Duality Theorem for a Minimax Vector Optimization Problem

Jacob Atticus Armstrong Goodall (✉ jaagoodall@gmail.com)

University of Warwick Mathematics Institute <https://orcid.org/0000-0002-3825-1898>

MPA - Full Length Paper

Keywords: Duality Theorem, Optimal Transport, Minimax

Posted Date: February 3rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-176902/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Duality Theorem for Minimax Vector Optimization Problem

Jacob A. Armstrong Goodall

Received: date / Accepted: date

Abstract A duality theorem is stated and proved for a minimax vector optimization problem where the vectors are elements of the set of products of compact Polish spaces. A special case of this theorem is derived to show that two metrics on the space of probability distributions on countable products of Polish spaces are identical. The appendix includes a proof that, under the appropriate conditions, the function studied in the optimisation problem is indeed a metric. The optimisation problem is comparable to multi-commodity optimal transport where there is dependence between commodities. This paper builds on the work of R.S. MacKay who introduced the metrics in the context of complexity science in [4] and [5]. The metrics have the advantage of measuring distance uniformly over the whole network while other metrics on probability distributions fail to do so (e.g total variation, Kullback–Leibler divergence, see [5]). This opens up the potential of mathematical optimisation in the setting of complexity science.

Keywords Duality Theorem · Optimal Transport · Minimax

Mathematics Subject Classification (2010) 49Q22 · 90C46 · 90C47

Acknowledgments

The main acknowledgment goes to my advisor Robert MacKay for his encouragement throughout the process of writing my dissertation and this paper. It was, after all, his suggestion alone that prompted me to attempt submitting this paper. Secondly my thanks goes to everyone in the mathematics department at the university of Warwick for making the environment so conducive to learning and research.

1 Introduction

Let S be a countable set and $X_{s \in S}$ a collection of Polish spaces with bounded diameters. Then define the product $\mathcal{X} := \prod_{s \in S} X_s$ and let μ and ν be probability distributions on $\mathcal{X}$, further define $\Pi(\mu, \nu)$ to be the

Jacob Armstrong Goodall
University of Warwick
Coventry
CV4 7AL
United Kingdom
Tel.: +44 (0) 795 110 5835
E-mail: jacob.armstronggoodall@warwickgrad.net

set of couplings of μ and ν . Then for $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ we consider the optimisation problem

$$\text{Find } \pi^* \text{ that realises } \inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) \quad (1)$$

where $d_s(\mathbf{x}, \mathbf{y}) = d_s(x_s, y_s)$, a pre-metric on the components of the elements of $\mathcal{X}$. More explicitly, $d_s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a semi-metric on X_s which is extended to a pre-metric on $\mathcal{X}$. Here “pre-metric” refers to a function satisfying the following conditions:

1. $d(x, y) \geq 0$;
2. $d(x, x) = 0$;
3. $d(x, y) = d(y, x)$.

While a semi-metric satisfies all the requirements of a metric apart from the triangle inequality.

For any functions $\phi : \mathcal{X} \rightarrow \mathbb{R}$ and $\psi : \mathcal{X} \rightarrow \mathbb{R}$ as well as the vector $\mathbf{e} \in \mathcal{E} = \{(e_s)_{s \in S} : \sum_s e_s < 1, e_s \geq 0\}$ meeting the condition below, which we refer to as the duality conditions,

$$\phi(\mathbf{y}) - \psi(\mathbf{x}) \leq \sum_s e_s d_s(\mathbf{x}, \mathbf{y}) \quad (2)$$

there exists the dual problem corresponding to (1), which is presented in (3). To make things tidier we will write $c_s = e_s d_s$.

$$\sup_{\phi, \psi} \left\{ \int_{\mathcal{X}} \phi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \quad \phi(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s; x_t = y_t \quad \forall t \neq s, x_s \neq y_s \right\} \quad (3)$$

The functions ϕ and ψ , and the vector $\mathbf{e}$ can be derived via linear programming as the Lagrange multipliers of (1). If we impose that ϕ and ψ be integrable: $\psi \in L^1(\mathcal{X}, \mu)$; $\phi \in L^1(\mathcal{X}, \nu)$. By linearity of the integral and (2), for a given $\mathbf{e}$ we have

$$\sup_{\substack{\phi - \psi \leq c_s \\ \mathbf{x} = \mathbf{y} \text{ off } s}} \left\{ \int_{\mathcal{X}} \phi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}) \right\} \leq \inf_{\pi \in \Pi} \left\{ \sup_{s \in S} \int d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) \right\} \quad (4)$$

In this paper it is proved that

$$\sup_{\substack{\phi - \psi \leq c_s \\ \mathbf{x} = \mathbf{y} \text{ off } s}} \left\{ \frac{\int_{\mathcal{X}} \phi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x})}{\sum_s e_s} \right\} = \inf_{\pi \in \Pi} \left\{ \sup_{s \in S} \int d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) \right\}$$

for a specific optimal $\mathbf{e}$ which is referred to as $\mathbf{e}^*$ (see theorem 2).

To illustrate the duality I will consider an analogy to information theory. Let $(\mathcal{X}, \mu)$ represent the set of input signals for a transmission network and $(\mathcal{X}, \nu)$ the set of possible output signals. Let the sites $s \in S$ represent the communication channels available in the network and interpret d_s as a measurement of error, with e_s the intrinsic error of the channel at s . Then the optimisation problem (1) requires us to find a transmission method that minimises the largest expected transmission error over all channels.

The pair of functions ϕ, ψ corresponds to a coding method, ψ representing the accuracy when encoding the input signals and ϕ the error decoding the output signals. This combined with intrinsic error allows us to interpret the duality condition (2), as the statement that the accuracy of the coding strategy is no greater than the sum of errors, each term scaled according to the intrinsic error of the channel. Of course, being more accurate than our measurements and communication system allow is unfeasible. The dual problem (3) requires us to find the most accurate feasible encoding strategy. This justifies my reference to this as the problem of optimal transmission.

2 Definitions

This section will contain all definitions specific to this problem.

Definition 1 (Steif's Metric)

$$\bar{d}(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int d_s(x_s, y_s) d\pi(\mathbf{x}, \mathbf{y})$$

This is an extension of a metric that was introduced by J. Steif in (Steif,1988) where he restricts in his definition to the discrete metric on the state space $\{0, 1\}$. In the appendix I prove that this is a complete metric on $\mathcal{P}(\mathcal{X})$.

Definition 2 (Dobrushin Metric)

Let F be the space of continuous functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that

$$\|f\|_F = \sum_{s \in S} \Delta_s(f) < \infty,$$

where

$$\Delta_s(f) = \sup \left\{ \frac{f(\mathbf{x}) - f(\mathbf{y})}{d_s(\mathbf{x}, \mathbf{y})} : x_t = y_t \ \forall t \neq s, x_s \neq y_s \right\}.$$

Then we define Dobrushin's metric as,

$$D(\mu, \nu) = \sup_{f \in F \setminus C} \frac{(\mu - \nu)(f)}{\sum_{s \in S} \Delta_s(f)}$$

Here C denotes the constant functions and $\mu(f) = \int f d\mu$.

Definition 3 Let $\mathcal{X} = \prod X_s$, where (X_s, d_s) are polish space. A subset $\Gamma \subset \mathcal{X} \times \mathcal{X}$ is said to be c_s -cyclically monotone if, for any collection $M \in \mathbb{N}$, and any family $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_M, \mathbf{y}_M)$ of points in Γ , holds the inequality

$$\sup_{s \in S} \sum_{i=1}^M c_s(\mathbf{x}_i, \mathbf{y}_i) \leq \sup_{s \in S} \sum_{i=1}^M c_s(\mathbf{x}_i, \mathbf{y}_{i+1}) \quad (5)$$

(with the conventions that $\mathbf{y}_{M+1} = \mathbf{y}_1$). A transmission method is said to be c_s -cyclically monotone if it is concentrated on a c_s -cyclically monotone set.

Definition 4 (Tight Coding Strategy) Let $|S| = N$. The coding strategy (ψ, ϕ) is tight if, for each s and $\forall(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X} : \mathbf{x} = \mathbf{y}$ off s ,

$$\phi(\mathbf{y}) = \sup_{e \in \mathcal{E}} \inf_x (\psi(\mathbf{x}) + e_{s^*} d_{s^*}(\mathbf{x}, \mathbf{y})), \quad \psi(\mathbf{x}) = \inf_{e \in \mathcal{E}} \sup_y (\phi(\mathbf{y}) - e_{s^*} d_{s^*}(\mathbf{x}, \mathbf{y})) \quad (6)$$

and we denote by e^* any $e \in \mathcal{E}$ for which the above equality holds. For convenience I will use the shorthand $e_{s^*}^* d_{s^*} = c^*$.

Definition 5 (c_s -convexity) Let $\mathcal{X} = \prod_s X_s$ be a set and $d_s : \mathcal{X} \times \mathcal{X} \rightarrow [0, +\infty]$ such that $\mathbf{x} = \mathbf{y}$ off s . A function $\psi : \mathbb{R} \cup \{+\infty\}$ is said to be c_s -convex if it is not identically $+\infty$, and there exists $\zeta : \mathcal{X} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ such that

$$\forall x \in \mathcal{X} \quad \psi(\mathbf{x}) = \inf_{e \in \mathcal{E}} \sup_{y \in \mathcal{X}} (\zeta(\mathbf{y}) - c_{s^*}(\mathbf{x}, \mathbf{y})) \quad (7)$$

Then its c_s -transform is the function $\psi^{c_s}(\mathbf{y})$ defined by

$$\forall \mathbf{y} \in \mathcal{X} \quad \psi^{c_s}(\mathbf{y}) = \sup_{e \in \mathcal{E}} \inf_{\mathbf{x} \in \mathcal{X}} (\psi(\mathbf{x}) + c_{s^*}(\mathbf{x}, \mathbf{y})) \quad (8)$$

and its c_s -sub-differential is the c_s -cyclically monotone set defined by

$$\partial_{c_s}\psi := \{(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X}; \psi^{c_s}(\mathbf{y}) - \psi(\mathbf{x}) = c^*(\mathbf{x}, \mathbf{y})\}$$

The functions ψ and ψ^{c_s} are said to be c_s -conjugate. Moreover, the c_s -sub-differential of ψ at point $\mathbf{x}$ is

$$\partial_{c_s}\psi(\mathbf{x}) = \{\mathbf{y} \in \mathcal{Y}; (\mathbf{x}, \mathbf{y}) \in \partial_{c_s}\psi\}$$

or equivalently

$$\forall \mathbf{z} \in \mathcal{X}, \quad \psi(\mathbf{x}) + c^*(\mathbf{x}, \mathbf{y}) \leq \psi(\mathbf{z}) + c^*(\mathbf{z}, \mathbf{y}) \quad (9)$$

since,

$$\begin{aligned} \psi^c(\mathbf{y}) - \psi(\mathbf{x}) &= c^*(\mathbf{x}, \mathbf{y}) \\ &\Rightarrow \sup_{e \in \mathcal{E}} \inf_{\mathbf{x} \in \mathcal{X}} (\psi(\mathbf{x}) + c_s^*(\mathbf{x}, \mathbf{y})) = \psi(\mathbf{x}) + c^*(\mathbf{x}, \mathbf{y}) \\ &\Rightarrow \psi(\mathbf{x}) + c^*(\mathbf{x}, \mathbf{y}) \leq \psi(\mathbf{z}) + c^*(\mathbf{z}, \mathbf{y}) \end{aligned}$$

3 Technical Lemmas

This section contains technical lemmas that are used in the proof of the main result.

Lemma 1 (Technical Lemma A) *If c_s satisfies the triangle inequality then ψ is c_s -convex if and only if $\psi(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s(\mathbf{x}, \mathbf{y})$; and in addition $\psi^{c_s} = \psi$.*

Lemma 2 (Technical Lemma B) *Let (X_s, d_s) be a collection of Polish spaces with bounded diameter. Then $\mathcal{X} = \prod_s X_s$ is also a Polish space by theorem 4. Let $d_s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a metric and $h_s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a continuous function such that for all $s \in S$, $d_s \geq h_s$. Let $(\pi_k)_{k \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{X} \times \mathcal{X}$, converging weakly to some $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{X})$, in such a way that $h_s \in L^1(\pi_k)$, $h_s \in L^1(\pi)$, and*

$$\int_{\mathcal{X} \times \mathcal{X}} h_s d\pi_k \xrightarrow{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{X}} h_s d\pi.$$

Then for every $s \in S$

$$\int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi \leq \liminf_{k \rightarrow \infty} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi_k.$$

In particular, $F : \pi \rightarrow \int d_s(\mathbf{x}, \mathbf{y}) d\pi$ is lower semi-continuous on $\mathcal{P}(\mathcal{X} \times \mathcal{X})$, equipped with the topology of weak convergence.

Lemma 3 (Tightness of transmission plans) *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Polish spaces. Let $P \subset \mathcal{P}(\mathcal{X})$ and $Q \subset \mathcal{P}(\mathcal{Y})$ be tight subsets of $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ respectively. Then the sets $\Pi(P, Q)$ of all transference plans whose marginals lie in P and Q respectively, is itself tight in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$.*

Lemma 4 (Technical Lemma C) *For any functions $\phi \in L^1(\mu)$ and $\psi \in L^1(\nu)$ and $e = (e_s)_{s \in S}$ a vector contained in $\mathcal{E}$. Then, we have*

$$\phi(\mathbf{x}) - \psi(\mathbf{y}) \leq \sum_{s \in S} c_s(\mathbf{x}, \mathbf{y}).$$

If and only if, for each $s \in S$ and all $(\mathbf{x}, \mathbf{y}) \in \mathcal{X} \times \mathcal{X}$ such that $\mathbf{x} = \mathbf{y}$ off s , we have

$$\phi(\mathbf{x}) - \psi(\mathbf{y}) \leq c_s(\mathbf{x}, \mathbf{y}).$$

Lemma 5 (Alternative Characteristic of c_s -convexity) *For any function $\psi : \mathcal{X} \rightarrow \mathbb{R} \cup \{+\infty\}$, let its c_s -convexification be defined by $\psi^{dd} = (\psi^d)^d$. More explicitly,*

$$\psi^{dd}(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{X}} \inf_{\mathbf{z} \in \mathcal{X}} (\psi(\mathbf{z}) + c^*(\mathbf{z}, \mathbf{y}) - c^*(\mathbf{x}, \mathbf{y})).$$

Then ψ is c_s -convex if and only if $\psi^{dd} = \psi$.

4 Proofs of Lemmas

Proof (Technical Lemma A) By rearranging and taking the infimum we can show that

$$\psi^{c_s}(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s(\mathbf{x}, \mathbf{y})$$

if and only if

$$\psi^{c_s}(\mathbf{y}) = \inf_x \{c_s(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{x})\} \text{ and } \psi(\mathbf{x}) = \sup_y \{\psi^{c_s}(\mathbf{y}) - c_s(\mathbf{x}, \mathbf{y})\}.$$

Now, if ψ is c_s -convex then there exists a function $\zeta : \mathcal{X} \rightarrow \mathbb{R} \cup \{-\infty\}$ such that

$$\psi(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{X}} \{\zeta(\mathbf{y}) - c_s(\mathbf{x}, \mathbf{y})\}.$$

So, if c_s obeys the triangle inequality then,

$$\begin{aligned} \psi^{c_s}(\mathbf{y}) - \psi(\mathbf{x}) &\leq c_s(\mathbf{x}, \mathbf{y}) \\ \inf_x \{c_s(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{x})\} - \psi(\mathbf{x}) &\leq c_s(\mathbf{x}, \mathbf{y}) \\ \inf_x \{c_s(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{x})\} &\leq c_s(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{x}) \\ \psi(\mathbf{z}) + c_s(\mathbf{x}, \mathbf{z}) &\leq c_s(\mathbf{x}, \mathbf{y}) + \psi(\mathbf{x}) \\ \psi(\mathbf{z}) + c_s(\mathbf{x}, \mathbf{z}) &\leq c_s(\mathbf{x}, \mathbf{z}) + c_s(\mathbf{z}, \mathbf{y}) + \psi(\mathbf{x}) \\ \psi(\mathbf{z}) - \psi(\mathbf{x}) &\leq c_s(\mathbf{z}, \mathbf{y}). \end{aligned}$$

Conversely, if $\psi(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s(\mathbf{x}, \mathbf{y})$ then

$$\begin{aligned} \psi(\mathbf{y}) - \psi(\mathbf{x}) &\leq c_s(\mathbf{x}, \mathbf{y}) \\ \psi(\mathbf{y}) - c_s(\mathbf{x}, \mathbf{y}) &\leq \psi(\mathbf{x}) \\ \sup_{\mathbf{y} \in \mathcal{X}} (\psi(\mathbf{y}) - c_s(\mathbf{x}, \mathbf{y})) &= \psi(\mathbf{x}). \end{aligned}$$

Hence $\psi = \psi^{c_s}$.

Proof (Technical Lemma B) Since d_s is non-negative it can be written as the point wise limit of a non-decreasing family $(d_s^k)_{k \in \mathbb{N}}$ of continuous real-valued functions. By monotone convergence,

$$\int d_s d\pi = \lim_{l \rightarrow \infty} \int (d_s)_l d\pi = \lim_{l \rightarrow \infty} \lim_{k \rightarrow \infty} \int (d_s)_l d\pi_k \leq \liminf_{k \rightarrow \infty} \int d_s d\pi_k.$$

This is the desired result.

Proof (Lemma 3) Let $\mu \in P$, $\nu \in Q$, and $\pi \in \Pi(\mu, \nu)$. By assumption, for any $\epsilon > 0$ there is a compact set $K_\epsilon \subset \mathcal{X}$, independent of the choice of μ in P , such that $\mu[\mathcal{X} \setminus K_\epsilon] \leq \epsilon$; and similarly there is a compact set $L_\epsilon \subset \mathcal{Y}$, independent of the choice of ν in Q , such that $\nu[\mathcal{Y} \setminus L_\epsilon] \leq \epsilon$. Then for any coupling (X, Y) of (μ, ν) ,

$$\mathbb{P}[(X, Y) \notin K_\epsilon \times L_\epsilon] \leq \mathbb{P}[X \notin K_\epsilon] + \mathbb{P}[Y \notin L_\epsilon] \leq 2\epsilon$$

The desired result follows since this bound is independent of the coupling, and $K_\epsilon \times L_\epsilon$ is compact in $\mathcal{X} \times \mathcal{Y}$.

Proof (Technical Lemma C) For sufficiency it is enough to simply specialize $\sum_s c_s$ to $\mathbf{x} = \mathbf{y}$ off s . For necessity, fix $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and put $J = \{s \in S : x_s \neq y_s\}$. Without loss of generality we may assume that $J = (j_n)_{n \in \mathbb{N}}$. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence of configurations such that $\mathbf{x}_0 = \mathbf{x}$, such that $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{y}$ and such that for all $n, m \in \mathbb{N}$:

$$x_{j_n}^n \neq x_{j_n}^{n+1}, x_b^n = x_b^{n+1} \ (b = j_n), x_{j_n}^{n+m+1} = y_{j_n}.$$

Which implies that

$$f(\mathbf{x}) - f(\mathbf{y}) \leq \sum_n f(\mathbf{z}_n) - f(\mathbf{z}_{n-1}) \leq \sum_n c_{j_n}(\mathbf{x}, \mathbf{y})$$

which is the desired conclusion.

$$x_{j_n}^n \neq x_{j_n}^{n+1}, x_b^n = x_b^{n+1} \ (b = j_n), x_{j_n}^{n+m+1} = y_{j_n}.$$

Proof (Alternative characteristic of c_s -convexity) For any function $\phi : \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ (not necessarily c_s -convex), it holds that $\phi^{c_s c_s c_s} = \phi^{c_s}$. More accurately we have

$$\phi^{c_s c_s c_s}(\mathbf{x}) = \sup_y \inf_z \sup_w [\phi(\mathbf{w}) - c^*(z, \mathbf{w}) + c^*(z, \mathbf{y}) - c^*(\mathbf{x}, \mathbf{y})].$$

Now, choose $z = \mathbf{x}$ which gives us $\phi^{c_s c_s c_s}(\mathbf{x}) \leq \phi^{c_s}(\mathbf{x})$; while the choice $\mathbf{w} = \mathbf{y}$ shows that $\phi^{c_s c_s c_s}(\mathbf{x}) \geq \phi^d(\mathbf{x})$.

If ψ is c_s -convex, then there is ζ such that $\psi = \zeta^{c_s}$, so $\psi^{c_s c_s} = \zeta^{c_s c_s c_s} = \zeta^{c_s} = \psi$. Conversely, if $\psi^{c_s c_s} = \psi$, then ψ is c_s -convex, as the c_s -transform of ψ^{c_s} .

5 The Main results

Theorem 1 (Existence of an optimal coupling) Let $(\mathcal{X}, \mu)$ and $(\mathcal{X}, \nu)$ be two Polish probability spaces. Let $d_s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a pre-metric on $\mathcal{X}$. Then there is a coupling of (μ, ν) which minimizes maximal distance $\sup_s \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y})$ over all sites, among all possible couplings (X, Y) .

Theorem 2 (MacKay's Duality) Let S be any countable set and consider a collection $\{X_s\}_{s \in S}$ such that each X_s is a complete separable metric space of bounded diameter. Let $\mathcal{X} = \prod_s X_s$ and $d_s : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ be a collection of pre-metrics on $\mathcal{X}$ defined by $d_s(\mathbf{x}, \mathbf{y}) = d_s(x_s, y_s)$. Denote the set of all probability measures on $\mathcal{X} \times \mathcal{X}$ with marginals μ and ν by $\Pi(\mu, \nu)$. Then there exists ϕ and ψ such that $\phi(\mathbf{x}) - \psi(\mathbf{y}) \leq c_s(\mathbf{x}, \mathbf{y})$ off s , for which there is Duality;

$$\inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) = \sup_{\psi \in L^1(\mu)} \sup_{\phi \in \mathcal{E}} \left(\int_{\mathcal{X}} \psi^{c_s}(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \psi^{c_s} - \psi \leq \sum_{s \in S} c_s(\mathbf{x}, \mathbf{y}) \right)$$

and in the above suprema one might as well impose that ψ be c_s -convex and ϕ be c_s -concave.

Further, if for all $s \in S$ the distance metric is real valued, the optimal cost is finite, and one has the pointwise upper bound

$$d_s(\mathbf{x}, \mathbf{y}) \leq f_s(\mathbf{x}) + g_s(\mathbf{y}), \quad (f_s(\mathbf{x}), g_s(\mathbf{y})) \in L^1(\mu) \times L^1(\nu),$$

where f_s and g_s are functions from X_s to $\mathbb{R}$ then both the primal and dual problems have solutions, so

$$\min_{\pi \in \Pi(\mu, \nu)} \max_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s d\pi = \max_{\psi \in L^1(\mu)} \left(\int_{\mathcal{X}} \psi^{c_s}(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}) ; \psi^{c_s} - \psi \leq \sum_{s \in S} c_s(\mathbf{x}, \mathbf{y}) \right)$$

Corollary 1 Steif's metric is identical to Dobrushin's metric

6 Proof of Main results

Proof The proof uses tightness of transmission plans and technical lemma A. Since $\mathcal{X}$ is Polish, $\{\mu\}$ and $\{\nu\}$ are tight in $\mathcal{P}(\mathcal{X})$; By Lemma 15, $\Pi(\mu, \nu)$ is tight in $P(\mathcal{X} \times \mathcal{X})$, and by Prokhorov's theorem this set has a compact closure. By passing to the limit in the equation for the marginals, we see that $\Pi(\mu, \nu)$ is closed, so it is compact. Then let $(\pi_k)_{k \in \mathbb{N}}$ be a sequence of probability measures on $\mathcal{X} \times \mathcal{X}$, such that $\sup_s \int d_s d\pi_k$ converges to the infimum transmission error. Extracting a subsequence if necessary, we may assume π_k converges to some $\pi \in \Pi(\mu, \nu)$. The function $h_s : (x_s, y_s) \mapsto (0, 0)$ lies in $L^1(\pi_k)$ and in $L^1(\pi)$, and $d_s \geq h_s$ by assumption; moreover, $\int h_s d\pi_k = \int h_s d\pi = 0$; so Lemma 4.3 implies

$$\int d_s d\pi \leq \liminf_{k \rightarrow \infty} \int d_s d\pi_k$$

and

$$\sup_s \int d_s d\pi \leq \sup_s \liminf_{k \rightarrow \infty} \int d_s d\pi_k.$$

Thus π is minimizing.

Proof Idea for Proof:

The main steps are as follows

1. Prove that the optimal plan is cyclically monotone if the distributions μ and ν are delta functions at each point in statespace. This would be sufficient to prove the finite case.
2. Use the central limit theorem to extend this existence theorem to countable S and X_s .
3. Show that if π is optimal it must lie in the sub-differential of ψ .
4. Show that ψ and ϕ must be bounded and by step 3 are optimal.
5. Show that ψ and ϕ are integrable and hence a feasible point exists for both problems.

We have proved that an optimal plan π indeed exists. For each $s \in S$ the state-space X_s is bounded so there exists $s^* \in S$ such that $\int d_{s^*} d\pi = \sup_s \int d_s d\pi$. Now, let (ψ, ϕ) a coding strategy satisfying. Of course, if $x_t = y_t \forall t : t \neq s$ then

$$\int c^*(x, y) d\pi(x, y) \geq \int \phi(x) d\nu - \int \psi(y) d\mu = \int [\phi(y) - \psi(x)] d\pi(x, y)$$

So if the inequality is an equality we have $\int [c^* - \phi + \psi] d\pi = 0$, and hence

$$\phi(y) - \psi(x) = c^* \quad \pi(dx dy) - a.s.$$

Intuitively speaking, whenever y is an allowable input and x an allowable output, the a coding strategy is chosen so that the accuracy is equal to the maximal error at any site scaled according to the worst case intrinsic error. Now let $(x_i)_{0 \leq i \leq m}$ and $(y_i)_{0 \leq i \leq m}$ be such that $d_{s^*}(x_i, y_i)$. Then we observe that

$$\begin{cases} \phi(y_0) - \psi(x_0) &= c^*(x_0, y_0) \\ \phi(y_1) - \psi(x_1) &= c^*(x_1, y_1) \\ \vdots &\vdots \\ \phi(y_m) - \psi(x_m) &= c^*(x_m, y_m) \end{cases}$$

On the other hand, if x is an arbitrary point,

$$\begin{cases} \phi(y_0) - \psi(x_1) &\leq c^*(x_1, y_0) \\ \phi(y_1) - \psi(x_2) &\leq c^*(x_0, y_1) \\ \vdots &\vdots \\ \phi(y_m) - \psi(x) &\leq c^*(x, y_m) \end{cases}$$

By subtracting these inequalities from the previous equalities and adding everything up, one obtains

$$\psi(x) \geq \psi(x_0) + [c^*(x_0, y_0) - c^*(x_1, y_0)] + \cdots + [c^*(x_m, y_m) - c^*(x, y_m)]$$

Of course, one can add an arbitrary constant to ψ , provided that one subtracts the same constant from ϕ ; so it is possible to decide that $\psi(x_0) = 0$, where (x_0, y_0) is arbitrarily chosen in the support of π . Then

$$\psi(x) \geq [c^*(x_0, y_0) - c^*(x_1, y_0)] + \cdots + [c^*(x_m, y_m) - c^*(x, y_m)]$$

and this should be true for all choices of (x_i, y_i) ($1 \leq i \leq m$) in the support of π , and for $m \geq 1$. So it becomes natural to define ψ as the supremum of all the function (in the variable x) appearing on the right hand side above. It will turn out that this ψ satisfies the equation

$$\psi^c(y) - \psi(x) = c^*(x, y) \quad \pi(dx dy) - a.s.$$

for all $(x, y) \in \mathcal{X} \times \mathcal{X} : x_t = y_t \iff t \neq s$. By lemma 8 we can generalize to

$$\psi^c(y) - \psi(x) = \sum c^*(x, y) \quad \pi(dx dy) - a.s.$$

for all $(x, y) \in \mathcal{X} \times \mathcal{X}$. Then, if ψ and ψ^c are integrable, one can write

$$\sup_s \int d_s d\pi = \frac{\int \psi^c d\pi - \int \psi d\pi}{\sum_s e_s^*}$$

This shows at the same time that π is optimal in the Primal problem, and the function ψ is optimal in the dual problem.

6.1 Rigorous Proof.

Throughout the proof let $|S| = N$ and $|X_s| = n_s$.

Proof Step 1: If $\mu = \frac{1}{\sum_s n_s} \sum_{s=1}^N \sum_{i=1}^{n_s} \delta_{(x_s)_i}$, $\nu = \frac{1}{\sum_s n_s} \sum_{s=1}^N \sum_{j=1}^{n_s} \delta_{(y_s)_j}$, where the distances $c_s(x_i, y_j)$ are finite, then there is at least one cyclically monotone transmission method.

In this particular case, a transmission method between μ and ν can be identified with a bi-stochastic $N \times N$ array of real valued matrices a_s with components $(a_s)_{ij} \in [0, 1]$: each $(a_s)_{ij}$ tells us the rate at which the point $(x_s)_i$, occurring with probability $\frac{1}{\sum_s n_s}$ will be interpreted as $(y_s)_j$ so the primal problem becomes

$$\inf_{(a_s)_{ij}} \sup_{s \in S} \sum_{ij} (a_s)_{ij} c_s(x, y)$$

Where the infimum is over all arrays $(a_s)_{ij}$ satisfying

$$\sum_s \sum_i (a_s)_{ij} = 1, \quad \sum_s \sum_j (a_s)_{ij} = 1. \quad (10)$$

Here we are minimizing a linear function on the compact set $[0, 1]^{(n_s \times n_s) \times (n_s \times n_s)}$, so there must exist a minimizer; the corresponding transmission method π can be written as

$$\pi = \frac{1}{\sum_s n_s} \sum_{s=1}^N \sum_{i,j=1}^{n_s} (a_s)_{ij} \delta_{((x_s)_i, (y_s)_j)}$$

and its support Γ is the set of all couples (x_i, y_j) such that $(a_s)_{ij} > 0$. Assume that $\Gamma = \text{Supp} \pi$ is not cyclically monotone: Then there exist $M \in \mathbb{N}$ and $(x_{i_1}, y_{i_M}), \dots, (x_{i_1}, y_{i_M})$ in Γ such that

$$\sup_s \{c_s(x_{i_1}, y_{i_2}) + \dots + c_s(x_{i_M}, y_{i_1})\} < \sup_s \{c_s(x_{i_1}, y_{i_1}) + \dots + c_s(x_{i_M}, y_{i_M})\}.$$

Let $a := \min((a_s)_{i_1 j_1}, \dots, (a_s)_{i_M j_M}) > 0$. Define a new transmission method $\tilde{\pi}$ by the formula

$$\tilde{\pi} = \pi + \frac{a}{\sum_s n_s} \sum_{s=1}^N \sum_{l=1}^{n_s} \left(\delta_{((x_s)_{i_l}, (y_s)_{j_{l+1}})} - \delta_{((x_s)_{i_l}, (y_s)_{j_l})} \right)$$

One can check that this has the same marginals, and the error rate associated with $\tilde{\pi}$ is strictly less than the error rate associates with π . This is a contradiction, so Γ is indeed (S, c_s) -cyclically monotone!

Step 2: If S and X_s are countably infinite, then there is a cyclically monotone transmission method.

To prove this, consider sequences of independent random variables $\mathbf{x}_i \in \mathcal{X}$, $\mathbf{y}_i \in \mathcal{X}$, with respective law μ, ν . According to Varadarajan's theorem, one has, with probability one,

$$\mu_{N,n} := \frac{1}{\sum_s n_s} \sum_{s=1}^N \sum_{i=1}^{n_s} \delta_{(x_s)_i} \rightarrow \mu, \quad \nu_{N,n} := \frac{1}{\sum_s n_s} \sum_{s=1}^N \sum_{j=1}^{n_s} \delta_{(y_s)_j} \rightarrow \nu \quad (11)$$

as $N, n_s \rightarrow \infty$, for all $s \in S$ in the sense of weak convergence of measures. In particular, by Prokhorovs theorem, μ_{N,n_s} and ν_{N,n_s} are tight sequences.

For each pair N, n_s , let π_{N,n_s} be a cyclically monotone transmission method between μ_{N,n_s} and ν_{N,n_s} . By Lemma 3 on page 4, $\{\pi_{N,n_s}\}_{N,n_s \in \mathbb{N}}$ is tight. By Prokhorov's theorem, there is a subsequence, still denoted (π_{N,n_s}) , which converges weakly to some probability measure π , i.e.

$$\int h(\mathbf{x}, \mathbf{y}) d\pi_{N,n_s}(\mathbf{x}, \mathbf{y}) \rightarrow \int h(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y})$$

for all bounded continuous functions h on $\mathcal{X} \times \mathcal{X}$. By applying the previous identity with $h(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})$ and $h(\mathbf{x}, \mathbf{y}) = g(\mathbf{y})$, we see that π has marginals μ and ν .

For each N and each n_s , the cyclical monotonicity of π_{N,n_s} implies that for all M and $\pi_{N,n_s}^{\otimes M}$ -almost all $(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_M, \mathbf{y}_M)$, the inequality (5) is satisfied; in other words, $\pi_{N,n_s}^{\otimes M}$ is concentrated on the set $\mathcal{C}(M)$ of all $((\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_M, \mathbf{y}_M))_{s \in S} \in (\mathcal{X} \times \mathcal{X})^M$ satisfying ((5)). Since d_s is continuous $\forall s \in S$, $\mathcal{C}(M)$ is a closed set, so the weak limit $\pi^{\otimes M}$ of $\pi_{N,n_s}^{\otimes M}$ is also concentrated on $\mathcal{C}(M)$. Let $\Gamma = \text{Supp}\pi$, then

$$\Gamma^M = (\text{Supp}\pi)^M = \text{Supp}(\pi^{\otimes M}) \subset \mathcal{C}(M),$$

and since this holds true for all M , Γ is cyclically monotone.

Step 3: If each d_s is real-valued and π is cyclically monotone, then there is an $e_s d_s$ -convex ψ such that $\partial_{e_s d_s} \psi$ contains the support of π .

Since π is optimal and by steps one and two is cyclically monotone we have the following implication. If $s^* \in S$ is such that $\int e_{s^*} d_{s^*} d\pi = \sup_s \int e_s d_s d\pi$ then from cyclical monotonicity

$$\sum_{i=1} c_{s^*}(\mathbf{x}_i, \mathbf{y}_i) \leq \sum c_{s^*}(\mathbf{x}_i, \mathbf{y}_{i+1})$$

which we can rewrite as

$$\sum_{i=1} [c_{s^*}(\mathbf{x}_i, \mathbf{y}_i) - c_{s^*}(\mathbf{x}_i, \mathbf{y}_{i+1})] \leq 0.$$

Indeed, let Γ again denote the support of π (a closed set). Pick any $(\mathbf{x}_0, \mathbf{y}_0) \in \Gamma$, and define

$$\begin{aligned} \psi(x) := & \inf_{e \in \mathcal{E}} \sup_{m \in \mathbb{N}} \sup \{ [c_{s^*}(\mathbf{x}_0, \mathbf{y}_0) - c_{s^*}(\mathbf{x}_1, \mathbf{y}_0)] + [c_{s^*}(\mathbf{x}_1, \mathbf{y}_1) - c_{s^*}(\mathbf{x}_2, \mathbf{y}_1)] + \\ & \dots + [c_{s^*}(\mathbf{x}_m, \mathbf{y}_m) - c_{s^*}(\mathbf{x}, \mathbf{y}_m)] \}; \quad (\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_m, \mathbf{y}_m) \in \Gamma \} \quad (12) \end{aligned}$$

By applying the definition with $m = 1$ and $(\mathbf{x}_1, \mathbf{y}_1) = (\mathbf{x}_0, \mathbf{y}_0)$, one immediately sees that $\psi(\mathbf{x}_0) \geq 0$. On the other hand, $\psi(\mathbf{x}_0)$ is the supremum of all the quantities $\inf_{e \in \mathcal{E}} \{[c_{s^*}(\mathbf{x}_0, \mathbf{y}_0) - c_{s^*}(\mathbf{x}_1, \mathbf{y}_0)] + \dots + [c_{s^*}(\mathbf{x}_m, \mathbf{y}_m) - c_{s^*}(\mathbf{x}_0, \mathbf{y}_m)]\}$ which by cyclical monotonicity are all non-positive. So actually $\psi(\mathbf{x}_0) = 0$.

By renaming $\mathbf{y}_m$ as $\mathbf{y}$, obviously

$$\begin{aligned} \psi(\mathbf{x}) = \sup_{\mathbf{y} \in \mathcal{X}} \inf_{e \in \mathcal{E}} \sup_{m \in \mathbb{N}} \sup_{(x^1, y^1), \dots, (x^{m-1}, y^{m-1}), x^m} \{ & [c_{s^*}(\mathbf{x}_0, \mathbf{y}_0) - c_{s^*}(\mathbf{x}_1, \mathbf{y}_0)] \\ & + [c_{s^*}(\mathbf{x}_1, \mathbf{y}_1) - c_{s^*}(\mathbf{x}_2, \mathbf{y}_1)] + \\ & e_s \dots + [c_{s^*}(\mathbf{x}_m, \mathbf{y}) - c_{s^*}(\mathbf{x}, \mathbf{y})]; (\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_m, \mathbf{y}) \in \Gamma \} \end{aligned} \quad (13)$$

So $\psi(\mathbf{x}) = \inf_{e \in \mathcal{E}} \sup_{\mathbf{y}} [\zeta(\mathbf{y}) - c_{s^*}(\mathbf{x}, \mathbf{y})]$, if ζ is defined by

$$\begin{aligned} \zeta(\mathbf{y}) = \sup \{ & [c^*(\mathbf{x}_0, \mathbf{y}_0) - c^*(\mathbf{x}_1, \mathbf{y}_0)] + [c^*(\mathbf{x}_1, \mathbf{y}_1) - c^*(\mathbf{x}_2, \mathbf{y}_1)] + \\ & \dots + c^*(\mathbf{x}_m, \mathbf{y}) \}; m \in \mathbb{N}, (\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_m, \mathbf{y}) \in \Gamma \} \end{aligned} \quad (14)$$

(with the convention that $\zeta = -\infty$ out of $\text{proj}_{\mathcal{X}}(\Gamma)$). Thus ψ is a c_s -convex function.

Now let $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \Gamma$. By choosing $\mathbf{x}_m = \bar{\mathbf{x}}$, $\mathbf{y}_m = \bar{\mathbf{y}}$ in the definition of ψ ,

$$\begin{aligned} \psi(\mathbf{x}) \geq \inf_{e \in \mathcal{E}} \sup_m \sup_{(\mathbf{x}_i, \mathbf{y}_i)_{0 \leq i < m}} \{ & [c_{s^*}(\mathbf{x}_0, \mathbf{y}_0) - c_{s^*}(\mathbf{x}_1, \mathbf{y}_0)] + \\ & \dots + [c_{s^*}(\mathbf{x}_{m-1}, \mathbf{y}_{m-1}) - c_{s^*}(\bar{\mathbf{x}}, \mathbf{y}_{m-1})] + [c_{s^*}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - c_{s^*}(\mathbf{x}, \bar{\mathbf{y}})] \}. \end{aligned}$$

In the definition of ψ , it does not matter whether one takes the supremum over $m-1$ or over m variables, since one also takes the supremum over m . So the Previous inequality can be recast as

$$\psi(\mathbf{x}) \geq \psi(\bar{\mathbf{x}}) + c^*(\bar{\mathbf{x}}, \bar{\mathbf{y}}) - c^*(\mathbf{x}, \bar{\mathbf{y}}).$$

In particular, $\psi(\mathbf{x}) + c^*(\mathbf{x}, \bar{\mathbf{y}}) \geq \psi(\bar{\mathbf{x}}) + c^*(\bar{\mathbf{x}}, \bar{\mathbf{y}})$. By technical lemma C this can be extended to the sum over $s \in S$;

$$\psi(\mathbf{x}) + \sum_{s \in S} c_s^*(\mathbf{x}, \bar{\mathbf{y}}) \geq \psi(\bar{\mathbf{x}}) + \sum_{s \in S} c_s^*(\bar{\mathbf{x}}, \bar{\mathbf{y}})$$

Where $c_s^* = e_s^* d_{s^*}$. Taking the infimum over $\mathbf{x} \in \mathcal{X}$ in the left-hand side, we deduce that

$$\psi^{c_s}(\bar{\mathbf{y}}) \geq \psi(\bar{\mathbf{x}}) + \sum_{s \in S} c_s^*(\bar{\mathbf{x}}, \bar{\mathbf{y}}).$$

Since the reverse inequality is always satisfied, actually

$$\psi^{c_s}(\bar{\mathbf{y}}) = \psi(\bar{\mathbf{x}}) + \sum_{s \in S} c_s^*(\bar{\mathbf{x}}, \bar{\mathbf{y}}),$$

and this means precisely that $(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \in \partial_{c_s} \psi$. So Γ does lie in the c_s -sub-differential of ψ .

Step 4: There is duality.

Let $\|d_{s^*}\| := \sup_{(x, y) \in \mathcal{X} \times \mathcal{X}} c^*$. By steps two and three, there exists a transmission method π whose support is included in $\partial_{c_s} \psi$ for some c_s -convex ψ , and which was constructed “explicitly” in Step three. Let $\phi = \psi^{c_s}$.

From equation (12), $\psi = \sup \psi_m$, where each ψ_m is a supremum of continuous functions, and therefore lower semi-continuous. In particular, ψ is measurable. The same is true of ϕ .

Next we check that ψ, ϕ are bounded. Let $(\mathbf{x}_0, \mathbf{y}_0) \in \partial_{c_s} \psi$ be such that $\psi(\mathbf{x}_0) < +\infty$; then consequently $\phi(\mathbf{y}_0) > -\infty$. So, for any $\mathbf{x} \in \mathcal{X}$,

$$\begin{aligned} \psi(\mathbf{x}) &= \inf_{e \in \mathcal{E}} \sup_{\mathbf{y}} [\phi(\mathbf{y}) - c_{s^*}(\mathbf{x}, \mathbf{y})] \geq \phi(\mathbf{y}_0) - c^*(\mathbf{x}, \mathbf{y}_0) \geq \phi(\mathbf{y}_0) - \|d_{s^*}\|; \\ \phi(\mathbf{y}) &= \sup_{e \in \mathcal{E}} \inf_{\mathbf{x}} [\phi(\mathbf{x}) + c_{s^*}(\mathbf{x}, \mathbf{y})] \leq \psi(\mathbf{x}_0) + c^*(\mathbf{x}_0, \mathbf{y}) \leq \psi(\mathbf{x}_0) + \|d_{s^*}\|. \end{aligned}$$

Re-injecting these bounds into the identities $\psi = \phi^{c_s}$, $\phi = \psi^{c_s}$ and applying Proposition 5, we get

$$\begin{aligned}\psi(\mathbf{x}) &\leq \sup_{\mathbf{y}} \phi(\mathbf{y}) \leq \psi(\mathbf{x}_0) + \|d_s\|; \\ \phi(\mathbf{y}) &\geq \inf_{\mathbf{x}} \psi(\mathbf{x}) \geq \phi(\mathbf{y}_0) - \|d_s\|.\end{aligned}$$

So both ψ and ϕ are bounded from above and below.

Thus we can integrate ϕ , ψ against μ , ν respectively, and, by the marginal condition,

$$\int \phi(\mathbf{y}) d\nu(\mathbf{y}) - \int \psi(\mathbf{x}) d\mu(\mathbf{x}) = \int [\phi(\mathbf{y}) - \psi(\mathbf{x})] d\pi(\mathbf{x}, \mathbf{y}).$$

Since $\phi(\mathbf{y}) - \psi(\mathbf{x}) = \sum_s c_s^*(\mathbf{x}, \mathbf{y})$ on the support of π , we may write

$$\int \phi(\mathbf{y}) d\nu(\mathbf{y}) - \int \psi(\mathbf{x}) d\mu(\mathbf{x}) = \int \sum_s c_s^*(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}).$$

It was shown in remark that $\phi = \psi^{c_s}$, so we can write

$$\int \psi^{c_s}(\mathbf{y}) d\nu(\mathbf{y}) - \int \psi(\mathbf{x}) d\mu(\mathbf{x}) = \int \sum_s c_s^*(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}).$$

This proves the duality.

Step 5: If $d_s(\mathbf{x}, \mathbf{y}) \leq f_s(\mathbf{x}) + g_s(\mathbf{y})$ then (ψ, ϕ) solves the dual problem.

The idea in this step is to prove that ψ and ϕ are integrable. The estimates in this step are similar to that of step 4, the difference being that we fix $(\mathbf{x}_0, \mathbf{y}_0)$ such that $\phi(\mathbf{y}_0)$, $\psi(\mathbf{x}_0)$, $f_s(\mathbf{x}_0)$ and $g_s(\mathbf{y}_0)$ are finite, and write

$$\begin{aligned}\psi(\mathbf{x}) + f_{s^*}(\mathbf{x}) &= \inf_{e \in \mathcal{E}} \sup_{y \in \mathcal{X}} [\phi(\mathbf{y}) - c_{s^*}(\mathbf{x}, \mathbf{y}) + f_{s^*}(\mathbf{x})] \\ &\geq \sup_{y \in \mathcal{X}} [\phi(\mathbf{y}) - g_{s^*}(\mathbf{y})] \\ &\geq \phi(\mathbf{y}_0) - g_{s^*}(\mathbf{y}_0)\end{aligned}$$

and;

$$\begin{aligned}\psi(\mathbf{x}) - g_{s^*}(\mathbf{y}) &= \sup_{e \in \mathcal{E}} \inf_{x \in \mathcal{X}} [\psi(\mathbf{x}) + c^*(\mathbf{x}, \mathbf{y}) - g_{s^*}(\mathbf{y})] \\ &\geq \inf_{x \in \mathcal{X}} [\psi(\mathbf{x}) + f_{s^*}(\mathbf{x})] \\ &\geq \phi(\mathbf{x}_0) + f_{s^*}(\mathbf{x}_0).\end{aligned}$$

So ψ is bounded below by the μ -integrable function $\phi(\mathbf{y}_0) - g_{s^*}(\mathbf{y}_0) - f_{s^*}$ and ϕ is bounded above by the ν -integrable function $\psi(\mathbf{x}_0) + f_{s^*}(\mathbf{x}_0) + g_{s^*}$; hence both $\int \psi d\mu$ and $\int \phi d\nu$ make sense in $\mathbb{R} \cup \{-\infty\}$. Further, both integrals are finite since $\int (\phi - \psi) d\pi = \int c^* d\pi > -\infty$, and so

$$\int c^* d\pi = \int \phi d\nu - \int \psi d\mu.$$

Hence, as a result of step 4 we can conclude that both π and (ψ, ϕ) are optimal in the primal and dual problems, respectively.

To prove the last part of the theorem, first note that d_s is continuous, so the sub-differential of any c_s -convex function is a closed (S, c_s) -cyclically monotone set.

Let π be an arbitrary optimal transmission method, and (ψ, ϕ) and an optimal coding strategy. we know that (ψ, ψ^{c_s}) is optimal in the dual problem, so

$$\int c^* d\pi = \int \psi^{c_s} d\nu - \int \psi d\mu.$$

Now, by the marginal condition we may rewrite this as

$$\int \psi^{c_s} - \psi - c^* d\pi = 0.$$

We know the integrand is non-negative so π must be concentrated on the pairs $(\mathbf{x}, \mathbf{y})$ for which

$$\psi^{c_s}(\mathbf{y}) - \psi(\mathbf{x}) - c^*(\mathbf{x}, \mathbf{y}) = 0.$$

But this is the sub-differential of ψ , so since π and ψ are arbitrary, any optimal plan must be concentrated on the sub-differential of any optimal ψ . Thus, if Γ is defined as the intersection of all sub-differentials of optimal functions ψ , then Γ also contains the support of all optimal plans.

For the converse, consider an arbitrary transfer plan $\tilde{\pi} \in \Pi(\mu, \nu)$ concentrated on Γ , then

$$\begin{aligned} \int c^* d\tilde{\pi} &= \int [\psi^{c_s} - \psi] d\tilde{\pi} \\ &= \int \psi^{c_s} d\nu - \int \psi d\mu. \end{aligned}$$

So $\tilde{\pi}$ is optimal. Similarly, if $\tilde{\psi}$ is a c_s -convex function such that $\partial_{c_s} \tilde{\psi}$ contains Γ , then by the previous estimates $\tilde{\psi}$ and $\tilde{\psi}^{c_s}$ are integrable against μ and ν respectively, and

$$\begin{aligned} \int c^* d\pi &= \int [\tilde{\psi}^{c_s} - \tilde{\psi}] d\pi \\ &= \int \tilde{\psi}^{c_s} d\nu - \int \tilde{\psi} d\mu \end{aligned}$$

We may conclude that $\tilde{\psi}$ is optimal.

Proof (of corollary) By MacKay's duality theorem we have

$$\inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) = \sup_{\psi \in L^1(\mu)} \sup_{e \in \mathcal{E}} \left(\int_{\mathcal{X}} \psi^{c_s}(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \psi^{c_s} - \psi \leq \sum_{s \in S} c_s(\mathbf{x}, \mathbf{y}) \right).$$

The duality condition ensures that $\|\psi\|_F < \infty$ and from technical lemma B, if d_s is a metric, we may conclude that $\psi^{c_s} = \psi$. So the duality becomes

$$\inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) = \sup_{\psi \in F \setminus C} \sup_{e \in \mathcal{E}} \left(\int_{\mathcal{X}} \psi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \psi(\mathbf{y}) - \psi(\mathbf{x}) \leq \sum_{s \in S} c_s(\mathbf{x}, \mathbf{y}) \right).$$

By technical lemma A this can be written

$$\begin{aligned} \inf_{\pi \in \Pi(\mu, \nu)} \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}) &= \sup_{\psi \in F \setminus C} \sup_{e \in \mathcal{E}} \left(\int_{\mathcal{X}} \psi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \psi(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s(\mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y}) \in \tilde{\mathcal{X}} \right) \\ &= \sup_{\psi \in F \setminus C} \sup_{e \in \mathcal{E}} \left(\int_{\mathcal{X}} \psi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \frac{\psi(\mathbf{y}) - \psi(\mathbf{x})}{d_s(\mathbf{x}, \mathbf{y})} \leq e_s(\mathbf{x}, \mathbf{y}), (\mathbf{x}, \mathbf{y}) \in \tilde{\mathcal{X}} \right) \\ &= \sup_{\psi \in F \setminus C} \left(\int_{\mathcal{X}} \psi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x}); \|\psi\|_F \leq 1 \right) \\ &= \sup_{\psi \in F \setminus C} \left(\frac{\int_{\mathcal{X}} \psi(\mathbf{y}) d\nu(\mathbf{y}) - \int_{\mathcal{X}} \psi(\mathbf{x}) d\mu(\mathbf{x})}{\|\psi\|_F} \right) \end{aligned}$$

where $\tilde{\mathcal{X}} = \{(\mathbf{x}, \mathbf{y}) : x_s \neq y_s, x_t = y_t \ \forall t \neq s\}$. □

Appendix

Proposition 1 $\bar{d}(\mu, \nu)$ is a metric

Proposition 2 $\bar{d}(\mu, \nu)$ is complete.

Lemma 6 There exists a metric ρ on $\mathcal{P}(\mathcal{X})$ such that $\rho \leq \bar{d}$ and the topology induced from ρ is the weak* topology on $\mathcal{P}(\mathcal{X})$.

Proof We must prove symmetry (1), identity of indiscernible (2) and the triangle inequality (3).

1. Symmetry is obvious
2. If $\bar{d}(\mu, \nu) = 0$ then for all $s \in S$ we have $\int_{X_s \times X_s} d_s(\mathbf{x}, \mathbf{y}) d\pi^* = 0$ which implies that $\mathbf{x} = \mathbf{y}$ π^* -a.s so $\mathbb{P}^{\pi^*}(\mathbf{x} = \mathbf{y}) = 1$ which implies that $\mu = \nu$. Conversely if $\mu = \nu$ we can trivially couple them with the identity map, then $\mathbf{x} = \mathbf{y}$ π -a.s and hence $\bar{d}(\mu, \nu) = 0$.
3. Let $(\mathcal{X}_i, \mu_i)$, $i = 1, 2, 3$, be Polish probability spaces. Let X_1, X_2, X_3 and (Z_1, Z_2, Z_3) be as in Lemma ???. Choose $\text{Law}(X_1, X_2) = \pi_1$ such that

$$\sup_{s \in S} \int_{\mathcal{X}_1 \times \mathcal{X}_2} d_s(\mathbf{x}, \mathbf{y}) d\pi_1 \leq \bar{d}(\mu_1, \mu_2) + \epsilon$$

and likewise choose $\text{Law}(X_2, X_3) = \pi_2$ such that

$$\sup_{s \in S} \int_{\mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{x}, \mathbf{y}) d\pi_2 \leq \bar{d}(\mu_2, \mu_3) + \epsilon.$$

By Lemma ?? we can choose a measure $\tilde{\pi}$ on $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ whose projection onto the first two factors is π_1 and whose projection onto the second two factors is π_2 . Let π_3 be the projection of $\tilde{\pi}$ onto the first and third factors then π_3 is a (μ_1, μ_3) coupling and $\phi(\mathbf{y}) - \psi(\mathbf{x}) \leq c_s(\mathbf{x}, \mathbf{y})$

$$\begin{aligned} \int_{\mathcal{X}_1 \times \mathcal{X}_3} d_s(\mathbf{x}, \mathbf{z}) d\pi_3 &= \int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{x}, \mathbf{z}) d\tilde{\pi} \\ &\leq \int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{x}, \mathbf{y}) + d_s(\mathbf{y}, \mathbf{z}) d\tilde{\pi} \\ &\leq \int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{x}, \mathbf{y}) d\tilde{\pi} + \int_{\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{y}, \mathbf{z}) d\tilde{\pi} \\ &= \int_{\mathcal{X}_1 \times \mathcal{X}_2} d_s(\mathbf{x}, \mathbf{y}) d\pi_1 + \int_{\mathcal{X}_2 \times \mathcal{X}_3} d_s(\mathbf{y}, \mathbf{z}) d\pi_2 \end{aligned}$$

Let $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}_3$ then we have

$$\begin{aligned} \bar{d}(\mu_1, \mu_3) &= \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{z}) d\pi_3 \\ &\leq \sup_{s \in S} \left(\int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi_1 + \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{y}, \mathbf{z}) d\pi_2 \right) \\ &\leq \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi_1 + \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{y}, \mathbf{z}) d\pi_2 \\ &= \bar{d}(\mu_1, \mu_2) + \bar{d}(\mu_2, \mu_3) + 2\epsilon. \end{aligned}$$

Since ϵ can be taken to be arbitrarily small this gives the result.

Let $S = \{s_1, s_2, \dots\}$ and let (X_s, d_s) be compact and Polish. Let $\sigma(s_1, \dots, s_N)$ be the sub σ -field of $\mathcal{X}$ generated by the coordinates $\{s_1, \dots, s_N\}$. Now define ρ in the following way

$$\rho(\mu, \nu) = \sum_{i=1}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} |\mu(F) - \nu(F)|.$$

Where $\mathcal{A}(\sigma(s_1, \dots, s_i))$ is the collection of subsets corresponding to the sub σ -field $\sigma(s_1, \dots, s_i)$. By Tychonoff's theorem $\mathcal{X}$ is compact and every open covering of $\mathcal{X}$ has a finite subcover. This implies that the Borel σ -algebra $\sigma(\mathcal{X})$ is finite. Hence, the second sum is finite since the cardinality of $\mathcal{A}$ is finite. If $\sup_s |\sigma(X_s)| = K$ then the second term is bounded above by K^i . It's obvious that ρ is a metric, we prove now that ρ corresponds to the weak* topology. Assume that $\mu_n \xrightarrow{w^*} \mu$. Let $\epsilon > 0$ be given and choose M so that

$$\sum_{i=M+1}^{\infty} \frac{1}{i2^i} \leq \frac{\epsilon}{2}.$$

Next, choose N such that for all $n \geq N$, μ_n and μ agree to within $\frac{\epsilon}{2}$ on all sets in the σ -field generated by $\{s_1, \dots, s_M\}$. Then if $n \geq N$

$$\begin{aligned} \rho(\mu_n, \mu) &= \sum_{i=1}^M \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} |\mu_n(F) - \mu(F)| + \sum_{i=1+M}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} |\mu_n(F) - \mu(F)| \\ &\leq \sum_{i=1}^M \frac{1}{i2^i} \frac{\epsilon}{2} + \sum_{i=M+1}^{\infty} \frac{1}{i2^i} \\ &< \epsilon \end{aligned}$$

so $\mu_n \xrightarrow{\rho} \mu$.

Conversely, assume $\mu_n \xrightarrow{\rho} \mu$. Let F be an arbitrary element of the Borel σ -algebra. Then

$$F \in \mathcal{A}(\sigma(s_i, \dots, s_i))$$

for some i sufficiently large. Choose N sufficiently large so that for all $n \geq N$, $\rho(\mu_n, \mu) < \frac{\epsilon}{i(2K)^i}$. Then for all $n \geq N$,

$$\frac{|\mu_n(F) - \mu(F)|}{i(2K)^i} \leq \rho(\mu_n, \mu) \leq \frac{\epsilon}{i(2K)^i}$$

and so $|\mu(F) - \mu_n(F)| \leq \epsilon$ and $\mu_n \xrightarrow{w^*} \mu$. Hence the weak topology and the weak*-topology coincide. We need to show that $\rho \leq \bar{d}$. Let π be an arbitrary μ, ν coupling. It suffices to show that $\rho(\mu, \nu) \leq \sup_s \int d_s(\mathbf{x}, \mathbf{y}) d\pi$.

$$\begin{aligned}
\rho(\mu, \nu) &= \sum_{i=1}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} \left| \int_F f d\mu - \int_F f d\nu \right| \\
&= \sum_{i=1}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} \left| \int_{F \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi - \int_{\mathcal{X} \times F} d_s(\mathbf{x}, \mathbf{y}) d\pi \right| \\
&\leq \sum_{i=1}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} \int_{(F \times \mathcal{X}) \Delta (\mathcal{X} \times F)} d_s(\mathbf{x}, \mathbf{y}) d\pi \\
&\leq \sum_{i=1}^{\infty} \frac{1}{i(2K)^i} \sum_{F \in \mathcal{A}(\sigma(s_1, \dots, s_i))} \sum_{j=1}^i \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi \\
&\leq \sum_{i=1}^{\infty} \frac{1}{2^i} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi \\
&= \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi \\
&\leq \sup_{s \in S} \int_{\mathcal{X} \times \mathcal{X}} d_s(\mathbf{x}, \mathbf{y}) d\pi
\end{aligned}$$

as required.

Proof Let $\{\mu_n\}_{n \in \mathbb{N}}$ be a $\bar{d}$ -Cauchy sequence. By the previous lemma it is also ρ -Cauchy and hence converges weakly to some μ . We show that $\mu_n \xrightarrow{\bar{d}} \mu$. To do this consider a coupling of μ_n and μ_m , say $\pi_{n,m}$, such that

$$\sup_s \int d_s(\mathbf{x}, \mathbf{y}) d\pi_{n,m} \leq 2\bar{d}(\mu_n, \mu_m).$$

We can use a standard diagonalization argument and the weak* compactness of the collection of probability measures (since $\mathcal{X} \times \mathcal{X}$ is compact) we can choose $m_k \rightarrow \infty$ such that, for all n

$$\pi_{n,m_k} \xrightarrow{w^*} \pi_n \in \mathcal{P}(\mathcal{X} \times \mathcal{X}), \quad k \rightarrow \infty.$$

This is a coupling of μ_n, μ and we now want to show that these $\{\pi_n\}_{n \in \mathbb{N}}$ yield good μ_n, μ couplings. Let $\epsilon > 0$ and choose N such that $\forall n, m \geq N$, $\bar{d}(\mu_n, \mu_m) \leq \epsilon$. Now, if $n, m_k \geq N$ and $s \in S$,

$$\begin{aligned}
\int d_s(\mathbf{x}, \mathbf{y}) d\pi_{n,m_k} &\leq \sup_s \int d_s(\mathbf{x}, \mathbf{y}) d\pi_{n,m_k} \\
&\leq 2\bar{d}(\mu_n, \mu_{m_k}) \\
&< 2\epsilon.
\end{aligned}$$

Letting $k \rightarrow \infty$ together with $\pi_{n,m_k} \xrightarrow{w^*} \pi_n$ and compactness of $\mathcal{X} \times \mathcal{X}$ we get that

$$\sup_s \int d_s(\mathbf{x}, \mathbf{y}) d\pi_n \leq 2\epsilon$$

and we can conclude that $\bar{d}(\mu_n, \mu) \leq 2\epsilon$ if $n \geq N$ and so $\mu_n \xrightarrow{\bar{d}} \mu$ as required.

References

1. Dobrushin, R. "Survey of Soviet Research in Information Theory." *IEEE Transactions on Information Theory* 18, no. 6 (1972): 703–24. <https://doi.org/10.1109/tit.1972.1054923>.
2. Föllmer, Hans, and Ulrich Horst. "Convergence of Locally and Globally Interacting Markov Chains." *Stochastic Processes and their Applications* 96, no. 1 (2001): 99–121. [https://doi.org/10.1016/s0304-4149\(01\)00110-7](https://doi.org/10.1016/s0304-4149(01)00110-7).
3. Liggett, T.M. *Interacting Particle Systems*. Berlin: Springer, 2005.
4. Mackay, R. S. "Management of Complex Dynamical Systems." *Nonlinearity* 31, no. 2 (2018). <https://doi.org/10.1088/1361-6544/aa952d>.
5. Mackay, R. S. "Robustness of Markov Processes on Large Networks." *Journal of Difference Equations and Applications* 17, no. 8 (2011): 1155–67. <https://doi.org/10.1080/10236190902976889>.
6. Steif, Jeffrey E. "The Ergodic Structure of Interacting Particle Systems," 1988.
7. Steif, Jeffrey E. " $\bar{d}$ -Convergence to Equilibrium and Space—Time Bernoullicity for Spin Systems in the $M < \epsilon$ Case." *Ergodic Theory and Dynamical Systems* 11, no. 3 (1991): 547–75. <https://doi.org/10.1017/s0143385700006337>.
8. Villani Cédric. "Chapters 1-6." In *Optimal Transport: Old and New*, 1–113. Berlin: Springer, 2009.