01 Dec 2020

A Brief on Optimal Transport

Austin G. Vandegriffe

A Brief on Optimal Transport

Special Lecture
Missouri S&T : Rolla, MO

Presentation by

Austin G. Vandegriffe

2020

Ackowledgement

I want to recognize my friend and colleague, Louis K. B. Steinmeister[1], for introducing and motivating my studies in pure mathematics. Without Louis, I likely would not have discovered the intricacies of measure theory, which lead me to optimal transport, for a while longer. I would also like to thank Dr. John R. Singler[2] and Dr. Jason C. Murphy[3] for guiding me through some proofs and for the sanity checks; their input was of great value.

[1] Louis K.B. Steinmeister: https://scholar.google.com/citations?user=BFSEaMkAAAAJ&hl=en
[2] John R. Singler: https://scholar.google.com/citations?user=0XXAypYAAAAJhl=en
[3] Jason C. Murphy: https://scholar.google.com/citations?user=32q4x_cAAAAJ&hl=en

Contents

Introduction to Optimal Transport

1 Coupling and Tranport Plans

Definition 1.1 (Coupling)**.** *Given two probability spaces,* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$*, a* coupling *of* (μ_1, μ_2)*, is a measure* π *on* $\Omega_1 \times \Omega_2$ *with marginals* μ_1 *and* μ_2*, that is,*

$$\pi(F_1 \times \Omega_2) = \mu_1(F_1) \quad \forall F_1 \in \mathcal{F}_1$$
$$\pi(\Omega_1 \times F_2) = \mu_2(F_2) \quad \forall F_2 \in \mathcal{F}_2$$

Alternatively, one can view a coupling as a pair of random variables $X_i : (\Omega_*, \mathcal{F}_*, \mathbb{P}) \to (\Omega_i, \mathcal{F}_i, \mu_i)$ *satisfying* $\mu_i = X_{i\#}\mathbb{P}$*. Just find a pair* (X_1, X_2) *satisfying* $\pi = \mathbb{P} \circ (X_1^{-1}, X_2^{-1})$*; there are existence theorems (Theorem 1.104 in [3]) which show that, for a given distribution, there exists a random variable which generates that distribution.*

Claim 1.2. π *is a coupling of the probability measures* (μ_1, μ_2) *iff* $\forall(\phi_1, \phi_2) \in L^1(\Omega_1, \mu_1) \times L^1(\Omega_2, \mu_2)$*, or equivalently* $L^\infty(\Omega_1, \mu_1) \times L^\infty(\Omega_2, \mu_2)$*, we have*

$$\int_{\Omega_1 \times \Omega_2} (\phi_1 \oplus \phi_2)(x, y) d\pi(x, y) = \int_{\Omega_1} \phi_1(x)\ d\mu_1(x) + \int_{\Omega_2} \phi_2(y)\ d\mu_2(y)$$

Proof. " $\Longrightarrow$ " Take $\phi_i = \mathbb{1}_{A^{(i)}}$ for some $A^{(i)} \in \mathcal{F}_i$, then

$$\int_{\Omega_1 \times \Omega_2} \phi_1 + \phi_2\ d\pi = \sup_{(F^{(1)}, F^{(2)}) \in \mathcal{F}_1 \times \mathcal{F}_2)} (\pi(A^{(1)}, F^{(2)}) + \pi(F^{(1)}, A^{(2)}))$$

by definition of Lebesgue integration. Since measures are monotone and $A^{(1)} \times F^{(2)} \subset A^{(1)} \times \Omega_2 \in \mathcal{F}_1 \times \mathcal{F}_2\ \forall F^{(2)} \in \mathcal{F}_2$ and similarly for $F^{(1)} \times A^{(2)}$, the integral is maximized when $(F^{(1)}, F^{(2)}) = (\Omega_1, \Omega_2)$, and by the marginal restrictions of π we have

$$\int_{\Omega_1 \times \Omega_2} \phi_1 + \phi_2\ d\pi = \mu_1(A^{(1)}) + \mu_2(A^{(2)})$$

Then proceed to simple functions and then limits of simple functions.

" $\Longleftarrow$ " Take $\phi_i = \mathbb{1}_{F^{(i)}}$ for $F^{(i)} \in \mathcal{F}_i$ and $\phi_2 = 0$, then

$$\pi(F^{(1)} \times \Omega_2) = \int_{\Omega_1 \times \Omega_2} \phi_1\ d\pi = \int_{\Omega_1} \phi_1\ d\mu_1 = \mu(F^{(1)})$$

Similarly with ϕ_2 we obtain the result. □

Definition 1.3 (Transport Plans)**.** *The set of transport plans is the set of couplings on* $\Omega_1 \times \Omega_2$ *for* (μ_1, μ_2)*, that is,*

$$\Pi(\mu_1, \mu_2) = \{\pi : \mathcal{F}_1 \times \mathcal{F}_2 \to \mathbb{R}_0^+ \cup \{\infty\} \ \ | \ \pi \text{ couples } (\mu_1, \mu_2)\}$$

Claim 1.4. *Given two probability spaces,* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$*, the set of transport plans is nonempty.*

Proof. $(\mu_1 \otimes \mu_2) \in \Pi(\mu_1, \mu_2)$ □

2 Kantorovich' O. T. & Basic Properties

Let $c : \Omega_1 \times \Omega_2 \to \mathbb{R}_0^+ \cup \{\infty\}$ be a loss metric and $(\Omega_i, \mathcal{F}_i, \mu_i)$ a probability space. The Kantorovich optimal transport problem is finding a π^* satisfying

$$\pi^* \in \underset{\pi \in \Pi(\mu_1,\mu_2)}{\arg\inf} \mathbb{K}_c(\pi) = \underset{\pi \in \Pi(\mu_1,\mu_2)}{\arg\inf} \int_{\Omega_1 \times \Omega_2} c(x,y)\ d\pi(x,y)$$

where the quantity '$c(x,y)\ d\pi(x,y)$' can be interpreted as " moving the amount $d\pi(x,y)$ from x to y at a cost $c(x,y)$." The minimal cost will be denoted $\mathcal{C}_c(\mu_1,\mu_2) = \mathbb{K}_c(\pi^*)$. The problem can also be posed with random variables, using the same notation as in Definition 1.1, we have

$$(X_1^*, X_2^*) \in \underset{\substack{X_i \in \mathcal{F}_i \\ \mu_i = X_{i\#}\mathbb{P}}}{\arg\inf} \mathbb{E}_{\mathbb{P}}\left[c(X_1, X_2)\right]$$

We want to prove the following:

Theorem 2.1 (Existence of an optimal coupling)**.** *Let $(\Omega_i, \mathcal{F}_i)$ $(i = 1, 2)$ be two Polish probability spaces, i.e. a separable, completely metrizable, topological, probability space; let $a_i \in L^1(\Omega_i, \mathbb{R} \cup \{-\infty\}, \mu_i)$ $(i = 1, 2)$ be two upper semicontinuous functions. Let $c : \Omega_1 \times \Omega_2 \to \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous cost function, such that $c(x, y) \geqq a_1(x) + a_2(y)$ for all x, y. Then there is a coupling of (μ_1, μ_2) which minimizes the total cost $\mathbb{E}\left[\ c(X_1, X_2)\ \right]$ among all possible couplings (X_1, X_2).*

Note. The lower bound for $c(\bullet, \bullet)$ in Theorem 2.1 guarantees a lower bound for the Kantorovich problem, this is because, by Claim 1.2,

$$\inf_{\pi \in \Pi(\mu_1,\mu_2)} \int a_1 + a_2\ d\pi = \int a_1 + a_2\ d(\mu_1 \otimes \mu_2) \leqq \mathbb{K}_c(\pi^*)$$

and since the 'a_i's are integrable, we obtain a lower bound. □

To prove Theorem 2.1, we will first need a few lemmas.

Lemma 2.2. *Let f be a nonnegative lower semicontinuous function on Ω. If $(\mu_n)_{n \geqq 1}$ converges narrowly to μ, then*

$$\int f\ d\mu \leqq \liminf_{n \uparrow \infty} \int f\ d\mu_n$$

Proof. Since g is lower semicontinuous $\exists (f_n)_{n \geqq 1} \subset \mathcal{C}_b^0(\Omega)$ such that $f_n \uparrow f$ by Lemma C.1. Let $\epsilon > 0$, by the Beppo-Levi lemma for nonnegative increasing measurable functions, we have that $\exists K \in \mathbb{N}$ s.t. $\forall k > K$

$$\left| \int f\ d\mu - \int f_k\ d\mu \right| = \int f\ d\mu - \int f_k\ d\mu \leqq \epsilon$$

rearranging we obtain

$$\int f \; d\mu \leqq \int f_k \; d\mu + \epsilon \tag{$\star$}$$

Now, by narrow convergence, we have that $\exists N \in \mathbb{N}$ s.t. $\forall n \geqq N$

$$\left| \int f_k \; d\mu - \int f_k \; d\mu_n \right| < \epsilon$$

by definition of the absolute value we have

$$-\epsilon < \int f_k \; d\mu - \int f_k \; d\mu_n < \epsilon$$

and then we add ϵ to both sides

$$0 < \int f_k \; d\mu - \int f_k \; d\mu_n < \int f_k \; d\mu - \int f_k \; d\mu_n + \epsilon < 2\epsilon \tag{$\star\star$}$$

"Adding zero" to ($\star$), applying ($\star\star$), then recalling that $\int f_k \; d\mu_n \leqq \int f \; d\mu_n$ since $f_k \leqq f \; \forall \; k \in \mathbb{N}$ we obtain

$$\begin{aligned}
\int f \; d\mu &\leqq \int f_k \; d\mu \pm \int f_k \; d\mu_n + \epsilon \\
&= \int f_k \; d\mu_n + \left(\int f_k \; d\mu - \int f_k \; d\mu_n \right) + \epsilon \\
&\leqq \int f_k \; d\mu_n + 3\epsilon \\
&\leqq \int f \; d\mu_n + 3\epsilon
\end{aligned}$$

Now, by taking the lim inf in n and taking $\epsilon \downarrow 0$ we obtain the result

$$\liminf_{n\uparrow\infty} \int f \; d\mu = \int f \; d\mu \leqq \liminf_{n\uparrow\infty} \int f \; d\mu_n$$

□

Lemma 2.3 (Lower semicontinuity of the cost functional)**.** *Let Ω_1 and Ω_2 be two Polish spaces, and $c : \Omega_1 \times \Omega_2 \to \mathbb{R} \cup \{\infty\}$ a lower semicontinuous cost function. Let $h : \Omega_1 \times \Omega_2 \to \mathbb{R} \cup \{-\infty\}$ be an upper semicontinuous function such that $c \geqq h$ everywhere. Let $(\pi_k)_{k\in\mathbb{N}}$ be a sequence of probability measures on $\Omega_1 \times \Omega_2$, converging weakly to some $\pi \in \mathscr{P}(\Omega_1 \times \Omega_2)$, in such a way that $h \in L^1(\pi_k) \cap L^1(\pi)$, and*

$$\int\limits_{\Omega_1\times\Omega_2} h \; d\pi_k \xrightarrow{k\uparrow\infty} \int\limits_{\Omega_1\times\Omega_2} h \; d\pi$$

Then

$$\int\limits_{\Omega_1\times\Omega_2} c \; d\pi \leqq \liminf_{k\uparrow\infty} \int\limits_{\Omega_1\times\Omega_2} c \; d\pi_k$$

Proof. Replace c by $c - h$, a non-negative lower semicontinuous function, and apply the previous lemma. $\square$

Lemma 2.4 (Tightness of transference plans)**.** *Let Ω_1 and Ω_2 be two Polish spaces. Let $\mathcal{P}_1 \subset \mathscr{P}(\mathcal{F}_1)$ and $\mathcal{P}_2 \subset \mathscr{P}(\mathcal{F}_2)$ be tight subsets of $\mathscr{P}(\mathcal{F}_1)$ and $\mathscr{P}(\mathcal{F}_2)$ respectively. Then the set $\Pi(\mathcal{P}_1, \mathcal{P}_1)$ of all transference plans whose marginals lie in $\mathcal{P}_1$ and $\mathcal{P}_2$ respectively, is itself tight in $\mathscr{P}(\mathcal{F}_1 \times \mathcal{F}_2)$.*

Proof. Let $\mu_1 \in \mathcal{P}_1$, $\mu_2 \in \mathcal{P}_1$, and $\pi \in \Pi(\mu_1, \mu_2)$. By Ulam's tightness theorem (and by assumption), we have that

$$\forall \epsilon > 0 \ \exists \text{ compact } \mathcal{K}_\epsilon^{(i)} \subset \Omega_i \ (\perp\!\!\!\perp \mu_i) \text{ s.t. } \mu_i(\Omega_i \setminus \mathcal{K}_\epsilon^{(i)}) \leqq \epsilon$$

Let (X_1, X_2) be a coupling of (μ_1, μ_2), that is, $law(X_i) = X_{i\#}\mathbb{P} = \mu_i$, then

$$\begin{aligned}
\implies\ & \mathbb{P}((X_1, X_2) \notin \mathcal{K}_\epsilon^{(1)} \times \mathcal{K}_\epsilon^{(2)}) \\
&= \mathbb{P}(\{\ \omega \mid X_1(\omega) \notin \mathcal{K}_\epsilon^{(1)}\} \cap \{\ \omega \mid X_2(\omega) \notin \mathcal{K}_\epsilon^{(2)}\}) \\
&= \mathbb{P}(X_1^{-1}(\Omega_1 \setminus \mathcal{K}_\epsilon^{(1)}) \cap X_2^{-1}(\Omega_2 \setminus \mathcal{K}_\epsilon^{(2)})) \\
&\leqq \mathbb{P}(X_1^{-1}(\Omega_1 \setminus \mathcal{K}_\epsilon^{(1)}) \cup X_2^{-1}(\Omega_2 \setminus \mathcal{K}_\epsilon^{(2)})) \\
&\leqq \mathbb{P}(X_1^{-1}(\Omega_1 \setminus \mathcal{K}_\epsilon^{(1)})) + \mathbb{P}(X_2^{-1}(\Omega_2 \setminus \mathcal{K}_\epsilon^{(2)})) \\
&= \mu_1(\Omega_1 \setminus \mathcal{K}_\epsilon^{(1)})) + \mu_2(\Omega_2 \setminus \mathcal{K}_\epsilon^{(2)})) \\
&\leqq \epsilon + \epsilon = 2\epsilon \perp\!\!\!\perp \mu_i
\end{aligned}$$

And since $\mathcal{K}_\epsilon^{(i)}$ is compact, we have, by Theorem A.7 (Tychonoff), that $\mathcal{K}_\epsilon^{(1)} \times \mathcal{K}_\epsilon^{(2)} \subset \Omega_1 \times \Omega_2$ too is compact; therefore, $\Pi(\mathcal{P}_1, \mathcal{P}_2)$ is tight. $\square$

Proof of Theorem 2.1. Since Ω_i is a Polish space, we have, by Theorem B.7 (Ulam), that μ_i is tight and, by Lemma 2.4, that $\Pi(\mu_1, \mu_2)$ is tight, and so by Theorem B.9 (Prokhorov) this set has a compact closure. Now, take $(\pi_k)_{k \geqq 1} \subset \Pi(\mu_1, \mu_2)$ s.t. $\pi_k \xrightarrow{k \uparrow \infty} \pi$ in the narrow sense; we want to show that $\pi \in \Pi(\mu_1, \mu_2)$, i.e. has margins μ_1 and μ_2. Let $(\phi_1, \phi_2) \in L^1(\mu_1) \times L^1(\mu_2)$, then we have

$$\int\limits_{\Omega_1 \times \Omega_2} \phi_1 + \phi_2 \ d\pi_k = \int\limits_{\Omega_1} \phi_1 \ d\mu_1 + \int\limits_{\Omega_2} \phi_2 \ d\mu_2 \xrightarrow{k \uparrow \infty} \int\limits_{\Omega_1 \times \Omega_2} \phi_1 + \phi_2 \ d\pi$$

hence

$$\int\limits_{\Omega_1 \times \Omega_2} \phi_1 + \phi_2 \ d\pi = \int\limits_{\Omega_1} \phi_1 \ d\mu_1 + \int\limits_{\Omega_2} \phi_2 \ d\mu_2$$

and by Claim 1.2 we conclude that $\pi \in \Pi(\mu_1, \mu_2)$; hence, $\Pi(\mu_1, \mu_2)$ is closed, and since it has a compact closure, we have that it is compact. Now let $(\pi_k)_{k \geqq 1}$ be the minimizing sequence for $\int c \ d\pi_k$ which converges to the optimal transport cost. Since $\Pi(\mu_1,\mu_2)$ is compact, take a narrowly convergent subsequence to $\pi \in \Pi(\mu_1,\mu_2)$. Notice that

$$h : \Omega_1 \times \Omega_2 \ni (x_1,x_2) \mapsto a_1(x_1) + a_2(x_2) = h(x_1,x_2) \in \mathbb{R}$$

is $L^1(\pi_k) \cap L^1(\pi)$ and, by assumption, $c \geqq h$ everywhere; moreover,

$$\int h \; d\pi_k = \int h \; d\pi = \int a_1 \; d\mu_1 + \int a_2 \; d\mu_2$$

Therefore, with Lemma 2.3 on $c-h$, we have

$$\int c \; d\pi \leqq \liminf_{k\uparrow\infty} \int c \; d\pi_k$$

thus π is minimizing. □

Theorem 2.5 (Optimality is inherited by restriction)**.** *Let* $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1,2)$ *be two Polish spaces,* $a_i \in L^1(\Omega_i,\mu_i)$*, and let* $c : \Omega_1 \times \Omega_2 \to \mathbb{R} \cup \{\infty\}$ *be a measurable cost function such that* $c \geqq a_1 + a_2$*; let* $\mathcal{C}_c(\mu_1,\mu_2)$ *be the optimal transport cost from* μ_1 *to* μ_2*. Assume* $\mathcal{C}_c(\mu_1,\mu_2) < \infty$ *and let* $\pi \in \Pi(\mu_1,\mu_2)$ *be an optimal transport plan. Let* $\tilde{\pi}$ *be a nonnegative measure on* $\mathcal{F}_1 \times \mathcal{F}_2$ *such that* $\tilde{\pi} \leqq \pi$ *and* $\pi(\Omega_1 \times \Omega_2) > 0$*. Then the probability measure*

$$\pi' = \frac{\tilde{\pi}}{\tilde{\pi}(\Omega_1 \times \Omega_2)} = \frac{\tilde{\pi}}{\tilde{Z}}$$

is an optimal transference plan between its marginals μ_1' *and* μ_2'*.*

Moreover, if π *is the unique optimal transference plan between* μ_1 *and* μ_2*, then* π' *is the unique optimal transference plan between* μ_1' *and* μ_2'*.*

Proof. Suppose π' is not optimal, then $\exists\ \pi''$ such that, for all $F^{(i)} \in \mathcal{F}_i$,

$$\pi''(\bullet \times F^{(2)}) = \mu_1', \qquad \pi''(F^{(1)} \times \bullet) = \mu_2'$$

and

$$\int c \; d\pi'' < \int c \; d\pi'$$

Now, consider

$$\begin{aligned}\hat{\pi} &= (\pi - \tilde{\pi}) + \tilde{Z}\pi'' \\ &= (\pi - \tilde{Z}\frac{\tilde{\pi}}{\tilde{Z}}) + \tilde{Z}\pi'' \\ &= (\pi - \tilde{Z}\pi') + \tilde{Z}\pi'' \\ &= \pi + \tilde{Z}(\pi'' - \pi')\end{aligned}$$

where $\tilde{Z} = \tilde{\pi}(\Omega_1 \times \Omega_2) > 0$ by assumption. It is clear that $\hat{\pi}$ is nonnegative since $\tilde{\pi} \leqq \pi$ and $\pi'' \geqq 0$. Note that $\hat{\pi} \in \Pi(\mu_1,\mu_2)$, that is, for all $F^{(i)} \in \mathcal{F}_i$

$$\begin{cases}\hat{\pi}(F^{(1)} \times \Omega_2) = \mu_1(F^{(1)} + \tilde{Z}\ \big(\ \mu_1'(F^{(1)}) - \mu_1'(F^{(1)})\ \big) = \mu_1(F^{(1)}) \\ \\ \hat{\pi}(\Omega_1 \times F^{(2)}) = \mu_1(F^{(2)} + \tilde{Z}\ \big(\ \mu_2'(F^{(2)}) - \mu_2'(F^{(2)})\ \big) = \mu_2(F^{(2)})\end{cases}$$

Since $\int c\ d(\tilde{Z}(\pi'' - \pi')) < 0$, we obtain

$$\int c\ d\hat{\pi} = \int c\ d\pi + \int c\ d(\tilde{Z}(\pi'' - \pi')) < \int c\ d\pi$$

which contracts the optimally of π; therefore, π' is optimal. Now suppose π is a unique optimal transference plan, let π' and $\pi'' \in \Pi(\mu_1', \mu_2')$ be optimal, define $\hat{\pi}$ as above and note that $\hat{\pi} \leqq \pi$ (since π is optimal), hence $\hat{\pi} = \pi$ yielding

$$\int \phi\ d\hat{\pi} = \int \phi\ d\pi + \int \phi\ d(\tilde{Z}(\pi'' - \pi')) = \int \phi\ d\pi$$

$\forall\ \phi \in Ł^1(\pi)$, and so, by Claim B.5 (bounded continuous equality, **not** $L^1(\pi)$, suppose ∞ at a point where $\pi \neq 0$ and $\pi' = 0$) on (μ_1', μ_2'), $\pi' = \pi''$ and so π' is unique. $\square$

3 The Wasserstein Distance

We want to be able to say that

$$\mathcal{C}_c(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1, \mu_2)} \mathbb{K}_c(\pi)$$

is the "distance between μ_1 and μ_2", but, in general, $\mathcal{C}_c(\bullet, \bullet)$ does not satisfy the axioms of a distance function, i.e. a metric; however, we obtain such a metric characteristic when c is a metric such as ℓ^p for some $p \in \mathbb{N}$.

Definition 3.1. *Let (Ω, d) be a Polish metric space, and let $p \in [1, \infty)$. For any two probability measures μ_1, μ_2 in $(\Omega, \mathcal{F})$, the* Wasserstein distance of order p *between μ_1 and μ_2 is defined by the formula*

$$\mathcal{W}_p(\mu_1, \mu_2) = \mathcal{C}^{\frac{1}{p}}_{d^p(\bullet,\bullet)}(\mu_1, \mu_2) = \left(\inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{\Omega} d^p(x, y) \; d\pi \right)^{\frac{1}{p}}$$
$$= \inf_{\substack{X_i \in \mathcal{F} \\ law(X_i) = \mu_i}} \left(\mathbb{E}\left[d^p(X_1, X_2)\right]^{\frac{1}{p}} \right)$$

Note: In the following I add a *compactness* assumption to Ω which has not been there up to now. As I work through [2] and build the presentation for a more general approach to optimal transport, I will add similar theorems with weaker assumptions.

Theorem 3.2 ($\mathcal{W}_p$ is a metric on $\mathscr{P}(\mathcal{F})$)**.** *Let $(\Omega, \mathcal{F}, d)$ be a measurable compact Polish metric space, then $\mathcal{W}_p$ is a metric on $\mathscr{P}(\mathcal{F})$.*

Lemma 3.3 (Gluing Lemma)**.** *Let $(\Omega_i, \mathcal{F}_i, \mu_i)$ $(i = 1, 2, 3)$ be a compact measured Polish space with associated transport plans $\pi_{12} \in \Pi(\mu_1, \mu_2)$ and $\pi_{23} \in \Pi(\mu_2, \mu_3)$, then $\exists\ \pi_{123} \in \mathscr{P}(\mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3)$ with marginals π_{12} and π_{23}.*

Proof. Let $V \subset C^0_b(\Omega_1 \times \Omega_2 \times \Omega_3)$ be the vector space

$$V = \{\phi_{12}(x_1, x_2) + \phi_{23}(x_2, x_3) : \phi_{12} \in C^0_b(\Omega_1 \times \Omega_2), \phi_{23} \in C^0_b(\Omega_2 \times \Omega_3)\}$$

and define a functions $G : V \to \mathbb{R}$ by

$$G(\phi_{12} + \phi_{23}) = \int_{\Omega \times \Omega} \phi_{12} \; d\pi_{12} + \int_{\Omega \times \Omega} \phi_{23} \; d\pi_{23}$$

We will now show that G is well defined. Let $\phi_{12} + \phi_{23} = \hat{\phi}_{12} + \hat{\phi}_{23}$, lets perturb x_1 by Δx_1

$$\phi_{12}(x_1 + \Delta x_1, x_2) - \tilde{\phi}_{12}(x_1 + \Delta x_1, x_2) = \tilde{\phi}_{23}(x_2, x_3) - \phi_{23}(x_2, x_3)$$
$$= \phi_{12}(x_1, x_2) - \tilde{\phi}_{12}(x_1, x_2)$$

and similarly for x_3 we obtain, with the equality restriction, that $\phi_{12}(x_1, x_2) - \tilde{\phi}_{12}(x_1, x_2)$ and $\tilde{\phi}_{23}(x_2, x_3) - \phi_{23}(x_2, x_3)$ are functions of x_2. Thus

$$\begin{aligned}\int\limits_{\Omega\times\Omega} \phi_{12} - \tilde{\phi}_{12}\ d\pi_{12} &= \int\limits_{\Omega} \phi_{12} - \tilde{\phi}_{12}\ d\mu_2 \\ &= \int\limits_{\Omega} \phi_{23} - \tilde{\phi}_{23}\ d\mu_2 \\ &= \int\limits_{\Omega\times\Omega} \phi_{23} - \tilde{\phi}_{23}\ d\pi_{23}\end{aligned}$$

and, by rearranging, we obtain

$$G(\phi_{12} + \phi_{23}) = G(\tilde{\phi}_{12} + \tilde{\phi}_{23})$$

and so G is well defined. Clearly G is bounded and linear (its an integral), we must show that it is *positive* linear. Let

$$\phi_{12}(x_1, x_2) + \phi_{23}(x_2.x_3) \geqq 0$$

then, with both sides being functions of x_2, we have

$$\phi_{12}(x_1, x_2) \geqq -\phi_{23}(x_2, x_3) \geqq -\inf_{x_3} \phi_{23}(x_2, x_3)$$

and that the infimum exists since ϕ_{23} is bounded. We now have

$$\begin{cases} \int\limits_{\Omega_1\times\Omega_2} \phi_{12}\ d\pi_{12} \geqq \int\limits_{\Omega_1\times\Omega_2} -\inf\limits_{x_3} \phi_{23}(x_2, x_3)\ d\pi_{12} \geqq -\int\limits_{\Omega_2} \inf\limits_{x_3} \phi_{23}(x_2, x_3)\ d\mu_2 \\ \int\limits_{\Omega_2\times\Omega_3} \phi_{23}\ d\pi_{23} \geqq \int\limits_{\Omega_2\times\Omega_3} \inf\limits_{x_3} \phi_{23}(x_2, x_3)\ d\pi_{23} \geqq \int\limits_{\Omega_2} \inf\limits_{x_3} \phi_{23}(x_2, x_3)\ d\mu_2 \end{cases}$$

Using the above lower bounds, we obtain

$$\begin{aligned}G(\phi_{12} + \phi_{23}) &= \int\limits_{\Omega_1\times\Omega_2} \phi_{12}\ d\pi_{12} + \int\limits_{\Omega_2\times\Omega_3} \phi_{23}\ d\pi_{23} \\ &\geqq -\int\limits_{\Omega_2} \inf_{x_3} \phi_{23}(x_2, x_3)\ d\mu_2 + \int\limits_{\Omega_2} \inf_{x_3} \phi_{23}(x_2, x_3)\ d\mu_2 \\ &= 0\end{aligned}$$

Thus, G is a positive linear functional. So, by Theorem C.5 (Hahn-Banach Positive Extension)[with $\Theta = \mathcal{C}_b^0(\Omega_1 \times \Omega_2 \times \Omega_3)$ and $\Theta' = V$, and $\rho(\bullet) = \sup\limits_{x\in\Omega}(\bullet)$

in the definition], $\exists\ \hat{G} : \mathcal{C}_b^0(\Omega_1 \times \Omega_2 \times \Omega_3) \to \mathbb{R}$, and by the Theorem C.6 (Riesz representation) we have $\exists\ \pi_{123} \in \mathscr{P}(\mathcal{F}_1 \times \mathcal{F}_2 \times \mathcal{F}_3)$ corresponding to $\hat{G}$ yielding

$$\begin{aligned}\int\limits_{\Omega^{\otimes 3}} \phi_{12} + \phi_{23}\ d\pi_{123} &= \hat{G}(\phi_{12} + \phi_{23}) \\ &= G(\phi_{12} + \phi_{23}) \\ &= \int\limits_{\Omega\times\Omega} \phi_{12}\ d\pi_{12} + \int\limits_{\Omega\times\Omega} \phi_{23}\ d\pi_{23} \quad \forall\ (\phi_{12} + \phi_{23}) \in V\end{aligned}$$

and we have, from Theorem B.5, that π_{123} has marginals π_{12} and π_{23} as desired. □

Proof of Theorem 3.2. We must show that $\mathcal{W}_p$ satisfies the properties of a metric in Definition A.1. It is clear that $\mathcal{W}_p$ is non-negative, symmetric, and finite (since the infimum is achieved). Now, suppose $\mu_1 = \mu_2 = \mu$, then there exists a random variable $X : (\Omega_*, \mathcal{F}_*, \mathbb{P}) \to (\Omega, \mathcal{F})$ such that $\mu = X_{\#}\mathbb{P}$, then, with $X_1 = X_2 = X$ in the definition of the Kantorovich problem, we obtain

$$\int\limits_{\Omega_*} d^p(X(\omega), X(\omega))\ d\mathbb{P}(\omega) = 0$$

so $\mathcal{W}_p(\mu, \mu) = 0\ \forall\ \mu \in \mathscr{P}(\mathcal{F})$. Now let $\mu_1, \mu_2 \in \mathscr{P}(\mathcal{F})$ (not necessarily equal). If $\mathcal{W}_p(\mu_1, \mu_2) = 0$, then π^* must concentrate all of its mass on the diagonal $\Delta_\Omega \subset \Omega \times \Omega$; suppose it didn't, then $\exists F \in \mathcal{F} \times \mathcal{F}\big|_{\Delta_\Omega^c}$ s.t. $\pi^*(F) > 0$ and $\sup\limits_{(x_1,x_2)\in F} (d(x_1, x_2)) > 0$, so we have

$$\left(\int\limits_F d^p(x_1, x_2)\ d\pi^*\right)^{\frac{1}{p}} \leqq \mathcal{W}_p(\mu_1, \mu_2)$$

which contradicts $\mathcal{W}_p(\mu_1, \mu_2) = 0$. With this, we have that $\forall\ u \in \mathcal{C}_b^0(\Omega)$

$$\int\limits_\Omega u(x)\ d\mu_1(x) = \int\limits_{\Omega\times\Omega} u(x)\ d\pi^*(x, y) = \int\limits_{\Omega\times\Omega} u(y)\ d\pi^*(x, y) = \int\limits_\Omega u(y)\ d\mu_2(y)$$

where the second equality comes from the concentration on Δ_Ω; thus, $\mu_1 = \mu_2$ by Theorem B.5.

Now let $\mu_i \in \mathscr{P}(\mathcal{F})$ $(i = 1, 2, 3), \pi_{12} \in \mathbb{K}(\mu_1, \mu_2), \pi_{23} \in \mathbb{K}(\mu_2, \mu_3)$, and, by the Lemma 3.3, $\pi_{123} \in \mathscr{P}(\mathcal{F}^{\otimes 3})$ coupling π_{12} and π_{23}. Letting $\pi_{13}(\bullet, \bullet) =$

$\pi_{123}(\bullet, \Omega, \bullet)$ (not necessarily optimal), we obtain

$$\begin{aligned}
\mathcal{W}_p(\mu_1, \mu_3) &\leqq \left(\int\limits_{\Omega\times\Omega} d^p(x_1, x_2)\ d\pi_{13} \right)^{\frac{1}{p}} \\
&= \left(\int\limits_{\Omega^{\otimes 3}} d^p(x_1, x_3)\ d\pi_{123} \right)^{\frac{1}{p}} \\
&\leqq \left(\int\limits_{\Omega^{\otimes 3}} [\ d(x_1, x_2) + d(x_2, x_3)\ ]^p\ \ d\pi_{123} \right)^{\frac{1}{p}} \\
&\leqq \left(\int\limits_{\Omega^{\otimes 3}} d^p(x_1, x_2)\ d\pi_{123} \right)^{\frac{1}{p}} + \left(\int\limits_{\Omega^{\otimes 3}} d^p(x_2, x_3)\ d\pi_{123} \right)^{\frac{1}{p}} \\
&= \left(\int\limits_{\Omega\times\Omega} d^p(x_1, x_2)\ d\pi_{12} \right)^{\frac{1}{p}} + \left(\int\limits_{\Omega\times\Omega} d^p(x_2, x_3)\ d\pi_{23} \right)^{\frac{1}{p}} \\
&= \mathcal{W}_p(\mu_1, \mu_2) + \mathcal{W}_p(\mu_2, \mu_3)
\end{aligned}$$

which proves the triangle inequality. □

Preliminaries*

*This section has not been developed in detail as the focus of this text is optimal transport. The details are left as a future work.

A Topology

Definition A.1 (Topology, Open/Closed Sets, and Metrizablility)**.** *Let Ω be a non-empty set, then a collection $\tau \subseteq 2^{\Omega}$ is a* topology *if*

i. $\emptyset, \Omega \in \tau$

ii. $T_1, T_2 \in \tau \implies T_1 \cap T_2 \in \tau$

iii. $(T_i)_{i \in \mathcal{I} \subset \mathbb{R}} \subset \tau \implies \bigcup_{i \in \mathcal{I}} T_i \in \tau$

Sets in τ are called open, *if $T^c \in \tau$ then T is closed, and if T is both open and closed it is* clopen. *If $\exists$ a metric on Ω which induces the topology, then (Ω, τ) is called metrizable, that is, $\exists d : \Omega \times \Omega \to \mathbb{R}$ satisfying $\forall\ x_i \in \Omega$ $(i = 1, 2, 3)$*

i. $d(x_1, x_2) \geqq 0$ *with* $d(x_1, x_2) = 0 \iff x_1 = x_2$

ii. $d(x_1, x_2) = d(x_2, x_1)$

iii. $d(x_1, x_3) \leqq d(x_1, x_2) + d(x_2, x_3)$

and $\tau = \{\ \{x \in \Omega : d(x_0, x) < r)\} : r \in \overline{\mathbb{R}}_0^+\ \&\ x_0 \in \Omega\ \}$.

Definition A.2 (Hausdorff Condition)**.** (Ω, τ) *is called* Hausdorff *if* $\forall\ x_1 \neq x_2 \in \Omega\ \exists U_i \in \tau\,|_{\ni x_i}$ *s.t.* $U_{x_1} \bigcap U_{x_2} = \emptyset$.

Definition A.3 (Polish Space)**.** *A topological space (Ω, τ) is Polish if it is a separable, completely metrizable, topological space, that is, a space homeomorphic to a complete metric space that has a countable dense subset.*

Definition A.4 (Lower Semicontinuity)**.** *Let (Ω, τ) be a topological space, a function $f : \Omega \to \mathbb{R}$ is lower semicontinuous if one of the following holds*

i. $\{x : f(x) > \alpha\} \in \tau \quad \forall\ \alpha \in \mathbb{R}$

ii. $\{x : f(x) \leqq \alpha\}^c \in \tau \quad \forall\ \alpha \in \mathbb{R}$

iii. If τ is metrized by d, $\forall\ (x_0 \in \Omega, \epsilon > 0)\ \exists\ \delta(x_0, \epsilon) = \delta > 0$ s.t.

$$d(x_0, x) < \delta \implies f(x) > f(x_0) - \epsilon$$

Definition A.5 (Limit Points and Convergence)**.** *A point x^* is a limit point of $A \subset \Omega$ if $\forall U \in \tau\,|_{\ni x^*}\ A \bigcap U \neq \emptyset$. A sequence $(x_n)_{n \geqq 1}$* converges *to a point $x^* \in \Omega$ if $\forall\ U \in \tau\,|_{\ni x^*}\ \exists N \in \mathbb{N}$ s.t. $x_n \in U\ \forall n \geqq N$; if τ is metrized by d, then the former can be formulated as*

$$\forall\ \epsilon > 0\ \exists N \in \mathbb{N}\ s.t.\ d(x^*, x_n) < \epsilon\ \forall\ n \geqq N$$

Theorem A.6 (Equivalence of Compactness (Theorem 28.2, pg. 179 [8]))**.** *If a set $\mathcal{C} \subset \Omega$ is compact, then the following are equivalent*

i. $\forall\, (\ \mathcal{I} \subseteq \mathbb{R}, (T_i)_{i\in\mathcal{I}} \subseteq \tau)\ \ s.t.\ \mathcal{C} \subseteq \bigcup_{i\in\mathcal{I}} T_i\ \exists (C_n)_{n=1}^{N<\infty} \subseteq (T_i)_{i\in\mathcal{I}}\ s.t.\ \mathcal{C} \subseteq \bigcup_{n=1}^{N} C_i$

ii. $\forall\ (x_n)_{n\geqq 1} \subseteq \mathcal{C}\ \exists (x_{n_k})_{k\geqq 1} \subseteq (x_n)_{n\geqq 1}\ \&\ x \in \mathcal{C}\ \ s.t.\ \ x_{n_k} \xrightarrow{k\uparrow\infty} x$

iii. Every infinite subset of $\mathcal{C}$ has a limit point

Theorem A.7 (Tychonoff's Compactness (Theorem 37.3, pg. 234 [8]))**.** *An arbitrary product of compact spaces is compact in the product topology.*

B Measure Theory

Definition B.1 (σ-Algebra)**.** *Let Ω be a non-empty set. A collection $\mathcal{F} \subset 2^{\Omega}$ is a σ-algebra if*

i. $\Omega \in \mathcal{F}$

ii. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$

iii. $(A_i)_{i \geqq 1} \subset \mathcal{F} \implies \bigcup_{i \geqq 1} A_i \in \mathcal{F}$

Definition B.2 (Measurable Maps / Random Variables)**.** *A map $X : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ is called* measurable *if*

$$X^{-1}(\mathcal{F}') \subset \mathcal{F}$$

Definition B.3 (Measure / Probability Measure)**.** *Let $(\Omega, \mathcal{F})$ be a measureable space, the set function $\mu : \mathcal{F} \to \mathbb{R}_0^+ \cup \{\infty\}$ is a measure if*

i. $\mu(\emptyset) = 0$

ii. $\mu(\biguplus A_i) = \sum_{i \geqq 1} \mu(A_i)$

The measure μ is called a probability measure *if $\mu(\Omega) = 1$. The space of all probability measures on $(\Omega, \mathcal{F})$ is denoted $\mathscr{P}(\mathcal{F})$.*

Definition B.4 (Image Measure / Distribution)**.** *Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\Omega', \mathcal{F}')$ be a random variable, then we can endow $(\Omega', \mathcal{F}')$ with the* distribution*, often called the* image measure*, $\mu = X_{\#}\mathbb{P} = \mathbb{P} \circ X^{-1}$*

Theorem B.5 (Eq. 1.2 pg. 18 in [1])**.** *Let $(\Omega, \mathcal{F})$ be a measurable Polish space and $\mu_i \in \mathscr{P}(\mathcal{F})$ $(i = 1, 2)$, then $\mu_1 = \mu_2$ if*

$$\int_{\Omega} \phi \, d\mu_1 = \int_{\Omega} \phi \, d\mu_2 \qquad \forall \phi \in \mathcal{C}_b^0(\Omega)$$

Proof. Assume that (X, d) is a metric space and let F be a closed subset of X. For $S \subset X$ and $x \in S$, define $d(x, S) := \inf\{d(x, y), y \in S\}$. Let $O_n := \left\{x \in X, d(x, F) < n^{-1}\right\}$. Then O_n is open and the map

$$f_n : x \mapsto \frac{d(x, X \setminus O_n)}{d(x, X \setminus O_n) + d(x, F)}$$

is continuous and bounded. It converges pointwise and monotonically to the characteristic function of F. So we get by monotone convergence that $\mu(F) = \nu(F)$ for all closed set F. Now given a Borel set B and $\varepsilon > 0$, we can find a closed set F and an open set O such that $F \subset S \subset O$ and $\mu(O \setminus S) \leqslant \varepsilon$. $\square$

Definition B.6 (Tightness)**.** *A family of finite measures $\mathcal{P} \subset \mathscr{P}_f(\mathcal{F})$ is called tight if*

$$\forall \epsilon > 0 \; \exists \; \textit{compact} \; \mathcal{K} \subset \Omega \; \textit{s.t.} \; \sup_{\mu \in \mathcal{P}} (\mu(\Omega \setminus \mathcal{K})) < \epsilon$$

Lemma B.7 (Ulam's theorem)**.** *If (Ω, τ) is a Polish space, and $(\Omega, \mathcal{B}(\tau), \mu)$ is a probability space, then μ is tight.*

Proof. Result from measure theory, see Theorem 2.49 in [4]. □

Definition B.8 (Weak Convergence in Measure)**.** *A sequence of measures $(\mu_n)_{n \geqq 1}$ converges* narrowly *to μ if $\forall \; f \in \mathcal{C}_b^0(\Omega)$*

$$\int f \; d\mu_n \xrightarrow{n \uparrow \infty} \int f \; d\mu$$

Lemma B.9 (Prokhorov's theorem)**.** *If (Ω, τ) is a Polish space, then a set $\mathcal{P} \subset \mathscr{P}(\mathcal{B}(\tau))$ is precompact for the weak topology if and only if it is tight.*

Proof. Result from measure theory, see Theorem 13.29 in [3]. □

C Analysis

Theorem C.1 (Baire's Theorem)**.** *If f is a nonnegative lower semicontinuous function on Ω, then $\exists (f_n)_{n \geqq 1} \subset \mathcal{C}^0(\Omega)$ such that $f_n \uparrow f$ pointwise.*

Proof. Let

$$f_n(\bullet) = \inf_{z \in \Omega} (f(z) + n \cdot d(\bullet, z))$$

Then all we must show is that f_n is (i) increasing, (ii) continuous, (iii) convergent to f.

(i) We have the inequality $f(z) + n \cdot d(x,z) \leqq f(z) + (n+1) \cdot d(x,z)$. Now taking the inf in z on both sides yields

$$\inf_{z \in \Omega} (f(z) + n \cdot d(x,z)) = f_n(x) \leqq f_{n+1}(x) = \inf_{z \in \Omega} (f(z) + (n+1) \cdot d(x,z))$$

Also note that z can be x, so we have $d(x,x) = 0$ and hence

$$f_n \leqq f \qquad (\star)$$

(ii) Choose $x, y, z \in \Omega$, then, by the triangle inequality, we have

$$f(z) + n \cdot d(z,x) \leqq f(z) + n \cdot d(z,y) + n \cdot d(y,x)$$

Taking the inf in z on both sides and subtracting $f_n(y)$ from both sides

$$f_n(x) - f_n(y) \leqq n \cdot d(y,x)$$

Since x and y were arbitrary one obtains

$$|f_n(x) - f_n(y)| \leqq n \cdot d(x,y)$$

Now let $\epsilon > 0$ and $\delta = \dfrac{\epsilon}{n}$, then we have

$$d(x,y) < \delta \implies |f_n(x) - f_n(y)| \leqq n \cdot d(y,x) < n \cdot \left(\frac{\epsilon}{n}\right) = \epsilon \quad \forall x, y$$

and we have uniform continuity.

(iii) Lower semicontinuity yields, for a fixed $x_0 \in \Omega$, that $\forall \epsilon > 0 \; \exists \; \delta > 0$ such that

$$d(x_0, z) < \delta \implies f(z) > f(x_0) - \epsilon \qquad (\star\star)$$

Now suppose $d(x_0, z) > \delta$, then $\exists \; N \in \mathbb{N}$ such that $\forall \; n \geqq N$

$$f(z) + n \cdot d(x_0, z) \geqq n \cdot \delta > f(x_0)$$

since $f > 0$. However, we have, from ($\star$), that $f_n \leqq f \; \forall \; n \in \mathbb{N}$, so

$$d(\; x_0 \;, \; \arg\inf_{z \in \Omega} (f(z) + n \cdot d(x_0, z)) \;) < \delta$$

But, for all $z \in \Omega$ s.t. $d(x_0, x) < \delta$ we have, from $(\star\star)$,

$$f(z) + n \cdot d(x_0, z) \geqq f(z) > f(x_0) - \epsilon$$

and so

$$\inf_{\substack{z\in\Omega \\ d(x_0,z)<\delta}} (f(z) + n \cdot d(x_0, z)) = f_n(x_0) > f(x_0) - \epsilon$$

Therefore,

$$f(x_0) - \epsilon \leqq f_n(x_0) \leqq f(x_0)$$

and so, taking $\epsilon \downarrow 0$ and $n \uparrow \infty$, we have $f_n(x_0) \to f(x_0)$. Since $x_0 \in \Omega$ was arbitrary, we have that $f_n \to f$ pointwise.

□

Theorem C.2 (Hahn-Banach Extension). *Let* $(\Theta, ||\bullet||)$ *be a normed linear space, let* $\Theta' \subset \Theta$ *be a linear subspace and let* $\ell \in (\Theta')^*$, *then* $\exists\ \tilde{\ell} \in \Theta^*$ *such that* $\ell(\omega) = \tilde{\ell}(\omega)\ \forall\ \omega \in \Theta'$.

Definition C.3 (Riesz Space). *A Riesz space* $\mathcal{R}$ *is a vector space endowed with a partial order, that is,* $\forall\ x_1, x_2, x_3 \in \mathcal{R}$

i. $x_1 \leqq x_2 \implies x_1 + x_3 \leqq x_2 + x_3$

ii. $\forall \alpha \geqq 0,\ x_1 \leqq x_2 \implies \alpha x_1 \leqq \alpha x_2$

iii. $\exists\ \sup(x_1, x_2)$

Definition C.4 (Positive Linear Functional & Sublinear). $\ell \in \Omega^*$ *is* positive linear *if*

$$\forall\ \omega \geqq 0 \implies \ell(\omega) \geqq 0$$

and is sublinear *if* $\forall \omega_1, \omega_2 \in \Omega$ *and* $\alpha \in \mathbb{R}_0^+$

i. $\ell(\omega_1 + \omega_2) \leqq \ell(\omega_1) + \ell(\omega_2)$

ii. $\ell(\alpha\omega_1) = \alpha\ell(\omega_1)$

Theorem C.5 (Hahn-Banach Positive Extension ([6], Theorem 8.31, pg. 330)). *Let* Θ *be a Riesz space,* $\Theta' \subset \Theta$ *be a Riesz subspace, and let* $\ell \in (\Theta')^*$ *be* ***positive linear****, then* ℓ *extends to a* ***positive linear functional*** *on all of* Θ *if and only if there is a monotone sublinear functional* $\rho \in \Theta^*$ *satisfying* $\ell(\theta') \leqq \rho(\theta')\ \forall\ \theta' \in \Theta'$.

Proof. "$\implies$": Let $\tilde{\ell} \in \Theta^*$ extend ℓ and set $\rho(\theta) = \hat{\ell}(\theta^+) = \hat{\ell}(\mathbb{1}(\theta \geqq 0) \cdot \theta)$
"$\impliedby$": Suppose $\rho : \Theta \to \mathbb{R}$ is monotone sublinear with $\ell(\theta') \leqq \rho(\theta')\ \forall\ \theta' \in \Theta'$. By the Hahn-Banach Extension Theorem $\exists\ \hat{\ell} \in \Theta^*$ extending ℓ and satisfying $\ell(\theta) \leqq \rho(\theta)\ \forall\ \theta \in \Theta$. Then for $\theta \geqq 0$

$$-\hat{\ell}(\theta) = \hat{\ell}(-\theta) \leqq \rho(-\theta) \leqq \rho(0) = 0$$

and multiplying both sides by -1 we obtain

$$\theta \geqq 0 \implies \hat{\ell}(\theta) \geqq 0$$

hence $\hat{\ell}$ is a positive extension of ℓ. □

Theorem C.6 (Riesz Representation ([5], Theorem 7.3, pg. 22))**.** *Let* (Θ, d) *be a metric space, then* $\forall$ ***positive*** $\ell \in \mathcal{C}_b^0(\Theta)^*$ $\exists$ *tight* $\mu \in \mathcal{P}(\mathcal{B}(\Theta, d))$ *s.t.*

$$\ell(f) = \int_{\Theta} f \ d\mu \ \ \forall \ f \in \mathcal{C}_b(\Theta)$$

References

[1] C. Villani, *Topics in Optimal Transport*. American Mathematical Society, 2003. ISBN: 0-8218-3312-X

[2] C. Villani, *Optimal Transport: Old and New*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. DOI: 10.1007/978-1-4471-5361-0 Available: https://ljk.imag.fr/membres/Emmanuel.Maitre/lib/exe/fetch.php?media=b07.stflour.pdf

[3] A. Klenke, *Probability Theory: A Comprehensive Course*. London: Springer, 2014. DOI: 10.1007/978-1-4471-5361-0

[4] P. Orbanz, *Probability Theory II*. London: Gatsby Computational Neuroscience Unit at UCL, 2016.

[5] O. van Gaans, *Probability measures on metric spaces*. Netherlands: Universiteit Leiden.

[6] C. Aliprantis, K. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2006. ISBN: 978-3540295860.

[7] O. Chodosh *Optimal Transport and Ricci Curvature: Wasserstein Space Over the Interval*. arXiv:1105.2883v2, 2012.

[8] J. Munkres *Topology*. Prentice Hall, 2000. ISBN: 0-13-181629-2