mathcal{Y})}\{C(x,p) - p(\psi\wedge k)\} &= \inf_{p\in\mathcal{P}_\rho(\mathcal{Y})}\inf_{k\in\mathbb{N}}\{C(x,p) - p(\psi\wedge k)\}\\ &= \inf_{p\in\mathcal{P}_\rho(\mathcal{Y})}\{C(x,p) - p(\psi)\} = \psi^C(x).\end{aligned}$$

We have

$$(\psi \wedge k)^C(x) \leq C(x, \delta_0) - \psi(0) \wedge k \leq C(x, \delta_0) + (\psi(0))^-,$$

where the right-hand side does not depend on k and is Lipschitz continuous with respect to x and therefore integrable with respect to the probability measure μ which belongs to $\mathcal{P}_\rho(\mathcal{X}) \subset \mathcal{P}_1(\mathcal{X})$. By monotone convergence, we deduce that

$$\mu\left((\psi \wedge k)^C\right) = \mu(C(\cdot, \delta_0)) + (\psi(0))^- - \mu\left(C(\cdot, \delta_0) + (\psi(0))^- - (\psi \wedge k)^C\right)$$

converges to $\mu(C(\cdot, \delta_0)) + (\psi(0))^- - \mu\left(C(\cdot, \delta_0) + (\psi(0))^- - \psi^C\right) = \mu\left(\psi^C\right)$ as $k \to \infty$. On the other hand, since $\nu \in \mathcal{P}_\rho(\mathcal{Y})$, by Lebesgue's theorem, $\nu(\psi \wedge k)$ converges to $\nu(\psi)$ as $k \to \infty$. Hence

$$\forall \psi \in C_\rho(\mathcal{Y}), \lim_{k \to \infty} \left\{\mu((\psi \wedge k)^C) + \nu(\psi \wedge k)\right\} = \left\{\mu(\psi^C) + \nu(\psi)\right\}.$$

With (2.5), we deduce that

$$\widetilde{V}_C(\mu, \nu) = \sup_{\psi \in C_\rho^\wedge(\mathcal{Y})} \{\mu(\psi^C) + \nu(\psi)\}.$$

When $\psi \in C_\rho^\wedge(\mathcal{Y})$, then ψ^C is $\mathbb{R}$ valued and Lipschitz continuous and therefore belongs to $C_1(\mathcal{Y})$ and satisfies $\psi^C(x) + p(\psi) \leq C(x, p)$ for all $(x, p) \in \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})$. On the other hand, when $\phi \in C_1(\mathcal{X})$ is such that $\phi(x) + p(\psi) \leq C(x, p)$ for all $(x, p) \in \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})$ (a condition which we abbreviate into $\phi \oplus \psi(\cdot) \leq C$) , then $\phi \leq \psi^C$. Therefore

$$\sup_{\substack{(\phi,\psi) \in C_1(\mathcal{X}) \times C_\rho^\wedge(\mathcal{Y}) \\ \phi \oplus \psi(\cdot) \leq C}} \{\mu(\phi) + \nu(\psi)\} = \sup_{\psi \in C_\rho^\wedge(\mathcal{Y})} \{\mu(\psi^C) + \nu(\psi)\} = \widetilde{V}_C(\mu, \nu).$$

General case. We use the sequence $(C_n)_{n \geq 1}$ of n-Lipschitz functions growing to C given by Lemma 1.13. For $P \in \mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$ such that $I(P) \in \Pi(\mu, \nu)$ and $(\phi, \psi) \in C_1(\mathcal{X}) \times C_\rho^\wedge(\mathcal{Y})$ such that $\phi \oplus \psi(\cdot) \leq C$, we have $\phi \leq \psi^C$ where, by definition of ψ^C, $\psi^C + \psi(\cdot) \leq C$ so that

$$P(C) \geq P(\psi^C \oplus \psi(\cdot)) = \mu(\psi^C) + \nu(\psi) \geq \mu(\phi) + \nu(\psi).$$

Taking the infimum over P and the supremum over (ϕ, ψ) we deduce that

$$\begin{aligned}\widetilde{V}_C(\mu, \nu) &\geq \sup_{\psi \in C_\rho^\wedge(\mathcal{Y})} \left\{\mu(\psi^C) + \nu(\psi)\right\} \geq \sup_{\substack{(\phi,\psi) \in C_1(\mathcal{X}) \times C_\rho^\wedge(\mathcal{Y}) \\ \phi \oplus \psi(\cdot) \leq C}} \{\mu(\phi) + \nu(\psi)\} \\ &\geq \sup_{n \in \mathbb{N}} \sup_{\substack{(\phi,\psi) \in C_1(\mathcal{X}) \times C_\rho^\wedge(\mathcal{Y}) \\ \phi \oplus \psi(\cdot) \leq C_n}} \{\mu(\phi) + \nu(\psi)\} = \sup_{n \in \mathbb{N}} \widetilde{V}_{C_n}(\mu, \nu), \end{aligned} \tag{2.6}$$

where we used the case when the cost function is Lipschitz for the last equality. Let for $n \geq 1$, $P_n \in \mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$ such that $I(P) \in \Pi(\mu, \nu)$ be optimal for $\widetilde{V}_{C_n}(\mu, \nu)$ (the existence of P_n was checked at the beginning of the proof). By Lemma 3.12 below, $\Pi(\mu, \nu)$ is compact in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$. Hence the sequence $(I(P_n))_{n \geq 1}$ is relatively compact in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{Y})$. By Lemma 2.9, we may extract a sequence $(P_{n_k})_k$ which converges weakly to P_∞ in $\mathcal{P}_\rho(\mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y}))$. Then, by the continuity of I stated in

Lemma 2.7, $I(P_{n_k})$ converges in $\mathcal{P}_\rho(\mathcal{X}\times\mathcal{Y})$ to $I(P_\infty)$ which belongs to $\Pi(\mu,\nu)$ since this set is closed. We have

$$\sup_{n\in\mathbb{N}} \widetilde{V}_{C_n}(\mu,\nu) = \lim_{n\to\infty} \nearrow P_n(C_n) = \lim_{k\to\infty} P_{n_k}(C_{n_k}).$$

For fixed $m\in\mathbb{N}^*$, when k is large enough, we have $n_k \geq m$ and $C_{n_k} \geq C_m$. By the Portmanteau theorem (see Theorem 5.2), $\liminf_{k\to\infty} P_{n_k}(C_m) \geq P_\infty(C_m)$. By the monotone convergence theorem, $\sup_{m\geq 1} P_\infty(C_m) = \pi_\infty(C)$. Therefore

$$\sup_{n\in\mathbb{N}} \widetilde{V}_{C_n}(\mu,\nu) = \lim_{k\to\infty} P_{n_k}(C_{n_k}) \geq \sup_{m\geq 1}\liminf_{k\to\infty} P_{n_k}(C_m) \geq \sup_{m\geq 1} P_\infty(C_m) = P(C) \geq \widetilde{V}_C(\mu,\nu).$$

With (2.6), we conclude that

$$\widetilde{V}_C(\mu,\nu) = \sup_{\psi\in C_\rho^\wedge(\mathcal{Y})} \left\{\mu(\psi^C)+\nu(\psi)\right\} = \sup_{\substack{(\phi,\psi)\in C_1(\mathcal{X})\times C_\rho^\wedge(\mathcal{Y})\\ \phi\oplus\psi(\cdot)\leq C}} \{\mu(\phi)+\nu(\psi)\}.$$

■

2.2 Strassen's theorem

Let us introduce the convex order on $\mathcal{P}_1(\mathbb{R}^d)$.

Definition 2.14. *For $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^d)$, we say that μ is smaller than ν for the convex order and denote $\mu\leq_{cx}\nu$ if*

$$\forall\varphi:\mathbb{R}^d\to\mathbb{R} \text{ convex},\ \mu(\varphi)\leq\nu(\varphi).$$

Remark 2.15. *When $\varphi:\mathbb{R}^d\to\mathbb{R}$ is convex, then φ is bounded from below by the affine functions $\mathbb{R}^d\ni x\mapsto\varphi(\tilde{x})+y.(x-\tilde{x})$ where y is any element of the subdifferential of φ at $\tilde{x}\in\mathbb{R}^d$. As a consequence, for each $\eta\in\mathcal{P}_1(\mathbb{R}^d)$, $\eta(\varphi^-)<\infty$, i.e. φ is semi-integrable with respect to η and $\eta(\varphi)$ makes sense in $\mathbb{R}\cup\{+\infty\}$.*

Theorem 2.16. *Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^d)$. Then*

$$\Pi_M(\mu,\nu)\neq\emptyset \Leftrightarrow \mu\leq_{cx}\nu.$$

The necessary condition follows from Jensen's inequality since for $\pi\in\Pi_M(\mu,\nu)$ and $\varphi:\mathbb{R}^d\to\mathbb{R}$ convex,

$$\mu(\varphi)=\int_{\mathbb{R}^d}\varphi(\bar{\pi}_x)\mu(dx)\leq\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\varphi(y)\pi_x(dy)\mu(dx)=\int_{\mathbb{R}^d\times\mathbb{R}^d}\varphi(y)\pi(dx,dy)=\nu(\varphi).$$

The proof of the sufficient condition is more difficult and relies on the next lemma

Lemma 2.17. *Let $\phi:\mathbb{R}^d\to\mathbb{R}$ be bounded from below by an affine function. Then the largest lower semi-continuous convex function $\phi^{\star\star}$ smaller than ϕ is given by*

$$\forall x\in\mathbb{R}^d,\ \phi^{\star\star}(x)=\inf\left\{p(\phi):p\in\mathcal{P}_1(\mathbb{R}^d) \text{ such that } \bar{p}=x\right\}.$$

Remark 2.18. *The supremum of lower semi-continuous and convex functions is lower semi-continuous and convex, which, together with the convexity and lower semi-continuity of the affine lower bound, explains why $\phi^{\star\star}$ exists.*

Proof: Let

$$\psi(x) = \inf\left\{p(\phi) : p \in \mathcal{P}_1(\mathbb{R}^d) \text{ such that } \bar{p} = x\right\},\ x \in \mathbb{R}^d.$$

For $p \in \mathcal{P}_1(\mathbb{R}^d)$ such that $\bar{p} = x$, we have by Jensen's inequality and $\phi^{\star\star} \leq \phi$,

$$\phi^{\star\star}(x) = \phi^{\star\star}(\bar{p}) \leq p(\phi^{\star\star}) \leq p(\phi).$$

We deduce that $\phi^{\star\star} \leq \psi$ and ψ is real valued. We have $\psi(x) \leq \phi(x)$ for the choice $p = \delta_x$. On the other hand, for $x, y \in \mathbb{R}^d$ and $\varepsilon > 0$, choosing p_x and p_y ε-optimal for $\psi(x)$ and $\psi(y)$ respectively, we get for $\alpha \in [0, 1]$,

$$\psi(\alpha x + (1-\alpha)y) \leq (\alpha p_x + (1-\alpha)p_y)(\phi) \leq \alpha\psi(x) + (1-\alpha)\psi(y) + \varepsilon.$$

Letting $\varepsilon \to 0$, we deduce that ψ is convex. It is therefore enough to check that ψ is lower semi-continuous to conclude that $\psi = \phi^{\star\star}$. For this purpose, we let $x \in \mathbb{R}^d$, $y_1, \cdots, y_{d+1} \in \mathbb{R}^d$ such that x is in the interior of the convex hull of $y_1, \cdots, y_{d+1}$. Let $(x_n)_{n\in\mathbb{N}} \subset \mathbb{R}^d$ converge to x as $n \to \infty$ and p_n be ε-optimal for $\psi(x_n)$. Up to removing the first terms of the sequence, we may suppose that for each $n \in \mathbb{N}$, $x_n + \frac{x-x_n}{\varepsilon}$ belongs to the convex hull of $y_1, \cdots, y_{d+1}$ so that there exists a probability measure q_n supported in $\{y_1, \cdots, y_{d+1}\}$ such that $\bar{q}_n = x_n + \frac{x-x_n}{\varepsilon}$. Then $\overline{(1-\varepsilon)p_n + \varepsilon q_n} = (1-\varepsilon)x_n + \varepsilon x_n + x - x_n = x$ so that

$$\psi(x) \leq (1-\varepsilon)p_n(-\phi) + \varepsilon q_n(\phi) \leq (1-\varepsilon)(\psi(x_n) + \varepsilon) + \varepsilon \max_{1\leq i\leq d+1} \phi(y_i).$$

We deduce that $\psi(x) \leq (1-\varepsilon)\liminf_{n\to\infty} \psi(x_n) + \varepsilon(1 - \varepsilon \max_{1\leq i\leq d+1} \phi(y_i))$ and conclude that ψ is lower semi-continuous by letting $\varepsilon \to 0$.

∎

Proof: The obvious necessary condition was proved using Jensen's inequality just after the statement. To check the sufficient condition, we define

$$C : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \ni (x, p) \mapsto |x - \bar{p}| \in \mathbb{R}.$$

The function C is continuous and convex in the measure argument as the composition of a convex function with the linear function $\mathcal{P}_1(\mathbb{R}^d) \ni p \mapsto \bar{p} \in \mathbb{R}^d$. By Theorem 2.2, there exists $\pi^\star \in \Pi(\mu, \nu)$ such that $V_C(\mu, \nu) = \int_{\mathbb{R}^d} |x - \bar{\pi}^\star_x|\mu(dx)$. Since, when $V_C(\mu, \nu) = 0$, then $\bar{\pi}^\star_x = x$, $\mu(dx)$ a.e., i.e. $\pi^\star \in \Pi_M(\mu, \nu)$, it is enough to check that $\mu \leq_{cx} \nu \Rightarrow V_C(\mu, \nu) = 0$, which we now do.
Let us check that when $\psi \in C_1(\mathbb{R}^d)$ (continuous with affine growth), then $\psi \leq -\psi^C$ and ψ^C is 1-Lipschitz convex. By Definition 2.3 and Lemma 2.17, we have

$$\begin{aligned}\psi^C(x) &= \inf_{p\in\mathcal{P}_1(\mathbb{R}^d)}\{|x - \bar{p}| - p(\psi)\} = \inf_{z\in\mathbb{R}^d}\ \inf_{p\in\mathcal{P}_1(\mathbb{R}^d):\bar{p}=z}\{|x - \bar{p}| + p(-\psi)\} \\ &= \inf_{z\in\mathbb{R}^d}\{|x - z| + (-\psi)^{\star\star}(z)\}. \qquad (2.7)\end{aligned}$$

For the choice $z = x$, we deduce also using the definition of $(-\psi)^{\star\star}$ that $\psi^C \le (-\psi)^{\star\star} \le -\psi$ so that $\psi \le -\psi^C$.

If $\psi^C(x) \le \psi^C(\tilde{x})$, then choosing z_n such that $\psi^C(x) \ge |x - z_n| + (-\psi)^{\star\star}(z_n) - \frac{1}{n}$ and $\tilde{z}_n$ such that $\psi^C(\tilde{x}) \ge |x - \tilde{z}_n| + (-\psi)^{\star\star}(\tilde{z}_n) - \frac{1}{n}$ for $n \in \mathbb{N}^*$, we get

$$\psi^C(\tilde{x}) - \psi^C(x) \le |\tilde{x} - z_n| + (-\psi)^{\star\star}(z_n) - \left(|x - z_n| + (-\psi)^{\star\star}(z_n) - \frac{1}{n}\right) \le |\tilde{x} - x| + \frac{1}{n}.$$

Moreover, for $\alpha \in [0,1]$, by convexity of $(-\psi)^{\star\star}$,

$$\begin{aligned}\psi^C(\alpha x + (1-\alpha)\tilde{x}) &\le |\alpha z_n + (1-\alpha)\tilde{z}_n - (\alpha x + (1-\alpha)\tilde{x})| + (-\psi)^{\star\star}(\alpha z_n + (1-\alpha)\tilde{z}_n) \\ &\le \alpha\left(|x - \tilde{z}_n| + (-\psi)^{\star\star}(\tilde{z}_n)\right) + (1-\alpha)\left(|x - \tilde{z}_n| + (-\psi)^{\star\star}(\tilde{z}_n)\right) \\ &\le \alpha\psi^C(x) + (1-\alpha)\psi^C(\tilde{x}) + \frac{1}{n}\end{aligned}$$

By taking the limit $n \to \infty$, we deduce that ψ^C is 1-Lipschitz convex.

If ψ is 1-Lipschitz concave, then $-\psi$ is 1-Lipschitz convex so that $(-\psi)^{\star\star} = -\psi$. With (2.7), we deduce that

$$\psi^C(x) = \inf_{z \in \mathbb{R}^d} \{|x - z| - \psi(z)\} \ge \inf_{z \in \mathbb{R}^d} \{|x - z| - \psi(x) - |x - z|\} = -\psi(x)$$

and $\psi^C(x) \le -\psi(x)$ by the choice $z = x$ in the first equality so that $\psi^C = -\psi$.

When $\psi \in C_1(\mathbb{R}^d)$, then the inequality $\psi^C \le -\psi$ ensures that μ is semi-integrable with respect to μ which belongs to $\mathcal{P}_1(\mathbb{R}^d)$. Therefore $\mu(\psi^C)$ makes sense in $\{-\infty\} \cup \mathbb{R}$ and so does $\mu(\psi^C) + \nu(\psi)$. With Theorem 2.4, we deduce that

$$V_C(\mu,\nu) = \sup_{\psi \in \mathcal{C}_1^{\wedge}(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(\psi)\right\} \le \sup_{\psi \in \mathcal{C}_1(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(\psi)\right\}.$$

Since the definition of ψ^C (see Definition 2.3) ensures that $\forall (x,\rho) \in \mathcal{X} \times \mathcal{P}_\rho(\mathcal{Y})$, $C(x,p) \ge \psi^C(x) + p(\psi)$, we have for $\psi \in C_1(\mathbb{R}^d)$ and $\pi \in \Pi(\mu,\nu)$,

$$\int_{\mathbb{R}^d} C(x,\pi_x)\mu(dx) \ge \int_{\mathbb{R}^d} \{\psi^C(x) + \pi_x(\psi)\}\mu(dx) = \mu(\psi^C) + \nu(\psi).$$

Taking the infimum over $\pi \in \Pi(\mu,\nu)$ and the supremum over $\psi \in C_1(\mathbb{R}^d)$, we conclude that

$$V_C(\mu,\nu) \ge \sup_{\psi \in \mathcal{C}_1(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(\psi)\right\} \ge V_C(\mu,\nu).$$

Using the inequality $\psi \le -\psi^C$ for the first inequality, the fact that ψ^C is 1-Lipschitz convex for the second inequality, the fact that $-\psi = \psi^C$ when ψ is 1-Lipschitz concave for the third equality, we deduce that

$$\begin{aligned}V_C(\mu,\nu) &= \sup_{\psi \in \mathcal{C}_1(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(\psi)\right\} \le \sup_{\psi \in \mathcal{C}_1(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(-\psi^C)\right\} \\ &\le \sup_{\phi \text{ 1-Lipschitz convex}} \{\mu(\phi) - \nu(\phi)\} = \sup_{\psi \text{ 1-Lipschitz concave}} \{\mu(-\psi) + \nu(\psi)\} \\ &= \sup_{\psi \text{ 1-Lipschitz concave}} \{\mu(\psi^C) + \nu(\psi)\} \le \sup_{\psi \in \mathcal{C}_1(\mathbb{R}^d)} \left\{\mu(\psi^C) + \nu(\psi)\right\} = V_C(\mu,\nu).\end{aligned}$$

Hence $V_C(\mu,\nu) = \sup_{\phi \text{ 1-Lipschitz convex}}\{\mu(\phi) - \nu(\phi)\}$ and

$$\mu \leq_{cx} \nu \Rightarrow \forall \phi \text{ 1 − Lipschitz convex, } \mu(\phi) - \nu(\phi) \leq 0 \Rightarrow V_C(\mu,\nu) = 0.$$

∎

Remark 2.19. *For the cost function* $C(x,\bar{p}) = |x - \bar{p}|$*,*

$$V_C(\mu,\nu) = \sup_{\phi \text{ 1-Lipschitz convex}} \{\mu(\phi) - \nu(\phi)\}$$

permits to measure in general the possible lack of convex order between μ *and* ν*. Note that*

$$\forall \phi \text{ 1 − Lipschitz convex, } \mu(\phi) \leq \nu(\phi) \Rightarrow V_C(\mu,\nu) = 0 \Rightarrow \Pi_M(\mu,\nu) \neq \emptyset \Rightarrow \mu \leq_{cx} \nu.$$

Since the converse implication $\mu \leq_{cx} \nu \Rightarrow \forall \phi$ *1−Lipschitz convex,* $\mu(\phi) \leq \nu(\phi)$ *is obvious, we conclude that*

$$\mu \leq_{cx} \nu \Leftrightarrow \forall \phi \text{ 1 − Lipschitz convex, } \mu(\phi) \leq \nu(\phi),$$

i.e. Lipschitz convex functions are enough to characterize the convex order.

Chapter 3

Martingale Optimal Transport

3.1 From Finance to Robust Finance

We consider the evolution in discrete time $t \in \{0, 1, 2, \cdots, T\}$ of a market with 2 assets :

- one riskless asset with constant value 1 i.e. we assume zero interest rates, a simplifying assumption which could be relaxed,
- one risky asset with successive values $S_0, S_1, \cdots, S_T$.

In a portfolio strategy, the amount H_t of risky asset held on $[t, t+1]$ is decided at time t in view of the information available on the market up to this time. Starting from the initial wealth 0 and choosing a self-financing strategy, we get that the value of the portfolio at time t is $(H.S)_t = \begin{cases} 0 \text{ if } t = 0 \\ \sum_{s=0}^{t-1} H_s(S_{s+1} - S_s) \text{ otherwise} \end{cases}$. Let us introduce a probabilistic setup $(\Omega, (\mathcal{F}_t)_{t=0}^T, \mathbb{P})$ where the historical probability measure $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F}_T)$ and $((S_t, H_t))_{t=0}^T$ is adapted to the filtration $(\mathcal{F}_t)_{t=0}^T$. The mathematical formulation of the Absence of Arbitrage Opportunities, which is a natural modeling assumption saying that there is no possibility to earn money without taking risk is

$$\text{(AAO)} \;\; \mathbb{P}((H.S)_T \geq 0) = 1 \Rightarrow \mathbb{P}((H.S)_T = 0) = 1.$$

The first fundamental theorem of asset pricing [8] is that (AAO) is equivalent to the existence of a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ under which, $(S_t)_{t=0}^T$ is a martingale. Such a probability measure $\mathbb{Q}$ is called an Equivalent Martingale Measure. When one considers an option with payoff F $\mathcal{F}_T$-measurable then

$$\sup_{Q \text{ EMM}} \mathbb{Q}(F) = \inf\{a \in \mathbb{R} : \exists (H_t)_{t=0}^T \text{ adapted self-financing s.t. } \mathbb{P}(a + (H.S)_T \geq F) = 1\}$$
$$\inf_{Q \text{ EMM}} \mathbb{Q}(F) = \sup\{a \in \mathbb{R} : \exists (H_t)_{t=0}^T \text{ adapted self-financing s.t. } \mathbb{P}(a + (H.S)_T \leq F) = 1\}.$$

In the robust finance approach, rather than starting with the sensitive choice of the historical probability $\mathbb{P}$, one only takes into account informations given by the market. Let for $K \geq 0$, $C(t, K)$ be the initial market price of the Call option with payoff $(S_t - K)^+$ at time t. The following are consequences of the Absence of Arbitrage Opportunities :

1) $\forall K \geq 0$, $C(t,K) \geq 0$,

2) by convexity of $K \mapsto (S_t - K)^+$, $\mathbb{R}_+ \ni K \mapsto C(t,K)$ is convex,

3) $C(t,0) = S_0$ since $(S_t - 0)_+ = S_t$,

4) $\lim_{K\to\infty} C(t,K) = 0$ since $\lim_{K\to\infty}(S_t - K)^+ = 0$,

5) for $0 \leq K \leq \tilde{K}$, $C(t,K) - C(t,\tilde{K}) \leq \tilde{K} - K$ since $(S_t - K)^+ - (S_t - \tilde{K})^+ \leq \tilde{K} - K$.

The following result is due to Breeden and Lizenberger [6].

Lemma 3.1. *Assume that* $\mathbb{R}_+ \ni K \mapsto C(t,K)$ *satisfies* 1)−5)*. Then there exists a unique probability measure* $\mu_t \in \mathcal{P}_1(\mathbb{R}_+)$ *such that* $\int_{\mathbb{R}_+} x\mu_t(dx) = S_0$ *and* $\forall K \in \mathbb{R}_+,\ C(t,K) = \int_{\mathbb{R}_+}(x-K)^+\mu(dx)$.

Proof: Let $\partial_{K+}C(t,K)$ denote the right-hand derivative of the convex function $K \mapsto C(t,K)$. Then $K \mapsto -\partial_{K+}C(t,K)$ is right-continuous and non-increasing, such that $-\partial_{K+}C(t,0) \leq 1$ by 5) and that $\lim_{K\to+\infty} -\partial_{K+}C(t,K) = 0$ by 4). We set

$$\mu_t(\{0\}) = 1 + \partial_{K+}C(t,0) \text{ and } \mu_t((K,+\infty)) = -\partial_{K+}C(t,K) \text{ for } K \geq 0.$$

By 4) then Fubini's theorem,

$$\begin{aligned}\forall K \geq 0,\ C(t,K) &= -\int_{(K,+\infty)} \partial_{K+}C(t,\ell)d\ell = \int_{(K,+\infty)}\int_{(\ell,+\infty)} \mu_t(dx)d\ell \\ &= \int_{[0,+\infty)}\int 1_{\{K<\ell<x\}}d\ell\mu(dx) = \int_{\mathbb{R}_+}(x-K)^+\mu_t(dx).\end{aligned}$$

For $K = 0$, we deduce with 3) that $\int_{\mathbb{R}_+} x\mu_t(dx) = C(t,0) = S_0$. ■

For $t \leq u$, since $(S_u - K) = (S_u - K)^+ - (K - S_u)^+$, the market price $(S_t - K)$ at time t of the wealth $(S_u - K)$ at time u is equal to the market price $C_t(u,K)$ at time t of the Call option with strike K and maturity u minus the market price at time t of the Put option with strike K and maturity u. Hence $C_t(u,K) \geq (S_t - K)^+$ and, still by Absence of Arbitrage Opportunities, $C(u,K) \geq C(t,K)$. With Remark 3.7 below, we deduce that for $t \leq u$, $\mu_t \leq_{cx} \mu_u$.

The extreme pricing values of an option with payoff $f(S_1,\cdots,S_T)$ are

$$\inf / \sup \left\{Q(f) : Q \text{ martingale measure on } R_+^T \text{ with marginals } Q_t = \mu_t \text{ for } t \in \{1,\cdots,T\}\right\}.$$

We will concentrate on the $T = 2$ case, denote $\mu = \mu_1$, $\nu = \mu_2$ and $c(x,y) = f(x,y)$ even if the results that we state have higher dimensional versions (sometimes more technical). We set

$$V_c^M(\mu,\nu) = \inf_{\pi\in\Pi_M(\mu,\nu)} \pi(c) \text{ (under the convention } \inf\emptyset = +\infty).$$

The results that we derive for the infimum hold true for the supremum, up to replacing the assumption c lower semi-continuous bounded from below by c upper semi-continuous bounded from above. We even consider $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$, which correspond to d risky assets, even if the joint distributions of these asset prices at times $t \in \{1,2\}$ cannot be fully derived from the market prices of traded options.

3.2 Existence of optimizers and duality

Remark 3.2.

- *Monge type couplings $(i_{\mathbb{R}^d}, T)\#\mu$ are martingale couplings iff $T(x) = x$, $\mu(dx)$ a.e.. Therefore, when $\mu \neq \nu$, $\pi_M(\mu,\nu)$ contains no Monge type coupling.*

- *For $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, $\Pi_M(\mu,\mu) = \{(i_{\mathbb{R}^d}, i_{\mathbb{R}^d})\#\mu\}$. Indeed, clearly, $(i_{\mathbb{R}^d}, i_{\mathbb{R}^d})\#\mu \in \Pi_M(\mu,\mu)$. On the other hand, the function $\varphi(x) = \sqrt{1+|x|^2}$ with gradient $\nabla\varphi(x) = \frac{x}{\sqrt{1+|x|^2}}$ and Heassian matrix $\nabla^2\varphi(x) = \frac{1}{(1+|x|^2)^{3/2}}\left((1+|x|^2)I_d - xx^T\right) \geq \frac{I_d}{(1+|x|^2)^{3/2}}$ is strictly convex and belongs to $C_1(\mathbb{R}^d)$. As a consequence, for $\pi \in \Pi_M(\mu,\mu)$, by Jensen's inequality,*

$$\mu(\varphi) = \int_{\mathbb{R}^d} \varphi(x)\mu(dx) = \int_{\mathbb{R}^d} \varphi(\bar{\pi}_x)\mu(dx) \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \varphi(y)\pi_x(dy)\mu(dx) = \int_{\mathbb{R}^d} \varphi(y)\mu(dy),$$

  so that $\mu(dx)$ a.e., $\varphi(\bar{\pi}_x) = \int_{\mathbb{R}^d} \varphi(y)\pi_x(dy)$ and, by strict convexity of φ, $\pi_x = \delta_{\bar{\pi}_x} = \delta_x$.

- *The equivalent for MOT of Proposition 1.4 is the result of the next exercise.*

Exercise 3.3. *Check that if any optimal coupling for $V_c^M(\mu,\nu)$ writes $\mu(dx)\left(p(x)\delta_{S(x)} + (1-p(x))\delta_{T(x)}\right)(dy)$ for some measurable functions $p : \mathbb{R}^d \to [0,1]$ and $S, T : \mathbb{R}^d \to \mathbb{R}^d$ such that $p(x)S(x) + (1-p(x))T(x) = x$, $\mu(dx)$ a.e., then there is at most one optimal coupling. You may first remark that for given $s, t \in \mathbb{R}^d$, there is at most one probability measure with expectation x which writes $q\delta_s + (1-q)\delta_t$ for some $q \in [0,1]$.*

According to Strassen's theorem (see Theorem 2.16),

$$\forall \mu,\nu \in \mathcal{P}_1(\mathbb{R}^d),\ \pi_M(\mu,\nu) \neq \emptyset \Leftrightarrow \mu \leq_{cx} \nu.$$

Example 3.4. *Let $d = 1$, $\mu = \frac{1}{2}(\delta_{-1} + \delta_1)$, $\nu = \frac{1}{3}(\delta_{-4} + \delta_0 + \delta_4)$. There are uncountably many martingale couplings with marginals μ and ν since for each $\alpha \in [-\frac{5}{24}, \frac{5}{24}]$, the coupling $\pi^\alpha(dx,dy) = \frac{1}{2}\left(\delta_{-1}(dx)\pi^\alpha_{-1}(dy) + \delta_1(dx)\pi^\alpha_1(dy)\right)$ with*

$$\pi^\alpha_{-1} = \left(\frac{11}{24} - \alpha\right)\delta_{-4} + \left(\frac{1}{3} + 2\alpha\right)\delta_0 + \left(\frac{5}{24} - \alpha\right)\delta_4 \text{ and}$$
$$\pi^\alpha_{1} = \left(\frac{5}{24} + \alpha\right)\delta_{-4} + \left(\frac{1}{3} - 2\alpha\right)\delta_0 + \left(\frac{11}{24} + \alpha\right)\delta_4,$$

belongs to $\Pi_M(\mu,\nu)$. Indeed,

$$\frac{1}{2}\left(\left(\frac{11}{24} \mp \alpha\right) + \left(\frac{5}{24} \pm \alpha\right)\right) = \frac{16}{48} = \frac{1}{3} = \frac{1}{2}\left(\left(\frac{1}{3} + 2\alpha\right) + \left(\frac{1}{3} - 2\alpha\right)\right)$$

and $\left(\frac{11}{24} \mp \alpha\right) \times (\mp 4) + \left(\frac{1}{3} \pm 2\alpha\right) \times 0 + \left(\frac{5}{24} \mp \alpha\right) \times (\pm 4) = \mp 1$.

It is not so easy to check whether $\mu \leq \mathcal{X}\nu$ when $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$. Of course, a necessary condition is the equality of the expectations $\bar{\mu} = \bar{\nu}$ (take $\varphi(x) = \pm x$ in Definition 2.14 or use that the marginals of a martingale coupling share the same expectation). Nevertheless, in dimension $d = 1$, the convex order can be characterized through the potential functions.

Proposition 3.5. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$. Then $\mu \leq_{cx} \nu \Leftrightarrow \forall x \in \mathbb{R},\ u_\mu(x) \leq u_\nu(x)$ where for $\eta \in \mathcal{P}_1(\mathbb{R})$, $u_\eta(x) = \int_{\mathbb{R}} |x-y|\eta(dy)$.*

Remark 3.6. *By convexity of $\mathbb{R} \ni x \mapsto |x-y|$, the potential function u_η of $\eta \in \mathcal{P}_1(\mathbb{R})$ is convex. Its right-hand derivative is equal to $\int_{\mathbb{R}} \left(1_{\{x\geq y\}} - 1_{\{x<y\}}\right)\eta(dy) = 2F_\eta(x) - 1$ so that it characterizes the probability measure η. Moreover, for $x, \tilde{x} \in \mathbb{R}$,*

$$|u_\eta(x) - u_\eta(\tilde{x})| \leq \int_{\mathbb{R}} ||x-y| - |\tilde{x}-y||\,\eta(dy) \leq |x-\tilde{x}|,$$

and u_η is 1-Lipschitz. Last, according to the proof of Proposition 3.5, $\lim_{x\to+\infty}|u_\eta(x) - x + \bar{\eta}| = \lim_{x\to-\infty}|u_\eta(x) + x - \bar{\eta}| = 0$.

Proof: The necessary condition follows from the convexity of $\mathbb{R} \ni y \mapsto |x-y|$ for fixed $x \in \mathbb{R}$. Let us prove the sufficient condition. Since $|x-y| = 2(y-x)^+ + x - y = 2(x-y)^+ + y - x$, we have for $\eta \in \mathcal{P}_1(\mathbb{R})$

$$\forall x \in \mathbb{R},\ u_\eta(x) = 2\int_{\mathbb{R}} (y-x)^+\eta(dy) + x - \bar{\eta} = 2\int_{\mathbb{R}} (x-y)^+\eta(dy) + \bar{\eta} - x \tag{3.1}$$

where $\int_{\mathbb{R}}(y-x)^+\eta(dy)$ goes to 0 as $x \to \infty$ and $\int_{\mathbb{R}}(x-y)^+\eta(dy)$ goes to 0 as $x \to -\infty$ by Lebesgue's theorem. Therefore the inequality $u_\mu(x) \leq u_\nu(x)$ yields $-\bar{\mu} \leq -\bar{\nu}$ in the limit $x \to +\infty$ and $\bar{\mu} \leq \bar{\nu}$ in the limit $x \to -\infty$, so that $\bar{\mu} = \bar{\nu}$. Therefore $u_\mu \leq u_\nu$ also implies

$$\forall x \in \mathbb{R},\ \int_{\mathbb{R}} (y-x)^+\mu(dy) \leq \int_{\mathbb{R}} (y-x)^+\nu(dy) \text{ and } \int_{\mathbb{R}} (x-y)^+\mu(dy) \leq \int_{\mathbb{R}} (x-y)^+\nu(dy). \tag{3.2}$$

For $\varphi : \mathbb{R} \to \mathbb{R}$ convex, the second order distribution derivative φ'' is a non-negative measure and

$$\forall y \in \mathbb{R},\ \varphi(y) = \varphi(0) + \varphi'_+(0)y + \int_{(0,+\infty)} (y-x)^+\varphi''(dx) + \int_{(-\infty,0]} (x-y)^+\varphi''(dx), \tag{3.3}$$

where φ'_+ denotes the right-hand derivative of φ. With $\bar{\mu} = \bar{\nu}$ and (3.2), we conclude that $\mu(\varphi) \leq \nu(\varphi)$. Let us finally check (3.3) for $y \geq 0$, the proof being analogous for $y < 0$. The second integral in the right-hand side vanishes and we have, using Fubini's theorem,

$$\begin{aligned}\int_{(0,+\infty)} (y-x)^+\varphi''(dx) &= \int_{x\in(0,+\infty)}\int_{z\in\mathbb{R}} 1_{\{x\leq z<y\}}dz\varphi''(dx) = \int_{z=0}^{y}\int_{x\in(0,z]}\varphi''(dx)dz \\ &= \int_{z=0}^{y}\varphi''((0,z])dz = \int_{z=0}^{y}(\varphi'_+(z) - \varphi'_+(0))dz = \varphi(y) - \varphi(0) - \varphi'_+(0)y.\end{aligned}$$

■

Remark 3.7. *In view of Proposition 3.5 and (3.1), for $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$,*

$$\begin{aligned}\mu \leq_{cx} \nu \Leftrightarrow \bar{\mu} = \bar{\nu} \text{ and } \forall x \in \mathbb{R}, &\int_{\mathbb{R}} (y-x)^+\mu(dy) \leq \int_{\mathbb{R}} (y-x)^+\nu(dy) \\ \Leftrightarrow \bar{\mu} = \bar{\nu} \text{ and } \forall x \in \mathbb{R}, &\int_{\mathbb{R}} (x-y)^+\mu(dy) \leq \int_{\mathbb{R}} (x-y)^+\nu(dy).\end{aligned}$$

Proposition 3.8. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$. The open set $\{u_\nu(x) > u_\mu(x)\}$ is the at most countable union of disjoint open intervals $(I_n)_{n\in N}$ (with $N \subset \mathbb{N}$). Moreover any $\pi \in \Pi_M(\mu,\nu)$ writes*

$$\pi(dx,dy) = \mu(dx)\pi_x(dy) = \sum_{n\in N} \mu|_{I_n}(dx)1_{\{\bar{I}_n\}}(y)\pi_x(dy) + 1_{\{u_\mu(x)=u_\nu(x)\}}\mu(dx)\delta_x(dy),$$

where $\bar{I}_n$ denotes the closed interval equal to the closure of I_n.

Remark 3.9. *This means that for each $n \in \mathbb{N}$, $1_{I_n}(x)\mu(dx)$ a.e., $\pi_x(\bar{I}_n) = 1$ and $1_{\{u_\mu(x)=u_\nu(x)\}}\mu(dx)$ a.e., $\pi_x = \delta_x$. In particular, we recover that the only element of $\Pi_M(\mu,\mu)$ is $\mu(dx)\delta_x(dy)$.*

The proof of Proposition 3.8 relies on the next lemmas.

Lemma 3.10. *Let $\eta \in \mathcal{P}_1(\mathbb{R})$ and $x \in \mathbb{R}$ be such that $u_\eta(x) = |x - \bar{\eta}|$. Then $\eta((-\infty,x))\eta((x,+\infty)) = 0$.*

Proof of Lemma 3.10: Let $\eta \in \mathcal{P}_1(\mathbb{R})$ be such that $\eta((-\infty,x))\eta((x,+\infty)) > 0$. Then $\eta_{<x}(dy) = \frac{1_{(-\infty,x)}(y)\eta(dy)}{\eta((-\infty,x))}$ and $\eta_{>x}(dy) = \frac{1_{(x,+\infty)}(y)\eta(dy)}{\eta((x,+\infty))}$ belong to $\mathcal{P}_1(\mathbb{R})$ and satisfy $\overline{\eta_{<x}} < x$ and $\overline{\eta_{>x}} > x$. We then have $\bar{\eta} = \eta((-\infty,x))\overline{\eta_{<x}} + \eta(\{x\})x + \eta((x,+\infty))\overline{\eta_{>x}}$ so that

$$|x - \bar{\eta}| = |\eta((-\infty,x))(x - \overline{\eta_{<x}}) + \eta((x,+\infty))(x - \overline{\eta_{>x}})|.$$

Since the first term in the absolute value in the right-hand side is positive and the second is negative,

$$|x - \bar{\eta}| < \eta((-\infty,x))(x - \overline{\eta_{<x}}) + \eta((x,+\infty))(\overline{\eta_{>x}} - x) = \int_{\mathbb{R}} |x-y|\eta(dy) = u_\eta(x).$$

By contraposition, we conclude that $u_\eta(x) = |x - \bar{\eta}| \Rightarrow \eta((-\infty,x))\eta((x,+\infty)) = 0$. ∎

Lemma 3.11. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$ be such that $\mu \leq_{cx} \nu$. Then*

$$u_\mu(z) = u_\nu(z) \Rightarrow \forall \pi \in \Pi_M(\mu,\nu),\ \pi\left((-\infty,z) \times (z+\infty)\right) = \pi\left((z+\infty) \times (-\infty,z)\right) = 0,$$

i.e. z is a barrier for the martingale couplings between μ and ν.

Proof of Lemma 3.11: Let us suppose that $u_\mu(z) = u_\nu(z)$ and choose some $\pi \in \Pi_M(\mu,\nu)$. We then have, using the consequence $\int_{\mathbb{R}} |z-y|\pi_x(dy) \geq |z - \bar{\pi}_x|$ of Jensen's inequality then the martingale property of π that

$$\begin{aligned} u_\nu(z) &= \int_{\mathbb{R}\times\mathbb{R}} |z-y|\pi(dx,dy) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |z-y|\pi_x(dy)\right)\mu(dx) \geq \int_{\mathbb{R}} |z-\bar{\pi}_x|\mu(dx) \\ &= \int_{\mathbb{R}} |z-x|\mu(dx) = u_\mu(z) = u_\nu(z). \end{aligned}$$

Therefore $\mu(dx)$ a.e. $\int_{\mathbb{R}} |z-y|\pi_x(dy) = |z - \bar{\pi}_x|$ and, by Lemma 3.10, $\pi_x((-\infty,z))\pi_x((z,+\infty)) = 0$. Since $\mu(dx)$ a.e. $\bar{\pi}_x = x$, we conclude that $1_{\{x<z\}}\mu(dx)$ a.e. $\pi_x((z,+\infty)) = 0$ so that $\pi((-\infty,z)\times(z,+\infty)) = 0$ and $1_{\{x>z\}}\mu(dx)$ a.e. $\pi_x((-\infty,z)) = 0$ so that $\pi((z,+\infty)\times(-\infty,z)) = 0$. We conclude that

$$\begin{aligned} \pi\left((-\infty,z)\times(z,+\infty)\right) = \int_{\mathbb{R}} \pi_x((z,+\infty))1_{\{x<z\}}\mu(dx) = 0 &= \int_{\mathbb{R}} \pi_x((-\infty,z))1_{\{x>z\}}\mu(dx) \\ &= \pi\left((z,+\infty)\times(-\infty,z)\right). \end{aligned}$$

∎

Proof of Proposition 3.8: Any open subset of the real line writes as an at most countable union of disjoint open intervals. Moreover, when finite, the boundaries of these intervals belong to the complementary of the open set. In particular, the boundaries $\ell_n < r_n$ of the intervals I_n are such that $\ell_n > -\infty \Rightarrow u_\mu(\ell_n) = u_\nu(\ell_n)$ and $r_n < +\infty \Rightarrow u_\mu(r_n) = u_\nu(r_n)$. With Lemma 3.11, this implies that for $\pi \in \Pi_M(\mu,\nu)$, $1_{I_n}(x)\mu(dx)$ a.e., $\pi_x(\bar{I}_n) = 1$. For $z \in \mathbb{R}$ such that $z \le \ell_n$ and $\pi \in \Pi_M(\mu,\nu)$, we deduce that

$$\begin{aligned}\int_{\mathbb{R}} |z-x|1_{I_n}(x)\mu(dx) &= \int_{\mathbb{R}}(x-z)1_{I_n}(x)\mu(dx) = \int_{\mathbb{R}}(\bar{\pi}_x - z)1_{I_n}(x)\mu(dx)\\ &= \int_{\mathbb{R}^2}(y-z)\pi_x(dy)1_{I_n}(x)\mu(dx) = \int_{\mathbb{R}^2}|z-y|\pi_x(dy)1_{I_n}(x)\mu(dx)\end{aligned}$$

and, by symmetry, the equality between the extreme sides is preserved when $z \in \mathbb{R}$ is such that $z \ge r_n$. We deduce that when $z \in \mathbb{R}$ is such that $u_\mu(z) = u_\nu(z)$ i.e. such that $z \in \left\{\bigcup_{n\in N} I_n\right\}^c$, then

$$\begin{aligned}\int_{\mathbb{R}} |z-x|1_{\{u_\nu(x)>u_\mu(x)\}}\mu(dx) &= \sum_{n\in N}\int_{\mathbb{R}}|z-x|1_{I_n}(x)\mu(dx)\\ = \sum_{n\in N}\int_{\mathbb{R}^2}|z-y|\pi_x(dy)1_{I_n}(x)\mu(dx) &= \int_{\mathbb{R}^2}|z-y|\pi_x(dy)1_{\{u_\nu(x)>u_\mu(x)\}}\mu(dx)\end{aligned}$$

so that

$$\begin{aligned}\int_{\mathbb{R}} |z-x|1_{\{u_\nu(x)=u_\mu(x)\}}\mu(dx) &= u_\nu(z) - \int_{\mathbb{R}}|z-x|1_{\{u_\nu(x)>u_\mu(x)\}}\mu(dx)\\ &= \int_{\mathbb{R}^2}|z-y|\pi_x(dy)\mu(dx) - \int_{\mathbb{R}^2}|z-y|\pi_x(dy)1_{\{u_\nu(x)>u_\mu(x)\}}\mu(dx)\\ &= \int_{\mathbb{R}^2}|z-y|\pi_x(dy)1_{\{u_\nu(x)=u_\mu(x)\}}\mu(dx).\end{aligned}$$

Therefore, when $\mu(\{u_\mu = u_\nu\}) > 0$, $\frac{1_{\{u_\nu(x)=u_\mu(x)\}}}{\mu(\{u_\mu=u_\nu\})}\mu(dx)\pi_x(dy)$ is a martingale coupling between two probability measures sharing the same potential functions and therefore equal by Remark 3.6. By the second point in Remark 3.2, we conclude that $1_{\{u_\nu(x)=u_\mu(x)\}}\mu(dx)$ a.e., $\pi_x = \delta_x$. ∎

Lemma 3.12. *Let $\rho \ge 1$ and $\mu,\nu \in \mathcal{P}_\rho(\mathbb{R}^d)$. Then $\Pi(\mu,\nu)$ is convex and compact in $\mathcal{P}_\rho(\mathbb{R}^d\times\mathbb{R}^d)$.*
When, moreover $\mu \le_{cx} \nu$, then $\Pi_M(\mu,\nu)$ is convex and compact in $\mathcal{P}_\rho(\mathbb{R}^d\times\mathbb{R}^d)$.
Last, if for $n \in \mathbb{N}$, $\pi_n \in \Pi_M(\mu_n,\nu_n)$ with $\mu_n \le_{cx} \nu_n$ and μ_n (resp. ν_n) converges to μ (resp. ν) in $\mathcal{P}_\rho(\mathbb{R}^d)$ as $n\to\infty$, then $\mu \le_{cx} \nu$ and we can extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ which converges to some $\pi \in \Pi_M(\mu,\nu)$ in $\mathcal{P}_\rho(\mathbb{R}^d\times\mathbb{R}^d)$.

Proof: By the equivalence of norms on $\mathbb{R}^d\times\mathbb{R}^d$ and Proposition 1.41, the definition of the convergence in $\mathcal{P}_\rho(\mathbb{R}^d\times\mathbb{R}^d)$ does not depend on the choice of the norm inducing the distance $d_{\mathbb{R}^d\times\mathbb{R}^d}$. When we choose $|(x,y)| = \left(\sum_{i=1}^d(|x_i|^\rho + |y_i|^\rho)\right)^{1/\rho}$ so that

$$d^\rho_{\mathbb{R}^d\times\mathbb{R}^d}((\tilde{x},\tilde{y}),(x,y)) = \sum_{i=1}^d |\tilde{x}_i - x_i|^\rho + |\tilde{y}_i - y_i|^\rho,$$

we see that the convergence in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$ amounts to the weak convergence and the respective convergence of the first and second marginals in $\mathcal{P}_\rho(\mathbb{R}^d)$. With Lemma 1.7, we deduce that when $\mu, \nu \in \mathcal{P}_\rho(\mathbb{R}^d)$, $\Pi(\mu, \nu)$ is compact in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$.

When $\mu \leq_{cx} \nu$ and $\pi \in \Pi_M(\mu, \nu)$, we have for each $h : \mathbb{R}^d \to \mathbb{R}^d$ measurable bounded,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} h(x).(y-x)\pi(dx,dy) = \int_{\mathbb{R}^d} h(x).\int_{\mathbb{R}^d}(y-x)\pi_x(dy)\mu(dx) = \int_{\mathbb{R}^d} h(x).(\bar{\pi}_x - x)\mu(dx)$$
$$= \int_{\mathbb{R}^d} h(x).0\mu(dx) = 0.$$

Let us conversely suppose that this equality holds for each $h : \mathbb{R}^d \to \mathbb{R}^d$ continuous and bounded. For $i \in \{1, \cdots, d\}$, the function $\mathbb{R}^d \ni x \mapsto 1_{\{(\bar{\pi}_x)_i > x_i\}} - 1_{\{(\bar{\pi}_x)_i < x_i\}}$ belongs to $L^1(|x - \bar{\pi}_x|\mu(dx))$. By density of $C_b(\mathbb{R}^d)$ in $L^1(|x - \bar{\pi}_x|\mu(dx))$, we deduce that

$$\int_{\mathbb{R}^d} \sum_{i=1}^d |x_i - (\bar{\pi}_x)_i|\mu(dx) = 0$$

so that $\mu(dx)$ a.e., $\bar{\pi}_x = x$ and $\pi \in \Pi_M(\mu, \nu)$. Therefore

$$\pi \in \Pi_M(\mu,\nu) \Leftrightarrow \forall h : \mathbb{R}^d \to \mathbb{R}^d \text{ continuous and bounded, } \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x).(y-x)\pi(dx,dy) = 0 \tag{3.4}$$

If $\mathcal{P}_\rho^M(\mathbb{R}^d \times \mathbb{R}^d) = \left\{\tilde{\pi}(dx,dy) = \tilde{\mu}(dx)\tilde{\pi}_x(dy) \in \mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d) : \tilde{\mu}(dx) \text{ a.e., } \bar{\tilde{\pi}}_x = x\right\}$ denotes the subset of $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$ consisting in martingale couplings, then we have

$$\mathcal{P}_\rho^M(\mathbb{R}^d \times \mathbb{R}^d) = \bigcap_{h \in C_b(\mathbb{R}^d,\mathbb{R}^d)} \left\{\pi \in \mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d) : \pi(H) = 0\right\}. \tag{3.5}$$

For $h \in C_b(\mathbb{R}^d, \mathbb{R}^d)$, the function $\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto h(x).(y - x)$ is continuous with affine growth i.e. belongs to $C_1(\mathbb{R}^d \times \mathbb{R}^d) \subset C_\rho(\mathbb{R}^d \times \mathbb{R}^d)$. By Proposition 1.41, we deduce that $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d) \ni \pi \mapsto \pi(H)$ is continuous. Therefore $\left\{\pi \in \mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d) : \pi(H) = 0\right\}$ is a closed and convex (the equality constraint involves $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d) \ni \pi \mapsto \pi(H)$ which is linear) subset of $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$. With (3.5), we deduce that $\mathcal{P}_\rho^M(\mathbb{R}^d \times \mathbb{R}^d)$ is a convex and closed subset of $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$. Moreover, $\Pi_M(\mu, \nu) = \Pi(\mu, \nu) \cap \mathcal{P}_\rho^M(\mathbb{R}^d \times \mathbb{R}^d)$, which is non empty iff $\mu \leq_{cx} \nu$ according to Strassen's theorem (see Theorem 2.16) is convex and compact in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$.

Let finally for $n \in \mathbb{N}$, $\pi_n \in \Pi_M(\mu_n, \nu_n)$ with $\mu_n \leq_{cx} \nu_n$ and μ_n (resp. ν_n) converging to μ (resp. ν) in $\mathcal{P}_\rho(\mathbb{R}^d)$ as $n \to \infty$. For $\varphi : \mathbb{R}^d \to \mathbb{R}$ Lipschitz continuous and convex, we have $\lim_{n\to\infty} \mu_n(\varphi) = \mu(\varphi)$ and $\lim_{n\to\infty} \nu_n(\varphi) = \nu(\varphi)$. By taking the limit $n \to \infty$ in the inequality $\mu_n(\varphi) \leq \nu_n(\varphi)$ consequence of $\mu_n \leq_{cx} \nu_n$, we obtain $\mu(\varphi) \leq \nu(\varphi)$. With Remark 2.19, we deduce that $\mu \leq_{cx} \nu$. By Lemma 1.7, we can extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ converging weakly to some $\pi_\infty \in \Pi(\mu, \nu)$ as $k \to \infty$. Since the convergence in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$ amounts to the weak convergence and the respective convergence of the first and second marginals in $\mathcal{P}_\rho(\mathbb{R}^d)$, π_{n_k} converges to π in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$. By closedness of $\mathcal{P}_\rho^M(\mathbb{R}^d \times \mathbb{R}^d)$ in $\mathcal{P}_\rho(\mathbb{R}^d \times \mathbb{R}^d)$, $\pi_\infty \in \Pi_M(\mu, \nu)$. ∎

Corollary 3.13. *Let c be lower semi-continuous and bounded from below. Then for $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ such that $\mu \leq_{cx} \nu$, there exists $\pi_\star \in \Pi_M(\mu, \nu)$ such that $V_c^M(\mu, \nu) = \pi_\star(c)$. Moreoever, V_c^M is lower semi-continuous on $\left\{(\mu, \nu) \in \mathcal{P}_1(\mathbb{R}^d) \times \mathcal{P}_1(\mathbb{R}^d) : \mu \leq_{cx} \nu\right\}$.*

Remark 3.14. *According to the proof of Lemma 3.12,* $\{(\mu,\nu)\in\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d):\mu\leq_{cx}\nu\}$ *is a closed subset of* $\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)$. *Since* V_c^M *is equal to* $+\infty$ *on its open complementary, we deduce that* V_c^M *is lower semi-continuous on* $\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)$.

Proof: Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^d)$ be such that $\mu\leq_{cx}\nu$. Let $(\pi_n)_{n\in\mathbb{N}}\subset\Pi_M(\mu,\nu)$ be a minimizing sequence for $V_c^M(\mu,\nu)$. By Lemma 3.12, we can extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ which converges weakly to some $\pi_\star\in\Pi_M(\mu,\nu)$ as $k\to\infty$. By the Portmanteau theorem (see Theorem 5.2 7)),

$$\pi_\star(c)=\liminf_{k\to\infty}\pi_{n_k}(c)=V_c^M(\mu,\nu),$$

so that $\pi_\star$ is optimal for $V_c^M(\mu,\nu)$.

Let now $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{P}_1(\mathbb{R}^d)$ with $\mu_n\leq_{cx}\nu_n$ converging respectively to μ and ν in $\mathcal{P}_1(\mathbb{R}^d)$ as $n\to\infty$ and such that

$$\lim_{n\to\infty}V_c^M(\mu_n,\nu_n)=\liminf_{(\tilde\mu,\tilde\nu)\overset{\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)}{\longrightarrow}(\mu,\nu):\tilde\mu\leq_{cx}\tilde\nu}V_c^M(\tilde\mu,\tilde\nu).$$

By Lemma 3.12, we can extract a subsequence $(\pi_{n_k})_{k\in\mathbb{N}}$ which converges in $\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d)$ to $\pi\in\Pi_M(\mu,\nu)$. By the Portmanteau theorem (see Theorem 5.2 7)), we conclude that

$$V_c^M(\mu,\nu)\leq\pi(c)=\lim_{k\to\infty}\pi_{n_k}(c)=\liminf_{(\tilde\mu,\tilde\nu)\overset{\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d)}{\longrightarrow}(\mu,\nu):\tilde\mu\leq_{cx}\tilde\nu}V_c^M(\tilde\mu,\tilde\nu).$$

∎

Remark 3.15. *Corollary 3.13 also is a consequence of Theorem 2.2 and the discussion at the beginning of Chapter 2 explaining the the MOT problem is a special case of the WOT problem.*

Example 3.16. *When* $d\geq 2$, *according to [7]* V_c^M *may fail to be continuous on* $\{(\mu,\nu)\in\mathcal{P}_1(\mathbb{R}^d)\times\mathcal{P}_1(\mathbb{R}^d):\mu\leq_{cx}\nu\}$ *even for continuous and bounded cost functions c. Indeed, let* $d=2$,

$$\mu=\frac{1}{2}\left(\delta_{(1,0)}+\delta_{(2,0)}\right)\ \text{and,}$$
$$\text{for }\theta\in[0,\pi),\ \nu_\theta=\frac{1}{4}\left(\delta_{(1+\cos\theta,\sin\theta)}+\delta_{(1-\cos\theta,-\sin\theta)}+\delta_{(2+\cos\theta,\sin\theta)}+\delta_{(2-\cos\theta,-\sin\theta)}\right).$$

For $\theta\in(0,\pi)$,

$$\frac{1}{4}\left(\delta_{((1,0),(1+\cos\theta,\sin\theta))}+\delta_{((1,0),(1-\cos\theta,-\sin\theta))}+\delta_{((2,0),(2+\cos\theta,\sin\theta))}+\delta_{((2,0),(2-\cos\theta,-\sin\theta))}\right)$$

is the only element of $\Pi_M(\mu,\nu_\theta)$ *so that* $V_c^M(\mu,\nu_\theta)=1$ *when* $c(x,y)=|x-y|\wedge 1$. *For* $\theta=0$, $\nu_0=\frac{1}{4}\left(\delta_{(0,0)}+\delta_{(1,0)}+\delta_{(2,0)}+\delta_{(3,0)}\right)$ *and*

$$\frac{1}{4}\left(\delta_{((1,0),(1,0))}+\delta_{((2,0),(2,0))}\right)+\frac{1}{6}\left(\delta_{((1,0),(0,0))}+\delta_{((2,0),(3,0))}\right)+\frac{1}{12}\left(\delta_{((1,0),(3,0))}+\delta_{((2,0),(0,0))}\right)$$

belongs to $\Pi_M(\mu,\nu_0)$ *so that*

$$V_c(\mu,\nu_0)\leq\frac{1}{4}\times(0+0)+\frac{1}{6}\times(1+1)+\frac{1}{12}\times(2+2)=\frac{2}{3}<1=\lim_{\theta\to 0+}V_c^M(\mu,\nu_\theta).$$

Remark 3.17. *In dimension $d = 1$, according to [4], it is possible to approximate any coupling $\pi \in \Pi_M(\mu, \nu)$ by a sequence of couplings $\pi_n \in \Pi_M(\mu_n, \nu_n)$ and check the continuity of V_c^M on $\{(\mu, \nu) \in \mathcal{P}_1(\mathbb{R}) \times \mathcal{P}_1(\mathbb{R}) : \mu \leq_{cx} \nu\}$ when the cost function c is continuous and bounded.*

Let us now give a dual formulation of the MOT problem.

Theorem 3.18. *Let $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous and bounded from below. Then for all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$,*

$$V_c^M(\mu, \nu) = \sup\{\mu(\phi) + \nu(\psi) : \phi, \psi \in C_1(\mathbb{R}^d),$$
$$\exists h \in C_b(\mathbb{R}^d, \mathbb{R}^d),\ \phi(x) + \psi(y) + h(x).(y - x) \leq c(x, y),\ \forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^d\}.$$

Remark 3.19.
- *Note that the couples $(\phi, \psi) \in C_b(\mathbb{R}^d) \times C_b(\mathbb{R}^d)$ such that $\phi \oplus \psi \leq c$ appearing in the dual formulation of the OT problem $V_c(\mu, \nu)$ given in Theorem 1.8 are admissible in the dual formulation of $V_c^M(\mu, \nu)$ so that $V_c^M(\mu, \nu) \geq V_c(\mu, \nu)$, an equality obvious from the primal formulations and the inclusion $\Pi_M(\mu, \nu) \subset \Pi(\mu, \nu)$.*
- *Returning to the original financial motivation, we see that the triplets (ϕ, ψ, h) which appear in the dual formulation provide the following model-free sub-hedging strategy for the option with exotic payoff $c(S_1, S_2)$ and maturity 2 :*
  - *with the initial wealth $\mu(\phi) + \nu(\psi)$ buy at time 0 an option with payoff $\phi(S_1)$ and maturity 1 and an option with payoff $\psi(S_2)$ and maturity 2,*
  - *buy at time 1, $h(S_1)$ units of the risky assets and invest $-h(S_1).S_1 + \phi(S_1)$ in the risk-free asset,*
  - *sell at time 2, the $h(S_1)$ units of the risky assets.*

Indeed, the terminal wealth is $-h(S_1).S_1 + \phi(S_1) + \psi(S_2) + h(S_1).S_2 \leq c(S_1, S_2)$.

The proof relies on the following minimax result (see Theorem 2.4.1 [2]).

Theorem 3.20. *Let K be a compact convex subset of a Hausdorff topological vector space, Y be a convex subset of an arbitrary vector space and $K \times Y \ni (x, y) \mapsto h(x, y) \in \mathbb{R}$ be convex lower semi-continuous in x on K for each fixed $y \in Y$ and concave in y on Y for each fixed $x \in K$. Then*

$$\min_{x\in K} \sup_{y\in Y} h(x, y) = \sup_{y\in Y} \min_{x\in K} h(x, y).$$

Remark 3.21. *For fixed $y \in Y$, the infimum on the compact set K of the lower semi-continuous function $x \mapsto h(x, y)$ is attained. Since $x \mapsto \sup_{y\in Y} h(x, y)$ also is lower semi-continuous as the supremum of lower semi-continuous functions, its infimum over K also is attained. Note that the inequality $\min_{x\in K} \sup_{y\in Y} h(x, y) \geq \sup_{y\in Y} \min_{x\in K} h(x, y)$ is obvious.*

Proof: Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ and $\pi \in \Pi(\mu, \nu) \setminus \Pi_M(\mu, \nu)$. For $\alpha \in \mathbb{R}_+$ and $i \in \{1, \cdots, d\}$, approximating in $L^1(|x - \bar{\pi}_x|\mu(dx))$ the function $\mathbb{R}^d \ni x \mapsto \alpha\left(1_{\{x_i > (\bar{\pi}_x)_i\}} - 1_{\{x_i < (\bar{\pi}_x)_i\}}\right)$ by functions in $C_b(\mathbb{R}^d)$ we check that

$$\sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x).(x - y)\pi(dx, dy) \geq \alpha \int_{\mathbb{R}^d} \sum_{i=1}^{d} |x_i - (\bar{\pi}_x)_i|\mu(dx).$$

Letting $\alpha \to \infty$, we conclude that $\sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \int_{\mathbb{R}^d\times\mathbb{R}^d} h(x).(x-y)\pi(dx,dy) = +\infty$. On the other hand, when $\pi \in \Pi_M(\mu,\nu)$, $\int_{\mathbb{R}^d\times\mathbb{R}^d} h(x).(y-x)\pi(dx,dy) = 0$ for each $h \in C_b(\mathbb{R}^d,\mathbb{R}^d)$. Therefore, setting $H(x,y) = h(x).(y-x)$, we have

$$V_c^M(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \pi(c-H).$$

Note that for $\pi \in \mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d)$ and $h \in C_b(\mathbb{R}^d,\mathbb{R}^d)$ so that $H \in C_1(\mathbb{R}^d\times\mathbb{R}^d)$, we have $H \in L^1(\pi)$. The cost function c being bounded from below, it is semi-integrable with respect to π and $\pi(c-H) = \pi(c) - \pi(H)$. For fixed $\pi \in \Pi(\mu,\nu)$, $C_b(\mathbb{R}^d,\mathbb{R}^d) \ni h \mapsto \pi(c-H) = \pi(c) - \pi(H)$ is affine and therefore concave. On the other hand, by Lemma 3.12, $\Pi(\mu,\nu)$ is a compact subset of $\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d)$ and, for fixed $h \in C_b(\mathbb{R}^d,\mathbb{R}^d)$, $\mathcal{P}_1(\mathbb{R}^d\times\mathbb{R}^d) \ni \pi \mapsto \pi(c) - \pi(H)$ is lower semi-continuous as the sum of the lower semi-continuous function $\pi(c)$ (see Theorem 5.2 7)) and the continuous function $-\pi(H)$ (see Proposition 1.41). By Theorem 3.20, we deduce that

$$V_c^M(\mu,\nu) = \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \inf_{\pi\in\Pi(\mu,\nu)} \pi(c-H).$$

Let $\tilde{H}(x,y) = \overline{|h|}(|x|+|y|)$ where $\overline{|h|} = \sup_{z\in\mathbb{R}^d} |h(z)|$. We have

$$V_c^M(\mu,\nu) = \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \left\{ \inf_{\pi\in\Pi(\mu,\nu)} \pi(c+\tilde{H}-H) - \overline{|h|}\,(\mu(|\cdot|)+\nu(|\cdot|)) \right\}.$$

The cost function $\tilde{c} = c + \tilde{H} - H$ being lower semi-continuous and bounded from below by the same constant as c, the duality Theorem for OT (see Theorem 1.8) ensures that

$$V_{\tilde{c}}(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \pi(\tilde{c}) = \sup_{\substack{(\tilde{\phi},\tilde{\psi})\in C_b(\mathbb{R}^d)\times C_b(\mathbb{R}^d)\\ \tilde{\phi}\oplus\tilde{\psi}\le\tilde{c}}} \{\mu(\tilde{\phi}) + \nu(\tilde{\psi})\}.$$

Therefore, using that $C_b(\mathbb{R}^d) \subset C_1(\mathbb{R}^d)$ for the first inequality then that $\phi = \tilde{\phi} - \overline{|h|}|\cdot|$ belongs to $C_1(\mathbb{R}^d)$ when $\tilde{\phi} \in C_1(\mathbb{R}^d)$, we obtain

$$\begin{aligned} V_c^M(\mu,\nu) &= \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \sup_{\substack{(\tilde{\phi},\tilde{\psi})\in C_b(\mathbb{R}^d)\times C_b(\mathbb{R}^d)\\ \tilde{\phi}(x)+\tilde{\psi}(y)\le c(x,y)+\overline{|h|}(|x|+|y|)-H(x,y)}} \{\mu(\tilde{\phi} - \overline{|h|}|\cdot|) + \nu(\tilde{\psi} - \overline{|h|}|\cdot|)\} \\ &\le \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \sup_{\substack{(\tilde{\phi},\tilde{\psi})\in C_1(\mathbb{R}^d)\times C_1(\mathbb{R}^d)\\ \tilde{\phi}(x)-\overline{|h|}|x|+\tilde{\psi}(y)-\overline{|h|}|y|\le c(x,y)-H(x,y)}} \{\mu(\tilde{\phi} - \overline{|h|}|\cdot|) + \nu(\tilde{\psi} - \overline{|h|}|\cdot|)\} \\ &\le \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \sup_{\substack{(\phi,\psi)\in C_1(\mathbb{R}^d)\times C_1(\mathbb{R}^d)\\ \phi\oplus\psi\le c-H}} \{\mu(\phi)+\nu(\psi)\}. \end{aligned}$$

Of course, these inequalities are equalities when $V_c^M(\mu,\nu) = +\infty$. Otherwise $\Pi^M(\mu,\nu) \neq \emptyset$ and for $\pi \in \Pi^M(\mu,\nu)$, $h \in C_b(\mathbb{R}^d,\mathbb{R}^d)$ and $(\phi,\psi) \in C_1(\mathbb{R}^d)\times C_1(\mathbb{R}^d)$ such that $\phi\oplus\psi \le c - H$, then $\pi(c) = \pi(c-H) \ge \pi(\phi\oplus\psi) = \mu(\phi) + \nu(\psi)$. By taking the infimum over π and the supremum over (ϕ,ψ) in this inequality, we conclude that

$$V_c^M(\mu,\nu) = \sup_{h\in C_b(\mathbb{R}^d,\mathbb{R}^d)} \sup_{\substack{(\phi,\psi)\in C_1(\mathbb{R}^d)\times C_1(\mathbb{R}^d)\\ \phi\oplus\psi\le c-H}} \{\mu(\phi)+\nu(\psi)\}.$$

∎

Chapter 4

Entropic Optimal Transport

The Entropic Optimal Transport problem consists in solving

$$V_{c,\varepsilon}(\mu,\nu) = \inf_{\pi\in\Pi(\mu,nu)} \left(\pi(c) + \varepsilon H(\pi|\mu\otimes\nu)\right),$$

where $c : \mathcal{X}\times\mathcal{Y}\to\mathbb{R}$ is a cost function, $\varepsilon > 0$ and

$$H(\pi|\mu\otimes\nu) = \begin{cases} \int_{\mathcal{X}\times\mathcal{Y}} \ln\left(\frac{d\pi}{d\mu\otimes\nu}(x,y)\right)\pi(dx,dy) \text{ if } \pi << \mu\otimes\nu \\ +\infty \text{ otherwise} \end{cases},$$

denotes the relative entropy of the coupling π with respect to $\mu\otimes\nu(dx,dy) = \mu(dx)\nu(dy)$. In this chapter, we will see that the Entropic Optimal Transport problem is an approximation of the classical Optimal Transport problem in the sense that $V_{c,\varepsilon}(\mu,\nu)$ converges to $V_{c,0}(\mu,\nu)$ as $\varepsilon\to 0$ and present the Sinkhorn algorithm which permits to solve efficiently this approximate problem.

4.1 Relative Entropy, primal and dual formulations

For $\mu,\nu\in\mathcal{P}(\mathcal{X})$, we say that μ is absolutely continuous with respect to ν and denote $\mu << \nu$ if $\nu(A) = 0 \Rightarrow \mu(A) = 0$ for each Borel subset A of $\mathcal{X}$. The relative entropy of μ with respect to ν is defined by

$$H(\mu|\nu) = \begin{cases} \mu\left(\ln\left(\frac{d\mu}{d\nu}\right)\right) = \nu\left(\frac{d\mu}{d\nu}\ln\left(\frac{d\mu}{d\nu}\right)\right) \text{ if } \mu << \nu \\ +\infty \text{ otherwise} \end{cases}.$$

Note that for $f(y) = y\ln(y)$, $f'(y) = \ln(y)+1$, $f''(y) = \frac{1}{y}$ and $\inf_{y\geq 0} f(y) = f(1/e) = -1/e$. This ensures that when $\mu << \nu$, $-1/e \leq \frac{d\mu}{d\nu}\ln\left(\frac{d\mu}{d\nu}\right)$ and the integral of the right-hand side with respect to ν makes sense. Moreover, if $H(\mu|\nu) < \infty$, then by strict convexity of f and Jensen's inequality,

$$H(\mu|\nu) = \nu\left(f\left(\frac{d\mu}{d\nu}\right)\right) \geq f\left(\nu\left(\frac{d\mu}{d\nu}\right)\right) = f(1) = 0 \tag{4.1}$$

with equality iff $\frac{d\mu}{d\nu}$ is ν a.e. constant i.e. $\mu = \nu$.

Proposition 4.1. *Let $c \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y})$. Then for each $\varepsilon > 0$, there exists a unique $\pi_\varepsilon \in \Pi(\mu, \nu)$ optimal for $V_{c,\varepsilon}(\mu, \nu)$.*

The proof relies on the next lemma which gives a variational formulation of the relative entropy.

Lemma 4.2.

$$\forall \mu, \nu \in \mathcal{P}(\mathcal{X}),\ H(\mu|\nu) = \sup_{f \in C_b(\mathcal{X})} \left\{ \mu(f+1) - \nu(e^f) \right\}, \tag{4.2}$$

and $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \ni (\mu, \nu) \mapsto H(\mu|\nu)$ is convex and lower semi-continuous for the product of the weak convergence topology.

Proof of Proposition 4.1: Let $R_\varepsilon \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be defined by $\frac{dR_\varepsilon}{d\mu \otimes \nu} = \frac{e^{-\frac{c}{\varepsilon}}}{\mu \otimes \nu (e^{-\frac{c}{\varepsilon}})}$. The probability measure R_ε is equivalent to $\mu \otimes \nu$ and $\pi << \mu \otimes \nu \Leftrightarrow \pi << R_\varepsilon$. When this holds, $\frac{d\pi}{dR_\varepsilon} = \mu \otimes \nu(e^{-\frac{c}{\varepsilon}}) e^{\frac{c}{\varepsilon}} \frac{d\pi}{d\mu \otimes \nu}$ so that

$$H(\pi|R_\varepsilon) = \ln\left(\mu \otimes \nu(e^{-\frac{c}{\varepsilon}})\right) + \frac{1}{\varepsilon}\left(\pi(c) + \varepsilon H(\pi|\mu \otimes \nu)\right). \tag{4.3}$$

The first term in the right-hand side is some constant not depending on the coupling π. Hence $\pi \in \Pi(\mu, \nu)$ is optimal for $V_{c,\varepsilon}(\mu, \nu)$ if and only if it is optimal for $\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R_\varepsilon)$. Let us check that there is a unique $\pi_\varepsilon \in \Pi(\mu, \nu)$ optimal for the latter problem. Note that $\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R_\varepsilon) \leq H(\mu \otimes \nu|R_\varepsilon) = \ln\left(\mu \otimes \nu(e^{-\frac{c}{\varepsilon}})\right) + \frac{1}{\varepsilon}\mu \otimes \nu(c) < \infty$. Lemma 1.7 ensures that from any minimizing sequence in $\Pi(\mu, \nu)$ for $\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R_\varepsilon)$, we may extract a subsequence $(\pi_n)_n$ converging weakly to some $\pi_\infty \in \Pi(\mu, \nu)$. Since, by Lemma 4.2, $\pi \mapsto H(\pi|R_\varepsilon)$ is lower semi-continuous, we deduce that π_∞ is optimal for $\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R_\varepsilon)$. For distinct $\pi, \tilde{\pi} \in \Pi(\mu, \nu)$ such that $H(\pi|R_\varepsilon) + H(\tilde{\pi}|R_\varepsilon) < \infty$, setting $g = \frac{d\pi}{dR_\varepsilon}$ and $\tilde{g} = \frac{d\tilde{\pi}}{dR_\varepsilon}$, we have, by strict convexity of $\mathbb{R}_+ \ni y \mapsto y \ln y$,

$$\begin{aligned} H\left(\frac{\pi + \tilde{\pi}}{2}|R_\varepsilon\right) &= R_\varepsilon\left(\frac{1}{2}(g + \tilde{g}) \ln\left(\frac{1}{2}(g + \tilde{g})\right)\right) \\ &< R_\varepsilon\left(\frac{1}{2}(g \ln(g) + \tilde{g} \ln(\tilde{g}))\right) = \frac{1}{2}(H(\pi|R_\varepsilon) + H(\tilde{\pi}|R_\varepsilon)). \end{aligned}$$

Since $\frac{\pi + \tilde{\pi}}{2} \in \Pi(\mu, \nu)$ by convexity of this set, this ensures uniqueness of optimizers for $\inf_{\pi \in \Pi(\mu,\nu)} H(\pi|R_\varepsilon)$.

∎

Proof of Lemma 4.2: Before proving (4.2), let us check that this variational formulation implies the second assertion. For fixed $f \in C_b(\mathcal{X})$, $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \ni (\mu, \nu) \mapsto \mu(f+1) - \nu(e^f)$ is linear and continuous. We deduce that $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X}) \ni (\mu, \nu) \mapsto H(\mu|\nu)$ is convex (resp. lower semi-continuous) as the supremum of convex (resp. lower semi-continuous) functions. Let us now prove that

$$\sup_{f \in C_b(\mathcal{X})} \left\{ \mu(f+1) - \nu(e^f) \right\} = \sup_{f \in \mathcal{M}_b(\mathcal{X})} \left\{ \mu(f+1) - \nu(e^f) \right\}.$$

Since $C_b(\mathcal{X}) \subset \mathcal{M}_b(\mathcal{X})$, the right-hand side is not smaller than the left-hand side and it is enough to check that for fixed $f \in \mathcal{M}_b(\mathcal{X})$ bounded by M_f, $\mu(f+1) - \nu(e^f) \leq \sup_{g \in C_b(\mathcal{X})} \{\mu(g+1) - \nu(e^g)\}$. By density of $C_b(\mathcal{X})$ in $L^1(\mu+\nu)$, there exists a sequence $(f_n)_{n\in\mathbb{N}}$ in $C_b(\mathcal{X})$ converging to f in $L^1(\mu+\nu)$. Since $\|f - (-M_f) \vee f_n \wedge M_f\|_{L^1(\mu+\nu)} \leq \|f - f_n\|_{L^1(\mu+\nu)}$, we may suppose that each function f_n is bounded by M_f. We extract a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that as $k \to \infty$, $f_{n_k} \to f$, $\mu+\nu$ a.e.. By dominated convergence, $\mu(f+1) - \nu(e^f) = \lim_{k\to\infty} \mu(f_{n_k}+1) - \nu(e^{f_{n_k}}) \leq \sup_{g\in C_b(\mathcal{X})} \{\mu(g+1) - \nu(e^g)\}$.

- When μ is not absolutely continuous with respect to ν, then there exists $A \in \mathcal{B}(\mathcal{X})$ such that $\mu(A) > 0$ and $\nu(A) = 0$. Choosing $f = c1_A$, we get $\mu(f+1) - \nu(e^f) = c\mu(A) + 1 - 0$ and letting $c \to \infty$, we deduce that

$$\sup_{f\in\mathcal{M}_b(\mathcal{X})} \left\{\mu(f+1) - \nu(e^f)\right\} = +\infty = H(\mu|\nu).$$

- When $\mu << \nu$, we set $g(x) = \frac{d\mu}{d\nu}(x)$. For $y > 0$, $f_y(x) = yx - e^x$ is strictly concave and maximal equal to $f_y(\ln(y)) = y\ln(y) - y$ when $f_y'(x) = 0$ i.e. $y = e^x$ i.e. $x = \ln(y)$. We deduce that for $y > 0$

$$y\ln(y) = \sup_{x\in\mathbb{R}}\{yx - e^x\} + y, \tag{4.4}$$

an equality which remains valid when $y = 0$. Hence for $f \in \mathcal{M}_b(\mathcal{X})$, $g\ln(g) \geq g + fg - e^f$ so that

$$H(\mu|\nu) = \nu(g\ln(g)) \geq \nu(g + fg - e^f) = \mu(1+f) - \nu(e^f).$$

We deduce that $H(\mu|\nu) \geq \sup_{f\in\mathcal{M}_b(\mathcal{X})} \left\{\mu(f+1) - \nu(e^f)\right\}$.
On the other hand, let for $n \in \mathbb{N}^*$, $f_n = \ln\left(\frac{1}{n} \vee g \wedge n\right)$. By monotone convergence, $\mu\left(1_{\{g>1\}} f_n\right) \to \mu\left(1_{\{g>1\}}\ln(g)\right) = \nu\left(1_{\{g>1\}} g\ln(g)\right)$ as $n \to \infty$. Since $-\frac{1}{e} = \inf_{y\in[0,1]} y\ln(y) \leq 1_{\{g\leq 1\}} g f_n \leq 0$, by Lebesgue's theorem, $\mu\left(1_{\{g\leq1\}} f_n\right) = \nu\left(1_{\{g\leq1\}} g f_n\right) \to \nu\left(1_{\{g\leq1\}} g\ln(g)\right)$. We deduce that

$$\begin{aligned}\mu(f_n) = \mu\left(1_{\{g>1\}} f_n\right) + \mu\left(1_{\{g\leq1\}} f_n\right) &\to \nu\left(1_{\{g>1\}} g\ln(g)\right) + \nu\left(1_{\{g\leq1\}} g\ln(g)\right)\\ &= \nu(g\ln g) = H(\mu|\nu).\end{aligned}$$

Moreover, by Lebesgue's theorem $\nu(e^{f_n}) = \nu(\frac{1}{n} \vee g \wedge n) \to \nu(g) = 1$. We deduce that $\mu(f_n+1) - \nu(e^{f_n}) \to H(\mu|\nu)$ so that $H(\mu|\nu) = \sup_{f\in\mathcal{M}_b(\mathcal{X})} \left\{\mu(f+1) - \nu(e^f)\right\}$.

∎

Let us now give the dual formulation of the entropic optimal transport problem.

Proposition 4.3. *Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $R \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$. Then*

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R) = \sup_{(\phi,\psi)\in L^1(\mu)\times L^1(\nu)} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R(e^{\phi\oplus\psi})\right\},$$

where $\phi\oplus\psi(x,y) = \phi(x) + \psi(y)$.

Proof of Proposition 4.3: By Lemma 4.2,

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R) = \inf_{\pi\in\Pi(\mu,\nu)} \sup_{f\in C_b(\mathcal{X}\times\mathcal{Y})} \left\{\pi(f+1) - R(e^f)\right\}.$$

Since $\Pi(\mu,\nu)$ is convex and compact by Lemma 1.7, $\pi \mapsto \pi(f+1) - R(e^f)$ is linear and continuous in π for fixed $f \in C_b(\mathcal{X}\times\mathcal{Y})$ and $f \mapsto \pi(f+1) - R(e^f)$ is concave in f by concavity of $\mathbb{R} \ni y \mapsto -e^y$, by the above minimax theorem (Theorem 3.20),

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R) = \sup_{f\in C_b(\mathcal{X}\times\mathcal{Y})} \inf_{\pi\in\Pi(\mu,\nu)} \left\{\pi(f+1) - R(e^f)\right\}.$$

For $f \in C_b(\mathcal{X}\times\mathcal{Y})$, by the duality theorem for the Optimal Transport problem with cost function f (see Theorem 1.8),

$$\inf_{\pi\in\Pi(\mu,\nu)} \left\{\pi(f+1) - R(e^f)\right\} = \sup_{\substack{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})\\ \phi\oplus\psi\leq f}} \left\{\mu(\phi) + \nu(\psi) + 1 - R(e^f)\right\}.$$

Since $\phi\oplus\psi \leq f$ implies $e^{\phi\oplus\psi} \leq e^f$ and $\phi\oplus\psi \in C_b(\mathcal{X}\times\mathcal{Y})$ for $(\phi,\psi) \in C_b(\mathcal{X})\times C_b(\mathcal{Y})$, we deduce that

$$\begin{aligned}
\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R) &\leq \sup_{f\in C_b(\mathcal{X}\times\mathcal{Y})} \sup_{\substack{(\phi,\psi)\in\mathcal{D}_f\\ \phi\in C_b(\mathcal{X}),\psi\in C_b(\mathcal{Y})}} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R\left(e^{\phi\oplus\psi}\right)\right\} \\
&= \sup_{(\phi,\psi)\in C_b(\mathcal{X})\times C_b(\mathcal{Y})} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R\left(e^{\phi\oplus\psi}\right)\right\} \\
&\leq \sup_{(\phi,\psi)\in L^1(\mu)\times L^1(\nu)} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R\left(e^{\phi\oplus\psi}\right)\right\}. \qquad (4.5)
\end{aligned}$$

For $\pi \in \Pi(\mu,\nu)$ such that $\pi << R$ and $(\phi,\psi) \in L^1(\mu)\times L^1(\nu)$, by the equality $y\ln(y) = \sup_{x\in\mathbb{R}}\{yx - e^x\} + y$ valid for $y \geq 0$, we have

$$\frac{d\pi}{dR}\ln\left(\frac{d\pi}{dR}\right) \geq \frac{d\pi}{dR}\phi\oplus\psi - e^{\phi\oplus\psi} + \frac{d\pi}{dR}.$$

We deduce that

$$H(\pi|R) = R\left(\frac{d\pi}{dR}\ln\left(\frac{d\pi}{dR}\right)\right) \geq \pi(\phi\oplus\psi) - R(e^{\phi\oplus\psi}) + \pi(1) = \pi(\phi\oplus\psi+1) - R(e^{\phi\oplus\psi})$$

so that

$$\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R) \geq \sup_{(\phi,\psi)\in L^1(\mu)\times L^1(\nu)} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R\left(e^{\phi\oplus\psi}\right)\right\}.$$

With (4.5), we conclude that both sides are equal. ∎

4.2 Relative entropy and projections

Let us recall that the total variation distance is defined by

$$d_{\mathrm{TV}}(\mu,\nu) = \sup_{A\in\mathcal{B}(\mathcal{X})} |\mu(A) - \nu(A)| \text{ for } \mu,\nu \in \mathcal{P}(\mathcal{X}).$$

Note that

$$d_{\mathrm{TV}}(\mu,\nu) = \sup_{f\in\mathcal{M}_b(\mathcal{X}):\|f\|_\infty\le\frac{1}{2}} |\mu(f)-\nu(f)|, \text{ and} \tag{4.6}$$

$$\mu << \eta \text{ and } \nu << \eta \Rightarrow d_{\mathrm{TV}}(\mu,\nu) = \frac{1}{2}\eta\left(\left|\frac{d\mu}{d\eta}-\frac{d\nu}{d\eta}\right|\right). \tag{4.7}$$

The Pinsker inequality provides an estimation of the total variation distance in terms of the relative entropy.

Lemma 4.4 (Pinsker's inequality)**.**

$$\forall \mu,\nu\in\mathcal{P}(\mathcal{X}),\ d_{\mathrm{TV}}(\mu,\nu)\le\sqrt{\frac{1}{2}H(\mu|\nu)}.$$

Proof: Since $d_{\mathrm{TV}}(\mu,\nu)\le 1$, the conclusion is obvious when $H(\mu|\nu)=\infty$ and we now suppose that $H(\mu|\nu)<\infty$. Let $\mu_0,\mu_1\in\mathcal{P}(\mathcal{X})$ be such that $H(\mu_0|\nu)+H(\mu_1|\nu)<\infty$, $\mu_t=(1-t)\mu_0+t\mu_1$ for $t\in[0,1]$, $f_t=\frac{d\mu_t}{d\nu}=(1-t)f_0+tf_1$ and $h(t)=H(\mu_t|\nu)=\nu(f_t\ln(f_t))$. For $t\in(0,1)$, we have

$$\partial_t(f_t\ln f_t)=(1+\ln f_t)\partial_t f_t=(1+\ln f_t)(f_1-f_0)$$

and

$$\partial_t^2(f_t\ln f_t)=\frac{\partial_t f_t}{f_t}(f_1-f_0)=\frac{(f_1-f_0)^2}{f_t}\le\frac{f_0+f_1}{f_t}|f_1-f_0|\le\frac{|f_1-f_0|}{(1-t)\wedge t}.$$

Using for the third inequality that for $y\in[0,1]$, $-1/e\le y\ln(y)\le 0$ so that $y|\ln(y)|\le 1/e\le 2/e+y\ln(y)$ then the convexity of $\mathbb{R}_+\ni y\mapsto y\ln(y)$ for the fourth, we obtain for $t\in(0,1)$

$$\begin{aligned}|(f_1-f_0)\ln f_t|&\le(f_0+f_1)|\ln(f_t)|\le\frac{f_t}{(1-t)\wedge t}|\ln(f_t)|\le\frac{2/e+f_t\ln(f_t)}{(1-t)\wedge t}\\&\le\frac{1/e+(1-t)(1/e+f_0\ln f_0)+t(1/e+f_1\ln f_1)}{(1-t)\wedge t}.\end{aligned}$$

Since $\partial_t(f_t\ln f_t)=(1+\ln f_t)(f_1-f_0)$, we deduce that for $\varepsilon\in(0,\frac{1}{2})$,

$$\sup_{t\in(\varepsilon,1-\varepsilon)}|\partial_t(f_t\ln f_t)|\le f_0+f_1+\frac{1/e+(1-\varepsilon)(1/e+f_0\ln f_0)+(1-\varepsilon)(1/e+f_1\ln f_1)}{\varepsilon}$$

where the right-hand side is integrable with respect to ν. By Lebesgue's theorem, we deduce that $h(t)$ is differentiable on $(\varepsilon,1-\varepsilon)$ and therefore on $(0,1)$ with derivative

$$h'(t)=\nu\left((1+\ln(f_t))(f_1-f_0)\right)=(\mu_1-\mu_0)(\ln(f_t)). \tag{4.8}$$

Since $\sup_{t\in(\varepsilon,1-\varepsilon)}|\partial_t^2(f_t\ln f_t)|\le\frac{f_0+f_1}{\varepsilon}$, h' is differentiable on $(0,1)$ with derivative

$$\begin{aligned}h''(t)=\nu\left(\frac{(f_1-f_0)^2}{f_t}\right)&=\mu_t\left(\frac{(f_1-f_0)^2}{f_t^2}\right)\\&\ge\left(\mu_t\left(\frac{|f_1-f_0|}{f_t}\right)\right)^2=(\nu(|f_1-f_0|))^2=4d_{\mathrm{TV}}^2(\mu_1,\mu_0),\end{aligned}$$

where we used (4.7) for the last equality. Therefore $[0,1] \ni h(t) + 2t(1-t)d_{\mathrm{TV}}^2(\mu_1,\mu_0)$ is convex and

$$\forall t \in [0,1],\ h(t) + 2t(1-t)d_{\mathrm{TV}}^2(\mu_1,\mu_0) \leq (1-t)h(0) + th(1)$$

so that $d_{\mathrm{TV}}^2(\mu_1,\mu_0) \leq \frac{h(0)}{2t} + \frac{h(1)}{2(1-t)}$ for $t \in (0,1)$. Choosing $\mu_0 = \mu$ and $\mu_1 = \nu$ so that $h(0) = H(\mu|\nu)$ and $h(1) = H(\nu|\nu) = 0$ and letting $t \to 1$, we conclude that $d_{\mathrm{TV}}^2(\mu_1,\mu_0) \leq \frac{1}{2}H(\mu|\nu)$.

∎

Lemma 4.5. *Let $\mu,\nu,\eta \in \mathcal{P}(\mathcal{X})$ with $\mu << \eta$. If either $H(\nu|\eta) < \infty$ or $H(\nu|\mu) < \infty$, then*

$$H(\nu|\eta) - H(\nu|\mu) = \nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right)$$

and $\ln\left(\frac{d\mu}{d\eta}\right)$ is semi-integrable with respect to ν (meaning that its positive or its negative part is integrable).

Proof: When $H(\nu|\eta) < \infty$ then $\nu << \eta$ and when $H(\nu|\mu) < \infty$, $\nu << \mu << \eta$. Let $f = \frac{d\nu}{d\eta}$ and $g = \frac{d\mu}{d\eta}$.

- When $\nu << \mu$, then $\frac{d\nu}{d\mu} = \frac{f}{g}$, η a.e. and therefore μ a.e.. Moreover, $H(\nu|\mu) = \nu(\ln(f/g)) \geq 0$ while $H(\nu|\eta) = \nu(\ln(f)) \geq 0$. When $H(\nu|\mu) < \infty$, then $\ln(f/g) \in L^1(\nu)$ and since $\nu(\ln(f)) \geq 0$, $(\ln(f))^- \in L^1(\nu)$. Moreover,

  $$(\ln(g))^- = (\ln(f) - \ln(f/g))^- \leq (\ln(f))^- + (-\ln(f/g))^- \leq (\ln(f))^- + |\ln(f/g)|$$

  implies that $(\ln(g))^- \in L^1(\nu)$. We deduce that

  $$\begin{aligned} H(\nu|\eta) - H(\nu|\mu) &= \nu(\ln f) - \nu(\ln(f/g)) = \nu(\ln f - \ln(f/g)) = \nu(\ln g) \\ &= \nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right). \end{aligned}$$

  When $H(\nu|\eta) < \infty$, the equality still holds since $\ln(f), (\ln(f/g))^- \in L^1(\nu)$. Moreover,

  $$(\ln(g))^+ = (\ln(f) - \ln(f/g))^+ \leq (\ln(f))^+ + (-\ln(f/g))^+ \leq |\ln(f)| + (\ln(f/g))^-$$

  implies that $(\ln(g))^+ \in L^1(\nu)$.

- When ν is not absolutely continuous with respect to μ, then $\eta(\{f > 0, g = 0\}) > 0$ so that $\nu(\{\ln(g) = -\infty\}) = \eta(1_{\{g=0\}}f) > 0$. Moreover, $H(\nu|\mu) = +\infty$ so that $H(\nu|\eta) = \eta(f\ln(f)) < \infty$ and $f\ln(f) \in L^1(\eta)$. Using that $\ln(y) \leq y - 1$ for $y > 0$, we have

  $$0 \leq f(\ln(g))^+ = \begin{cases} 0 \text{ if } f = 0 \text{ or } g \leq 1 \\ f(\ln(g/f) + \ln(f)) \leq g - f + f\ln(f) \text{ otherwise} \end{cases},$$

  so that $f(\ln(g))^+ \in L^1(\eta)$ and $(\ln(g))^+ \in L^1(\nu)$. Hence $\nu(\ln(g)) = -\infty = H(\nu|\eta) - H(\nu|\mu)$.

■

Proposition 4.6. *Let $\mathcal{E} \subset \mathcal{P}(\mathcal{X})$ be convex and such that $\mu \in \mathcal{E}$ and $\eta \in \mathcal{P}(\mathcal{X})$ be such that $H(\mu|\eta) < \infty$. Then the following are equivalent*

(i) $H(\mu|\eta) = \inf_{\nu \in \mathcal{E}} H(\nu|\eta)$,

(ii) $\forall \nu \in \mathcal{E}$ *with* $H(\nu|\eta) < \infty$, $\nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right) \geq H(\mu|\eta)$,

(iii) $\forall \nu \in \mathcal{E}$, $H(\nu|\eta) \geq H(\nu|\mu) + H(\mu|\eta)$.

Remark 4.7. *Under the equivalent conditions* $(i) - (ii) - (iii)$,

- *if* $\eta \in \mathcal{E}$, *then* $\mu = \eta$,
- *if* $\nu \in \mathcal{E}$ *satisfies* $H(\nu|\eta) < \infty$, *then* $H(\nu|\mu) < \infty$ *and* $\nu << \mu$.

Proof: Since $H(\nu|\mu) \geq 0$, we have $(iii) \Rightarrow (i)$.

Let us now check that $(ii) \Rightarrow (iii)$. When $H(\nu|\eta) = \infty$ then the inequality in (iii) holds. Otherwise, by Lemma 4.5 then (ii),

$$H(\nu|\eta) - H(\nu|\mu) = \nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right) \geq H(\mu|\eta).$$

Let us finally check that $(i) \Rightarrow (ii)$. Let $\nu \in \mathcal{E}$ be such that $H(\nu|\eta) < \infty$ and let $\nu_t = (1-t)\mu + t\nu$ for $t \in [0,1]$. Like in the derivation of (4.8) in the proof of Lemma 4.4, we obtain

$$\forall t \in (0,1),\ \frac{d}{dt}H(\nu_t|\eta) = (\nu - \mu)\left(\ln\left((1-t)\frac{d\mu}{d\eta} + t\frac{d\nu}{d\eta}\right)\right).$$

For all $t \in (0,1)$, since $\nu_t \in \mathcal{E}$ by convexity of this set, $H(\nu_t|\eta) \geq H(\mu|\eta)$. Hence there exists a decreasing sequence $(t_n)_{n\in\mathbb{N}}$ in $(0,1)$ such that $\lim_{n\to\infty} t_n = 0$ and

$$\begin{aligned}\forall n \in \mathbb{N},\ 0 \leq& (\nu - \mu)\left(\ln\left((1-t_n)\frac{d\mu}{d\eta} + t_n\frac{d\nu}{d\eta}\right)\right)\\ \leq& \nu\left(1_{\{\frac{d\mu}{d\eta} \geq \frac{d\nu}{d\eta}\}}\ln\left(\frac{d\mu}{d\eta}\right)\right) + \nu\left(1_{\{\frac{d\mu}{d\eta} < \frac{d\nu}{d\eta}\}}\ln\left((1-t_n)\frac{d\mu}{d\eta} + t_n\frac{d\nu}{d\eta}\right)\right)\\ &- \mu\left(\ln(1-t_n) + \ln\left(\frac{d\mu}{d\eta}\right)\right).\end{aligned}$$

Since $\lim_{n\to\infty} t_n = 0$, the third term in the right-hand side converges to $\mu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right) = H(\mu|\eta)$ as $n \to \infty$.

By monotone convergence, $\nu\left(1_{\{\frac{d\mu}{d\eta} < \frac{d\nu}{d\eta}\}}\left(\ln\left(\frac{d\nu}{d\eta}\right) - \ln\left((1-t_n)\frac{d\mu}{d\eta} + t_n\frac{d\nu}{d\eta}\right)\right)\right)$ converges to $\nu\left(1_{\{\frac{d\mu}{d\eta} < \frac{d\nu}{d\eta}\}}\left(\ln\left(\frac{d\nu}{d\eta}\right) - \ln\left(\frac{d\mu}{d\eta}\right)\right)\right)$. Since $H(\nu|\eta) < \infty$ implies $\nu\left(\left|\ln\left(\frac{d\nu}{d\eta}\right)\right|\right) < \infty$, we deduce that the second term in the right-hand side converges to $\nu\left(1_{\{\frac{d\mu}{d\eta} < \frac{d\nu}{d\eta}\}}\ln\left(\frac{d\mu}{d\eta}\right)\right)$. By Lemma 4.5, $\ln\left(\frac{d\mu}{d\eta}\right)$ is semi-integrable with respect to ν so that $\nu\left(1_{\{\frac{d\mu}{d\eta} \geq \frac{d\nu}{d\eta}\}}\ln\left(\frac{d\mu}{d\eta}\right)\right) + \nu\left(1_{\{\frac{d\mu}{d\eta} < \frac{d\nu}{d\eta}\}}\ln\left(\frac{d\mu}{d\eta}\right)\right) = \nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right)$ and we conclude that $0 \leq \nu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right) - H(\mu|\eta)$. ■

Corollary 4.8. *Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and $R \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$. If there exists $(\phi_\star, \psi_\star) \in L^1(\mu) \times L^1(\nu)$ such that $\pi_\star$ with density $\frac{d\pi_\star}{dR} = e^{\phi_\star \oplus \psi_\star}$ belongs to $\Pi(\mu,\nu)$, then $H(\pi_\star|R) = \inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R)$ and $(\phi_\star, \psi_\star)$ is a pair of dual optimizers.*

Proof: For $\pi \in \Pi(\mu,\nu)$, we have

$$\pi\left(\ln\left(\frac{d\pi_\star}{dR}\right)\right) = \pi(\phi_\star \oplus \psi_\star) = \mu(\phi_\star) + \nu(\psi_\star) = \pi_\star(\phi_\star \oplus \psi_\star) = \pi_\star\left(\ln\left(\frac{d\pi_\star}{dR}\right)\right) = H(\pi_\star|R).$$

Hence statement (ii) in Proposition 4.6 holds with $(\eta, \mu, \nu) = (R, \pi_\star, \pi)$ and $\mathcal{E} = \Pi(\mu,\nu)$ which is convex. By this proposition, statement (i) also holds : $H(\pi_\star|R) = \inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R)$. Moreover, since $R(e^{\phi_\star\oplus\psi_\star}) = R(\frac{d\pi_\star}{dR}) = 1$,

$$\mu\otimes\nu(\phi_\star\oplus\psi_\star+1) - R(e^{\phi_\star\oplus\psi_\star}) = H(\pi_\star|R) = \sup_{(\phi,\psi)\in L^1(\mu)\times L^1(\nu)} \left\{\mu\otimes\nu(\phi\oplus\psi+1) - R(e^{\phi\oplus\psi})\right\},$$

where the last equality follows from Proposition 4.3. ∎

For $\mu \in \mathcal{P}(\mathcal{X})$, we denote by $\Pi(\mu,\cdot) = \bigcup_{\nu\in\mathcal{P}(\mathcal{Y})} \Pi(\mu,\nu)$ the subset of $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ which consists in probability measures with first marginal equal to μ.

Proposition 4.9. *Let $\mu, \eta \in \mathcal{P}(\mathcal{X})$ with $\mu \sim \eta$ (i.e. $\mu << \eta$ and $\eta << \mu$) and $\frac{d\mu}{d\eta} = e^{\phi_\star} \in L^1(\mu)$. For $R \in \Pi(\eta,\cdot)$ with desintegration $R(dx,dy) = \eta(dx)R_x(dy)$,*

$$\inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R) = \mu(\phi_\star) = \sup_{\phi\in L^1(\mu)} \left\{\mu(\phi+1) - R(e^\phi)\right\}.$$

In particular $\pi_\star(dx,dy) = \mu(dx)R_x(dy) = \frac{d\mu}{d\eta}(x)R(dx,dy)$ is the primal optimizer and $\phi_\star$ a dual optimizer.

Remark 4.10. *The fact that $\mu \sim \eta$ implies that $\frac{d\mu}{d\eta} > 0$, η a.e. (and therefore μ a.e.) and writes $e^{\phi_\star}$.*

Proof: For $\pi \in \Pi(\mu,\cdot)$ such that $\pi << R$,

$$\begin{aligned} H(\pi|R) = H(\mu(dx)\pi_x(dy)|\eta(dx)R_x(dy)) &= \int_{\mathcal{X}\times\mathcal{Y}} \ln\left(\frac{d\mu}{d\eta}(x) \times \frac{d\pi_x}{dR_x}(y)\right)\mu(dx)\pi_x(dy) \\ &= H(\mu|\eta) + \int_{\mathcal{X}\times\mathcal{Y}} H(\pi_x|R_x)\mu(dx) \geq H(\mu|\nu), \end{aligned}$$

since $H(\pi_x|R_x) \geq 0$. As a consequence, $\inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R) \geq H(\mu|\nu)$. Since $\pi_{\star x} = R_x$ $\mu(dx)$ a.e., $H(\pi_\star|R) = H(\mu|\eta) = \inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R)$ and $\pi_\star$ is a primal optimizer. The uniquess of the primal optimizer is a consequence of the strict convexity of $\mathbb{R}_+ \ni y \mapsto y\ln(y)$ like in the proof of Corollary 4.1. Finally,

$$\inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R) = H(\mu|\eta) = \mu\left(\ln\left(\frac{d\mu}{d\eta}\right)\right) = \mu(\phi_\star) = \mu(\phi_\star+1) - R\left(e^{\phi_\star}\right), \tag{4.9}$$

where we used that $R\left(e^{\phi_\star}\right) = \eta\left(e^{\phi_\star}\right) = \eta\left(\frac{d\mu}{d\eta}\right) = 1$ for the last equality.

Reasoning like in the end of the proof of Proposition 4.3, we check that $\inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R) \geq \sup_{\phi\in L^1(\mu)} \left\{\mu(\phi+1) - R(e^\phi)\right\}$. With (4.9), we conclude that $\inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|R) = \mu(\phi_\star+1) - R\left(e^{\phi_\star}\right) = \sup_{\phi\in L^1(\mu)} \left\{\mu(\phi+1) - R(e^\phi)\right\}$ and that $\phi_\star$ is a dual optimizer. ∎

4.3 Stability of the Entropic Optimal Transport problem

Proposition 4.11. *Let $c \in C_b(\mathcal{X} \times \mathcal{Y})$. Then $\lim_{\varepsilon\to 0} V_{c,\varepsilon}(\mu,\nu) = V_c(\mu,\nu)$. Moreover, if $(\pi_n)_{n\in\mathbb{N}}$ is a weakly convergent sequence of optimizers for $V_{c,\varepsilon_n}(\mu,\nu)$ with $\lim_{n\to\infty}\varepsilon_n = 0$, then its limit is optimal for $V_c(\mu,\nu)$.*

Remark 4.12. *From any sequence $(\pi^n)_{n\in\mathbb{N}}$ of optimizers we can extract a weakly convergent subsequence according to Lemma 1.7.*

The proof relies on the data processing inequality :

Lemma 4.13 (Data processing inequality)**.** *Let $\kappa : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ be a probability kernel. For $\eta \in \mathcal{P}(\mathcal{X})$, we denote by $\eta\kappa \in \mathcal{P}(\mathcal{Y})$ the probability measure defined by $\eta\kappa(dy) = \int_{x\in\mathcal{X}} \kappa_x(dy)\eta(dx)$. Then*

$$\forall \mu,\nu \in \mathcal{P}(\mathcal{X}),\ H(\mu\kappa|\nu\kappa) \leq H(\mu|\nu).$$

Proof of Proposition 4.11: Clearly, $\varepsilon \mapsto V_{c,\varepsilon}(\mu,\nu)$ is non-decreasing and bounded from below by $V_c(\mu,\nu)$ so that $\lim_{\varepsilon\to 0} V_{c,\varepsilon}(\mu,\nu)$ exists and satisfies $\lim_{\varepsilon\to 0} V_{c,\varepsilon}(\mu,\nu) \geq V_c(\mu,\nu)$. Let $\pi_\star$ be optimal for $V_c(\mu,\nu)$. To check that $\lim_{\varepsilon\to 0} V_{c,\varepsilon}(\mu,\nu) = V_c(\mu,\nu)$, it is enough to exhibit a sequence $(\pi_n)_{n\in\mathbb{N}}$ in $\Pi(\mu,\nu)$ that converges weakly to $\pi_\star$ and such that $H(\pi_n|\mu\otimes\nu) < \infty$ for each $n \in \mathbb{N}$. Indeed setting $\varepsilon_n = \frac{1}{n(H(\pi_n|\mu\otimes\nu)\vee 1)}$ we have $\lim_{n\to\infty}\varepsilon_n = 0 = \lim_{n\to\infty}\varepsilon_n H(\pi_n|\mu\otimes\nu)$ and

$$\lim_{n\to\infty} V_{c,\varepsilon_n}(\mu,\nu) \leq \lim_{n\to\infty}\{\pi_n(c) + \varepsilon_n H(\pi_n|\mu\otimes\nu)\} = \pi_\star(c) = V_c(\mu,\nu).$$

To construct π_n with $H(\pi_n|\mu\otimes\nu) < \infty$, we rely on finitely supported probability measures. Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\nu_n)_{n\in\mathbb{N}}$ be sequences of finitely supported probability measures in $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ such that μ_n converges weakly to μ and ν_n converges weakly to ν as $n\to\infty$ (existence follows for instance from the almost sure convergence of the empirical measure of i.i.d. copies deduced from the strong law of large numbers). Let

- $\mu\otimes\kappa^{n,1} = \mu_n\otimes\tilde{\kappa}^{n,1}$ be some optimal coupling for $\mathcal{W}_1(\mu,\mu_n)$ with the distance $d_{\mathcal{X}}\wedge 1$,
- $\nu\otimes\kappa^{n,2} = \nu_n\otimes\tilde{\kappa}^{n,2}$ be some optimal coupling for $\mathcal{W}_1(\nu,\nu_n)$ with the distance $d_{\mathcal{Y}}\wedge 1$,
- $\hat{\pi}_n = \pi_\star(\kappa^{n,1}\otimes\kappa^{n,2}) \in \Pi(\mu_n,\nu_n)$ and $\pi_n = \hat{\pi}_n(\tilde{\kappa}^{n,1}\otimes\tilde{\kappa}^{n,2}) \in \Pi(\mu,\nu)$.

The weak convergence is equivalent to convergence in $\mathcal{W}_1$ for the distance capped by 1 so that $\lim_{n\to\infty}\mathcal{W}_1(\mu_n,\mu) + \mathcal{W}_1(\nu_n,\nu) = 1$. When $\mathcal{X}\times\mathcal{Y}$ is equipped with the distance $(d_{\mathcal{X}}\wedge 1)\oplus(d_{\mathcal{Y}}\wedge 1)$,

$$\begin{aligned}
\mathcal{W}_1(\pi_\star,\hat{\pi}_n) &\leq \int_{\mathcal{X}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{Y}} (d_{\mathcal{X}}(x,\tilde{x})\wedge 1) + (d_{\mathcal{Y}}(y,\tilde{y})\wedge 1)\pi_\star(dx,dy)\kappa_x^{n,1}(d\tilde{x})\kappa_y^{n,2}(d\tilde{y}) \\
&= \int_{\mathcal{X}\times\mathcal{X}} (d_{\mathcal{X}}(x,\tilde{x})\wedge 1)\mu(dx)\kappa_x^{n,1}(d\tilde{x}) + \int_{\mathcal{Y}\times\mathcal{Y}} (d_{\mathcal{Y}}(y,\tilde{y})\wedge 1)\nu(dy)\kappa_y^{n,2}(d\tilde{y}) \\
&= \mathcal{W}_1(\mu,\mu_n) + \mathcal{W}_1(\nu,\nu_n).
\end{aligned}$$

In a symmetric way, $\mathcal{W}_1(\hat{\pi}_n, \pi_n) \le \mathcal{W}_1(\mu_n, \mu) + \mathcal{W}_1(\nu_n, \nu)$ so that by the triangle inequality

$$\mathcal{W}_1(\pi_\star, \pi_n) \le \mathcal{W}_1(\pi_\star, \hat{\pi}_n) + \mathcal{W}_1(\hat{\pi}_n, \pi_n) \le 2\left(\mathcal{W}_1(\mu, \mu_n) + \mathcal{W}_1(\nu, \nu_n)\right) \stackrel{n\to\infty}{\longrightarrow} 0,$$

and π_n converges weakly to $\pi_\star$ as $n \to \infty$. On the other hand, since $\hat{\pi}_n$ and $\mu_n \otimes \nu_n$ are finitely supported with the support of $\hat{\pi}_n$ included in the one of $\mu_n \otimes \nu_n$, $\hat{\pi}_n << \mu_n \otimes \nu_n$, and since the relative entropy amounts to a finite sum, $\infty > H(\hat{\pi}_n|\mu_n \otimes \nu_n)$. Since $\pi_n = \hat{\pi}_n(\tilde{\kappa}^{n,1} \otimes \tilde{\kappa}^{n,2})$ and $\mu \otimes \nu = (\mu_n \tilde{\kappa}^{n,1}) \otimes (\nu_n \tilde{\kappa}^{n,2}) = (\mu_n \otimes \nu_n)(\tilde{\kappa}^{n,1} \otimes \tilde{\kappa}^{n,2})$, the data processing inequality stated in Lemma 4.13 ensures that $H(\pi_n|\mu \otimes \nu) \le H(\hat{\pi}_n|\mu_n \otimes \nu_n) < \infty$. The sequence $(\pi_n)_{n\in\mathbb{N}}$ has the desired properties and therefore $\lim_{\varepsilon\to 0} V_{c,\varepsilon}(\mu, \nu) = V_c(\mu, \nu)$.

Last, if $(\pi_n)_{n\in\mathbb{N}}$ is sequence of optimizers for $V_{c,\varepsilon_n}(\mu, \nu)$ with $\lim_{n\to\infty} \varepsilon_n = 0$ that converges weakly to π_∞, then $\pi_\infty \in \Pi(\mu, \nu)$ and taking the limit $n \to \infty$ in the inequality $\pi_n(c) \le V_{c,\varepsilon_n}(\mu, \nu)$, we get $\pi_\infty(c) \le V_c(\mu, \nu)$, so that π_∞ is optimal for $V_c(\mu, \nu)$. ∎

Proof of Lemma 4.13: It is enough to deal with the case when $H(\mu|\nu) < \infty$, which implies $\mu << \nu$. Using the desintegrations $\mu(dx)\kappa_x(dy) = \mu\kappa(dy)\kappa^\mu_y(dx)$ and $\nu(dx)\kappa_x(dy) = \nu\kappa(dy)\kappa^\nu_y(dx)$, we have

$$\frac{d\mu}{d\nu}(x) = \frac{d\mu\kappa_x}{d\nu\kappa_x}(x, y) = \frac{d\mu\kappa}{d\nu\kappa}(y) \times \frac{d\kappa^\mu_y}{d\kappa^\nu_y}(x).$$

Then

$$\begin{aligned} H(\mu|\nu) = H(\mu\kappa_x|\nu\kappa_x) &= \int_{\mathcal{X}\times\mathcal{Y}} \ln\left(\frac{d\mu\kappa}{d\nu\kappa}(y)\right) + \ln\left(\frac{d\kappa^\mu_y}{d\kappa^\nu_y}(x)\right) \kappa^y_\mu(dx)\mu\kappa(dy) \\ &= H(\mu\kappa|\nu\kappa) + \int_{\mathcal{Y}} H(\kappa^y_\mu|\kappa^y_\nu)\mu\kappa(dy) \ge H(\mu\kappa|\nu\kappa), \end{aligned}$$

where we used the non-negativity of the relative entropy for the inequality. ∎

Proposition 4.14. *Let $c \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y})$, $\mu, \tilde{\mu} \in \mathcal{P}(\mathcal{X})$, $\nu, \tilde{\nu} \in \mathcal{P}(\mathcal{Y})$. Then*

$$|V_{c,\varepsilon}(\mu, \nu) - V_{c,\varepsilon}(\tilde{\mu}, \tilde{\nu}) \le 2\|c\|_\infty \left(d_{\mathrm{TV}}(\mu, \tilde{\mu}) + d_{\mathrm{TV}}(\nu, \tilde{\nu})\right).$$

Moreover, the optimizers $\pi_\star \in \Pi(\mu, \nu)$ and $\tilde{\pi}_\star \in \Pi(\tilde{\mu}, \tilde{\nu})$ satisfy

$$d_{\mathrm{TV}}(\pi_\star, \tilde{\pi}_\star) \le d_{\mathrm{TV}}(\mu, \tilde{\mu}) + d_{\mathrm{TV}}(\nu, \tilde{\nu}) + \sqrt{2\|c\|_\infty \left(d_{\mathrm{TV}}(\mu, \tilde{\mu}) + d_{\mathrm{TV}}(\nu, \tilde{\nu})\right)}.$$

Proof: The coupling

$$\mu(dx)\kappa^1_x(d\tilde{x}) = \mu \wedge \tilde{\mu}(dx)\delta_x(d\tilde{x}) + 1_{\{\mu\wedge\tilde{\mu}(\mathcal{X})<1\}} \frac{(\mu - \mu \wedge \tilde{\mu})(dx)(\tilde{\mu} - \mu \wedge \tilde{\mu})(d\tilde{x})}{1 - \mu \wedge \tilde{\mu}(\mathcal{X})}$$

is optimal for $d_{\mathrm{TV}}(\mu, \tilde{\mu})$. Let $\nu(dy)\kappa^2_y(d\tilde{y})$ defined similarly be optimal for $d_{\mathrm{TV}}(\nu, \tilde{\nu})$. For $\pi \in \Pi(\mu, \nu)$, $\tilde{\pi} := \pi(\kappa^1 \otimes \kappa^2) \in \Pi(\tilde{\mu}, \tilde{\nu})$. Let $I^{\mathcal{Y}}_y(dw) = \delta_y(dw)$ and $\hat{\pi} = \pi(\kappa^1 \otimes I^{\mathcal{Y}})$. By the triangle inequality, $d_{\mathrm{TV}}(\pi, \tilde{\pi}) \le d_{\mathrm{TV}}(\pi, \hat{\pi}) + d_{\mathrm{TV}}(\hat{\pi}, \tilde{\pi})$. We have $\hat{\pi}(dx, dy) = \tilde{\mu}(dx)\pi_x(dy) = \nu(dy)\tilde{\pi}_y(dx)$. For $f \in \mathcal{M}_b(\mathcal{X} \times \mathcal{Y})$ bounded by 1/2, the function $g(x) = \int_{\mathcal{Y}} f(x, y)\pi_x(dy)$ also is bounded by 1/2 and, by (4.6),

$$|\pi(f) - \hat{\pi}(f)| = |\mu(g) - \tilde{\mu}(g)| \le d_{\mathrm{TV}}(\mu, \tilde{\mu}).$$

With (4.6), this implies that $d_{\mathrm{TV}}(\pi,\hat\pi)\le d_{\mathrm{TV}}(\mu,\tilde\mu)$. In a symmetric way, $d_{\mathrm{TV}}(\hat\pi,\tilde\pi)\le d_{\mathrm{TV}}(\nu,\tilde\nu)$ and therefore $d_{\mathrm{TV}}(\pi,\tilde\pi)\le d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)$. With (4.6), we deduce that

$$|\tilde\pi(c)-\pi(c)|\le 2\|c\|_\infty d_{\mathrm{TV}}(\pi,\tilde\pi)\le 2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right). \tag{4.10}$$

Since $\tilde\mu\otimes\tilde\nu=(\mu\kappa_1)\otimes(\nu\kappa_2)=(\mu\otimes\nu)(\kappa^1\otimes\kappa^2)$, by the data processing inequality stated in Lemma 4.13, $H(\tilde\pi|\tilde\mu\otimes\tilde\nu)\le H(\pi|\mu\otimes\nu)$. Hence

$$\tilde\pi(c)-\pi(c)+\varepsilon H(\tilde\pi|\tilde\mu\otimes\tilde\nu)-\varepsilon H(\pi|\mu\otimes\nu)\le 2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right).$$

For the choice $\pi=\pi_\star$, we deduce that

$$V_{c,\varepsilon}(\tilde\mu,\tilde\nu)-V_{c,\varepsilon}(\mu,\nu)\le 2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right).$$

The first inequality follows by symmetry. For $\tilde R_\varepsilon\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ such that $\frac{dR_\varepsilon}{d\tilde\mu\otimes\tilde\nu}=\frac{e^{-\frac{c}{\varepsilon}}}{\tilde\mu\otimes\tilde\nu\left(e^{-\frac{c}{\varepsilon}}\right)}$, $\tilde\pi_\star$ minimizes $H(\tilde\pi|\tilde R)$ over $\Pi(\tilde\mu,\tilde\nu)$ and by Proposition 4.6, $H(\tilde\pi|\tilde R)\ge H(\tilde\pi|\tilde\pi_\star)+H(\tilde\pi_\star|\tilde R)$ for each $\tilde\pi\in\Pi(\tilde\mu,\tilde\nu)$. We choose $\tilde\pi=\pi_\star(\kappa^1\otimes\kappa^2)$ to deduce with (4.10), $H(\tilde\pi|\tilde\mu\otimes\tilde\nu)\le H(\pi_\star|\mu\otimes\nu)$ (consequence of the data processing inequality stated in Lemma 4.13) and the first assertion that

$$\begin{aligned}
H(\tilde\pi|\tilde\pi_\star)&\le H(\tilde\pi|\tilde R)-H(\tilde\pi_\star|\tilde R)\\
&=\tilde\pi(c)-\tilde\pi_\star(c)+\varepsilon\left(H(\tilde\pi|\tilde\mu\otimes\tilde\nu)-H(\tilde\pi_\star|\tilde\mu\otimes\tilde\nu)\right)\\
&\le\pi_\star(c)-\tilde\pi_\star(c)+2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right)+\varepsilon\left(H(\pi_\star|\mu\otimes\nu)-H(\tilde\pi_\star|\tilde\mu\otimes\tilde\nu)\right)\\
&=V_{c,\varepsilon}(\mu,\nu)-V_{c,\varepsilon}(\tilde\mu,\tilde\nu)+2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right)\\
&\le 4\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right).
\end{aligned}$$

With Pinsker's inequality stated in Lemma 4.4, we conclude that

$$\begin{aligned}
d_{\mathrm{TV}}(\pi_\star,\tilde\pi_\star)\le d_{\mathrm{TV}}(\pi_\star,\tilde\pi)+d_{\mathrm{TV}}(\tilde\pi,\tilde\pi_\star)&\le d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)+\sqrt{\frac{1}{2}H(\tilde\pi|\tilde\pi_\star)}\\
&\le d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)+\sqrt{2\|c\|_\infty\left(d_{\mathrm{TV}}(\mu,\tilde\mu)+d_{\mathrm{TV}}(\nu,\tilde\nu)\right)}.
\end{aligned}$$

∎

4.4 The Sinkhorn algorithm

Let $c\in\mathcal{M}_b(\mathcal{X}\times\mathcal{Y})$, $\varepsilon>0$ and $R_\varepsilon\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be defined by $\frac{dR_\varepsilon}{d\mu\otimes\nu}=\frac{e^{-\frac{c}{\varepsilon}}}{\mu\otimes\nu(e^{-\frac{c}{\varepsilon}})}$. The probability measure R_ε is equivalent to $\mu\otimes\nu$ and for $\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, $\pi<<\mu\otimes\nu\Leftrightarrow\pi<<R_\varepsilon$. When this holds, $\frac{d\pi}{dR_\varepsilon}=\mu\otimes\nu(e^{-\frac{c}{\varepsilon}})e^{\frac{c}{\varepsilon}}\frac{d\pi}{d\mu\otimes\nu}$ so that

$$\varepsilon H(\pi|R_\varepsilon)=\left(\pi(c)+\varepsilon H(\pi|\mu\otimes\nu)\right)+\varepsilon\ln\left(\mu\otimes\nu(e^{-\frac{c}{\varepsilon}})\right). \tag{4.11}$$

Up to adding $\varepsilon\ln\left(\mu\otimes\nu(e^{-\frac{c}{\varepsilon}})\right)$ to c we suppose from now on that

$$\frac{dR_\varepsilon}{d\mu\otimes\nu}=e^{-\frac{c}{\varepsilon}}\text{ and }V_{c,\varepsilon}(\mu,\nu)=\varepsilon\inf_{\pi\in\Pi(\mu,\nu)}H(\pi|R_\varepsilon).$$

By Corollary 4.8, we deduce that if we find $(\phi, \psi) \in L^1(\mu) \times L^1(\nu)$ such that π with density $\frac{d\pi}{dR_\varepsilon} = e^{\phi \oplus \psi}$ belongs to $\Pi(\mu, \nu)$, then π is the minimizer of $V_{c,\varepsilon}$ (uniqueness is a consequence of Corollary 4.1). Note that then

$$\nu(dy) = \int_{x \in \mathcal{X}} \pi(dx, dy) = \int_{x \in \mathcal{X}} e^{\phi(x)+\psi(y)-\frac{c}{\varepsilon}(x,y)} \mu(dx)\nu(dy)$$

so that

$$\nu(dy) \text{ a.e., } e^{\psi(y)} \int_{x \in \mathcal{X}} e^{\phi(x)-\frac{c}{\varepsilon}(x,y)} \mu(dx) = 1$$

and by symmetry,

$$\mu(dx) \text{ a.e., } e^{\phi(x)} \int_{y \in \mathcal{Y}} e^{\psi(y)-\frac{c}{\varepsilon}(x,y)} \nu(dy) = 1.$$

Hence

$$\begin{aligned} \nu(dy) \text{ a.e., } \psi(y) &= -\ln\left(\int_{x \in \mathcal{X}} e^{\phi(x)-\frac{c}{\varepsilon}(x,y)} \mu(dx)\right) \\ \text{and } \mu(dx) \text{ a.e., } \phi(x) &= -\ln\left(\int_{y \in \mathcal{Y}} e^{\psi(y)-\frac{c}{\varepsilon}(x,y)} \nu(dy)\right). \end{aligned}$$

Definition 4.15. *The Sinkhorn algorithm consists in alternatively solving these two equations starting from $\phi_0(x) = 0,\ x \in \mathcal{X}$: for $n \in \mathbb{N}$,*

$$\begin{aligned} \psi_n(y) &= -\ln\left(\int_{x \in \mathcal{X}} e^{\phi_n(x)-\frac{c}{\varepsilon}(x,y)} \mu(dx)\right),\ y \in \mathcal{Y} \\ \phi_{n+1}(x) &= -\ln\left(\int_{y \in \mathcal{Y}} e^{\psi_n(y)-\frac{c}{\varepsilon}(x,y)} \nu(dy)\right),\ x \in \mathcal{X}. \end{aligned}$$

Since the cost function c is bounded, we check by induction on n that so are ϕ_n and ψ_n. Let $\pi_0 = R_\varepsilon = e^{-\frac{c}{\varepsilon}} \mu \otimes \nu$ and for $n \in \mathbb{N}$, π_{2n+1} and π_{2n+2} be defined by

$$\frac{d\pi_{2n+1}}{d\mu \otimes \nu} = e^{\phi_n \oplus \psi_n - \frac{c}{\varepsilon}} \text{ and } \frac{d\pi_{2n+2}}{d\mu \otimes \nu} = e^{\phi_{n+1} \oplus \psi_n - \frac{c}{\varepsilon}}. \tag{4.12}$$

Since, by definition of ψ_n and ϕ_{n+1},

$$\int_{x \in \mathcal{X}} e^{\phi_n(x)+\psi_n(y)-\frac{c}{\varepsilon}(x,y)} \mu(dx) = 1 \text{ and } \int_{y \in \mathcal{Y}} e^{\phi_{n+1}(x)+\psi_n(y)-\frac{c}{\varepsilon}(x,y)} \nu(dy) = 1,$$

we have $\pi_{2n+1} \in \Pi(\cdot, \nu)$ and $\pi_{2n+2} \in \Pi(\mu, \cdot)$ with $\Pi(\cdot, \nu) = \bigcup_{\eta \in \mathcal{P}(\mathcal{X})} \Pi(\eta, \nu)$ denoting the subset of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ consisting in probability measures with second marginal equal to ν. For $n \in \mathbb{N}$, the first marginal of π_{2n+1} is

$$\eta(dx) := e^{\phi_n(x)} \int_{y \in \mathcal{Y}} e^{\psi_n(y)-\frac{c}{\varepsilon}(x,y)} \nu(dy)\mu(dx) = e^{\phi_n(x)-\phi_{n+1}(x)} \mu(dx) \tag{4.13}$$

and the second marginal of π_{2n+2} is

$$e^{\psi_n(y)} \int_{y \in \mathcal{Y}} e^{\phi_{n+1}(x)-\frac{c}{\varepsilon}(x,y)} \mu(dx)\nu(dy) = e^{\psi_n(y)-\psi_{n+1}(y)} \nu(dy). \tag{4.14}$$

Since $\frac{d\mu}{d\eta} = e^{\phi_{n+1}-\phi_n}$ and, by (4.12), $\pi_{2n+2}(dx,dy) = \frac{d\mu}{d\eta}(x)\pi_{2n+1}(dx,dy)$, by Proposition 4.9 applied with $R = \pi_{2n+1}$, $H(\pi_{2n+2}|\pi_{2n+1}) = \inf_{\pi\in\Pi(\mu,\cdot)} H(\pi|\pi_{2n+1})$ and $\phi_{n+1}-\phi_n \in \operatorname{argmax}\left\{\mu(\phi) - \pi_{2n+1}\left(e^{\phi}\right) : \phi \in L^1(\mu)\right\}$. Hence

$$\mu(\phi_{n+1}-\phi_n) - \pi_{2n+1}\left(e^{\phi_{n+1}-\phi_n}\right) \geq \mu(0) - \pi_{2n+1}\left(e^0\right),$$

so that with (4.12),

$$\mu(\phi_{n+1}) - \mu\otimes\nu\left(e^{\phi_{n+1}\oplus\psi_n-\frac{c}{\varepsilon}}\right) \geq \mu(\phi_n) - \mu\otimes\nu\left(e^{\phi_n\oplus\psi_n-\frac{c}{\varepsilon}}\right).$$

In a symmetric way, for $n\geq 1$, $H(\pi_{2n+1}|\pi_{2n}) = \inf_{\pi\in\Pi(\cdot,\nu)} H(\pi|\pi_{2n})$ and

$$\nu(\psi_n) - \mu\otimes\nu\left(e^{\phi_n\oplus\psi_n-\frac{c}{\varepsilon}}\right) \geq \nu(\psi_{n-1}) - \mu\otimes\nu\left(e^{\phi_n\oplus\psi_{n-1}-\frac{c}{\varepsilon}}\right).$$

Hence, for $n\geq 1$,

$$\begin{aligned}\mu(\phi_n)+\nu(\psi_n) - \mu\otimes\nu\left(e^{\phi_n\oplus\psi_n-\frac{c}{\varepsilon}}\right) &\geq \mu(\phi_n)+\nu(\psi_{n-1}) - \mu\otimes\nu\left(e^{\phi_{n-1}\oplus\psi_{n-1}-\frac{c}{\varepsilon}}\right)\\ &\geq \mu(\phi_{n-1})+\nu(\psi_{n-1}) - \mu\otimes\nu\left(e^{\phi_{n-1}\oplus\psi_{n-1}-\frac{c}{\varepsilon}}\right).\end{aligned}$$

This rewrites $\mu\otimes\nu(\phi_n\oplus\psi_n+1) - R_\varepsilon\left(e^{\phi_n\oplus\psi_n}\right) \geq \mu\otimes\nu(\phi_{n-1}\oplus\psi_{n-1}+1) - R_\varepsilon\left(e^{\phi_{n-1}\oplus\psi_{n-1}}\right)$. Therefore, in view of the dual formulation of the entropic optimal transport problem stated in Proposition 4.3, we may expect convergence of the Sinkhorn algorithm.

Theorem 4.16. *The sequence $(\pi_n)_n$ converges in total variation to the optimizer of $V_{c,\varepsilon}(\mu,\nu)$.*

Proof: Let $\pi_\star \in \Pi(\mu,\nu) = \Pi(\mu,\cdot)\cap\Pi(\cdot,\nu)$ denote the optimal coupling for $V_{c,\varepsilon}(\mu,\nu) = \varepsilon\inf_{\pi\in\Pi(\mu,\nu)} H(\pi|R_\varepsilon)$. For $n\in\mathbb{N}$, since $\pi_{2n+1}\in\operatorname{argmin}\{H(\pi|\pi_{2n}) : \pi\in\Pi(\cdot,\nu)\}$ and $\pi_{2n+2}\in\operatorname{argmin}\{H(\pi|\pi_{2n+1}) : \pi\in\Pi(\mu,\cdot)\}$, by Proposition 4.6,

$$\begin{gathered}H(\pi_\star|\pi_{2n}) \geq H(\pi_\star|\pi_{2n+1}) + H(\pi_{2n+1}|\pi_{2n})\\ \text{and } H(\pi_\star|\pi_{2n+1}) \geq H(\pi_\star|\pi_{2n+2}) + H(\pi_{2n+2}|\pi_{2n+1}).\end{gathered}$$

We deduce that $H(\pi_\star|\pi_{2n}) \geq H(\pi_\star|\pi_{2n+2}) + H(\pi_{2n+1}|\pi_{2n}) + H(\pi_{2n+2}|\pi_{2n+1})$ and that

$$H(\pi_\star|\pi_0) \geq \sum_{n\in\mathbb{N}}(H(\pi_{2n+1}|\pi_{2n}) + H(\pi_{2n+2}|\pi_{2n+1})). \tag{4.15}$$

Since $\pi_0 = R_\varepsilon$, $H(\pi_\star|\pi_0) = H(\pi_\star|R_\varepsilon) \leq H(\mu\otimes\nu|R_\varepsilon) = \mu\otimes\nu\left(\frac{c}{\varepsilon}\right) < \infty$.

Let μ_n and ν_n denote the first and second marginals of π_n. We have $\nu_{2n+1} = \nu$ and $\mu_{2n+2} = \mu$ for $n\in\mathbb{N}$ and

$$H(\mu_n|\mu) + H(\nu_n|\nu) = \begin{cases} H(\mu_n|\mu_{n-1}) \text{ if } n\geq 3 \text{ is odd}\\ H(\nu_n|\nu_{n-1}) \text{ if } n\geq 2 \text{ is even}\end{cases}.$$

Note that

$$\begin{aligned}H(\pi_n|\pi_{n-1}) &= \int_{\mathcal{X}\times\mathcal{Y}} \ln\left(\frac{d\mu_n}{d\mu_{n-1}}(x)\times\frac{d\pi_{nx}}{d\pi_{n-1x}}(y)\right)\pi_{n-1x}(dy)\mu_{n-1}(dx)\\ &= H(\mu_n|\mu_{n-1}) + \int_{\mathcal{X}} H(\pi_{nx}|\pi_{n-1x})\mu_{n-1}(dx) \geq H(\mu_n|\mu_{n-1})\end{aligned} \tag{4.16}$$

and, in a symmetric way, $H(\pi_n|\pi_{n-1}) \geq H(\nu_n|\nu_{n-1})$. Therefore, for $n \geq 2$, $H(\mu_n|\mu) + H(\nu_n|\nu) \leq H(\pi_n|\pi_{n-1})$. With (4.15) then Pinsker's inequality (see Lemma 4.4), we deduce that

$$H(\pi_\star|\pi_0) \geq \sum_{n\geq 2} (H(\mu_n|\mu) + H(\nu_n|\nu)) \geq 2\sum_{n\geq 2} (d^2_{\mathrm{TV}}(\mu_n, \mu) + d^2_{\mathrm{TV}}(\nu_n, \nu)).$$

Therefore $\lim_{n\to\infty}(d_{\mathrm{TV}}(\mu_n, \mu) + d_{\mathrm{TV}}(\nu_n, \nu)) = 0$.
Since, by (4.12) and the definition of R_ε,

$$\frac{d\pi_{2n+1}}{dR_\varepsilon} = e^{\phi_n \oplus \psi_n} \text{ and } \frac{d\pi_{2n+2}}{dR_\varepsilon} = e^{\phi_{n+1} \oplus \psi_n},$$

Corollary 4.8 ensures that for $n \geq 1$, π_n is optimal for $\inf_{\pi \in \Pi(\mu_n, \nu_n)} H(\pi|R_\varepsilon)$. For $\pi \in \Pi(\mu_n, \nu_n)$, since $\frac{d\pi}{dR_\varepsilon} = \frac{d\pi}{d\mu_n \otimes \nu_n} \times \frac{d\mu_n \otimes \nu_n}{d\mu \otimes \nu} \times e^{\frac{c}{\varepsilon}}$, we have

$$\begin{aligned} H(\pi|R_\varepsilon) &= \pi\left(\ln\left(\frac{d\pi}{dR_\varepsilon}\right)\right) = H(\pi|\mu_n \otimes \nu_n) + \pi\left(\ln\left(\frac{d\mu_n}{d\mu}\right) \oplus \ln\left(\frac{d\nu_n}{d\nu}\right)\right) + \frac{1}{\varepsilon}\pi(c) \\ &= \frac{1}{\varepsilon}\left(\pi(c) + \varepsilon H(\pi|\mu_n \otimes \nu_n)\right) + H(\mu_n|\mu) + H(\nu_n|\nu). \end{aligned}$$

We deduce that for $n \geq 1$, π_n is optimal for $V_{c,\varepsilon}(\mu_n, \nu_n)$. With Proposition 4.14, we conclude that $\lim_{n\to\infty} d_{\mathrm{TV}}(\pi_n, \pi_\star) = 0$.

∎

Remark 4.17. *By (4.12), (4.13) and (4.14),* $\frac{d\pi_{2n}}{d\pi_{2n+2}} = \frac{e^{\phi_n \oplus \psi_{n-1}}}{e^{\phi_{n+1} \oplus \psi_n}}$, $\frac{d\mu_{2n+1}}{d\mu} = e^{\phi_n - \phi_{n+1}}$ *and* $\frac{d\nu_{2n}}{d\nu} = e^{\psi_{n-1} - \psi_n}$. *Therefore*

$$\begin{aligned} H(\pi_{2n}|\pi_{2n+2}) &= \pi_{2n}\left((\phi_n - \phi_{n+1}) \oplus (\psi_{n-1} - \psi_n)\right) = -\mu\left(\ln\left(\frac{d\mu}{d\mu_{2n+1}}\right)\right) + \nu_{2n}\left(\ln\left(\frac{d\nu_{2n}}{d\nu}\right)\right) \\ &= H(\nu_{2n}|\nu) - H(\mu|\mu_{2n+1}) \\ H(\pi_{2n+2}|\pi_{2n}) &= \pi_{2n+2}\left((\phi_{n+1} - \phi_n) \oplus (\psi_n - \psi_{n-1})\right) \\ &= H(\mu|\mu_{2n+1}) + \nu_{2n+2}\left(2\psi_n - \psi_{n-1} - \psi_{n+1}\right) - \nu_{2n+2}\left(\psi_n - \psi_{n+1}\right) \\ &= H(\mu|\mu_{2n+1}) + H(\nu_{2n+2}|\nu_{2n}) - H(\nu_{2n+2}|\nu). \end{aligned}$$

The first equality implies that $H(\nu_{2n}|\nu) \geq H(\mu|\mu_{2n+1})$ and the second combined with $H(\pi_{2n+2}|\pi_{2n}) \geq H(\nu_{2n+2}|\nu_{2n})$ proved like (4.16) ensures that $H(\mu|\mu_{2n+1}) \geq H(\nu_{2n+2}|\nu)$. Therefore $H(\nu_{2n}|\nu) \geq H(\nu_{2n+2}|\nu)$. In a symmetric way, we can check that the sequence $(H(\mu_{2n+1}|\mu))_{n\in\mathbb{N}}$ also is non-increasing.

Remark 4.18. *By studying the dual problem, it is possible to check that $\|\phi_\star - \phi_n\|_{L^2(\mu)} + \|\psi_\star - \psi_n\|_{L^2(\nu)} \leq Ch^n$ for some rate $h \in (0, 1)$ depending on $\|c\|_\infty$.*

Chapter 5

Weak convergence

Let $\mathcal{X}$ be a metric space endowed with its Borel σ-field $\mathcal{B}(\mathcal{X})$ i.e. the smallest σ-field which contains all the open subsets of $\mathcal{X}$ for the distance $d_{\mathcal{X}}$. A subset $O \subset \mathcal{X}$ is open if $\forall x \in O,\ \exists \varepsilon > 0,\ B(x,\varepsilon) := \{y \in \mathcal{X} : d_{\mathcal{X}}(x,y) < \varepsilon\} \subset O$. Let $\mathcal{P}(\mathcal{X})$ denote the set of probability measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $C_b(\mathcal{X}) = \{f : \mathcal{X} \to \mathbb{R}$ continuous and bounded$\}$.

Definition 5.1.

- *A sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ converges weakly to $\mu \in \mathcal{P}(\mathcal{X})$ if*

$$\forall f \in C_b(\mathcal{X}),\ \lim_{n\to\infty} \mu_n(f) = \mu(f).$$

- *A function $f : \mathcal{X} \to \mathbb{R}$ is upper (resp. lower) semi-continuous if*

$$\forall x \in \mathcal{X},\ f(x) \geq \limsup_{y\to x} f(y)\ (\textit{resp. } f(x) \leq \liminf_{y\to x} f(y)).$$

Note that a function is continuous iff it is both upper and lower semi-continuous.

Theorem 5.2 (Portmanteau)**.** *The following assertions are equivalent*

1) *The sequence $(\mu_n)_{n\in N}$ converges weakly to μ as $n \to \infty$.*

2) $\forall f : \mathcal{X} \to \mathbb{R}$ *Lipschitz and bounded,* $\lim_{n\to\infty} \mu_n(f) = \mu(f)$.

3) $\lim_{n\to\infty} \sup_{\substack{f:\mathcal{Y}\to\mathbb{R} \\ 1-\text{ Lipschitz bounded by } 1}} \{\mu_n(f) - \mu(f)\} = 0.$

4) $\forall F \subset \mathcal{X}$ *closed,* $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$.

5) $\forall f : \mathcal{X} \to \{-\infty\} \cup \mathbb{R}$ *upper semi-continuous and bounded from above,* $\limsup_{n\to\infty} \mu_n(f) \leq \mu(f)$.

6) $\forall O \subset \mathcal{X}$ *open,* $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$.

7) $\forall f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ *lower semi-continuous and bounded from below,* $\liminf_{n\to\infty} \mu_n(f) \geq \mu(f)$.

8) $\forall A \in \mathcal{B}(\mathcal{X})$ *with* $\mu(\partial A) = 0$, $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.

9) $\forall f : \mathcal{X} \to \mathbb{R}$ *bounded and continuous μ a.e.,* $\lim_{n\to\infty} \mu_n(f) = \mu(f)$.

We refer to Theorem 2.1 [5] for the equivalence between 1), 4), 6) and 8). We will frequently use 7). Let us check that it is equivalent to 6). Since for $O \subset \mathcal{X}$ open, 1_O is lower semi-continuous and bounded from below by 0, we have 7) $\Rightarrow$ 6). Let $f : \mathcal{X} \to \mathbb{R}$ be lower semi-continuous and bounded from below. Up to replacing, f by $f - \inf f$, we may suppose that the function is non-negative. For $\eta \in \mathcal{P}(\mathcal{X})$, we have

$$\eta(f) = \int_{x\in\mathcal{X}} \int_{\alpha=0}^{f(x)} d\alpha \eta(dx) = \int_{\alpha=0}^{+\infty} \int_{x\in\mathcal{X}} 1_{\{f(x)>\alpha\}} \eta(dx) d\alpha = \int_{\alpha=0}^{+\infty} \eta(\{x \in \mathcal{X} : f(x) > \alpha\}) d\alpha.$$

Since f is lower semi-continuous, the set $\{x \in \mathcal{X} : f(x) \leq \alpha\}$ is closed and its complementary $\{x \in \mathcal{X} : f(x) > \alpha\}$ is open. Hence, by 6) and Fatou lemma,

$$\begin{aligned}\mu(f) = \int_{\alpha=0}^{+\infty} \mu(\{x \in \mathcal{X} : f(x) > \alpha\}) d\alpha &\leq \int_{\alpha=0}^{+\infty} \liminf_{n\to\infty} \mu_n(\{x \in \mathcal{X} : f(x) > \alpha\}) d\alpha \\ \leq \liminf_{n\to\infty} \int_{\alpha=0}^{+\infty} \mu_n(\{x \in \mathcal{X} : f(x) > \alpha\}) d\alpha &= \liminf_{n\to\infty} \mu_n(f).\end{aligned}$$

The following result is the combination of Theorems 5.1 and 5.2 [5].

Theorem 5.3 (Prokhorov)**.** *Let $\mathcal{L} \subset \mathcal{P}(\mathcal{X})$.*

- *If $\mathcal{L}$ is tight i.e.*

$$\forall \varepsilon > 0,\ \exists K \subset \mathcal{X} \text{ compact such that } \sup_{\mu\in\mathcal{L}} \mu(K^c) \leq \varepsilon,$$

  then $\mathcal{L}$ is relatively compact (i.e. its closure is compact) for the weak convergence topology.

- *Let $\mathcal{X}$ be Polish (i.e. separable and complete in addition to metric). Conversely, if $\mathcal{L}$ is relatively compact, then $\mathcal{L}$ is tight.*

Proposition 5.4 (Skorokhod representation of the weak convergence)**.** *Let $\mathcal{X}$ be a Polish space and $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ converge weakly to $\mu \in \mathcal{P}(\mathcal{X})$ as $n \to \infty$. Then there exists on a probability space, $X_n \sim \mu_n$, $n \in \mathbb{N}$ and $X \sim \mu$ such that* $\lim_{n\to\infty} X_n = X$ *a.s..*

We refer to Theorem 6.7 [5] for a proof of this result which is a consequence of Lemma 1.43 when $\mathcal{X} = \mathbb{R}$.

Bibliography

[1] B. Acciaio and G. Pammer. Advances in optimal transport and stochastics.

[2] D. R. Adams and L. I. Hedberg. *Function spaces and potential theory*, volume 314 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1996.

[3] A. Alfonsi and B. Jourdain. Quadratic Wasserstein distance between Gaussian laws revisited with correlation. *Preprint arXiv 2506.23742*, 2025.

[4] M. Beiglböck, B. Jourdain, W. Margheriti, and G. Pammer. Approximation of martingale couplings on the line in the adapted weak topology. *Probab. Theory Related Fields*, 183(1-2):359–413, 2022.

[5] 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.

[6] D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51, 1978.

[7] M. Brückerhoff and N. Juillet. Instability of martingale optimal transport in dimension $d \geq 2$. *Electron. Commun. Probab.*, 27:Paper No. 24, 10, 2022.

[8] R. C. Dalang, A. Morton, and W. Willinger. Equivalent martingale measures and no-arbitrage in stochastic securities market models. *Stochastics Stochastics Rep.*, 29(2):185–201, 1990.

[9] D. C. Dowson and B. V. Landau. The Fréchet distance between multivariate normal distributions. *J. Multivariate Anal.*, 12(3):450–455, 1982.

[10] C. R. Givens and R. M. Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Math. J.*, 31(2):231–240, 1984.

[11] N. Gozlan, P.-M. Samson, and P.-A. Zitt. Notes de cours sur le transport optimal. 2022.

[12] L. Kantorovich. On the translocation of masses. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 37:199–201, 1942.

[13] G. Monge. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'académie Royale des Sciences de Paris*, 1781.

[14] I. Olkin and F. Pukelsheim. The distance between two random vectors with given dispersion matrices. *Linear Algebra Appl.*, 48:257–263, 1982.

[15] R. T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.

[16] C. Zălinescu. *Convex analysis in general vector spaces*. World Scientific Publishing Co., Inc., River Edge, NJ, 2002.